Winter 1997

The Effects of Human-Computer Communication Mode, Task Complexity, and Desire for Control on Performance and Discourse Organization in an Adaptive Task

Cristina Bubb-Lewis
Old Dominion University

Recommended Citation

Bubb-Lewis, Cristina. "The Effects of Human-Computer Communication Mode, Task Complexity, and Desire for Control on Performance and Discourse Organization in an Adaptive Task" (1997). Doctor of Philosophy (PhD), dissertation, Psychology, Old Dominion University, DOI: 10.25777/k248-4361
https://digitalcommons.odu.edu/psychology_etds/252

THE EFFECTS OF HUMAN-COMPUTER COMMUNICATION MODE, TASK COMPLEXITY, AND DESIRE FOR CONTROL ON PERFORMANCE AND DISCOURSE ORGANIZATION IN AN ADAPTIVE TASK

by

Cristina Bubb-Lewis
B.S., May 1991, Duke University
M.S., December 1993, Old Dominion University

A Dissertation submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 1997

Approved by:

Mark W. Scerbo (Director)

Glynn D. Coates (Member)

James R. Comstock, Jr. (Member)

Raymond H. Kirby (Member)

Danielle S. McNamara (Member)

ABSTRACT

THE EFFECTS OF HUMAN-COMPUTER COMMUNICATION MODE, TASK COMPLEXITY, AND DESIRE FOR CONTROL ON PERFORMANCE AND DISCOURSE ORGANIZATION IN AN ADAPTIVE TASK

Cristina Bubb-Lewis
Old Dominion University, 1997
Director: Dr. Mark W. Scerbo

The present study examined how different communication patterns affected task performance with an adaptive interface. A Wizard-of-Oz simulation (Gould, Conti, & Hovanyecz, 1983) was used to create the impression of a talking and listening computer that acted as a teammate to help participants interact with a computer application.

Four levels of communication mode were used which differed in the level of restriction placed on human-computer communication. In addition, participants completed two sets of tasks (simple and complex). Further, a personality trait, Desire for Control (DC), was measured and participants were split into high and low groups for analysis. Dependent measures included number of tasks completed in a given time period as well as subjective ratings of the interaction. In addition, participants' utterances were assessed for verbosity, disfluencies, and indices of common ground.

The largest performance differences were found between the groups that could communicate freely and those where communication was restricted or denied. As the level of restriction increased, performance decreased. Further, as communication restriction increased, the computer assumed greater control and levels of verbosity decreased. Performance on the simple tasks declined as communication restriction increased, but no differences were observed among communication modes for complex tasks. There were no performance effects due to DC, however high-DC participants rated their ability to communicate as easier than low-DC participants. The results of the present study are discussed with respect to differences between human-human and human-computer communication as well as research on adaptive environments.

This is dedicated to my grandparents, Helen and John Macomber and Nettie and Harry Bubb, who waited patiently for great-grandchildren and got this instead. I love you and thank you for your support throughout the years.

ACKNOWLEDGMENTS

Never ending thanks to my dissertation director, Dr. Mark W. Scerbo, for your time, liberal use of many red pens, and for generously mentoring a sometimes neurotic graduate student. I will always appreciate your contributions to my education.

Thanks also to my committee members: Dr. Glynn D. Coates, Dr. James R. Comstock, Jr., Dr. Raymond H. Kirby, and Dr. Danielle S. McNamara. Your input improved the research immeasurably, and helped to make this an invaluable learning experience.

Thank you to Jamie LoVerde, my angelic research assistant, you know I couldn't have done it without you. My gratitude to Julie MacMurdo, the voice of the computer, I know how much you love to have your voice taped.

To my husband, Mel Lewis, what can I say? I could not have done this without your periodic boosts to my wavering self-esteem and your unconditional love and support. And yes, I will get a job now!

Momma, Opie, Uncle Geof, and Grandpa. You have always supported me in everything I have chosen to do. These past two years have been hard for us all, but we have never forgotten the importance of family. I love you all very much and only hope to make you proud.

Gramma, you know how much I wanted you to see this day. You were always my best cheerleader. I hope you are watching and smiling down on us. Could you talk the Big Guy into another National Championship for Duke? I love you.

TABLE OF CONTENTS

PAGE

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Adaptive Automation

Adaptive automation refers to dynamic systems which adjust their methods of operation in response to changes in situational demands (Gluckman, Morrison, & Deaton, 1991; Rouse 1988). In an adaptive automation system, the human and the machine must work together as partners in order to maintain optimal operation of the system (Scerbo, 1994). The idea is that as operator workload increases the system can take over some tasks, and when workload demands are reduced, tasks are returned to the operator in order to maintain optimal situation awareness (Rouse, 1988). For example, fighter pilots can sometimes sustain G-forces which will render them unconscious for periods of up to 12 seconds (Buick, 1989; Whinnery, 1989). In this kind of situation it would be beneficial for a computer to take over and stabilize the plane until the pilot can resume control. Since adaptive automation is still in its early stages, researchers and designers have the opportunity to consider how the technology might be successfully implemented, before it is fully developed.

Hammer and Small (1995) worked on the design and implementation of the

---

The Publication Manual of the American Psychological Association (4th ed.) was used in the preparation of this manuscript.

Pilot's Associate (PA), an adaptive decision aiding program for tactical aircraft. The PA was designed to help pilots cope with the increased complexity and inherent difficulty of tactical air combat. Because many of the difficulties of operating the aircraft are related to problems with the interface, a goal of the program was to utilize the full capabilities of the aircraft while also simplifying the interface.

The PA (Hammer & Small, 1995) used intelligent adaptive automation to overcome pilot limitations and enhance pilot abilities. The system was not meant to simply take tasks away from the human, but rather to share the responsibility of flying the plane so that both human and computer abilities were used to full advantage. In addition, the PA was designed to keep the pilot aware of the flight situation by filtering large amounts of data and generating and displaying the right information in the appropriate form at the right time. The aircraft avionics provided data to the assessors, which produced descriptions for the planners and intelligent interface. The intelligent interface might then execute a task on behalf of the pilot or instruct the display generator to produce displays for the pilot. The intelligent interface also monitored pilot error, determined pilot intentions, and recommended responses to the pilot. The pilot read the displays and issued commands to the aircraft and the display system.

The PA (Hammer & Small, 1995) was a mixed-initiative system. It could perform actions on behalf of the pilot in overload conditions or allocate tasks of low importance to automation. As Hammer and Small point out, the PA is, "more like an electronic crew member than conventional automation," resulting in "a demand for new types of knowledge in the design of the interaction between intelligent automation (associate systems) and human operators of complex systems (p. 3)." Hammer and Small see the potential for adaptive automation in a large number of areas where complex systems are used (e.g., aerospace systems, weapon systems, control systems, process control, manufacturing, design, and medical technology). The capabilities of a fully developed "electronic crew member" would have a great impact on the control of complex systems in areas such as error reduction, enhanced human-computer communication, and less complexity. In fact, Hammer and Small see a day when the behavior of adaptive systems, "will be indistinguishable from that of another human crew member (p. 42)."

Other adaptive systems are currently being designed in areas such as supervisory control, intelligent tutoring, and on-line documentation (Bushman, Mitchell, Jones, & Rubin, 1993; Chu, Mitchell, & Jones, 1995; Mason, 1986). For instance, Mason (1986) describes a technique called adaptive command prompting and its application to an enhanced version of the UNIX on-line manual. The system automatically adjusts a set of prompts in order to suit the individual user. They found that the adaptive capabilities of the system were not intrusive to the user and did not appear to change on an arbitrary basis. However, they point out the importance of considering these issues in the design of adaptive systems. They predict that users may have trouble with more complex systems if the adaptive behavior is not easily understood.

Bushman et al. (1993) described the design, implementation, and evaluation of ALLY, an operator's associate for cooperative supervisory control of a simulated satellite ground control system. ALLY used intent inferencing (representations of operator plans based on operator actions) in order to function as an assistant to the human operator of the system, and used the metaphor of human-human cooperation to develop the human-computer interaction. ALLY actively monitored the system and made recommendations and initiated troubleshooting when appropriate. The operator had the ability to decide how much responsibility to delegate to ALLY. In an empirical analysis of the system, human-ALLY teams performed comparably to human-human teams. Bushman et al. (1993) conclude that ALLY provides strong support for the effective functioning of a computer-based associate in a supervisory control team. They point out the need for a more refined theory of human-machine cooperation to guide the development of future systems.

Chu et al. (1995) also used intent inferencing as the basis for an intelligent tutoring system which was meant to act as both a tutor for novices and an aid for expert operators of supervisory control systems. This system was used to train operators in a simulator environment where operational skills including rare and catastrophic system conditions can be practiced. In addition, the system allowed the operators to form relationships with their computer partner over the training period which then carried over into the actual control setting.

These examples highlight the growing importance of adaptive automation technology. As Bushman et al. (1993) pointed out, there is a need for greater understanding of human-machine cooperation in order to enhance the usability of these systems. Many adaptive automation systems, such as the PA (Hammer & Small, 1995), demonstrate the use of current technology in aiding humans; they do not, however, investigate the best way to implement adaptive technology.

In addition, although Hammer and Small (1995) envision a time when adaptive technology will be indistinguishable from a human partner, the predominant strategy used in the development of adaptive systems to date has been to put the human in charge with the computer acting as a subordinate. The human decides when the computer can intervene, in what areas, and for how long. This does not truly reflect team interaction processes. If the human and the computer are to be true teammates responsibility will have to be shared. This is not to say that the human will no longer be in control, but that computer behavior will not always be limited to a checklist of behaviors filled out by the human.

The Human-Machine Team

Some researchers have recently begun to look at adaptive automation from a team perspective (Hammer & Small, 1995; Malin & Schreckenghost, 1992;

Malin, Schreckenghost, Woods, Potter, Johannesen, Holloway, & Forbus, 1991; Scerbo, 1994). Malin and Schreckenghost (1992) suggest that an intelligent system must meet four criteria in order to be considered a team member. First, the system must be reliable and modifiable. Second, the system must communicate effectively with other team members. Third, the system must coordinate activities with other team members. Fourth, teams must be coached, meaning that members are responsible for the behavior of other team members as well as their own. In order for these criteria to be met, the system and other team members must be able to exchange information freely, and team members must be aware of the capabilities and limitations of the system (Scerbo, 1996). Scerbo (1994) has suggested that an understanding of team dynamics should guide the development of adaptive automation technology, and he has identified analogs for many team functions in adaptive automation technology. This paper will concentrate on Malin and Schreckenghost's (1992) second criterion for a computer team member, effective communication with team members.

Communication

The exchange of information is essential to an efficiently functioning team (Fleishman & Zaccaro, 1992; Salas, Dickinson, Converse, & Tannenbaum, 1992; Scerbo, 1996). However, in human teams this flow of information is often less than perfect. For example, a recent survey of pilots indicated that over half of all pilot errors result from failures of information transfer (Nagel, 1988).

Communication problems occur in human-machine systems as well. Wiener (1989) identifies the three most commonly asked questions on the highly automated flight deck as, “What is it doing?”, “Why is it doing that?”, and “What will it do next?” Sarter and Woods (1995) add, “How in the world did we get into that mode?” to the list. Therefore, from a team perspective it is essential that we understand the issues associated with communication and information exchange and how they will apply to adaptive automation technology. Scerbo (1996) suggests that the success of adaptive automation will depend largely on the methods of information exchange that are available to the human-machine team, that is, the interface.

When humans communicate with each other they can use spoken language (which includes not only words, but also tone of voice) or written language, they can draw pictures, they can use nonverbal information such as body movements and facial expressions, and they can even use physical contact. Scerbo (1996) points out that because humans make use of all of these methods when communicating with each other, an adaptive system which uses only one method of information exchange (for example, an alphanumeric interface) will severely limit the quality of communication between the human and the system, and thus limit the ability of the team to work effectively. This highlights the importance of research on the effects of communication mode on human-computer interaction. It is hoped that the current study will provide useful information for implementing successful human-computer communication in adaptive automation.

Currently there are no adaptive systems which could be said to communicate with humans on a human teammate level. Communication between humans and machines is still very rudimentary. However, in some cases it is possible and beneficial to study the human factors requirements of technology before the technology itself is fully developed. These studies can guide the development of technology from a human usability perspective instead of addressing these issues after the fact. This has been done in the past using a Pay No Attention to the Man Behind the Curtain (PNAMBiC) or Wizard-of-Oz method (Brennan, 1991; Gould, Conti, & Hovanyecz, 1983; Guindon, Shuldberg, & Connor, 1987; Newell, Arnott, Carter, & Cruickshank, 1990; Newell, Arnott, Dye, & Cairns, 1991).

Gould and his colleagues (1983) were pioneers of this method and their efforts will be described briefly here to illustrate the merit of this paradigm. Gould et al. (1983) wanted to study the usefulness of a listening typewriter (a typewriter that would change speech input into a textual format) at a time when speech recognition was not yet a viable technology. They accomplished this using a microphone which transmitted the subject's voice to a skilled typist who then typed what the participant said according to certain rules which would simulate either a limited (1000 or 5000 words) or unlimited vocabulary. The typed information was then displayed on a screen in front of the participant. The simulation was so convincing that some participants refused to believe that they were interacting with another human even after they were introduced to the typist. The results suggested that some versions of the listening typewriter could be as good as traditional methods of handwriting and dictating, and provided useful information for the future implementation of the technology.

The success of Gould et al. (1983) and the other investigators cited above led to the decision to use a PNAMBiC adaptive interface in the present study. Although the interfaces for this study could not be built with current technology, they can be simulated using a PNAMBiC method. Using this method will result in information which may affect the way this technology is implemented once it becomes technically feasible.

Communication Modes

The study of human communication in various modes (i.e., communication using varied input and output channels) began as an investigation of the effects of new developments in telecommunications (e.g., the telephone, teleconferencing, and electronic mail). The researchers believed that an understanding of human communication would be essential for the development of truly interactive technology.

Chapanis and his associates performed a number of studies comparing different modes of communication (Chapanis & Overbey, 1974; Chapanis, Ochsman, Parrish, & Weeks, 1972; Chapanis, Parrish, Ochsman, & Weeks, 1977;

Krueger & Chapanis, 1980; Ochsman & Chapanis, 1974; Weeks & Chapanis, 1976). A typical protocol in the series involved two-person teams solving problems by one of four communication modes: (a) handwriting, (b) typewriting, (c) voice, and (d) face-to-face. The problems were "real-world" problems for which computer assistance could be useful such as geographic orientation problems or equipment assembly problems. The problems required more than one person to solve. Performance was assessed using three dependent variables: (a) time to solution, (b) behavioral measures of activity, and (c) linguistic measures. Large differences were found between the nonvoice and voice modes in all three classes of dependent variables and Chapanis et al. (1977) reported several conclusions from their series of studies:

1. Problems requiring the exchange of factual information can be solved twice as fast in voice modes than in nonvoice modes.

2. When using voice modes participants are better able to engage in multiple activities. This is very difficult in nonvoice modes where typing is required.

3. Only about one third of the time spent solving these problems was used for communicating. Searching for information was the predominant behavior in most modes of communication.

4. Natural human communication is apparently unruly. It is full of errors and irregularities which makes it difficult to measure objectively. If human-

computer interaction is ever going to approach human communication, computers will have to cope with these irregularities.

5. Although natural human communication appears to follow no standard rules, the fact that we can solve difficult problems so efficiently shows that it must.

6. Voice modes of communication are fast, but they are also wordy. There is a lot of redundancy built into the communication.

7. There were no practical differences in the efficiency of voice only and face-to-face modes for the problems tested and variables measured.

8. Participants in face-to-face conditions spoke more than participants in voice only modes.

Two studies from this series are particularly important to the present experiment and will be discussed in more detail. Chapanis and Overbey's (1974) experiment compared free and restricted interrupt options in the voice only mode. Participants in the free interchange condition could interrupt each other at any time while participants in the restricted interchange condition were prevented from transmitting a message until the person in control of the channel voluntarily gave up that control. This condition had no effect on the time taken to solve the problems, the total number of words exchanged, or the rate at which words were exchanged. However, the interruption manipulation did impact how communicators "packaged" their messages. When there was freedom to interrupt more messages were exchanged, messages were shorter and were exchanged faster. This ability to exchange information freely may be important in solving complex or time-constrained problems.

Ochsman and Chapanis (1974) studied a more extensive set of communication modes designed to provide a hierarchy of communication richness from a mode in which participants could only use typing to a mode that approached face-to-face communication. They used five communication channels in various combinations (typing, handwriting, voice, video without voice, and visual contact through a glass panel) to produce 10 communication modes. The protocol was similar to the one described earlier.

The results showed that the largest difference in the modes of communication was between those that had a voice channel and those that did not. The typing and writing modes did not approach the speed or efficiency of voice modes, suggesting that speech will be necessary for effective communication between a human and a machine where complex problems must be solved under time pressure. In addition, there was no evidence that the addition of a video channel had any significant effect on communication behavior or times.

There was evidence that communication times were inversely related to the richness of communication modes. Overall, communication times decreased as the number and quality of communication channels increased. This supports Scerbo's (1996) suggestion that the success of adaptive automation will be affected by the methods of information exchange available to the human-machine team.

Other research involving communication modes has been conducted in the area of computer-mediated communication. O'Conaill, Whittaker, and Wilbur (1993) examined how the spoken aspects of video-mediated communication differ from face-to-face interaction. Using a series of real meetings they evaluated two wide-area conferencing systems. One was an ISDN system that had transmission lags, a half-duplex audio line, and poor quality video, while the other was a broadcast system with negligible delays, full duplex audio, and broadcast quality video. Hypotheses were generated by comparing the channel properties of the conferencing systems with those of face-to-face communication (i.e., low transmission delays, two-way, multiple modalities). As predicted, communication using the ISDN system had longer conversational turns, fewer interruptions, less overlaps (simultaneous speaking), less backchannel feedback, and increased formality when switching speakers. Communication using the broadcast system was similar to, but did not replicate, face-to-face communication. Formal techniques were still used to achieve speaker switching and the authors suggest that these may have been necessary because of the absence of certain speaker-switching cues (e.g., directional sound, unrestricted vision). They conclude that certain basic communication processes are disrupted by the channel properties of mediated communication systems, and that these disruptions result in differences from face-to-face communication. This study points to some of the difficulties that may be encountered when a human is communicating with a computer. The disruptions caused by mediated communication may result in differences in communication ability.

The research on communication modes has been successful in highlighting the large differences between modes that have a voice channel and those that do not. The presence or absence of a voice channel leads to differences in solution time, participant behavior, and verbal output. Solution times are faster, information transfer is quicker, and more information is exchanged in modes with a voice channel. Although these differences did not seem to hamper performance in the studies described above, the problems used were relatively simple. The problems required that partners exchange factual information, but not necessarily work together as a team. More complex interactions might benefit more from using a voice mode.

The freedom to interrupt affected communication in Chapanis and Overbey's (1974) study. Again, the importance of information exchange is highlighted. There is also evidence that the richness of the communication mode affects the communication process (Ochsman & Chapanis, 1974), and that this process can be affected by disruptions caused by mediated communication (O'Conaill et al., 1993).

The studies on communication mode illustrate the importance of understanding the limitations that occur in human communication with a computer. There will undoubtedly be differences in the richness of the communication process which may result in differences in communication ability and the need to minimize any undesirable results of those differences. In addition, the research on voice versus nonvoice modes has contributed to the decision to use a spoken interface in the present study. It is believed that adaptive automation systems of a complex nature will require a voice interface in order to successfully exchange information.

Feedback

Visual and verbal feedback are very important elements in the coordination of conversation. During conversations listeners provide concurrent feedback in the forms of auditory backchannels (e.g., “uhuh”, “yeah”) and visual feedback (e.g., headnod, smile). When this feedback is absent or delayed the speaker’s ability to communicate efficiently is reduced (Krauss & Bricker, 1967; Krauss & Fussell, 1990). The speaker cannot determine if a message has been understood and might reiterate points unnecessarily to ensure understanding, thus resulting in longer communications (Krauss & Bricker, 1967; Oviatt & Cohen, 1991). Visual and verbal feedback are also used to regulate conversational turns (O’Conaill et al., 1993; Sacks, Schegloff, & Jefferson, 1974). Feedback may also affect human-computer interaction.

Multiple nonverbal cues such as gaze, facial expression, posture, and physical proximity often accompany verbal messages. Research has shown that these cues may help the listener to identify the meaning of the message (Argyle, Lalljee, & Cook, 1968; Jaffe & Feldstein, 1970), support smooth speaker transitions (Rutter & Stephenson, 1977), and offer the speaker information about the effects her speech is having on the listener (Short, Williams, & Christie, 1976).

Kiesler and Sproull (1992) compared face-to-face meetings with real-time computer-mediated discussions. The groups were asked to reach consensus on several decision tasks. They found that real-time computer conference decisions took four times as long as face-to-face decisions. They speculated that one of the causes of these delays was lack of nonverbal backchannel feedback. Kiesler and Sproull (1992) asserted that diminished nonverbal backchannel feedback led to more difficulty establishing a mutual understanding of the problem and thus increased time to solve the problem.

Krauss and Weinheimer (1966) found that progressive noun phrase reduction (when an object is referred to repeatedly during a task, the referring noun phrase will become shorter) was influenced by the presence or absence of concurrent feedback from the listener. They concluded that backchannel feedback plays an important role in helping the speakers to converge on a reduced noun phrase. In addition, other studies have shown that during typical interactive dialogues, confirmations are used for reducing the descriptive detail needed between speakers, thus increasing communication efficiency (Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987).

The effects of listener responsiveness on conversational effectiveness were examined by Kraut, Lewis, and Swezey (1982). They had speakers summarize the plot of a movie to one or two listeners. The results showed that as speakers received more feedback from a partner, listeners (active and eavesdroppers) understood their summaries better. In addition, active listeners' summaries were better than eavesdroppers', suggesting that feedback helped to tailor the conversation to the individual. Kraut et al. (1982) point out that feedback influenced conversational process and outcome even in this constrained laboratory setting. The influences may be much stronger in more interactive, natural communication.

Johannesen, Cook, and Woods (1994) conducted a field study with anesthesiologists in order to examine common ground in dynamic fault management applications. Common ground refers to the set of mutual beliefs and knowledge developed and updated during a conversation (Clark & Schaefer, 1989). The grounding process is affected by factors such as the medium and purposes of communication (Clark & Brennan, 1991), and is essential for understanding how team members work efficiently in evolving situations.

Johannesen et al. (1994) identified several methods used to maintain situation awareness as conditions changed. Updates occur when a team member returns to the situation and must be informed of what has happened during his absence. He is given the necessary information to reestablish common ground.

Team members also provide spontaneous (unrequested) reports of their activities and assessments to keep all members aware of what is happening. In addition, when team members notice something that does not fit with their expectations, a dialogue will ensue that serves to realign a common ground that may have been diverging. It was also observed that in cases when the information provider did not know the questioner's goals, uninterpreted information was provided instead of interpretations of that data (e.g., "120 over 80" instead of "blood pressure normal"). This allowed the questioner to form his own interpretation according to his goals. The exchanges between team members were very brief and they used domain specific language. Johannesen et al. (1994) believe that this is partly due to shared domain knowledge, and partly due to mutual knowledge about the history of the process, and about goals and expectations. This mutual knowledge serves as context for the communication. Johannesen et al. (1994) point to the need for intelligent systems to establish and maintain common ground with human partners through cooperative exchanges that occur within a common frame of reference. All of the methods of maintaining common ground identified by Johannesen et al. (1994) might be useful for human-computer interaction.

In a study of human advisory interactions between computer system help desk consultants and system users, Aaronson and Carroll (1987) found that advice was frequently modified in response to verification requests. The interactions were more like negotiations where the two participants would trade knowledge back and forth and come to a mutual understanding of the problem. Aaronson and Carroll (1987) also found that verification requests were used more often by experienced users than less sophisticated users. They suggest that designers might exploit this tendency by supporting the verification strategy in intelligent help systems.

Oviatt and Cohen (1991) studied how limitations on speaker interaction influence spoken discourse patterns. The purpose of the study was to analyze the differences in discourse organization, referential characteristics, and performance efficiency for dialogues and monologues during a task-oriented exchange, and to examine the implications for the development of future speech systems. Dialogues and monologues were used because they represent opposite ends on the spectrum of speaker interaction (interactive and noninteractive). Experts were asked to provide spontaneous instructions either by telephone (dialogue) or audiotape (monologue) to help a novice partner assemble a water pump.

The interactive telephone dialogues had a unique discourse structure with many clarification subdialogues between the expert and novice. The telephone dialogues also had a distinct confirmation structure. Listeners regularly confirmed that instructions had been received and understood with an average rate of one confirmation every five to six seconds.

The organization of the noninteractive audiotape monologues differed in several ways from the interactive telephone dialogues. It was theorized that these differences were attempts on the part of the experts to compensate for the lack of interactive feedback by relying more on organizational strategies to clarify their instructions. Audiotape experts made significantly more explicit introductions of upcoming actions before they began relaying instructions. Although summaries were common in both modalities, they occurred significantly more often in the noninteractive monologues. In addition, audiotape experts often made parallel introductions and summaries of small sections of assembling the water pump, perhaps to provide structural bracketing of a group of steps.

Telephone and audiotape experts also differed in their descriptions of the water pump pieces and what to do with them, which made up the bulk of the task instructions. It was hypothesized that audiotape experts would provide more extensive descriptions because step-by-step confirmations were not available. The results showed that the audiotape experts did produce significantly more and longer spontaneous elaborative descriptions of parts and actions. The hypothesis that audiotape speakers would continue to elaborate their descriptions because they could not receive feedback was also supported. In other words, audiotape experts kept describing the piece to be assembled even after they had given the assembly instructions for it.

Spontaneous phrase and sentence repetitions were significantly more common in the audiotape mode. These repetitions tended to occur during difficult assembly segments and may have been due to the experts' inability to receive feedback about whether the novice understood the segment. Audiotape and telephone experts also introduced piece descriptions differently. Telephone experts tended to decompose instructions into two parts: identifications and actions. In contrast, audiotape experts often first referred to a piece by telling the novice to act on it in some way. Telephone novices took a significantly shorter time to assemble the pump than audiotape novices, but the task appeared to be relatively easy with most teams completing it in less than 10 minutes. Oviatt and Cohen (1991) attempted to uncover the discourse factors that correlated with assembly time. In both modes, elaborated descriptions, frequent use of personal pronouns, and advance introductions of upcoming actions were positively correlated with assembly time. This highlights the relative inefficiency of excessive elaboration (and therefore the audiotape mode) as a discourse strategy.

During the assembly task, all telephone teams engaged in frequent confirmations. This continual confirmation is the primary method for the listener to signal to her partner that the partner's communicative goals have been achieved in a task-oriented dialogue. In addition, access to concurrent feedback has been linked to increased dialogue efficiency in the form of reduced noun phrases with repeated reference (Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987; Krauss & Weinheimer, 1964 (as cited in Oviatt & Cohen, 1991), 1966). Because audiotape experts did not have access to confirmatory feedback their extensive elaboration was a conservative strategy which, while ultimately successful, sacrificed efficiency.

Oviatt and Cohen (1991) discussed the results of their study with respect to the design of interactive speech systems. They point out that although the goal for spoken language systems is the development of fully interactive speech, the current capabilities of these systems could be considered interactive in only a very limited sense. Therefore, it is important to consider the effects that limited interactivity will have on communication between the machine and the user.

For example, system delays are currently longer than those encountered in human communication. Experimental research on telephone conversations has shown that transmission and access delays as small as .25 to 1.8 seconds tend to disrupt the normal pattern of conversation and reduce referential efficiency (Krauss & Bricker, 1967). In addition, research on human-computer dialogue (VanKatwijk, VanNes, Bunt, Muller, & Leopold, 1979) has shown that language systems that have delays in processing can result in user input that has characteristics of noninteractive speech. This research supports further the importance of confirmation feedback in promoting optimal conversational efficiency, and highlights the importance of finding ways to compensate for disruptions in the normal feedback channels of communication.

Another area where current interactive speech systems are limited is prosodic analysis (e.g., intonation, pauses). For example, in order to analyze and respond to a request for confirmation, a system might have to detect rising intonation, pauses, and other characteristics of the speech signal (Pierrehumbert, 1983; Waibel, 1988). Because current systems cannot reliably perform these types of analyses, supplying appropriate and properly timed confirmations will be difficult. Similar to the effects of transmission delays, this lack of prosodic analysis may lead to some of the characteristics of noninteractive speech.

Oviatt and Cohen (1991) also point out that there is no well developed model of human-machine communication to use in designing human-machine systems. Further research is needed on the extent to which human-computer speech differs from that between humans.

The studies discussed above highlight the importance of feedback in the communication process. Feedback availability affects discourse structure as well as communication efficiency and will be an important issue for human-computer communication. In addition, the lack of research on feedback availability in human-computer interaction has contributed to the decision to study different levels of human-computer spoken interaction in the present study. It is hypothesized that the different modes of interaction used in the present study will affect the ability of the human to receive feedback from the computer and therefore affect discourse structure and performance.

Human-Computer Communication

To date, the few studies comparing human-human and human-machine communication during task completion have used keyboard input (Brennan, 1991;

Guindon et al., 1987; Kennedy, Wilkes, Elder, & Murray, 1988), as have those experiments that explored only human-machine communication (Carroll & Aaronson, 1988; Chin, 1984; Malhotra & Sheridan, 1976).

Guindon et al. (1987) found that keyboard users of a limited interaction system frequently produced complex noun phrases. These phrases were similar to the elaborative noun phrases found in noninteractive speech by Oviatt and Cohen (1991). These expressions may have been used to emphasize referential precision because feedback was limited and users were not sure about the degree of common ground with their partner. Guindon et al. (1987) also showed that users request help with simple, restricted English that resembles informal speech except for referring expressions, completeness, and formality which were more like formal written language (complex referring expressions, no sociability, and few fragments). The authors suggested that users believe there is poor shared context with the machine and that the system cannot handle fragmentary language. They concluded that unrestricted natural language is not necessary for efficient advisory systems.

Other studies using the keyboard modality have also pointed to the possibility of developing successful limited natural language systems (Chin, 1984; Malhotra & Sheridan, 1976). Malhotra and Sheridan (1976) used a simulation of an order-writing and invoicing system to study the requirements for natural language capabilities. They were able to classify over half of users' statements using a small set of structural templates. However, more than a third of the utterances were classified as not analyzable syntactically. Chin (1984) found that over a quarter of English queries to a simulated advisor used contextual constructs (e.g., ellipses, anaphor, fragments), but participants querying a human used nearly twice as many contextual constructs. This suggests that users may be able to voluntarily restrict the complexity of their queries when interacting with an advisory system.

Carroll and Aaronson (1988) simulated an active help system, whereby help is provided without requiring a request from the user, for a database program. When users made a mistake a help message was displayed on the screen. They found that users sometimes expected the help system to know their intentions and that providing intelligent help could be an asset as well as a hindrance. They acknowledge that even human advisors are less than perfect and suggest that the real problem is how to implement a less than perfect computer advisor. Carroll and McKendree (1987) have also pointed to the need to understand human advisory strategies as well as restricted natural language capabilities in order to allow for empirical selection of implementation strategies.

Kennedy et al. (1988) report on three experiments in which participants carried on typed dialogues with what they believed to be either a computer system or another person. The transcripts were analyzed in order to examine the use of anaphor (a reference which points back to elements already mentioned or implied during the conversation) and lexical choice (word choice). Anaphor is an important index of common ground because it allows participants to communicate without the continual reintroduction of topics. Lexical choice is an indicator of what the speaker believes the audience can understand. When participants had a computer partner (real or simulated by a human) the resulting dialogues were composed of focused content, shorter utterances, less lexical choice, and less use of pronominal anaphor (e.g., they, she, it). These dialogue characteristics were persistent over lengthy periods of interaction even when there was no evidence to support the need for limited interaction (i.e., the computer understood and responded to everything the participant typed).

Brennan (1991) performed a similar experiment which varied type of partner (human or simulated computer) and the style of responses (short, sentence, lexical change) in a database query task. She observed fewer acknowledgments in human-computer dialogue. In contrast to Kennedy et al. (1988), Brennan (1991) found that the style of the partner's response shaped the form of the subject's subsequent queries (i.e., short responses led to short queries, full sentence responses led to sentence queries) and that the subject's expected "connectedness" across conversational turns (as shown by the use of anaphor by specifically referring to something once and subsequently using a pronoun to refer to the same participant). As in the Kennedy et al. (1988) experiment, there was no reason for the participants to adapt because the computer understood everything the participant typed.

Cohen, Perrault, and Allen (1982) also found that in interactions with a question answering system, users expect more than just answers to unrelated questions. They expect the system to carry on a conversation which includes understanding of the user's goals and use of common ground that has been developed in the course of the interaction. In contrast to Brennan (1991) and in agreement with Kennedy et al. (1988), Cohen and his colleagues (1982) found that users did not change their expectations and style of responding as the interaction continued.

Oviatt, Cohen, and Wang (1994) used a speech interface to analyze how users' linguistic complexity is influenced by the modality and presentation structure used during human-computer interaction. A simulated service transaction system was used that could assist users with conference registration and car rental exchanges. Users could speak naturally, but the feedback provided by the computer was displayed on the screen and was not conversational. The feedback consisted of filling in the fields on an electronic registration receipt as the information was processed. Presentation (structured and unconstrained) and communication modality (spoken, written, or both) each had an impact on linguistic variability. A more structured interface reduced the number of words, length of utterances, and amount of information per utterance. It also resulted in a restricted range of syntactic structures and reduced their ambiguity. The structured format also eliminated 70 percent of all speech disfluencies or mistakes (Oviatt, 1995). Similar to Brennan (1991), Oviatt et al. (1994) came to the conclusion that presentation format can influence the nature of the interaction. Although Oviatt (1995) compared the disfluencies in human-computer speech in her study to the results of human-human speech studies there was no actual computer speech involved. As mentioned above, the computer simply gave feedback by filling in the registration receipt.

Collectively, these studies have shown that users may take a conservative linguistic approach (e.g., more elaboration to ensure understanding) to communication with a computer which could lead to or even amplify the patterns found in noninteractive speech. Future research is needed to develop comprehensive models appropriate for human-machine spoken interaction. Designers of future systems will also need to consider the possibility that an application may elicit noninteractive speech phenomena, and that these phenomena may have adverse consequences for the human-machine interaction. There is a need for research which examines human-computer communication using modalities other than the keyboard. In addition, there is a need to expand and clarify the research on whether users modify their style of interaction according to the presentation format. Consequently, consideration of these issues contributed to the decision to use a spoken interface in the present study, as well as to the choice to examine discourse structure. It was hypothesized that the different levels of human-computer interaction would affect discourse structure.

Desire for Control

When looking at adaptive automation from a team perspective and considering the human and computer as partners, it may also be helpful to consider personality variables which may be related to this partnership. Burger and Cooper (1979) introduced the Desirability of Control (DC) Scale which is designed to measure individual differences in the general desire for control over the events in one's life. People who score high on the scale are described as decisive, assertive, and active (Burger & Cooper, 1979). People who score low on the DC Scale are described as "generally nonassertive, passive, and indecisive. These people are less likely to attempt to influence others and may prefer that many of their daily decisions be made by others" (Burger & Cooper, 1979, p. 383).

Desire for control as measured by the DC Scale has been found to affect achievement-related behavior (Burger, 1985). High-DC participants have been shown to display higher levels of aspiration, have higher expectancies for their performance, and set more realistic expectations than low-DC participants. In addition, high-DC participants responded to a challenging task with more effort, persisted longer, and performed better than low-DC participants.

Although desire for control has not been studied with regard to team dynamics, it seems to have relevance in this area. A high-DC person might be less willing to act as a team member in solving problems because they have a need to control the situation. On the other hand, a low-DC person might rely too heavily on their partner. Either of these effects within a team could have a detrimental impact on the efficiency of the interaction. Desire for control could also affect human-computer interaction in adaptive automation in the same way.

Desire for control might also affect the performance of the human-computer team on challenging tasks. Because low-DC participants are less likely to respond well to a challenge, they may not do as well on more complicated tasks when paired with a computer partner. A high-DC person might show better performance in this kind of situation. In other words, low- and high-DC participants might show different patterns of interaction with the computer partner. Because desire for control may have an effect on human-computer interaction in adaptive automation it is important to study this variable as it relates to performance and human-computer interaction variables.

Present Research

The present study was designed to investigate the effects of human-computer communication mode, task complexity, and desire for control in an adaptive task on performance and discourse organization. An adaptive interface was chosen for the present study due to the growing importance of adaptive technology, the need for additional understanding of human-computer cooperation, and the need for an expanded definition of the human-computer team (see Bushman et al., 1993; Hammer & Small, 1995).

The importance of information flow to an efficiently functioning team emphasizes the importance of research on the effects of communication mode on human-computer interaction (see Fleishman & Zaccaro, 1992; Salas et al., 1992; Scerbo, 1996). The current study was designed to provide information for implementing successful human-computer communication in adaptive automation.

The success of using the PNAMBiC method (Gould et al., 1983) to study the human factors requirements of technology before the technology itself is developed led to the decision to use a PNAMBiC adaptive interface in the present study. Using this method will result in information which may affect the way this technology is implemented once it becomes technically feasible.

It is important to understand the limitations that occur in human communication with a computer. There will undoubtedly be differences in the richness of the communication process which may result in differences in communication ability and the need to minimize any undesirable results due to those differences (see Ochsman & Chapanis, 1974; O'Conaill et al., 1993). In addition, the research on voice versus nonvoice modes contributed to the decision to use a speech interface in the present study. It is believed that adaptive automation systems of a complex nature will require a voice interface in order to successfully exchange information.

The studies on feedback highlight its importance in the communication process. Feedback availability affects discourse structure as well as communication efficiency and will be an important issue for human-computer communication (see Kiesler & Sproull, 1992; Krauss & Weinheimer, 1966; Kraut et al., 1982; Oviatt & Cohen, 1991). The lack of research on feedback availability in human-computer interaction contributed to the decision to study different levels of human-computer speech interaction in the present study. It is hypothesized that the different modes of interaction used in the present study will affect the ability of the human to receive feedback from the computer and therefore affect discourse structure and performance.

Studies on human-computer communication have shown that users may take a conservative linguistic approach (e.g., more elaboration to ensure understanding) to communication with a computer which could lead to or even amplify the patterns found in noninteractive speech (see Guindon et al., 1987; Kennedy et al., 1988). In addition, there is a need to expand and clarify the research on whether users modify their style of interaction according to the presentation format. This need contributed to the decision to use a speech interface, as well as to the choice to examine discourse structure. It is hypothesized that the different levels of human-computer interaction will affect discourse structure.

The research reviewed above emphasizes the need to understand communication processes between humans and computers if they are to work together as teammates. Because current technology does not allow for fully interactive speech with a computer, and speech interaction is likely to be a critical component of adaptive systems, it is essential to determine how limited interaction will affect performance and human-computer communication in an adaptive task. The present study was designed to investigate the effects of human-computer communication mode, task complexity, and desire for control in an adaptive task on discourse organization and performance.

Four levels of communication mode were used. Each differed in the level of restriction placed on communication between the participant and computer partner. Two levels of task complexity were used, with all participants completing both simple and complex tasks. It was hypothesized that task complexity would affect the dependent measures due to the differing need for assistance on easy and difficult problems. Desire for control (DC) was measured and participants were split into high-DC and low-DC groups for analysis. As stated above, desire for control was hypothesized to affect the interaction pattern.

Dependent measures included task score as well as participant opinions as measured by a questionnaire. In addition, transcripts were prepared in order to examine discourse organization. Participants' utterances were assessed for verbosity, disfluencies, and indices of common ground.

The following relationships were hypothesized a priori:

Communication Restriction

It was hypothesized that as more restrictions were placed on communication, performance would decrease, computer control would increase, and participant opinions would become more negative. It was expected that restricting communication would make the interaction less efficient, and make it more difficult for participants to complete the tasks. This would lead to lower scores and less satisfaction with the interaction.

As communication restriction increased, it was expected that measures of verbosity, disfluencies, and indices of common ground would decrease. Increased restriction would result in less verbosity. A decrease in disfluencies was expected as communication restriction increased due to the increased need to plan utterances, resulting in fewer mistakes. Indices of common ground were expected to decrease due to the less "human" nature of the communication as restriction increased. Participants would have more difficulty communicating with the computer resulting in less understanding.

Correlations with Performance

It was hypothesized that performance would be correlated with words per minute and anaphoric reference and that performance would be inversely related to computer control and complex referring expressions. Performance should increase as the participant communicates more with the computer. Because anaphor is an indicator of common ground, performance should increase as anaphor increases. As computer control increases, performance should decrease because the participant is having trouble completing the tasks. Finally, as the number of complex referring expressions increases, performance should decrease because complex referring expressions indicate that the participant does not think there is common ground with their partner.

Correlation of Mean Length of Utterances and Disfluencies

It was hypothesized that mean length of utterances would be correlated with disfluencies. This was based on Oviatt & Cohen's (1991) finding that disfluencies increased as utterance length increased.

Task Complexity

Task complexity was expected to interact and amplify the basic effects described above for communication restriction. It was hypothesized that task score would be higher for simple tasks than complex tasks for all communication groups, and that as restriction increased, it would lower scores for complex tasks, but would have a minimal effect for simple tasks. This interaction was expected because participants might be able to complete the simple tasks regardless of how well they could communicate with the computer, but that communication would be more important for completing the complex tasks. In addition, it was hypothesized that computer control would be lower for simple tasks than for complex tasks.

Measures of verbosity and disfluencies were expected to be higher for complex tasks than for simple tasks. The complex nature of the tasks would lead to more communication and less time to plan utterances, resulting in more mistakes. Indices of common ground were not expected to differ according to task complexity. Although there would be less verbosity for simple tasks, communication mode would have the overriding effect on indices of common ground resulting in similar per minute measures of these indices for simple and complex tasks.

Desire for Control

Desire for control was expected to amplify the basic effects described above for communication restriction and task complexity. It was hypothesized that high-DC participants would score higher on complex tasks than low-DC participants, but there would be little difference in task scores for simple tasks. This interaction was expected because of the experimental evidence that high-DC individuals respond better to challenging tasks than low-DC individuals (Burger, 1985).

It was also expected that high-DC participants would have more negative opinions about the interaction than low-DC participants due to their desire to have more control over the outcome of a situation.

CHAPTER II

METHOD

This study used a 4 (communication mode) x 2 (task complexity) x 2 (desire for control) mixed-model factorial design with participants nested in communication mode and desire for control.

Participants

Desire for Control Pre-Screening

Participants were 64 university students who received course credit for their participation. Three hundred and two possible participants were prescreened for the desire for control variable using the Desirability of Control (DC) Scale (Burger & Cooper, 1979; see Appendix A). The DC Scale identifies the extent to which people are motivated to be in control. The DC Scale has been found to have adequate internal consistency ($r_{yy} = .80$) and test-retest reliability ($a = .75$), as well as discriminant validity from measures of locus of control ($r = -.19$; Rotter, 1966) and social desirability ($r = .11$; Crowne & Marlowe, 1960). Construct validation evidence was provided by studies on learned helplessness, hypnosis, and illusion of control (Burger & Cooper, 1979).

The mean DC score for females was $\underline{M} = 102.96$, $\underline{SD} = 12.51$, and $\underline{M} = 106.17$, $\underline{SD} = 11.23$ for males. A $\underline{t}$ test showed these means to be significantly different ($\underline{t} = -2.20$, $\underline{p} < .05$), therefore separate cutoff scores were used for selecting the 64 male and female participants for the main study. The cutoff scores for females were a DC score less than 93 (Low) or greater than 112 (High). The cutoff scores for males were a DC score less than 98 or greater than 114.

Main Study

Participants who met the cutoff scores were contacted for participation in the main study. 16 participants were assigned to each communication mode in a quasi-matched groups design. Half of the participants in each communication mode condition scored high on desire for control and half scored low. Participants with the highest (or lowest) scores were assigned to separate groups and then the next set of highest (or lowest) scores were assigned to separate groups. This continued until all potential participants had been assigned to a group. The object of this type of group assignment was to reduce the possibility of group differences for DC score.

The experimenters were blind to the DC level (high or low) and score of each participant. Participant group assignments were made by another individual who used a coding scheme which allowed the experimenters to assign participants to the correct groups.

Computer Task

Each participant was asked to complete a series of tasks using the Expert Travel Planner program by Expert Software on the computer (see Appendix B). This is a commercial program that allows users to plan trips including mileage, gas consumption, type of route, and itineraries. The participant used the keyboard to complete the tasks. The mouse was reserved for the computer partner and was used when the computer took over according to the adaptive rules (see below). When the computer partner was not using the mouse, the pointer was moved to the top right of the screen and was not visible.

Apparatus

Figure 1 illustrates the experimental setup and hardware configuration. The testing site was divided into two rooms. The participant's room contained a WIN IBM compatible computer running MS-DOS with a WIN 13 inch color monitor and a WIN keyboard. A mouse was also connected to the participant's computer, but was placed in the experimenter's room by a cable that ran through the wall. A Radio Shack 2-station wired intercom (#43-222A) ran between the two rooms. The intercom was placed in an inconspicuous location and allowed the experimenters to hear the participants' comments. A QVS MSV604 VGA monitor signal splitter was connected to the participant's computer. The signal splitter sent a copy of the images on the participant's monitor to a second WIN 13 inch color monitor in the experimenter's room allowing the experimenters to follow the particpant's progress.

The experimenter's room also contained a separate WIN IBM compatible computer running Windows 95 with a TVM MD System 13 inch Super Sync 2A+ monitor, a WIN keyboard, a mouse, and Sun CP55 speakers. This computer was used to generate the computer partner's audio responses. Thus, the speakers

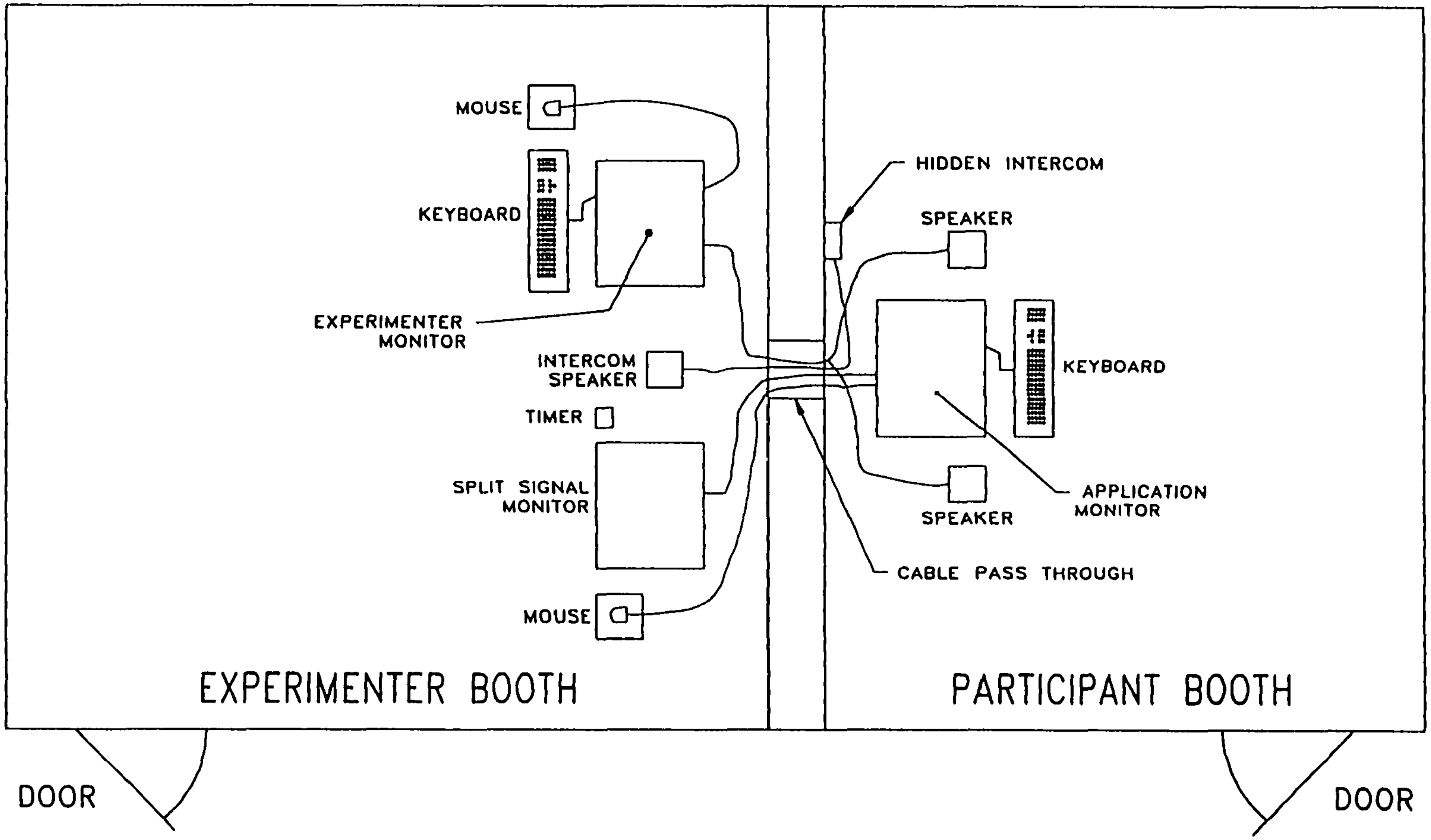

FIGURE 1. EXPERIMENTAL HARDWARE CONFIGURATION

connected to this computer were placed in the participant's room.

Interface

A PNAMBiC (Pay No Attention to the Man Behind the Curtain) method was used to simulate the computer partner (Gould et al., 1983; Newell et al., 1990; Newell et al., 1991). This was done with the experimenter in a separate room from the participant. She could hear what the participant said and responded according to the communication condition assigned to the participant. All responses were prerecorded and the experimenter chose the appropriate .WAV file which was played to the subject. In addition, she could see the participant's actions on a separate monitor which displayed the same image as seen on the participant's monitor. A confederate used the mouse to take over the task from the participant when necessary.

Procedure

The participant was tested in a small, sound-attenuated room with no windows and the experimenter and confederate were in a separate room out of sight of the participant. Participants were tested individually and all sessions were audio taped. Each participant was preassigned to a communication mode and the appropriate set of formalized instructions was read aloud (see Appendix C) by the confederate. Participants were told that they were testing a computerized system that would act as a partner in completing the tasks. They were asked to interact with this computer partner to help test the system. They were given appropriate instructions for communicating with the partner and told that the computer would act just like a human partner in that it might give advice or even take over the tasks. They were asked to cooperate with the computer in completing the tasks as quickly and accurately as possible. Participants had 10 minutes to complete the simple tasks and 20 minutes to complete the complex tasks (see below). Pilot work indicated that participants would be unable to complete all the tasks given in the allotted time. Participants were asked to work as quickly as possible to complete the tasks in the given time period. They were put under time pressure to minimize the chances they would try to complete the tasks without working with the computer partner.

The participant and the confederate then completed an "ice breaking" session designed to show the participant how to speak to the computer and how the computer might take over the task. Participants were asked to change the system colors by asking the computer for help as required by their communication mode assignment. Next, they were asked to refrain from doing anything for a while in order to make the computer "think" they were having trouble. The computer then took over the task and told them what to do. The experimenter controlled both the mouse and the sound files for the ice breaker session.

Any questions were answered and the confederate left the room. Participants were told when to start and stop the first set of tasks. During the session, the computer partner answered the participant's questions according to the rules for her group assignment. In addition, the adaptive rules stated that the computer partner would take over the task from the participant if she did not speak or press any keys for 30 seconds. If the participant spent more than 3 minutes working on one task, the computer partner would interrupt every 30 seconds until the task was completed. Each computer intervention consisted of completing the next step in the task solution and informing the participant what was being done. After the appropriate time period (10 minutes for simple tasks, 20 minutes for complex tasks), the confederate returned and asked the participant to rate the tasks they had just completed for difficulty, and then set up the tasks for the next session. Upon completion of the second session, the confederate returned and asked the participant to rate the second set of tasks for difficulty. Next, she asked the participant to fill out a questionnaire which addressed the participant's ability to communicate with the computer partner, the helpfulness of the computer partner, enjoyment of the interaction, and the participant's computer experience (see Appendix D). Participants were then debriefed and thanked for their time.

Independent Measures

Communication Mode

There were four levels of communication mode:

1. Context Sensitive Interaction. In this communication mode the participant was permitted to speak normally. The computer partner chose a response from a list that included context sensitive responses. For example, "Use

the Facts function on the File menu to find out about Fayetteville's history."

2. Limited Response Interaction. In this mode the participant could speak normally and the computer partner chose a response from a limited list that did not include context sensitive responses. For example, "Use the Facts function on the File menu."

3. Limited Human-Computer Interaction. In this mode the participant was required to use keywords to formulate utterances and the computer partner chose responses from a limited list that did not include context sensitive responses. A list of the keywords was provided for the participant to use as a reference. For example, participants were told they could use the keyword "Help" plus a menu function to receive information about that function.

4. Control Group. The control group was not able to communicate with the computer partner. As in all the other groups, however, the computer partner would take over the tasks according to the adaptive rules (see above).

Task Complexity

Tasks were designed to represent the program's capabilities, divided into simple and complex sets, and then rated for difficulty on a scale of 1 to 7 during a pilot test (see Appendix B). The simple tasks had a lower difficulty rating of $\underline{M}_S = 2.57$, $\underline{SD} = .77$, while the complex tasks had a higher difficulty rating of $\underline{M}_C = 4.05$, $\underline{SD} = .92$.

In addition, a GOMS analysis (Kieras, 1988) was performed to ensure that the task goals were evenly distributed throughout the simple and complex tasks (see Appendix E). The analysis showed that the goals were evenly distributed for halves and thirds of both the simple and complex tasks.

In the main study, after participants completed each set of tasks they were asked to rate the difficulty of the set from 1 (easy) to 5 (difficult) as a manipulation check. The order of presentation of the task sets was counterbalanced.

Dependent Measures

Task Score

Task completion was scored by tabulating the number of tasks that were completed in the allotted time period for both simple and complex tasks. This task score was then converted to a common metric by dividing the number of tasks completed by the minutes on task to obtain tasks per minute.

Computer Control

The level of computer control was measured by tabulating the number of times the computer partner took over the tasks from the participant. This number was then converted to a common metric by dividing by the minutes on task to obtain the number of interruptions per minute. This measure was derived from the analysis of participants' transcripts. As a result, there are no computer control results for the control group.

Discourse Organization

Transcripts of the communication between the partners were prepared and analyzed. Each subject's speech was transcribed from the audiotapes of the sessions. Attention was paid to transcribing speech verbatim without editing it in any way. This included transcribing the speech as well as non-speech sounds, disfluencies, and confirmations. The following dependent measures were coded:

1. Total Words. The total number of spoken words was tabulated for each participant. Total words was converted to words per minute for use as a dependent variable. The total number of words also provided a baseline for converting other dependent measures to a rate per 100 words.

2. Mean Length of Utterance. The average number of words per utterance was calculated for each participant by dividing total words by the number of utterances. Utterance boundary judgments were assisted by cues indicating participant disengagement such as pausing, sentence intonation, and change in intensity caused by the participant moving away from the microphone. Mean length of utterance was used primarily for examining the relation between utterance length and disfluency rate.

3. Disfluencies. Spontaneously occurring disfluencies were tabulated for each participant and included the following: (a) content self-corrections - errors that are corrected as the participant speaks (e.g., "Virginia . . . Maryland"), (b) uncorrected miscommunications - errors that are not corrected as the participant

speaks (e.g., speaking the wrong number or state), (c) false starts - changes in the grammatical structure of an utterance that occur as the participant speaks (e.g., "I need to . . . I want to go to Portland"), (d) verbatim repetitions (e.g., "of the . . . of the"), (e) filled pauses - non-word sounds that fill pauses in running speech (e.g., "uh," "um"), and (f) simultaneous speech - speech that overlaps that of the computer partner. The total number of disfluencies per condition was then converted to a rate per 100 words.

4. Confirmations. Confirmations included: (a) repetition of a portion of a partner's utterance, (b) explicit acknowledgment, and (c) a relevant next conversational turn and were converted to a rate per 100 words.

5. Anaphoric Reference. Anaphors are words which point back to elements (events, objects, people, places, times) already mentioned or implied in a conversation. It is an important index of common ground or shared knowledge (Brennan, 1991; Kennedy et al., 1988). The use of anaphor was calculated per 100 words and included two measures. The first was ellipses - the omission of one or more words that are needed to make an utterance grammatically complete (e.g., "How do I find Maine?. . .What about Vermont?" In this case the first question must be remembered in order to correctly interpret the second question.). The second measure was pronominal anaphor - the replacement of a noun or noun phrase with another referent (e.g., "How do I get to the Florida map?. . .What do I do when I get there?"). In this case "Florida map" is replaced by "there".

6. Complex Referring Expressions. This measure is also an index of common ground (Guindon et al., 1987; Oviatt & Cohen, 1991) and includes: (a) noun phrases with prepositional attachments - complex substitutes for nouns (e.g., "the part of the map with the wide blue line bordering it") and (b) elaborations - continued explanation of an earlier description in the discourse (e.g., "The 95-85 intersection. . . The spot where 95 and 85 cross each other."). The use of these types of expressions was calculated per 100 words.

Participant Ratings

Participants were asked to complete a questionnaire about the interaction (see Appendix D). They were asked to rate their ability to communicate with the computer partner, the helpfulness of the computer partner, and how much they enjoyed the interaction. Participants were also asked about their computer experience and a manipulation check was included to confirm that the participant believed he was interacting with a computer.

CHAPTER III

RESULTS

The data from the study were analyzed using a series of ANOVAs and correlations. In all cases an alpha level of .05 was used to determine statistical significance. Analyses of simple effects and Student Newman-Keuls (SNK) post hoc tests were used to examine significant effects.

Desirability of Control Scores

A 4 (communication mode) x 2 (gender) ANOVA was used to analyze the Desirability of Control (DC) scores of the participants in order to determine if there were group differences in DC score. The analysis of variance summary table for DC score is presented in Table 1. There were no significant differences in DC scores for communication mode, gender, or communication mode by gender. Thus, DC scores were comparable across communication groups. Due to scheduling difficulties, the number of high- and low-DC subjects in each communication group was not equal as was planned in the experimental design.

Table 1 - Summary of ANOVA for DC Scores

Source	**DF**	**SS**	**MS**	**F**	**Eta^2**
CM	3	169.258	56.419	0.19	.
GEN	1	432.571	432.571	1.43	.
CM*GEN	3	244.805	81.602	0.27	.
SUB(CM*GEN)	55	16688.131	303.420	.	.
Total		17680.984			

Interrater Reliability

All participant transcripts were coded twice for accuracy. One participant from the limited response group had to be dropped from the analysis due to an audiotape failure, leaving a total of 63 participants. In addition, approximately 25% of the transcripts for each communication mode were selected randomly and independently coded by another rater. The results of the two codings were used to calculate interrater reliability for each of the dependent measures by correlating the number of occurrences reported by the first rater with the number of occurrences reported by the second rater for each measure over this subset of transcripts. All of the reliabilities were above 0.90 except for two measures, confirmations and complex referring expressions. The two raters discussed the ratings for these measures while reanalyzing the transcripts. Discrepancies were resolved and the recalculated reliabilities were above 0.90. The reliabilities for the transcript variables are shown in Table 2.

Task Score and Task Difficulty Ratings

A 4 (CM) x 2 (TC) x 2 (DC) mixed-model factorial ANOVA with participants nested in communication mode and desire for control was used to analyze the number of questions completed per minute of time on task (task score) and the task difficulty ratings.

Table 2 - Interrater Reliabilities for Transcript Dependent Measures

Variable	Reliability
Number of Interruptions	0.996
Number of Utterances	0.999
Total Words	1.000
Words per Minute	1.000
Mean Length of Utterances	0.999
Disfluencies	0.932
Confirmations	0.951
Anaphor	0.997
Complex Referring Expressions	0.936

Task Score

Task score refers to the number of questions completed per minute of time on task. The analysis of variance summary table for task score is presented in Table 3.

Communication Mode

There was a significant main effect for CM, F (3, 55) = 8.16. A post hoc test showed that task scores for the context sensitive interaction group (M = 0.772, SD = 0.440) and the limited response interaction group (M = 0.768, SD = 0.465)

Table 3 - Summary of ANOVA for Task Score

Source	DF	SS	MS	F	Eta^2
CM	3	1.503	0.501	8.16*	0.066
DC	1	0.000	0.000	0.00	.
CM*DC	3	0.170	0.057	0.92	.
TC	1	14.403	14.403	317.71*	0.634
CM*TC	3	0.519	0.173	3.82*	0.023
DC*TC	1	0.000	0.000	0.00	.
CM*DC*TC	3	0.098	0.033	0.72	.
SUB(CM*DC)	55	3.377	0.061	.	.
SUB*TC(CM*DC)	55	2.493	0.045	.	.
Total		22.701			

were significantly higher than scores for the limited human-computer interaction group (M = 0.572, SD = 0.373) and the control group (M = 0.536, SD = 0.386). No other comparisons reached statistical significance.

Task Complexity

There was a significant main effect for TC, F (1, 55) = 317.71. An examination of the means indicated that task scores were significantly higher for simple tasks (M = 1.000, SD = 0.350) than for complex tasks (M = 0.321, SD = 0.095).

Communication Mode by Task Complexity

There was a significant interaction between communication mode and task complexity, F (3, 55) = 3.820. Figure 2 illustrates this interaction. An analysis of the simple effects showed that there was a significant difference in task score among the communication groups for simple tasks, F (3, 55) = 11.88. A post hoc

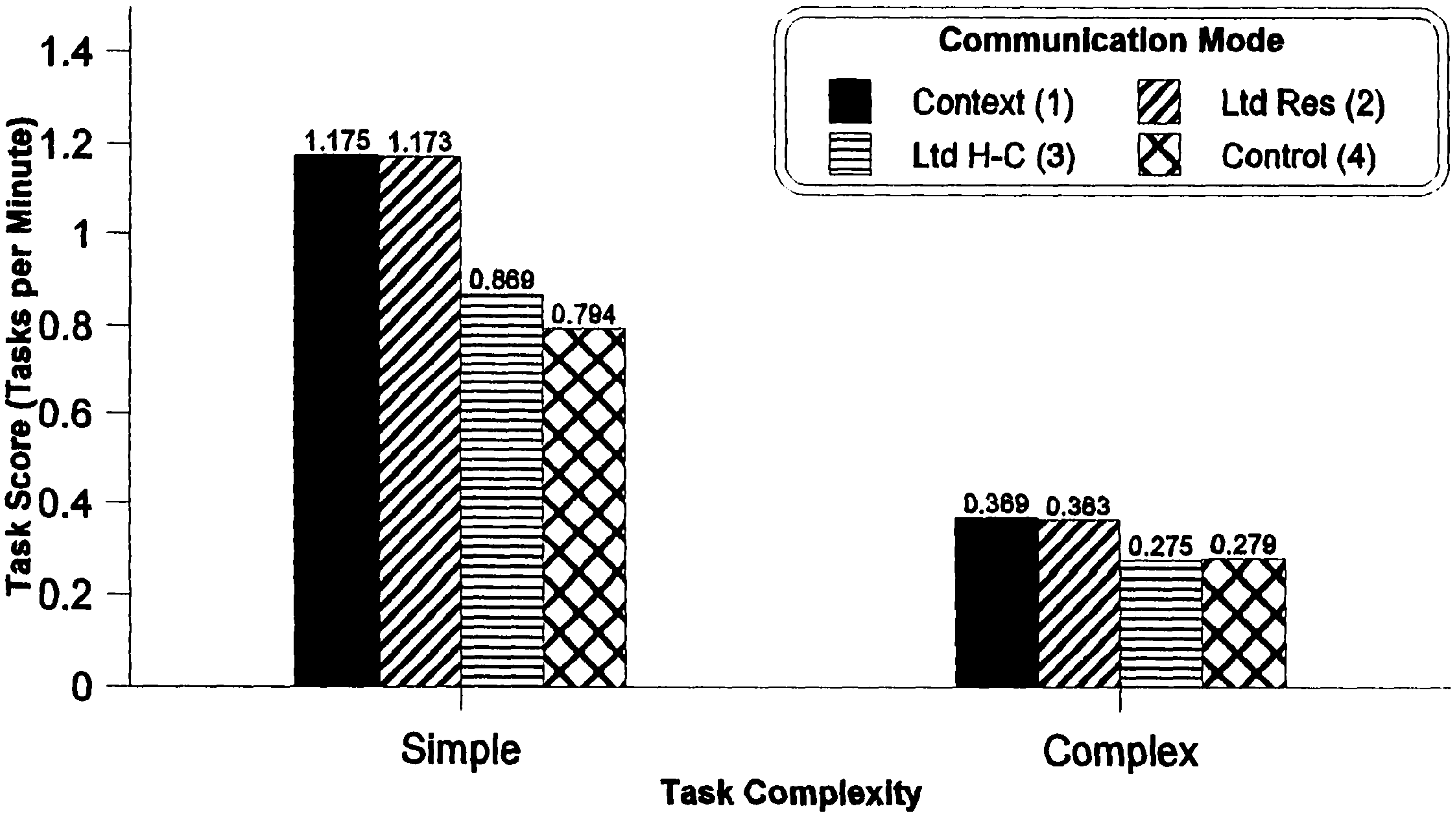

Figure 2. Task Score for Communication Mode and Task Complexity

test showed that task score for simple tasks was significantly higher in the context sensitive (M = 1.175, SD = 0.221) and limited response (M = 1.173, SD = 0.294) groups than in the limited human-computer (M = 0.869, SD = 0.309) and control (M = 0.794, SD = 0.394) groups. No other comparisons reached statistical significance.

Task Difficulty Ratings

Participants rated task difficulty using a scale of 1 (easy) to 5 (difficult). The analysis of variance summary table for task difficulty ratings is presented in Table 4.

Communication Mode

There was a significant main effect for CM, F (3, 55) = 4.05. A post hoc test showed that difficulty ratings for the context sensitive interaction group (M = 2.594, SD = 0.911) were significantly lower than ratings for the limited human-computer interaction group (M = 3.406, SD = 0.756) and the control group (M = 3.312, SD = 1.120). The mean for the limited response group was, M = 3.067, SD = 1.048. No other comparisons reached statistical significance.

Task Complexity

There was a significant main effect for task complexity (TC), F (1, 55) = 20.77. An examination of the means indicated that difficulty ratings were significantly lower for simple tasks (M = 2.794, SD = 0.919) than for complex tasks (M = 3.397, SD = 1.009).

Table 4 - Summary of ANOVA for Task Difficulty Ratings

Source	DF	SS	MS	F	Eta^2
CM	3	13.470	4.490	4.05*	0.106
DC	1	0.495	0.495	0.45	.
CM*DC	3	3.760	1.253	1.13	.
TC	1	11.937	11.937	20.77*	0.094
CM*TC	3	4.802	1.601	2.79*	0.038
DC*TC	1	0.557	0.557	0.97	.
CM*DC*TC	3	0.991	0.330	0.57	.
SUB(CM*DC)	55	60.999	1.109	.	.
SUB*TC(CM*DC)	55	31.602	0.574	.	.
Total		126.857			

Communication Mode by Task Complexity

There was a significant interaction between communication mode and task complexity, F (3, 55) = 2.79. Figure 3 illustrates this interaction. An analysis of the simple effects showed that in the limited response group, ratings were significantly higher for complex tasks (M = 3.667, SD = 0.900) than for simple tasks (M = 2.467, SD = 0.834), (F (1,55) = 6.747). No other comparisons reached statistical significance.

Computer Control and Discourse Organization

A 3 (CM) x 2 (TC) x 2 (DC) mixed-model factorial ANOVA with participants nested in communication mode and desire for control was used to analyze computer control, mean length of utterances, disfluencies, confirmations, anaphoric reference, and complex referring expressions.

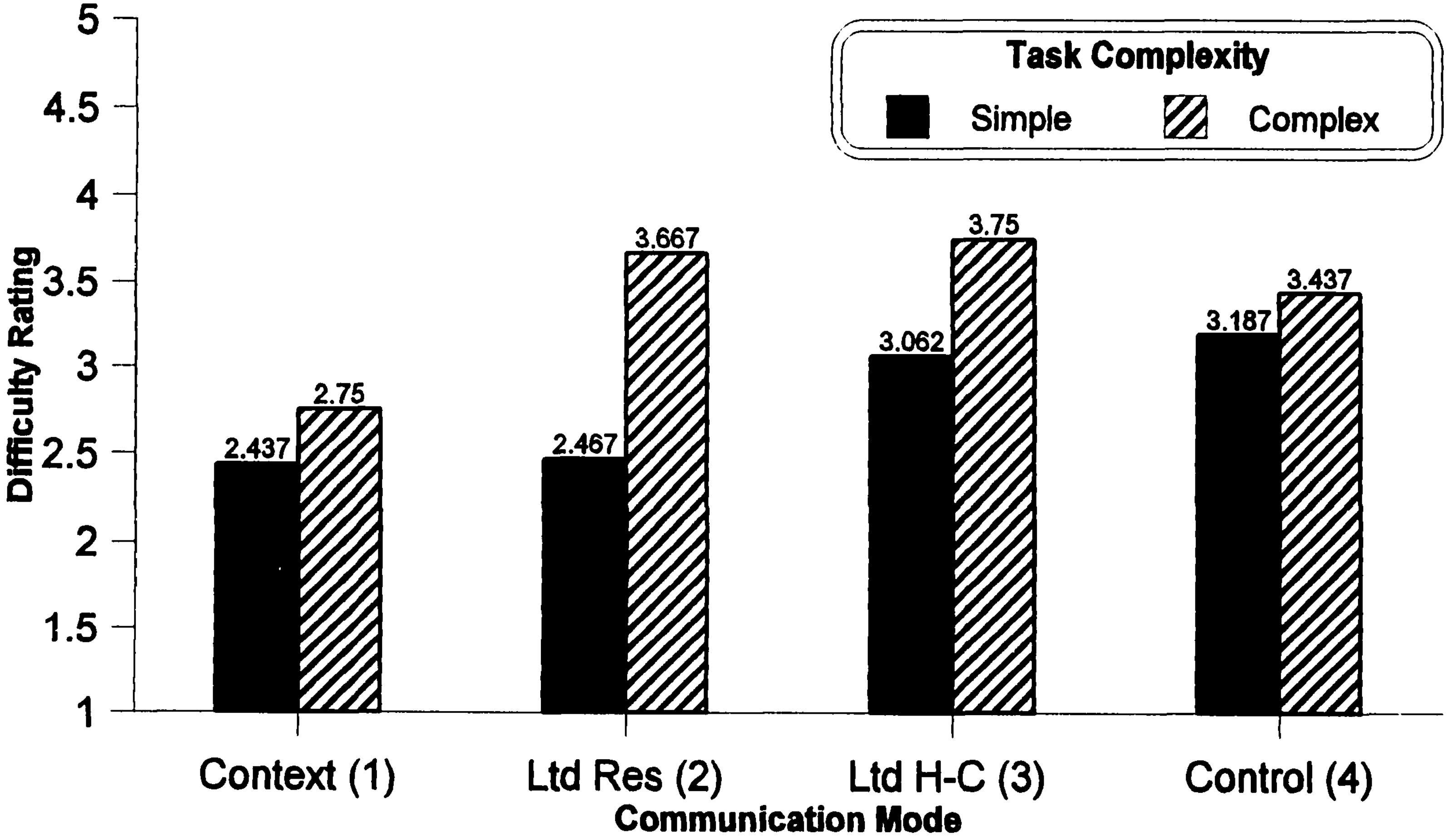

Figure 3. Task Difficulty Ratings for Communication Mode and Task Complexity

Computer Control

Computer control refers to the number of times the computer partner took over the task according to the adaptive rules divided by time on task. The analysis of variance summary table for computer control is presented in Table 5.

Table 5 - Summary of ANOVA for Computer Control

Source	**DF**	**SS**	**MS**	**F**	**Eta^2**
CM	2	0.980	0.490	15.90*	0.238
DC	1	0.004	0.004	0.12	.
CM*DC	2	0.049	0.025	0.80	.
TC	1	0.729	0.729	31.59*	0.177
CM*TC	2	0.047	0.023	1.02	.
DC*TC	1	0.000	0.000	0.01	.
CM*DC*TC	2	0.070	0.035	1.51	.
SUB(CM*DC)	41	1.263	0.031	.	.
SUB*TC(CM*DC)	41	0.946	0.023	.	.
Total		4.116			

Communication Mode

There was a significant main effect for CM, F (2, 41) = 15.90. A post hoc test showed that computer control for only the context sensitive interaction group (M = 0.161, SD = 0.122) and the limited response interaction group (M = 0.210, SD = 0.171) was significantly lower than computer control for the limited human-computer interaction group (M = 0.395, SD = 0.243).

Task Complexity

There was a significant main effect for TC, F (1, 41) = 31.59. An examination of the means indicated that computer control was significantly lower for simple tasks (M = 0.166, SD = 0.171) than for complex tasks (M = 0.347, SD = 0.209).

Words per Minute

Words per minute (WPM) refers to the number of words spoken by the participant divided by time on task. The analysis of variance summary table for WPM is presented in Table 6.

Table 6 - Summary of ANOVA for Words per Minute

Source	DF	SS	MS	F	Eta^2
CM	2	750.549	375.275	14.38*	0.326
DC	1	7.342	7.342	0.28	.
CM*DC	2	75.654	37.827	1.45	.
TC	1	78.626	78.626	12.60*	0.034
CM*TC	2	24.223	12.111	1.94	.
DC*TC	1	19.975	19.975	3.20	.
CM*DC*TC	2	9.930	4.965	0.80	.
SUB(CM*DC)	41	1069.942	26.096	.	.
SUB*TC(CM*DC)	41	255.764	6.238	.	.
Total		2298.072			

Communication Mode

There was a significant main effect for CM, F (2, 41) = 14.38. A post hoc test showed that WPM for the context sensitive interaction group (M = 8.251, SD = 4.505) and the limited response interaction group (M = 7.573, SD = 5.509) was significantly higher than WPM for the limited human-computer interaction group

(M = 2.041, SD = 1.287). No other comparisons reached statistical significance.

Task Complexity

There was a significant main effect for TC, F (1, 41) = 12.60. An examination of the means indicated that WPM was significantly higher for simple tasks (M = 6.898, SD = 5.636) than for complex tasks (M = 4.944, SD = 4.030).

Mean Length of Utterances

Mean length of utterances (MLU) refers to the number of words spoken divided by the number of utterances. The analysis of variance summary table for MLU is presented in Table 7.

Table 7 - Summary of ANOVA for Mean Length of Utterances

Source	DF	SS	MS	F	Eta^2
CM	2	401.759	200.880	28.47*	0.496
DC	1	10.539	10.539	1.49	.
CM*DC	2	18.344	9.172	1.30	.
TC	1	3.725	3.725	2.51	.
CM*TC	2	1.203	0.602	0.41	.
DC*TC	1	0.116	0.116	0.08	.
CM*DC*TC	2	3.907	1.953	1.32	.
SUB(CM*DC)	41	289.302	7.056	.	.
SUB*TC(CM*DC)	41	60.736	1.481	.	.
Total		810.303			

Communication Mode

There was a significant main effect for CM, F (2, 41) = 28.47. A post hoc test showed that MLU for the context sensitive interaction group (M = 7.161, SD = 2.288) and the limited response interaction group (M = 7.947, SD = 2.603) was significantly higher than MLU for the limited human-computer interaction group (M = 3.113, SD = 0.950). No other comparisons reached statistical significance.

Disfluencies

Disfluencies is the number of mistakes in the participant's speech divided by total words. The analysis of variance summary table for disfluencies is presented in Table 8.

Table 8 - Summary of ANOVA for Disfluencies

Source	DF	SS	MS	F	Eta^2
CM	2	0.001	0.001	0.65	.
DC	1	0.000	0.000	0.26	.
CM*DC	2	0.007	0.003	3.63*	0.073
TC	1	0.000	0.000	0.21	.
CM*TC	2	0.004	0.002	2.14	.
DC*TC	1	0.001	0.001	0.66	.
CM*DC*TC	2	0.001	0.001	0.63	.
SUB(CM*DC)	41	0.040	0.001	.	.
SUB*TC(CM*DC)	41	0.042	0.001	.	.
Total		0.096			

Communication Mode by Desire for Control

There was a significant interaction between communication mode and desire for control, F (2, 41) = 3.63. Figure 4 illustrates this interaction. An

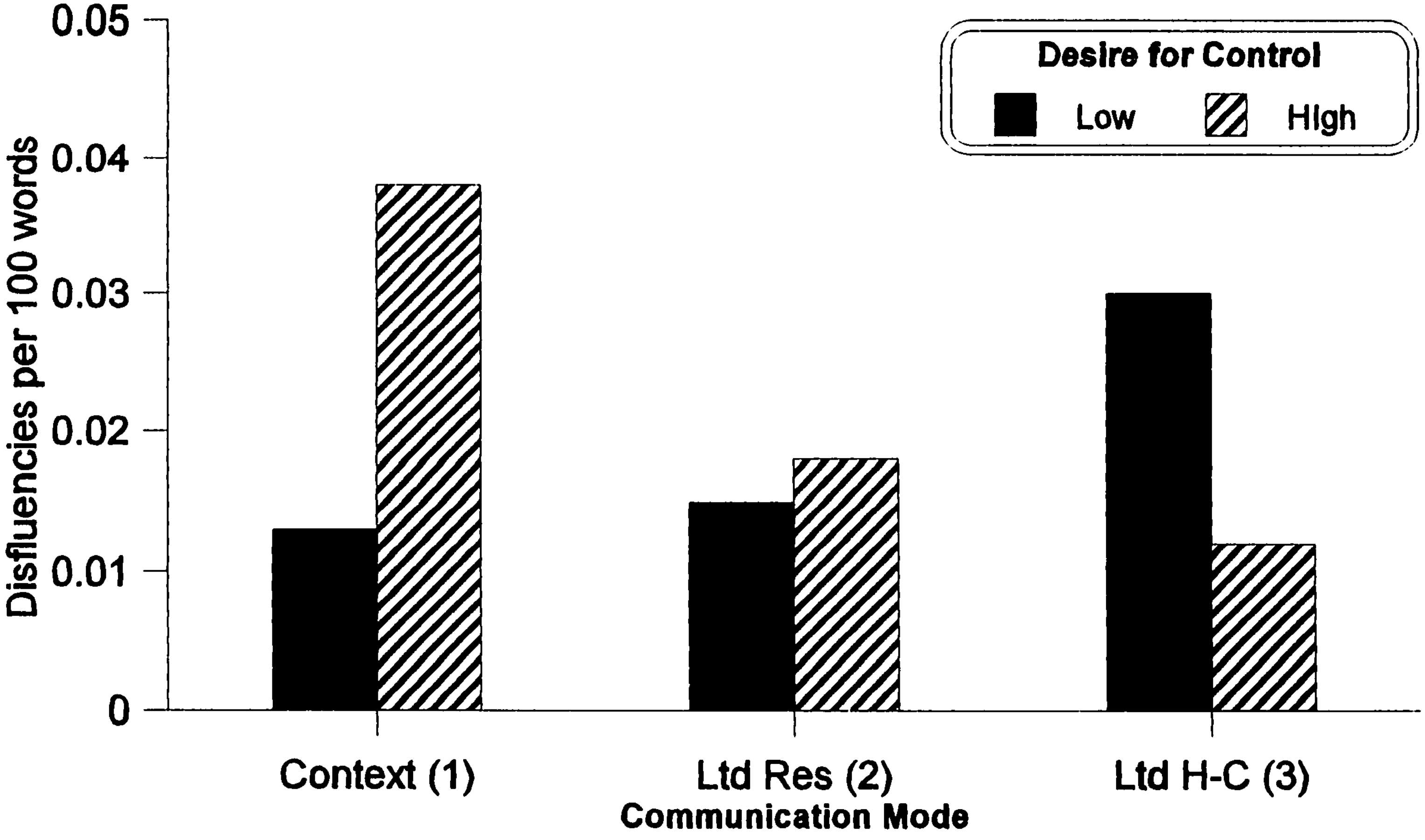

Figure 4. Disfluencies for Communication Mode and Desire for Control

analysis of the simple effects, however indicated that none of the differences for this interaction reached statistical significance. This is probably the result of unequal sample sizes for low- and high-DC subjects reducing the power of the post hoc test. In addition, a planned correlation for mean length of utterances and disfluencies did not reach significance.

Confirmations

Confirmations refers to the number of times the participant confirmed the computer partner's last utterance divided by total words. The analysis of variance summary table for confirmations is presented in Table 9.

Table 9 - Summary of ANOVA for Confirmations

Source	DF	SS	MS	F	Eta^2
CM	2	0.024	0.012	6.95*	0.175
DC	1	0.001	0.001	0.30	.
CM*DC	2	0.002	0.001	0.70	.
TC	1	0.000	0.000	0.00	.
CM*TC	2	0.003	0.001	1.75	.
DC*TC	1	0.000	0.000	0.06	.
CM*DC*TC	2	0.000	0.000	0.03	.
SUB(CM*DC)	41	0.072	0.002	.	.
SUB*TC(CM*DC)	41	0.033	0.001	.	.
Total		0.137			

Communication Mode

There was a significant main effect for CM, F (2, 41) = 6.95. A post hoc test showed that confirmations for the context sensitive interaction group

(M = 0.025, SD = 0.030) and the limited response interaction group (M = 0.013, SD = 0.025) was significantly lower than confirmations for the limited human-computer interaction group (M = 0.052, SD = 0.046). No other comparisons reached statistical significance.

Anaphoric Reference

Anaphoric reference is the number of elements in the participant's speech that refer to earlier elements in the conversation divided by total words. The analysis of variance summary table for anaphor is presented in Table 10.

Table 10 - Summary of ANOVA for Anaphoric Reference

Source	DF	SS	MS	F	Eta^2
CM	2	0.018	0.009	2.94	.
DC	1	0.006	0.006	2.03	.
CM*DC	2	0.009	0.004	1.45	.
TC	1	0.010	0.010	11.95*	0.049
CM*TC	2	0.003	0.001	1.59	.
DC*TC	1	0.001	0.001	0.65	.
CM*DC*TC	2	0.002	0.001	1.24	.
SUB(CM*DC)	41	0.124	0.003	.	.
SUB*TC(CM*DC)	41	0.036	0.001	.	.
Total		0.204			

Task Complexity

There was a significant main effect for TC, F (1, 41) = 11.95. An examination of the means indicated that anaphor was significantly lower for simple tasks (M = 0.012, SD = 0.034) than for complex tasks (M = 0.033, SD = 0.055).

Complex Referring Expressions

Complex referring expressions (CRE) refers to the number of times the participant elaborated or used complex noun phrases divided by total words. The analysis of variance summary table for CRE is presented in Table 11. There were no significant effects for CRE.

Table 11 - Summary of ANOVA for Complex Referring Expressions

Source	DF	SS	MS	F	Eta^2
CM	2	0.007	0.004	2.26	.
DC	1	0.001	0.001	0.31	.
CM*DC	2	0.002	0.001	0.50	.
TC	1	0.002	0.002	1.25	.
CM*TC	2	0.002	0.001	0.56	.
DC*TC	1	0.004	0.004	3.05	.
CM*DC*TC	2	0.002	0.001	0.79	.
SUB(CM*DC)	41	0.068	0.002	.	.
SUB*TC(CM*DC)	41	0.060	0.001	.	.
Total		0.149			

Participant Ratings

Ability to Communicate

A 3 (CM) x 2 (DC) ANOVA with participants nested in communication mode and desire for control was used to analyze ability to communicate with the computer partner. Participants were asked to rate their ability to communicate with the computer on a scale of 1 (easy) to 5 (difficult). The analysis of variance summary table for ability to communicate is presented in Table 12.

Table 12 - Summary of ANOVA for Ability to Communicate

Source	**DF**	**SS**	**MS**	**F**	**Eta^2**
CM	2	20.471	10.235	9.87*	0.297
DC	1	6.093	6.093	5.88*	0.088
CM*DC	2	0.901	0.450	0.43	.
SUB(CM*DC)	41	42.520	1.037	.	.
Total		68.936			

Communication Mode

There was a significant main effect for CM, F (2, 41) = 9.87. A post hoc test showed that ability to communicate for the context sensitive interaction group (M = 1.625, SD = 0.806) and the limited response interaction group (M = 2.000, SD = 1.195) was rated as significantly easier than ability to communicate for the limited human-computer interaction group (M = 3.125, SD = 1.147). No other comparisons reached statistical significance.

Desire for Control

There was a significant main effect for DC, F (1, 41) = 5.88. An examination of the means indicated that ability to communicate was rated as significantly easier by high-DC participants (M = 1.905, SD = 1.179) than by low-DC particpants (M = 2.538, SD = 1.208).

Helpfulness, Enjoyment, and Computer Experience

A 4 (CM) x 2 (DC) ANOVA with participants nested in communication mode and desire for control was used to analyze helpfulness of the computer partner, enjoyment of the interaction, self-rated computer ability, and self-reported years of computer experience.

Helpfulness

Participants were asked to rate the helpfulness of the computer partner on a scale of 1 (helpful) to 5 (unhelpful). The analysis of variance summary table for helpfulness is presented in Table 13. There were no significant effects found for participants' helpfulness ratings.

Table 13 - Summary of ANOVA for Helpfulness

Source	**DF**	**SS**	**MS**	**F**	**Eta^2**
CM	3	1.520	0.506	0.67	.
DC	1	0.707	0.707	0.93	.
CM*DC	3	2.454	0.818	1.08	.
SUB(CM*DC)	55	41.686	0.758	.	.
Total		46.413			

Enjoyment

Participants were asked to rate their enjoyment of the interaction on a scale of 1 (enjoyable) to 5 (unenjoyable). The analysis of variance summary table for enjoyment is presented in Table 14.

Table 14 - Summary of ANOVA for Enjoyment

Source	DF	SS	MS	F	Eta^2
CM	3	4.298	1.433	1.59	.
DC	1	0.013	0.013	0.01	.
CM*DC	3	10.271	3.423	3.79*	0.160
SUB(CM*DC)	55	49.635	0.902	.	.
Total		64.317			

Communication Mode by Desire for Control. There was a significant interaction between communication mode and desire for control, F (3, 55) = 3.79. Figure 5 illustrates this interaction. An analysis of the simple effects showed that enjoyment was rated as significantly lower (F (1,55) = 6.925) for low-DC than high-DC participants in the limited human-computer interaction group (M = 3.000 > M = 1.750). Enjoyment ratings were significantly higher (F (1, 55) = 4.432) for low-DC participants in the control group (M = 2.000 < M = 3.000).

Computer Ability

Participants were asked to rate their computer ability on a scale of 1 (beginner) to 5 (expert). The analysis of variance summary table for computer ability is presented in Table 15. There were no significant differences found for computer ability.

Years of Computer Experience

A 4 (CM) x 2 (DC) ANOVA with participants nested in communication mode and desire for control was used to analyze participants' self-reported years

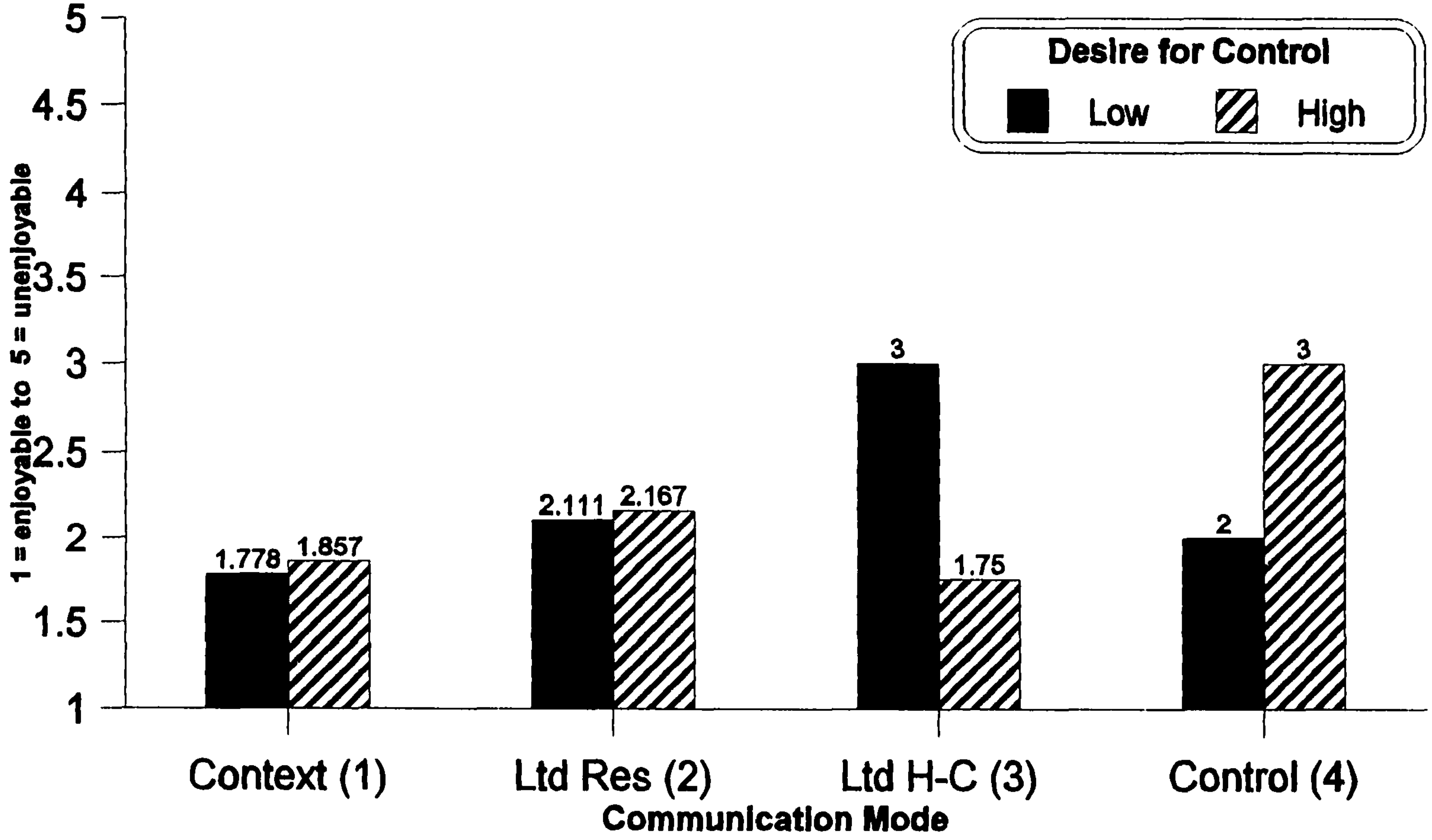

Figure 5. Enjoyment for Communication Mode and Desire for Control

Table 15 - Summary of ANOVA for Computer Ability

Source	DF	SS	MS	F	Eta^2
CM	3	1.934	0.644	0.80	.
DC	1	0.819	0.819	1.02	.
CM*DC	3	1.863	0.621	0.77	.
SUB(CM*DC)	55	44.367	0.807	.	.
Total		49.079			

of computer experience. Participants were asked to report how many years of computer experience they had. The analysis of variance summary table for years of computer experience is presented in Table 16.

Table 16 - Summary of ANOVA for Years of Computer Experience

Source	DF	SS	MS	F	Eta^2
CM	3	107.487	35.829	2.92*	0.132
DC	1	1.465	1.465	0.12	.
CM*DC	3	33.480	11.160	0.91	.
SUB(CM*DC)	55	675.136	12.275	.	.
Total		812.417			

Communication Mode. There was a significant main effect for CM, F (3, 55) = 2.92. A post hoc test showed that years of computer experience for the limited response interaction group (M = 7.167, SD = 3.589) was significantly higher than years of computer experience for the limited human-computer interaction group (M = 3.644, SD = 2.492). No other comparisons reached statistical significance.

Correlations with Task Score

A significant correlation between task score and words per minute was observed ($r = 0.271$). Task score was also inversely related to anaphor ($r = -0.203$), and computer control ($r = -0.627$). The correlation for task score with complex referring expressions was not significant.

The correlations for task score with computer ability and years of computer experience were used as manipulation checks and were not significant. Computer ability did not differ among the communication groups and was not related to task score. In addition, although the limited response interaction group had more years of computer experience, this experience was not correlated with task performance.

CHAPTER IV

DISCUSSION

The present study was conducted in order to examine communication processes between human-computer teammates. Because speech interaction is likely to be a critical component of adaptive systems, it is essential to determine how limited interaction will affect performance and human-computer communication in an adaptive task. Desire for control was identified as a personality variable that might be related to the communication process and was therefore included in the research design. Task complexity was also included in the design because it was expected to interact with communication mode and desire for control in its effects on performance and discourse structure.

Desirability of Control Scores

The quasi matched groups assignment of participants in the present study was used to avoid group differences in DC scores so that equivalent distributions of DC scores would be present in each communication mode. An analysis of the DC scores confirmed that there were no significant differences in DC score for communication mode or gender.

Participant Performance

Communication Restriction

The primary goal of the present study was to examine the effects of communication restriction on adaptive task interactions. It was hypothesized that as more restrictions were placed on communication, task scores would decrease and computer control would increase.

The hypothesis that task scores would decrease as communication restriction increased was partially supported. The results for task score showed that the context sensitive interaction and limited response interaction groups scored higher than the limited human-computer interaction and control groups. Moreover, this finding is supported by the results from the task difficulty ratings (see below). The tasks were rated as easier in the context sensitive interaction group and participants scored better in this group.

The results of the present study are consistent with those of Ochsman and Chapanis (1974). Their study examined time to solution, behavioral measures of activity, and linguistic measures for two-person teams working in 10 different communication modes. They found large differences for all three classes of dependent variables between groups with a voice channel and those with no voice channel. Participants in the voice mode conditions had shorter solution times, spent less time sending and receiving messages, created more messages, and had faster message rates than those in the nonvoice modes. The study also showed that communication time was inversely related to the richness of the communication mode.

The results from the present study show that when participants could communicate freely (i.e., context sensitive and limited response) they completed more tasks than participants whose communication was restricted or denied (i.e., limited human-computer interaction and control). This supports Ochsman and Chapanis' (1974) findings that voice and nonvoice modes of communication differ significantly, and that the richness of the communication mode affects communication as well.

The findings from the present study go beyond those of Ochsman and Chapanis (1974) because this is the first study in which the global task was decomposed into smaller subgoals providing a more fine-grained analysis of performance. In the past, researchers (Chapanis & Overbey, 1974; Chapanis et al., 1972; Chapanis et al., 1977; Krueger & Chapanis, 1980; Ochsman & Chapanis, 1974; O'Conaill et al., 1993; Oviatt & Cohen, 1991; Weeks & Chapanis, 1976) have used global completion time as a dependent measure. The use of this measure showed large differences between voice and nonvoice modes, but because all groups eventually solved the problem and the task was not broken into smaller components, analyses of the specific activities performed by the participants to solve the problems were not provided.

The present study used a finer-grained analysis of performance by including many individual activities in each task set. Further, in order to be able to simulate the computer responses, the tasks used in the present study had to be of a definable nature. The tasks were partitioned using a GOMS analysis (Kieras, 1988) and all solution paths were known in advance in order to prepare the computer responses.

Therefore, the knowledge domain for the tasks was fully described and was also used to decide when a participant had successfully completed each task. The opportunity to more fully describe performance was one of the reasons that the individual computer tasks were selected for use in the present study. Consequently, it was possible to demonstrate more precisely the effects of restricted human communication on adaptive computer interactions. Specifically, the unrestricted groups (context sensitive and limited response) completed 74.5 percent of the tasks, while the restricted groups (limited human-computer interaction and control) completed only 53.6 percent of the tasks. These results show that restricted communication lowered performance efficiency by more than 20 percent over that of unrestricted communication.

Although a difference among the communication groups was observed in the present study, significant differences among all modes were not found. It is possible that the low number of participants in the current study may have been responsible for the small differences among the communication modes. It is true that the differences among the four modes were in the expected direction, but only the context sensitive and limited response groups differed from the limited human-computer interaction and control groups. Thus, the need to screen such a large portion of the participants for the DC variable may have limited the pool of participants available for the main study, thereby reducing the power to detect these differences.

One could postulate that the performance differences observed in the present study were due to the computer responses. The results, however, show that this was not the case. The differences in the computer's responses, context sensitive responses for the context sensitive interaction group and limited responses for the other three groups, did not have an impact on performance. Thus, participants' task scores were not affected by differences in how the computer communicated with them. This suggests that there may be some level of quality of feedback that is required for efficient communication between a human and a computer, but that there may be no appreciable benefit to exceeding this level. This argument is supported by evidence from the present study which showed that context sensitive responses and limited responses were comparable for performance when participant communication was unrestricted.

It is also possible that the context sensitive feedback was not necessary for this type of interaction. The context sensitive responses were designed to provide the participants with more detailed information in order to give the impression that the computer knew their plans and goals. It was expected that this would improve performance because researchers studying the effects of feedback in communication have shown that feedback influences the formation of common ground between speakers (Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987; Krauss & Weinheimer, 1966; Oviatt & Cohen, 1991). In addition, Kraut et al. (1982) examined the effects of listener responsiveness on conversational effectiveness and found that feedback helped to tailor the conversation to the individual. However, this type of feedback may not be necessary or beneficial for the kinds of tasks that were used in the present study. The structure and order inherent in the experimental stimuli may have provided adequate cues for task completion that minimized the incremental value of context sensitive feedback. Participants completed the questions in a specific order and the computer partner corrected them if they attempted to deviate from that order. It is possible that the participants understood this structure and the computer's role in solving the tasks, thereby minimizing the importance of context sensitive computer responses. The context sensitive responses may not have provided any additional benefit because the participant already believed there was common ground with the computer. In the future, researchers should examine further the role of structure in task scenarios and feedback content in human-computer communication.

Task Complexity

A second goal of the current study was to examine the effects of task complexity on adaptive task interactions. It was expected that task scores would be higher for simple tasks in all communication groups, and that scores across communication modes would be similar for simple tasks, but decrease with greater restriction on communication for complex tasks. The results showed that participants did score higher on simple tasks than on complex tasks and that there was an interaction between complexity and communication mode. Analysis of this interaction showed that for simple tasks, the context sensitive and limited response groups scored significantly higher than the limited human-computer interaction and control groups. There were no significant differences in task scores among the communication groups for complex tasks, although the group differences for both simple and complex tasks followed the same pattern as the main effect (see Figure 2).

The observed pattern was the opposite of what was expected. This may have been due to the participants' ability to communicate effectively about the tasks. It was hypothesized that communication ability would be less important for simple tasks, but this did not happen. Instead, the wording of the simple tasks may have made it relatively easy for the participants to identify and request the information they needed from the computer. Therefore, for the simple tasks, the benefits of increased communication in the context sensitive and limited response groups showed up in higher task scores.

For example, in one of the simple tasks, participants were required to find the phone number for Days Inn Hotels (see Appendix B). A typical exchange between the human and the computer in the context sensitive group was:

> Participant: How do I find hotels?
>
> Computer: You can find the number for Days Inn by using Hotels on the File menu. (Transcript 1C183)

The wording of the simple questions allowed the participants to easily identify the goal of the question and often they would simply rephrase the question that had been presented to them.

In the complex tasks, however, there were more goals and the participants could not simply rephrase the question to obtain the solution. Instead, they had to ask many more questions about the subgoals of the task in order to complete the overall task. For instance, consider the following typical exchange for one of the questions in the complex tasks (see Appendix B) for the context sensitive group:

> Participant: Do I put my name in here?
>
> Computer: No.
>
> Participant: Should I hit enter?
>
> Computer: No.
>
> Participant: How do I get the map?
>
> Computer: You have to use Calculate to make this trip the active route.
>
> Computer: **(Computer takes control of tasks.)** You can display the current route by using Draw Entire Route on the Display Menu.
>
> (Transcript 1C082)

In this example the computer eventually takes control and completes the task because the participant spent too much time trying to figure out how to phrase her request in order to complete the task. These kinds of problems may explain why the interaction with task complexity was not as hypothesized. It is possible that participants were unable to ask for the information they needed to complete the complex tasks effectively. Again, this points to the need to consider the structure of the tasks in human-computer interaction. In addition, it is also possible that given more experience with the tasks, participants might have been able to communicate more effectively. In the future, it may be necessary for researchers to allow the participants to become more experienced with the computer and tasks in order to get a clearer picture regarding the effects of complexity.

Computer Control

Computer control refers to the number of times per minute the computer partner usurped control of the task from the participant. It was expected that as communication restriction increased computer control would also increase. As hypothesized, computer control was lower for the context sensitive and limited response groups than for the limited human-computer interaction group. This is consistent with the results for task score and supports the notion that communication was more efficient in the context sensitive and limited response groups, resulting in less need for the computer partner to take over the tasks.

A finding such as this also highlights the importance of Malin and Schreckenghost's (1992) second criterion for a computer team member - effective communication with human team members. When the participant's ability to communicate was restricted, computer control increased. Unfortunately, the increase in control did not result in improved task scores. In fact, performance suffered.

It is also possible that the nature of the adaptive part of the task affected the computer control results. In the present study, the computer interrupted when the participant was inactive for a period of 30 seconds or had reached a limit of 3 minutes on one task. These rules resulted in a computer partner that was not very aggressive in its behavior and may have contributed to the decrease in performance as computer control increased (see below). When the computer partner eventually did step in and take control, it was after significant periods of time had elapsed. Because the task score reflects the number of tasks completed in a block of time, these lapses would have contributed to lower task scores. When the human and computer could communicate without restriction, computer control was lower and performance was higher. The computer was able to work effectively with the human to complete the tasks with limited need to usurp control.

The hypothesis that computer control would be lower for simple tasks than for complex tasks was also supported. Complex tasks contained more goals and were expected to be more difficult to finish within the allotted time, thus resulting in more computer control. As noted above, participants may have had difficulty figuring out what to ask the computer while completing the complex tasks, resulting in more computer control. This points to the need for efficient communication between the human and the computer (Malin & Schreckenghost, 1992) in order to avoid having the computer dominate interactions in adaptive environments.

It is important to understand that one of the objectives of the present study was to examine discourse patterns between the human and computer. In order to ensure that communication occurred, the computer partner had to allow time for the participant to create a dialogue. If the computer partner had been overly aggressive, it might have completed all the tasks without leaving much opportunity for the participant to communicate. In the future, the effects of various computer intervention times on communication and performance need to be examined.

Correlation of Task Score and Computer Control. It was expected that task score would have a negative correlation with computer control because increased computer control indicates that the participant is having trouble completing the tasks. The results confirmed this hypothesis. The more the computer had to intervene according to the adaptive rules, the lower the task score. As was mentioned above, this correlation may have been influenced by the aggressiveness of the computer partner. However, when this correlation is considered in the context of a positive relationship between words per minute and task score (see below), it lends additional support to the hypothesis that unrestricted communication is beneficial to task completion.

Task Difficulty Ratings

Task difficulty ratings were included primarily to confirm that there was a difference in difficulty between the simple and complex tasks, although they produced some interesting results of their own. The analysis of task difficulty ratings showed that participants rated the tasks as easier in the context sensitive interaction group than in the limited human-computer interaction and control groups. This finding supports the results for task score and computer control (see above). The freedom to speak naturally in the context sensitive interaction group as opposed to restricting speech in the limited human-computer interaction and control groups may have caused participants to feel the tasks were easier. This result agrees with other research that has shown information exchange to be essential to team functioning (Fleishman & Zaccaro, 1992; Salas et al., 1992).

As expected, participants found the simple tasks to be easier than the complex tasks. However, this effect was modified by an interaction with communication mode which showed that the task difficulty ratings differed significantly only for the limited response group (see Figure 3). In this group, complex tasks were rated as more difficult than simple tasks. Thus, although the differences in difficulty between the simple and complex tasks across all the communication modes were in the expected direction, only the difference for the limited response group was significant.

One possible explanation for this result is that it is an artifact of the way the measure was taken. When the tasks were developed, participants in a pilot study provided difficulty ratings after each task was completed. For the pilot study the simple tasks had a lower difficulty rating of M = 2.57, while the complex tasks had a higher difficulty rating of M = 4.05 (see Method Section). In the main study, however, participants could not be interrupted after each task because of the time limit to complete the set of tasks. Therefore, participants rated each set of tasks after the time for working on them had elapsed. Thus, the difference in rating procedure between the pilot and main study may have contributed to the differences in rating scores.

For example, participants tended to rate the first set of tasks as more difficult regardless of whether it was the simple or complex set, because they were just starting to learn the program and how to interact with the computer. In addition, the complex tasks involved the same kinds of activities as the simple sets except that more goals were involved in each task. When making their ratings, pilot participants paid attention to task boundaries because they rated each task separately. The participants in the main study, however, may have been less able to recognize the differences because they rated the tasks collectively at the end of each set. Thus, the results of the difficulty ratings for the pilot study along with the trend in the expected direction for all of the communication groups in the main study provides sufficient evidence for differences between the simple and complex tasks.

Discourse Organization

Communication Restriction

Discourse organization variables were measured in order to gain more information about the effects of communication restriction on adaptive task interactions. It was expected that measures of verbosity, disfluencies, and indices of common ground would decrease as communication restriction increased.

Measures of Verbosity - Words per Minute

The number of words spoken per minute was used as a measure of verbosity and was expected to decrease as communication restriction increased. As expected, the number of words spoken per minute was higher for the context sensitive and limited response groups. This is consistent with the findings of Chapanis and Overbey (1974) who studied different modes of communication while allowing half of the participants to interrupt their partner. They found that participants altered the content of their messages when they had the freedom to interrupt. Specifically, participants exchanged more utterances, the utterances were shorter, and they were exchanged at a faster rate. Thus, the evidence from Chapanis and Overbey's (1974) study along with that of the present study supports the conclusion that the ability to exchange information freely may be important in problem solving.

Correlation with Task Score. It was also expected that task score would increase as communication increased and, in fact, the number of words per minute was correlated with task score. The more the participant communicated with the computer partner, the higher the task score. This correlation and the correlation between task score and computer control (see above) support the hypothesis that conversation between the human and computer benefits successful task completion.

Task Complexity. The results for task complexity showed that words per minute was higher for simple tasks than for complex tasks. This effect was not hypothesized and might have occurred for the same reasons already discussed for the communication mode by task complexity interaction for task score (see above). The participants seemed to have more trouble communicating with the computer about the complex tasks, resulting in more planning of utterances and fewer overall words spoken during the complex tasks.

Measures of Verbosity - Mean Length of Utterances

The mean length of utterances was used as another measure of participants' verbosity and was expected to decrease as communication restriction increased. As hypothesized, the mean length of utterances for the context sensitive and limited response groups was higher than for the limited human-computer interaction group. This was clearly the result of restricted communication in the limited human-computer interaction group. When participants could speak freely, they chose to use longer sentences rather than restrict their speech. These results support the findings of Kennedy et al. (1988) and Cohen et al. (1982) who found that participants do not change their style of interaction in response to the computer's responses and that they expect the computer to have conversational ability. In other words, even though the computer's responses were restricted in the limited response group, the participants did not voluntarily modify their style of speech to match that of the computer. They seemed to assume the computer could understand what they were saying.

Disfluencies

Disfluencies, or mistakes in the participants' speech, were used in examining discourse structure. It was predicted that disfluencies would decrease as communication restriction increased and that disfluencies would be lower for simple tasks than complex tasks. These hypotheses were based on Oviatt's (1995) findings that increased planning demands and longer utterances were related to increases in spoken disfluencies during human-computer interaction. The results showed no effects for disfluencies except for an interaction between communication mode and desire for control which was not supported by an analysis of the simple effects (see Figure 4).

One could argue that these results were due to very low levels of disfluencies. Disfluency levels in the present study, however, averaged 2 per 100 words spoken and were comparable to those found by Oviatt (1995) in her study of human-computer communication on a simulated service transaction computer system. Moreover, as in the present study, Oviatt (1995) also found disfluency levels that were much lower for human-computer communication than for comparable human-human communication. One possible explanation for why disfluency effects were not observed in the present study is that participants may not have generated enough long utterances to produce disfluencies due to the nature of the tasks. On the other hand, it is also possible that participants were more careful when planning their utterances due to the novel nature of the tasks.

Indices of Common Ground - Confirmations

Confirmations refer to the number of times the participant confirmed the computer partner's last utterance. This measure was used as an index of common ground. Confirmations were significantly lower for the context sensitive and limited response groups than for the limited human-computer interaction group. However, it was hypothesized that there would be more confirmations (indicating more common ground) in the context sensitive and limited response groups due to the more "human" nature of the interaction resulting from less restriction on communication. It is possible, however, that the high levels of confirmations in the limited human-computer interaction group were the result of the experimental protocol.

In the ice breaking session for the limited human-computer interaction group, participants used two statements that were designed to illustrate the differences in the keywords ("Help system colors?" and "How system colors?"). In the experimental session, many of the participants continued to use this rather inefficient way of asking for information; they used both statements each time they wanted a piece of information. In the transcript analysis, the second statement of this interaction was coded as a confirmation if it immediately followed the computer's answer to the first statement because the participant was confirming that he understood the first answer by following it with a relevant next turn. Consequently, this type of interaction may have resulted in inflated confirmation levels for the limited human-computer interaction group.

Oviatt and Cohen (1991) found confirmation levels to be about 18 percent of total verbal output in their study of human-human dialogues. In contrast, confirmation levels in the present study averaged about 3 percent of total verbal output. Other studies on human-computer interaction have not examined confirmation levels, therefore the hypotheses for confirmations were based on the results of human-human communication studies. In the present study, the prerecorded computer responses may have limited opportunities for confirmations, resulting in lower levels than those observed in human-human communication. Therefore, it may be that confirmations are not a central component of human-computer communication. On the other hand, human-computer communication might actually benefit if the use of confirmations is encouraged through system design. In studies of human communication this type of feedback has been shown to be important for communication efficiency (Krauss & Bricker, 1967; Oviatt & Cohen, 1991). The use of confirmation as a feedback tool in human-computer communication should be examined in more detail in the future.

Indices of Common Ground - Anaphoric Reference

Anaphoric reference reflects elements in the participants' speech that refer back to earlier elements in the conversation. Anaphor was used as an indication of common ground and was expected to decrease as communication restriction increased. Contrary to expectations, anaphoric reference did not differ for the communication modes. The means, however, were in the expected direction.

Kennedy et al. (1988) found that anaphor was used less by participants who believed they were interacting with a computer system than by participants who believed they were interacting with another human. However, in their study anaphor was measured only for early and late conversational exchanges. Therefore, the total amount of anaphor is not available from their study. The levels of anaphor in the present study were low compared to human-human communication (Brennan, 1991) and lower than those found in studies of human-computer communication (Brennan, 1991; Guindon et al., 1987). Perhaps, the nature of the tasks reduced the need for anaphor in the present study. The scenarios were relatively independent and did not require participants to refer back to earlier tasks in order to complete the task on which they were presently working. The task complexity effect discussed below also supports the hypothesis that task structure affected anaphoric reference.

Correlation of Task Score and Anaphoric Reference. Contrary to expectations, there was an inverse relationship between task score and the occurrence of anaphor. It was hypothesized that anaphor would be positively related to task score because indications of more common ground should lead to higher task scores.

Again, this result might be due to the nature of the tasks. In the present study if a participant used anaphor it often meant he was having trouble solving a problem and therefore kept referring to it. This could explain why task score decreased as anaphor increased. The use of anaphor revealed that the participant thought he had established common ground with the computer, but the inability to complete the task in question led to lower task scores. For instance, in the following exchange a participant in the context sensitive group attempts to complete one of the complex tasks:

Participant: Okay, I'm tryin a; find a road my friend lives in in Missouri from [Springfield]

Computer: [You need to] use State Road Network on the Display Menu

Participant: Use what again? **(Anaphor - Ellipse)**

Computer: You need to use SRN on the D Menu

Participant: Thank you.

Participant: Is that I-(8)? **(Anaphor)**

Participant: Where does it tell you the name of the road? **(Anaphor - 2)**

Computer: The information is on the screen.

(Transcript 1C034)

In this example, the participant continues to refer to the same task because he cannot determine how to do the task correctly. If the conversations in the present study had been longer or not focused on relatively short task scenarios then the use of anaphor might have resulted in higher task scores. This aspect of task structure clearly needs to be examined in the future.

Task Complexity. There was an unexpected effect for task complexity that showed that anaphoric reference was lower for simple tasks than for complex tasks. Again, this could be due to the task structure. Most of the simple tasks had only one goal; therefore, once the participant asked the computer about the task it was not referred to again, resulting in lower levels of anaphoric reference.

The main effect for task complexity supports this assertion because anaphor was higher for complex tasks which had more goals than simple tasks. This would lead to more inquiries in order to complete a particular problem and possibly a higher incidence of anaphor. The use of anaphor could also mean a participant was having trouble with a problem and kept referring to it. It would be reasonable to expect that the complex tasks would have higher levels of anaphor for this reason as well.

Indices of Common Ground - Complex Referring Expressions

Complex referring expressions occur when the participant elaborates or uses complex noun phrases and is an index of common ground. It was hypothesized that complex referring expressions would increase as communication restriction increased. There were, however, no significant effects for complex referring expressions although the means were in the expected direction. Guindon et al. (1987) found that over 50 percent of the utterances in their study of human dialogues with a simulated computer advisor contained complex referring expressions. They theorized that users generated these expressions because the users believed they shared little context with the computer. Oviatt and Cohen (1991) found a similar phenomenon in human monologues. Therefore, it was hypothesized that complex referring expressions would increase as communication restriction increased. In the present study, however, the levels of complex referring expressions were very low. There were only about 3 complex referring expressions for every 100 words spoken by the participants. This is extremely low when compared to the levels found by Guindon et al. (1987) and Oviatt and Cohen (1991).

It is possible that the levels of complex referring expressions were also influenced by the structure of the tasks. For example, participants in the human-computer interaction group were required to follow rules for generating utterances, however, the resulting utterances would have contained no complex noun phrases. If the participants in that group generated every utterance correctly the level of complex noun phrases would have been zero because the keywords that participants were required to use did not contain complex noun phrases. However, the levels of complex noun phrases were greater than zero, and the complex referring expressions level (complex noun phrases and elaborations) for the limited human-computer interaction group was no different from that of the other groups. This indicates that the participants in the limited human-computer interaction group found it difficult to follow the rules they were given and may have believed that they shared little context with the computer. It is important to remember, however, that the levels for all the communication groups were very low.

Correlation of Task Score and Complex Referring Expressions. It was hypothesized that there would be a negative relationship between task score and complex referring expressions because as the incidence of complex referring expressions increases, task score should decline due to a decrease in common ground. There was no relationship found between task score and complex referring expressions. Levels of complex referring expressions were very low and there may not have been enough variability to produce a relationship.

Discourse Organization Summary

On the whole, the discourse organization variables all had low levels of occurrence. However, as expected, measures of verbosity (words per minute and mean length of utterances) decreased as communication restriction increased. In addition, the correlation between words per minute and task score confirmed that performance improved with increases in verbosity. As discussed above, some of the discourse organization results may have been affected by task structure. For example, the computer responses may have provided fewer opportunities for confirmations thereby resulting in low levels of this variable. In addition, the task scenarios may have reduced the need for anaphoric reference, thus affecting the levels of this variable as well as its correlation with task score. The levels of complex referring expressions may have been affected similarly. Therefore, the predictions for the effects of communication restriction on measures of verbosity were supported. The other discourse organization measures, however, showed unexpected results, some of which may have been due to task structure as explained above.

In the present study, it was expected that communication in the context sensitive group would be similar to human communication, but that the other groups would differ more from human communication as restriction increased. This was not observed, however. In all groups, the levels of discourse that are common in human conversation (Brennan, 1991; Oviatt & Cohen, 1991) were uncommon in human-computer communication. Although the present study did not include a human-human communication comparison group, the results suggest that some of the processes that make human communication so efficient are absent in human-computer communication in its present form. For example, researchers have shown that feedback, such as confirmations, is related to the speaker's ability to communicate efficiently (Krauss & Bricker, 1967; Oviatt & Cohen, 1991). Indices of common ground, such as anaphoric reference, have also been linked to communication efficiency (Brennan, 1991; Kennedy et al., 1988). Given the low occurrence of these variables in the present study, it is necessary to expand the research on human-computer communication in order to better understand the organization of human-computer dialogue.

Another important finding from the discourse organization results is that the participants voluntarily restricted their use of language (as compared to human-human communication) in the context sensitive and limited response groups. This may support the possibility of developing successful limited natural language systems (Chin, 1984; Malhotra & Sheridan, 1976). On the other hand, it is important not to ignore the poor performance of the limited human-computer interaction group. When participants were asked to restrict their language according to certain rules, performance suffered. It is very important to determine when and how people restrict their language when communicating with a computer and the potential impact it might have in an adaptive environment.

The modifications that participants in the present study made in their discourse when working with a computer partner as opposed to a human partner, supports the conclusion of O'Conaill et al. (1993) that the disruptions caused by mediated communication may result in differences in communication ability. In addition, as other researchers have observed (Cohen et al., 1982; Kennedy et al., 1988), participants in the present study did not modify their style of interaction as they gained more experience with the computer. They adopted a way to communicate with the computer and stayed with it as the experiment progressed. It may be that people find it difficult to modify their "style" of speaking. For example, in the present study there were participants in every communication group who stopped communicating with the computer rather than modify a strategy that wasn't working for them. This may have also contributed to the difficulties experienced by those in the limited human-computer interaction group. In order to use the keywords, participants in the limited human-computer interaction group had to modify their normal speaking style in a regimented way. Many of the participants in this group continued to make mistakes in formulating their utterances throughout the entire experiment.

Participant Ratings

Communication Restriction

Participant ratings of their interaction with the computer were gathered to provide additional evidence about the effects of communication restriction in an adaptive task environment. It was predicted that as communication restriction increased participant ratings would become more negative.

Ability to Communicate

Participants were asked to rate their ability to communicate with the computer partner. As hypothesized, their ability to do so was rated as easier by the context sensitive and limited response groups than by the limited human-computer interaction group. These ratings provide additional evidence that restricted communication was more difficult for participants.

Helpfulness

Participants also rated the helpfulness of the computer partner. There were no significant effects for helpfulness. It was expected that ratings of helpfulness would decrease as communication restriction increased. The ratings may have been similar across communication groups because all participants were comparing how well they did with the computer's help to how well they thought they would have done without any help at all. Moreover, because each participant only experienced one kind of help from the computer they all rated the computer as fairly helpful. Greater discrepancies might have been observed if participants had the opportunity to experience different modes of communication.

Enjoyment

Participants were also asked to rate how well they enjoyed interacting with the computer. There was no communication mode effect for enjoyment ratings, however, the means were in the expected direction. It was hypothesized that enjoyment would decrease as communication restriction increased. Again, because communication mode was a between-subjects variable all of the participants may have found their particular interactions with the computer to be enjoyable. Most participants found the idea of a talking computer very interesting and therefore enjoyed the session. In addition, there was an interaction with desire for control which is discussed below.

Desire for Control

A third goal of the present study was to examine the effects of desire for control on adaptive task interactions. It was hypothesized that high-DC participants would score higher on complex tasks than low-DC participants, but there would be little difference in task scores for simple tasks. In addition, it was hypothesized that high-DC participants would have more negative opinions about the interaction than low-DC participants.

Communication Restriction

Task Score

Based on Burger's (1985) research on achievement-related behavior in high- and low-DC individuals, it was hypothesized that there would be an interaction between task complexity and desire for control such that high-DC participants would perform better on complex tasks than low-DC participants, but there would be little difference in performance for simple tasks. There is no evidence, however, of any effect of desire for control on task score in this experiment.

Confirmation of these findings would be desirable because the results do not support Burger's (1985) research on desire for control and there does not seem to be an explanation for participants approaching these tasks any differently than other achievement-related tasks. In the past, researchers studying desire for control (Burger, 1985; Burger & Cooper, 1979) have used a less powerful median split method and have found significant effects. This suggests that the lack of effects here may be valid.

It is possible that the complex tasks in the present study did not provide enough challenge to illicit greater effort from the high-DC participants. Past research, however, has found differences for DC with only minor differences between tasks. For example, Burger (1985) used a proofreading task to examine the responses of high- and low-DC participants. The more challenging condition in his experiment consisted of the proofreading task plus a word counting task. The results showed that high-DC participants proofread more lines in the more challenging condition, but that there were no differences for DC in the less challenging condition. The manipulation used in Burger's (1985) study does not seem particularly strong, yet he still found significant effects due to DC. Given the established differences in difficulty between the task sets used in the present study, the complex tasks should have been challenging enough to reveal a difference in DC.

It should also be noted, however, that participants in the present study did not perceive differences between the simple and complex tasks to be excessive. In fact, the differences for task complexity were attributed solely to the limited response group, and the rating differences between simple and complex tasks were not very large for any of the other groups. Therefore, if high-DC subjects did not perceive the complex tasks as challenging, they might not have been motivated to work harder on them. Perhaps if participants had been specifically cued as to which task sets were simple and complex, a performance difference would have emerged.

Another possibility for the lack of differences in performance for DC is that the high-DC participants may not have viewed the computer as a threat to their control. If high-DC participants considered the computer to be a tool and not another person to whom they were relinquishing control, they might have been more willing to accept its help. In fact, the participants' ratings of the computer's helpfulness support this explanation and are discussed below. Perhaps desire for control does not apply to working with a computer partner in the same way it does with a human. This is an issue that should be explored in more detail.

Although there were no performance differences found for DC in the present study, it should be noted that there were differences in the way high- and low-DC participants perceived the tasks.

Ability to Communicate

Participants rated their ability to communicate with the computer partner as part of a questionnaire designed to provide additional evidence about the effects of communication restriction on adaptive task interactions. It was predicted that as communication restriction increased, participant ratings would become more negative. Therefore, it was expected that high-DC participants would have more negative opinions of the interaction due to their desire to control the situation. Contrary to this expectation, high-DC participants rated ability to communicate as easier than low-DC participants. This might be due to the high-DC tendency to want to master a situation. High-DC participants might have had more motivation to do well. However, because their motivation did not translate into increased performance, the higher ratings of the high-DC participants over the low-DC participants were likely the result of perceptions of performance.

Helpfulness

Participants also rated the computer partner's helpfulness. It was expected that high-DC participants would perceive the computer as less helpful than low-DC participants because of their unwillingness to accepting help from others (Burger, 1985); however, there were no significant effects for helpfulness. As discussed above, this supports the assertion that high-DC participants did not see the computer as a threat to their control because they did not reject its help and rated its helpfulness similar to that of low-DC participants.

Enjoyment

Participants were asked to rate their enjoyment of the interaction with the computer as well. The results showed an unexpected interaction between communication mode and desire for control on the enjoyment ratings (see Figure 5). High-DC participants reported more enjoyment than low-DC participants in the limited human-computer interaction group and high-DC participants reported less enjoyment than low-DC participants in the control group.

Although this interaction was not predicted, it might be explained by the way high- and low-DC participants perceived the tasks. The limited human-computer interaction group was faced with the most challenging condition (as shown by task score) in which participants were still able to actively communicate with the computer. It is possible that high-DC participants in this group may have enjoyed the interaction more because of the challenge of formulating appropriate utterances to communicate with the computer. By contrast, the control group had no control over their interaction with the computer because they could not communicate in any way. This inability to control when and how the computer would intervene may have caused high-DC participants to rate enjoyment as lower in this group.

Conclusions

Communication Restriction

The present study examined the effects of communication restriction on adaptive task interactions. The context sensitive and limited response interaction groups performed better than the limited human-computer interaction and control group. In addition, the present study is one of the first to use a finer-grained performance measure which demonstrated large performance differences for unrestricted and restricted communication modes. In the present study, the groups who could communicate freely performed better than those where communication was restricted or denied. This large difference between unrestricted and restricted groups was apparent for other dependent measures as well (i.e., task difficulty ratings, computer control, words per minute, mean length of utterances, confirmations, and ability to communicate).

Another important finding from the present study was that differences in how the computer responded (context sensitive or limited) did not affect task score. This suggests that participants do not expect computer feedback to be exactly like that of communicating with another human. In the present study, the addition of context sensitive information to the computer responses did not result in changes in performance or discourse structure. Thus, there may be a level of feedback quality that is required for efficient communication between a human and a computer, but exceeding this level may not be beneficial.

The results for computer control showed that as communication restriction increased, computer control also increased. These results support the assertion that effective communication between the human and computer is essential (Malin & Schreckenghost, 1992). The results also showed that performance decreased as computer control increased. It is possible that the nature of the adaptive part of the task contributed to this relationship because the computer partner was not very aggressive in its behavior.

The results for discourse organization highlight the differences between human communication and human-computer communication. Measures of verbosity showed the expected effects; however, the other discourse measures produced a pattern that was different than that of human communication (Brennan, 1991; Oviatt & Cohen, 1991). It is possible that some of the processes that contribute to successful human communication may be absent in human-computer communication.

Although human-computer communication in the present study was different from that typically observed among humans, participants did not modify their style of interaction to match that of the computer. In addition, when participants were asked to produce utterances according to specific rules in the limited human-computer interaction group, performance was similar to the control group where no communication occurred at all.

Task Complexity

Task complexity was also examined and was expected to amplify the basic effects for communication restriction. Participants scored higher on simple tasks than complex tasks. However, the communication mode by task complexity interaction showed that there were no differences in task score among the communication modes for complex tasks, but that the context sensitive and limited response groups scored higher than the limited human-computer and control groups on simple tasks. This may have been due to the difficulty participants had in formulating questions for the complex tasks. The results for words per minute showed a similar pattern with words per minute being higher for simple tasks than complex tasks, possibly due to the same problem.

Desire for Control

Desire for control was also expected to amplify the basic effects for communication restriction. There were, however, very few effects for this variable. There were no performance effects observed and, as discussed above, this may mean that high-DC participants did not perceive the computer to be a threat to their control. In addition, high-DC participants rated their ability to communicate as easier than low-DC participants, possibly due to a preference to master challenging situations. There was an interaction of communication mode and desire for control for enjoyment as well. Again, this interaction may have been the result of differences in the DC response to challenges and degree of control over the situation as explained above. Future research on this trait would be beneficial to determining its utility in assigning partners to adaptive systems.

Final Thoughts

To date, adaptive systems have not relied on explicit communication between the human and the computer (Bushman et al., 1993; Chu et al., 1995; Hammer & Small, 1995; Mason, 1986). Instead, the systems have used intent inferencing which relies on representations of operator plans based on operator actions. Although researchers are beginning to explore the possibility of a human-computer team (Hammer & Small, 1995; Malin & Schreckenghost, 1992; Malin et al., 1991; Scerbo, 1994) there has been little exploration of explicit communication between the human and computer teammates.

Researchers have shown that information exchange is essential to an efficient team (Fleishman & Zaccaro, 1992; Salas et al., 1992). In addition, the differences in performance that result from different modes of communication have been documented (Chapanis & Overbey, 1974; Chapanis et al., 1972; Chapanis et al., 1977; Krueger & Chapanis, 1979; Ochsman & Chapanis, 1974; Weeks & Chapanis, 1976). To date, however, there has been no research on human-computer communication in an adaptive interface using speech as the medium of communication.

The results of the present study show that human-computer communication in an adaptive environment appears to be different from human communication. Despite the fact that the context sensitive interaction condition was designed to simulate the richness of human communication, participants' speech differed significantly from the discourse organization that has been observed in human communication (Brennan, 1991; Oviatt & Cohen, 1991). In addition, differences in how participants could communicate produced large differences in performance. In fact, when participants were asked to use a particular style of modified speech (i.e., limited human-computer interaction), the results were similar to the control group where no communication occurred at all.

It was also noted that the timing of the computer's interventions may have affected performance. The adaptive rules that were required to allow a dialogue to develop between the participant and the computer resulted in a computer partner that was not very aggressive and, ultimately, this may have contributed to the decrease in performance as computer control increased.

Therefore, the findings from the present study show that both communication restriction and adaptive timing contribute to performance. In the future, it will be necessary to consider further the differences between human cooperation and human-computer cooperation in an adaptive environment with a speech interface. Such information will no doubt be valuable because optimizing human-computer communication will be necessary to produce efficient teamwork between the human and the computer.

The results of restricting communication in the present study support Scerbo's (1996) assertion that the success of adaptive automation will depend on the methods of information exchange that are available to the human-computer team. It seems likely that the inclusion of other methods of communication in addition to speech would increase further the efficiency of communication between the human and computer. For example, humans are often able to use gestures and other types of nonverbal communication. Perhaps adding a touch screen or a mouse to the speech interface would improve human-computer communication. Support for this hypothesis has already been shown by Ochsman and Chapanis (1974) who found that performance in human communication increased as the richness of the communication modes available increased.

The results of the present study point to the need for further exploration of the issues surrounding human-computer communication and its implementation in adaptive systems. Adaptive systems rely on the human and machine working together as partners in order to maintain optimal operation of the system (Scerbo, 1994) and the interface between the human and computer will undoubtedly contribute significantly to the success of that partnership.

Researchers and designers are currently in the position to consider how the technology of adaptive automation might be best implemented because it is still in the early stages of development. Thus, it may be possible to build the proper "electronic crew member" of the future by considering the appropriate usability issues today.

REFERENCES

Aaronson, A., & Carroll, J. (1987). The answer is in the question: A protocol study of intelligent help. Behaviour and Information Technology, 6, 393-402.

Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in the dyad. Human Relations, 28, 289-304.

Brennan, S. E. (1991). Conversation with and through computers. User Modeling and User-Adapted Interaction, 1, 67-86.

Buick, F. (1989). +Gz protection in the future - Review of scientific literature. Technical report DCIEM 89-RR-47. Defense and Civil Institute of Environmental Medicine, Downsview, Ontario, Canada.

Burger, J. (1985). Desire for control and achievement-related behaviors. Journal of Personality and Social Psychology, 48, 1520-1533.

Burger, J. & Cooper, H. (1979). The desirability of control. Motivation and Emotion, 3, 381-393.

Bushman, J., Mitchell, C., Jones, P., & Rubin, K. (1993). ALLY: An operator's associate for cooperative supervisory control systems. IEEE Transactions on Systems, Man, & Cybernetics, 23, 111-128.

Carroll, J., & Aaronson, A. (1988). Learning by doing with simulated intelligent help. Communications of the ACM, 31, 1064-1079.

Carroll, J., & McKendree, J. (1987). Interface design issues for advice-giving expert systems. Communications of the ACM, 30, 14-31.

Chapanis, A., & Overbey, C. M. (1974). Studies in interactive communication III: Effects of similar and dissimilar communication channels and two interchange options on team problem solving. Perceptual and Motor Skills, 38, 343-374.

Chapanis, A., Ochsman, R. B., Parrish, R. N., & Weeks, G. D. (1972). Studies in interactive communication I: The effects of four communication modes on the behavior of teams during cooperative problem solving. Human Factors, 14, 487-509.

Chapanis, A., Parrish, R. N., Ochsman, R. B., & Weeks, G. D. (1977). Studies in interactive communication II: The effects of four communication modes on the linguistic performance of teams during cooperative problem solving. Human Factors, 19, 101-126.

Chin, D. (1984). An analysis of scripts generated in writing between users and computer consultants. In Proceedings of the National Computer Conference: Vol. 53 (pp. 637-642). Reston, VA: AFIPS Press.

Chu, R., Mitchell, C., & Jones, P. (1995). Using the operator function model and OFMspert as the basis for an intelligent tutoring system: Towards a tutor/aid paradigm for operators of supervisory control systems. IEEE Transactions on Systems, Man, & Cybernetics, 25, 1054-1075.

Clark, H., & Brennan, S. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley (Eds.), Socially shared cognition. Washington, DC: APA.

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

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

Cohen, P., Perrault, C., & Allen, J. (1982). Beyond question answering. In W. Lehnert & M. Ringle (Eds.), Strategies for natural language processing (pp. 245-274). Hillsdale, NJ: Erlbaum.

Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. Journal of Consulting Psychology, 24, 349-354.

Fleishman, E. A., & Zaccaro, S. J. (1992). Toward a taxonomy of team performance functions. In R. W. Swezey & E. Salas (Eds.), Teams: Their training and performance (pp. 31-56). Norwood, NJ: Ablex Publishing.

Gluckman, J., Morrison, J., & Deaton, J. (1991). Complex task performance as a basis for developing cognitive engineering guidelines in adaptive automation. In Proceedings of the Human Factors Society 35th Annual Meeting (pp. 116-120). San Francisco, CA: Human Factors Society.

Gould, J., Conti, J., & Hovanyecz, T. (1983). Composing letters with a simulated listening typewriter. Communications of the ACM, 26, 295-308.

Guindon, R., Shuldberg, K., & Connor, J. (1987). Grammatical and ungrammatical structures in user-advisor dialogues: Evidence for sufficiency of restricted languages in natural language interfaces to advisory systems. In Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics (pp. 41-44). Stanford, California.

Hammer, J. M., & Small, R. L. (1995). An intelligent interface in an associate system. In W. B. Rouse (Ed.), Human/technology interaction in complex systems, Vol. 7 (pp. 1-44). Greenwich, CT: JAI Press.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. Journal of Experimental Psychology: General, 116, 26-37.

Jaffe, J. & Feldstein, S. (1970). Rhythms of dialogue. New York: Academic.

Johannesen, L. J., Cook, R. I., & Woods, D. D. (1994). Cooperative communications in dynamic fault management. Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting (pp. 225-229). Santa Monica, CA.

Kennedy, A., Wilkes, A., Elder, L., & Murray, W. S. (1988). Dialogue with machines. Cognition, 30, 37-72.

Kieras, D. (1988). Towards a practical GOMS model methodology for user interface design. In M. Helander (Ed.), Handbook of human-computer interaction (pp. 135-157). Amsterdam: North Holland.

Kiesler, S. & Sproull, L. (1992). Group decision making and communication technology. Organizational Behavior and Human Decision Processes, 52, 96-123.

Krauss, R. M., & Bricker, P. D. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. The Journal of the Acoustical Society of America, 41, 286-292.

Krauss, R. & Fussell, S. (1990). Mutual knowledge and communication effectiveness. In J. Galegher, R. Kraut, & C. Egido (Eds.), Intellectual teamwork (pp. 111-146). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Krauss, R. M., & Weinheimer, S. (1964). The effect of feedback on changes in reference phrase. Conference presentation, Psychonomic Society, Niagra Falls, Ontario, Canada.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. Journal of Personality and Social Psychology, 4, 343-346.

Kraut, R., Lewis, S., & Swezey, L. (1982). Listener responsiveness and the coordination of conversation. Journal of Personality and Social Psychology, 43, 718-731.

Krueger, G. & Chapanis, A. (1980). Conferencing and teleconferencing in three communication modes as a function of the number of conferees. Ergonomics, 23, 103-122.

Malhotra, A., & Sheridan, P. (1976). Experimental determination of design requirements for a program explanation system (Res. Rep. RC 5831). Yorktown Heights, NY: IBM.

Malin, J. T., & Schreckenghost, D. L. (1992). Making intelligent systems team players: Overview for designers. NASA Technical Memorandum 104751, Johnson Space Center, Houston, TX.

Malin, J., Schreckenghost, D., Woods, D., Potter, S., Johannesen, L., Holloway, M., & Forbus, K. (1991). Making intelligent systems team players: Case studies and design issues. Vol. 1: Human-computer interaction design; Vol. 2: Fault management system cases. NASA Technical Report 104738, Johnson Space Center, Houston, TX.

Mason, M. (1986). Adaptive command prompting in an on-line documentation system. International Journal of Man-Machine Studies, 25, 33-51.

Nagel, D. C. (1988). Human error in aviation operations. In E.C. Wiener & D.C. Nagel (Eds.), Human Factors in Aviation. San Diego, CA: Academic Press.

Newell, A., Arnott, J., Carter, K., & Cruickshank, G. (1990). Listening typewriter simulation studies. International Journal of Man-Machine Studies, 33, 1-19.

Newell, A., Arnott, J., Dye, R., & Cairns, A. (1991). A full speed listening typewriter simulation. International Journal of Man-Machine Studies, 35, 119-131.

Ochsman, R. B., & Chapanis, A. (1974). The effects of ten communication modes on the behavior of teams during co-operative problem-solving. International Journal of Man-Machine Studies, 6, 579-619.

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

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. Computer Speech and Language, 9, 19-35.

Oviatt, S. L., & Cohen, P. R. (1991). Discourse structure and performance efficiency in interactive and non-interactive spoken modalities. Computer Speech and Language, 5, 297-326.

Oviatt, S., Cohen, P., & Wang M. (1994). Toward interface design for human language technology: Modality and structure as determinants of linguistic complexity. Speech Communication, 15, 283-300.

Pierrehumbert, J. (1983). Automatic recognition of intonation patterns. In Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics. (pp. 85-90). Cambridge, MA.

Rotter, J. (1966). Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 80, Whole Number 609.

Rouse, W. B. (1988). Adaptive aiding for human/computer control. Human Factors, 30, 431-443.

Rutter, D. & Stephenson, G. (1977). The role of visual communication in synchronising conversation. European Journal of Social Psychology, 7, 29-37.

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

Salas E., Dickinson, T.L., Converse, S.A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R.W. Swezey & E. Salas (Eds.), Teams: Their training and performance (pp. 3-29). Norwood, NJ: Ablex Publishing.

Sarter, N., & Woods D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Human Factors, 37, 5-19.

Scerbo, M. W. (1994). Implementing adaptive automation in aviation: The pilot-cockpit team. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 249-255). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 37-63). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. London: Wiley.

VanKatwijk, A. F., VanNes, F. L., Bunt, H. C., Muller, H. F., & Leopold, F. F. (1979). Naive subjects interacting with a conversing information system. IPO Annual Progress Report, 14, 105-112.

Waibel, A. (1988). Prosody and Speech Recognition. London: Pitman Publishing, Ltd.

Weeks, G. & Chapanis, A. (1976). Cooperative versus conflictive problem solving in three telecommunication modes. Perceptual and Motor Skills, 42, 879-917.

Whinnery, J. E. (1989). Observations on the neurophysiologic theory of acceleration (+Gz) induced loss of consciousness. Aviation, Space, and Environmental Medicine, 6, 589-593.

Wiener, E. (1989). Human factors of advanced technology ("glass cockpit") transport aircraft (Technical Report 177528). Moffett Field, CA: NASA Ames Research Center.

APPENDIX A
DESIRABILITY OF CONTROL SCALE

Using the 7-point scale below, indicate the extent to which each of the following statements applies to you. That is, if the statement always applies to you, place a 7 in the appropriate space. If the statement doesn't apply to you at all, place a 1 in the appropriate space. Use the numbers 2 through 6 to indicate partial agreement.

1	2	3	4	5	6	7
Doesn't Apply To Me At All						Always Applies To Me

_____ 1. I prefer a job in which I have a lot of control over what I do and when I do it.

_____ 2. I enjoy political participation because I want to have as much say in running government as possible.

_____ 3. I try to avoid situations where someone else tells me what to do.

_____ 4. I would prefer to be a leader rather than a follower.

_____ 5. I enjoy being able to influence the actions of others.

_____ 6. I am careful to check everything on an automobile before leaving on a long trip.

_____ 7. Others usually know what is best for me.

_____ 8. I enjoy making my own decisions.

_____ 9. I enjoy having control over my own destiny.

_____ 10. I would rather someone else took over the leadership role when I'm involved in a group project.

1	2	3	4	5	6	7
Doesn't Apply To Me At All						Always Applies To Me

_____ 11. I consider myself to be generally more capable of handling situations than others are.

_____ 12. I'd rather run my own business and make my own mistakes than listen to someone else's orders.

_____ 13. I like to get a good idea of what a job is all about before I begin.

_____ 14. When I see a problem, I prefer to do something about it rather than sit by and let it continue.

_____ 15. When it comes to orders, I would rather give them than receive them.

_____ 16. I wish I could push many of life's daily decisions off on someone else.

_____ 17. When driving, I try to avoid putting myself in a situation where I could be hurt by someone else's mistake.

_____ 18. I prefer to avoid situations where someone else has to tell me what it is I should be doing.

_____ 19. There are many situations in which I would prefer only one choice rather than having to make a decision.

_____ 20. I like to wait and see if someone else is going to solve a problem so that I don't have to be bothered by it.

APPENDIX B
COMPUTER TASKS

SIMPLE TASKS:

Please complete the 18 tasks below as quickly and accurately as possible. Do the tasks in the order they are given. You will have 10 minutes to work on this section. You may not complete all the tasks. You will be told when to stop.

1. When was Fayetteville, AK settled?____________

2. What is the total number of miles for the active route?____________

3. What is the phone number for Days Inn Hotels?____________

4. Update your traveling speeds so that Travel Planner can estimate your travel time. You travel at 70 mph on the Interstate, 55 mph on toll roads, 50 mph on US/State Roads, and 40 mph on local roads.

5. What is the total number of road segments for the active route?_________

6. What is the phone number for Avis Car Rental?____________

7. Update your gas consumption so that Travel Planner can estimate your travel costs. Gas currently costs $1.50 per gallon and your car gets 30 miles per gallon.

8. What is the number of states traveled in for the active route?___________

9. Locate Buffalo, WY and make an "**X**" on the map below where it is located.

10. Display the entire active route on the screen and then draw it on the map below.

11. What is the fuel cost for the active route?____________

12. Using the map below, fill in the names of the cities in Florida that the active route goes through.

13. Does I95 cross the active route in Georgia?____________

14. What is the total travel time for the active route?______________

15. Locate road I70 in Colorado and draw a line on the map below where it runs.

16. What road runs between Trinidad, CO and Raton, NM?____________

17. In what city does I25 cross I70 in Colorado?____________

18. What method of route classification was used to calculate the active route?____________

COMPLEX TASKS:

Please complete the 13 tasks below as quickly and accurately as possible. Do the tasks in the order they are given. You will have 20 minutes to work on this section. You may not complete all the tasks. You will be told when to stop.

1. Ilene Adeb recently made a trip from Grafton, WV to Dodgeville, WI. She saved her trip as ROAD.TRP. Open the file for her trip and display the entire route on the screen. Make an "**X**" on the map below where her trip ends.

2. Ilene used the Quickest Method of calculating her route. How many miles is her trip?___________

   Recalculate her route using the Shortest Method. How many miles is her trip now?___________

3. In what city in South Dakota do US212 and US281 intersect?___________

4. You are thinking of taking a trip from Santa Fe, NM to Ann Arbor, MI and want to know how far it is. You don't need to save the information at this point.

   Calculate this trip using the Shortest Method. How far is it?_________

   Calculate this trip using the Quickest Method. How far is it?__________

5. What is the name of the first road you will travel on during this trip?___________

   How much time will you spend on this road?____________

   You realize that your grandmother will be driving and she only travels at 40 mph on the Interstate, 30 mph on toll roads, 25 mph on US/State roads, and 20 mph on local roads. Update your traveling speeds to reflect this. How long will you be on the first road now?___________

6. How much will the trip cost?___________

   You forgot to take into account the recent gas crisis. Gas currently costs $2.50 per gallon. Update your gas consumption to reflect this. How much will the trip cost now?___________

7. You'd like to visit a friend in Missouri on the way. She lives in Collins. What road will you take from Springfield, Missouri to Collins, Missouri?________

8. You receive a visit from the police. Your friend John is a suspect in a crime and they want to know if you know where he is. He saved a trip under the name JOHN.TRP. Where is he going?___________

   What is his License Plate #?____________

   The police would like you to show them his route on the screen. Display it and make an "**X**" on the map below where his trip ends.

9. The police would like to intercept John in Texas. Using the map below, fill in the names of the cities John will pass through in Texas.

10. The police will travel 90 mph on all roads. Update your traveling speeds to reflect this and calculate a route from Virginia Beach, VA to Canadian, TX using the Shortest Method.

    How long will it take the police?____________

11. If the police take US287 in Texas and John takes I40, in what city will they intercept John?__________

12. Your friend, Max Smith, would like to make a trip from Albany, NY to Orient, NY.
    Since he will be making this trip often you should save it as MAX.TRP. You only need to fill in his name in the information section. He would like to take the Quickest route.

    How many miles is his trip?____________

13. Max would like to know what cities he will go through on Long Island, NY. Fill in the names of the cities on the map below.

APPENDIX C
PARTICIPANT INSTRUCTIONS

GENERAL DIRECTIONS:

During this task you will be working with a new computer system designed to help you plan road trips. You might want to think of the computer you will be working with as an expert partner. It will be able to help you complete the tasks more quickly. The computer partner has the ability to take over the task for you and give you advice if you run into serious difficulty. It cannot, however, complete all the tasks for you. You don't have to follow the advice of the computer partner although it is designed to help you.

Appropriate communication group directions here.

You will do two sets of tasks. You will have 10 minutes to complete one of the sets and 20 minutes to complete the other. You will be told when to begin and when to stop. Please complete the tasks in the order they are given as quickly and accurately as you can. Do not move on to the next task until you have successfully completed the one before it. You will be monitored to make sure you are working on the tasks and doing them in order. You may not complete all the tasks.

Please do not use the help function you see on the computer screen. Using the function may cause the program to crash and this experiment is designed to evaluate the computer partner so you should receive your help from the computer partner.

Communication Group Instructions:

Control (4)

We will be audio taping this session in order to record any comments you make while completing the tasks.

Context Sensitive (1)
Limited Response (2)

You will also be able to communicate with the computer partner by speaking in your normal voice to ask questions or make comments. Just speak as you would to another person. Don't hesitate to speak to the computer because this can make your activities go much more smoothly. The computer has a voice recognition system as well as speech capabilities and will attempt to respond to you in an appropriate manner. You may experience some brief system delays while the computer processes your speech. This is normal. For example, if you need to change your traveling speed, you might say, “How do I change my speed?”

The computer will inform you if it does not understand your speech and you can try again.

We will be audio taping this session in order to record your conversation with the computer and any additional comments you make while completing the tasks.

Limited Human-Computer (3)

You will also be able to communicate with the computer partner by speaking in your normal voice to ask questions or make comments, but you will have to make your speech as simple as possible. Here are the instructions for talking to the computer. Don’t hesitate to speak to the computer because this can make your activities go much more smoothly. The computer has a limited vocabulary, but will understand words that are related to this task. The computer also has speech capabilities and will attempt to respond to you in an appropriate manner. You may experience some brief system delays while the computer processes your speech. This is normal. For example, if you need to change your traveling speed, you might say, “How traveling speeds?”

The computer will inform you if it does not understand your speech and you can try again.

We will be audio taping this session in order to record your conversation with the computer and any additional comments you make while completing the tasks.

APPENDIX D
PARTICIPANT QUESTIONNAIRE

1. Using the scale below, rate your **ability to communicate** with the computer partner during the session.

1	2	3	4	5
Easy		Neither Easy or Difficult		Difficult

2. Using the scale below, rate the **helpfulness of the computer partner** in helping you to complete the computer tasks.

1	2	3	4	5
Helpful		Neither Helpful or Unhelpful		Unhelpful

3. Using the scale below, rate your **enjoyment** of working with the computer partner on the computer tasks.

1	2	3	4	5
Enjoyable		Neither Enjoyable or Unenjoyable		Unenjoyable

4. Using the scale below, rate your **computer ability**.

1	2	3	4	5
Beginner		Intermediate		Expert

5. How many **years** of computer experience do you have?___________

6. Do you have any comments about the computer system you worked with?

APPENDIX E
GOMS ANALYSIS OF COMPUTER TASKS

Summary of GOMS Analysis

	Simple Tasks	Complex Tasks
1.	1 goal / 2 subgoals	3 goals / 2,1,1 subgoals
2.	1 goal / 1 subgoal	4 goals / 1,2,1,1 subgoals
3.	1 goal / 1 subgoal	3 goals / 2,2,2 subgoals
4.	1 goal / 1 subgoal	3 goals / 1,1,1 subgoals
5.	1 goal / 1 subgoal	3 goals / 1,1,1 subgoals
6.	1 goal / 1 subgoal	3 goals / 1,1,1 subgoals
7.	1 goal / 1 subgoal	2 goals / 2,1 subgoals
8.	1 goal / 1 subgoal	3 goals / 2,1,1 subgoals
9.	1 goal / 2 subgoals	2 goals / 2,1 subgoals
10.	1 goal / 1 subgoal	4 goals / 1,2,1,1 subgoals
11.	1 goal / 1 subgoal	2 goals / 2,2 subgoals
12.	1 goal / 1 subgoal	5 goals / 1,2,1,1,1 subgoals
13.	1 goal / 2 subgoals	2 goals / 2,2 subgoals
14.	1 goal / 1 subgoal	-----
15.	2 goals / 2,2, subgoals	-----
16.	1 goal / 1 subgoal	-----
17.	1 goal / 2 subgoals	-----
18.	1 goal / 1 subgoal	-----
Total:	**19 goals**	**39 goals**

Goals in halves of tasks:

Simple	**9**	**10**
Complex	**20**	**19**

Goals in thirds of tasks:

Simple	**6**	**6**	**7**
Complex	**13**	**13**	**13**

VITA

Cristina Bubb-Lewis was born on April 25, 1969. She graduated cum laude from Duke University in May 1991 with a Bachelor of Science degree in psychology and a second major in religious studies. She earned her Master of Science in psychology from Old Dominion University in December 1993. In December 1997, she earned her Doctor of Philosophy in industrial/organizational psychology with a major in human factors and a minor in marketing.

In 1995, Cristina completed an internship at AT&T Bell Laboratories in the Wireless Human Factors Group - Cellular Division. Cristina has presented papers at various conferences including the 1st and 2nd Mid-Atlantic Human Factors Conferences and the 2nd Automation Technology and Human Performance Conference. She has accepted a position with Lucent Technologies, Bell Laboratories in the Advanced Technologies Organization.

Correspondence may be sent to:

Old Dominion University - Psychology Department
Mills Godwin Building, Room 250
Norfolk, VA 23529