# EMPIRICAL METHODS FOR EVALUATING VIDEO-MEDIATED COLLABORATIVE WORK

Jonathan K. Kies

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Industrial and Systems Engineering

Dr. Robert C. Williges, Chair

Dr. Robert J. Beaton

Dr. Roger W. Ehrich

Dr. Mary Beth Rosson

Dr. Walter W. Wierwille

March 26, 1997
Blacksburg, Virginia

Keywords: human factors, desktop video conferencing,
research methods, psychophysics, ethnography

# EMPIRICAL METHODS FOR EVALUATING VIDEO-MEDIATED COLLABORATIVE WORK

Jonathan K. Kies

(ABSTRACT)

Advancements in computer technology are making video conferencing a viable communication medium for desktop computers. These same advancements are changing the structure and means by which information workers conduct business. From a human factors perspective, however, the study of new communication technologies and their relationships with end users presents a challenging research domain. This study employed two diverse research approaches to the problem of reduced video frame rate in desktop video conferencing. In the first study, a psychophysical method was used to evaluate video image quality as a function of frame rate for a series of different scenes. Scenes varied in terms of level of detail, velocity of panning, and content. Results indicate that for most scenes, differences in frame rate become less detectable above approximately 10 frames per second (fps), suggesting a curvilinear relationship between image quality and frame rate. For a traditional conferencing scene, however, a linear increase in frame rate produced a linear improvement in perceived image quality. High detail scenes were perceived to be of lower quality than the low detail scenes, while panning velocity had no effect. In the second study, a collection of research methods known as ethnography was used to examine long-term use of desktop video by collaborators in a real work situation. Participants from a graduate course met each week for seven weeks and worked on a class project under one of four communication conditions: face-to-face, 1 fps, 10 fps, and 25 fps. Dependent measures included interviews, questionnaires, interaction analysis measures, and ethnomethodology. Recommendations are made regarding the utility and expense of each method with respect to uncovering human factors issues in video-mediated collaboration. It is believed that this research has filled a significant gap in the human factors literature of advanced telecommunications and research methodology.

This work received support from the National Science Foundation and
the Southeastern University and College Coalition for Engineering Education (SUCCEED).

# Acknowledgments

I would like to take this opportunity to thank the many individuals who helped make my graduate study a successful and rewarding experience. First, grateful recognition is due to my advisor, Dr. Robert C. Williges. He furnished expert technical guidance throughout all the stages of my graduate education, but more importantly, he provided me with countless opportunities to further my professional career.

My other committee members were also of invaluable service during the course of this project. Dr. Walter Wierwille was always available and a reliable resource for thoughtful feedback. Dr. Robert Beaton frequently provided positive encouragement and is acknowledged for taking a personal interest in this research. Dr. Roger Ehrich is greatly indebted for helping procure the computer equipment essential to this research and for clarifying the intricacies of video compression techniques. Dr. Mary Beth Rosson is thanked for taking the time to supply insightful feedback and providing much-appreciated encouragement and recognition of hard work. She has also been an inspiring example of quality scholarship and technical writing. Though not on my committee, John Kelso, a friend and colleague, is acknowledged for providing the technical prowess, programming skills, and constant enthusiasm without which this research would not have been possible. The second study of my dissertation was greatly enhanced through the efforts of my two undergraduate research assistants, Edward George and William Baum to whom I am greatly indebted.

I would also like to acknowledge the other human factors professionals who provided career opportunities and exemplified academic excellence and professionalism: Marvin Dainoff of Miami University; Bob Lysaght and Bob Springer of NYNEX Science and Technology; Dennis Price, Bev Williges, Pat Koelling, and Jack Carroll, all from Virginia Tech.

My friends at Virginia Tech have also been a source of encouragement and a wellspring of collegiality. I would like to acknowledge the support of Dennis and Vicki Neale, Mike McGee, Mike Snow, Andy Gellatly, Mark McMulkin, Terry Fairbanks, Jose Pesante, and Wayne Neale. Of course, none of this would be possible without the unwavering love and support of Kristine Brady, an outstanding researcher and true friend who showed me the way.

Finally, I would like to thank my parents, Frederick and Josephine Kies, who encouraged and facilitated academic pursuits, intellectual curiosity, and hard work.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1. Introduction

## Problem Statement

As the information era arrives, the fundamental nature of how we "do work" is shifting dramatically. Much of this change results from the ubiquitous implementation of new technology, namely, the networked computer. As a result of this changing work environment, human factors engineers need to develop new approaches to designing and evaluating systems. This research embodies an ambitious goal: to contribute to the understanding of an advanced communication technology - desktop video conferencing - and how it is used in a bandwidth-constrained network. New methods of design and evaluation are not fully understood in terms of effective selection, use, and expected results. Equally problematic is the realization that some traditional evaluation procedures have not been used extensively in designing and evaluating the changing work environment. This research will help to clarify the use of emerging as well as traditional design and evaluation practices by applying several to the examination of human interaction with a desktop video conferencing system and comparing the results to established research.

### Changing Work Environments

Computing devices have permeated many working environments, requiring task performance to depend on cognitive skills, as opposed to physical capabilities (Zuboff, 1988). An example is a service station that performs routine maintenance and oil changes on automobiles. Often, a computer with diagnostic capabilities and maintenance histories is consulted throughout an appointment. Our study of work must shift to account for these new cognitive task requirements. Recently, this shift has taken another interesting turn with the ubiquitous networking of computing devices. This trend changes work yet again by establishing shared working environments for use by remote group members.

Two significant characteristics of this change are relevant. First, a common collaborative working environment is established. For example, the use of Lotus Notes™ in corporate business environments as a shared database and document manipulation technology greatly changes the process of doing work as well as the organizational structure that must contend with a radically leveling influence (Orlikowski, 1992). The second important consequence of this changing work environment is the ability and requirement for communicating over computer networks. Computer networks afford the transmission of text (i.e. email, chat, and MUDs), audio (i.e. internet telephony and audio streaming), and video (i.e. desktop video conferencing and multicast video). The traditional communication media, including memos, letters, and telephony is being augmented or completely replaced by computer-mediated communication. Kies, Williges, and Rosson (in press, a) claimed that external communication is critical to collaborating successfully in networked environments. For example, computer-supported cooperative work applications such as Lotus Notes™ cannot be used efficiently or productively without coordination between participants. Grudin (1990) stated that technology-driven research worked relatively well for single-user environments, but fails miserably for the design and evaluation of multi-user applications. A better understanding of how groups work, how organizations operate, and how individual differences impact the use of computer-supported cooperative work is needed. These changes require that we re-examine the suitability of traditional work design and evaluation methods and explore new techniques or methods from other disciplines.

## Changing Technology

Desktop video conferencing systems (DVC) have become available in recent years due to developments in compression methods, high-bandwidth networks, and faster computer processing power (Gale, 1992). Compression methods include the standard International Telephone Union (ITU) H.261/Px64 and Motion Pictures Expert Group (MPEG), as well as a variety of proprietary products, such as Cell B. Each uses one of a variety of schemes such as discrete cosine transform, predictive techniques, and interpolative techniques (Luther, 1994). Network technologies capable of transmitting video and audio include plain old telephone service (POTS), 10BaseT ethernet, 100BaseT ethernet, switched ethernet, Asynchronous Transfer Mode (ATM), and fiber optics. Desktop computers can take advantage of the multitasking power and Reduced Instruction Set Computing (RISC) in UNIX workstations as well as the increasingly fast PowerPC and Pentium chips. Due to these readily available, albeit disparate technologies, the video image quality experienced by the end user has the potential to approach television broadcast quality, but can nonetheless vary considerably.

Frame rate is one of the most influential variables impacting the demand on networks, as well as the success of a video conferencing session, and can fluctuate over a wide range of settings according to the system constraints listed above. Unfortunately, little research has been conducted evaluating how and to what degree these crucial variables impair the communication process when reduced to levels below that which is considered acceptable by the user. Of particular interest are the potential effects of reduced frame rate on task performance and user acceptance.

## Problems with studying Collaborative Work

Collaborative work environments are defined as involving physically-dispersed workers communicating about a shared activity through computer networks. This situation has become more common, and therefore warrants the attention of human factors engineers. Unfortunately, studying these work environments is not a straightforward task. The degree of variability increases exponentially as more workers become involved. Multi-user applications are difficult to study for a variety of reasons (Grudin, 1990). First, the success of system usage is to a large degree contingent upon users' personalities and backgrounds. For example, as with all new technology, there are early adopters and those who resist its use. A collaborative application will be successful if all group members choose to use it. Second, the social, motivational, economic, and political dynamics associated with groups working in organizations are difficult to account for. These factors, which are normally eliminated in traditional studies, can have a profound impact on the use of a groupware application. Finally, our methods of design and evaluation are lacking. For example, research methods (which will be discussed in Chapter 2) from the fields of social psychology and anthropology that are often applied to multi-user environments are less precise and generalizable than the methods of traditional human factors engineering or experimental psychology.

We are no longer confronting human-machine interaction, but human-machine-human interaction. The computer interface has transcended the role of an object to be communicated with to that of a medium through which two or more humans can interact, communicate, and collaborate. As a result of this transformation, the highly-interdisciplinary field of human factors engineering must become even more interdisciplinary by borrowing on the research methods, vocabulary, and tools of communications experts, anthropologists, ethnographers, and information systems experts.

## Design Methods

The field of human factors concerns itself with the design of systems to account for the capacities and limitations of the user. To this end, the discipline has long benefited from the symbiosis of diverse methods and fields. The intent of this section is to outline the historical progression of human factors design methods from its early years in military applications to the current state of affairs. Particular attention will be paid to the genealogy of current methods and the ever-increasing diversity which propels the field.

The earliest attempt at improving the match between humans and machines was through the application of scientific methods to the design of manual tasks by Frederick Taylor. He leveraged both job re-design and personnel selection to maximize productivity and worker compensation (Konz, 1990). Later, the optimization of human-system performance was studied by Frank and Lillian Gilbreth through the refinement and application of time and motion studies. They set out to assess human performance and fatigue so that job tasks could be designed most optimally (Konz, 1990). The next significant advance came during World War II during which psychologists (i.e. Donald Broadbent and Paul Fitts) were called in to alleviate the operator workload associated with fighter pilot crashes. These contributions set the stage for modern human factors engineering as it applied to large-scale military systems, and later, a range of industries and scales, including consumer products.

Central to successful human factors design is the systems approach (Hopkins, et al., 1982; Meister, 1991; Sanders and McCormick, 1993). The systems approach implies attention paid to the phases of a system lifecycle (i.e. from concept to disposal) as well as a systematic procedure for considering the human user. The most fundamental means for initiating design is through a generalized procedure known as task analysis. Meister (1985) defined task analysis as a method by which tasks are studied in terms of design issues, personnel, training, and testing and evaluation. An earlier description of task analysis emphasized the description of component sub-tasks and the skills and human characteristics necessary to perform the tasks (Van Cott and Kinkade, 1972).

In describing direct observation techniques, Chapanis (1959) outlined methods of task analysis. The method of taking operator opinions in a survey or interview format was shown to provide valuable information regarding the design of truck cabs and airport control towers. Another analysis involves activity sampling. This technique permits the collection of data in either a gross fashion (i.e. over several hours) or a microscopic level (i.e. over a few seconds). Videotaped recordings can be made to permit post-hoc evaluation of the tasks and the task requirements. Another method outlined by Chapanis is that of process analysis, which can be conducted in a variety of different ways, depending upon the objectives of the designer. One such method is link analysis, which involves the statistical representation of relationships among system components. Process analysis can also be manifested in flow diagrams and process charts which track the flow of materials or information among machines and people.

Task analysis remains one of the oldest and most effective means for informing the system design process. Today, these initial techniques have blossomed into a wide range of methods, each with different objectives, advantages, and limitations. For example, Kiran and

Ainsworth (1992) described 25 task analysis procedures which were divided into five primary categories: task data collection methods, task description methods, simulation, behavioral assessment, and requirements evaluation. Three methods described are relevant to this study. Within the category of task data collection, "observational techniques" were suggested as a means for collecting data in an informal, naturalistic setting. Its intention is to permit external influences such as social interaction and to avoid the constraints associated with precise task measurements. Kiran and Ainsworth (1992) also described the use of questionnaires and interviews in collecting task analysis information.

In addition to task analysis, human factors engineers rely on physical assessments of humans, such as anthropometry, biomechanical characteristics, physiological measures of workload, and psychological assessments, such as mental workload, manual control, and basic sensory perceptual capabilities. These methods, in combination with sound task analysis embedded within the systems approach have proven to be an effective and time-tested approach for enhancing the design of systems to account for human use. The fields of industrial engineering and experimental psychology (from behaviorism through cognitive schools of thought) combined to form the basis of human factors design methodology.

While the ultimate goal of human factors is to inform system design by accounting for the human operator, it is now being recognized that psychology and engineering are not the only two fields which have the potential to make a contribution. As systems become more complex, and the influence of organizations and groups are viewed as significant factors, human factors is being influenced by other fields with a stake in human behavior and technology. On the human side, such fields include anthropology, sociology, social psychology, education, management systems engineering, and the Scandinavian workplace democracy movement. On the technology side, fields such as computer science, graphic design, electrical engineering, and computer networking are taking an interest in mapping human abilities to systems. These fields are combining research methods in a creative manner by cross-fertilizing philosophical orientations. Some of the new research methods developed in parallel to those used by traditional human factors engineers (i.e. information systems; Kling, 1995), while others developed before human factors (i.e. ethnography; Hughes, et al., 1995a), as a result of human factors (i.e. contextual inquiry; Wixon, et al, 1990), or from the influence of other fields, such as computer science (i.e. the cognitive walkthrough; see Polson, et al., 1992).

Traditional human factors research and design methods have come under intense scrutiny in the past 10 years. Numerous researchers who share the same goal as those dating back to W.W.II are now encouraging the application of other fields to help balance the limitations associated with older methods. Two parallel trends are motivating what can arguably be seen as a Kuhn-like paradigm shift in user interface design. First, designers from Scandinavia, influenced by the worker-democratization movement, have outlined the need to include users in the design process, and have had an influential voice in doing so (Greenbaum and Kyng, 1991; Schuler and Namioka, 1993). Second, cognitive scientists from both Europe and the U.S. have reflected upon the design of systems in recent years and concluded that formal research methods (i.e. laboratory testing; see Carroll, 1997) have been frequently ineffectual, and psychological theory (i.e. GOMS; Card, Moran, and Newell, 1983) has had minimal impact on the design of usable systems (Carroll, 1991). As a result, a powerful movement is underfoot to consider the social and organizational factors which may determine the success of new technologies (Carroll, 1997).

Bannon (1991), in a critical review of traditional human factors proposed a change in perspective:

> Within the HF (human factors) approach, the human is often reduced to being another system component with certain characteristics, such as limited attention span, faulty memory, etc., that need to be factored into the design equation for the overall human-machine system. This form of piecemeal analysis of the person as a set of components de-emphasizes important issues in work design. Individual motivation, membership in a community of workers, and the importance of the setting in determining human action are just a few of the issues that are neglected. By using the term *human actors* emphasis is placed on the person as an autonomous agent that has the capacity to regulate and coordinate his or her behavior, rather than being simply a passive element in a human-machine system.

People have goals, personalities, and political relationships which are generally difficult to characterize or control in traditional design methods, such as task analysis. These factors are virtually ignored by human factors researchers who have embraced the systems approach in which nearly every other factor (i.e. cognitive demands, physiology, maintenance, illumination, training, etc.; see Folley, 1960) has been accounted for in the design process. Bannon claimed that we need to do research in the workplace (although Chapanis, 1959 recognized this useful approach as well), to focus on group work, as opposed to solitary tasks, and to consider human interaction with technology over time as competency develops. In the past, systems required long development cycles and were intended for long-term use by select user populations (i.e. military personnel), permitting heavy investment in training resources. Today, technology changes much more rapidly, skilled performance is expected quickly and without significant assistance, and ease-of-use is a requirement, particularly for the demands of mass market applications and consumer products.

The abundance of research activity in HCI reflects a field in which intense creativity is being harnessed to further the design methods of human factors practitioners. For example, the notion of design rationale - that body of knowledge implicit in a design - is viewed as essential to the iterative design process (Moran and Carroll, 1996). Articulating the reason behind the selection of alternative features in a design can enhance user-designer communication and the development of future product versions. Means by which this knowledge is captured, stored, and communicated among designers have been studied by those who find design methodology a pertinent research endeavor. Likewise, other researchers have focused on refining the design process in terms of usability sample size (Virzi, 1992; Nielsen, 1994), prototype fidelity (Virzi, Sokolov, and Karis, 1996), usability walkthroughs (Nielsen and Mack, 1994), cost-effectiveness (Bias and Mayhew, 1994), and methods for including the user in the design process (Schuler and Namioka, 1993), to name a few. Clearly, human factors design is being improved and refined as organizations recognize its value and other perspectives are welcomed into the fold.

One way to characterize the dichotomy between the methods outlined by Chapanis (1959) and the newer, more multi-disciplinary methods is in terms of perspective. While older methods seek to improve design by studying human behavior and making lists of design recommendations, the focus is on cognitive modeling and human performance data, such as the calculation of mental workload and the application of multiple processing theory in GOMS. The human is perceived as a system component which can be studied under the proper conditions and subsequently "known". For many applications this has been and will continue to be a fruitful and valuable approach to system design. The other school of thought places the emphasis on context. Examples include studies which seek to account for social relationships, communities of learning, the ability of users to become experts over time, and organizational influences. The majority of methods which have arisen within this school of thought evolved independently of traditional human factors methods. For example, participatory design derived from the socialist worker movement in Scandinavia in the 1970's (Greenbaum and Kyng, 1991) and ethnography was used as an anthropological tool since the 1920's. The later field predated most human factors work, but has only been applied to human-technology studies in the past ten years.

Despite this independence, a number of other "new" methods which focus on context and the design of computer interfaces, have, in fact, evolved from progenitors in human factors. For example, the structured interview process described by Holtzblatt and Jones (1993) is a formal method for surveying workers in their environments, a method also advocated by Chapanis (1959) and Meister (1985). The activity analysis described by Chapanis (1959) which is clearly influenced from earlier time and motion studies has itself motivated design team task analysis (Olson, et al., 1992). While not a contextual method per se, usability testing has evolved into a rich set of methods for conducting formative and summative evaluations in the context of rapid prototyping and iterative product development from the earlier "test and evaluation" work done on large scale military systems (Meister, 1986). Likewise, the methods of critical incident taking used in user interface evaluation have descended directly from the method in which trained observers systematically recorded predefined events of particular importance (Shattuck and Woods, 1994). The technique has evolved throughout its 40-year history to meet the demands of new human-system problems.

This section has set out to describe a few basic human factors design methods and how they have evolved over time through a rich body of research and the confluence of diverse fields all concerned with studying human interaction with technology. Neither the cognitive side nor the contextual side is being condemned. It is hoped that human factors designers can embrace both camps intelligently and apply the most appropriate methods for unique design problems. Human factors engineers must not shun the potential benefit from other fields' methodologies, nor should they ignore the successful techniques from the past. By knowing where we have been while keeping our minds open to new ideas, the design toolkit for human factors professionals will grow. This research seeks to explore the use of both contextual and cognitive approaches in studying video-mediated collaborative work.

## Research Methods

In addition to designing systems, human factors engineers are concerned with advancing the science of human behavior so that system design can be informed more effectively. While research methods are often distinct from design methods, they often overlap. For the sake of discussion, however, they will be treated as independent activities, replete with their own idiosyncratic criteria.

Within traditional human factors engineering, we have used the methods of experimental psychology as they embody the scientific method. Such methods include controlled laboratory studies making use of factorial experimental designs and psychophysical approaches. While this collection of research methods has been invaluable to our growing understanding of human behavior with respect to technology, the problems of studying collaborative work environments are not well-addressed. We must begin to consider methods that take into account the *context* of a work environment, as well as the collaborative nature of multi-user computer work. It is here where ethnography and field studies can make a valuable contribution to filling the gaps of traditional methods and enhancing our understanding of this new work paradigm. In effect, the trend for design methods to consider social and organizational factors is also reflected in our research methodology.

Research on group work can occur within a variety of frameworks, each with advantages and disadvantages. McGrath (1995) discussed research in terms of three fundamental characteristics: generalizability, precision, and realism. No research paradigm can maximize all three simultaneously. Different approaches offer advantages and disadvantages over others in terms of these dimensions. Ethnographic methods (to be discussed in detail with other methods in Chapter 2) offer a high degree of realism, but utilize imprecise measurements. Alternatively, controlled experimental testing offers a high degree of control and precision, but often fails to embrace a realistic task replete with contextual influences.

The emerging technology used to support distributed work groups has received much research attention, little of which embraces the traditional approaches of human factors engineers. Psychophysics, in particular, has an opportunity to be "resurrected" as a viable method for understanding the technological characteristics of Computer-Supported Cooperative Work (CSCW) applications and how they are used by the contemporary work force. Together with ethnographic methods, psychophysics may result in a more complete picture of human-computer interaction.

Unfortunately, the research domain of CSCW is relatively new and few explicit guidelines are available to help the researcher select from the various methods. The purpose of this research is to outline where the different research methods fall in terms of their ability to answer research problems typical of human factors engineers studying computer-supported cooperative work. Two experiments were conducted to help fill the gaps in the literature in terms of methods applied to video-mediated collaborative work.

## Research Goals

This research seeks to contribute to both our knowledge of human performance and the research methodology of evaluating the new work paradigm. The following goals are addressed by this research.

- Outline the advantages and disadvantages of psychophysics, traditional laboratory research, ethnographic methods, and field studies as techniques for studying human interaction in collaborative work settings. [Chapter 2]

- Establish a ratio scale using the psychophysical method of free-modulus magnitude estimation for video image quality at different frame rates and for different scene types. [Chapter 3]

- Conduct a rigorous ethnographic analysis based on questionnaires, contextual inquiry, interaction analysis, and ethnomethodology. [Chapter 4]

- Determine how reduced frame rate and the effects of time influence the nature of work activities for desktop video conferencing systems used by collaborative workers. [Chapter 4]

- Determine the relative effectiveness of the above methods for human factors research, design, and evaluation. [Chapter 5]

# Chapter 2.  Background Literature

A complete understanding of the motivation for this research cannot be gained without a thorough review of the relevant research literature. Due to the focus of the proposed research on methodology, literature describing the use of various research methods is particularly germane. Specifically, psychophysics, controlled experimental testing, ethnography, and field studies are described. Also covered in this section are the historical development of video-mediated collaboration, the details of video communication, and the tasks that have been studied. From the research in these areas, one can identify the gaps in the research and learn from the methods and mistakes of other researchers.

## Methods for Studying Computer-Mediated Communication

The question of research method selection is highly contingent upon the objectives and philosophy of the researcher. Through the many scientific disciplines, we have developed a large body of methods by which we can learn about the world. In brief, these methods can be divided into two primary classes, each with its own philosophical boundaries. In one class of methods, *positivism*, the researcher's objective is to build or test a theory using the scientific method. Positivism generally works by comparing two or more situations in which all variables are kept constant except the independent variable of interest. The advantage to this method is that, if applied correctly, one can make claims about a theory: prove, disprove, or build an explanation for some behavior. There are drawbacks to positivism, however. These include costly experimental designs (i.e. large subject pools and inconvenient access points), low ecological validity, and uncertainty that all extraneous variables have, in fact, been controlled.

Another set of research methods have been designed and used to counter some of the problems found with positivist methods. These methods, known under the umbrella of *naturalism*, seek to understand behaviors without seeking a unifying theoretical explanation. Three primary methods in the social sciences embrace this objective: symbolic interactionism, hermeneutics, and phenomenology. Symbolic interactionism embodies a flat rejection of behaviorism by claiming that people respond to stimuli differently over time, often changing their models Behavior is constantly changing as these models are redefined and reconstructed. Hermeneutics, is a similar, but broader concept which entails the science of interpretation. The third method, phenomenology, which encompasses ethnography, seeks to understand daily living activities from the perspective of the person, or user. Figure 2.1 displays the relationship between these "ways of knowing".

As previously mentioned, the most suitable - and by far most popular - method for building and testing theories is that of the positivist tradition and the scientific method. However, it has been shown that naturalism, specifically, ethnography, can also be used to test hypotheses (Hammersly and Atkinson, 1995). This accounts for the dotted line connecting the two traditions in Figure 2.1 and is embodied by the method of analytic induction, which will take on greater importance in the results section of the second study.

Human factors engineers have long relied on a variety of methods for conducting research on human-machine interaction. Each method poses challenges to the researcher and inherently limits the findings which can be gained. McGrath (1995) emphasized the mindful selection of method to address the research questions of interest. Each has its own advantages in terms of generalizability, realism, and precision. The mechanics as well as benefits and drawbacks are outlined in this section.

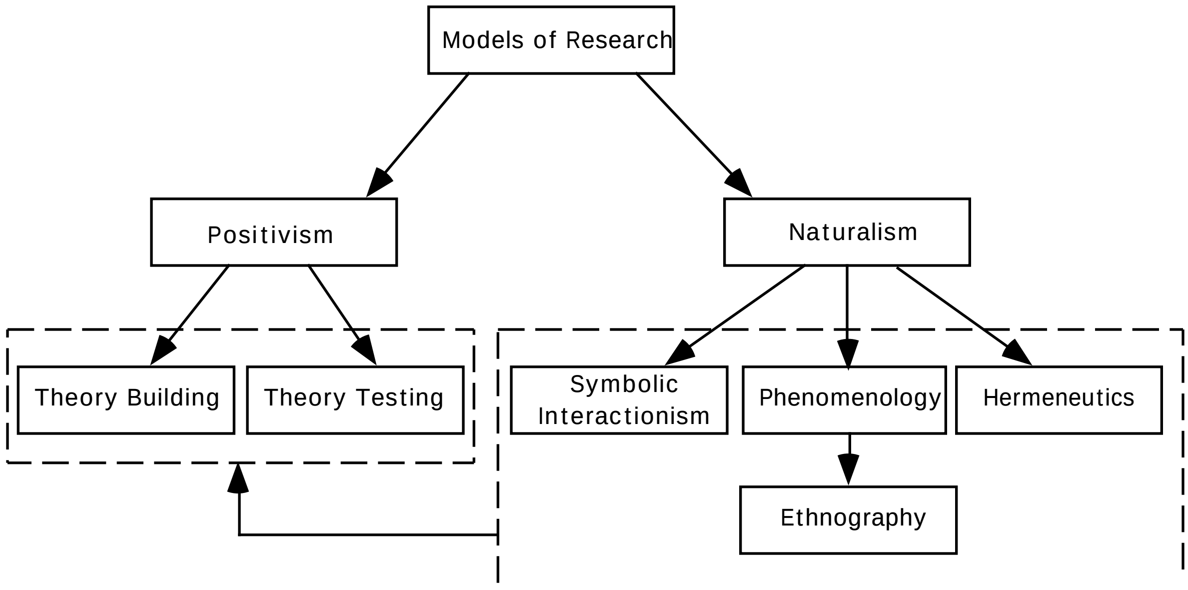



FIGURE 2.1  Research Methods for the Behavioral Sciences

## Psychophysics

Psychophysical methods attempt to map psychological experiences to physical stimuli or seek to determine perceptual thresholds.  Many methods are available, but magnitude estimation and category scaling are of interest in this study.  For a comprehensive review of these methods, see Stevens (1975) and Gescheider (1985).

The method involves the subjective rating of a physical stimulus by having the subject assign a numeric value to each stimulus, with smaller values indicating weaker stimuli, and higher values indicating stronger stimuli.  Thus, the subject is applying the well-established cognitive model of the number system, which is used in daily life.  This model, known as "numerosity" describes the relationship between numbers and has been shown to be a robust construct in adults (Stevens, 1975).  Subjects are exposed to a randomly presented sample of approximately 6-10 stimuli (although 60 samples have been used in some experiments) for which they are to make numeric judgments of their relative intensity.  Generally, the stimuli need to be presented no more than twice each for approximately 10 subjects.  Additional subjects are necessary if the variability of judgments is considerable or the stimuli are complex.  In the case of video image quality acceptability, the stimulus is, in fact, complex, and Chuang and Haines (1993) found considerable intra-subject variability, warranting a sample size greater than 10.  The obtained values are then corrected to a common modulus, the geometric mean is calculated and plotted on log-log paper against the physical stimulus.  The specific calculations are covered in Chapter 3.

Stimuli are often grouped into one of two categories:  prothetic and metathetic stimuli (Gescheider, 1985).  Prothetic stimuli are those which tap additive neural processes and demonstrate additivity.  Examples include sound pressure level, light, and other sensory modalities.  Metathetic stimuli are those which tap qualitative changes in magnitude, such as image quality. Most prothetic stimuli have been found to meet Steven's Power Curve, while many metathetic stimuli do not.

Stevens (1975) showed how magnitude estimation has proven to be an efficient tool for determining subjective impressions of metathetic and prothetic physical stimuli. Stevens (1975) also discussed the application of magnitude estimation to "social psychophysics" for a variety of dimensions, including wrist watch preference, music and handwriting aesthetics, pleasantness of odors, and prestige of occupations, among others. For example, Snow (1996) measured "presence" in a virtual environment using magnitude estimation. It is in this realm that the question, "Is this video quality acceptable?" (per this study) is situated rather than traditional sensory threshold questions such as "Can you hear it?".

Another consideration for conducting a magnitude estimation study is the definition of comparative modulus. In experimenter-determined magnitude estimation, a standard stimulus is provided to the subject for which all subsequent judgments are to be based. Free-modulus magnitude estimation allows the observer to use any number based on a randomly determined initial stimulus level. The later method has been shown to reduce biases associated with modulus selection. For example, Gescheider (1985) showed how a modulus selected from the center of the stimulus distribution will produce higher psychophyscial functions than those selected from the extremes. For this reason, free-modulus magnitude estimation will be used in this study.

Category scaling is another psychophysical method similar to magnitude estimation. Due to the potential usefulness of category scale data, and the demonstrated ease with which such data can be collected in conjunction with the magnitude estimation data, the methods section will outline the procedure for conducting such an experiment. First, however, a few words of caution on the subject of category scales are necessary. Stevens (1975) claimed that the fundamental difference between magnitude estimation and category scales is that the category scale imposes a limit on the number of adjectives or numbers that can be used. For this reason, relative magnitudes cannot be judged on a ratio scale continuum. Another limitation of the category scale is that the assumption of equal-appearing intervals is not necessarily valid, resulting in an ordinal scale at best (which may require non-parametric statistics). The values of the scale shape depend upon the stimulus set and order of presentation. This effect may be due to the tendency for subjects to use all scale values equally often (Engen, 1971). Jones and Marks (1985) compared category scaling with magnitude estimation in a video image quality experiment and found that the category scale suffered from context effects. In other words, subjects were constrained by the range of values in the provided scale, unlike magnitude estimation in which subjects are free to use any positive number they feel is appropriate. Finally, Stevens points out that the results of the magnitude estimation technique may have a different mid-point than that of the category scale.

The advantage of psychophysical studies is the establishment of a direct relationship between physical stimuli and psychological experience. These studies reflect a high degree of precision and have been found to produce functions consistent across many physical dimensions. In terms of desktop video conferencing, psychophysics can lead to a rich understanding of how people perceive varying image quality across a continuum of values.

The disadvantage to this paradigm is that findings may not reflect realistic tasks and may not generalize well to other applications. Additionally, it fails to account for subjective satisfaction with a particular stimulus. For example, suppose frame rates above 5 fps were found to be significantly different than those below 5 fps on a psychological experience scale. In reality, users may have little objection with the lower frame rates. Finally, psychophysical methods do not account well for individual differences, despite the establishment of a well-developed scale from a wide range of subjects.

## Controlled Experimental Studies

Controlled experimental studies fall firmly into the positivism tradition of research and are well-known to human factors engineers. To summarize, subjects are selected randomly and assigned to experimental conditions in which categorical treatments are administered, reflecting the manipulation of independent variables. The effects of those variables are gauged in terms of pre-determined dependent measures. Dependent measures may be quantitative as in task performance times and error rates, or qualitative as in subjective satisfaction. Clearly, there is much overlap with psychophysics since magnitude estimation data can be collected in a controlled factorial design. The design of experiments can reflect full-factorial designs (Winer, Brown, and Michels, 1991; Keppel, 1992), efficient designs involving fractional factorials (Martin, Williges, and Williges, 1990; Winer, et al., 1991), and sequential experimental designs and empirical model building (Williges, Williges, and Han, 1992).

The advantages of this paradigm of research include generalizability (although not necessarily ecological validity), a high degree of precision, and the ability to compute descriptive and inferential statistics. From these studies, predictions can be made of how humans will perform in similar environments, enabling systems to be designed for optimal human-system compatibility.

The drawback of this paradigm is the low degree of realism and removal of contextual influences. By its very nature, controlled experimental studies remove contextual biases so that the effect of manipulated variables can be judged without the confusion of confounding factors. However, when studying group behavior, many of the confounding factors often controlled for (i. e. organizational pressures, effects of time on group cohesion and performance, and meaningful task assignment) are hypothesized to have an effect on the use of computer-mediated communication (Tang and Isaacs, 1993).

## Ethnography

In contrast to controlled experimental studies, ethnography is a research method rooted in the naturalism class of phenomenology. Ethnographic methods comprise a loose family of techniques aimed at developing rich characterizations of individuals and groups as they interact with each other and with supporting artifacts in realistic work settings. The techniques derived from the research methods of sociology and anthropology incorporate a variety of data collection procedures. One early researcher, Bales, developed a method of studying conversation through meticulously precise metrics (Bales, 1950). Another influential researcher, Harold Garfinkle, described these methods under the rubric of "ethnomethodology", and used them to understand the mundane behavior of individuals by purposely breaking established social norms (Garfinkle, 1967). Since then, other methods have evolved that account for the work context in the study and design of systems, some of which are less rigorous than others. These methods include participatory design (Greenbaum and Kyng, 1991; Bødker, Grønbæk, and Kyng, 1993), contextual inquiry (Wixon, Holtzblatt, and Knox, 1990; Holtzblatt and Jones, 1993), scenario-based design (Carroll and Rosson, 1990; Carroll, 1995), and Activity Theory (Nardi, 1996).

The general theme of many of these methods is to understand the world as a collection of socially-constructed phenomena, accomplished through the study of daily activities. Central to ethnography is the user perspective and the intention to uncover meaning in social patterns and artifacts. This section will briefly describe three types of ethnography relevant to the research: ethnomethodology, interaction analysis, and contextual inquiry.

### *Ethnomethodology*

Ethnomethodology involves the systematic and detailed study of ordinary tasks in daily life from the perspective of a subject. This focus on ordinary tasks separates ethnomethodology from the larger rubric of ethnography. Indeed, there appears to be some confusion among the use of these terms within the CSCW community, as they are often used interchangeably (Dooley, 1995). Shapiro (1992) suggested that when most CSCW researchers refer to "ethnography", what they really mean is "ethnomethodological ethnography". For this reason, "ethnography" will be used synonymous with ethnomethodology in this section, in keeping with the terminology used in the literature cited.

Ethnomethodology is primarily concerned with the study of mundane, everyday tasks from the perspective of the subject. To achieve this understanding, several approaches are available to the ethnographer. Garfinkle (1967) popularized the violation of social norms to study how people behave in such situations. Reactions of unsuspecting subjects were intended to highlight or clarify social behavior. For example, a researcher might purposely invade personal space to understand the dimensions of this phenomenon and the means by which people establish and protect that space. A more conventional approach - and one that is more useful for studying CSCW - is to videotape users interacting with a system so that detailed observations can be made. Along with this method, researchers can rely on an informant or a penetrating researcher to help explain an activity from the perspective of the subjects. The Chicago School in the 1920's was the first prominent and widespread use of ethnography in sociology (Hughes, O'Brien, Rouncefield, and Sharrock, 1995). These researchers advocated the use of an inside source, and underscored its advantage with the following example: To study the urban working class in Chicago, researchers were interested in Taxi-Dance Halls in which men could purchase 10 cent dances. Clearly, interviewing patrons or giving them surveys would be an inadequate means for understanding this urban activity. Instead, researchers covertly penetrated the clubs and interacted with the patrons anonymously. Hughes, et al. (1995) and Hughes, King, Roden, and Andersen (1994) maintained that anonymous inside information is helpful, but not crucial, especially when dealing with computer applications which are not of a sensitive social nature.

Hughes, et al., (1994) described four classes of ethnography which can be used for various purposes in system design. *Concurrent* ethnography involves long term study and parallels the entire system development process, informing design throughout. *Quick and dirty* ethnography is used to inform initial design stages by employing short studies of representative work situations. *Evaluative* ethnography is also short, but is used to assess a design in a summative evaluation sense. Finally, *reassessment* ethnography involves the re-examination of previous studies to look for commonalties relevant to a new system design project. Each of these variations retain the same focus, principles, and general procedures of ethnomethodology. While traditional anthropologic ethnography is usually conducted over many months or years, these researchers and others (Rose, Shneiderman, and Plaisant, 1995) emphasize that user interface ethnography can be conducted over shorter periods of time.

The most extensive research conducted on and with ethnomethodology has occurred in Great Britain in the past five years under a multi-institution collaboration program known as COMIC. These researchers have used the method for the study of an air traffic control computer system (Bentley, Hughes, Randall, Rodden, Sawyer, Shapiro, and Sommerville, 1992; Twidale, Randall, and Bentley, 1994), a banking system (Hughes, Sharrock, Rodden, Kristoffersen, O'Brien, Rouncefield, and Calvey, 1995a), the workflow of a technology center (Hughes, et al., 1995a), a networked virtual environment (Hughes, et al., 1995a), and desktop video conferencing (Hughes, et al., 1995a). While other ethnographies have been influential (Nardi and Miller, 1991, Suchman, 1983, Suchman and Trigg, 1991), the U.K. researchers offered the most flexible and established approach to collecting, analyzing, and presenting ethnomethodological ethnographic data.

Hughes, et al. (1995a) suggested centering analysis around three main themes: distributed coordination, plans and procedures, and awareness of work. These three themes are intended to establish a balance between idiosyncratic, happenstance observation and an overly systematic and inflexible approach. Distributed coordination refers to the means by which participants organize the workflow and the process of an activity. This aspect of the analysis is crucial to groups communicating through a computer system while working on a shared task. Plans and procedures refer to the formal structures which enable distributed coordination to take place. Examples include schedules, manuals, organization charts, and workflow diagrams. Awareness of work describes the mechanisms by which workers make their activities visible to others. Hughes, et al. (1995a) described two methods by which workers achieve work awareness. The first, ecologies of awareness, refers to the physical layout of an environment or interface which affords knowledge of others' work status. For example, a radar screen which displays territory beyond the operator's sector can facilitate communication and cooperation among air traffic controllers. The other method for supporting work awareness is "representation of the work from within the work". This concept involves using representations within an interface to maintain one's place in a complex work system.

Applying this framework requires several steps. First, the organization, its politics, employees, general work practices, and purpose must be understood. Given this general background, the practitioner must observe behavior over an extended period of time, often making video and audio recordings. Hughes, et al. (1995b) outlined a useful series of procedures by which an ethnography can be performed, including steps such as gaining access to the site, selecting objectives, and organizing a study team. The analysis of data, however, is the most tricky aspect to conducting an ethnography. Hughes, et al. (1995a) provided several examples which follow a similar format. A series of text boxes containing actual transcriptions of an activity of interest were presented. These transcriptions were much less detailed than a thorough Interaction Analysis, but contained the important statements and actions of the people involved. From these representative transcripts, explanations for the behavior were posited. In general, instances of support or conflict with respect to each of the three themes of the analytic framework were discussed. It is argued that by adhering to this framework, some commonality will exist among different studies, but that enough flexibility is permitted such that detailed analyses can be performed to meet the objectives the study.

A word of caution should be mentioned regarding a reliance on ethnography. In no way is this method being put forth as a paragon research method for the design of systems. Hughes, et al. (1995a) recommended using it in conjunction with other methods, depending upon the research needs and the stage of system design. Clearly, ethnography has numerous shortcomings, just as all research methods do. For example, its ability to test theories is limited, the analytic induction method (Hammersley and Atkinson, 1995) notwithstanding. Similarly, ethnography is influenced by the observers' biases and preconceptions, making data difficult to interpret objectively. Finally, it is difficult to compare results among studies due to the numerous formats for reporting and interpreting data. It is therefore advisable to use a mix of approaches which seek to complement each other and balance the pros and cons of each.

### *Interaction Analysis*

Conversation Analysis was one of the first techniques used in ethnography, dating back to Bales' (1950) study of small group interaction. This technique allows the researcher to quantify and describe conversation between individuals in an exacting and tedious fashion. Example measures include the number of turns taken by an individual, the duration of utterances, the amount of "back channel" utterances and so forth. From this data, researchers can gain an understanding of conversation efficacy, what influences the nature of conversation, and how different personalities affect verbal interaction.

Since then, conversation analysis has become one technique of many used in the more general study known as interaction analysis (IA). IA allows researchers to analyze conversation as well as the interaction of individuals with each other and the physical artifacts in their environment. The collection of techniques has gone through many changes, with few clear guidelines on how to conduct the method. The research at Xerox PARC and the Institute for Research on Learning (IRL) have been leaders in the use of ethnographic techniques as they relate to human-computer interaction and learning environments and have contributed much to the methodology.

The influential study of an accounting office by Suchman (1983) was one of the first to apply ethnographic methods to a task analysis in an attempt to establish models of work for HCI practitioners. Suchman and Trigg (1991) claimed that studying workers in their actual environments is ideal, but re-creations can be made in the laboratory so that prototype systems (i.e. expensive electronic whiteboards or sophisticated video conferencing facilities) can be evaluated using real workers performing real tasks prior to the release of the system. Suchman and Trigg described four classes of records which can be collected during a session. Setting-oriented records make use of video tape to preserve the physical environment of the workspace. Person-oriented records seek to understand the work from a person's point of view by tracking the movement and position of workers relative to each other and the technology of interest. Object-oriented records track the movement and control of a technological artifact such as a mouse. Finally, task-oriented records involve the use of multiple recordings of distributed personnel working toward a common goal (i. e. dispersed workers collaborating using video conferencing technology).

Fafchamps (1991) provided more detail than Suchman and Trigg (1991) for conducting an ethnographic work flow analysis; five data collection techniques were presented. Thinking aloud procedures require the worker to describe verbally his or her activity while it is being performed. A guided tour is an explanation of the workspace to the evaluator. Structured observations include video and audio recordings of interactions between the workers with respect to their task. Written artifact analysis requires the collection of formal reports and informal documents such as PostIt Notes and descriptions of these documents by the workers. Finally, focused interviews are conducted which explore aspects of the work not otherwise identified in the analysis. After collecting the data, a themes analysis can be conducted which classifies the data into one or more of several general themes: breakdowns, routines or processes, one-of-a-kind actions, comments on actions, artifacts, procedures, and descriptions of the work environment.

Other studies involving ethnographic methods have been used by Tang (1991) in the study of collaborative design meetings and Nardi and Miller (1991) in the study of spreadsheet users. Tang classified three categories of behavior used for each of three functions. A matrix of behavior and functions was developed to understand the process of the design meetings. Nardi and Miller employed semi-structured interviews and audio recordings of conversation to analyze spreadsheet usage and the degree of collaboration among the users.

In a comprehensive method-oriented paper, representatives from Xerox PARC and IRL labs described their years of experience with these methods, provide suggestions for making the methods more effective, and gave examples of data collection and analysis (Jordan and Henderson, 1995). Interaction Analysis, as Jordan and Henderson described it, is a dynamic, malleable set of interests or "foci for analysis". While the study of each particular focus depends on the objectives of the researcher, seven relevant foci were presented.

First, the structure of events is one unit of analysis. Here, researchers look for when activities start and end, what occurs immediately adjacent to those end points, and how the group transitions between disparate segments of activity. The second focus is that of the temporal organization of an activity. Here, researchers look for periodicity and rhythms to the activity. For example, activities that keep re-surfacing or seem to be repetitive are noted. The third focus is that of turn-taking. Long the interest of conversation analysts, the analysis of how transitions in conversation are made can be divided into two end points on a continuum: Talk-driven interaction and instrumental interaction. In talk-driven interaction, the primary activity is speech. Determining conversation transitions involves watching gesture, gaze, and other clues. In the case of instrumental interaction, the attention is on physical objects, tools, and workspaces, as in the case of students sharing input and output devices in front of a computer workstation. In studying turn-taking, researchers should focus on the mouse, a common object of control. The fourth focus is on participation structures. Researchers must study the way participants use their bodies during the work process. For example, the positioning of workers in a group has much to do with attitudes toward the task. Related to this focus is that of the spatial organization of activity. This focus is concerned with the effects of positioning on social comfort, power struggles, and task involvement. Here, furniture and technology have a major impact. The sixth focus involves the study of communication breakdowns and the means by which they are repaired. In computer-mediated communication, users often experience communicative difficulties due to limited bandwidth and partial anonymity. The final focus is on artifacts and documents. Jordan and Henderson (1995) suggested tracking the control and use of key artifacts, like a mouse on a computer workstation or the marker on a whiteboard. Such behavior has the power to reveal a good deal about the process of communication and collaboration in these work settings. Jordan and Henderson outlined a series of steps for conducting this type of analysis; a modified technique will be outlined in the Method Section of this paper.

Olson, Olson, Carter, and Storrøsten (1992) studied small group design meetings in which they applied a method for classifying the activities of design meetings and the transitions among these activities. Their scheme was used to examine the relationships among these activities by calculating 1st through 6th order transitions and hierarchical clustering analyses. The design meeting activities, which were cataloged, included issues, alternatives, criterion, project management, meeting management, summary, clarification, digression, goal, walkthrough, and other; each is highly defined in the paper. This analysis fits well into the framework of Jordan and Henderson (1995).

The advantages of Interaction Analysis are the collection of data in highly-realistic work settings which account for organizational pressures, the effects of time on groups, physical work spaces, realistic task scenarios, and other factors stemming from a highly contextual analysis. Many studies of work (i. e. contextual inquiry) rely heavily on interviews and questionnaires. Interaction analysis provides a more direct analysis which does not rely on the post hoc recollection of participants (Suchman, 1983).

Any published paper on the use of conversation analysis, interaction analysis, or ethnography in general will provide different data collection techniques, analysis procedures, and suggestions for effective implementation of the method. This is one of the weaknesses of these methods, according to Nardi (1996a). Comparisons across studies are difficult because the variables being measured differ, making its repeatability problematic. Similarly, the degree of precision afforded by the method does not allow researchers to control extraneous factors. While this is also considered a strength, the use of this analysis in a factorial experiment is problematic. Finally, carrying out this method is cumbersome and highly time-consuming. Some estimates put the analysis of one hour of video tape at 60 hours (Jordan and Henderson, 1995).

### *Contextual Inquiry*

Interviewing workers as part of a task analysis is an established method for collecting data (Chapanis, 1959) A method which is highly-related to ethnography is that of contextual inquiry (CI). CI was established by researchers at Digital Equipment Corporation (DEC) as a formal means for analyzing work activity in the context of the actual work environment. Wixon, Holtzblatt, and Knox (1990) discussed the method in detail as it relates to the design of new computer systems or consumer products. The method provides a comparatively structured framework for conducting an interview with representative users, expanding upon users' comments by establishing a shared knowledge base of the work activity in its context, and the use of this information to guide the initial designs of a system. Holtzblatt and Jones (1993) provided more detail about how one should conduct the interview, including how to arrange a visit, select representative users, structure an interview, interpret the information, and implement the method throughout the design cycle of the system. The method is similar to ethnography in that it looks for patterns of work flow, disruptions to the flow, and the relationship between the system and its support for the task activity. The method's strength is the focus on users within the context of work, but appears to offer little advantage over the methods described by Suchman (1983), Fafchamps (1991), and Jordan and Henderson (1995) other than providing guidelines for conducting an interview with a user. Therefore, the disadvantages and advantages to CI parallel that of ethnography.

## Field Studies

The study of work in the field often employs the techniques described in the previous section. Participatory design allows systems to be designed with the cooperation of workers at all stages of the development process. Contextual Inquiry makes use of detailed interviews with workers, in situ, so that the process of work can be understood for the design of new systems. Finally, scenario-based techniques focus on design from the perspective of descriptive narrative of system use so that all stake holders in the design process can envision the system from a common viewpoint. All of these methods can be conducted in a casual fashion or more rigorous manner. The problem arises when researchers publish papers on prototype systems in which usage was evaluated and described in a manner subordinate to the discussion of the system itself. Such is the case with a great deal of research on "media spaces" and collaborative technologies which have been developed by well-funded corporate research organizations who emphasize design over evaluation. The data collected often involves embedded metering of usage (e.g. number of email messages sent per day, time using the system, or network traffic) subjective satisfaction, and limited video taping.

The advantage to these research methods is the focus on technology. Without the sharing of knowledge about costly prototype systems, our design efforts in this area may become duplicated, or worse, restricted. Such methods are best conducted when technologies are just beginning to mature.

The disadvantage of this research is that the developer of the system is often the user of the system, precluding objective analysis and evaluation. Often the system was developed by organizations with a large financial stake in proving its commercial viability and acceptance by "prospective" users. These research methods lack generalizability and precision, and offer a taste of realism only as it pertains to the limited subject sample selected. Despite these criticisms, properly conducted field studies can offer a glimpse of the real problems associated with implementing a new technology.

## History of Studying Video-mediated Collaborative Work

The study of video-mediated collaborative work dates back to the early 1970's when Chapanis (1971) set out to determine how humans would communicate with computers in the future. Although interaction modalities have remained largely limited to the keyboard 25 years later, a research program was conducted for nearly a decade, contributing much to our knowledge of human-human communication. At roughly the same time, researchers in Great Britain were examining the human factors of communication media, documented in the landmark book, The Social Psychology of Telecommunications (Short, Williams, and Christie, 1976). The role of communications media gained more attention in the early 1990's as desktop video conferencing systems became viable products due to advanced computing power and ubiquitous networking of computer workstations. Studies often examined prototype video systems, known as "media spaces" which rarely resulted in consumer products. Other studies were conducted on more viable systems available on today's computers.

In addition to these differences, the research in this area is decidedly diverse in terms of research methods. Some studies were concerned with the mapping of physical stimuli to psychological experience, using psychophysical techniques. Others utilized traditional experimental designs. Still others used various ethnographic methods and field study designs. This section is divided into four subsections which present the relevant research findings according to the research methods which were used.

### Psychophysical Studies

The majority of psychophysical studies on image quality have examined human detectability of image degradations with respect to spatial and temporal resolution. Because the psychophysical methods are diverse, yielding different scales and due to the differences in standards between Europe, Asia, and the U.S., much of the relevant research is difficult to compare. Similarly, the degree of context is highly variable across the studies - some have examined motion detectability of vertical lines, while others involve more realistic tasks, such as lab rat behavior detection. Often the type of method used and the task examined are a function of the question asked by the researcher. Some studies were conducted in the interest of basic human vision capabilities while others examined actual human performance on a likely task scenario. This section will briefly describe literature which has used psychophysical scaling and the variables of interest for the purposes of demonstrating the use of psychophysics in visual display evaluation. Studies relevant to frame rate will be discussed in more detail in the video communication section of this chapter.

The only authors to directly and formally examine the effect of frame rate on subjective evaluation in a realistic task were Chuang and Haines (1993). These researchers employed the method of limits to evaluate the ability of biologists to evaluate rat behavior in a proposed space station project. Data were collected and reported for both increasing and decreasing stimulus values for nine subjects. Another study was conducted by Whiting, Honig, Carterette, and Eigler (1991) in which the effect of frame rate was measured on the detection of ambiguous signals in noisy medical images. These researchers employed the forced-choice maximum likelihood estimation analysis using the method of scoring.

In a comprehensive study, Watson, Ahumada, and Farrell (1986) reported the use of the forced-choice variation of the method of limits for evaluating factors which influence the perception of apparent motion. It was determined that a "Window of Visibility" for perceived motion in terms of frame rate was a function of the spatial and temporal acuity of the observer and the velocity and spatial frequency content of the image. This finding directly influences the stimuli described in the Method Section.

Miller (1993) evaluated a range of display parameters on the perception of horizontal image motion. Free modulus magnitude estimation was used in conjunction with a factorial experimental design. Jones and Marks (1985) also used magnitude estimation to measure the general image quality of NTSC static images at varying signal-to-noise ratios. The study was conducted to compare magnitude estimation with a category scaling method to determine which is best-suited to map psychophysical phenomena. It was determined that magnitude estimation produced values with less variability in addition to the advantage of permitting parametric statistical analyses. Despite the weaknesses of the category scaling method, Westerink and Teunissen (1990) used it measure the effects of velocity on the perceived sharpness of images in complex scenes.

In summary, these studies have enabled researchers to understand how the physical parameters of displays correspond to human observers' psychological experience. It has been demonstrated that the methods are useful for resolving display issues, and can be an informative means for maximizing display quality when various characteristics (e.g. spatial resolution, color, frame rate) are competing for limited bandwidth (Schreiber, 1984).

## Controlled Experimental Studies

Controlled laboratory studies making use of experimental designs have been one of the most common methods for exploring video-mediated communication. The first and most well-known studies to examine human communication modalities were conducted by Chapanis and his colleagues throughout the 1970's (Chapanis, 1971; Ochsman and Chapanis, 1974; Chapanis, 1975; Weeks and Chapanis, 1976). These authors used rigorously controlled factorial designs examining a wide range of communication modalities and their combinations on the performance of tasks between two individuals. The most significant finding consistent across these studies was that the audio channel was most critical to task performance. Video-mediated communication was found to be more similar to audio-only communication than face-to-face communication.

Since these studies, others have borrowed Chapanis's research paradigm to test technical configurations of video-mediated communication as well as new communication technologies. For example, Gale (1990, 1991) manipulated communication medium (data sharing, audio, and audio/video) over information exchange, creative cooperative work, and meeting scheduling tasks. Another series of studies conducted by Prussog, Mühlbach, and Böcker (1994) researched the effects of video size and contextual cues in video-mediated communication by having dyads (groups of two users) perform negotiation and decision-making tasks and complete questionnaires. Similarly, Mühlbach, Böcker, and Prussog (1995) examined stereoscopy and visual angle on related tasks for dyadic groups. Green and Williges (1995) used factorial laboratory studies to determine the benefits of different communication media in a shared word processing application by measuring task performance on a collaborative writing task. Experimental designs can examine metrics other than task performance. For example, Anderson, Newlands, Mullin, Fleming, Doherty-Sneddon, and Van der Velden (1996) measured task performance, conversation, and user satisfaction. Conclusions from this study indicated that a combination of methods provides the most complete picture of video-mediated communication.

These controlled laboratory methods offer several advantages. First, variability among different subjects can be minimized. Green and Williges (1995) for example, matched subjects according to prior familiarity, gender, dominance, and writing ability. Whether these pre-test measures controlled all or most of the variability associated with group work is not clear, but such measures are possible when using this experimental technique. In addition to subject variability, parameters such as physical environment, task structure and timing, scheduling, and rewards can be under some reasonable control of the experimenter.

Unfortunately, the advantages to this method can also be viewed as disadvantages. By controlling all the factors mentioned above, the study has the tendency to lose external validity. Tang and Isaacs (1993) and Isaacs and Tang (1994) claimed that such studies involve highly artificial working arrangements which may not extend to real situations. These studies make use of contrived tasks between strangers over unrealistically brief working conditions. Additionally, the motivation to use the technology and perform the tasks well is greatly reduced compared to organizational pressures and commitments ordinarily found in the workplace. Finally, these methods often measure an artifact of the communication (task performance and preference) as opposed to the communication itself, although these metrics are beginning to converge (Watts, Monk, and Daly-Jones, 1996). Ethnographic methods make a tradeoff for control and generalizability by measuring the detailed elements of communication and task performance under highly realistic working arrangements.

## Ethnographic Studies

The published research on studying video-mediated collaboration with ethnographic methods is sparse and varied. A few authors have employed fairly rigorous techniques, such as conversation analysis and task analysis, while most were lax in their data collection and method.

In an attempt to understand how tasks and the configuration of groups impacted communication in video conferencing applications, Kies, Kelso, and Williges (1995) evaluated a series of scenarios in which real-world meetings were conducted. By varying task and group configuration, the researchers were able to become acquainted with the usefulness of the technology for different group sizes, different visually-dependent tasks, and the role of audio. Equally important was the ability to suggest theoretical underpinnings of the observed behavior. While the method was not formal and the ethnographic procedures were less than rigorous, the study enabled the researchers to familiarize themselves firsthand with the limitations of the technology and consider areas in which research remains to be conducted.

Tang and Isaacs (1993) and Isaacs and Tang (1994) conducted a series of studies with the intention of understanding the role of video in CSCW and its effects on collaboration among group members. Data was collected on usage patterns of the desktop video conferencing prototype through embedded metering techniques and information was gleaned through the use of questionnaires and video tapes. The researchers were able to theorize about the behavior of the groups in the video-mediated meetings and suggest ways in which the technology could be improved and applied to appropriate tasks. The data were collected in an ethnographic manner by studying real groups performing real activities in their working environment. Despite the utility of the study, the sample was small and only limited data were collected for analyzing the activities, the interaction with artifacts, or the conversation between members.

An example of rigorous ethnographic methods applied to the study of video-mediated communication can be found in the work of Sellen (1992), O'Conaill, et al. (1993), Doerry (1995), and O'Malley, Langton, Anderson, Doherty-Sneddon, and Bruce (1996). These studies made use of conversation analysis to study the impact of various video communication media on the flow of conversation, establishing a direct interpretation of communication effectiveness.

Sellen (1992) conducted a one-factor repeated measures experiment, varying the communication medium between two prototype systems and face-to-face. Five conversation analysis measures were collected: turns, groups turns, speaker switches, switch time, and simultaneous speech. Twelve groups of four individuals participated at four separate nodes in the video-mediated conditions. No differences were found in the mean duration of turns per session or the distribution of turns among speakers. More interruptions, simultaneous speech, and speaker switches were found to occur in the face-to-face conditions. It appeared as though video-mediated communication inhibited participants from seizing control of the floor. From this relatively detailed analysis of conversation, Sellen was able to analyze the effects of the media with parametric statistical tests; from these analyses, direct interpretations of communication behavior were drawn with respect to the different media in a manner precluded by other research methods.

O'Conaill, et al. (1993) conducted a similar study, employing similar analyses in which ISDN and broadcast quality audio and video were compared to face-to-face interaction. The study examined 20 minute segments from 4-5 meetings under each communication medium, measuring conversation components such as backchannels, interruptions, overlaps, handovers, turn size, and turn distribution. While these formal measures allowed inferential statistics to be calculated, the study was highly contextual and realistic, but therefore failed to control for number of sessions under each medium, the familiarity of participants with each other and the medium, group size, and type of task. Equally problematic was that the majority of differences between the different communication media were based on the audio quality differences. The narrow bandwidth of ISDN (128 kb/sec of which only 90 kb/sec is allocated to the video channel) severely inhibits video performance, but also transmits audio at half-duplex. Face-to-face communication is obviously full-duplex in that anyone can speak at any time without interference, and the broadcast quality video was transmitted using standard full-duplex audio similar to POTS telephony using a handset. The implication of the hardware choices created a situation in which radically different video qualities were compared, but the differences between audio quality were primarily responsible for the differences found in the conversation parameters measured. Relevant to the present study, it was found that there were more speaker switches in a face-to-face condition than high and low quality video, supporting Sellen's (1992) finding. It was also found that there was no significant difference between conditions for overlapping speech. Despite the lack of control over experimental conditions and audio performance, this study demonstrated the use of rigorous conversation analysis in a highly contextual and realistic study on different communication media.

O'Malley, et al. (1996) conducted a series of controlled experimental studies and examined both task performance and conversation metrics. Results showed that audio-only communication produces fewer speaker turns and words per dialog than high-quality video-mediated communication, supporting Sellen's (1992) work. When delays of 500ms were introduced, though, performance suffered on a collaborative navigation task, but no differences were found in terms of conversation metrics. Watts, et al. (1996) also examined conversation analysis metrics in an experimental design. They claim that the benefits of ethnographic methods and experimental designs are being sought by HCI researchers who recognize the need for more holistic research approaches.

A recent ethnographic analysis for a computer-mediated communication study was conducted by Doerry (1995) in which communication breakdowns were the central focus. Recording problematic events in communication was found to be an effective means for analyzing the efficacy of different communication media. For example, it was found that face-to-face communication had significantly fewer breakdowns than audio-only communication and audio/video communication.

From these studies, one can see that ethnographic research methods apply rigorous and not-so rigorous data collection procedures and experimental control in the evaluation of different communication media. There appears to be a need for more meticulous research methodology in this area so that a more thorough understanding of realistic work practices can be established in conjunction with more precise measurements and analyses regarding the effects of different technologies on the nature of work.

## Field Studies of Media Spaces

Over the past ten years, a great deal of research has been conducted on prototype video conferencing systems or "media spaces" which consist of expensive combinations of high bandwidth networks, analog video, and large displays. While this work has been conducted at a variety of educational and corporate research organizations, four institutions stand out in terms of research generated and technical prowess: Rank Xerox EuroPARC, the University of Toronto, Bellcore, and NTT, Japan. Nearly all of this research has concentrated on the design and technical functionality of the systems, theories of communication and interaction, and observations of use by the designers.

Various research groups at Rank Xerox EuroPARC (Boring and Travers, 1991; Heath and Luff, 1991; Gaver, 1992; Gaver, Moran, MacLean, Lovstand, Dourish, Carter, and Buxton, 1992; Heath and Luff, 1992; Gaver, Sellen, Heath, and Luff, 1993; Dourish, 1993; Dourish, Adler, Bellotti, and Henderson, 1994) have evaluated the use of several different systems used at their site, including the prototypes RAVE, Polyscope, and VROOMS. Some of the issues examined through their systems include the importance of gesture and eye contact in video conferencing sessions, the affordances of video communication from an ecological point of view, privacy concerns, and the use of multiple views of a remote site.

Several researchers at the University of Toronto (Mantei, 1991; Mantei, Baecker, Sellen, Buxton, Milligan, and Wellman, 1991; Buxton, 1992) have implemented different systems in their research facilities for the study of group communication, telepresence, gaze, and conversation fluidity, among other issues. These researchers built systems published as "Hydra", "Picture in a Picture", and "CAVECAT" which were designed to understand the theoretical foundations of video communication.

The groups at Bellcore (Fish, Kraut, and Chalfonte, 1990; Cool, Fish, Kraut, and Lowry, 1992; Fish, Kraut, Root, and Rice, 1992; Hollan and Stornetta, 1992; ) have also studied video-mediated communication systems, known as "VideoWindow" and "CRUISER". These systems were developed and utilized by researchers at Bellcore to understand evaluation techniques, social artifacts, the design of video conferencing systems, and the theories behind making video a better medium than face-to-face contact.

Researchers in Japan (Ishii, Kobayashi, and Grudin, 1992 and Ishii and Miyke, 1994) have developed video conferencing systems which attempt to overcome the common eye contact problems with impressive technical solutions such as drawing on an angled video monitor displaying a remote participant. These systems, known as "Team Workstation" and "ClearBoard" provide future designers with notions of how traditional desktop video conferencing paradigms can be altered so that distributed meetings can minimize adverse video communication side effects.

In general, this collection of research has expanded our understanding of video-mediated communication in several ways. First, the technical feasibility associated with implementing high-end video networks and display systems has been presented for future designers to learn from the mistakes and accomplishments of early prototypes. Second, social scientists working at these institutions have elaborated on theories of communication and interaction between participants remotely located. Finally, guidelines for the configuration of video-mediated communication systems have been established, providing insightful suggestions for issues such as eye contact, audio localization, contextual cues, and the conveyance of gesture. Despite these achievements, the methodology associated with these studies is weak and lacks formalization, repeatability, and precision. Limited objective data have been collected which can be compared to similar studies. Few suggestions have been made for how best to study working groups using video technology, and the effects of experimentally manipulated technical parameters were not measured on repeatable and objective dependent measures.

Research on video-mediated communication has been reviewed according to the general method used. It has been shown that psychophysics can augment our understanding of specific technical parameters such as display resolution and frame rate. Controlled experimental designs can be used to understand which configurations increase a sense of presence or which communication media are best suited for certain tasks. Ethnography can allow group meetings to be characterized in a direct and formal fashion as well as provide a framework for studying the effects of different media on conversation dynamics. Finally, field studies of media spaces push the envelope of technology and provide an environment in which theories of communication and interaction can be suggested and tested. This review has also demonstrated that a variety of methodological approaches can often be conducted within the same piece of research. Communication can be measured in terms of task performance, work structure, conversation analysis, and subjective ratings. The next section will discuss what has been learned from these techniques in terms of video-mediated communication.

## Video Communications

The study of video communications has been an ongoing interest for human factors engineers since the introduction of the Video Phone at the 1964 World's Fair. Video has often been viewed as a means by which the relatively impoverished communication through the audio channel can be greatly augmented. This section will explore the relevant research on video communication in terms of theory, visual channel characteristics, and the study of technical parameters related to video-based communication.

### Theoretical Background

Three main theories have been set forth to explain human communication under differing modalities. The social presence theory and the media richness theory seek to distinguish media in terms similar to classic information theory. For example, higher bandwidth media is theorized to improve communication relative to low bandwidth media. These theories view the human as relatively isolated and context-free. Another paradigm, the social information processing theory has sought to explain the phenomenon of communication media selection and the role of group and organizational influences on the use of different media.

#### *Social Presence Theory*

One goal of video conferencing systems is to establish a sense of "presence" akin to that of a face-to-face meeting. Social presence (Short, Williams, and Christie, 1976) is defined as a quality of a communication medium that depends upon its capacity to transmit verbal and nonverbal cues as well as aspects such as the apparent distance or realness of the participants. Telepresence, a related concept defined by Böcker and Mühlbach (1993), is the transferability of cues that allow one to get the impression of sharing space with a remote site. Finally, communicative presence (Böcker and Mühlbach, 1993) is the capacity of a system to transfer mutual verbal and non verbal communicative signals of participants. Taken together, these concepts define a general impression by which participants feel as if they are sharing space with a remote site. It is hypothesized that for situations in which participants feel more "presence", communication effectiveness will be increased. Furthermore, it is hypothesized that constraints such as frame rate can enhance or facilitate "presence", thereby improving the communication process.

The social presence theory claims that media such as audio are offer less presence than video which offers less presence than face-to-face interaction. Therefore, an extension of the theory would predict that low frame rate conditions allow less social presence than high frame rate conditions.

*Media Richness Theory*

Related to the social presence theory, the media richness theory (Daft and Lengel, 1986), explains how the richness of the medium affects the affordances for feedback, number of channels used, personalization, and variety of language. Video, which allows facial cues and body gestures to be transmitted is considered a rich medium, while text is on the opposite end of the spectrum. The degree of richness is theorized to impact the groups' ability to resolve uncertainty or equivocality among members in working toward a shared goal.

These two theories are relevant models for discussing the manipulation of video quality parameters such as frame rate. For example, low frame rate conditions are less-rich than high frame rate conditions, as the media-richness theory would suggest.

*Social Information Processing Theory*

In response to the social presence theory and the media richness theory, which focus on the characteristics of the media, Fulk, Steinfeld, Schmitz, and Power (1987) postulated the social information processing theory. They propose that the dynamics and past experiences of the group play a large role in the acceptance of various communication media; thus, a shared view of socially-constructed meaning is established from the use of a particular medium. The related theory of symbolic interactionism proposed by Trevino, Lengel, and Daft (1987) suggests that the symbolic messages inherent in the use of a medium combine with its richness to explain the managerial choice of different media in organizations. These perspectives are particularly germane to the proposed research in that individuals will be using a video conferencing system simultaneously, and the utility, quality, and status of the system as viewed by one participant may influence or permeate the opinions of the others.

While these theories have been shown to support observations of group behavior, they do not fully account for all the studies on communication in CSCW. For example, they fail to account for the consistent finding that video is more similar to audio communication than face-to-face communication. Similarly, these theories have trouble accounting for large individual differences in degraded video acceptability. Although more research is needed to help establish new theories and validate or discard existing theories, these models can be used in designing new CSCW systems as well as helping to explain findings from summative evaluations and basic group behavior research.

## The Visual Channel in Dyadic Communications

To understand the communication process between two individuals, it is necessary to ground the discussion in the fundamental aspects of human communication. Once these components are understood, the means by which a video conferencing system degrades communication can be investigated better.

The role of the head and face is crucial in face-to-face communication. Harrison (1974) claimed that four factors are most critical to influencing how the head and face are involved in the communication process. First, the face provides general appearance cues such as demographic information, status and occupation, fleeting facial expressions, and emotion. Similarly, subtle changes in facial expression and head nods can evoke powerful influences on the communication process. Finally, the behavior of the eyes has the ability to influence conversation fluidity and perceptions of the other participant.  The importance of gaze in human communication should not be underestimated.  Argyle (1975) and Argyle and Cook (1976) claimed that the face is the most important area for non-verbal signaling.  It was estimated that gaze was used 75% of the time for listening tasks and 40% of the time for speaking tasks.

Harrison (1974) discussed the importance of the hands and body in the communication process, and postulated four key factors.  The appearance of the body can project salient cues such as demographics, and more subtle information, such as personality.  Similarly, the posture of an individual can reveal the confidence level, interest, and tension level of a participant.  Harrison also argues that hand movements can play a vital role in the communication process.  For example, the hands can be used for illustration, the regulation of conversation, affect, and emblems (icon-like gestures such as a thumbs up to signify something positive).

Understanding the role of the aforementioned visual cues will enable one to understand how a video conferencing session may degrade the transmission of visual information.  First, visual cues promote smooth transitions between speakers (Rutter and Stephenson, 1977).  Some of the signals that are used to take the floor include paralinguistic cues such as audible inhalation and over-loudness and body motion cues such as head shifts and the initiation of gesticulation (Duncan and Niederehe, 1974).  Duncan (1972) found that body motions play a critical part for the promotion of turn-yielding.  However, there is some controversy in the literature in that some studies found no communication difference between audio-only and visual-audio media (Cook and Lalljee, 1972).  Indeed, the work of Chapanis (1975) confirmed that the most critical component for human communication is the auditory channel.

## Visual Cues in Video Conferencing

The investigation on the importance of visual cues in a video conferencing environment can be understood within the context of the ecological affordances offered by the medium.  Gaver (1992) emphasized this theoretical perspective - first advocated by J. J. Gibson (1979) in basic perception studies - and outlined three primary affordances for vision.  First, video offers a restricted field-of-view on remote sites which limits peripheral vision and constrains perceptual exploration of the scene.  Second, video has resolution artifacts not encountered in real life.  For example, a closer examination of the video display does not yield more detail, rather, it reveals the grainy structure of the image.  Finally, video conveys a limited amount of information about the three-dimensional structure of the remote site, which limits our exploration, inspection, and peripheral awareness.

In light of these video-induced limitations, it is important to understand how visual cues are transmitted and received through a video-mediated technology.  Conversational communication between individuals relies to a large extent on visual modalities, subtle and otherwise.  Facial cues and expressions are a source of considerable information when communicating face-to-face.  For example, Bruce (1996) found that viewing the face can aid the understanding of degraded speech quality.  Body movements and gestures add much to the communication session.  A primary advantage of video conferencing systems over traditional voice-only communication is that it attempts to transmit these visual cues to create a richer communication session.  Many researchers of late have examined various aspects of desktop video conferencing systems and how the social and technical aspects of these systems impact human communication.  One study by Heath and Luff (1991), using the extensive video conferencing facilities at Rank Xerox EuroPARC, emphasized the role of gesture in a multi-media office environment.

According to Heath and Luff (1991), gesture in face-to-face communication settings is constructed of subtle nuances, is contextually-utilized, and can perform a variety of functions (at times, simultaneous). Gesture plays several important roles, such as illustration, attention solicitation and maintenance, and conversation control. Contrasting the methods by which people utilize gesture in face-to-face communication settings to those in a video-mediated system allows the influence of video-mediated communication to be studied.

Heath and Luff (1991) found that altered gesture behaviors occurred when participants in the desktop video conferencing experiment attempted to garner the attention of other participants. Typically, in a face-to-face situation, one attempts to make eye contact by glancing at the other participant prior to speaking. In a video-mediated system, such fine gestures are not adequately transmitted, necessitating more drastic measures such as hand waving. Once a session was established, gesture played other roles. Participants often attempted to implement gesture in a manner similar to that which they would normally do so in a face-to-face situation. However, it was found that often these gestures which promoted "co-participation" were not received well by the remote party, resulting in a variety of coordination problems. When gesture was needed to illustrate a point, the communication effectiveness was found to break down. Similarly, when gesture was used to gauge the attentiveness of the other participant, conversation became stilted or abandoned when the acknowledgment cues were not returned.

Another interesting observation made by Heath and Luff (1991) was that gesture in face-to-face settings tends to be located in the visual periphery, not the direct line of sight. Video-mediated gesture, however, is much smaller in the field of view, and thus, only gross motor movements can be transmitted effectively. The technology was found to alter the size, shape, and pace of the gesture. It is here where the frame rate factor is most important. By slowing the frame rate, such gross motor movements can be exaggerated or distorted to affect communication adversely. As such, it is critical that the minimal gesture cues capable of being transmitted be displayed as faithfully as possible.

Despite all the research suggesting the crucial role visual cues play in communication, the bulk of evidence supports the notion that task performance is not adversely affected by a lack of visual channel. The studies citing poor transmission of visual cues through video-mediated communication are cast in doubt due to considerable literature which reports that only the audio channel is crucial for successful communication (Ochsman and Chapanis, 1974; Chapanis, 1975; Conrath, Dunn, Bloor, and Tranquada, 1977; Gale, 1990, 1991; Masoodian, Apperley, and Fredrickson, 1995). The picture is, however, not complete without a careful examination of the tasks measured and the user opinions garnered from these and other studies of video-mediated communication.

## Degraded Video Quality

Video-based communication is more complex than audio-only communication, and therefore is impacted by numerous technical characteristics of the specific hardware and software used to deliver the signal. Digital video requires a massive amount of bandwidth, much more than common high-speed computer networks can support. For example, one second of uncompressed, full-motion digital video using a standard image size and color palette, requires upwards of 15 megabits per second, which is more bandwidth than most networks in use today. Despite the use of compression, which can reduce that requirement tremendously, technical parameters such as the temporal and spatial sampling, video window size, color, and number of participants can become restricted. This section will discuss one especially critical parameter, frame rate.

The frame rate of a video system is critical to the transmission of visual cues. Common desktop video conferencing systems operate at reduced frame rates due to relatively low bandwidth data connections (e.g., traditional LAN ethernet operates at 10 Megabits per second), the current state of the art in compression algorithms, and the limiting factor of current computer processors. Lower frame rates result in "choppy" or "broken" motion, which may preclude complete understanding of visually-presented information or decreased user satisfaction.

The successful perception of motion is a key ingredient in the design of multimedia computer systems. The type of motion referred to here is not the natural or biological motion of everyday life. Rather, it is apparent motion, which uses the presentation of sequential static images to create the illusion of movement. The technique requires that successive images be displayed at a proper rate and are not significantly different from each other (Ramachandran and Anstis, 1986). Research on apparent motion was begun in the late 1800's, but landmark research was done in the early half of the 20th century by the Gestalt psychologist, Wertheimer. Wertheimer studied the apparent movement of two lines displayed rapidly in sequence, and concluded that the phenomenon was due to a short circuit between the retina and higher brain processing which did not make attentional demands. He was incorrect in this conclusion (Hochberg, 1986) as the current explanation describes the responses of direction-sensitive neurons and the role of the neuron's receptive field (Sekuler and Blake, 1990).

Designing multimedia systems to adequately incorporate realistic apparent movement has special considerations for both animation and video. First, animation sequences can be developed in two manners. The Disney technique of drawing Donald Duck, for example many times with minor changes in each successive frame is highly effective, but also expensive and time-consuming. Another approach is know as the Fred Flintstone Effect (Bregman and Mills, 1982), which capitalizes on the brain's ability to synthesize discrete pictures in context, allowing animators to suffice on drawing only three positions of Fred's legs. Although the motion is somewhat jerky, it is effective and relatively inexpensive.

In designing video displays, a critical factor is the frame rate at which the video is refreshed on the screen. The National Television Standards Committee (NTSC) which governs television broadcast in the United States and Japan has set a 30 frame per second (fps) standard. By using interlace, the refresh rate is effectively 60 Hz, which is above the critical flicker fusion (CFF) frequency for all but very bright images. To avoid flicker in nearly all cases, 80 or 90 Hz refresh rates are required (Schreiber, 1984). Despite flicker and jerky motion, many studies show that frame rates around 5 fps are acceptable for a variety of video conferencing tasks (Tang and Issacs, 1993; Pappas and Hinds, 1995). To make the issue more confusing, studies have shown that people are most sensitive to flicker around 10 Hz (Sekuler and Blake, 1990). It is not clear how the subjective degradations of frame rate interact with CFF in video displays.

Few studies have examined the relationship between frame rate and task performance or subjective assessment of the video conferencing session. Masoodian, Apperley, and Fredrickson (1995) conducted an experiment in the spirit of the communication studies of the 1970's (Ochsman and Chapanis, 1974; Chapanis, 1975) and concluded that for an intellectual task (based on McGrath's (1984) taxonomy), the reduced frame rate video condition had no impact on task performance. However, they failed to detect any difference between the various communication media on their choice of task and failed to observe tasks indicative of normal work requirements.

Although Tang and Isaacs (1993) did not vary frame rate in their studies of video conferencing in collaborative activities, they did use a reduced video frame rate (5 fps). They learned that although the reduced frame rate was clearly noticeable, it did not adversely affect the users' opinions of the system. Few studies have directly examined user performance or preference in reduced video frame rate conferencing applications.

Pappas and Hinds (1995) examined the tradeoff between frame rate and level of grayscale. They found that when frame rates dropped below 5 fps, subjects were willing to make tremendous compromises such as choosing binary gray level over full gray scale in order to avoid these low frame rates. They concluded that 5 fps is the critical level at which video quality becomes highly objectionable.

Another study, conducted by Apteker, Fisher, Kisimov, and Neishols (1994) attempted to assess the watchability of various types of video imagery under different frame rate degradations. They manipulated the temporal nature of the video and the importance of the video and audio information to understanding the message. Compared to a benchmark 100% "watchability" for 30 fps, video clips high in visual importance dropped to a range of 43% to 64% watchability when displayed at 5 fps, depending upon the audio importance and temporal nature. Overall, subject assessments of the video clips dropped significantly with each 5 fps decrease in frame rate (5, 10, 15 fps). These data suggest that although the video will certainly degrade in subjective watchability over decreased frame rates, the ratings are highly contingent upon factors such as the importance of the audio or video information to message comprehension as well as the temporal nature of the imagery.

Due to potentially limited bandwidth availability in a space station project, Chuang and Haines (1993) were interested in determining the frame rate necessary for monitoring laboratory rats through a video link. Using the method of limits, a psychophysical method, the authors assessed frame rate for slow scenes (depicting minimal rat movement) and fast scenes (depicting considerable rat movement). Two primary measures of acceptability were reported. The first measure was the lowest frame rate which yielded an “acceptable” rating by subjects. No statistically significant difference resulted between the slow and fast scenes. The other measure was the highest frame rate which was deemed unacceptable by subjects. In this case, slow scenes resulted in an average rating of 6.2 fps and the fast scenes resulted in an average rating of 4.8 fps, a difference which was not reported in terms of statistical significance. This report provided a poor explanation of data analyses, may have produced results biased by a methodological artifact, and did not seek to explain why a difference between slow and fast scenes occurred. Nonetheless, the study suggests that between 4 and 6 fps may be a crude estimate for video quality acceptability.

Their research also supports the tendency for highly experienced subjects to require slower frame rates to obtain the needed information from the video. In other words, considerable individual differences are present and may be explained by the prior experience the subject has had with the task domain. Although these data may not extrapolate well to other task domains, frame rates of approximately 6 fps are considered acceptable. These data agree with Tang and Isaacs (1993) and suggest that video conferencing sessions could also be conducted at rates well below the 30 fps NTSC standard.

A recent study by Kies, Williges, and Rosson (in press, b) employed laboratory and field research to study the effects of reduced bandwidth on image quality in a distance learning application. In a simulated college course, frame rate was manipulated and dependent measures were taken in terms of quiz performance and subjective assessments. The results indicated that performance does not suffer under reduced image quality conditions, but subjective satisfaction plummets. A field study was then conducted to validate these findings in a shared lecture series using desktop video conferencing over the internet. Quiz performance was relatively unchanged, but again user satisfaction of low frame rates was unacceptably low for continued use of the system.

The constraints of limited bandwidth are of concern and have been studied in other domains such as undersea telerobotics (Deghuee, 1980), interactive educational multimedia (Christel, 1991) and unmanned aerial vehicles (Swartz, Wallace, and Tkacz, 1992). These studies may provide some clues regarding how users will be able to communicate successfully via sub-optimal frame rates. From these studies, it is apparent that research is needed to investigate a potential task performance-frame rate relationship, as well as how reduced frame rates may affect the work process. No research has systematically varied frame rate of a range of levels in desktop video conferencing settings.

# Chapter 3.  Psychophysical Study

## Purpose

A thorough review of the research on video frame rate degradations has shown that few studies have measured user perceptions of image quality over a comprehensive range of stimuli. Likewise, few studies have employed rigorous psychophysical methods to determine image quality for degraded video. It has also been shown that desktop video conferencing will operate under bandwidth conditions which will significantly impair video quality. As a result, there is a clear gap in the human factors literature. The purpose of this study was threefold:

- To establish a psychophysical scale of perceived image quality for an extensive range of frame rates.
- To determine if level of detail, speed of panning, or scene content affects perceived image quality.
- To determine if video image quality perception fits Steven's Power Function, the most robust psychophysical law.

It is hoped that this study will allow network administrators and system engineers to make decisions regarding bandwidth support for desktop video conferencing systems.

## Method

The first study utilized the classic psychophysical methods, free-modulus magnitude estimation and category scaling to create a ratio scale and an ordinal scale for video image quality across a range of frame rates and scenes. Subjects, materials and equipment, procedures, and data analysis will be discussed.

### Subjects

Thirty college-aged subjects participated in this study. Although Stevens (1975) claimed that only 10 subjects are needed, the large sample size was used due to the complexity of the stimulus and the variability in acceptability judgments found in other studies (i. e. the psychophysical scaling study by Chuang and Haines, 1993; distance learning simulation study by Kies, Williges, and Rosson, in press, b). The average age of the subjects was 26.43 (standard deviation of 6.95). Nineteen participants were male and eleven were female. Subjects were recruited through Virginia Tech Usenet Groups, signs posted around campus, and word-of-mouth. Participants were paid $5 per hour for their time and were screened for normal color vision and 20/20 acuity.

### Materials and Equipment

Video clips were digitized and displayed on a Sun SPARC 5 UNIX workstation equipped with a video digitizing board and Sun 17” display. Sun’s Show Me conferencing software was used to manipulate the frame rate. All scenes were compressed using the Cell B method developed by Sun Microsystems. The method divides the image into four-by-four regions of pixels represented by two colors and a 16-bit mask which indicates which of the two colors to place at each of the resulting 16 pixel positions. The colors are chosen to preserve the mean and variance luminance and the average chrominance for the region (Sun Microsystems, 1997). The compression scheme is designed for rapid decompression and display of images by using both block truncation coding and interframe compression. The number of bits used to represent each pixel ranges from approximately .33 bits/pixel to approximately .75 bits/pixel and is dependent upon the amount of motion in the image. The system uses 8 bits per pixel when no compression is used. The resulting compression ratio is roughly 1:15, but varies depending on scene detail and movement. The bandwidth needed to display images with 320 x 240 addressability and 200 colors (the quality levels used and kept constant throughout the study) with Cell B compression at 30 fps is approximately .70-1.5 megabits per second. The bandwidth needed to display the same image with no compression is approximately 17 megabits per second. The compression scheme appeared to refresh the screen with a complete image, as opposed to some schemes which update pixels as they change, resulting in images with black blocks. As far as could be determined, however, the compression scheme was not dependent on frame rate.

The details of the color quantization scheme, the method by which colors are encoded into a table of coefficients, was not known for Cell B compression. It is possible that local discolorization could have resulted from the assignment of chrominance values to the quantizing table. However, these compression artifacts were not expected to unduly bias observers' evaluations of image quality. In addition to compression, a number of other considerations have the potential to influence the perception of digital video quality. The illuminance in the room and the brightness on the display were both held constant throughout the course of the study. Likewise, the contrast setting on the display was maintained at the same level during the study.

The configuration is depicted in Figure 4.1. A VCR was connected to the workstation via coaxial video cable at which point the analog video signal was digitized. An instruction sheet (see Appendix A) was furnished to each of the subjects. Table 3.1 provides descriptions of the independent variables (video clips typology). The fast scenes panned two times the speed of the slow scenes. The lighting, focal length, and distance between the object of interest (bicycle) from the camera was the same in all conditions. The slow scenes panned across the object(s) twice in approximately 12 seconds. The fast scenes panned across the object four times in approximately 12 seconds (two times faster than the slow scene). A fifth scene depicting a close-up of a talking head was used to represent a traditional video conferencing task. All scenes were silent except for the talking head. Figure 3.1 depicts the high and low detail scenes, as well as the "talking head" scene (all were actually shown in color for the experiment). Figure 3.2 depicts the approximate panning angles for filming the object.

## Procedure

The first step was to administer a visual acuity eye test with a Bausch and Lomb Orthrater. Each subject was screened to have 20/20 corrected or uncorrected vision and normal color vision to participate. The experimental procedure began by orienting the subjects to the purpose of the experiment and having them practice the magnitude estimation technique on lines, as suggested by Stevens (1975) and described in the instruction sheet supplied in Appendix A. Following this procedure, the subject was queried for questions and asked to begin the experiment. Each subject was randomly assigned to an order of scene presentation according to the Latin Square depicted in Table 3.2.

The first scene was presented 20 times. The first ten presentations were randomly presented, differing from 1 to 28 fps, for every third frame rate value (1, 4, 7, ...28). The second ten presentations were identical, but randomized again. After each video clip, the subject responded to two questions. The first question asked for a standard magnitude estimation value of acceptable video image quality. Secondly, the subjects were asked to rate the acceptability of the video image quality according to the scale depicted in Figure 3.3. This is a standard category scaling exercise similar to that implemented by Chuang and Haines (1993). The response alternatives in Figure 3.3 were selected so that intervals between phrases were as equal as possible, according Meister (1985).

After the subject completed the two rating questions, the next video clip was shown. After all 20 video clips had been shown, the subject was given a brief break before the next scene is initiated. After all five scene conditions had been presented, the subject was asked to respond to a brief post-experiment questionnaire. Finally, the subject was paid and dismissed. Each session lasted approximately 1.5 hours.

TABLE 3.1 Scene Types

| *Speed of Panning* | *Level of Detail* | |
|---|---|---|
| | *Low Detail* | *High Detail* |
| *Slow Scene* | 1) Slow pan of a bicycle in front of white wall. | 2) Slow pan of bicycle in front of a cluttered garage. |
| *Fast Scene* | 3) Fast pan of bicycle in front of white wall. | 4) Fast pan of bicycle in front of a cluttered garage. |

FIGURE 3.1 Low and high detail scenes (top) and "talking head" scene (bottom).

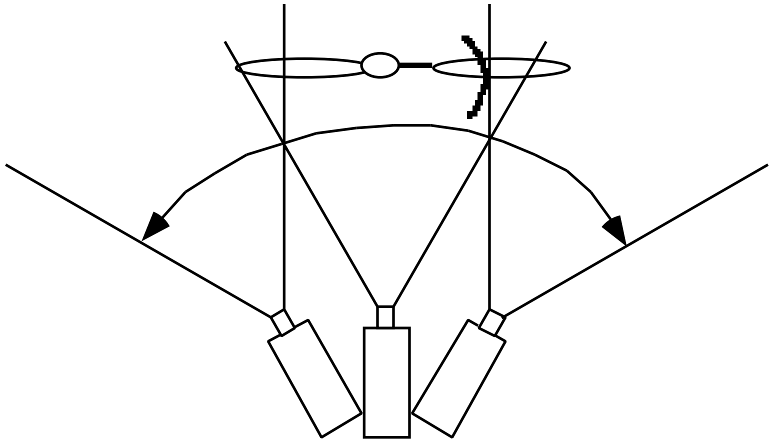


FIGURE 3.2 Filming arc of object(s) for experimental scenes. Cameras are shown with approximate angles. The distance between the camera and the bicycle was approximately 15' The figure is not drawn to scale.

TABLE 3.2 Latin Square for Stimulus Presentation Order. Two Latin Squares are required because there is an odd number of treatments (5).

| | *Subjects* | | | | |
|---|---|---|---|---|---|
| | *S1-S3* | *S4-S6* | *S7-S9* | *S10-S12* | *S13-S15* |
| | 1 | 2 | 3 | 4 | 5 |
| | 2 | 3 | 4 | 5 | 1 |
| | 5 | 1 | 2 | 3 | 4 |
| | 3 | 4 | 5 | 1 | 2 |
| | 4 | 5 | 1 | 2 | 3 |
| Treatment | | | | | |
| | *S16-S18* | *S19-S21* | *S22-S24* | *S25-S27* | *S28-S30* |
| | 4 | 5 | 1 | 2 | 3 |
| | 3 | 4 | 5 | 1 | 2 |
| | 5 | 1 | 2 | 3 | 4 |
| | 2 | 3 | 4 | 5 | 1 |
| | 1 | 2 | 3 | 4 | 5 |

Note: Number indicates scene type according to Table 3.1; scene 5 is the talking head scene.

**The video quality was acceptable:**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree |

FIGURE 3.3  Category Scale

# Results

The results from the psychophysical study have been divided into two major analyses. First, the magnitude estimation data will be discussed in terms of data standardization, trend analysis, the calculation of Steven's Power Function, and analysis of variance procedures. Consequently, the supplemental analyses involving category scaling will be presented both separately and with respect to the findings from the magnitude estimation procedure.

## Data Standardization

Collecting data using the magnitude estimation procedure generates a significant amount of data quickly.  However, a set of complex mathematical transformations must first be conducted to prepare the data for analysis of trend, Steven's Power Function, and parametric statistical procedures.  Standardization is particularly necessary with magnitude estimation data because the method dictates that each participant use a subjective scale free of modulus. Therefore, estimates of a given stimulus vary considerably across participants.  Several studies have suggested methods by which the standardization process can be conducted (Stevens, 1975, Gescheider, 1985; Engen, 1971, and Morrison, 1995).  On the basis of these studies, the following standardization process was conducted.  The data shown in all forthcoming examples is from line length estimations used by all participants for the sake of learning the method.  The same procedure was performed on the data from each scene type.

1.  Calculate the geometric mean of each of the two estimates for each subject. This data is displayed in Figure 3.4.

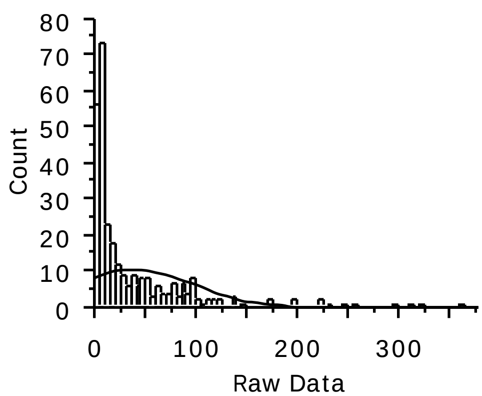



FIGURE 3.4  Raw Data (Geometric Mean of each pair of estimates for each subject)

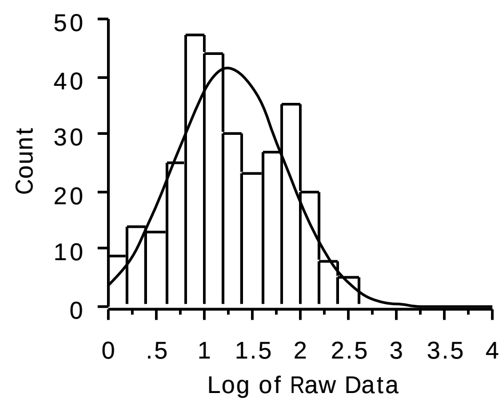



FIGURE 3.5  Log of Raw Data (1)

2. Calculate the log of each estimate in (1). Figure 3.5 depicts this data set.

3. Calculate the arithmetic mean of (2) for each subject.

4. Calculate the arithmetic mean of (3), the arithmetic grand mean.

5. Subtract the arithmetic grand mean (4) from each subjects' arithmetic mean (3).

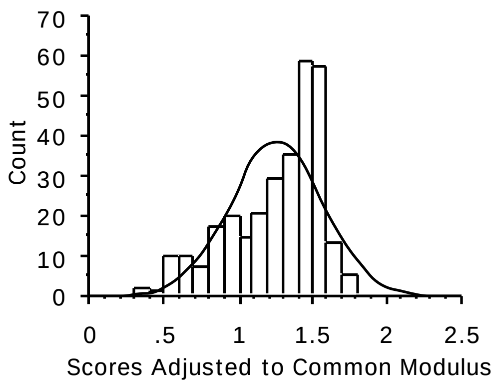



FIGURE 3.6 Log of raw data (2) - ((3) - (4)).

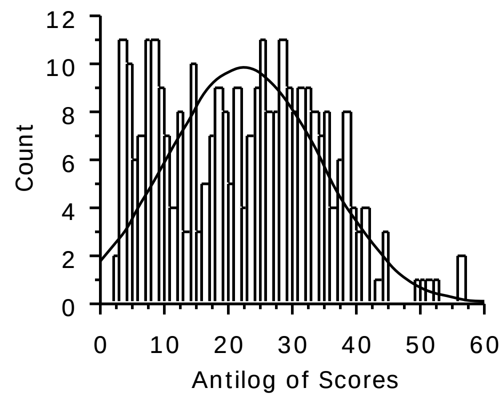



FIGURE 3.7 Antilog of (6).

6. Subtract (5) from each score in (2) so that all scores are adjusted to a common modulus. This data set is depicted in Figure 3.6.

7. Calculate the antilog of (6) so that all scores are put in positive units fit for parametric statistics. Figure 3.7 depicts this data set.

8. Depending upon the analysis to be conducted, a final step of calculating the arithmetic mean for each stimulus across all subjects can be taken.

While the data in Figure 3.7 does not appear to be distributed normally, the analysis of variance procedure has been shown to be robust to such violations (Keppel, 1991). This data set is now ready for trend analysis, the calculation of Steven's Power Function, and analysis of variance.

## Trend Analysis

The purpose for conducting a trend analysis on each scene's psychophysical function is to characterize the relationship between the stimulus and the psychological evaluation. This procedure does not test the significance of adjacent treatment means, rather, it seeks to explain the overall trend of the data across all levels of the independent variable. By employing the method of orthogonal polynomials, one can isolate different trend components from a function to determine their significance (Keppel, 1991). The following analysis uses this approach to test the significance of a linear, quadratic, and cubic trend for each of the five scenes tested in this study. Tables 3.2-3.6 depict the significant trends for each scene, the residual F statistics (the amount of variation which can be further isolated), and the percent of variation which can be accounted for by that trend.

TABLE 3.2 Scene 1 Trend Analysis Summary Table

| *Trend* | *Component Significance* | *Residual Significance* | *Percent Variation Between Conditions Explained* |
|---|---|---|---|
| Linear | F (1,290) = 245.44, p<.001 | F (8,290) = 9.55, p<.001 | 76% |
| Quadratic | F (1,290) = 54.00, p<.001 | F (7,290) = 3.20, p<.01 | 16% |
| Cubic | F (1,290) = 19.01, p<.001 | F (6,290) = .56, Not Significant | 6% |

TABLE 3.3 Scene 2 Trend Analysis Summary Table

| *Trend* | *Component Significance* | *Residual Significance* | *Percent Variation Between Conditions Explained* |
|---|---|---|---|
| Linear | F (1,290) = 398.58, p<.001 | F (8,290) = 14.96, p<.001 | 77% |
| Quadratic | F (1,290) = 97.55, p<.001 | F (7,290) = 3.16, p<.01 | 19% |
| Cubic | F (1,290) = 13.12, p<.001 | F (6,290) = 1.50 Not Significant | 2% |

TABLE 3.4 Scene 3 Trend Analysis Summary Table

| *Trend* | *Component Significance* | *Residual Significance* | *Percent Variation Between Conditions Explained* |
|---|---|---|---|
| Linear | F (1,290) = 240.78, p<.001 | F (8,290) = 13.56, p<.001 | 69% |
| Quadratic | F (1,290) = 67.81, p<.001 | F (7,290) = 5.81, p<.001 | 20% |
| Cubic | F (1,290) = 28.97, p<.001 | F (6,290) = 1.95 Not Significant | 8% |

TABLE 3.5 Scene 4 Trend Analysis Summary Table

| *Trend* | *Component Significance* | *Residual Significance* | *Percent Variation Between Conditions Explained* |
|---|---|---|---|
| Linear | F (1,290) = 254.21, p<.001 | F (8,290) = 11.11, p<.001 | 74% |
| Quadratic | F (1,290) = 54.80, p<.001 | F (7,290) = 4.87 p<.01 | 16% |
| Cubic | F (1,290) = 19.83, p<.001 | F (6,290) = 2.38 p<.05 | 6% |

TABLE 3.6 Scene 5 Trend Analysis Summary Table

| *Trend* | *Component Significance* | *Residual Significance* | *Percent Variation Between Conditions Explained* |
|---|---|---|---|
| Linear | F (1,290) = 263.51 p<.001 | F (8,290) = .636, Not Significant | 98% |
| Quadratic | -- | -- | |
| Cubic | -- | -- | |

From tables 3.2-3.6, several themes are evident. First, scene 5 appears to depart from scenes 1 through 4 in that it has only one significant trend component, the linear trend. The residual F statistic was not significant, indicating that there are no other trend components. The first four scenes had significant linear, quadratic, and cubic components. The linear component accounted for between 69% and 76% of the variance between the conditions. The quadratic component accounted for between 16% and 20% of the variance. The cubic component accounted for between 2% and 8% of the variance. The residual F statistic for the cubic component was only significant for scene 4, suggesting that higher order trends do not characterize the data for these scenes. A quartic trend was not calculated for scene 4 due to the difficulty of interpreting trends higher than a quadratic.

Figure 3.8 plots the standardized response values across subjects for each of the five scenes. The plot supports the results from the trend analysis in that the first four scenes appear to be characterized by at least a quadratic function, and the fifth scene, a linear function.

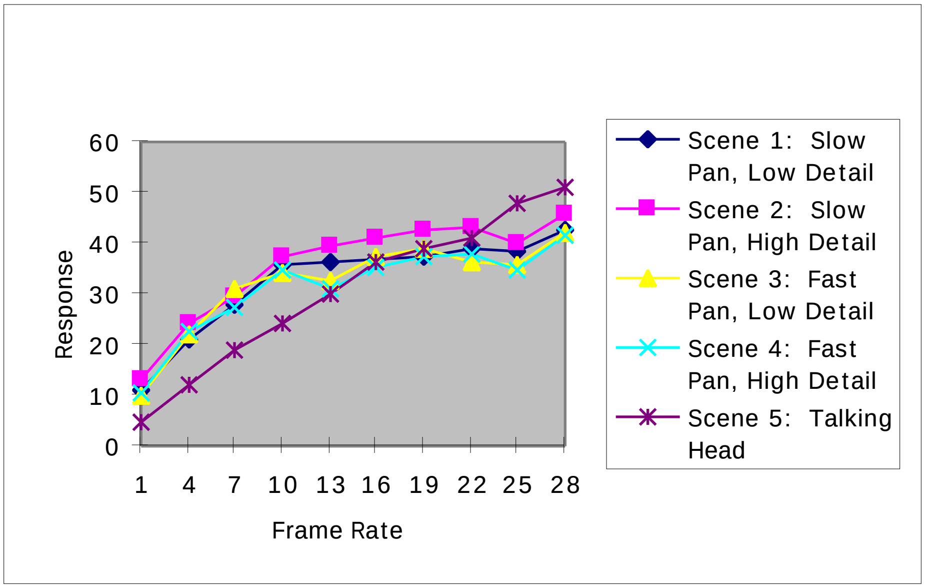



FIGURE 3.8. Magnitude Estimation Data for Each Scene.

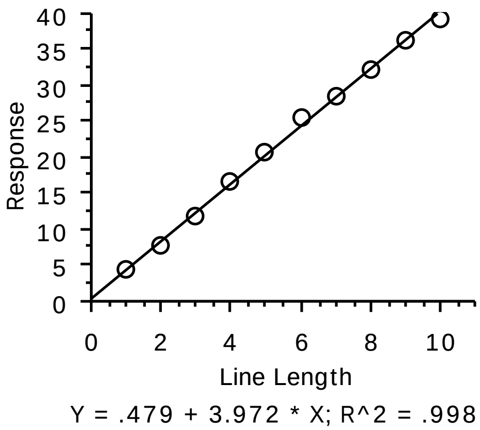
### Steven's Power Function

The most robust mathematical function for relating a wide range of physical stimuli to psychological scales is through Steven's Power Function. Stevens (1975) defined the relationship in the following equation:

$$\Psi = \kappa \Phi^{a}$$

where $\Psi$ is the sensation magnitude or psychological response, $k$ is a scaling constant, $\Phi$ is the physical stimulus, and $a$ is the power exponent which is contingent upon the modality and stimulus conditions. One objective of this study is to determine $a$ for different video scenes relative to frame rate. By mapping the physical stimulus of frame rate to the psychological experience of video image quality, one can use the magnitude estimation procedure to develop a relationship between these two variables suitable for regression analysis.

Figure 3.9 depicts the relationship between the standardized magnitude estimates across participants for each level of line length (frame rate will be depicted shortly). Clearly, there is a direct linear relationship between these two variables ($r^2$=.998). However, to determine the power function exponent, $a$, one must plot the log of both the stimulus and response and calculate the linear regression equation. The coefficient of this equation yields the power function coefficient. Figure 3.10 depicts the log-log relationship based on Figure 3.9, along with the linear regression equation. If the log-log plot shows a linear relationship, regardless of the non-log plot, the function fits the Steven's Power Function (Gescheider, 1985). Such is the case with the line length estimates.



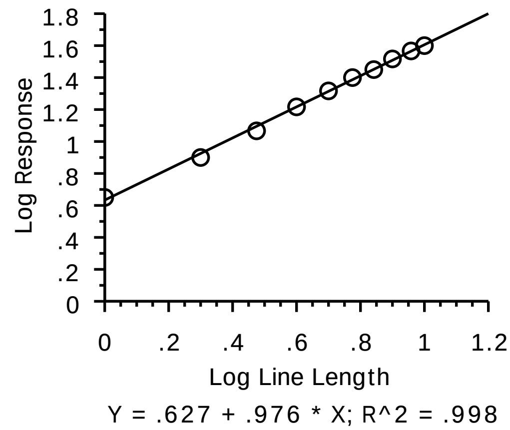




FIGURE 3.9 Line Length estimates plotted against actual lengths. Also shown is the linear regression equation of best fit.

FIGURE 3.10 Figure 5 plotted with the log of each axis. Also shown is the linear regression equation of best fit.

From these plots, one can determine the psychophysical power function:

$$\Psi = 4.23\,\Phi^{.976}$$

where 4.23 is a scalar (the antilog of the log-log plot intercept) and .976 is the power function exponent (the regression equation coefficient). This exercise can be conducted for each of the scenes viewed. Figures 3.11-3.20 depict the plots for each of the five scenes viewed in this study. All regression equations for non-log plots were fitted with no y intercept due to the ratio scale nature of the magnitude estimation method: a stimulus of 0 must be estimated with a psychological value of 0. Linear regression was conducted on the fifth scene, but all other scenes were fitted with a second order polynomial. The trend analysis in the previous section demonstrated that scenes 1-4 are characterized with at least a quadratic trend, while scene five is only characterized with a linear trend. The log-log plots, however, were not forced through the 0 point because the log of 1 equals 0 and to demonstrate a linear relationship between the variables for Steven's Power Function. The implications of this fit will be discussed in the following section.

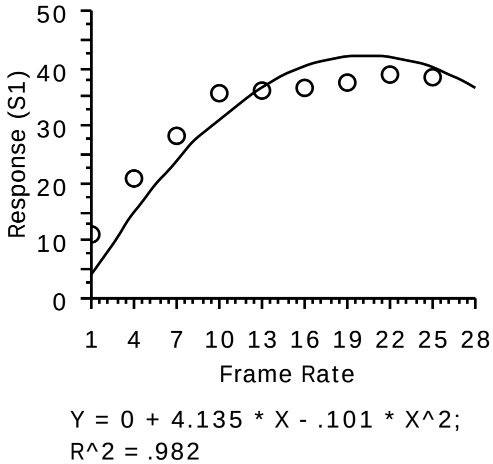



FIGURE 3.11 Subject estimates for scene 1 (slow pan, low detail).

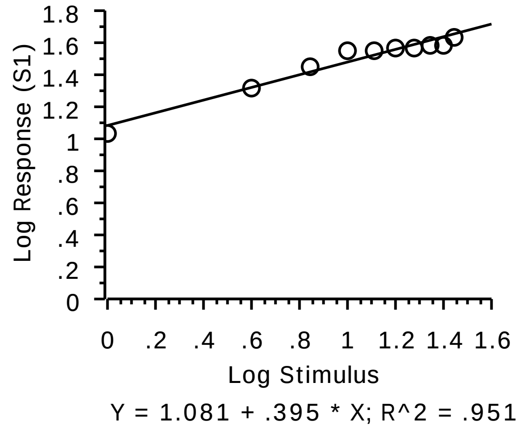



FIGURE 3.12 Log of subject estimates for scene 1 (slow pan, low detail).

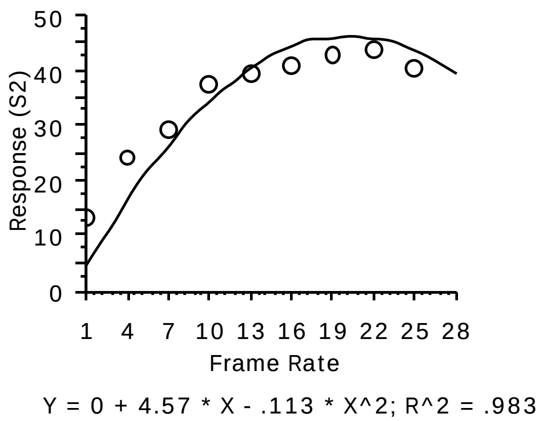



FIGURE 3.13 Subject estimates for scene 2 (slow pan, high detail).

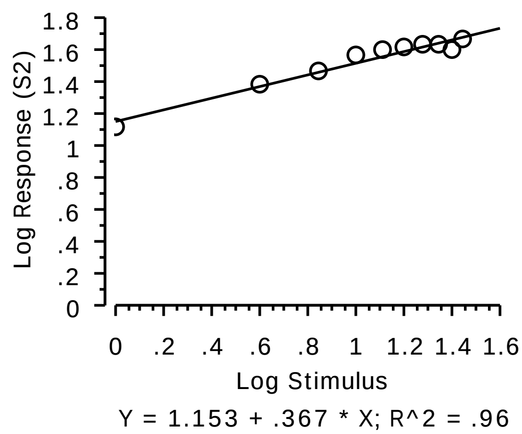



FIGURE 3.14 Log of subject estimates for scene 2 (slow pan, high detail).

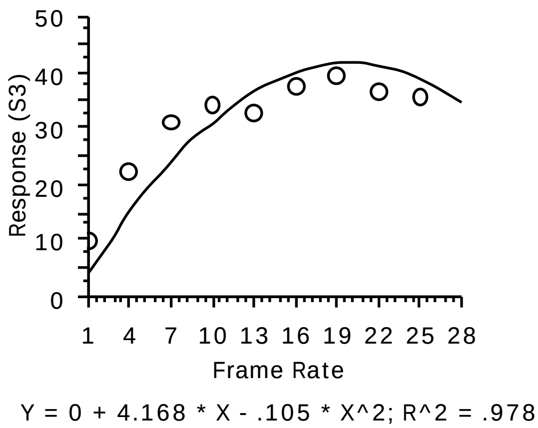



FIGURE 3.15 Subject estimates for scene 3 (fast pan, low detail).

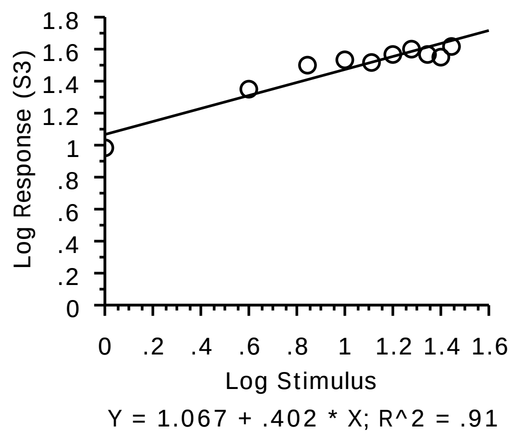



FIGURE 3.16 Log of subject estimates for scene 3 (fast pan, low detail).

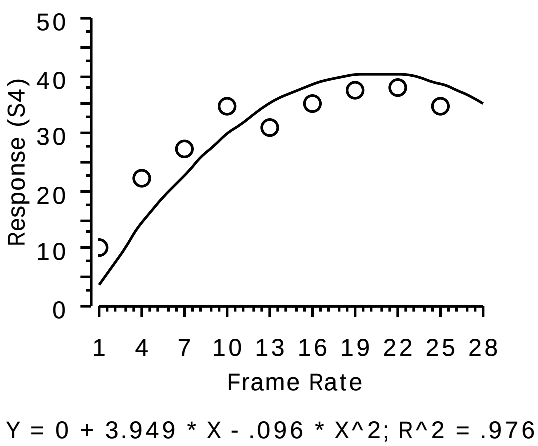



FIGURE 3.17 Subject estimates for scene 4 (fast pan, high detail).

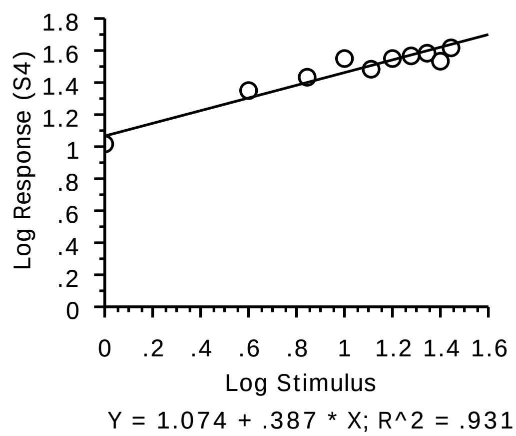



FIGURE 3.18 Log of subject estimates for scene 4 (fast pan, high detail).

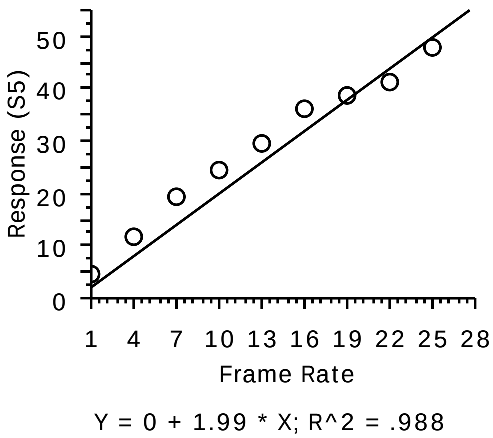



Figure 3.19 Subject estimates for scene 5 (talking head).

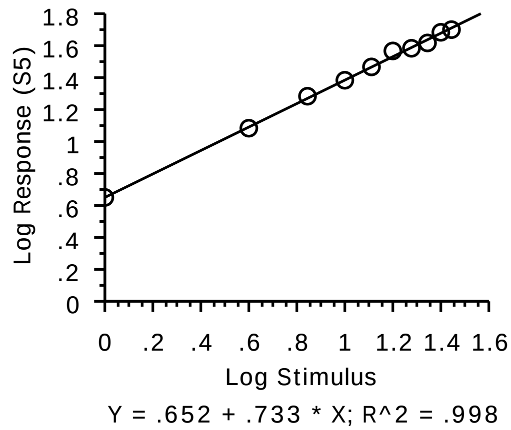



Figure 3.20 Log of subject estimates for scene 5 (talking head).

From figures 3.11-3.20, the high $r^2$ values on the log-log plots suggest that each of the scenes fits Steven's Power Function. Consequently, one can determine the psychophysical function for each scene. Table 3.7 shows each scene's power function exponent along with the line length estimate exponent. Also displayed are the means and standard deviations for each scene's standardized ratings. This table highlights an interesting finding not easily discerned from each of the above figures.

TABLE 3.7 Power Function Exponents and Rating Means and Standard Deviations

| *Stimulus* | *Description* | *Power Function Exponent* | *Mean* | *Standard Deviation* |
|---|---|---|---|---|
| Scene 1 | Slow pan, low detail | .395 | 34.73 | 13.64 |
| Scene 2 | Slow pan, high detail | .367 | 37.69 | 12.93 |
| Scene 3 | Fast pan, low detail | .402 | 34.06 | 12.96 |
| Scene 4 | Fast pan, high detail | .387 | 33.42 | 12.69 |
| Scene 5 | Talking head | .733 | 30.47 | 21.02 |
| Line Length | | .976 | | |

Table 3.7 suggests that the exponents for scenes 1-4 are similar. Scene 5 has an exponent much closer to 1, resulting in a more linear relationship. Likewise, the line length data produced an exponent very close to 1, resulting in a nearly perfect linear relationship. Exponents less than 1 produce curves which when the non-log values are plotted, results in a negatively accelerated function. Exponents near 1 are linear, and those greater than one are positively accelerated. Figure 3.21 depicts unitless data entered into each stimulus's power function.

Scenes 3 and 4 (the fast pan scenes) are nearly identical, while scenes 1 and 2 (the slow pan scenes) are higher on the y-axis. Clearly, the line length and scene five are the most linear of the curves.

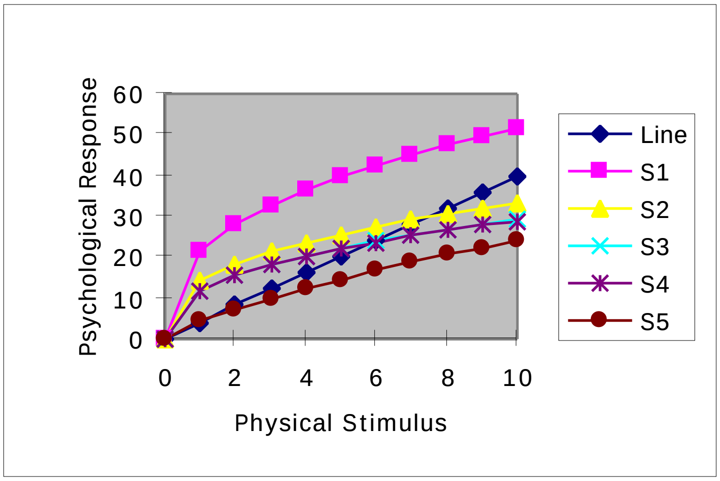



FIGURE 3.21 Psychophysical power function plotted for unitless data for each stimulus.

By calculating the functions for each stimulus, it has been shown that Steven's Power Function characterizes each psychophysical relationship. These data will now be used to determine which if any frame rates are statistically significant within and between the different scenes.

## Analysis of Variance and A Priori Contrasts

In order to determine if there was an effect of frame rate or type of scene on subjects' perceptions of video image quality in the magnitude estimation task, a 5 (scene) x 10 (frame rate) within-subjects analysis of variance was conducted and is displayed in Table 3.8. Table 3.7 suggests that the variance among the five scenes was not similar. To ensure that the homogeneity of variance assumption of ANOVA was not violated, an Fmax test was performed. Results indicated that the assumption was not violated. Significant main effects for scene and frame rate are reported, along with a significant interaction.

TABLE 3.8. ANOVA Summary Table for Magnitude Estimation Task

| *Source* | *df* | *Sum of Squares* | *Mean Square* | *F-Value* | *P-Value* |
|---|---|---|---|---|---|
| Subject | 29 | 18042.58 | 622.16 | | |
| Scene | 4 | 8063.72 | 2015.93 | 16.51 | .0001 |
| Scene x Subject | 116 | 14166.51 | 122.13 | | |
| Frame Rate | 9 | 165315.55 | 18368.40 | 85.13 | .0001 |
| Frame Rate x Subject | 261 | 56312.70 | 215.76 | | |
| Scene x Frame Rate | 36 | 13091.79 | 363.66 | 5.48 | .0001 |
| Scene x Frame Rate x Subject | 1044 | 69234.30 | 66.32 | | |

Due to the nature of the scenes, three comparisons were planned a priori: talking head scene vs. non-talking head scenes, fast panning scenes vs. slow panning scenes, and low detail scenes vs. high detail scenes. To determine if differences between these combinations of scenes were significant, linear combinations of treatment means were constructed and formal contrasts were computed. Table 3.9 displays the results from these analyses. Greenhouse-Geyser and Hundt-Feldt corrections were made to help control the problem of correlated means on repeated measures (Abacus Concepts, 1989). Likewise, an alpha level of .01 was established to reduce the impact of inflated Type I error probability associated with conducting multiple comparisons (Howell, 1992).

TABLE 3.9 A Priori Contrasts

| *Contrast* | *F Statistic* | *Greenhouse-Geyser Correction* | *Hundt-Feldt Correction* | *Significant at alpha=.01* |
|---|---|---|---|---|
| Talking Head vs. No Talking Head | F (1,116) = 39.83, p<.0001 | .0001 | .0001 | Yes |
| Slow vs. Fast Scenes | F (1, 116) = 14.99, p=.0002 | .0041 | .0036 | Yes |
| Low Detail vs. High Detail | F (1, 116) = 3.27, p = 073 | .0944 | .094 | No |

The results from the contrasts displayed in Table 3.9 indicate a significant difference between the talking head condition and the others, as well as a significant difference between the slow pan scenes and the fast pan scenes.

## Category Scaling Analyses

The category scaling portion of the experiment can be analyzed by averaging the two values for each frame rate for each subject, averaging across subjects, and plotting against frame rate on non-log paper. The following steps must be taken to transform the raw data:

1) Calculate the arithmetic mean of the two judgments for each frame rate for each participant.

2) Calculate the arithmetic means of the frame rate judgments across subjects.

3) Plot the mean frame rate judgments against each frame rate on non-log paper.

These data need not be standardized in the same fashion as for the magnitude estimation data due to the scale provided in the method. Figure 3.22 depicts the category scale data for each scene. The difference in trends between scene 5 and the other scenes evidenced in the magnitude estimation data appears to be replicated for the category scale. No statistical analyses will be conducted, as these data reflect an ordinal scale.

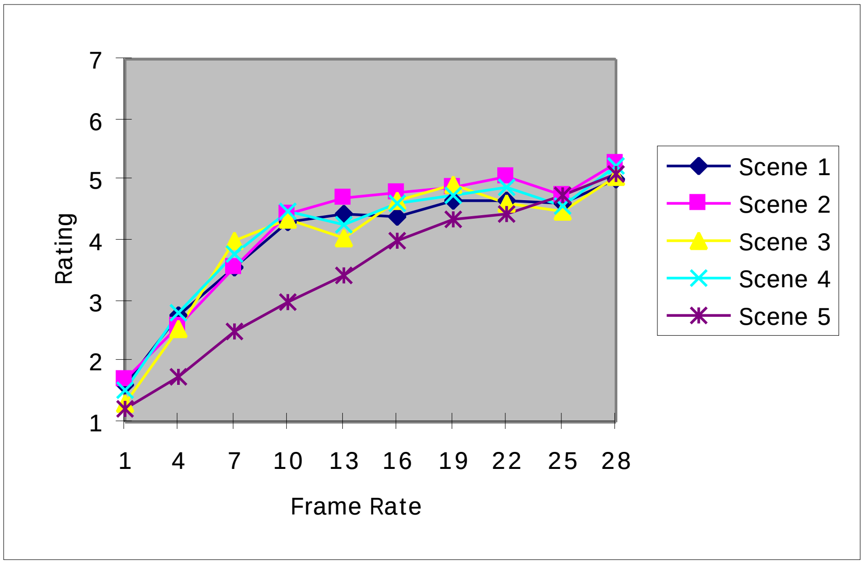



FIGURE 3.22 Category Scale Data for Each Scene.

## Discussion

The purpose of this study was to determine the psychophysical relationship between video frame rate and image quality. A free-modulus magnitude estimation study was conducted using thirty subjects. The results of this study point to several interesting findings. First, Steven's Power Function was upheld for this stimulus, providing evidence that metathetic stimuli will conform to the law. Second, the non-talking head scenes conformed to higher-order polynomial trends, while the talking head scene could only be characterized by a linear trend. Finally, an analysis of variance and a priori contrast suggested that the speed of panning and scene with a talking head affected subject's evaluations of the video quality.

Trend Analysis is a procedure which tests the significance of various trends in a data set using the method of orthogonal polynomials. The method first tests where a significant linear trend is present, it then tests the residual variance to determine if other trend components are present. If this second test is significant, the analysis looks for a quadratic trend and so on. The results from this analysis indicate that the non-talking head scenes could all be characterized by three trend components: linear, quadratic, and cubic. The talking head scene, however, could only be characterized by a linear trend. These results are interesting for a number of reasons. In the first four scenes, an elbow appeared at 10 frames per second, after which point the evaluations reached an approximate asymptote. At 25 frames per second, the curve appears to dip slightly, followed by an increase at 28 frames per second. This final curve likely resulted in the significant cubic component but is difficult to explain. One hypothesis is that the video conferencing software failed to produce frame rates at that level. However, evaluations of the talking head condition increased at this frame rate. Alternatively, the dip may derive from the variability in the data set; although the evaluations for scenes one through four all dipped slightly at this frame rate.

The curve at 10 frames per second, however, is more robust and easier to explain. It appears that subjects are better able to detect video quality improvements at frame rates less than 10 fps than at frame rates greater than 10 fps for the panning scenes. One explanation suggests that the subjects had no pre-conceived notion of how fast the camera was panning, and therefore, were not influenced by the increase in video quality. This interpretation is better understood in light of the talking head function. In this scene, a close-up view of a person talking (with audio) was shown. People have extensive experience watching others speak. Therefore, minor degradations in video quality can be detected in the subtle frame rate differences at the upper end of the frame rate range. Conversely, people have no prejudice regarding the proper speed at which a scene is filmed. Another factor - perhaps more significant - is the inclusion of an audio track for the talking head scene. At frame rates below the critical flicker fusion rate (all frame rates in this study), people are able to detect slight synchronization discrepancies between the audio and the video. The combination of these two factors likely explains the different trend components found in these two classes of video.

The results from this analysis have implications for the design of video-based communication systems. If the task involves a person talking as in traditional desktop video conferencing applications or distance learning, increasing the frame rate to the highest level supported by the system will be noticed. If the task involves viewing scenery (architectural walkthrough, surveillance), frame rates beyond 10 fps may not be detected by the user. Because these evaluations were made in terms of subjective evaluations and no task performance measure was taken, these results cannot be extended to predict how people will perform specific tasks. Nonetheless, these results can have implications for network engineers who view bandwidth as a precious commodity: it is suggested that for non-traditional uses (i.e. those not involving close-up views of people speaking), video over 10 fps is not worth the added cost. However, for most desktop video conferencing applications, increasing bandwidth allocation to support upwards of 30 fps may be noticed by the users, if not reflected by improvements to task performance. Further research needs to address specific task performance issues in relation to video frame rate.

In summary, non-talking head scenes were characterized by linear and higher-order trends. The linear and quadratic components were discussed in terms of subject's inability to detect increased frame rate beyond 10 fps for the panning scenes. The cubic trend was not discussed due to the difficulty interpreting higher-order trends. A linear trend characterized 98% of the variance for the talking head scene, suggesting that frame rate can be detected for the entire range of values tested in this study.

Following the trend analysis, the log of the functions were plotted to determine the fit of Steven's Power Function. The line length estimates were plotted in this fashion, with $r^2 = .998$, demonstrating that subjects understood the magnitude estimation task. The video evaluations produced $r^2$ values between .91 and .99, demonstrating adherence to the Steven's Power Function. This is significant in that some researchers (Gescheider, 1985) question the ability of metathetic stimuli to conform to this psychophysical law. Recall that stimuli which tap the additive neural processes and address questions of "what kind" and "where" are prothetic. Metathetic stimuli are substitutive, and address questions of quality and location. Because the question of interest in this study was metathetic (video quality), it is interesting that it did indeed conform to Steven's Power Function. However, one might argue that while the question was of a metathetic class, the actual stimulus was additive in nature: one frame rate is perceived by the same neural system as a higher frame rate. More research may be required to resolve the differences between these two classes of stimuli.

For each scene, Steven's Power Function was calculated. The exponent of the function is derived from the coefficient of the log-log regression equation. Coefficients less than one are negatively accelerated, while those greater than one are positively accelerated. A coefficient of one, characterizes a linear relationship between the stimulus and the psychological response. In this study, the panning scenes had coefficients between .367 and .402. These values suggest highly similar functions which are negatively accelerated. This result supports the trend analysis finding that a significant linear and quadratic component characterize these functions. The talking head scene produced a coefficient of .733 - closer to 1 - supporting the trend analysis finding of a significant linear trend.

For each scene, the non-log plot was fitted with a regression curve with no y-intercept (see figures). Due to the trend analysis findings, the panning scenes were fitted with a second order polynomial and the talking head scene was fitted with a simple regression line. The $r^2$ of these regressions ranged from .976-.983 for the panning scenes and .998 for the talking head scene. This result again reinforces the conclusions drawn from the trend analysis.

Finally, an analysis of variance and a priori contrast analysis were conducted on these data. By standardizing the data and demonstrating Steven's Power Function fit (a ratio scale), parametric statistics are an appropriate analysis tool. The within-subject ANOVA resulted in highly significant effects of frame rate, scene, and the interaction between the two. The slow panning scenes were rated higher than the fast panning scenes and the talking head scene was rated lower than the panning scenes. The effect of frame rate shows that frame rate significantly impacted subjects' evaluations for the videos. Again, the trend analyses support this finding, and suggest the 10 fps asymptote for the panning scenes. The effect of scene is also supported by the trend analysis in that the panning scenes had different significant trend components than the talking head scene.

Although there exists ample evidence to show that the panning scenes differ from the talking head scene, the ANOVA does not reveal which if any of the panning scenes differ from each other. To answer this question, a priori contrasts were conducted. It was hypothesized that the fast pan scenes would differ from the slow pan scenes and that the low detail scenes would differ from the high detail scenes. The panning speed contrast resulted in a significant difference $F$ (1, 116) = 14.99, p=.0002. The effect of scene detail, however, did not reach significance $F$ (1, 116) = 3.27, p = .073 (a Greenhouse Geyser and Hundt-Feldt correction yielded a p-value of .09). These findings present interesting implications for the use of desktop video conferencing systems at low frame rates. It appears that the faster panned scenes received significantly lower ratings than the slower panned scenes. In other words, faster panning was rated lower overall than slower panning. However, a close look at the plot of these functions (Figure 3.21) reveals little difference between the fast and slow panned scenes at frame rates less than 10 fps. The difference does not appear until the frame rate is increased, which is curious since subjects were unable to detect large differences overall above 10 fps for the panning scenes. For desktop video conferencing, the implication is that systems supporting frame rates higher than 10 fps will be viewed more favorably if scenery is depicted in a slower pan than a faster pan. However, more research needs to address this finding as only two levels of panning speed were used and practical significance is questionable. It is likely that a more complex relationship would result with increased levels of panning speed.

The contrast between low detail scenes and high detail scenes was not statistically significant. Such a result may derive from the narrow range of this variable. As with panning, only two levels of detail were tested, and it is possible that the levels selected did not differ enough themselves to produce an effect. However, under the assumption that these levels were representative of a typical "low detail" scene and a typical "high detail" scene, the lack of a significant difference suggests that video conferencing users need not be concerned with the level of detail depicted by their systems. However, this finding is specific to the video image size and resolution used in this study. A smaller video image window and lower resolution may impact the perception of highly-detailed scenes. More research is required to explore these issues.

The final contrast that was conducted was between the panning scenes and the talking head scenes. While the trend analysis and Steven's Power Function analysis suggest significant differences between these scenes, the contrast validated these findings $F\ (1,116) = 39.83$, $p<.0001$. Again, more useful information can be gleaned from examining the plot of these functions. It appears that below 16 fps, the talking head scene was rated much lower than the panning scenes. Between 16 and 22 frames per second, there is little difference between the ratings of these scenes. At 25 fps, the talking head scene appears to be rated higher than the panning scenes. This relationship suggests that below 16 fps, subjects are less tolerant of the talking head scene. Perhaps the audio - video mismatch had a large impact under these conditions. Once the frame rate increases to 25 and 28 fps, the video quality may approach that of subjects' pre-conceived notions of video frame rate, and therefore is rated higher than the panning scenes for which they had no preconceived notion of quality.

One limitation of this study is that judgments of video image quality can not be related to task performance. While evaluations of image quality have been shown to be task-dependent (Beaton, 1984), no task performance measures were taken in this study to make such claims. Task performance measures will certainly shed some light on this issue, but preferences can also have a significant impact on people's desire to use a system. Regardless of whether a frame rate supports successful completion of a task, a low opinion of the quality may increase over time, forcing users to select a different communication medium. This issue is addressed in the second study.

This study has contributed to the understanding of low-bandwidth desktop video conferencing usage. Several studies have systematically varied the frame rate of a video to determine user preferences. Chuang and Haines (1993) varied video between 1.5 and 30 fps and used the method of limits to determine the lowest frame rate acceptable for monitoring laboratory rats in a proposed space station experiment. They concluded that 4 fps was the lowest frame rate deemed acceptable by trained observers. This finding cannot directly be compared to this study as the question posed by the experimenters was highly task-specific. Chuang and Haines (1993) also varied the speed of the action in the scene. This variable had no effect on subjects' perceptions, in contrast to the present study. This discrepancy may be the result of different levels of speed manipulated or the fact that the space station study tested subject matter experts in a specific task environment.

Apteker, et al. (1994) sought to develop a dynamic model of Quality of Service (QoS) by measuring preferences for various scene types under different frame rate conditions. Using category scaling (7-point Likert-style scales) these researchers found that relatively static scenes were more susceptible to frame rate changes than high velocity scenes. Overall, high velocity scenes were rated higher than low velocity scenes. This outcome is not consistent with the present study's finding that fast panning scenes were rated significantly lower than slow panning scenes. Because the panning speed effect was statistically significant, but still small (mean rating for slow scenes = 36.11; mean rating for fast scenes = 33.74) practical significance is cast in doubt. The present study extended the Apteker, et al. finding by suggesting that higher frame rates produce more significant differences between slow and fast panning than low frame rates. Overall, Apteker, et al. found that each drop in frame rate measured (15, 10, 5 fps) produced a significant drop in subjective evaluation. The present study produced a similar effect, albeit more so for the talking head scene than for the panning scenes over this particular range.

Pappas and Hinds (1995) tested frame rates between 1 and 7.5 fps to determine the lowest acceptable frame rate before subjects selected a 30 fps binary resolution image over the full grayscale. These researchers report that most subjects (exact number not reported) chose to view a video at 30 fps/binary resolution when grayscale images drop below 5 fps. However, 25% of the subjects opted to view the image at full grayscale for all levels of frame rate. Because their range was restricted at 1 - 7.5 fps, they were unable to detect the differences found in the present study at higher frame rates. For example, this research demonstrates that the increase from 4 to 7 fps is approximately equal to the jump from 7 to 10 fps.

These three studies recommend that digital video conferencing require a minimum of 4-6 fps. By testing a wider range over more levels of frame rate, the present study has demonstrated that significant jumps in image quality improvement are made up to 10 fps for the panning scenes, and up to 28 fps for the talking head scene - a more conventional use for desktop video conferencing. Counter to the findings from other research, this study demonstrates the need for higher bandwidth networks to support desktop video conferencing. For non-traditional conferencing applications, such as real estate walkthroughs and product demonstrations, bandwidth required for frame rates greater than 10 fps is probably not needed. However, for traditional conferencing applications involving the depiction of the face, head, and shoulders, benefits may be detected higher than 28 fps. More research is needed to determine whether high frame rates are necessary to support the transmission of facial cues, gaze, and gesture, all of which are needed for efficient and effective conversation. To summarize the findings, the following conclusions can be made from the first study:

- Four panning scenes were characterized by significant linear and quadratic trends (bend at 10 fps).
- Talking head scene was characterized by a significant linear trend only.
- All scenes were characterized by Steven's Power Curve.
- There was a significant effect of scene content: panning scenes were rated higher than talking head scenes.
- There was a significant effect of panning: slow scenes were rated higher than fast scenes.
- There was no significant effect of detail: high detail scenes vs. low detail scenes.

The final purpose for developing a sound psychophysical function of frame rate was to set levels appropriate for the second study. Because 10 fps represented a bend in the curve for the panning scenes, it was chosen as one level. The other two levels were based on the lower limit of the system (1 fps) that is typical of internet-based desktop video conferencing today, and a stable upper limit of the system (25 fps). The following sections will describe this study.

# Chapter 4.  Ethnography

## Purpose

A review of the literature on ethnographic research methods has shown that little guidance is available regarding the proper application of the various methods. It has also been shown that few relatively long-term studies on the use of desktop video conferencing have been conducted. To help fill this gap in the human factors literature, this study seeks to address several objectives.

- Evaluate the use of desktop video conferencing by real work groups under varying frame rate conditions over an extended time frame.

- Employ and evaluate several ethnographic research approaches.

It is intended that this study will provide useful information for those considering the use of desktop video conferencing as a primary communication medium. This study also seeks to inform human factors researchers and designers regarding the selection of ethnographic research approaches to study complex multi-user communication technologies.

## Method

The second study employed a collection of ethnographic analyses, including elements of interaction analysis, contextual inquiry, questionnaires, and ethnomethodology. Extensive use of interviews, videotaped sessions, surveys, and real-time observation were used to collect data. The intention of this study was to gain insight into desktop video conferencing use not available from traditional human factors studies within the cognitive psychology framework, and to evaluate these methods from the standpoint of efficiency, scientific merit, and system design utility.

Although this study's objective is to overcome the limitations of traditional laboratory research, it was conducted in a quasi-experimental fashion in a laboratory setting. This section will describe the subjects, equipment, and procedure used to collect data.

### Subjects

The participants consisted entirely of students enrolled in ISE 5634: Training System Design (Fall, 1996). During the first class meeting at the beginning of the semester, the study was presented and participation was solicited. Care was taken not to coerce the students into participating; compensation consisted of a t-shirt and pizza lunch. The class was organized around a large project with multiple deliverables. Students were to form groups of approximately four, to meet each week on the project.

### Materials and Equipment

This study utilized two Sun SPARCStation 5 UNIX workstations equipped with a video board, color CCD video camera, and the video conferencing software, ShowMe™, as depicted in Figure 4.1 and Figure 4.2. The same color pallet (Sun platform, 200 colors), addressability (320 x 240), and compression scheme (Sun's Cell B) were all used as in the first study. As in the first study, the same considerations regarding compression artifacts apply.

Audio communication was supported by the conferencing system. Digital audio transmitted at full-duplex was used. Full-duplex audio is a format which permits speakers at both ends of a conference to share the channel simultaneously, as in a traditional telephone call. Half-duplex, incidentally, allows only one party to use the line at one time. The full-duplex, digital audio system used in this study, however, suffered from quality problems. Because both parties shared the line at the same time, the audio from the speaker could be channeled back through the microphone, resulting in echo. Additionally, the digitized audio required a brief period of time to be encoded, transmitted, and subsequently un-encoded, resulting in lags longer than those of a modern telephone network. Care was taken to minimize the echo by adjusting the microphone sensitivity and speaker volume and placing the two as far apart as possible.

Each workstation was placed in a separate room in the HCI Lab (room 530 Whittemore) and networked on an isolated 10BaseT ethernet local area network through an Asante 8 port ethernet hub, as shown in Figures 4.3 and 4.4. The rooms were located far enough apart such that no unwanted sound could travel between them.

FIGURE 4.1 Video Conferencing System


FIGURE 4.2 Video Conferencing Software User Interface

In addition to the computer network, a video observation network was built so that the two rooms could be monitored. A separate video camera was placed in the corner of each room such that the best view of the work group and computer setup could be obtained. These two video signals were spliced and displayed on a video cube monitor in the hallway of the lab. The composite video signal and one audio signal were recorded on a standard VHS VCR. The single auditory signal picked up the sound from both rooms (one direct, the other through the video conferencing system). In addition to the computer network and video network, other materials included interview forms, interaction analysis forms, and questionnaires.

## Procedure

The class members formed four teams during the first week of classes. Each team consisted of four participants (one team had five). All class members agreed to participate in the study. Each team met for approximately forty minutes each week in the lab to work on anything related to their design project. The actual time spent in the lab varied from week to week and from team to team, but averaged 37 minutes. The first week, all four teams met face-to-face in one of the lab's rooms. Each team was then chosen at random to fill one of four experimental conditions: face-to-face, 1 fps, 10 fps, and 25 fps. All subsequent meetings were based on the condition for which the team was assigned.

At the start of the second session, those teams using the video conferencing systems were instructed briefly on how to use the system and its capabilities. Because the participants all had considerable experience with graphical user interfaces and computer communication systems, minimal instruction was required. The participants were told that they could use the software in any way they felt was appropriate, with the exception that the video quality parameters could not be adjusted. Specifically, the electronic whiteboard was made available to the users and the sizing and location of the two video windows could be adjusted. The addressability of the video was kept constant at 320 x 240, the highest quality supported by the system at 25 fps.

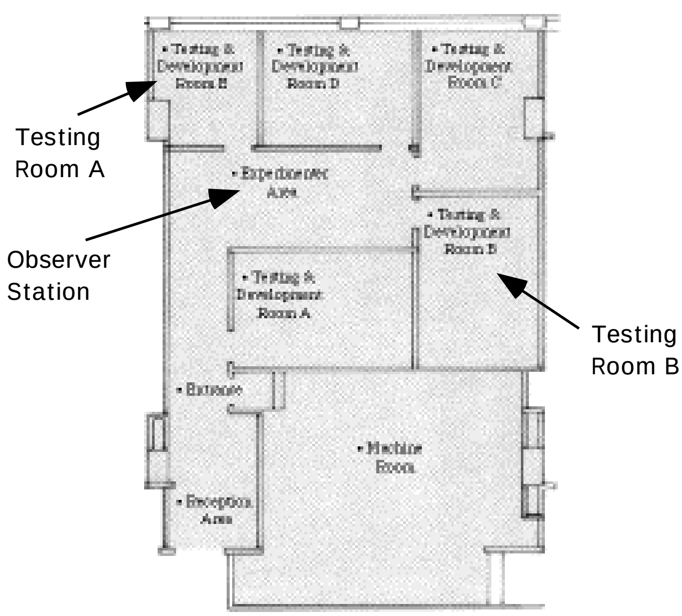



FIGURE 4.3 Floor plan of the HCI lab indicating the rooms in which the study was conducted.

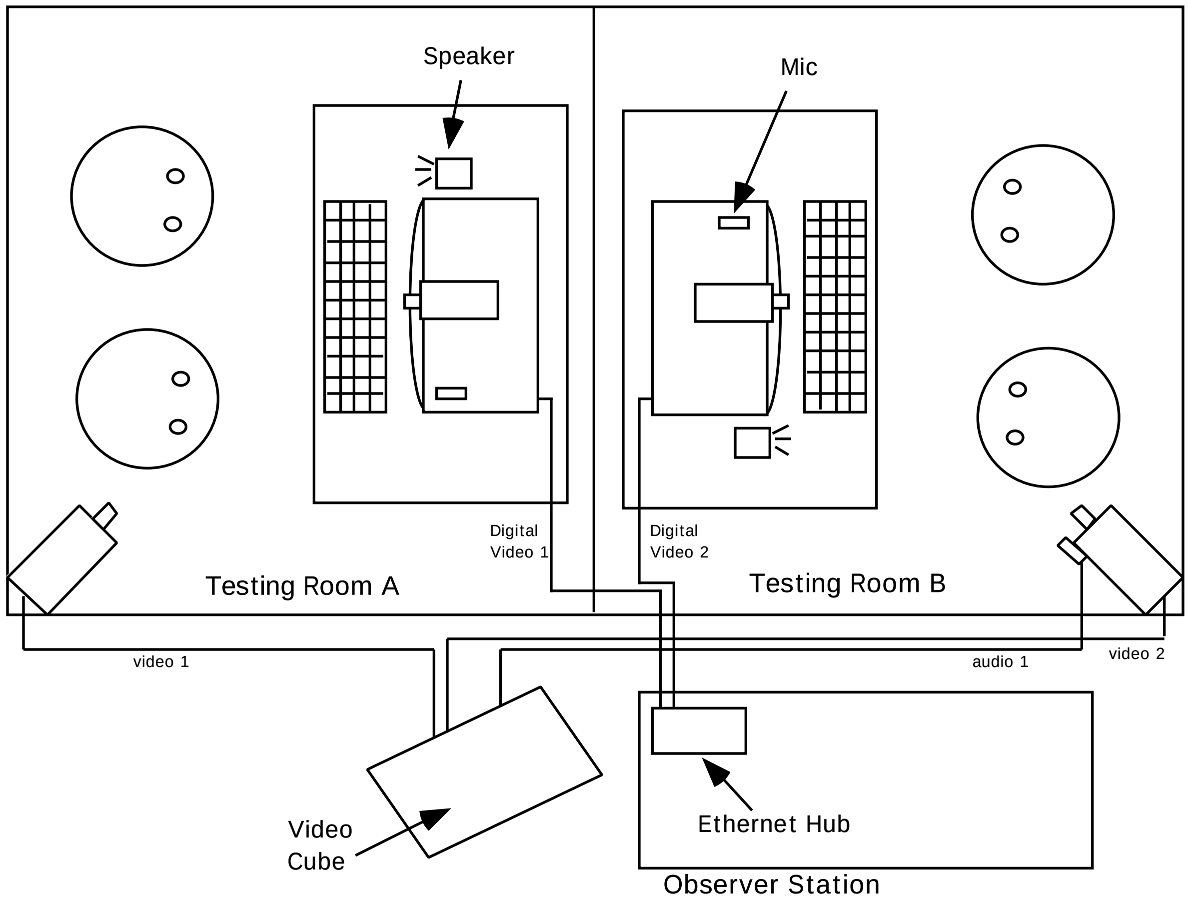



FIGURE 4.4 Physical layout of the experimental setup for phase II. In reality, the two testing rooms were not adjacent to prevent unwanted sound sources.


FIGURE 4.5 Recorded video signal for a face-to-face meeting and a video-mediated meeting.

Each session was videotaped with the split screen composite (except the face-to-face conditions and group which was videotaped with the entire screen). During the session, the author monitored the meetings and took detailed notes on the discussion, any communication breakdowns that seemed critical, and the general activity structure of the session. Each group ended the meeting when they felt the need to do so. Following the meeting, each team member was asked to complete a comprehensive questionnaire and then to answer collectively a series of interview questions during which notes were taken by the experimenter.

## Results

The results of this study will be presented in subsections for each major class of analysis: contextual inquiry, questionnaire, interaction analysis, and ethnomethodology.

### Contextual Inquiry

At the end of each session, participants answered survey questions in an open-ended style interview. Participants were asked several questions aloud, and were encouraged to verbalize any feedback they had regarding the use of the video conferencing system. After several sessions, the face-to-face group ceased participating in this activity as most of the questions concerned the use of the video conferencing system. While the survey had consistent questions for each week (see Appendix C), other issues were often discussed at the time, and responses did not necessarily fall under a particular item of the survey. The purpose of this method is to view the video conferencing experience from the perspective of the participants - to understand what problems they encountered and to provide a comfortable setting by which participants could reflect upon how the session went.

Reporting qualitative data is difficult, especially since participant responses did not necessarily correspond to the survey items. For this reason, these data will be presented as general comments by each group for each week in Tables 4.1-4.4. The interview items were collapsed into the two with the most responses: problems communicating and changes since the previous week.

### Questionnaire

Questionnaires were administered to all participants in the video conferencing conditions at the conclusion of each session (not including the initial face-to-face session). The data from these questionnaires were not analyzed statistically due to small sample sizes, missing data cells (three participants each missed a session), and fact the that one group dropped out from the study after five weeks. Despite these problems, plots of the data for each question over the six weeks of video conferencing usage are useful in that trends and differences can be noted, albeit without the clout of a statistical test for significant difference. Each question is displayed in a separate graph, depicting the group responses over the course of the study. Figures 4.6-4.11 display these graphs. Figures are divided into the four major categories of items: listener, speaker, technology, and environment questions.

TABLE 4.1  Interview Responses for Face-to-Face Group Compiled by Week

| | *Questionnaire Items* | |
|---|---|---|
| *Week* | *Problems communicating with each other* | *Changes in behavior since last week* |
| 1 | -- | -- |
| 2 | -- | More efficient when being watched |
| 3 | -- | Ignored the camera more than previous week, less self-conscious; more productive than last week |
| 4 | -- | Much less productive since last week; no clear direction on project; probably would not have met today if not for the study |
| 5 | -- | Much more productive this week |
| 6 | -- | Very productive this week Everyone did a lot of work over the week |
| 7 | -- | -- |

TABLE 4.2 Interview Responses for 1 fps Group Compiled by Week

| | *Questionnaire Items* | |
|---|---|---|
| *Week* | *Problems communicating with each other* | *Changes in behavior since last week* |
| 1 | -- | -- |
| 2 | Could not see one member (the third person in a group); video added nothing to communication process; video was more amusing and source of distraction than a help; whiteboard has potential; echo was a little annoying; low frame rate was a problem; whiteboard was distracting | |
| 3 | Takes too long to read off what is being written; Video was not used - did not help; communication was most like talking on a phone; one person was not visible (the third person in one group); if frame rate was faster, they would use video more; it is frustrating knowing that they are only 20 feet away and have to use this system; add a text window for collaboratively writing documents; faster frame rate; wider field of view | More productive b/c they had an assignment due and did not get distracted with the whiteboard; they are getting used to the system |
| 4 | Session went much easier with the smaller group; impossible to make eye contact; locate the camera behind the monitor somehow and look into the camera lens | Stopped looking at screen - just occasionally glanced up if something happened |
| 5 | Better visual b/c all fit into the field of view of the camera; easier to determine who was talking; one on one would be much easier to communicate than 2 on 2 or 3 | Shorter session; group has given up on getting lots of work done - they simply try to get more done during the week when they can meet face-to-face; Did not look at video much |
| Final Interview | It was torture; It was not productive; I dreaded it every week; The audio was ok, but the video was useless; Could not read the facial expressions; Very easy to lose your train of thought - for example when you are talking, you pause to see if they are listening, then you hear your own echo and get lost; Having 5 in the group was very difficult - the times when 3 or 4 group members were present were much more productive than when the whole group was there; Knowing that they could meet face to face right after the meeting was frustrating; Referring to shared documents was tough -instead of just handing someone a book turned to a certain page; A shared text editor would have made it *much* easier and better. - they would not have had to read off notes to each other this way; Group grew tired of repeating themselves; They could never see all three participants on that side of the conference | |

TABLE 4.3 Interview Responses for 25 fps Group Compiled by Week

| | *Questionnaire Items* | |
|---|---|---|
| Week | *Problems communicating with each other* | *Changes in behavior since last week* |
| 1 | -- | -- |
| 2 | Tough to get used to talking to a computer monitor; Some gestures are missed - not in FOV; Tough to speak loud to the mic; Group claimed that they seemed to form their own verbal turntaking protocol; Impossible to make eye contact; Felt compelled to speak in long statements - they wanted to use up their time on the floor one it had been gained; To make eye contact, group members would lean in and look at monitor, thereby invading the personal space of the co-located person; Often, they would speak louder than normal to make the remote group hear; Problems could be remedied: Develop own protocol more; Eliminate self view video window | -- |
| 3 | Unsure if side conversations were important enough to repeat to rest of group; Control of conversation is switched from listener to speaker; Difficult with humor: Tough to mumble a joke under your breath without interrupting the speaker; Some communication cues are lost - its tough to judge how your remark was received; Silence is awkward | Grew more comfortable speaking to computer; Feedback was not as bad; Volume could have been louder; One guy was not purposely speaking louder; Overall comfort level increased; When talking louder so remote group can hear, it inadvertently emphasizes the statement - gives it more weight |
| 4 | Sound/system crashed at one point; Moving windows on one side caused the other side to temporarily freeze for some reason - the moment the video stopped, they stopped listening | Echo was worse; Lots of laughter; Feedback was tough for the speaker - very distracting: "Screws up your ability to interrupt"; Clearing the throat can cause accidental interruptions (perhaps due to lack of visual cues?) |
| 5 | Tough to work when not having structure; Video seemed grainier | Toured a video conferencing room the day before which may have biased opinions or expectations of quality; Not as much feedback this time |
| 6 | -- | More motivated this week; More active this week |
| 7 | Fewer problems than before; Echo in the auditory - 2 group members got used to it over time (1 reported that they did not) | Video image flashed several times - more than normal; Guy who was alone was much more comfortable by himself; Easier talking to one guy as opposed to two (i.e. no side conversations to deal with); Group definitely got used to the system over the past six weeks - in general they got more comfortable with it; They really relied on this session to be their regular meeting each week; The group definitely used the video channel each week - some, however, claimed they did not look at the video window too much |

TABLE 4.4  Interview Responses for 10 fps Group Compiled by Week

| | *Questionnaire Items* | |
|---|---|---|
| *Week* | *Problems communicating with each other* | *Changes in Behavior since last week* |
| 1 | -- | -- |
| 2 | Could not see both remote participants; Echo problem in audio; Lag time was difficult to deal with; Feels like you are talking into space; Problems could be remedied:  Eliminate the echo and lag; Camera angles posed a minor problem | All the differences/effects associated with the video system |
| 3 | Feedback problems; Video angles; One group had much more feedback than the other - probably due to the fact that one side turned up volume considerably; Problems remedied:  More camera angles, zoom feature | -- |
| 4 | Tough to interrupt another speaker - like dealing with 2-way radio; Group members were not always sure if they were heard; Tough to read the body language - delays were tricky; Feedback was bothersome | Sat further back from camera to get everyone in the view |
| 5 | Very few problems communicating; Whiteboard was lost on one side; Lots of talking at the same time; Group went out of the way to not meet during the week | Auditory quality was improved; Echo was not as bothersome on one end, but annoying on the other; Liked larger video window that they used |
| 6 | Not good seating for the people in the same room - the don't face each other | Co-located group snuggled up so that they would both appear in the video window |
| 7 | No major problems; Still tough to interrupt - audio seems to only be going one way | Lag time seemed longer; Seemed like echo was worse; More non-productive use of whiteboard (group drew an elaborate picture); Overall, the group said that they got used to the system over the 6 weeks of use |

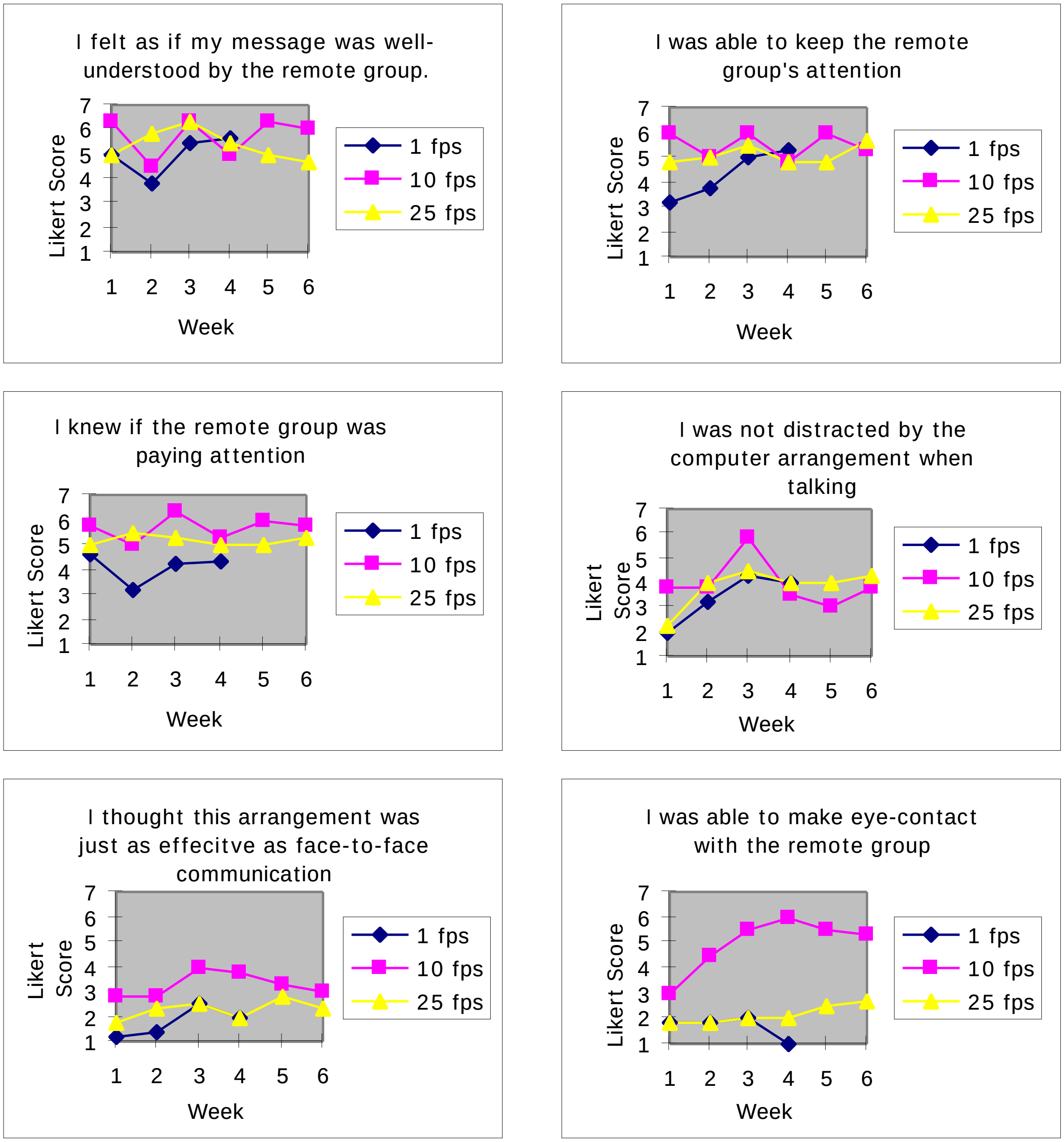



FIGURE 4.6  Speaker Questions

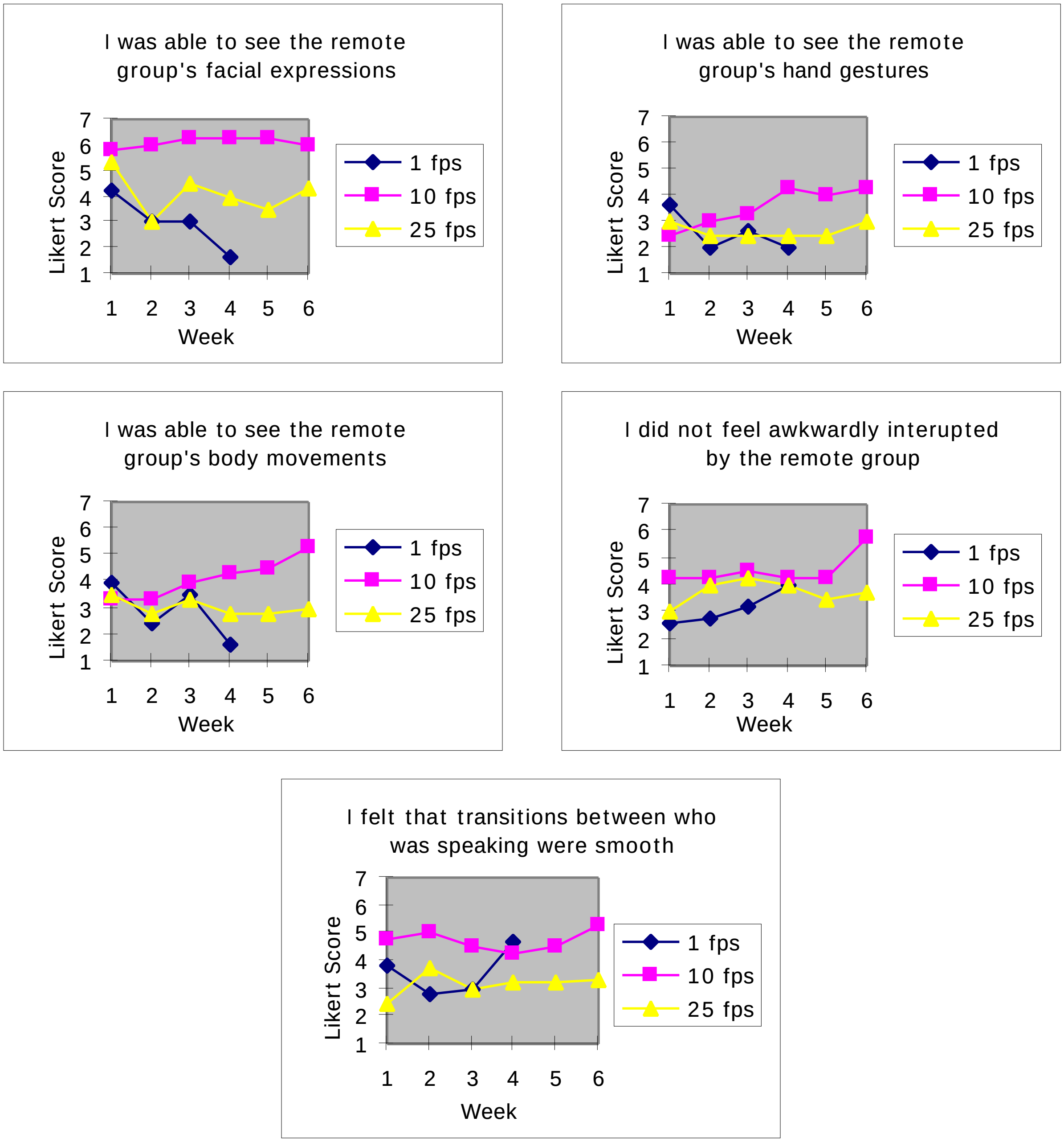



FIGURE 4.7 Speaker Questions (Continued)

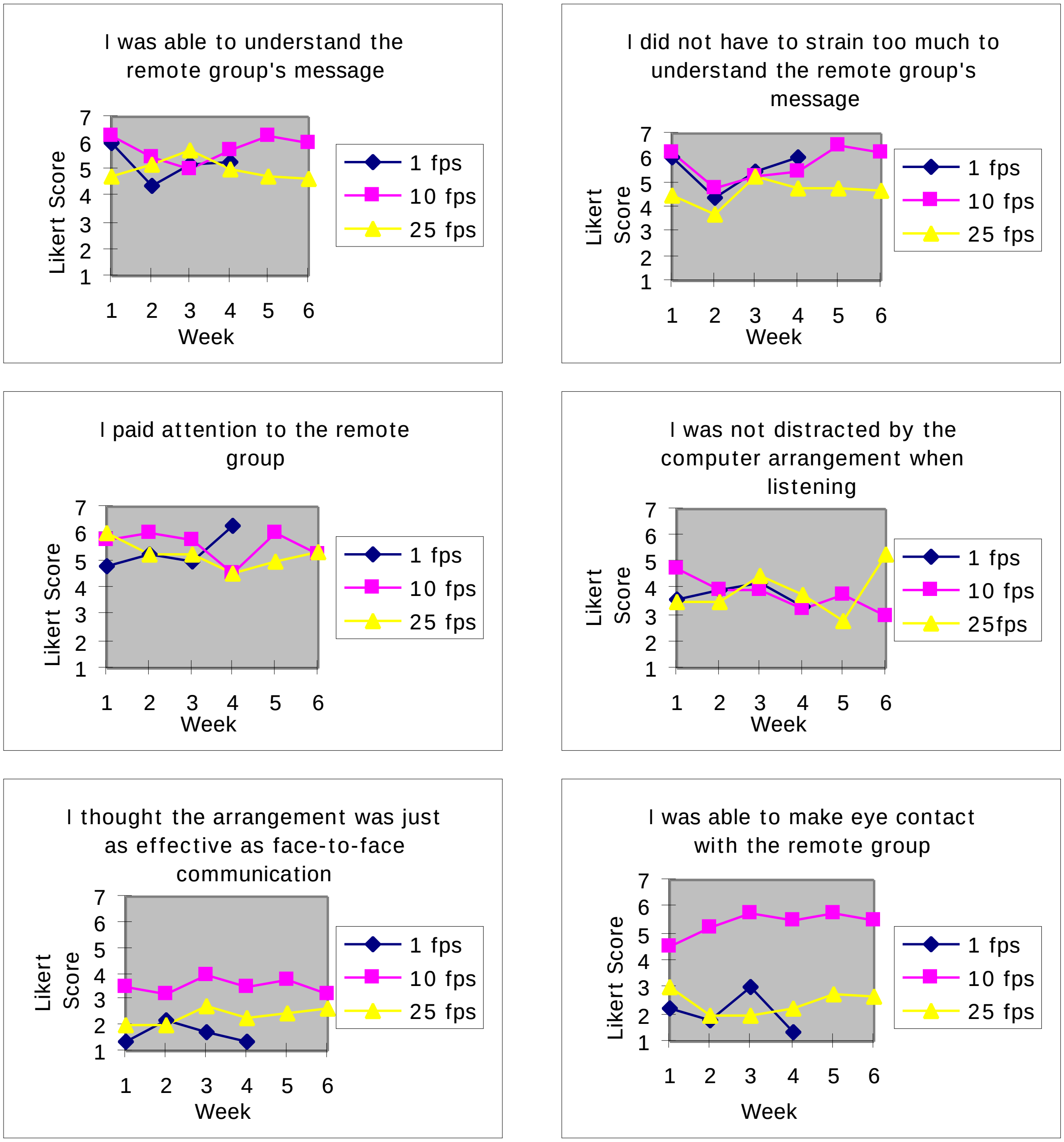



FIGURE 4.8 Listener Questions

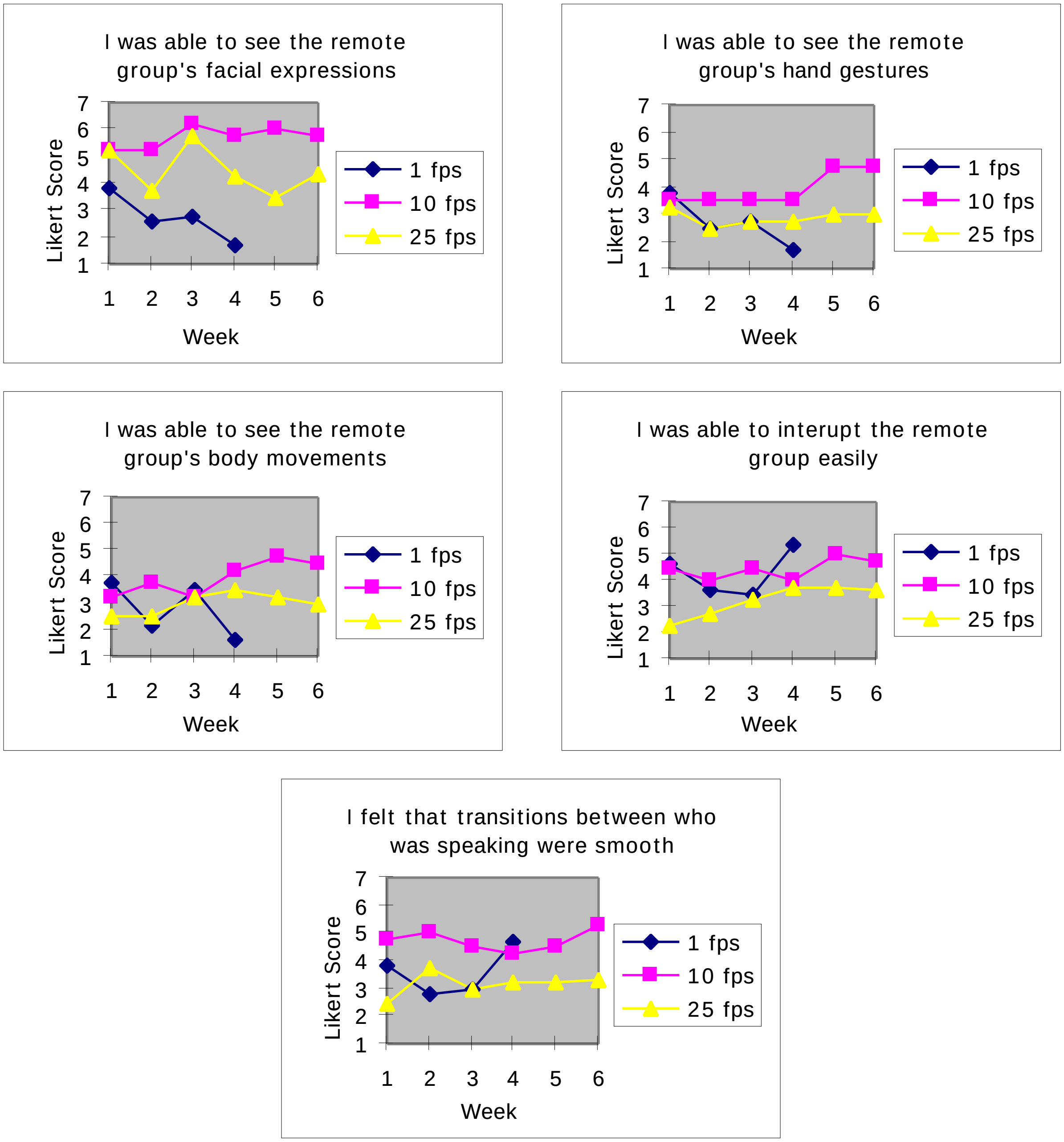



FIGURE 4.9 Listener Questions (Continued)

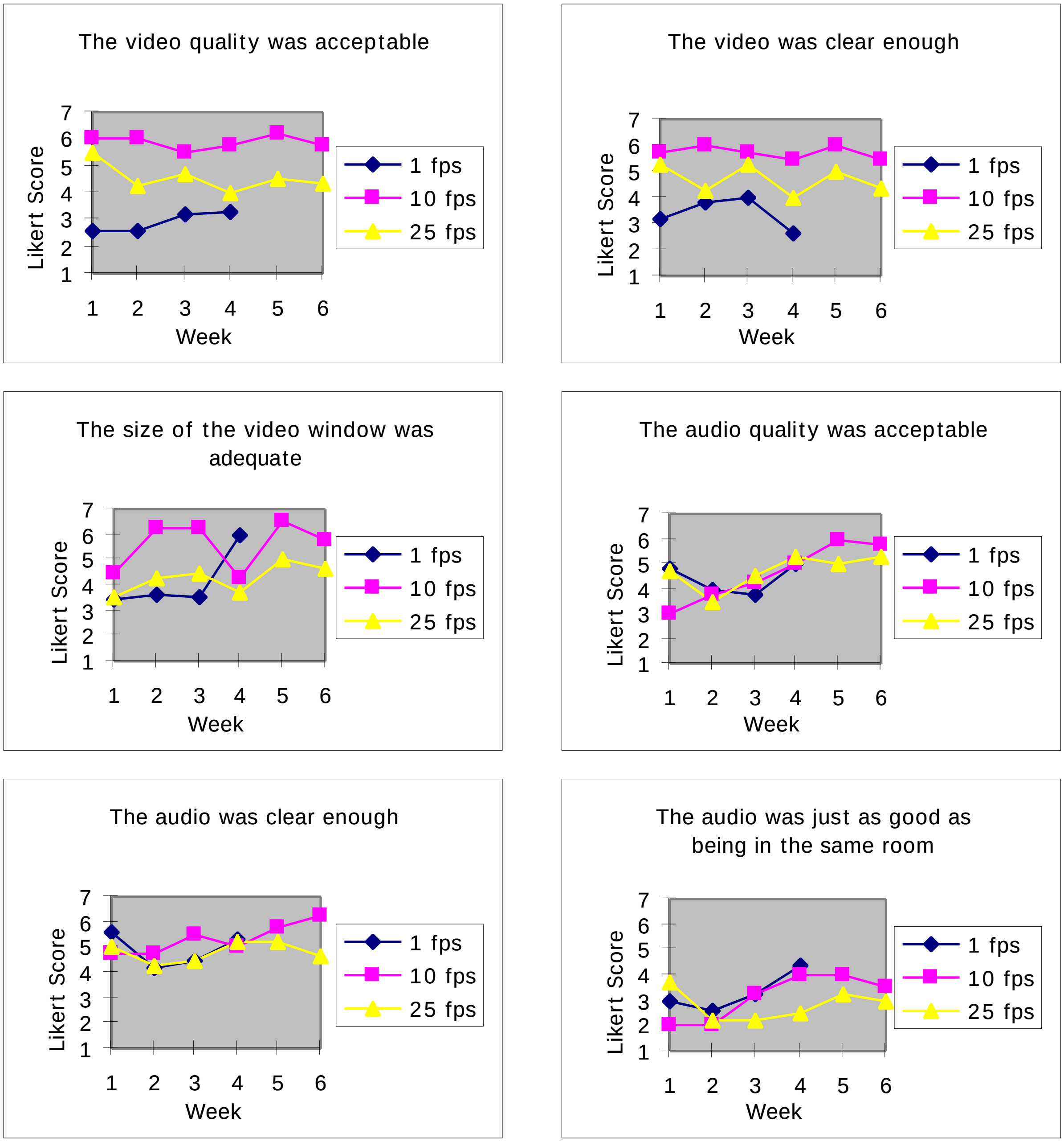



FIGURE 4.10 Technology Questions

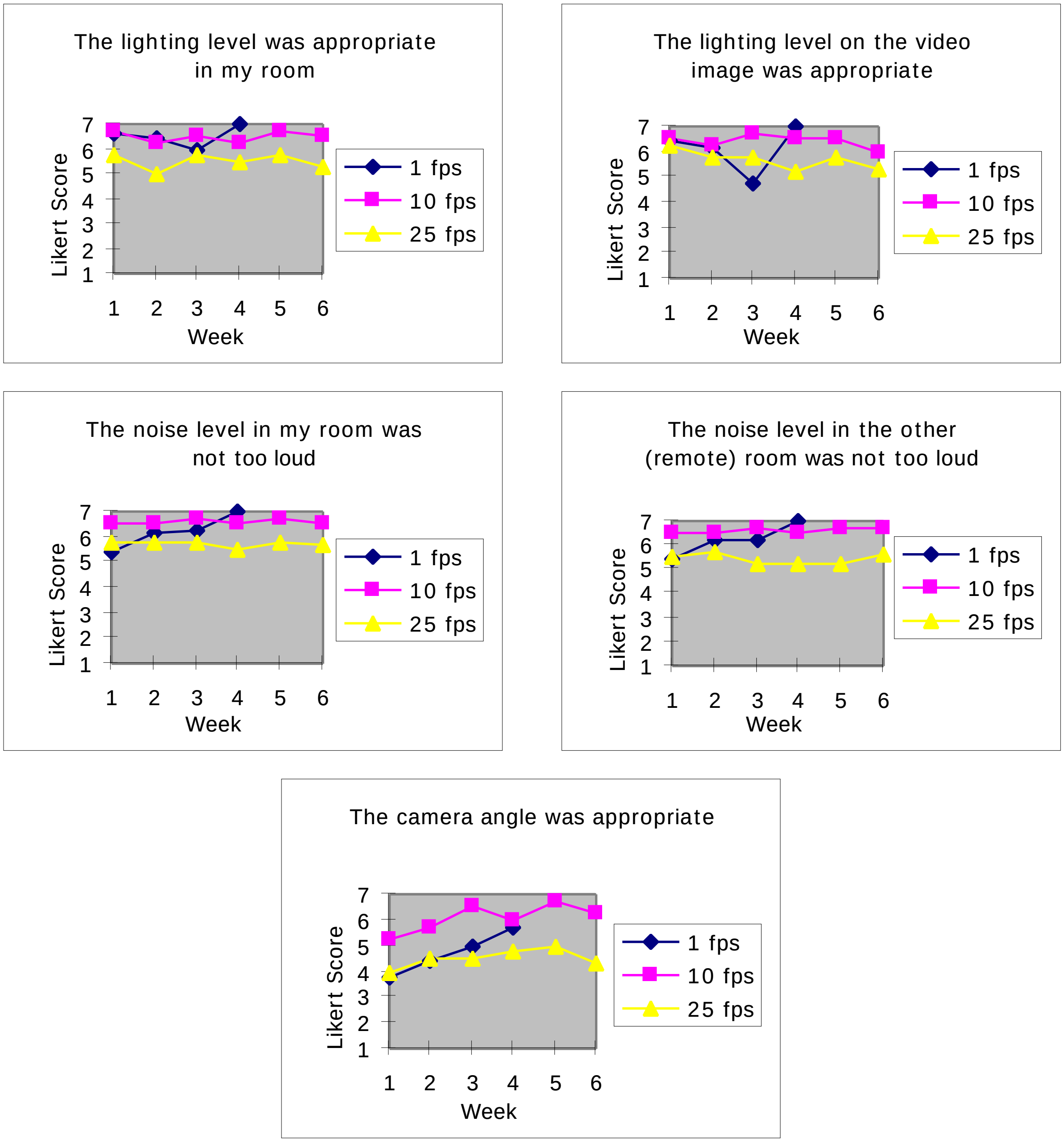



FIGURE 4.11 Environmental Questions

## Interaction Analysis

The ethnographic method used is based heavily on the research of Jordan and Henderson (1995). Jordan and Henderson (1995) present an overview of their method based on many years of conducting extensive ethnographic studies on the relationship between people and computers systems. Their method was selected due to its thoroughness, extensibility, and flexibility. Interaction Analysis provides a framework within which certain aspects of a work environment are focused upon. Table 4.5 describes these "foci of analysis". This section presents the results from two foci: turntaking and trouble and repair, as these were the most applicable aspects of the video conferencing sessions. the other foci are addressed in the ethnomethodology section.

A subset of interaction analysis, known as conversation analysis, has been developed to quantify speech patterns in conversational systems. Because speech is an integral aspect of communication within this study, two conversation analysis methods were employed. Breakdown analysis, based on the work of Dorrey (1995) was used to measure the degree of communication efficacy engendered in each conferencing situation. The intent of this analysis is to quantify and characterize the type of communication breakdown that occurs. Turntaking analysis, on the other hand, seeks to characterize the flow of conversation. Particular attention is paid to understanding who is most likely to speak following a given speaker. This question is important in understanding video conferencing usage difficulties and side conversations.

TABLE 4.5. Foci of Analysis (based on Jordan and Henderson (1995)

| *Focus of Analysis* | *Details* | *Measurement Method* |
|---|---|---|
| Event Structure | Timing of beginning and ending of events, description of transitions between events, periodicity of events, sequencing | -- |
| Turn Taking | Verbal transitions, Artifact transitions (who controls the mouse), Non-verbal transitions (gestures, gaze) | Turn taking Analysis |
| Participation Structures | Identify positioning of participants with respect to co-located group and remote group. Also examine positioning with respect to artifacts (distance from display, use of artifacts) | -- |
| Trouble and Repair | Identify communication breakdowns and means by which they are repaired. | Breakdown Analysis |
| Artifacts and Documents | Inventory of artifacts: using, controlling, sharing. | -- |

### *Breakdown Analysis*

All sessions were analyzed to determine the frequency of communicative breakdown. This was accomplished by having two undergraduate research assistants watch video tapes of the sessions and record incidents of communicative breakdown. The data sheet displayed in Appendix D was used by the research assistants to record six classes of breakdown: verbal turntaking, reference, topic, audio, video, and shared conversation. Each was operationally defined for the research assistants according to the descriptions listed in Table 4.6.

TABLE 4.6 Communicative Breakdown Operational Definitions

| *Breakdown Type and Subtype* | | *Definition* |
|---|---|---|
| Verbal Turntaking | | An incident in which two or more participants are talking at the same time. This does not include specific interruptions (i.e. "...Excuse me, but...") |
| A | One Stopped, one continued | A turntaking breakdown which was resolved by one participant ceasing to speak while the other continues |
| B | Request for other to stop | A turntaking breakdown in which a participant asked another to stop talking |
| C | Two continue longer than 3 seconds | A turntaking breakdown in which both participants talking simultaneously continue for longer than 3 seconds |
| D | Other | Any other example of a turntaking breakdown not described above |
| Reference | | An incident in which the sharing of an object (papers, pencil, notes, mouse, etc.) becomes problematic (i.e. two users "battle" for use of the mouse) |
| Topic | | An incident in which there is a failure to maintain topical orientation by one or more of the participants |
| A | Request to be updated, result of side conversation | A topic breakdown in which one or more participants request to be updated. The breakdown is the result of a side conversation |
| B | Request to be updated, result of other activity | A topic breakdown in which one or more participants request to be updated. The result of an activity other than a side conversation |
| C | Other | An other topic breakdown not described above |
| Visual | | An incident in which visual communication suffers |
| A | Request for shift in seating | A visual breakdown evidenced by one participant requesting another to shift seating |
| B | Request for camera movement on participant | A visual breakdown evidenced by a request for re-positioning the camera on a participant |
| C | Unable to see a participant | A visual breakdown in which a participant is unable to view another participant |
| D | Request for an object to be re-positioned | A visual breakdown in which a participant requests that an object of interest be re-positioned in the camera field-of-view |
| E | Request for camera movement on an object | A visual breakdown evidenced by a request for re-positioning the camera on an object |
| F | Other | Any other visual breakdown not described above |

| | | |
|---|---|---|
| Audio | | An incident in which a breakdown results due to loudness, feedback, or static difficulties. Requests for repeating due to these problems |
| A | Request for repeat, other repeats | An audio breakdown in which one participant asks to repeat a statement and another participant honors the request |
| B | Request for repeat, other does not repeat | An audio breakdown in which one participant asks to repeat a statement and no other participant honors the request |
| Shared Conversation | | An incident in which a statement or question is directed at an adjacent (or co-located) participant with no attempt to include the rest of the group |
| A | One side conversation | A shared conversation breakdown in which only one side conversation occurs |
| B | Two side conversations | A shared conversation breakdown in which two side conversations occur |

The six communication breakdown types were based on Doerry (1995) and inferences regarding what communication problems would occur in desktop video conferencing systems (i.e. audio-related problems or visual-related problems). The sub-types were based on pilot testing with a sample video clip. Table 4.7 displays the total number of each breakdown which occurred across all sessions, adjusted per hour.

TABLE 4.7 Communicative Breakdown Totals (adjusted per hour) Percent of total breakdown type is provided in parentheses.

| *Breakdown Type* | | *Group* | | | |
|---|---|---|---|---|---|
| | | Face-to-face | 1 fps | 10 fps | 25 fps |
| Verbal Turntaking (Total) | | 305.06 | 59.31 | 134.76 | 146.38 |
| A | One Stopped, one continued | 158.86 (52%) | 41.19 (69%) | 118.70 (88%) | 123.14 (84%) |
| B | Request for other to stop | 3.67 (1%) | 2.62 (4%) | 1.67 (1%) | 0 |
| C | Two continue longer than 3 seconds | 134.80 (44%) | 9.99 (17%) | 12.32 (9%) | 22.04 (15%) |
| D | Other | 8.90 (3%) | 5.41 (9%) | 2.07 (2%) | 0 |
| Reference (Total) | | 0 | 1.67 | 0 | 1.03 |
| Topic (Total) | | 1.25 | 4.44 | 11.67 | 7.90 |
| A | Request to be updated, result of side conversation | 1.25 (100%) | 1.90(43%) | 8.33 (71%) | 1.30 (17%) |
| B | Request to be updated, result of other activity | 0 | 2.53 (57%) | 3.33 (29%) | 5.61 (71%) |

| | | | | | |
|---|---|---|---|---|---|
| C | Other | 0 | 0 | 0 | .98 (12%) |
| Visual (Total) | | 0 | 3.43 | 9.81 | 3.22 |
| A | Request for shift in seating | 0 | 1.76 (51%) | 4.39 (45%) | 0 |
| B | Request for camera movement on participant | 0 | 0 | 0 | 0 |
| C | Unable to see a participant | 0 | 0 | 0 | 0 |
| D | Request for an object to be re-positioned | 0 | 0 | 0 | 0 |
| E | Request for camera movement on an object | 0 | 0 | 0 | 0 |
| F | Other | 0 | 1.67(49%) | 5.42 (55%) | 3.22 (100%) |
| Audio (Total) | | 3 | 6.43 | 29.81 | 19.44 |
| A | Request for repeat, other repeats | 0 | 5.47 (85%) | 26.06 (87%) | 16.69 (86%) |
| B | Request for repeat, other does not repeat | 3 (100%) | .95 (15%) | 3.75 (13%) | 2.75 (14%) |
| Shared Conversation | | 38.48 | 63.41 | 94.35 | 65.64 |
| A | One side conversation | 14.75 (29%) | 45.69 (72%) | 88.89 (94%) | 52.84 (81%) |
| B | Two side conversations | 36.58 (71%) | 17.71 (28%) | 5.45 (6%) | 12.80 (19%) |

The data for this analysis were compiled by having two observers view the video tapes and record when specific incidents of communication breakdown occurred. Each observer was responsible for half the videotapes. To test whether there was positive agreement between the observers, four sessions were observed redundantly. Specifically, each observer watched two video tapes which were originally viewed by the other observer. This set of data was analyzed with a Pearson's product moment correlation coefficient and is displayed in Table 4.8. Bordens and Abbott (1991) demonstrate the use of this statistic in determining inter-rater reliability. As can be seen from Table 4.7, only two classes of communication breakdown were common: verbal turntaking breakdowns and shared conversation breakdowns. There was considerable positive agreement between the two observers on the other breakdown classes, but the small data cells do not warrant correlation analysis.

The data in Table 4.8 show that there was fairly good correlation among the two observers. The first session was not analyzed for shared conversation as only one incident was observed by each observer, resulting in perfect correlation. This analysis demonstrates that reliability was probably adequate. P-values indicate Fisher’s r to z transformation, testing the null hypothesis that the $r^2$ is significantly different from 0.

TABLE 4.8  Reliability Analysis for Breakdown Analysis Data

| | *Verbal Turntaking* | | *Shared Conversation* | |
|---|---|---|---|---|
| *Session* | $r^2$ | *p* | $r^2$ | *p* |
| 1 | .73 | .02 | -- | -- |
| 2 | .76 | .01 | .96 | .0001 |

One objective of this study was to determine if time influenced the manner in which desktop video conferencing was used.  To this end, Figures 4.12-4.23 are provided to show the number of each type of communication breakdown per hour for each group over each week. Verbal turntaking, audio, and shared conversations were found to be the most common forms of communicative breakdown.  As a result, each of these were plotted according to their subtypes. For verbal turntaking breakdown, only two sub types were found to occur frequently; therefore, only these data were plotted.  No statistical analyses were conducted on these data for the same reason the questionnaire data were not analyzed with parametric statistics.

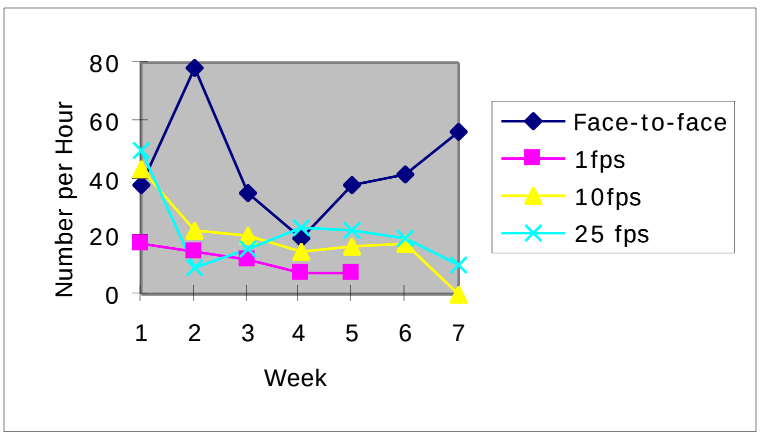



FIGURE 4.12  Verbal Turntaking Breakdown - TOTAL

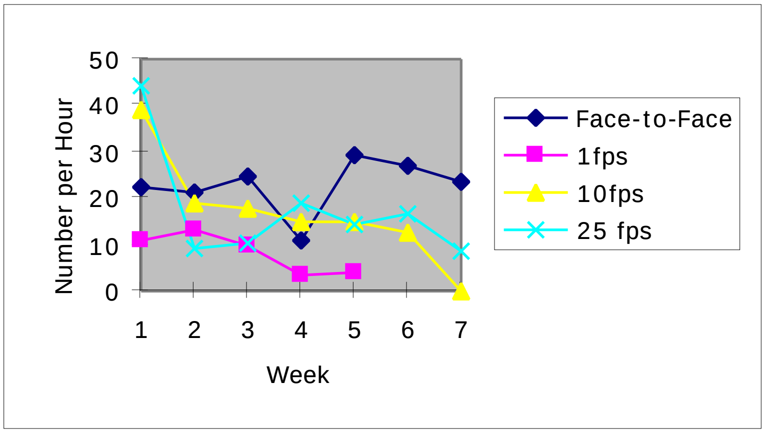



FIGURE 4.13 Verbal Turntaking Breakdown - TYPE A

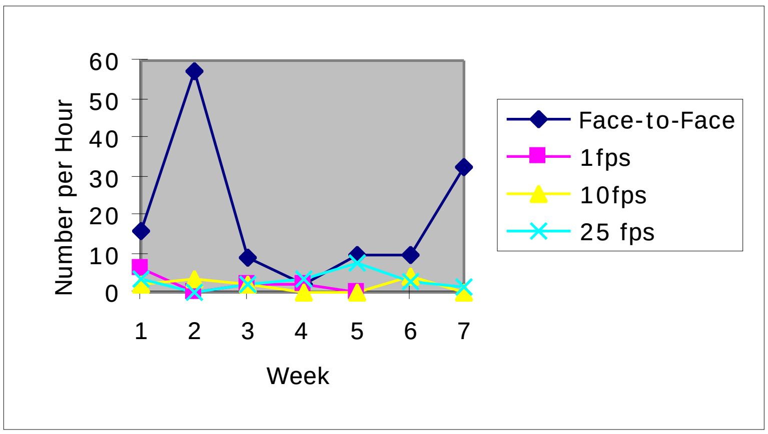



FIGURE 4.14 Verbal Turntaking Breakdown - TYPE B

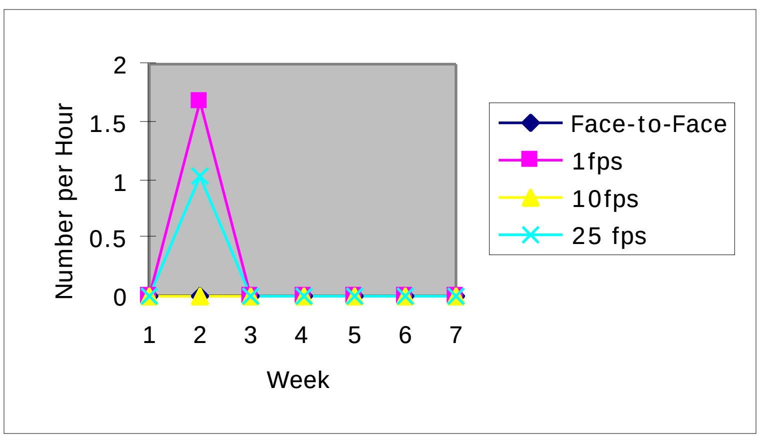



FIGURE 4.15 Reference Breakdowns

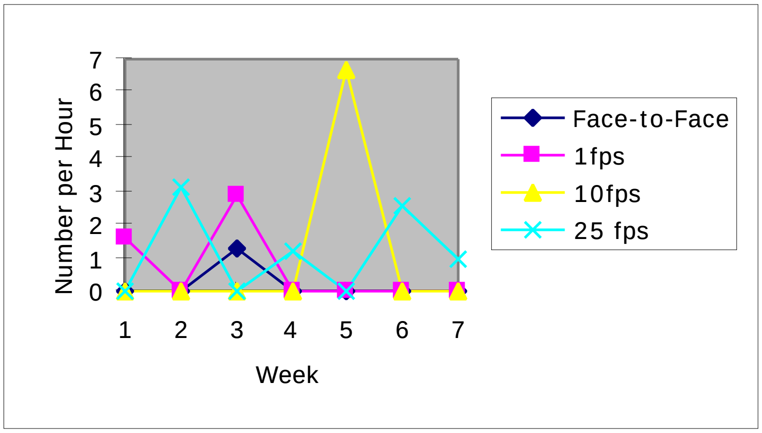



FIGURE 4.16 Topic Breakdowns

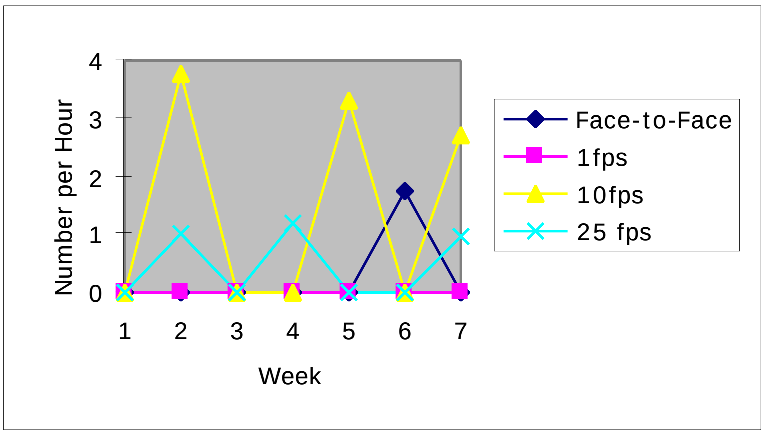



FIGURE 4.17 Visual Breakdowns

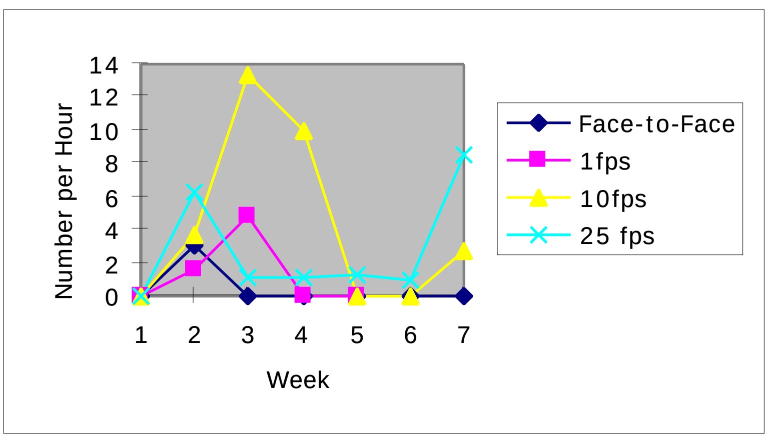



FIGURE 4.18 Audio Breakdowns - TOTAL

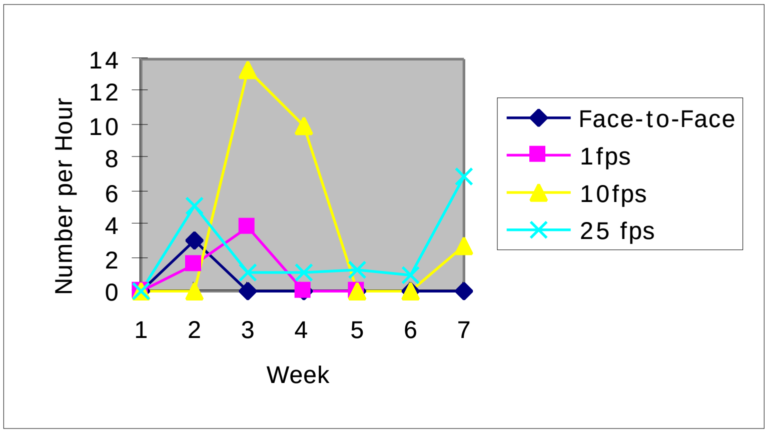



FIGURE 4.19 Audio Breakdowns - TYPE A

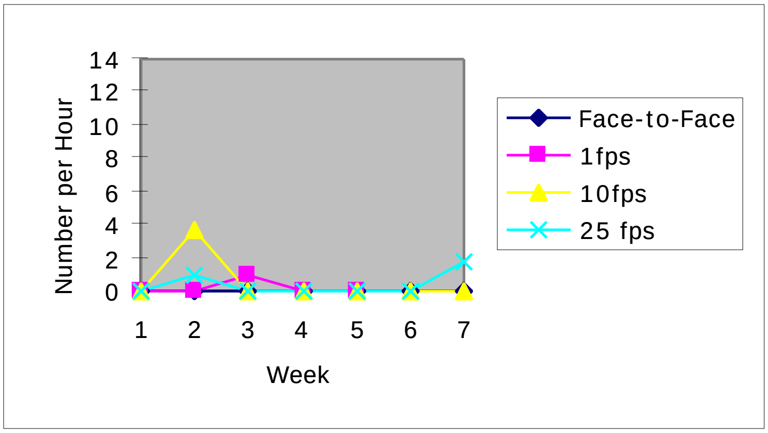



FIGURE 4.20 Audio Breakdowns - TYPE B

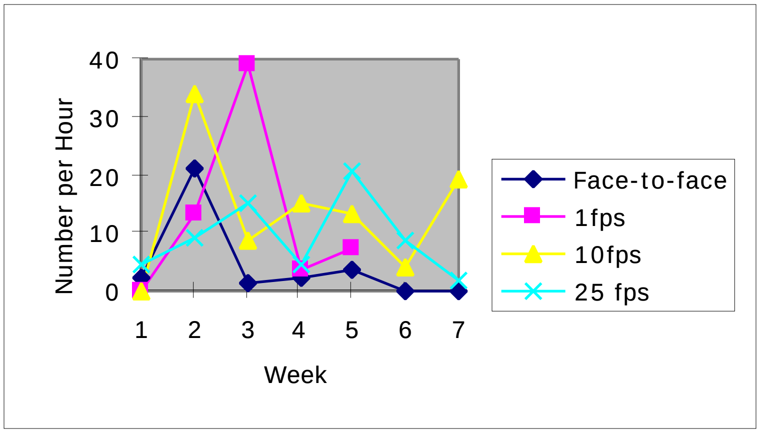



FIGURE 4.21 Shared Conversation Breakdowns - TOTAL

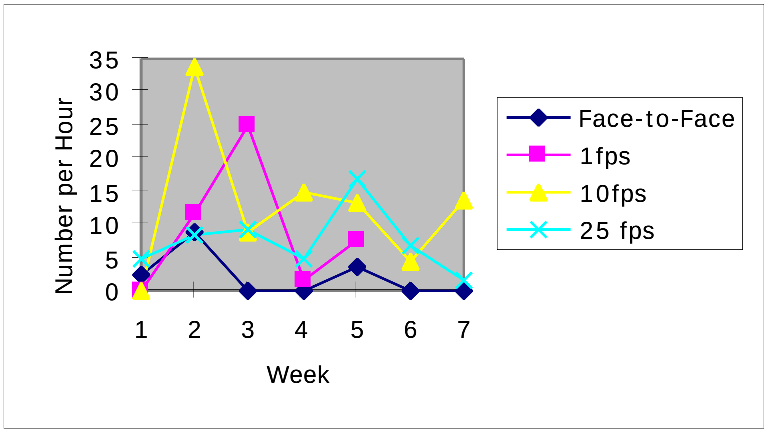



FIGURE 4.22 Shared Conversation Breakdowns - TYPE A

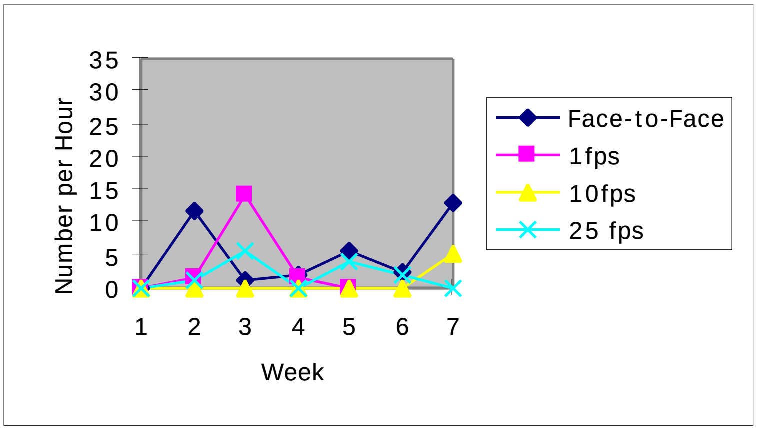



FIGURE 4.23 Shared Conversation Breakdowns - TYPE B

*Turntaking Analysis*

To understand the conversational flow of the sessions, the order and frequency of taking turns speaking was compiled in the following manner. Two undergraduate assistants watched the videotapes and recorded who was speaking by pressing a coded sequence of keys on computer keyboard. For example, a "1" was pressed if subject 1 was speaking, then if subject 2 took the floor, a "2" was pressed. This analysis therefore collected the order and frequency of the speakers; it did not collect the time each speaker held the floor.

Permutations of turntaking were formed by grouping each speaker with the speaker who followed him or her. Consider Figure 4.24. Each box in the diagram represents the two speakers who shared a video conferencing unit. On the left half, subject 1 could be followed by subject 2 or vice versa. The lines joining the two halves of the conference represent when speaker 1 or 2 was followed by speaker 3 or 4 and vice versa. Because the two groups did not represent differences in task allocation and discussion was expected to occur equally among the participants, the model in Figure 4.24 describes expected speaker distributions based on equal probabilities.

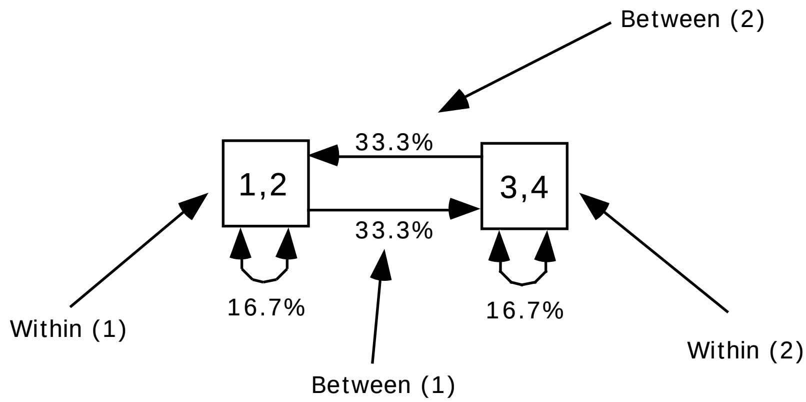



FIGURE 4.24 Turntaking model with expected frequency percentages for a four-person conference.

Twelve permutations of two speakers are possible with four speakers arranged in this manner and the percentages of grouping these conversation patterns are provided in the Figure 4.24, based on the probabilities associated with their permutations. A chi-square analysis was conducted to determine if the probability of a speaker followed by another occurred randomly or not, according to the groupings shown in Figure 4.24. For sessions in which one or more participant was absent, expected percentages were recalculated based on the number of possible speaker turn permutations. For each group's initial, face-to-face session and for the face-to-face group, the same grouping was made, albeit artificially. In other words, even though all participants were seated together, subjects 1 and 2 were treated as a co-located group so that generalizations across sessions could be made. Tables 4.9-4.12 display the chi-square analysis for each group for each session. The "Within" and "Between" categories refer to turntaking within each group and between each group. Subjects remained constant in groups across all sessions.

TABLE 4.9 Face-to-Face Group Turntaking Data

| | *Within (1)* | | *Within (2)* | | *Between (1)* | | *Between (2)* | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Session* | *Obs* | *Exp* | *Obs* | *Exp* | *Obs* | *Exp* | *Obs* | *Exp* | *X^2* | *Significance* |
| 1 | 41 | 50.43 | 49 | 50.43 | 106 | 100.67 | 106 | 100.67 | 2.36 | N/S |
| 2 | 40 | 41.42 | 38 | 41.42 | 84 | 82.67 | 86 | 82.67 | .48 | N/S |
| 3 | 41 | 53.44 | 55 | 53.44 | 111 | 106.67 | 113 | 106.67 | 3.51 | N/S |
| 4 | 56 | 51.5 | 49 | 51.5 | 102 | 103 | 102 | 103 | .52 | N/S |
| 5 | 51 | 78.17 | 71 | 78.17 | 175 | 156.33 | 172 | 156.33 | 13.90 | p<.005 |
| 6 | 96 | 72.5 | 31 | 72.5 | 154 | 145 | 154 | 145 | 32.5 | p<.005 |
| 7 | 30 | 51.5 | 61 | 51.5 | 109 | 103 | 109 | 103 | 11.43 | p<.01 |

TABLE 4.10 1 fps Group Turntaking Data

| | *Within (1)* | | *Within (2)* | | *Between (1)* | | *Between (2)* | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Session* | *Obs* | *Exp* | *Obs* | *Exp* | *Obs* | *Exp* | *Obs* | *Exp* | *X^2* | *Significance* |
| 1* | 4 | 47.167 | 83 | 47.167 | 96 | 94.33 | 100 | 94.33 | 67.10 | p<.005 |
| 2 | 54 | 80.1 | 13 | 26.7 | 100 | 80.1 | 100 | 80.1 | 25.42 | p<.005 |
| 3 | 157 | 157.2 | 50 | 52.4 | 160 | 157.2 | 157 | 157.2 | .16 | N/S |
| 4** | 16 | 27.72 | 7 | 27.72 | 71 | 53.33 | 72 | 53.33 | 32.84 | p<.005 |
| 5*** | -- | -- | 19 | 9.67 | 20 | 19.67 | 20 | 19.67 | .017 | N/S |

* Within Group (1) missing group member #5
** Within Group (1) missing group member #1
*** Within Group (1) missing group members #2 and #5

TABLE 4.11 25 fps Group Turntaking Data

| | *Within (1)* | | *Within (2)* | | *Between (1)* | | *Between (2)* | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Session* | *Obs* | *Exp* | *Obs* | *Exp* | *Obs* | *Exp* | *Obs* | *Exp* | *X^2* | *Significance* |
| 1 | 44 | 56.95 | 29 | 56.95 | 133 | 113.67 | 135 | 133.67 | 23.95 | p<.005 |
| 2 | 39 | 57.78 | 59 | 57.78 | 125 | 115.33 | 123 | 115.33 | 7.45 | N/S |
| 3 | 57 | 51.60 | 37 | 51.60 | 106 | 103 | 109 | 103 | 5.14 | N/S |
| 4 | 46 | 44.59 | 31 | 44.59 | 90 | 89.33 | 101 | 89.33 | 1.52 | N/S |
| 5 | 42 | 46.59 | 37 | 46.59 | 100 | 93 | 100 | 93 | 3.47 | N/S |
| 6 | 110 | 63.46 | 23 | 63.46 | 123 | 126.67 | 124 | 126.67 | 60.00 | p<.005 |
| 7* | -- | -- | 91 | 93 | 94 | 93 | 94 | 93 | .09 | N/S |

* Within Group (1) missing group member #1

TABLE 4.12 10 fps Group Turntaking Data

| | *Within (1)* | | *Within (2)* | | *Between (1)* | | *Between (2)* | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Session | Obs | Exp | Obs | Exp | Obs | Exp | Obs | Exp | X^2 | Significance |
| 1 | 44 | 44.42 | 44 | 44.42 | 89 | 88.67 | 89 | 88.67 | .01 | N/S |
| 2 | 5 | 10.35 | 15 | 10.35 | 21 | 20.67 | 21 | 20.67 | 4.86 | N/S |
| 3 | 15 | 28.22 | 30 | 28.22 | 62 | 56.33 | 62 | 56.33 | 7.45 | N/S |
| 4 | 27 | 18.87 | 9 | 18.87 | 38 | 37.67 | 39 | 37.67 | 9.04 | p<.05 |
| 5 | 34 | 40.25 | 21 | 40.25 | 93 | 80.33 | 93 | 80.33 | 14.17 | p<.005 |
| 6 | 11 | 10.02 | 6 | 10.02 | 22 | 20 | 21 | 20 | 1.96 | N/S |
| 7 | 15 | 20.04 | 12 | 20.04 | 47 | 40 | 46 | 40 | .90 | N/S |

One of the largest concerns when conducting a chi-square is violating the assumption that observations are independent from each other. An interesting and historic debate on this issue (Lewis and Burke, 1949; Peters, 1950; Pastore, 1950; Lewis and Burke, 1950; Howell, 1992) helped clarify the implications of violating this assumption. The realization that a large portion of chi-square analyses in the published literature violated numerous assumptions - particularly the independence of observations assumption - was made first by Lewis and Burke (1949). Peters (1950) however, makes the point that if the research objective is to make claims about the sample, and not to generalize to the wider population, then violations of this assumption are not problematic. Such is the case for the present study, in which an understanding of one group's behavior within a particular session is not intended to describe how other similar groups might behave. By describing the turntaking behaviors of the participants in this study, claims regarding the communication effectiveness do not extend beyond this study.

In addition to studying the nature of conversation turntaking, the number of turns per minute per person for each group was calculated. Table 4.13 shows these data for all sessions except the first, in which all groups met face-to-face. A 1-factor, between-subjects ANOVA determined that there was a significant effect of communication condition ($\underline{F}$ (3,18) = 5.77, p=.006). A Newman-Keuls post hoc test revealed that the face-to-face group differed significantly from the other three groups and that the video groups did not differ significantly from each other.

TABLE 4.13 Number of Speaker Turns per Minute per Person

| *Condition* | *Avg. Number of Turns* | *Standard Deviation* |
|---|---|---|
| Face-to-face | 2.47 | .51 |
| 1 fps | 1.40 | .21 |
| 10 fps | 1.51 | .54 |
| 25 fps | 1.71 | .48 |

As with the communication breakdown analysis, the turntaking analysis was subjected to a reliability test to determine whether or not the two observers agreed well on the recording of turn-taking. One session was analyzed in common by both observers. A Pearson's product moment correlation coefficient was calculated on the data for the four types of communication (two within group and two between group). An $r^2$ of .995 indicated that there was excellent agreement between the observers.

## Ethnomethodology

Ethnomethodology is often conducted by consulting with an informant or an insider. While this is not necessary according to Hughes, et al. (1995a), the researcher in this study was in a unique position to understand the group being studied. The researcher has taken the course from which the participants were enrolled and has nearly the same background in the human factors program. This commonality makes understanding the pressures, motivation, working styles, and goals of the participants easier and more accurate than a researcher from a different background.  Additionally, the researcher knows nearly all of the participants personally, so is better equipped to understand their work from their perspective.

The extensive ethnographic studies described by Hughes, et al. (1995a) provide the framework for this analysis.  In general, three themes are examined:  distributed coordination, plans and procedures, and work awareness (each was discussed in depth in Chapter 2). Observations were based on verbatim evidence culled from the videotaped sessions, and are presented as data from which discussion evolves.  Prior to this analysis, however, a general background of the groups of interest is presented.

### *General Background*

As mentioned in the Method Section, the participants in this study were enrolled in ISE 5634:  Training System Design, and were mostly first and second year graduate students in the Human Factors Engineering option of the Industrial and Systems Engineering department at Virginia Tech.  The primary assignment for the semester-long course was to complete a training program using the systems approach.  Groups were formed the first week of class, based on mutual project interest and personality.  No restrictions were made on who could form a group other than approximately four students be members.  The project accounted for 50% of the final grade, and the class was required for all students in the human factors engineering option.  The grading for the project was based on a final report submitted as a group (30%), group presentations to the class and class participation (10%) and participation within the group (10%). The group participation portion was based on evaluations made by the other members of the group.

TABLE 4.14  Work Flow Activities

| | |
|---|---|
| *Scope Discussion* | Determining the scope of the project; reviewing and discussing the assignment details; compare scope to other groups/semesters |
| *Work Discussion* | Updating each other on individual work done over the past week; discussing the technology (object of training program); discussing class presentation details; compare progress to other groups/semesters; review old copies of previous projects; discuss what remains to be done; brainstorming; discuss current training methods; discuss what resources are available |
| *Work* | Writing a requirements specification document |
| *Overhead* | Setting the agenda for the meeting; assigning work duties for the next week |
| *Digression* | Off-subject discussion |

*Distributed Coordination*

The theme of distributed coordination refers to how the participants organized the work process. The goal is to describe the actual work being done in the sessions. Table 4.14 depicts the many activities which made up the sessions, not all of which were associated with all groups. Many of these activities occurred each session. An attempt was made to classify the activities into five major types: scope discussion, work discussion, work, overhead, and digression. These categories are loosely related to the taxonomy of design group activities derived by Olson, et al. (1992).

*Artifacts, Plans and Procedures*

While Table 4.14 seeks to provide a general picture of what activities comprise the work process, this section will describe what artifacts, plans, and procedures were used to support the coordination of those activities between the participants. Hughes, et al. (1995a) presented case studies in which artifacts are included in this portion of the analysis and others in which is separated. Because artifacts such as notes and handouts are intimately tied to the procedure for which many of the design meetings were conducted, these elements will be treated together in this analysis.

Each group chose to coordinate their discussion in different ways. It was found that several of the groups (face-to-face, 10 fps, and 25 fps) implemented a relatively formal structure for the group sessions. In general, this organization involved each group member taking a turn to report back to the group what had been accomplished on an individual basis since the last meeting. For example, a group member might have been assigned in the previous meeting to meet with users of the technology of interest during the week to determine what training method is currently in place. In the subsequent meeting, that individual would relate the knowledge gained to the group. In the case of the 25 fps group, this would be accomplished with the aid of a handout. This artifact would contain a description of what was learned and copies would be made for all group members. This was especially crucial for a non-co-located group, as they were unable to share a common document. Figure 4.25 describes such an example. In this case, P1 and P2 are co-located participants; P3 and P4 are co-located. This episode takes place at the beginning of the first video session for the 25 fps group.

1. P1: Facing camera while others are looking at papers: "We can just go through this now, if you like."
2. P3: Looking up at camera: "You're talking about the page you just gave us?"
3. P1: Nods, looking into camera: "Yeah."
4. P3: "OK": Holds up page directly in front of camera.
5. P4: Laughs: "That was funny": P3 laughs with P4.
6. P1: Smiles into camera: "What's so funny?"
7. P3: Scratches head, smiles: "Nothin'."
8. P4: Looking down at handout as others do the same: "You wanna talk us through this for a second?"
9. P1: "Sure": Gesturing with hand along one side of the handout: "On the left hand side you've got the general approach...uh, very general...just ways we can approach the problem."
10. P2: Interrupts P1: unintelligible.
11. P1: Uh, ok, there's a little bit of reverb...we can hear ourselves..kind of a delay..."
12: P3: "Just keep talkin'."
13. P1: Continues to talk about the handout, as other members direct attention toward it.

FIGURE 4.25 25 fps Group, Session 2

Three observations can be drawn from this episode. First, this is a typical interaction of coordinating work between participants through the use of an artifact (pre-distributed handout). Such an artifact allows all participants to maintain topical orientation and be aware of the direction of the ensuing discussion. Second, this example suggests how the communication system can, itself, obstruct communication. When P3 held the page up to the camera, he and P4 found this novel use humorous. However, only the laughter - and not the source of the humor - was made apparent to the remote participants, causing them to ask, "What's so funny?" Finally, this initial exposure to the video system and the subsequent troubles with the audio feedback are discovered by the speaker, P1. P1 comments on the audio trouble, is requested to continue speaking, and finally continues with the work at hand. The main point of this example is that by distributing a handout prior to the session, the workflow is structured and topic orientation is maintained.

For all groups except the 1 fps group, the actual work was accomplished in isolation and the meeting was used primarily to discuss what was done and to update the rest of the group. In one instance, group members worked on components of the larger task of writing requirements specifications for the training program. Each group member then discussed the section for which he or she was responsible. In doing so, issues were raised which required discussion from the entire group. An interesting by-product of this artifact-driven procedure resulted, however. Often, participants would refer to specific aspects of a shared document. Because they were distributed, however, they had to describe which part of the document they were referring to (or which document). Figure 4.26 depicts an interaction for the 25 fps group in which orientation between the groups is maintained in terms of both inter and intra document references.

1. P3: Looking over a manual and describing one part of it: "...There's a data log, not just a calibration log, but a data log...and it has things on it for..."
2. P2: Thumbing through the manual: "Where in the sheet?"
3. P3: "Its not...Oh, no, there it is...um..."
4. P4: Thumbing through the manual.
5. P3: "A couple pages back from where you're lookin' at...I was lookin' at it in one of the books we've been passin' around..."
6. P4: Finds the page and holds the manual up to the camera for the others to see.
7. P1: Flips to the page in his copy of the manual: "OK, and in which part of the handout that you gave us?"
8. P3: "After the last horizontal rule."
9. P1: Picks up handout, turns page: "OK"

FIGURE 4.26 25 fps, Session 2

If these documents were viewable in a shared window on the computer screen, two problems would be solved. First, a cursor could be used to focus attention on the part of the document of interest, avoiding unnecessary description and resulting in more efficient communication. Second, gaze could be maintained in one location - the screen - avoiding transitions between papers held in the hand and the visual portion of the conferencing system. Regardless of the limitations of using handouts, they still retain topical focus and assist the coordination of the work (in this case, review of prior work).

While the 25 fps group tended to structure their sessions around shared handouts, the 10 fps group relied on taking turns reporting the results of the previous week's individual work. For example, one group member would take approximately 15 minutes describing what was done, what issues were resolved, and then ask questions of the other group members, facilitating and directing discussion.

The 1 fps group differed from the others in this regard. They sought to *perform some of the work* while in session. For example, questions to be posed to a subject matter expert were formulated during the sessions in a group format, rather than assigning the task to an individual to be accomplished off-line before the next session. Because this task relies heavily on verbal coordination and participation among group members, the 1 fps group had a difficult experience. Figure 4.27 illustrates this process. In the example, P1 and P2 are co-located, and P3, P4, and P5 are co-located.

1. P4: After trying to collaboratively write a set of requirements specifications for the training system using the whiteboard: "I think its easier to write it... than type it."
2. P1: Grabbing a pencil and paper: "Allright, I'll be the writer."
3. P1: Some time is spent trying to interpret what was typed into the whiteboard by all group members, brief discussion of the video being slow: "Allright, what else do we want to ask him?"
4. P4: Looking at video: "Well, what kind a chemicals are in there, I mean, uh what's the nastiest it can get?"
5. P1: Takes notes, looks up at video: "So, maybe a syllabus?"
6. P3: Looks at co-located group members in surprise: "A Syllabus?...No, like some kind of a chemical data sheet or inventory list or something like that."
7. P2: Tries to speak, but is drown out by P4.
8. P4: Looking at video: "Just like a list of what would be used...What would be used in a student lab and what would be used in a GRA Lab, something like that.": P1 taking notes.
9. P2: Looking at video: "We could ask him, um...which area he would want us to concentrate on.": P1 taking notes.
10. P4: Looking at video: "Right, what is the uh, organizational goal."
11.: P1: Nods: "Right."
12. P4: Looking at video and around the co-located group: "Hey does anyone have that list of stuff that we were going to ask?"
13. P1: Reaches down into a bag at floor: "I did": Places document on keyboard for both co-located members to see.
14. P2: Reads off items from document: P4 and P5 (remote) don't listen and talk to themselves. P1 and P3 are listening.
15: P4: Looking at video: "Yeah, those are questions...Hey, maybe we could ask if..."
16: P1: Tries to interrupt.
17: P4: Looking at video: "Sorry."
18. P1: Looking at video: "Go ahead."
19. P4: Looking at video: "I mean, maybe we could ask if there were any accidents in the past..." Goes on to state point: P1 takes notes.

FIGURE 4.27 1 fps Group, Session 2.

This example highlights the difficulty the group had in collaboratively writing a document through the video system. The whiteboard was experimented with, but deemed too cumbersome for the task. Once a group member chose to write the ideas down, conversation grew stilted and inefficient, but the work was accomplished more readily than if they were taking notes via the whiteboard.

The following week, the group decided that having a text window on at least one of the screens would solve the problems encountered earlier with taking notes and using the whiteboard. To this end, one group member who was adept at UNIX, opened a text editor on one machine. This window was not visible to the other group, however. Figure 4.28 displays the interaction among group members as they sought to write the formal specifications for the training system. This document is intended to describe in detail the behavioral goals of the training system. In Figure 4.28, P1 and P2 are co-located, and P3, P4, and P5 are co-located with the text window. P5 is not in the camera field of view.

1. P1: Looking at video: "...OK, I guess first we need to define a skill that falls into each of those categories."
2. P5: Looking at video: "OK"
3. Both groups talk among themselves, preventing the other group from hearing.
4. P5: Quietly: "OK, the first one would be...": Continues to describe a specification in a soft voice while P3 types it into the text window. P1 and P2 are talking quietly among themselves and reviewing a book.
5. P5: "This would be so much easier if we were all in the same room."
6. P3: "Uh huh, yeah."
7. P5: "I mean, this is a challenge."
8. P1: Looking into camera: "Allright, so a psychomotor skill might be something like knowing how to appropriately use an apparatus in the lab?"
9. P2: Looking at video: "Uh huh."
10 P3: Looking at video: "OK, so...
11. P5: "Well, how about for each of our objectives, like remember, identification...uh...[voice is lowered, hand gestures become more prominent] what would be a psychomotor skill?"
12. P1 and P2 are talking privately: P5: "...Like should we have for each of those objectives... a performance objective for each of those objectives?"
13: P1 and P2 are talking privately: P3: Looking at video: "OK, P5 is talking about, for we have the four things: identify, store, transport, and handle..."
14: P5: "And dispose"
15: P3: Looking at video: "And, uh, maybe to ...go through...and listen to myself twice..." P3, P4, and P5 snicker.
16. P1: Looking at video: "That's a good idea, I think, as long as you put performance objectives, intellectual skill, and verbal information..."
17. Groups continue to debate format of the document, how to word specifications, and what ought to be included in the document. P3 types, based on the input primarily from P4 and P5.
18. P1: Looking at video: "What are you working on?"

FIGURE 4.28 1 fps Group, Session 3

Figure 4.28 highlights the difficulty in carrying out actual work (a specification document) through the video conferencing system. The work is organized informally around the task of developing this document. One group member (P3) is the designated typist for the group, but two of the group members are unable to see what is being typed. Clearly, a group text editor would aid the task tremendously by making visible the work output to all members and providing control and input capability to all. In Figure 4.28, not only does the one-sided text window isolate one group, but the poor audio quality and lack of visual cues contributes to the problem. For example, in line 11, P5, who is not visible in the video window lowers her voice and uses hand gestures to illustrate her point. This has the effect of disengaging the other group, who begin to talk privately. The group is reunited in topical orientation once P3, in a booming voice, commands control by repeating what P5 had just said.

At this point, the group has organized their work in this and the previous session around the central task of developing text documents, and in doing so, have attempted three different work styles, all of which failed. Toward the end of the third session, the two groups decide to adopt yet another strategy to developing the specification document. In this case, each group works separately on the same issue, and then reads the language back to the other group to re-work. Then the best of the two is chosen, as illustrated in Figure 4.29.

1. P1 and P2 work silently, writing down their version of the specification. P3, P4, and P5 do the same, but type in the text window and discuss the language aloud to each other.
2. P3: Finishes typing: "Cool? OK, P5 is going to read it now."
3. P5: Reading off of text window: "All right...Given a hazardous material, er...a...no...Given a hazardous chemical, choose the appropriate type of container to store the chemical and then place the container in the appropriate location for that chemical (i.e. the proper temperature, humidity, height, etc.).
4. P1: Looking at video: "Read it again?"
5. P5: Sighs: "Given a hazardous chemical, choose the appropriate type of container to store the chemical and then place the container in the appropriate location for that chemical (i.e. the proper temperature, humidity, height, etc.).
6. P1: Pause: "Well..."
7. P5: "Or would you rather have environment?"
8. Silence: P1 and P2 talk privately.
9. P1: "choose the appropriate container in which to store the chemicals?"
10. P3 and P5: "Yes"
11. P2: "Read it again?"
12. P5: "What?"
13: P4: In heavily-accented English: "Ah right. Um...I'll read again...practice my English...Ahright...Given a hazardous chemical, choose the appropriate type of container to store the chemical and then place the container in the appropriate location for that chemical (i.e. the proper temperature, humidity, height, etc.).
14. Silence: P5: Pointing to the text window: "Can they not see this?"
15: P3: "What?"
16. P1 and P2 talking privately: P5: Again pointing to the text window: "We're the only ones who have this?"
17. P3: "Oh, yeah."
18. P1: "Allright, P2 is gonna tell you how to change it."
19. P3: "Oh, she is?": Laughter among both groups.
20. P2: Reading off a notepad: "OK, given a hazardous chemical...
21. P3: "Uh huh."
22. P2: "Chose to store it in the appropriate container under safe environmental conditions, such as humidity, temperature...
23: P4: "OK"
24: P2: "...etc."
25: P4: "OK, can you read that again, please?"
26: P2: "Given a hazardous chemical...
27: P4: "Uh huh"
28: P2: "Chose the storage...
29. P3: Typing in what is being said: "OK, hold on...
30. P4: OK, stop, time out...
31. P4: "Choose the storage.": P3 types.
32. P5: "OK"
33. P2: "In the appropriate container... under safe environmental conditions...
34. P4: "OK, allright..." P3 types.
35. P4: "To store it..."
36: P5: "Under safe environmental conditions.": P3 types.
37. P4: Points to text window: "Such as temperature, humidity, etc.".
38. P3 types while P4 and P5 point to the text window and make quiet suggestions for wording changes. P1 and P2 talk privately.
39. P5: "OK...what's the next one now?"

FIGURE 4.29 1 fps Group, Session 3

The example in Figure 4.29 took place over six minutes. Clearly, this method is unproductive and frustrating to all group members. The wording of one statement in the document is read aloud five times, several of which would have been unnecessary had the audio quality been better. Also of interest is the lack of awareness afforded by the text editor. In line 13, after using the text editor for over half an hour, one group member finally realizes that the other group does not have access. After trying four different methods of writing a text document collaboratively, it appears as though all were failures.

It is interesting to compare the previous examples with an example of document sharing within the face-to-face group. In Figure 4.30, the group is discussing the scope of the project. Specifically, they have obtained a copy of a previous semester's project to determine how much work is required.

1. P1: Pulls out a copy of a previous semester's project, flips to a certain page, points at the text as the others look on: "But for example, you go through...what they do is they talk about the rationale...you come down here, and the result of the task analysis...all they do is...": Continues to point out what another project encompassed, while other group members look on.

FIGURE 4.30 Face-to-face group, Session 5.

While this example does not demonstrate the use of materials to structure a session, it does highlight the ease with which all group members maintain mutual understanding. There is no need to expend time or resources explaining what part of the document - or even which document is being referenced. The face-to-face group did not attempt to write a document together, nor did the 10 fps or 25 fps groups. In one situation, the face-to-face group had worked on presentation slides during another, non-taped session. A print out of the slides was brought to a meeting at which point they flipped through the slides to divide the responsibilities for the presentation. Figure 4.31 depicts this interaction.

1. P1: "OK, I thought we were doing it..."
2. P1: Points to three piles of papers on the desk: "...we had talked about lumping this [pats hand on a stack of papers] together with the objective [pats hand on another stack of papers] and the target audience [pats hand on yet another stack of papers]..."
3. P2: "P1 talks about this [points at one stack] and then I talk about this [points at another stack]: All talk at once.
4. P3: Flips through presentation slides: "Wa - Wait a minute....start from the top"
5. P2: Leans over and points at the different slides as P3 flips through them and the other two look on: "P1":
6. P3: Flips page: "P1"
7. P2: Points at next page: "P1"
8. P3: Flips page: "P1"
9. P2: Points at next page: "Me"
10. P3: Flips page: "You"
11. P2: Points at next page: "Me"
12. P3: Flips page: "You"
13. P2: Points at next page: "You"
14. This process of assigning presentation slides continued as all group members looked on.

FIGURE 4.31 Face-to-face group, Session 2

From the example in Figure 4.31, it is apparent that the work was highly coordinated, with no interference in either the auditory or visual channels. Much of the overhead needed to maintain topical orientation in a distributed group (i.e. "At the top of page 3, second paragraph, third line...") is not necessary for a small, co-located group which can rely heavily on visual cues.

From the analysis thus far, it appears that only the 1 fps group had difficulty organizing, coordinating, and conducting work sessions. It has been shown that the methods employed and the goals sought from this group were not supported by the video conferencing system. The examples provided were of the first two video sessions for the 1 fps group. The next two sessions were characterized by a somewhat different approach. In the third video session, the group had just four members and executed the work process much the same as the other groups. They abandoned the attempt to create formal text documents collaboratively and settled for high-level discussion and note-taking. Issues such as project scope, resources available, and questions to ask a subject-matter-expert were discussed. Despite this new tack, remnants of the old working style were still evident. For example, in one case, a group member read off a document while a remote group member took notes on it. In another instance, a group member sought to take notes on all ideas generated from the discussion. In both cases, however, notes were informal and precise language was not a concern. Requests for repeating were kept to a minimum.

In the fourth (and final) video session for the 1 fps group, only three group members were present, creating an interesting situation. As in the previous session, discussion was maintained at a high-level. Although detailed work was not being performed, the group member who was isolated grew highly disengaged from the other two group members. At times, the interface was distracting, evidenced by tinkering with window positions and sizing. Reading and consulting with text documents remained the focus of the session, but writing them was not.

The face-to-face, 10 fps, and 25 fps groups conducted most of their sessions in a similar format. Group members often reported on what had been accomplished since they last met, which often sparked discussion on issues which remained to be resolved. This discussion was usually high-level and frequently concerned the scope of the system, the technology of interest, and resources available. To assist in answering these questions, group members often consulted a range of books, notes, class handouts, and documents which they had created earlier. The end of these sessions were characterized by assigning tasks to be completed individually, as well as specific deliverables to be finished by the next session. Often notes were taken during these sessions, but these were kept informal and brief. The 10 fps group and the 25 fps group often relied on handouts distributed prior to the session to maintain a shared orientation throughout the meeting.

#### *Awareness of Work*

The means by which group members make their work visible - or salient - to others is of interest in distributed groups. Hughes, et al. (1995a) presented the idea of "ecologies of work", physical configurations which support work awareness. In the case of the video conferencing environment, there is an obvious impediment to making work activities salient to all participants. For example, the field of view can prevent group members' activities from being transmitted to a remote group.

One of the most frequent problems in maintaining work awareness is when a remote group fails to understand what the other group is discussing. Such an event occurs when a document is being referenced and only one copy is available, putting the remote group at a disadvantage. In Figure 4.32, the 1 fps group suffers from this problem, seeks to overcome it, and finally, abandons the activity in favor of a task which does not require a document to be shared. In this example, P1 and P2 are co-located and P3 is alone at the remote side.

1. P1 and P2: Discussing details of a document which is laid on the keyboard in front of them. P3 looks into video window, listening.
2. P3: Looking at video: "Uh, I'm lost again."
3. P1 and P2 laugh.
4. P1: "I wish I could show you this"
5. P1: Holds up document to camera: "This is what I'm looking at": Document slowly enters the video window of P3.
6. P3: Moves closer to the video window: "Where's that from?"
7. P1: Still holding document to camera: "Its a little thing that I typed up."
8. P3: "Not with me...I don't think...let me look.": Moves back to look through bag.
9. P3: Returns to workstation: "No I don't."
10. P1: "OK...Well, still want to meet for five more minutes then?"
11. P3: Looking into video: "Sure...Its just hard to... I mean... I have learned a bunch of stuff about UHCP: University Hazardous Chemical Plant, er something..."
12. P1: "Tell us what you've learned about that."
13. P3 goes on to relate what he's learned about the UHCP.

FIGURE 4.32 1 fps Group, Session 5.

The example in Figure 4.32 demonstrates how easily a remote group member can become unaware or lost in a discussion. Once the problem is identified, the camera is used to show what the center of discussion is. Despite the remote group member's efforts to discern what is being shown in the video window, he still must ask that it be identified. Subsequently, a futile search is made for a copy of the document. Finally, the discussion is abandoned - presumably due to the difficulty of not having two copies of the document - in favor of another activity. This new activity involves reporting back to the group on work done individually during the week and, ironically, is what the other groups found most effective.

At times, tools designed to facilitate the work process and maintain awareness of work across distributed groups failed to accomplish this task. Such was the case in Figure 4.27 when the 1 fps group sought to write a text document using the whiteboard. The 10 fps group also experimented with using the whiteboard to make brief high level notes. In this case, however, the whiteboard connection became disabled, preventing the remote groups from sharing information. In the first case, the whiteboard failed to meet the requirements of the session simply because it is an inadequate technology for extended text editing. In the second case, the whiteboard was suitable for short text messages, but the lost connection precluded its ability to maintain awareness between the groups. Again, these examples point to the need for a shared text editor to support the primary activity of the group sessions.

Another example in which shared work awareness was maintained was through the unique uses of the video channel. For example, often the groups would make reference to a document or book. To make the remote group understand what is being referenced, the document would be held up to the camera. Because the majority of the work done during the sessions was in the form of discussion, making the work aware to the other group members was accomplished through two mechanisms: maintaining a sufficient audio level and maintaining interest or attention level.

## Discussion

This section will provide some interpretive discussion of the results presented in the previous section. First, the findings from each of the four main research methods will be presented separately. Next, the methods will be compared in terms of the quality of information gathered and cost to conduct the analysis. The significance of these findings with respect to the literature on video-mediated communication will be discussed, and finally, research implications for continued research will be outlined.

### Contextual Inquiry

Interviewing the participants at the conclusion of each session provided excellent information regarding meeting success, the attitudes of the participants toward the communication system, and the general productivity level. Unfortunately, this type of qualitative, user-generated data is difficult to discuss objectively. Determining whether a comment reflected a true sentiment or was a poorly-considered statement biased by an external factor is difficult to discern. Equally problematic is deciding the magnitude of importance of a particular attitude. For example, it may not be clear whether a statement regarding the use of the video window was felt strongly by a participant, or if it was merely a mild opinion. In an effort to minimize inaccurate interpretations of these comments, only those statements which were made by more than one group member more than one time will be considered valid.

To help organize the many comments which were made during the interviews, Table 4.15 was formulated. Comments are organized according to issues related to the video, audio, and the general physical environment, including the user interface. While some comments may overlap or be related to others, this distinction provides a mechanism by which discussion can be focused. The comments received from the face-to-face group were not useful in that they did not describe any communication problems. Most comments from the face-to-face group were related to productivity levels during and between the sessions and the feeling of being observed.

By studying table 4.15, one can discern several themes. Because all three groups experienced the same audio quality, it is not surprising that there was a degree of consensus regarding the troubles encountered with this channel. Most significant throughout the study, were comments regarding the lag, and to a greater extent, the echo. The audio quality was full duplex, allowing two parties to speak simultaneously, but completely suppressing the echo in the system was not possible. At the start of all sessions, the speaker volume and microphone sensitivity were set to a level at which echo was minimized (these levels were determined after extensive subjective testing by the researcher and an assistant). Despite these efforts, the echo was always present, and could vary from session to session; for example, if a group decided to increase the volume, the echo problem would be worsened for the other group. It is felt that audio decrements such as those imposed by this system are common to internet conferencing. Regardless, audio is the most crucial channel in a communication system (Chapanis, 1975), and should be maintained at a level similar to telephone quality to ensure communication success and technology acceptance. All findings from this study may be impacted by the quality of the audio system.

A comment was made by two of the groups regarding the loss of control by the listener in a conversation due to the audio channel. It was stated that the speaker maintained control because of the difficulty in interrupting the speaker to take the floor. Furthermore, it was stated that once the floor had been obtained, the speaker felt compelled to speak in longer statements, taking advantage of the opportunity. This problem appears recursive as frustrated group members compete to take the floor, maintain it longer than normal due to the difficulty in originally obtaining it, and further frustrate the other group members. This finding is supported by the interaction analysis in which the face-to-face group was determined to have more speaker switches than the video groups. The problem of interrupting speakers' fluidity may also result from a lack of visual cues. For example, much research has established that facial expressions, gesture, and eye contact are often used to take the floor from a speaker in a fluid, unspoken manner (Duncan, 1972; Duncan and Niederehe, 1974; Rutter and Stephenson, 1977). The video quality may have contributed to the speaker interruption problem experienced by the participants in this study.

TABLE 4.15 Interview Themes

| | |
|---|---|
| *Audio Issues* | |
| Lag in audio transmission was annoying and cumbersome | All Three Groups |
| Echo, reverberation, and feedback were annoying and cumbersome | All Three Groups |
| Speaker has control of the conversation, not the listener; Difficult to interrupt | 10 fps, 25 fps |
| *Video Issues* | |
| Field of view was inadequate | 10 fps, 25 fps |
| Video was distracting | 1 fps |
| Eye contact was impossible | 25 fps |
| Facial expressions were lost | 25 fps |
| *Environmental/User Interface Issues* | |
| Text Editor was needed | 1 fps |
| Whiteboard was distracting | 1 fps |
| Two people on one end was too many | All Three Groups |
| *Other* | |
| Talking to computer monitor took some getting used to | 25 fps |
| The system was very frustrating, especially knowing the group was nearby | 1 fps |

The video-related comments raise several interesting findings. It is curious that the 25 fps group commented on a host of limitations of the video channel; for instance, the restricted field-of-view, inability to make eye contact, and the loss of facial expressions were mentioned. These limitations would be evident at any frame rate, suggesting that any of the groups may have experienced these problems. It is possible that the 25 fps group used the video more than the others, and therefore had more exposure to these problems. The limited field-of-view was found to restrict the transmission of gesture as well as the depiction of all participants. Often, the camera was not positioned to capture both participants or the group was situated too far apart to be conveyed well.

The 1 fps group noted that the video was often distracting. It seems as if such a low frame rate was more a hindrance than a help in communication between the two groups. This finding supports the results of a low frame rate distance learning study conducted by Kies, Williges, and Rosson (in press, b), in which students suggested that poor quality video prevented full attention to the auditory channel, which was most crucial to the message.

The environmental issues associated with using the video conferencing system were crucial to its success. All three groups commented in one fashion or another that having two participants on one side of the video conference was too many. The 1 fps and 25 fps groups had direct experience with this phenomenon when group members failed to attend a session. The 1 fps group had five members, but only four were in attendance for one session and only 3 were in attendance for another session. Following these sessions, the remaining members commented extensively on how much easier it was to conduct a conference. The 25 fps group also held a session with three members, after which, the ease of conducing the session was commented upon. Indirectly, the 10 fps group - which had full participation throughout the study - commented on the difficulty in having a group of two on one end of the conference. This group suggested that the co-location seating configuration was cumbersome. It was deemed difficult to speak to the co-located group member when both were facing the video monitor. A number of explanations can be set forth to account for the difference between one and two group members on one end of a conference. First, reducing the number of participants in any session will reduce the frequency of simultaneous speech and attempted interruptions. Second, directing conversation toward a specific participant is simplified. Finally, the video camera's field-of-view depicts a single participant better than two, permitting more gesture, and other visual cues to be conveyed by a group member.

The other environmental comments were made by the 1 fps group. This group repeated often the need for a text editor and the tendency of the whiteboard to be distracting. This finding is likely the result of the type of work conducted by the 1 fps group. As was evidenced in the ethnomethodology section, the 1 fps group sought to prepare text documents during the sessions, as opposed to the other groups who focused their energies on high-level discussion.

The final comments dealt with attitudes toward the system. The 25 fps group noted that it felt strange to converse with a remote group by looking and talking at a computer monitor. They noted in a later session, however, that they were able to get used to this feeling, and no longer viewed the medium as an awkward communication system. This finding supports the need for relatively long-term studies with new communication technologies in that shorter studies may fail to reveal such effects. The 1 fps group also expressed an interesting attitude toward the system. They felt frustrated throughout the study for a number of reasons. Aside from the communication difficulties they experienced, they felt that knowing the other group was located a short distance away made the session all the more frustrating. They often felt that their work would be accomplished much more efficiently at the end of the session when they were free to meet together in one room.

The results from the contextual interviews at the end of each session provided the participants with a chance to explain the experience from their perspective while their feelings were fresh. The information gained from the interviews was collected in an open-ended, semi-structured format, allowing the participants to say anything that was on their mind. Additionally, they were free to spark each other's imaginations by making comments based on what another group member had said. This information was highly contextual as the interviews were conducted near the conferencing system and resulted in user-generated, qualitative data. The questionnaires, which were also completed at the close of each session and provided user-generated information, were structured and quantitative in nature. These data will be discussed in the next section.

## Questionnaires

The extensive questionnaire provided in Appendix B was given to each participant in the video conferencing conditions at the conclusion of each session. Likert-style questionnaire items provide a time-tested method for gathering quantitative user-generated data on a system. The questionnaire was divided into four sections: questions concerning the session from the perspective of a speaker, questions concerning the session from the perspective of a listener, the technology, and the environment. These sections were intended to reflect four of the most crucial elements of a communication system.

The disadvantage to the questionnaire data in this study was the fact that statistical analyses were not appropriate for the data set. As mentioned earlier, one group dropped out of the study after five weeks and some group members failed to attend several sessions for two of the groups. In light of these problems, the data were simply plotted for each question for each group across time. Four themes were used to interpret the data: large changes in response over time, universally negative or positive responses, large differences between groups, and differences between similar question items from the perspective of a speaker and the perspective of a listener. Each question will be discussed with respect to these considerations. The following items pertain to the perspective of the speaker.

"I felt as if my message was well-understood by the remote group": The responses for this item yielded little change over time and were relatively similar between groups. Most response reflected agreement with this statement, suggesting that participants did not feel as though conversation was lost in the communication medium.

"I was able to keep the remote group's attention": This item also yielded fairly stable responses over time with no large differences among groups. Again, the responses were generally positive, indicating that speakers perceived little trouble in maintaining the remote group's attention.

"I knew if the remote group was paying attention": The responses for this item were somewhat lower for the 1 fps group than for the other two groups. Little change was evident over the course of the study. Although there appeared to be a difference between the 1 fps and the other two groups, the 1 fps group provided responses in the neutral range of the scale, indicating that there were no strong feelings that the remote group was not paying attention.

"I was not distracted by the computer arrangement when talking": This item yielded responses that were consistent between groups and over time. Responses were slightly low, indicating a mild sentiment that the computer arrangement was a distraction. This sentiment was perhaps most felt by the 1 fps group whose members were easily distracted by the electronic whiteboard.

"I thought this arrangement was just as effective as face-to-face communication": In general, there was a strong feeling that the conferencing system was not nearly as effective as speaking in person. This position did not change over time for the groups.

"I was able to make eye-contact with the remote group": This item reflected a large difference between the 10 fps group and the other two groups. Curiously, the 1 fps group and the 25 fps group provided similar responses. Such a finding appears to reflect the fact that improved frame rate will not enable eye contact to be made. The high responses of the 10 fps group may be indicative that group's higher marks on most items of the questionnaire. Such a finding is not surprising given the small sample sizes and the ability for one or two group members to influence the group average on an item.

"I was able to see the remote group's facial expressions": This item also reflected large differences between the 10 fps group and the other two groups. This result fits with the contextual inquiry finding that the 25 fps group made complaints regarding the inability to perceive facial expressions. Additionally, the 1 fps seemed to make lower responses over time. This finding may be due to the frustration felt by the 1 fps group.

"I was able to see the remote group's hand gestures": The responses for this item were fairly low, and did not reflect much difference among the groups or over time. This item shows that participants had more trouble perceiving hand gestures than facial expressions. This finding can be explained by noting that the camera was focused on the participant's heads - not their hands - making gesture much less prominent than facial cues in the video window.

"I was able to see the remote group's body movements": In general, this item reflected the same responses as from the previous item regarding gesture. Such a finding is not surprising since body movements are related to hand gestures, neither of which were the primary focus of the camera.

"I did not feel awkwardly interrupted by the remote group": This item reflected similar responses between groups and over time. Responses tended to lie in the "indifferent" range, suggesting no strong opinions of being interrupted awkwardly by the remote group. This finding is interesting in that the contextual inquiry interviews revealed that the 10 fps and the 25 fps groups had difficulty interrupting other speakers. It is possible that the speakers were unaware that other group members were attempting to gain control of the floor. The lack of visual cues such as gesture and eye contact support this claim.

"I felt that transitions between who was speaking were smooth": For the most part, responses to this item reflected a large difference between the 10 fps group and the other two groups with little change over time. While the reason for this is difficult to assess, it may be the result of the 10 fps group consistently providing higher responses than the other two groups. This issue of speaker transition will be discussed in more depth in the Interaction Analysis section.

The next set of items pertain to the perspective of the listener. The same considerations will be mentioned for these items; however, it is interesting to note the differences between an item in this set and a nearly identical item in the previous set to determine if participants had different attitudes based on being a speaker or a listener.

"I was able to understand the remote group's message": This item reflected relatively similar responses across time and between groups. In general, the groups felt as though they were able to understand what the remote group was saying. The responses corresponded fairly well with those of the related question from the perspective of the speaker.

"I did not have to strain too much to understand the remote group's message": The responses for this item were similar to those of the previous item, suggesting that the communication process did not require a large amount of effort.

"I paid attention to the remote group": Little difference between groups and over time was evidenced in this item. The response were generally positive and corresponded well with the speaker's perspective.

"I was not distracted by the computer arrangement when listening": Again, little difference between groups and over time is reflected in this item. Similarly, the responses appear to correspond well with those of the speaker-oriented item.

"I thought the arrangement was just as effective as face-to-face communication": This item, worded verbatim as the speaker-perspective item reflected nearly identical responses. These responses were low, and the 10 fps group was generally higher than the other groups. An interesting artifact of the similarity between these two items is that evidence is generated which suggests a degree of inter-item reliability. The item also reinforces the claim that the conferencing system was not as effective as face-to-face communication.

"I was able to make eye contact with the remote group": This item, also verbatim with that from the speaker perspective, indicates a large difference between the 10 fps group and the other two groups. Again, the best explanation for this finding may be that the 10 fps group simply rated all items higher than the other two groups.

"I was able to see the remote group's facial expressions", "I was able to see the remote group's hand gestures", and "I was able to see the remote group's body movements": These items reflect similar findings as that of the speaker-oriented items. Again, differences appeared between all three groups for the perception of facial expressions, with the 10 fps group providing the highest responses. Facial expressions were better transmitted than other visual cues such as gesture and body movements. The explanation for this finding is likely based on the fact that the cameras were trained on the upper torso of the participants, and much of the peripheral visual information was restricted. The low frame rate experienced by the 1 fps group probably did not faithfully convey facial expressions which have the tendency to change rapidly, but made little impact on the perception of gesture and body movements.

"I was able to interrupt the remote group easily": This item showed no large differences between groups or changes over time, in keeping with the responses for the speaker-oriented item. However, the 25 fps group had the lowest response of all groups, corresponding well with the data from the contextual inquiry analysis in which many complaints were made by this group regarding the difficulty in seizing control of the floor from another speaker.

"I felt that transitions between who was speaking were smooth": Again, this item reflected strong correspondence with the related speaker-oriented item. In general, the 10 fps group was slightly more positive, and there was little change over time.

The next set of items is concerned with the conferencing communication channels. Specifically, the video and audio channels of the conferencing system were addressed.

"The video quality was acceptable": This item reflected large differences between groups but little change over time. In particular, the 10 fps group rated the video quality higher than the other two groups; the 25 fps group rated the video quality higher than the 1 fps group. While the 10 fps group and the 25 fps group were in the positive range of the scale (i. .e. the video quality was deemed acceptable), the 1 fps group was well within the negative range of the scale, suggesting that the quality was not acceptable.

"The video was clear enough": Because the resolution of the video did not change across groups, the results from this item are interesting. In general, there is good correspondence between the results from this item and those from the previous item. One explanation for this finding may be that the question was interpreted to mean that clarity was related to frame rate. Alternatively, reduced frame rate inhibits the ability to perceive small detail in motion such as facial expressions (this was confirmed by a previous questionnaire item), and may therefore have accounted for the lower responses by the 1 fps group.

"The size of the video window was adequate": The size of the video window could be customized by the participants, which was a simple task given the computer experience of the groups. The fact that different window sizes were used throughout the study may explain the varied responses to this item by the 1 fps and 10 fps groups. The 25 fps group maintained constant responses throughout the study.

"The audio quality was acceptable": There was a tight coupling of responses between all three groups for this item. Over time, the responses increased from the "indifferent" range to more positive responses. This finding indicates that the groups may have grown used to the audio quality and its side effects (feedback and echo). The audio quality experienced by each group remained constant throughout the study. The microphone level and speaker volume were set by the experimenter at the start of each session. Despite this control, participants adjusted the speaker volume occasionally, contributing to increased feedback for the remote group.

"The audio was clear enough": This item resulted in responses which closely corresponded to those of the previous item. The only difference, however, was that the increase over time in the agreement with this statement was much less salient. This difference may be explained by the broader definition of audio quality. For example, audio clarity may be a component of overall quality, but does not encompass the negative characteristics possibly responsible for the initially low responses to the quality item (echo and feedback). Since clarity was never deemed a problem, it was not a characteristic which could be "grown accustomed to".

"The audio was just as good as being in the same room": This item produced similar responses across all groups, but reflected negative opinions of the audio. All responses across all weeks were within the negative range, suggesting dissatisfaction with the audio quality compared to what is normally experienced in face-to-face conversation.

The next set of questions relate to the physical configuration of the conferencing system and other environmental issues. These factors were, for the most part, held constant across all groups and for all sessions.

"The lighting level was appropriate in my room"; "The lighting level on the video image was appropriate": The responses to these items were high for all groups and changed little over time. This finding suggests that the lighting level was adequate and well-suited to a video conferencing system and that there was no significant glare on the monitor detracting from the quality of the image.

"The noise level in my room was not too loud"; "The noise level in the other (remote) room was not too loud" The responses to these items were universally high and consistent over time and across groups. This finding indicates that there was little external noise affecting the subjective impressions of the system.

"The camera angle was appropriate": Because the camera was adjustable, it is not surprising that there was little difference in the ratings of this item over time. The 25 fps group again indicated less positive responses than the 10 fps group, most likely due to the trend that the 10 fps group consistently rates the items more positively than the 25 fps group.

The results from the questionnaires can be summarized in several themes. First, there was a relatively high correspondence between the speaker-oriented questions and the listener-oriented questions. This finding suggests that there was good inter-item reliability. One interesting implication of this consistency was the finding that speakers perceived interruptions and transitions to occur as easily as listeners. This is counter to the finding from the contextual inquiry interview in which members of the 25 fps group mentioned the difficulty in interrupting the speaker and the feeling that the speaker preferred to retain floor control once it was obtained.

The second theme from the questionnaire results was that the 10 fps group consistently rated items more positively than the 25 fps group. This is a peculiar finding given the quality difference between these two video conditions. This is particularly strange given the results from the first study which show that for a talking head scene, increases in frame rate from 10 fps to 25 fps reflect significant increases in perceived video image quality. One explanation for this finding may lie in the variability inherent in small sample sizes. For example, it is possible that the group member in the 10 fps condition maintained a higher frame of reference on the Likert scale. It has been suggested that Likert scales are subject to context effects which can constrain the range of values which subjects are willing to use (Jones and Marks, 1985). Additionally, it has been shown that rating scales can suffer from a variety of biases, including halo and leniency effects (Meister, 1985) Halo effects result from a tendency by subjects to respond favorably to one item after responding favorably to another. Leniency effects concern the occasional finding that observers respond favorably to items in an effort to please the experimenter.

Another theme from the questionnaire concerns the content of the items. The items most sensitive to differences among groups were those that related to the quality of the video and the video's ability to convey details of the face and eyes. This is not surprising since frame rate was the only variable which differed among the groups. It is interesting to note that the items concerning the ability to detect facial expressions and make eye contact were sensitive to differences in frame rate. But, the gesture and body movement items failed to reflect differences in frame rate. Most other items showed no apparent differences between the groups.

The final theme from the questionnaire relates to the degree of change over the course of the study. For most items, there was little change in responses over time. This is interesting, since the 10 fps and 25 fps groups made comments in the contextual inquiry interviews that they grew accustomed to the system over time. The only item to show a more favorable opinion over the course of the study was for the audio quality acceptability. While the audio quality was unchanged during the study, it is possible that participants became accustomed to the feedback and echo in the audio system.

In light of the observations regarding the questionnaire data, a note of caution is necessary. As mentioned earlier, no statistics were calculated on these data due to missing data cells and treatment conditions. This makes conclusions regarding possible "effects" and "trends" uncertain. This analysis does not claim to assert universal truths on the basis of this data. Rather, general observations are drawn about several characteristics of the data set: large differences between groups, changes over time, and position on the scale with respect to the item (i. e. negative, neutral, or positive responses). Obviously these data would be insufficient if taken in isolation. However, they shall be used in conjunction with the other measures taken, providing what is intended to be a more complete picture of the video conferencing use.

## Interaction Analysis

The third major class of measures taken from the video taped sessions was the Interaction Analysis. Measures within this category include the communication breakdown analyses and the turntaking analyses. Each of the six types of communication breakdown will be discussed first, followed by the turntaking analysis.

### *Breakdown Analysis*

The first type of communication breakdown to be discussed is that of verbal turntaking. This measure was defined as "An incident in which two or more participants are talking at the same time, not including specific interruptions." In the literature, this is also referred to as “simultaneous speech”. Nearly all incidents of this type fell into two categories: when one person stopped speaking while the other continued, and when both continued to speak for longer than three seconds. For the video groups, between 69% and 88% of all verbal turntaking incidents involved the first type, when one person stopped speaking. Between 9% and 17% of these breakdowns were characterized by incidents in which both speakers continued to speak for longer than three seconds. This would not be surprising, given the way in which we converse in face-to-face settings, if not for the results from the face-to-face group. The face-to-face group had a more equal distribution of turntaking breakdown types: 52% were of the first type and 44% were of the later type. One explanation may lie in the small sample sizes used to take these measures; perhaps the face-to-face group was idiosyncratic. Alternatively, the co-located setting of this group may have permitted longer periods of simultaneous speech while avoiding the tendency to disengage a group member.

This measure also suggests that the face-to-face group had many more instances of simultaneous speech than the other groups (305 vs. 59-146, respectively). The face-to-face group had twice as many of these occurrences than the 25 fps group and nearly six times as many as the 1 fps group. One explanation for this finding may be that the audio system caused the conversation in the video groups to be more stilted. This explanation supports the findings of the interviews and the questionnaires in which it was indicated that speech was often stilted and participants often had difficulty in controlling the floor. Short utterances are much easier to make in a face-to-face setting where casual comments are not perceived as attempts to control the conversation. Such comments may have been lost in the video sessions.

Sellen (1992) compared conversation in different communication media and found that a higher percentage of time involved simultaneous speech for face-to-face conditions than video conditions. The present study supports this finding. Conversely, O'Conaill, Whittaker, and Wilbur (1993), found no difference in the total number of speaker overlaps between face-to-face speech, high quality video, and low quality video. These researchers did, however, find differences between these groups in terms of the type of simultaneous speech which occurred, a measure not taken in the present study. Doerry (1995) also measured the frequency of overlapping speech and concluded that no differences occurred between video-mediated communication and face-to-face communication. The quality of the audio channel may help account for the different conclusions made from these studies. In the present study, the audio was digitized and compressed full-duplex, resulting in lags and echoes. Sellen (1992) did not describe the audio system, but probably used analog, based on the description of the video system. Doerry (1995) used high-quality full-duplex audio. O'Connail, et al. (1993) used both half-duplex analog audio and high-quality full-duplex audio. It is likely that these researchers did not encounter lag and echo problems to the same magnitude as those experienced by participants in the present study.

The second type of breakdown was that of reference. The definition of a reference breakdown is "an incident in which the sharing of an object (papers, pencil, notes, mouse, etc.) becomes problematic. Very few examples of this breakdown were recorded. Such breakdowns are more likely to occur in situations which require the use of keyboard or mouse to interact with the system. Such was the case in Dorrey (1995) who found that more reference breakdowns occurred in video-mediated communication as compared to face-to-face communication. Although the whiteboard was controlled through these devices, the central activity was conversation, which did not rely upon external devices, unlike Doerry's study in which teams collaborated on a shared computer application.

The third class of breakdown, topic breakdowns, involved "incidents in which there is a failure to maintain topical orientation by one or more of the participants." Relatively few of these breakdowns occurred, and no discernible pattern appeared to characterize the four groups. This finding supports that of Doerry (1995) who found no significant differences between face-to-face groups and high-quality video-mediated communication in terms of topical breakdowns.

Visual breakdowns comprised the fourth class of communicative breakdown. Six types were classified, but the general definition was defined in Table 4.6 as "an incident in which visual communication suffers". Surprisingly few of these incidents occurred in the video sessions. No such incidents occurred in the face-to-face group. There were no requests for camera movements, or repositioning of objects to fill the camera field-of-view. This is surprising given the numerous complaints about the inability to see gesture and remote participants. It is possible that participants never grew fully comfortable with the system to adjust the camera or request that it be adjusted. Several times, however, requests were made for remote group members to shift seating so that the camera would encompass the group better.

Audio breakdowns consisted of "incidents in which a breakdown results due to loudness, feedback, or static difficulties, including requests for repeating due to these difficulties." Nearly all incidents of this breakdown class occurred in the video conditions. Between 85% and 87% of these breakdowns were of Type A: "requests for the another to repeat, and the other does so." Between 13% and 15% of these breakdowns were of Type B: "requests for the other to repeat, but the other does not do so." Because the majority of such breakdowns occurred in the video groups, it can be surmised that the audio/video system cause these breakdowns. Interestingly, most of the breakdowns were repaired, demonstrating that communication and the work process was not severely impacted as a result of the breakdown; it was simply less efficient.

The final breakdown class, shared conversations, were defined as "incidents in which a statement or question is directed at an adjacent (or co-located) participant with no attempt to include the rest of the group." Many more of these conversations occurred in the video conditions (63 - 94 for video, 38 for face-to-face). This finding indicates that the groups who were separated by the video conferencing system tended to have many more side conversations than the group who was co-located throughout the study. A look at the plot of the total number of side conversations over time (Figure 4.21) indicates that in the first session when all groups met face-to-face, there were very few side conversations. For the video groups, between 72% and 94% of these breakdowns occurred when just one side conversation took place. For the face-to-face group, 71% of these breakdowns took place when two side conversations was taking place. It is possible that remote groups who were cognizant of a side conversation at the other end sought to listen to that conversation, and hence, avoided speaking themselves. In the face-to-face group, side conversations may have evolved out of the chaotic nature of design team discussion, in which nearly all participants frequently spoke at the same time.

The analysis of communication breakdown provides an opportunity to measure directly the communication between the groups. This analysis has shown that few reference, topic, and visual breakdowns occurred in the study. It also indicated that more verbal turntaking breakdowns (i.e. simultaneous speech) occurred in the face-to-face group than the video groups, a finding which supports the work of Sellen (1992) but counters conclusions by Doerry (1995) and O'Conaill, et al. (1993). More audio breakdowns and shared conversations occurred in the video groups than the face-to-face group. While no statistics were conducted, making generalizations problematic, it is believed that this analysis has shed some light on the true nature of communicative efficacy between distributed and co-located groups. Furthermore, this analysis has built upon the work of Doerry (1995) by demonstrating other relevant classes of communication breakdown and providing another application by which this method has been successfully applied.

*Turntaking Analysis*

The other interaction analysis measure focused on the probability that a speaker would be followed by a co-located speaker or a remote speaker. A chi-square statistic was calculated for each session, based on the probabilities associated with each permutation of turntaking. Figure 4.24 depicts this model for the analysis. A significant chi-square suggests that the conversation flow was not balanced. For example, a high expected frequency compared with the observed frequency provides an indication that person A did not speak as often as expected. The results of this analysis are largely inconclusive. No pattern among the significant and non-significant chi-square statistics appeared to suggest a relationship between turntaking and group or time spent in the study. However, one note can be made on this subject. In the face-to-face group, 43% of the sessions resulted in significant chi-squares. In the 1 fps group, half the video sessions resulted in significant chi-squares. For the 10 fps group, only 17% of the video sessions resulted in significant chi-squares. Finally, the 25 fps group resulted in 33% significant chi-squares. This analysis may suggest that the 1 fps group had a more uneven conversation flow than the other two video groups.

O'Conaill, Whittaker, and Wilbur (1993) also sought to determine the distribution of turns between speakers in video and face-to-face communications. Their hypothesis was that poor video quality would result in a protocol in which a dominant speaker on each end of the video conference would emerge as the communication "manager" through whom most conversation would flow. To assess this notion, they compared the number of turns by the most dominant speakers in each communication condition. No significant differences were found, which, although using a different measure, supports the finding from this study. Still, this analysis is vague and inconclusive, due in part to the small sample sizes.

In addition to this analysis, the number of turns per person per minute was calculated for each session and group. A significant effect of condition was found and a post hoc test revealed the only significant difference was between the face-to-face group and the video groups. This finding suggests that face-to-face interactions have more speaker switches than video-mediated interactions. While the small sample size makes generalizing to other situations difficult, this finding both supports and refutes published reports on speaker turn frequency in different communication contexts. Sellen (1992) compared two different video-mediated communication systems with face-to-face communication and found no significant difference in terms of the number of speaker switches per unit of time. O'Conaill, Whittaker, and Wilbur (1993), however, found that more speaker switches tend to occur in both high-quality and low-quality video-mediated communication systems compared with face-to face communication. O'Malley, Langton, Anderson, Doherty-Sneddon, and Bruce (1996) found significantly more speaker switches in video-mediated systems compared to audio communication, suggesting that their video system (high-quality analog video) was capable of transmitting visual cues necessary for mediating conversation.

The present study supports these results. Explanations for the smaller number of turns in video-mediated systems - compared with face-to-face - may result from both the video and the audio characteristics of the system. The audio transmission system produced lags and echo which made verbal communication cumbersome and stilted. Participants often paused while the echo faded before interrupting a speaker or continuing their own statements. It is likely that this prevented speaker switches from occurring more often. It has also been shown that the visual channel is used in a variety of ways to take the floor and interrupt speakers (Duncan, 1972; Duncan and Niederehe, 1974; Rutter and Stephenson, 1977). In the video conditions, this channel was disrupted by restricted field-of-view, and at times, poor camera positioning - a negative artifact much worse than the high-quality video system used by O'Malley, et al. (1996).

## Ethnomethodology

The final major class of measures fell under the heading of ethnomethodology. As mentioned earlier, the format advocated by Hughes, et al. (1995a) was used to structure this analysis. Because the results section provided detailed findings from the sessions, this section addresses the major themes and discusses implications for desktop video conferencing use under varying frame rates over time.

One aspect of this analysis addressed the various activities which were conducted during the design sessions. It was found that the activities loosely matched those found by Olson, et al. (1992). Because only the first seven weeks of the design project were observed, the activities were heavily weighted by those which characterized initial design work. For example, much time was devoted to defining the scope of the project and discussing the work which was required. Only a small amount of time was devoted to performing actual work on the project (i.e. writing specifications documents), in part due to the organization of the work by the group. For example, most groups opted to divide the work such that individuals completed tasks outside the observed sessions. As with the Olson, et al. (1992) findings, overhead activities such as scheduling, and digression activities such as off-subject conversation were evident.

The major portion of this analysis sought to understand what artifacts, plans, and procedures were used to support the work activities. Several themes emerged from this analysis. First, it was found that different work organization strategies were utilized by the groups. The 1 fps group tended to use the session to complete specific work activities such as writing specification documents. In attempting to write a document collaboratively, the group switched tactics among various ineffective writing artifacts (note sharing, one-sided text editor, whiteboard). Each tactic altered the task, demonstrating a portion of the task-artifact cycle (Carroll, Kellogg, and Rosson, 1991) and the Activity Theory notion that developer's goals and values can influence a user's task (Kaptelinin, 1996). The other groups, however, used the session for more general discussion such as project scope and to report individually on the work activities which occurred since the last session. This finding had acute consequences for the overall success of the system. The 1 fps group, which was exposed to the worst quality video also used the most inefficient organization strategy. Additionally, this group had five group members, all of whom were present for the first two video sessions. As a result, three factors likely contributed to the low questionnaire ratings and interview responses by this group: group size, organization strategy, and low frame rate. The final two video sessions did improve for this group, as the group size shrunk and the organization strategy was altered. However, these changes were not sufficient to boost morale and keep the group in the study; the group dropped out of the study after four weeks in the video sessions.

Another major theme which was derived from this analysis was the need for a shared text editor. The major work activity involved writing text documents. As such, group members often took notes, referred to texts and notes, and created their own documents. The conferencing software supported a shared whiteboard, but this tool was not designed for extended text editing. Numerous communication difficulties would have been avoided if a shared text editor was available. For example, preventing disengagement from a group and maintaining topical orientation may have resulted from allowing all group members to view and participate in the reading or writing of a document.

Maintaining awareness of work activities was another major focus of this analysis. In the case of the video conferencing system, work awareness was often compromised by the relatively narrow field-of-view afforded by the video camera. This was overcome by unique positioning of the camera. For example, to show quickly a remote group which book or set of notes was being referenced, a group member would hold the object up to the camera. The face-to-face group had an easier time conveying such information as referenced documents and objects were often placed in the middle of the table, establishing a smooth and transparent discussion. The 1 fps group did not use the camera in this manner as often as the other two groups, and probably was less successful in doing so given the low frame rate.

This analysis demonstrated that video conferencing systems have implications for the type of work activities which can be conducted and the degree to which those activities are supported by the technology. It also highlighted some of the differences between face-to-face interaction and video-mediated interaction, as well as between the different frame rate conditions. Finally, this analysis showed how work activities can change over time and how added technological support can further enhance the work process.

## Comparison of Methods

Four major classes of research methods were used in this study. These methods are outlined in Table 4.16 according to the source of the information and the type of information. This section will seek to determine the relative usefulness of each method as well as the effort required to perform the analysis.

The contextual inquiry interviews provided several interesting pieces of information. Notions of audio quality perceptions, reactions to the use of video, suggestions for improving the system, and the tendency to accommodate or reject the system over time were discovered in this method. This information is considered highly valuable from a design and a scientific perspective and was relatively easy to gather, organize, and interpret. The disadvantage to this analysis is the questionable validity and reliability of information gained from the interviews. As mentioned earlier, it is difficult to determine whether a comment resulted from a true sentiment or was the result of a poorly-considered statement.

TABLE 4.16 Research Methods Used in the Ethnographic Study

| | *User-Generated* | *Observer-Generated* |
|---|---|---|
| *Quantitative* | Questionnaires | Interaction Analysis |
| *Qualitative* | Contextual Inquiry | Ethnomethodology |

The questionnaire data also contributed several compelling pieces of information. This method showed that perceptions of the various communication system components were relatively stable over the course of the study. The one exception to this finding was that audio quality perceptions tended to improve with time. Other information gleaned from this measure included differences among the groups in terms of video quality perception and transmission of visual cues, and the similarity of perceiving interruptions from speaker and listener perspectives. This information is useful for the advancement of scientific conclusions as well as feedback for the design of video conferencing systems, although some of the items were of questionable utility. Unfortunately the design of the study permitted different sample sizes and changes in sample size over time, precluding statistical analysis. However, the similarity of responses for like items indicated a degree of reliability helpful in drawing conclusions from the data.

The interaction analysis measures also provided interesting information regarding the use of video as a communication medium. For example, it was found that few reference, topic, or visual communication breakdowns occurred in any of the conditions. Likewise, there was more simultaneous speech and speaker turns in the face-to-face conditions. Finally, there were more shared conversation breakdowns and audio breakdowns in the video groups. These breakdown measures are useful in that they further develop our understanding of video-mediated communication and suggest mechanisms by which packet-switched communication modes can be improved. Of lesser utility was the turntaking analysis. While counting the number of turns per session was useful and interesting, seeking to quantify different communication patterns among the different sub-groups proved inconclusive. Given the utility this data, it is probably not worth the large effort required to collect it.

The ethnomethodology analysis also contributed some interesting behavioral insight. For example, the structure and organization of the work was determined, the degree to which the communication system supported this work was assessed, and the ability to maintain work awareness through the system was judged. This sort of information would have been difficult to extract from the other methods, yet enhanced the overall understanding of how the system was used. While this data is not necessarily repeatable or generalizable, it does assist in understanding the findings from the other methods. For example, the rationale for the 1 fps group dropping out of the study was only partially articulated in the interviews and questionnaire (nothing in the interaction analysis helped explain this event). By reviewing the videotapes of these sessions and learning how the group organized its work activities, one could gain more insight regarding the source of this group's frustrations with the communication medium.

One ethnographer has sought to incorporate the perceived benefits of Activity Theory to further HCI study (Nardi, 1996). She calls for the need to use long-term studies, examine complex patterns of activity, use a variety of data collection techniques, and to understand HCI from the user's point-of-view. This study sought to address all four of these concerns.

After considering the quality, usefulness, and cost-effectiveness of the various evaluation methods, it is clear that each contributed both unique and redundant information to the overall analysis. Any one of these methods in isolation may have led to somewhat different conclusions or emphases in evaluating the video conferencing system. By looking at the events of interest from several different points of view, a much more thorough and accurate understanding is gained. Applying multiple research methods to complex HCI problems (i. e. multi-user interfaces and new communication technologies) is becoming more common in the literature (Sellen, 1992; O' Conaill, et al., 1993; Watts, et al., 1996) as research domains addressing cognitive issues are being combined with those that address social issues. It is this broadening of study that Kaptelinin (1996) suggests is necessary for HCI to make a more effective contribution to the design of usable systems.

While this study has shown how various ethnographic research methods can provide valuable information in terms of design, evaluation, and scientific merit, it is not suggested that human factors practitioners and researchers abandon the traditional paradigm of controlled experimental designs. As the literature review section indicates, a tremendous body of useful contributions have been made through such research designs. It is hoped that considering ethnographic approaches in combination with rigorously controlled designs, we can gain a more complete understanding of human interaction with artificial systems.

## Research Contribution and Implications

In addition to the contributions outlined in the previous section, this study also contributed to the understanding of computer-supported cooperative work systems. McGrath (1991) presents a critique of prior research on working groups. Table 4.17 depicts the specific problems with past research and indicates which areas the present research has contributed.

TABLE 4.17 McGrath's (1995) Critique of Group Research and Contribution of this Research

| *Criticisms of Prior Group Research* | *Contribution of This Research* |
|---|---|
| Limited range of group types | |
| Highly controlled experimental conditions | X |
| Very small groups (2-4) | |
| Membership arbitrarily assigned by experimenter | X |
| Limited lifespan | X |
| No past or future as a group | X |
| Groups not embedded in a larger social organization or community | X |
| Performance of single, simple tasks | X |
| Tasks not in inherent activity of the group | X |
| Context-free conditions | X |

Research often poses more questions than answers, and this study is no exception. Due to the limitations of the research design, many of the findings from this study do now allow generalizations to be extended to similar situations. It is recommended that controlled experimental designs be used to study a number of issues which were raised from this study.

The question of whether or not users habituate or grow intolerant of poor video quality over time remains to be conclusively studied. This study suggested that users of poor video quality grow less tolerant with time and that high-quality allows users to acclimate. A more thorough investigation of these variables should be conducted before firm conclusions are drawn.

The interviews and ethnomethodology suggested that the communication system lacked specific features to support the work at hand, namely, a shared text editor. While this finding was fairly conclusive, more research is needed to address ways in which such collaborative distributed systems can be integrated with the communication modality. The issue of integrating video and a text editor should be investigated such that an understanding of how best to use video without the distraction factor can be determined.

The interaction analysis of this study provided interesting observations regarding the frequency and type of communication breakdown, as well as the methods for repair. While other studies have examined this issue (Sellen, 1992, O'Conaill, et al., 1993, Doerry, 1995), some factors remain inconclusive. For example, there is still disagreement concerning the impact audio and video systems have on the frequency and duration of simultaneous speech. A more rigorous study examining this issue, as well as its role in conversation fluidity would be beneficial to understanding how these systems impact speech.

Finally, the ethnomethodology showed that some groups - namely the 1 fps group - organized their work activity in a way which was least supported by the communication technology. Although Olson, et al. (1992) conducted a thorough analysis of design team task activities, more research is needed to understand how various communication media support or detract from these various tasks.

# Chapter 5.  Conclusion

This research used two distinct methods to help understand the human factors issues of degraded image quality in desktop video conferencing. Networked digital video can be degraded in a variety of manners, but the issue of reduced frame rate was examined in these studies. The first study used two psychophysical methods: free-modulus magnitude estimation and category scaling. The second study used four ethnographic methods: questionnaires, contextual inquiry, interaction analysis, and ethnomethodology.

The psychophysical study examined the relationship between frame rate and perceived video image quality for a number of different scenes. Magnitude estimation and category scales enabled a psychophysical function to be calculated for a ratio scale. Subsequent analyses showed that the scenes were perceived in significantly different ways across a range of frame rates. For scenes depicting a static image which was filmed at varying speeds and which included varying levels of detail, functions were characterized by linear and quadratic trends. These scenes produced an asymptote at approximately 10 fps. The implication for this finding is that similar applications (i.e. real estate walkthroughs) can be conducted effectively at this frame rate. For a scene in which a traditional person-to-person conference was simulated, the resulting function was quite different. This curve produced no asymptote, and was characterized by only a linear trend. The implication for this finding is that traditional desktop video conferencing applications may benefit from frame rates as high as 28 fps.

The second study used ethnographic methods to examine another class of questions relating to frame rate degradations in video-mediated communication. This study sought to address the issues of long-term use, the impact of contextual forces such as realistic groups and tasks, and the differences between video and face-to-face communication in terms of direct conversation measures. Contextual interviews collected user perceptions of audio quality, reactions to communicating through a video channel, suggestions for improving the user interface, and the tendency for the high-quality video groups (10 fps and 25 fps) to acclimate to video-mediated communication. Questionnaire data showed little change in perceptions of conversation fluidity, video quality, or physical environment over time. Perceptions of audio quality did, however, appear to improve slightly with time. The interaction analysis provided information on the frequency and type of communication breakdowns which occurred, the finding that more simultaneous speech and speaker turns occurred in the face-to-face condition, and the finding that there were more audio and shared conversation breakdowns in the video conditions. Finally, ethnomethodology enabled the work structure and organization of the groups to be understood better with respect to technology support. It was found that the 1 fps group organized its work differently than the other groups and that this activity was not well-supported by the communication technology. This analysis also provided some insight regarding the differences between low and high quality video communication and face-to-face communication for the awareness of work between group members.

Based on the findings from the two studies, a number of design guidelines are suggested for the implementation and use of desktop video conferencing systems. These guidelines are displayed according to primary system components in Figure 5.1. None of these suggestions reflect absolute truths, and as a result, further research is necessary and contextual and organizational factors must always be considered carefully prior to the implementation of a new communication system. Other reports have also suggested design guidelines for desktop video conferencing, and those should be consulted as well (Fussell and Benimoff, 1995; Kies, Williges, and Williges, 1997)

**Interface**
- Provide re-sizable video windows and ability to turn video off.
- Allow users to control audio and video characteristics (i.e. speaker volume, frame rate).

**Workspace Layout and Support Tools**
- More than one person per conferencing unit is too many, unless a wide-angle lens is used.
- Consider tendency for side conversations if more than one person per conferencing unit.
- Provide space for documents and other materials.
- Allow users to control camera positioning.
- Provide software applications to support the task (i.e. group text editor if writing documents; shared whiteboard, if making illustrations).

**Video Quality**
- Traditional conferencing applications should use the highest frame rate supported by the network, greater than 10 fps if possible.
- Non-traditional conferencing applications not depicting people speaking need not use a frame rate greater than 10 fps.
- Poor quality video (i.e. 1 fps) may be distracting, rather than helpful.
- Field-of-view may not transmit gesture.
- Image resolution may not transmit facial expressions.

**Audio Quality**
- Use high quality audio if possible; it is the most crucial channel for communication.
- Systems with digital full-duplex audio may result in more stilted conversation (less simultaneous speech, fewer speaker turns).

**Social and Organizational Issues**
- Account for tendency to grow tired of low-quality video with time.
- Account for break-in time; some users may require time to become comfortable communicating through a video link.
- Task structure may interfere with communication system (i.e. demands of task may necessitate face-to-face communication).

FIGURE 5.1 Design Guidelines for Desktop Video Conferencing

This research sought to help clarify the use of different design and research methods by employing two which have been neglected in the published literature on the use of desktop video conferencing. The need for this research was stated in a comprehensive review of the literature on this technology. The literature review also established the need for this research by describing how office environments are currently undergoing a dramatic paradigm shift which may require a variety of research methods for the successful design, evaluation, and implementation of CSCW technology. This research sought to clarify methodology and in the process, contribute to the understanding of video-mediated communication.

Each of the methods used in this study required different levels of resources (i. e. time, cost, etc.) and had different characteristics with respect to McGrath's (1995) research criteria: generalizability, precision, and realism. Table 5.1 depicts the research methods used in this study and their degree of generalizability, precision, and realism. McGrath suggests using a variety of methods and selecting them so that their characteristics are complimentary. As shown in Table 5.1, traditional experimental designs and ethnographic methods achieve this qualification.

TABLE 5.1  Critical Research Method Dimensions

| *Research Dimensions* | *Traditional Experimental Designs* | *Ethnography* | | | |
|---|---|---|---|---|---|
| | | *Contextual Inquiry* | *Questionnaire* | *Interaction Analysis* | *Ethno-methodology* |
| Generalizability | Depends | Low | Low | Low | Low |
| Precision | High | Low | Medium | Medium | Low |
| Realism | Low | High | Medium | High | High |
| Cost | Depends | Low | Depends | High | Depends |

From Table 5.1, it can be seen that ethnographic approaches offer little generalizability or precision, while experimental designs such as the one used in the first study, provide a high degree of generalizability and precision.  In terms of realism, the opposite trend holds true: experimental designs offer a low degree of realism, while ethnographic approaches are characterized by a high degree of realism.  The fourth dimension, cost, depends greatly on extraneous factors.  In the present study, magnitude estimation data were relatively inexpensive to collect (2.5 hours per subject for 30 subjects), and some researchers have claimed that only 10 subjects are required to collect sufficient data (Stevens, 1975).  More advanced factorial designs, however, can require numerous subjects over many conditions, creating an expensive research program.  Contextual inquiry and questionnaire data are relatively inexpensive research approaches.  Contextual inquiry may require repeat visits with users over an extended period of time, but few subjects are required.  The efficiency of questionnaire data depends greatly on the means by which the survey was created.  If proper factor analytic methods are used to generate questionnaire items from a larger sample, the cost and effort can be enormous.  Most human factors research, however, does not use such formal methods, and therefore keeps costs low.  The collection of interaction analysis data requires videotapes to be made of human-computer interaction.  These tapes are then painstakingly analyzed by paid observers who code pre-defined behaviors.  These data can be used in a traditional experimental design or as part of an ethnographic study.  In either case, the expense and time required to train observers, generate the data, and calculate reliability, is large.  It is recommended that unless specific hypotheses are being tested in a formal fashion, this research tool not be used.  Finally, ethnomethodology also requires a fairly high cost.  Unlike interaction analysis, this method can be performed in real time at a low cost (i.e. "quick and dirty ethnography", Rouncefield, Hughes, Rodden, and Viller, 1994), or by meticulously reviewing videotapes individually and with other researchers.

Human factors engineering activities encompass two major objectives:  scientific advancement and system design.  Scientific advancement involves applied behavioral research problems which are solved through the scientific method.  System design incorporates user-centered design principles at all stages of the development process.  Table 5.2 outlines each of the ethnographic methods used in this study, as well as the experimental method, with respect to the four main stages of scientific endeavor as discussed by Yaremko, Harari, Harrison, and Lynn (1982).  The experimental method includes traditional formal research approaches, including psychophysics as it was applied in this study.  Table 5.2 shows that most ethnographic methods can be applied in the first two stages of scientific endeavor: observing phenomena and formulating hypotheses. This was born out by the second study in which ethnographic methods produced a large set of data and subsequent hypotheses. Experimental designs may best be applied in the third and fourth stages: conducting experiments and re-testing hypotheses. Science is based on the ability to control and manipulate variables, activities addressed only by traditional experimental approaches. Ethnographic methods are patently inappropriate for these later stages. However, they serve highly useful roles when seeking to observe behaviors and formulate explanations. Perhaps they are overlooked as tools for uncovering interesting areas for which more traditional experimental approaches can be applied.

TABLE 5.2 Scientific Activities for Various Research Methods

| *Stages* | *Traditional Experimental Designs* | *Ethnography* | | | |
|---|---|---|---|---|---|
| | | *Contextual Inquiry* | *Questionnaire* | *Interaction Analysis* | *Ethno-methodology* |
| Observing a Phenomenon | Pilot Studies, Previous Research | Collect initial user data, in situ | Collect initial user data, in situ | Collect initial user data, in situ | Collect initial user data, in situ |
| Formulating Tentative Explanations | Develop hypothesis, Define variables of interest | Develop hypothesis, Define variables of interest | Develop hypothesis, Define variables of interest | Develop hypothesis, Define variables of interest | Develop hypothesis, Define variables of interest |
| Further Observing and Experimenting | Design and Execute a Formal Experiment | Not Appropriate | Supplemental Data | Primary or Supplemental Data | Not Appropriate |
| Revising and Re-testing Explanations | Follow-up Experiment, Replicate Experiment, Refine hypothesis | Support/discredit findings | Support/discredit findings | Primary or Supplemental Data | Support/discredit findings |

The other primary activity of human factors engineering is that of system design. Many design and development cycles have been proposed (the waterfall method, Royce, 1970; the spiral model, Boehm, 1988; the star lifecycle, Hartson and Hix, 1989; the usability engineering lifecycle model, Nielsen, 1993). Although the system development process is highly complex and mercurial and all these cycles have merit, general stages and activities can be outlined. Table 5.3 adapts the primary components of Hartson and Hix's (1989) model for the purpose of discussing the value of each research method at each stage. Traditional experimental designs are most suited to formative and summative evaluation in the context of usability testing. The ethnographic methods can serve as means to collect data on users, their tasks, and their working environments. They are also appropriate for conducting formative and summative evaluation. The choice of method depends on its applicability at each development stage, as well as the objectives of the designer, the nature of the project, and the resources available.

TABLE 5.3  System Development Activities for Various Research Methods

| *Stages* | *Traditional Experimental Designs* | *Ethnography* | | | |
|---|---|---|---|---|---|
| | | *Contextual Inquiry* | *Questionnaire* | *Interaction Analysis* | *Ethno-methodology* |
| System/task/ functional/user Analysis | Review previous research | User, needs, task analysis | User, needs, task analysis | Review previous research | User analysis, task analysis |
| Requirements/ Usability Specifications | Not appropriate | Inform specifications, review with users | Inform specifications, review with users | Not Appropriate | Not appropriate |
| Design and Design Representation | Not appropriate | Participatory design activities | Not appropriate | Not appropriate | Not appropriate |
| Rapid Prototyping | Usability testing, formative evaluation | Usability testing, formative evaluation | Usability testing, formative evaluation | Not appropriate | Contextual evaluation on advanced prototype |
| Software and System Production | Not appropriate | User-designer communication | Not appropriate | Not appropriate | Not appropriate |
| Deployment | Summative evaluation | Summative evaluation | Supplemental analysis for summative evaluation | Summative evaluation | Summative evaluation |

Although this research may have raised more questions than it answered, it is hoped that human factors researchers and practitioners will have a more thorough understanding of the new problems which collaborative technologies pose.  It is also anticipated that the need to consider a variety of research and design methods in order to assess and study the interaction between groups and technology will be more salient.  Through the literature review, this study was grounded by questioning the reliance of traditional experimental designs to solve all human factors problems.  However, it also cautions against the overuse and limitations inherent in a more contextual approach.  The most accurate picture of the interactions between humans and technology will only be obtained if several methods are used intelligently and appropriately in conjunction with each other, and in accordance with the research and design objectives of the human factors practitioner.

# References

Abacus Concepts (1989) *SuperANOVA Manual.* Berkeley, CA: Abacus Concepts, Inc.

Anderson, A. H., Newlands, A., Mullin, J., Fleming, A. M., Doherty-Sneddon, G., and Van der Velden, J. (1996) Impact of video-mediated communication on simulated service encounters. *Interacting with Computers*, 8 (2), 193-206.

Apteker, R. T., Fisher, J. A., Kisimov, V. S., and Neishols, H. (1994). Distributed multimedia: user perception and dynamic QoS. In *Proceedings of SPIE,* 2188, 226-234. Bellingham, WA: The International Society for Optical Engineering.

Argyle, M. (1975). *Bodily communication.* New York, NY: International Universities Press.

Argyle, M. and Cook, M. (1976). *Gaze and mutual gaze.* Cambridge, U. K.: Cambridge University Press.

Bales, R. F. (1950) *Interaction process analysis: a method for the study of small groups.* Cambridge, MA: Addison-Wesley.

Bannon, L. (1991) From human factors to human actors: the role of psychology and human-computer interaction studies in system design. In Greenbaum, J., and Kyng, M. (Eds.) *Design at work.* Hillsdale, NJ: Lawrence Erlbaum & Associates.

Beaton, R. J. (1984) *A human-performance based evaluation of quality metrics for hard-copy and soft-copy digital imaging systems.* Unpublished doctoral dissertation. Blacksburg, VA: Virginia Polytechnic Institute and State University.

Bentley, R., Hughes, J., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., and Sommerville, I. (1992) Ethnographically informed system design for air traffic control. In *Proceedings of CSCW '92: Computer-Supported Cooperative Work,* (pp.123-129). New York, NY: Association for Computing Machinery.

Bias, R. G. and Mayhew, D. J. (1994) *Cost-justifying usability.* New York, NY: Academic Press.

Böcker, M. and Mühlbach, L. (1993). Communicative presence in video communications. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting,* (pp. 249-253). Santa Barbara, CA: Human Factors and Ergonomics Society.

Bødker, S., Grønbæk, K., and Kyng, M. (1993) Cooperative design: Techniques and experiences from the scandinavian scene. In Schuler, D. and Namioka, A. (Eds.) *Participatory design: Principles and practices.* Hillsdale, NJ: Lawrence Erlbaum & Associates.

Boehm, B. W. (1988) A spiral model of software development and enhancement. *IEEE Computer,* 21(2), 61-72.

Bordens, K. S. and Abbott, B. B. (1991) *Research design and methods: A process approach, 2nd Edition.* Mountain View, CA: Mayfield Publishing Company.

Boring, A. and Travers, M. (1991) Two approaches to casual interaction over computer and video networks. In *Proceedings of CHI '91: Human Factors in Computing Systems.* (pp. 13-19). New York, NY: Association for Computing Machinery.

Bregman, A. S. and Mills, M. I. (1982) Perceived movement: the Flintstone constraint. *Perception*, 11, 201-206.

Bruce, V.  (1996)  The role of the face in communication implications for videophone design.  *Interacting with Computers*, 8 (2), 166-176.

Buxton, W. A. S.  (1992)  Telepresence:  Integrating shared task and person spaces.  In *Proceedings of Graphic Interface '92* (pp.123-129)  San Francisco, CA:  Morgan Kaufman Publishers.

Card, S. K., Moran, T. P., and Newell, A.  (1983)  *The psychology of human-computer interaction.*  Hillsdale, NJ:  Lawrence Erlbaum & Associates.

Carroll, J. M.  (1997)  Human-computer interaction:  Psychology as a science of design.  *Annual Review of Psychology, Volume 48.*  Palo Alto, CA:  Annual Reviews.

Carroll, J. M., Ed.  (1995)  *Scenario-based design:  Envisioning work and technology in system design.*  New York, NY:  John Wiley & Sons.

Carroll, J. M., Ed.  (1991)  *Designing interaction:  Psychology at the human-computer interface.*  Cambridge, UK:  Cambridge University Press.

Carroll, J. M., Kellogg, W. A., and Rosson, M. B.  (1991)  The task-artifact cycle.  In Carroll, J. M. (Ed.)  *Designing Interaction:  Psychology at the Human-Computer Interface.*  New York, NY:  Cambridge University Press.

Carroll, J. M., & Rosson, M. B. (1990). Human-computer interaction scenarios as a design representation. In *Proceedings of HICSS-23: Hawaii International Conference on System Sciences*,  (pp. 555-561). Los Alamitos, CA: IEEE Computer Society Press.

Chapanis, A.  (1959)  *Research techniques in human engineering.*  Baltimore, MD:  The Johns Hopkins Press.

Chapanis, A.  (1971)  Prelude to 2001:  Explorations in human communication.  *American Psychologist*, 26, 949-961,

Chapanis, A. (1975). Interactive human communication. *Scientific American*, 232(*3*), 36-42.

Christel, M. G.  (1991)  *A comparative evaluation of digital video interactive interfaces in the delivery of a code inspection course.*  Unpublished Doctoral Dissertation.  Atlanta, GA:  Georgia Institute of Technology.

Chuang, S. L. and Haines, R. F. (1993).  A study of video frame rate on the perception of compressed dynamic imagery. In *Society for Information Display '93 Digest*.  (pp.  524-527).

Conrath, D. W., Dunn, E. V., Bloor, W. G., and Tranquada, B.  (1977)  A clinical evaluation of four alternative telemedicine systems.  *Behavioral Science*, 22, 12-21.

Cook, M. and Lalljee, M. G. (1972). Verbal substitutes for visual signals in interaction.  *Semiotics*, 3, 221-221.

Cool, C., Fish, R. S., Kraut, R. E., and Lowery, C. M. (1992)  Iterative design of video communication systems.  In *CSCW '92:  Computer-Supported Cooperative Work.*  (pp. 25-32)  New York, NY:  Association for Computing Machinery.

Daft, R. L., Lengel, R. H.  (1986)  Organizational information requirements, media richness, and structural design.  *Management Science*, 32(5), 554-571.

Deghuee, B. J.  (1980)  *Operator-adjustable frame rate, resolution, and gray scale tradeoff in fixed-bandwidth remote manipulator control.*  Unpublished Master's Thesis.  Cambridge, MA:  Massachusetts Institute of Technology.

Doerry, E. (1995) *An empirical comparison of copresent and technologically-mediated interaction based on communicative breakdown.* Unpublished Doctoral Dissertation. Eugene, OR: University of Oregon.

Dooley, D. (1995) *Social research methods, 3rd Edition.* Englewood Cliffs, NJ: Prentice Hall.

Dourish, P. (1993) Culture and control in a media space. In *Proceedings of the European Conference on CSCW.*

Dourish, P., Adler, A., Bellotti, V., and Henderson, A. (1994) *Your place or mine? Learning from long-term use of video communication.* Technical Report EPC-94-105. Cambridge, UK: Rank Xerox EuroPARC.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(*2*), 283-292.

Duncan, S. and Niederehe, G. (1974). On signaling that it's your turn to speak. *Journal of Experimental Social Psychology*, 10, 234-247.

Duncanson, J. P. and Williams, A. D. (1973). Video conferencing: reactions of users. *Human Factors*, 15(*5*), 471-485.

Engen, T. (1971) Psychophysics II. scaling methods. In Kling, J. W. and Riggs, L. A. (Eds.) *Woodworth & Schlosberg's Experimental Psychology, Third Edition.* New York, NY: Holt, Rinehart, and Winston.

Fafchamps, D. (1991) Ethnographic workflow analysis: Specifications for design. In Bullinger, H-J. (Ed.) *Human aspects in computing: Design and use of interactive systems and work with terminals.* (pp. 709-715). Elsevier.

Fish, R. S., Kraut, R. E., and Chalfonte, B. L. (1990) The videowindow system in informal communications. In *Proceedings of CSCW '90: Computer-Supported Cooperative Work.* (pp.1-11). New York, NY: Association for Computing Machinery.

Fish, R. S., Kraut, R. E., Root, R. W. and Rice, R. E. (1992). Evaluating Video as a Technology for Informal Communication. In *CHI '92: Human Factors in Computing Systems* (pp. 37-48). New York, NY: Association for Computing Machinery.

Folley, J. D., Ed. (1960) *Human factors methods for system design.* Washington, DC: American Institute for Research.

Fulk, J., Steinfeld, C. W., Schmitz, J., and Power, J. G. (1987) A social information processing model of media use in organizations. *Communication Research*, 14(5), 529-552.

Fussell, S. R. and Benimoff, N. I. (1995) Social and cognitive processes in interpersonal communication: Implications for advanced telecommunications technologies. *Human Factors*, 37(2), 228-250.

Gale, S. (1990) Human aspects of interactive multimedia communication. *Interacting with Computers*, 2 (2), 175-189.

Gale, S. (1991) Adding audio and video to an office environment. In Bowers, J. M. and Benford, S. D. (Eds.) *Studies in Computer Supported Cooperative Work.* New York, NY: Elsevier Science Publishing Company, Inc.

Gale, S. (1992) Desktop video conferencing: Technical advances and evaluation issues. *Computer Communications*, 15 (8), 517-526.

Garfinkle, H. (1967) *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Gaver, W. W. (1992) The affordances of media spaces for collaboration. In *CSCW '92: Computer Supported Cooperative Work.* New York, NY: The Association for Computing Machinery.

Gaver, W., Moran, T., MacLean, A., Lovstrand, L., Dourish, P., Carter, K. and Buxton, W. (1992). Realizing a Video Environment: Europarc's RAVE System. In *Proceedings of CHI '92: Human Factors in Computing Systems* (pp. 27-35). New York, NY: Association for Computing Machinery.

Gaver, W., Sellen, A., Heath, C. and Luff, P. (1993). One is not Enough: Multiple Views in a Media Space. In *INTERCHI '93* (pp. 335-341). New York, NY: Association for Computing Machinery.

Gescheider, G. A. (1985) *Psychophysics: Method, Theory, and Application.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gibson, J. J. (1979) *The Ecological Approach to Visual Perception.* Hillsdale, NJ: Lawrence Erlbaum Associates. Inc.

Green, C. and Williges, R. C. (1995). Evaluation of alternative media used with a groupware editor in a simulated telecommunications environment. *Human Factors*, 37(*2*), 283-289.

Greenbaum, J. and Kyng, M., Eds. (1991). *Design at work: Cooperative design of computer systems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Grudin, J. (1990) Groupware and cooperative work: Problems and prospects. In Laurel, B. (Ed.) *The art of human-computer interaction.* Reading, MA: Addison-Wesley

Hammersley, M. and Atkinson, p. (1995) *Ethnography: Principles in practice*. New York, NY: Routledge.

Harrison, R. P. (1974). *Beyond words: an introduction to nonverbal communication*. Englewood Cliffs, NJ: Prentice-Hall.

Hartson, H. R. and Hix, D. (1989) Toward empirically derived methodologies and tools for human-computer interface development. *International Journal of Man-Machine Studies*, 31, 477-494.

Heath, C. and Luff, P. (1991). Disembodied conduct: communication through video in a multi-media office environment. In *Proceedings of CHI '91: Human Factors in Computing Systems* (pp. 99-103) New York, NY: Association for Computing Machinery.

Heath, C. and Luff, P. (1992) Media space and communicative asymmetries: preliminary observations of video-mediated interaction. *Human-Computer Interaction*, 7, 315-346.

Hochberg, J. (1986) Representation of motion and space in video and cinematic displays. In Boff, K. R., Kauffman, L., and Thomas, J. P. (Eds.) *Handbook of Perception and Human Performance, Vol. 1*. New York, NY: John Wiley & Sons.

Hollan, J. and Stornetta, S. (1992) Beyond being there. In *Proceedings of CHI '92: Human Factors in Computing Systems* (pp. 119-125) New York, NY: Association for Computing Machinery.

Holtzblatt, K. and Jones, S. (1993) Conducting and analyzing a contextual interview. In Schuler, D. and Namioka, A. (Eds.) *Participatory design: Principles and practices.* Hillsdale, NJ: Lawrence Erlbaum & Associates.

Hopkins, C. O., Snyder, H. L., Price, H. E., Hornick, R. J., Mackie, R. R., Smillie, R. J., and Sugarman, R. C. (1982) *Critical human factors issues in nuclear power regulation and a recommended comprehensive human factors long-range plan.* (NUREG/CR-2833 Vol.2) Washington, DC: U.S. Nuclear Regulatory Commission.

Howell, D. C. (1992) *Statistical methods for psychology, Third Edition.* Belmont, CA: Duxbury Press.

Hughes, J. A., King, V., Rodden, T., and Andersen, H. (1994) Moving out from the control room: Ethnography in system design. In *CSCW '94: Computer Supported Cooperative Work.* (pp. 429-439) New York, NY: Association for Computing Machinery.

Hughes, J. A., Sharrock, W., Rodden, T., Kristoffersen, S., O'Brien, J., Rouncefield, M, and Calvey, D. (1995a) *CSCW Requirements Development.* COMIC Technical Report D2.4. Lancaster, UK: Lancaster University.

Hughes, J. A., O'Brien, J., Rouncefield, M. and Sharrock, W. (1995b) An Ethnography Handbook. In *CSCW Requirements Development.* COMIC Technical Report D2.4. Lancaster, UK: Lancaster University.

Isaacs, E. A. and Tang, J. C. (1994). What video can and cannot do for collaboration: a case study. *Multimedia Systems*, 2, 63-73.

Ishii, H., Kobayashi, M., and Grudin, J. (1992) Integration of inter-personal space and shared workspace: Clearboard design and experiments. In *Proceedings of CSCW '92: Computer Supported Cooperative Work* (pp. 33-42). New York, NY: Association for Computing Machinery.

Ishii, H., and Miyake, N. (1994) Multimedia groupware: Computer and video fusion approach to open shared workspace. In Buford, J. F. K. (Ed.) *Multimedia Systems.* New York, NY: ACM Press.

Jones, B. L. and Marks, L. E. (1985) Picture quality assessment: A comparison of ratio and ordinal scales. *SMPTE Journal*, Dec., 1244-1248.

Jordan, B. and Henderson, A. (1995) Interaction analysis: foundations and practice. *The Journal of the Learning Sciences*, 4(1), 39-103.

Kaptelinin, V. (1996) Computer-mediated activity: Functional organs in social and developmental contexts. In Nardi, B. A. (Ed.) *Context and Consciousness: Activity Theory and Human-Computer Interaction.* Cambridge, MA: The MIT Press.

Keppel, G. (1991). *Design and analysis: A researcher's handbook.* (3rd Ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kies, J. K., Kelso, J., and Williges, R. C. (1995) The use of scenarios to evaluate the effects of group configuration and task on video-teleconferencing communication effectiveness. In *Proceedings of the Third Annual Mid-Atlantic Human Factors Conference*. (pp. 22-28) Blacksburg, VA.

Kies, J. K., Williges, R. C., and Rosson, M. B. (in press, a) Studying computer-supported cooperative work: A review of research issues and strategies. *Journal of the American Society for Information Science.*

Kies, J. K., Williges, R. C., and Rosson, M. B. (in press, b) Evaluating desktop video conferencing for distance learning applications. *Computers and Education.*

Kies, J. K., Williges, R. C., and Williges, B. A. (1997) Desktop video conferencing: A systems approach. In Helander, M. G., Landauer, T. K., and Prabhu, P. (Eds.) *Handbook of Human-Computer Interaction, 2nd Edition.* Amsterdam, The Netherlands: Elsevier Science.

Kiran, B. and Ainsworth, L. K. (1992) *A guide to task analysis.* London, UK: Taylor and Francis.

Kling, R. (1995) Controversies about computerization and the organization of white collar work. In Baecker, R. M., Grudin, J., Buxton, W. A. S., and Greenberg, S. (Eds.) *Readings in Human-Computer Interaction: Toward the Year 2000.* San Francisco, CA: Morgan Kaufmann Publishers.

Konz, S. (1990) *Work design: Industrial ergonomics.* Scottsdale, AZ: Publishing Horizons, Inc.

Lewis, D. and Burke, C. J. (1949) The use and misuse of the chi-square test. *Psychological Bulletin,* 46, 433-489.

Lewis, D. and Burke, C. J. (1950) Further discussion of the use and misuse of the chi-square test. *Psychological Bulletin,* 47, 347-355.

Luther, A. C. (1994) Digital video and image compression. In Kogel-Buford, J. F. (Ed.) *Multimedia Systems.* New York, NY: Addison-Wesley Publishing Company.

McGrath, J. E. (1984). *Groups: interaction and performance*. Englewood, NJ: Prentice-Hall.

McGrath, J. E. (1995) Methodology matters: Doing research in the behavioral and social sciences. In Baecker, R. M., Grudin, J., Buxton, W. A. S., and Greenberg, S. (Eds.) *Readings in human-computer interaction: Toward the year 2000.* San Francisco, CA: Morgan Kaufman.

Mantei, M. M. (1991) Adoption patterns for media space technology in a university research environment. In *Friend '21*, Tokyo, Japan.

Mantei, M. M., Baecker, R. M., Sellen, A. J., Buxton, W. A. S., Milligan, T., and Wellman, B. (1991). Experiences in the Use of a Media Space. In *CHI '94: Human Factors in Computing Systems* (pp. 203-208). Association for Computing Machinery.

Martin, M. M.., Williges, B.H., and Williges, R.C. (1990). Improving the design of telephone-based information systems. In *Proceedings of the Human Factors Society 34th Annual Meeting,* (pp. 198-202) Santa Monica: Human Factors Society.

Masoodian, M, Apperley, M., and Fredrickson, L. (1995) Video support for shared work-space interaction: an empirical study. *Interacting with Computers,* 7(3), 237-253.

Meister, D. (1985) *Behavioral Analysis and Measurement Methods.* New York, NY: John Wiley and Sons, Inc.

Meister, D. (1986) *Human factors testing and evaluation.* Amsterdam, The Netherlands: Elsevier.

Meister, D. (1991) *Psychology of system design.* Amsterdam: The Netherlands: Elsevier.

Miller, M. E. (1993) The effect of display characteristics on perception of motion from remotely sensed imagery. In *SID Digest* (pp. 118-119). Society for Information Display.

Moran, T. P. and Carroll, J. M., Eds. (1996) *Design rationale: Concepts, techniques, and use.* Mahwah, NJ: Lawrence Erlbaum Associates.

Morrison, H. B. (1995) *The influence of depth and stutter on consumer preference for static three-dimensional lenticular-sheet images.* Unpublished Doctoral Dissertation. Blacksburg, VA: Virginia Polytechnic Institute and State University.

Mühlbach, L., Böcker, M, and Prussog, A. (1995) Telepresence in videocommunications: a study on stereoscopy and individual eye contact. *Human Factors,* 37 (2), 290-305.

Nardi, B. A. (1996a) Activity theory and human-computer interaction. In Nardi, B. A. (Ed.) *Context and Consciousness: Activity Theory and Human-Computer Interaction.* Cambridge, MA: The MIT Press.

Nardi, B. A. (1996b) Studying context: A comparison of activity theory, situated action models, and distributed cognition. In Nardi, B. A. (Ed.) *Context and Consciousness: Activity Theory and Human-Computer Interaction.* Cambridge, MA: The MIT Press.

Nardi, B. and Miller, J. (1991). An ethnographic study of distributed problem solving in spreadsheet development. In *Proceedings of CSCW '90: Computer-Supported Cooperative Work.* (pp. 197-208), New York, NY: ACM Press.

Nielsen, J. (1994) Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies,* 41, 385-397.

Nielsen, J. (1993) *Usability Engineering.* New York, NY: Academic Press.

Nielsen, J. and Mack, R. L., Eds. (1994) *Usability inspection methods.* New York, NY: John Wiley & Sons, Inc.

Ochsman, R. B. and Chapanis, A. (1974). The effects of 10 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., and Wilbur, S. (1993) Conversations over video conferences: an evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction*, 8, 389-428.

Olson, G. M., Olson, J. S., Carter, M. R., and Storrøsten, M. (1992) Small group design meetings: An analysis of collaboration. *Human-Computer Interaction,* 7, 347-374.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., and Bruce, V. (1996) Comparison of face-to-face and video-mediated interaction. *Interacting with Computers*, 8(2), 177-192.

Orlikowski, W. J. (1992) Learning from notes: Organizational issues in groupware implementation. In *CSCW '92: Computer-Supported Cooperative Work.* (pp. 362-369) New York, NY: Association for Computing Machinery.

Pappas, T. N. and Hinds, R. O. (1995). On video and audio integration for conferencing. In *Proceedings of SPIE,* Vol. 2411. Bellingham, WA: The International Society for Optical Engineering.

Pastore, N. (1950) Some comments on "the use and misuse of the chi-square test". *Psychological Bulletin,* 47, 338-340.

Peters, C. C. (1950) The misuse of chi-square - a reply to Lewis and Burke. *Psychological Bulletin,* 47, 331-337.

Polson, P. G., Lewis, C. H., Rieman, J., and Wharton, C. (1992) Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies*, 36, 741-773.

Prussog, A., Mühlbach, L., and Böcker, M. (1994) Telepresence in videocommunications. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting,* (pp. 180-184) Santa Monica, CA: Human Factors and Ergonomics Society.

Ramachandran, V. S. and Anstis, S. M. (1986) The perception of apparent motion. *Scientific American,* 254, 102-109.

Rose, A., Shneiderman, B., and Plaisant, C. (1995) An applied ethnographic method for redesigning user interfaces. In *Proceedings of DIS '95: Symposium Designing Interactive Systems.* (pp.115-122). New York, NY: ACM Press.

Rouncefield, M., Hughes, J. A., Rodden, T., and Viller, S. (1994) Working with "constant interruption": CSCW and the small office. In *Proceedings of CSCW '94: Computer-Supported Cooperative Work* (pp. 275-285). New York, NY: ACM Press.

Royce, W. W. (1970) Managing the development of large software systems: Concepts and techniques. In Proceedings of *Wescon*, Aug. 1970.

Rutter, D. R. and Stephenson, G. M. (1977). The role of visual communication in synchronizing conversation. *European Journal of Social Psychology*, 7(*1*), 29-37.

Sanders, M. S. and McCormick, E. J. (1995) *Human factors in engineering and design.* New York, NY: McGraw-Hill Publishing Company.

Schreiber, W. F. (1984) Psychophysics and the improvement of television image quality. *SMPTE Journal,* Aug., 717-725.

Schuler and Namioka (1993) *Participatory design: Principles and practices.* Hillsdale, NJ: Lawrence Erlbaum & Associates.

Sekuler, R. and Blake, R. (1990) *Perception*, 2nd Edition. New York, NY: McGraw-Hill, Inc.

Sellen, A. J. (1992). Speech Patterns in Video-Mediated Conversations. In *CHI '92: Human Factors in Computing Systems* (pp. 49-59). The Association for Computing Machinery.

Shapiro, D. (1994) The limits of ethnography: Combining social sciences for CSCW. In *CSCW '94: Computer Supported Cooperative Work.* (pp. 417-428) New York, NY: Association for Computing Machinery.

Shattuck, L. G. and Woods, D. D. (1994) The critical incident technique: 40 years later. In *Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1080-1084). Santa Monica, CA: The Human Factors and Ergonomics Society.

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

Snow, M. P. (1996) *Charting presence and performance in virtual environments.* Unpublished Doctoral Dissertation. Blacksburg, VA: Virginia Polytechnic Institute and State University.

Stevens, S. S. (1975) *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects.* New York, NY: John Wiley & Sons.

Suchman, L. (1983) Office procedure as practical action: Models of work and system design. *ACM Transactions on Office Information Systems,* 1(4), 320-328.

Suchman, L. and Trigg, R. H. (1991). Understanding practice: Video as a medium for reflection and design. In J. Greenbaum & M. Kyng (Eds.), *Design at work*, pp. 65-90. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sun Microsystems (1997) Sun Glossary. *Located at http://www.sun.com/glossary/glossary.html*. Mountain View, CA: Sun Microsystems, Inc.

Swartz, M., Wallace, D., and Tkacz, S. (1992). The influence of frame rate and resolution reduction on human performance. In *Proceedings of the Human Factors and Ergonomics Society 36th Annual Meeting* (pp. 1440-1444). Santa Monica, CA: Human Factors and Ergonomics Society.

Tang, J. C. and Isaacs, E. (1993).Why do users like video? studies of multimedia-supported collaboration. *Computer Supported Cooperative Work (CSCW)*, 1, 163-193.

Tang, J. C. (1991) Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, 34 (2), 143-160.

Trevino, Lengel, and Daft (1987) Media symbolism, media richness, and media choice in organizations. *Communications Research*, 14(5), 553-574.

Twindale, M., Randall, D., and Bentley, R. (1994) Situated evaluation for cooperative systems. In *Proceedings of CSCW '94: Computer Supported Cooperative Work.* (pp. 441-452) New York, NY: Association for Computing Machinery.

Van Cott, H. and Kinkade, R. (1972) *Human engineering guide to equipment design.* Washington, DC: U.S. Government Printing Office.

Virzi, R. A. (1992) Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors,* 34, 457-468.

Virzi, R. A., Sokolov, J. L., and Karis, D. (1996) Usability problem identification using both low- and high-fidelity prototypes. In *CHI '96: Human Factors in Computing Systems* (pp. 236-243). New York, NY: Association for Computing Machinery.

Watson, A. B., Ahumada, A. J., and Farrell, J. E. (1986) Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America (A)*, 3(3), 300-307.

Watts, L., Monk, A., and Daly-Jones, O. (1996) Inter-personal awareness and synchronization: assessing the value of communication technologies. *International Journal of Human-Computer Studies*, 44, 849-873.

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

Westerink, J. H. D. M. and Teunissen, C. (1990) Perceived sharpness in moving images. In *SPIE, 1249: Human Vision and Electronic Imaging: Models, Methods, and Applications.* (pp. 78-87) The International Society for Optical Engineering.

Whittig, J., Honig, D., Carterette, E., and Eigler, N. (1991) Observer performance in dynamic displays: the effect of frame rate on visual signal detection in noisy images. In *SPIE, 1453: Human Vision, Visual Processing, and Digital Display II.* (pp. 165-176) The International Society for Optical Engineering.

Williams, E. (1977) Experimental comparisons of face-to-face and mediated communication: a review. *Psychological Bulletin*, 84(5), 963-976.

Williges, R. C., Williges, B. H., and Han, S. H. (1992). Developing quantitative guidelines using integrated data from sequential experiments. *Human Factors*, 34, 399-408.

Winer, B.J., Brown, D. R.., Michels, K. M.. (1991) *Statistical principles in experimental design (3rd Edition*). New York, NY: McGraw-Hill.

Wixon, D., Holtzblatt, K., and Knox, S. (1990) *Contextual design: An emergent view of system design.* In *Proceedings of CHI '90: Human Factors in Computing Systems.* New York, NY: Association for Computing Machinery.

Yaremko, R. M., Harari, H., Harrison, R. C., and Lynn, E. (1982) *Reference handbook of research and statistical methods.* New York, NY: Harper & Row.

Zuboff, S. (1988) *In the age of the smart machine: The future of work and power.* New York, NY: Basic Books.

# Appendix

# Appendix A

## Participant Instructions (Study #1)

Welcome! In this experiment, you will be asked to evaluate a series of video clips displayed on a computer monitor. Each one will differ slightly according to frame rate. The frame rate of a video clip is the number of frames redrawn on the screen each second. The higher the frame rate, the more smoothly the images move across the screen. Slower frame rates, therefore, tend to be somewhat "jerky" or stilted.

Each video clip is approximately 10 seconds in length. After each one, you will be asked to answer to brief questions about the quality of the video. The first question will ask you to assign a number corresponding to the quality level of the video clip. You can chose any number you like, but higher numbers indicate better image quality, and lower numbers indicate more degraded image quality. For each successive video clip, you are to assign a number reflecting your subjective impression. There is no limit to the range of numbers you may use. Whole numbers, fractions, and decimals are all acceptable. Try to make each number match the image quality as you perceive it.

The second question will have you evaluate the image quality acceptability according to a 5-point scale. Simply circle the number on the scale you believe corresponds to the image quality of the video clip you just observed. Following the completion of the two questions, the next video clip will be presented, and the sequence will be repeated.

At the end of the experiment, you will be given a brief questionnaire. Following the completion of the questionnaire, any questions you may have will be addressed, and you will be paid for you time.

To accommodate you to the nature of the task, we will practice a few times by judging the length of a line. When you are ready, you will be shown 10 lines (one at a time). For each line you are to assign a value based on its length relative to the others. You may begin with any number you wish.

Do you have any questions?

# Appendix B

## Questionnaire for Communication Effectiveness

Please fill out the items on this questionnaire by circling the number which best corresponds to your opinion. If the question is not appropriate, then circle n/a.

**Part A. As a Speaker:**

1. I felt as if my message was well-understood by the remote group.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

2. I was able to keep the remote group's attention.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

3. I knew if the remote group was paying attention.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

4. I was not distracted by the computer arrangement when talking.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

5. I thought this arrangement was just as effective as face-to-face communication.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

6. I was able to make eye-contact with the remote group.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

7. I was able to see the remote group's facial expressions.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

8. I was able to see the remote group's hand gestures.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

9. I was able to see the remote group's body movements.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

10. I did not feel awkwardly interupted by the remote group.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

11. I felt that transitions between who was speaking were smooth.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

**Part B. As a Listener:**

1. I was able to understand the remote group's message.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

2. I did not have to strain too much to understand the remote group's message.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

3. I paid attention to the remote group.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

4. I was not distracted by the computer arrangement when listening.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n / a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

5. I thought this arrangement was just as effective as face-to-face communication.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n / a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

6. I was able to make eye-contact with the remote group.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n / a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

7. I was able to see the remote group's facial expressions.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n / a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

8. I was able to see the remote group's hand gestures.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n / a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

9. I was able to see the remote group's body movements.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n / a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

10. I was able to interupt the remote group easily.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n / a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

11. I felt that transitions between who was speaking were smooth.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n / a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

**Part D. Transmission System (audio and video):**

1. The video quality was acceptable.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

2. The video was clear enough.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

4. The size of the video window was adequate.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

5. The video was just as good as a live lecture in the same room.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

7. The audio quality was acceptable.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

8. The audio was clear enough.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

9. The audio was just as good as being in the same room.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

**Part E. Physical Environment:**

1. The lighting level was appropriate in my room.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

2. The lighting level on the video image was appropriate.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

3. The noise level in my room was not too loud.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

4. The noise level in the other (remote) room was not too loud.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

5. The camera angle was appropriate.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | n/a |
|---|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Mildly Disagree | Indifferent | Mildly Agree | Moderately Agree | Strongly Agree | |

Any additional Comments you wish to share:

Many thanks for your participation.

## Appendix C
### Structured Interview for Contextual Inquiry

1) What problems did you have communicating with each other?

2) How might those problems be remedied?

3) Did you notice any changes in your behavior since last week?

6) Do you have any other comments about the system and its' use?

# Appendix D
## Video Analysis Sheet

**Video Tape Analysis -- Ethnographic Study -- Fall, 1996**

Date of Video: ________ Time of Video: ___________ Session #: ___________
Date of Analysis: _________ Time of Analysis: ___________ Team: A B C D Analyzer: ___________

---

1) Verbal Turntaking Breakdown
i.e. Overlapping talk (except explicit interruptions)
Time: Repair Method: Notes:

---

2) Reference Breakdowns
i.e. Breakdowns over shared objects (papers, notes, pencils, mouse, keyboard, etc.)
Time: Repair Method: Notes:

---

3) Topic Breakdowns
i.e. Failure to maintain topical orientation, when one member is confused as to the topic
Time: Repair Method: Notes:

---

4) Visual Breakdowns
i.e. Requests to shift seating, re-position camera, show a sketch better, etc.
Time: Repair Method: Notes:

---

5) Audio Breakdowns
i.e. Loudness, feedback, and static problems, requests for repeating due to these problems
Time: Repair Method: Notes:

---

6) Shared Conversation Breakdown
i.e. When conversation is directed at a co-located group member only.
Time: Repair Method: Notes:

# Vita

Jonathan K. Kies was born March 4, 1970 in Summit, New Jersey. He received a B. A. in psychology from Miami University in Oxford, Ohio in 1992. The summers of 1993 and 1994 were spent prototyping and testing a screen-based telephone and other telecommunications products at NYNEX Science and Technology in White Plains, New York. In 1994, he completed an M. S. in Industrial and Systems Engineering in the human factors option area at Virginia Tech. While working in the Human-Computer Interaction Lab at Virginia Tech, he has conducted research on desktop video conferencing, the world-wide-web, scenario-based design methods, and multi-user domains (MUDs and MOOs). In addition to these projects, he was involved with the design of a new usability testing facility, two world-wide-web sites, an on-line hypermedia technical report series, and a multimedia CD-ROM. His research has been funded by the National Institute for Occupational Safety and Health, NYNEX Science and Technology, the National Science Foundation, and the Southeastern University and College Coalition for Engineering Education (SUCCEED). He is a member of the Human Factors and Ergonomics Society, the Association for Computing Machinery (Special Interest Group on Computer-Human Interaction) and Alpha Pi Mu, the National Industrial Engineering Honor Society. He has accepted a position with Lucent Technologies in Holmdel, New Jersey where he will conduct research on wireless communication devices. A sample of his publications include:

Kies, J. K., Kelso, J., and Williges, R. C. (1995) The use of scenarios to evaluate the effects of group configuration and task on video-teleconferencing communication effectiveness. In *Proceedings of the Third Annual Mid-Atlantic Human Factors Conference*. (pp. 22-28) Blacksburg, VA.

Neale, D. C. and Kies, J. K. (1996). Scenario-based design for human-computer interface development. In *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society,* (pp. 338-342) Santa Monica, CA: Human Factors and Ergonomics Society.

Snow, M. P., Kies, J. K., Neale, D. N., and Williges, R. C. (1996) A case study in participatory design. *Ergonomics in Design.* Santa Monica, CA: Human Factors and Ergonomics Society, 18-24.

Kies, J. K., Williges, R. C., and Williges, B. A. (1997) Desktop video conferencing: A systems approach. In Helander, M. G., Landauer, T. K., and Prabhu, P. (Eds.) *Handbook of Human-Computer Interaction, 2nd Edition.* Amsterdam, The Netherlands: Elsevier Science.

Kies, J. K., Williges, R. C., and Rosson, M. B. (in press) Studying computer-supported cooperative work: A review of research issues and strategies. *Journal of the American Society for Information Science.*

Kies, J. K., Williges, R. C., and Rosson, M. B. (in press) Evaluating desktop video conferencing for distance learning applications. *Computers and Education.*