The Connection Project: Toward supporting virtual collocation

Emilee J. Rader, Erik C. Hofer and Thomas A. Finholt

School of Information
University of Michigan
Ann, Arbor, MI
{ejrader, ehofer, finholt }@umich.edu

ABSTRACT

This paper describes the Connection Project, an effort at the University of Michigan to support an academic unit divided between two buildings on campus using high-quality, high-bandwidth conferencing technologies. This paper describes the systems used in the project and provides an overview of a number of human-factors related research activities that are being conducted as part of the project.

1. INTRODUCTION

Social studies of science and science policy researchers have long observed the increasing scale of scientific research and argued that the increasing capital and intellectual demands of cutting edge problems will lead to a qualitative change in how science is conducted, "big science [1]." While big science is very real for a small number of domains, most notably high-energy physics, much of science is still conducted in relatively small teams. Increasingly common however, are teams that are highly interdisciplinary, involving groups of researchers from different primary communities of practice who work together as part of a center or emerging field of research. At the University of Michigan, for instance, we see this through emerging research and academic programs on the environment, biomedical engineering and nanoscale systems.

Organizing members of these research programs presents a challenge because the underlying problems are inherently interdisciplinary. The traditional approach – geographic collocation – does not work because each researcher already has a primary community of practice, often associated with field-specific facilities or pools of expertise. Similar challenges to collocation exist in organizations that simply lack space. We believe that *virtual collocation* -- the realization of at least some of the benefits of collocated work [2] -- can be achieved through a combination of organizational practices and technical systems, but the design space is not well understood. In this paper we describe the Connection Project, an effort at the University of Michigan School of Information, funded by the University of Michigan Office of the Provost, to explore this design space by conducting laboratory studies, building new systems, and studying the effects of their use.

2. PROJECT OVERVIEW

While there is considerable interest in designing collaborative environments to support highly distributed, multi-disciplinary project teams, these settings are extremely complex and include a number of social, technical and organizational challenges that make it difficult for even the best technologies to succeed [3]. We believe that organizations that are divided by smaller distances, such as between multiple buildings at a single university, provide an opportunity to study the potential impacts of collaborative technologies in settings where they are much more likely to succeed, yielding insights into the technologies themselves, rather than just the difficulties of distributed work. In these settings, videoconferencing emerges as a particularly attractive option for virtual interaction because, at least in theory, video-mediated communication engages the same channels as face-to-face conversation. That is, people can both see and hear each other – and feedback is immediate.

Unfortunately, the potential of videoconferencing is often missed through a combination of technological and ergonomic deficiencies. For example, users are very sensitive to even subtle changes in picture or audio quality, while image and audio fidelity is often significantly less than what is experienced in face-to-face meetings. In addition, lack of gaze contact, reduced perception of body language, and absence of broader visual access all combine to make video interaction disorienting relative to face-to-face interaction [3]. The faults of contemporary videoconferencing can be summarized in terms of the following description. Resolution is poor and lighting is inadequate, peripheral participants (who appear less engaged) sit closer to the camera, while the principal

speaker sits deep in the room to be fully within the camera's field-of-view (and is therefore apparently smaller) – but must shout to be heard over the poor audio channel.

The initial phase of the Connection Project seeks to improve this conferencing experience, focusing on the design of a high-quality conferencing system to connect conference rooms in two buildings, roughly three miles apart, occupied by the School of Information. A program of basic and applied research, described below, informs the design of this system.

3. SYSTEM DESCRIPTION

The first Connection Project systems focus on two conference rooms, one at each School of Information location. These conference rooms were selected because they are already heavily used as meeting places for projects within the school, with some prior use of commercial video conferencing systems to connect the two places. While the video connection between the two rooms was functional, it was far from ideal. Each site had a number of ergonomic problems that made conferencing between the rooms awkward an unnatural. Prior to the installation of the Connection Project systems, the Connector Conference Room (CCR, see Figure 1) was equipped with an extremely wide, light-colored table contributed to video of the room appearing washed-out and made it possible for a single camera to capture all the participants in the room. Video was displayed on a single, 10-foot projection screen using a projector that was very dim, causing remote participants to appear very large and washed out (if the lights were on) and forced local participants to choose between viewing video or data. The camera for this system was mounted to the left of the screen, high on the wall, making for an extremely awkward viewing angle – participants in a conference appeared small, and very far away. Any type of approximation of gaze or eye contact was simply impossible in this setup.

FIGURE 1.: Conference room prior to redesign.

To improve the ergonomics of the conferencing environment in this room, a number of changes were made to the room (see Figure 2). The single projection screen was replaced with two 50" plasma displays, providing a more life-size image of the remote site, displays that work well in high ambient light conditions, and the capability to simultaneously display separate data and video streams. The wide table was replaced with a narrow, wooden table, to rebalance illumination levels and ensure that all participants could appear in a single shot. Cameras for the new system were mounted on top of the video display, centered at the end of the conference table, providing a much better approximation of eye contact and gaze awareness. The data and video displays were positioned in such a way at the two sites that it is possible to tell which display remote participants are looking at. In addition to these ergonomic improvements, the existing conferencing is being replaced, replacing the legacy H.323 capability with upgraded H.323 equipment, an Access Grid capability and DV feeds between the two sites. Remote data display is handled through a dedicated system that encodes a VGA signal as an MPEG4 stream. We are currently testing a bidirectional high definition video link and are evaluating Teravision as a possible replacement for the existing data link. We replicated this setup at the other School of Information site.

Figure 2: Upgraded conference room.

4. RESEARCH OVERVIEW

An essential part of the work underway as part of the Connection Project is the exploration of usability, human factors, and social impact issues that are part of the development and deployment of the system. In this paper, we will describe several activities that are currently underway in these three categories of user research.

4.1 Usability research

During the early stages of systems design we conducted a number of interviews of users and potential users of the system. One theme that emerged from these interviews as that a major barrier to the adoption of videoconferencing systems is that average users who want to participate in meetings using these systems are not technically expert enough to operate them without the assistance of someone who is specifically trained to do so. For many conferencing systems, this results from a number of subsystems that are independently well documented and easy to use, but extremely difficult to master when taken together. For the existing conferencing setups, a user had to master remote controls for a projector, for the video conferencing endpoint and the controls for the audio subsystem in the room. While each of these controls is not terribly complex on its own, understanding each of and the order of the necessary commands for setting up a call was simply overwhelming.

In response to these problems of system usability, we are currently evaluating a number of different control strategies for the conferencing systems. One approach involves the design and construction of a number of customized serial control devices. Each of these devices includes a small number of hardwired buttons, tailored to the common actions needed to set up a call. Developing these systems from scratch allows us to use a common look and feel across devices, give users a tangible interface for initiating calls, and allows us to easily execute common command sequences through a single button push. In addition to this embedded system approach, we are designing and testing a set of touch-screen interfaces that will allow for more comprehensive system control through a unified interface. Using this combination of tangible and touch screen interfaces, we hope to reduce the workflow of setting up a call to a single button press, rather than a long sequence involving two or more remote controls.

4.2 Documentation development

Another area of usability research and development has involved the creation of usable and effective training materials for the systems. Initial user studies suggested that a lack of familiarity with conferencing systems and the uncertainty about how to complete a call were both major barriers to use. Comments from users suggested that they needed documentation designed around their potential actions, rather than the comprehensive capability sets of each component. These comments highlight the difference between system documentation and user documentation and suggest that both are necessary.

We generated training and documentation materials for the system, including a training video and a brief step-by-step user guide. Our motivation for developing these materials was to mitigate the reluctance potential users might feel in the face of unfamiliarity with the system. By walking them through the steps required in order to hold a virtual meeting, we hoped to empower them to use the system without our assistance.

The primary documentation for the system includes screenshots and annotated photos of the equipment to assist users in completing the steps to activating the system correctly. This documentation is designed to guide users towards completing a specific action – connecting to the other School of Information site – rather than serving as an instruction manual for how the system works.

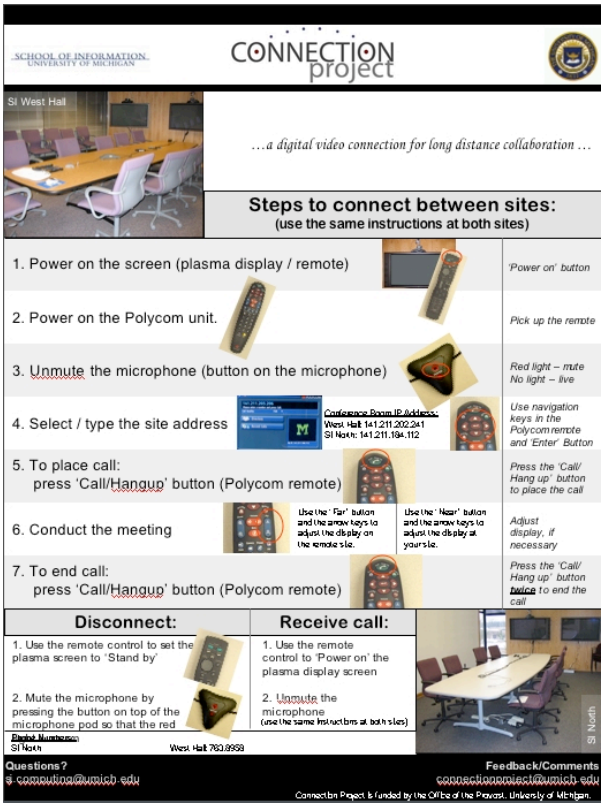

FIGURE 3: Quickstart Documentation

In addition to the step-by-step user guide, we produced a training video that provides an overview of the equipment that is used in each of the Connection Project rooms in an effort to increase users' knowledge of the system so they feel more comfortable using it.

The video provides a step-by-step walkthrough of setting up and adjusting the system for a call, along with a brief description of common video conferencing etiquette and procedures. Both the documentation and training video were user-tested before being released to faculty and staff members of the school. The user testing of the documentation consisted of informal think-aloud usability studies with several representative users of the system. Care was taken to select users who had not previously used the system for a videoconference so that the test would more realistically represent actual use. Several improvements were made based on feedback from these tests, including producing a minimal instruction set designed to fit on one page to supplement other training materials (See Figure 3). The training video was put through several dry runs with staff members of the School of Information, who raised several issues with unclear statements in the script and helped to generate a list of questions about the system such as a high level overview of how the technology works, and whether or not use of the system must be billed like a telephone call. The user testing and constant refinement of documentation was a significant amount of work, but was necessary to ensure that the system is accessible to the School of Information community.

4.3 Video Quality Research

The Connection Project systems are able to connect the two spaces using a number of video codecs at many different quality and bandwidth levels, ranging from 128 Kbps H.264 streams to 30 Mbps DV streams. Given the different quality choices and bandwidth costs of these different methods, it is important to understand when and if the higher quality codecs will justify their bandwidth cost. Another area of Connection Project research seeks to understand the impact of different codec choices on subjective quality in order to better understand the implications of codec choice.

Many factors have the potential to influence video quality. Gili et al. [3] identified seven factors relating to encoding and playback algorithms that influence quality. Due to the similarities in commonly used algorithms, three factors are generally understood to impact video quality between the systems heavily used for conferencing: frame resolution, frame rate and frame quantization. Frame resolution refers to the number of pixels that make up a given frame in a video clip. Higher resolution frames are able to display more detail, resulting in a crisper image. Frame rate refers to the number of frames presented per second. NTSC television is composed of 30 frames per second, resulting in smooth video. Many studies and standards emphasize the frame rate as a key determinant of video quality. Frame quantization refers generally to the quantization factor of the discrete cosine transformation algorithm used in block-based video compression algorithms like JPEG, DV and the intraframe compression portions of H.261 and H.263. McCarthy et al. [4] recently identified quantization as the key determinant of quality in certain cases – more important than frame rate.

Accurate measurement of video or image quality is a challenge, though many products and metrics have been developed for quality assessment of television video. Internet-based video however, still heavily relies on subjective experiments for quality evaluations because of the wider range of impairments that result from different encoding techniques and formats. A typical subjective experiment involves having several non-expert viewers watch short clips of video shot from a constant camera angle and rate the quality of those clips on a 5-point Likert scale [5].

We conducted a study to measure the differences in perceived video quality of four test scenes under the following conditions: H.263 at 384 kbps, H.263 at 1920 kbps, DV, and NTSC playback of a DVD (control). Four test scenes were obtained from the Video Quality Experts Group [6]. Each scene is 8 seconds in duration. We selected scenes that differed in the amount of motion depicted in the video, and in subject matter. A 4 x 4 x 4 mixed design experiment was conducted with 20 participants. The subjective evaluation procedure used in this experiment was based on ITU-R Rec. BT.500-11 Double Stimulus Continuous Quality Scale [5].

DV test scenes with a lot of motion were perceived to be of equal quality to the control scenes. However, high-motion scenes were rated significantly worse than the control in the 382 kbps and 1920 kbps conditions. Clips with little motion were not significantly different from the control clips.

These findings provide some insight into the bandwidth/quality tradeoffs in video conferencing systems. The higher-bandwidth systems resulted in a higher perceived quality, but the differences were most dramatic in cases when there was a large amount of motion in the scene. This suggests that the higher-bandwidth cost is justified in cases where there is a lot of detailed motion in the scene. In a conferencing context, this commonly occurs when a scene involves a number of people instead of a closeup of one person.

In cases where a conference involves a smaller number of people or only a single participant, the medium-bandwidth systems are likely acceptable, realizing a large quality increase over low-bandwidth systems, but a lower-magnitude disadvantage to the high-bandwidth systems. This finding is somewhat surprising, as it is generally accepted that low bandwidth systems work quite well for scenes of individual's faces.

This study highlights a number of opportunities for future work that will strengthen the recommendations for users and developers of conferencing systems. One area for future research is to explore the impact of simulated network disturbances on the different codecs. Introducing network emulation into the hypothetical reference circuits will allow us to simulate real world conditions and make the findings more applicable to deployments of conferencing technologies.

4.4 Social Ergonomics Field Trial

Another area of human factors research involves developing a better understanding of the social ergonomics of conferencing technologies. "Social ergonomics" refers to the changes in social behavior that result from different technical design decisions. For example, tension often exists between public and private use of the technology when videoconferencing systems are deployed. Choices must be made between installing the equipment in a public (lounge or hallway) vs. a private (conference room or office) space, and there are different consequences for the subsequent use of the system depending on the environment in which it is installed. We are currently preparing to conduct a field trial of a system that will support both public and semi-private conversations. This system will connect staff lounge areas in the two School of Information locations, and a variety of qualitative research methods will be used to collect usage, usability, and satisfaction data from end users. We want to collect data on the real-world usage of such a system, in order to learn more about adoption and use, as well as investigate technical solutions to the problem of supporting private mediated conversations in a public space.

The field trial will examine the utility of mechanisms to simulate privacy cues for video-mediated communication in public settings. We will observe use of a system that uses proximity to modify audio levels and visual access within a point-to-point video communication installation in the School of Information. Data will be collected through interviews, observations, and communication logbooks filled out by participants, to determine the number of interaction opportunities, total number of interactions, and the number of interactions that move from public to private. We will also obtain feedback about system usability and about whether the designed cues signaled privacy. Use of the system will be contrasted with observation of the same public spaces (e.g., lounges within the West Hall and North Campus locations) without the video technology.

The study will be conducted in three stages. The baseline stage takes place before the planned system installation. Participants will complete communication logs for one week, and afterward will be interviewed (Time 1). Observations of the public areas in which the system will be installed will also take place. The second stage of the study takes place after the system is installed. Participants will complete communication logs during the third week of the installation, and then in the fourth week be interviewed again (Time 2). Observations of the public areas will also be conducted while the system is in operation. Finally, after the system has been removed, participants will fill out communication logs for a final week, and be interviewed a third time (Time 3). Observations of the public areas where the system was installed will be conducted post-installation.

Results from this study will be used to better understand the social processes at work in situations where videoconferencing systems are installed in public spaces, and to elaborate on implications for the design of systems to support semi-private conversations in public spaces.

4.5 Adoption and Organizational Impact

Another area or research involves the assessment of the adoption and use of these conferencing systems. Fully understanding the organizational impact of technologies is extremely difficult. Common metrics of technology success are often based on frequency measures of technology use, focusing on first order use of the technology rather than second order effects, such as changes in social networks or other human-focused outcomes. We are using a variety of data collection methods to assess the degree of adoption and social impact of the Connection Project in the School of Information, including interviews, usage logs and social network analysis instruments.

We are currently in the process of conducting an initial analysis of the social network data we have collected. Social network analysis [9] describes a set of analytical methods used to enumerate and interpret patterns of communication between individuals within an organization. These methods highlight the frequency of communication between individuals and have been used to understand the spread of innovation [10],

productivity [11] and the creation of energy [12] in organizations. Using the social network data that we have collected, we will be able to see changes in the social network of the school at different time points, watch how patterns of use change and watch how overall perceptions of the technology develop and change over time. Over time, we hope to be able to understand the role the Connection Project systems and other collaborative tools play in facilitating the flow of information and innovation between School of Information sites.

5. ACKNOWLEDGEMENTS

This project was generously funded through the University of Michigan. We are also indebted to the rest of the Connection Project team and project alumni, Rasika Ramesh, Jina Huh, Il-Hwan Kim, Jason Lin, Turquoise Archie and Pritpaul Mahal.

6. REFERENCES

[1] Weinberg, A. (1961). Impact of large-scale science on the United States. *Science, 134,* pp. 161-164.

[2] Teasley, S. D., Covi, L., Krishnan, M.S., Olson, J. S. S. (2002). Rapid Software Development through Team Collocation. *IEEE Transactions on Software Engineering, 28,* 7, 671-683.

[3] Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human Computer Interaction*, 15, 139-179.

[4] Whittaker, S. (1995). Rethinking video as a technology for interpersonal communications: theory and design implications. *International Journal of Man-Machine Studies, 42,* pp. 501-529.

[5] Gili, M.J., Janez, E.L., Hernandez, L.M. and Szymanski, M. (1991). Subjective image quality assessment and prediction in digital videocommunications. COST 212 HUFIS Report.

[6] McCarthy, J.D., Sasse, M.A., and Miras, D. (2004). Smooth or sharp? Comparing the effects quantization vs. frame rate for streaming video. In *Proceedings of the ACM Conference on Human Factors in Computing (CHI 2004)*, Vienna, Austria.

[7] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures." International Telecommunication Union, Geneva, Switzerland, 2002.

[8] VQEG. (2000). Final report form the Video Quality Experts Group on the validation of objective models of video quality assessment. http://www.vqeg.org

[9] Wellman, B., & Wasserman, S. (2000). Social networks. In Kazdin. A. (ed.), Encyclopedia of Psychology, Volume 7, pages 351-353. New York: American Psychological Association and Oxford University Press.

[10] Wu, F., B. A. Huberman, L. A. Adamic, and J.R. Tyler (2003). Information flow in social groups. *Physica A*.

[11] Cummings, J. & Cross, R. (2003). Structural Properties of Work Groups and their Consequences for Performance. Social Networks 25(3), pp. 197-210.

[12] Cross, R., Baker, W. & Parker, A. (2003). What Creates Energy in Organizations? Sloan Management Review 44(4), pp. 51-57.