Experiments Comparing Face-to-Face with Virtual Collaborative Learning

Randall B. Smith

Sun Microsystems Laboratories

Michael J. Sipusic, Robert L. Pannoni

SERA Learning Technologies

Abstract: We report on set of studies conducted over two years involving over 1000 students at two universities. The main study compares three conditions: conventional classroom lecture, a face-to-face collaborative learning technique called Tutored Video Instruction (TVI), and the virtual-world counterpart of TVI, Distributed Tutored Video Instruction (DTVI). The main study involved over 700 students in 6 courses. When using final course grade as an outcome measure, it has been previously established that TVI students outperform lecture students. Therefore the comparison of interest for us is between DTVI and TVI: would the audio and video technology used to support a distributed group enable DTVI students to achieve the higher grades attainable in TVI? We found no statistical difference between the grades of the DTVI and TVI students, and both groups outperformed the lecture students.

We also summarize extensive interaction process data and survey data, then report on some more informal studies assessing the usability and effectiveness of an "Enhanced DTVI" system, in which distributed students can not only see and talk over digital networked media, but can take notes together in real time.

Keywords: computer-mediated communication, multiuser virtual environments, videoconferencing.

Introduction

One-by-one, the technological and cost barriers to high-bandwidth networking are falling. The increasing bandwidth availability inspires visions of a new world where people can work collaboratively without being held hostage to physical proximity. But lost in the frenetic rush toward virtualization is the possibility that there will be an additional cost to the substitution of virtual communication for face-to-face communication. While networks have proven quite adept as a medium for asynchronous and broadcast forms of communication, highly interactive, real-time, multi-way communication has proven more problematic (Sellen 1992).

Problems with technology-mediated communication would seem to be particularly significant in the domain of distance learning. Distance learning is likely to be among the first large-scale uses of high-bandwidth networking technology, because of the obvious savings in time and travel costs. The simplest way to apply this technology is to recreate the familiar classroom lecture environment. In fact for many, the broadcast of standard lecture courses is synonymous with "distance learning." There is, however, an overwhelming body of educational research showing that instructional methods that foster interpersonal discourse and the social construction of knowledge are more effective than methods that rely on the broadcast of information (Cohen 1994). Therefore the rush to virtualize standard lectures may in fact be a retreat toward outdated and less-effective instructional methods.

We attempt to "raise the bar" for distance learning by moving from a classroom transmission metaphor to a collaborative learning metaphor. Collaborative learning techniques have been shown to be consistently superior to traditional classroom lecture both in effectiveness and student satisfaction (Cohen 1994; Johnson and Johnson 1994). However, collaborative learning is highly dependent on communication or *discourse*. Past research shows communication breakdowns in video-mediated settings. We designed this research project to find out whether video-mediated communication could support the rich social discourse required for collaborative learning.

The collaborative learning method chosen for this study is Tutored Video Instruction (TVI). TVI was invented at Stanford University over twenty years ago. In TVI, a small group of students play a pre-recorded videotape of a classroom lecture. During the playing of the tape, a facilitator encourages the students to pause the tape to ask questions or discuss topics. As with other forms of collaborative learning, TVI students have been shown to outperform students who physically attended the lectures (Gibbons et. al., 1977).

Duplicating the success of TVI in a video-mediated environment would seem to be a significant challenge. Like other forms of collaborative learning, TVI groups exhibit frequent and complex social interaction among group members. Students in TVI use verbal and non-verbal communication to negotiate meaning and relationships with group members. Video has been shown to be much less important than audio in broadcast learning environments, but because of the need to support non-verbal communication, the visual channel may play a crucial communication role in TVI. The purpose of this study is to determine whether these rich communication patterns and relationships can be created in a video-mediated environment specifically designed to mimic face-to-face TVI. The virtual version of TVI created for this experiment was called "Distributed Tutored Video Instruction," or DTVI.

Over 700 hundred students from 6 courses and 2 universities were involved in the main experiment. Details of the protocols, procedures, and statistical analysis used in this experiment can be found in a technical report from Sun Laboratories (Sipusic et. al., 1999), though we summarize the main findings here.

We also discuss our experiences with a follow-on system called "Enhanced DTVI." Audio and video conferencing are only the beginning of what computer networks can do, and the Enhanced DTVI system, built in the Kansas shared application construction environment (Smith et. al. 1998, Smith et. al. 1997, Maloney & Smith 1995), is designed to investigate how we might better utilize the technology. In Enhanced DTVI, students watch a tape and interact as they would under normal DTVI, but they can collaboratively take notes as the lecture plays. The notes, which form a hierarchical outline, automatically appear as web pages so that the group members, the instructor, and even students from other groups can view the notes later. Our sample size was too small for significant statistical analysis in this part of our study, but we found the system well-received, and were able to demonstrate that shared note taking among up to 7 students works well. Because the notes form a shared task whose outcome is of potential value to the students, collaboration might be elevated in systems like Enhanced DTVI.

Video-Mediated Communication

Given that the collaborative learning effect is built on content and group-oriented discourse, the ability of technology to support all of the subtle aspects of human communication is a serious issue for collaborative distance learning. Recent studies of collaborative work done at a distance over a video link have found that there are subtle and unexpected communication difficulties encountered by participants attempting to coordinate their efforts through these systems. Isaacs and Tang (1993) reported that participants using the prototype *ShowMe* from Sun Microsystems had difficulty in the real-time management of a number of social behaviors that support conversations, such as turn-taking and the coordination of joint attention through eye-gaze. For many of the same reasons, Fish, Kraut, Root, and Rice (1992) in an evaluation of AT&T's video telephone, *Cruiser,* state "users said they used face-to-face communication rather than the Cruiser system because the Cruiser system was not able to support all the communication demands of conventional work activities."

Communication breakdowns resulting from technology-mediated communication can typically be attributed to three causes:

Network transmission artifacts. Low frame rate, audio and video latency, and artifacts from information compression have been shown to affect people's preference for the media, as well as how they interact through it (Kies et. al.,1996, Isaacs & Tang, 1993). We did not want to position this work so that it might give a negative result valid only for the few years it will take for these transmission artifacts to become a negligible factor. Consequently, in our main DTVI experiment, we used only analog video and audio links so that latency was negligible and image quality was relatively high. In our Enhanced DTVI experiment, we used the purely digital test bed provided by the Kansas system. Even here, the audio quality was high (16 bits at 22 kHz), the latency was barely noticeable, and the video image quality, though perceptibly lower than the analog system, was still quite good (the digital version averaged about 6 Mb per sec of multicast data for the 9 audio and video streams).

Reduced image size. Image size may adversely effect conversation patterns (Monk & Watts, 1995) and the willingness of individuals to interact with (Fish et. al., 1990) or trust (Rocco 1998) remote collaborators. Storck and Sproull (1995) found that students who view their peers through video give them lower competency ratings than they give their co-present peers.

Diminished directionality cues from eye gaze. Difficulty in establishing the location of participants' gaze has been linked to various perturbations in the communication process. These, include exaggerated gestures and staring (O'Conaill et. al., 1993; Colston & Shiano 1995), more formal speech with fewer overlapping utterances (Monk & Watts 1995; Gale 1998; Cohen 1994; Sellen 1992) and difficulty in managing group speech issues such as turn-taking and maintaining or acquiring the floor (Isaacs & Tang 1993).

To the extent that the DTVI system reproduces these problems, we would expect that the collaborative learning processes of DTVI groups would be compromised when compared to their face-to-face TVI equivalents.

The DTVI system

The DTVI prototype was designed to simulate a face-to-face TVI session as closely as possible. The system uses an open microphone architecture in which anyone can speak and be heard at anytime. Students wear stereo headphones with a built-in microphone attached. The group discussion plays in one ear while audio from the course videotape plays in the other. Students can adjust the volume of the two channels independently. They can also mute the microphone.

Analog video distribution was chosen for the prototype to achieve full 30-frames per second (TV quality) video with no compression artifacts or perceptible latency. This idealized environment allowed us to study the learning process without worrying about the effect of digital transmission artifacts. We compared the analog version with a purely digital networked equivalent version, and will discuss these findings in the section on Enhanced DTVI.

Real-time video of each participant is delivered from a camera positioned beside the monitor. These images are arrayed in individual cells of a 3x3 "Brady Bunch" matrix (see Figure 1). For the main DTVI experiment, the matrix displays as an analog video overlay on a 20-inch computer screen. Videotape of the course lecture plays in the lower right cell. Because of the 3x3 matrix limitation, a maximum of seven students (and one tutor) can participate in a DTVI session. Individual cells of the matrix cannot be resized, but students can alternate between watching all nine cells or exploding a single cell to take up the full window. The video window itself is also fully movable and resizable.

Figure 1. The DTVI system presents each user with this "Brady Bunch" view of the group. A videotape image of the lecture is at lower right, and the facilitator (lower center) and up to 7 students participate

The tutor controls the VCR that plays the course video. Students can verbally request that the tutor pause the tape or can send a pause request message to the tutor by pressing a button on the user interface. (The verbal method was by far more common). Later versions of the DTVI software allow students to send private text messages to the tutor. This feature was provided primarily as a way for students to notify the tutor of technical problems when the microphone wasn't functioning correctly. We resisted the temptation to add additional features to the software because we wanted to maintain the direct comparison between TVI and DTVI.

Academic Performance: Lecture vs. TVI vs. DTVI

California State Polytechnic University at San Luis Obispo (Cal Poly) and California State University at Chico were the two institutions in the study. There were six courses used in the experiment. The courses are described briefly here.

Course Title	Symbol	Location	Number of students Lecture / TVI / DTVI		
			Lecture	TVI	DTVI
Architecture, Materials of Construction	ARCH	Cal Poly	34	17	15
Media Esthetics, introduction to recognition and interpretation of media production.	CDES 40	Chico	134	25	26
Telecommunications Industry, covers the technological, historical, and legal development of electronic media in America	CDES 65	Chico	97	37	32
Materials Engineering, introduction to the physical properties of materials.	MATE	Cal Poly	30	45	42
Business Management, graduate level course in the MBA program.	MGT	Cal Poly	0	20	18
Management Information Systems, introduction to the application of computer systems to businesses.	MIS	Cal Poly	24	63	57
		Total	319	207	190

To assess academic performance, we converted final letter grades for the course into numbers based on a 4.0 grading scale. We then ran an analysis of variance (ANOVA) using course and method as experimental factors and incorporating overall GPA (as measured at the start of the course) as a covariant. As the table below shows, the mean grade for TVI students and DTVI students is nearly identical while the mean grade for lecture students is lower.

Estimates with GRADE as Dependent Variable

| Method	Mean Grade	Std. Error
Lecture	2.83	.052
TVI	3.14	.057
DTVI	3.13	.060

A pairwise comparison shows that TVI and DTVI means are not significantly different from each other at the .05 level (p=.861) and that both the TVI and DTVI mean are significantly different from the lecture mean (p=.000). Virtualizing TVI does not appear to lessen the collaborative learning effect, at least as measured by course grade.

Pairwise Comparisons based on estimated marginal means

Method I	Method J	Mean Difference (I-J)	Significance	95% Confidence Interval	
				Lower Bound	Upper Bound
Lecture	TVI	-.305 *	.00	-.491	-.120
	DTVI	-.291 *	.00	-.481	-.101
TVI	DTVI	.0144	.861	-.183	.212
	Lecture	.305 *	.00	.120	.491
DTVI	TVI	-.0144	.861	-.212	.183
	Lecture	.291 *	.00	.101	.481

* Significant at the .05 level

Besides looking at grades by course, we looked at whether students of varying academic ability responded differently to TVI and DTVI. We sorted students into quartiles based on their GPA relative to other students in the same course. The chart below shows the mean grades for students in each quartile:
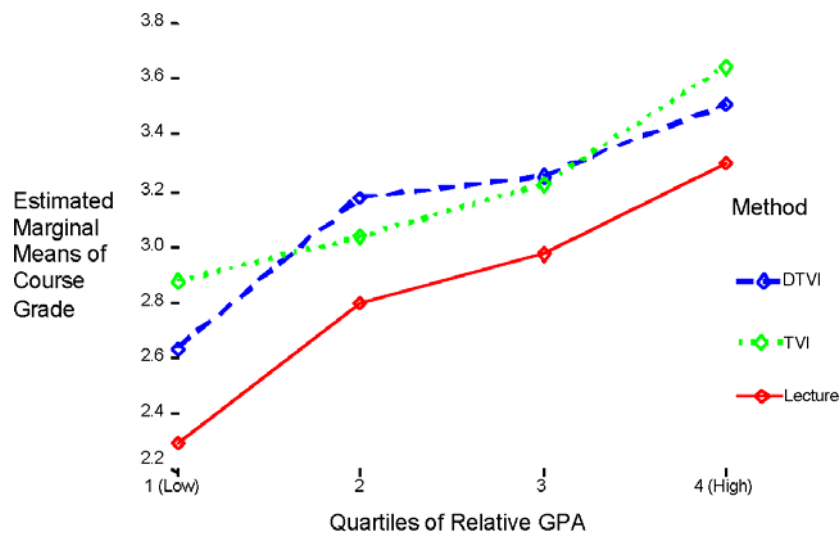
<P

For both TVI and DTVI students, the collaborative learning effect held for students of all levels of academic ability.

Other observations

In addition to student grades, grade point averages, and SAT scores, our main DTVI / TVI comparison data includes a catalogue of over 3000 conversational exchanges, hundreds of survey questionnaires, and several hundred hours of videotaped sessions. The survey questionnaires, the analytic framework for the conversational exchanges, and the statistical treatment of these data are detailed in the Sun Laboratories technical report (Sipusic et. al., 1999). We used this data to search for evidence of process differences

between the face-to-face and DTVI conditions that might illuminate the impact of the technology on collaborative learning, and summarize some of the findings here.

Among other things, survey questionnaires assessed student satisfaction. Both TVI and DTVI students reported higher levels of satisfaction with collaborative learning than with traditional course methods. There was no difference in self-reported learning or likelihood of recommending the method to others. However, the TVI students reported enjoying the process slightly more and were slightly more likely to say they would take another course using this method.

We hypothesized that the DTVI technology would influence discourse patterns. However, based on statistical analysis of the conversational exchanges, we found there was a remarkable similarity in discourse patterns between TVI and DTVI. Where we were able to detect small differences, they were in the opposite direction of our original hypothesis. Students in the TVI and DTVI conditions spent the same total amount of time in conversation, but DTVI groups had more conversational turns than their TVI counterparts. We described them as "chattier." DTVI groups also reported higher amounts of humor. Most of this increased interaction occurred in conversation topics outside the content domain. When limiting our analysis to exchanges with positive ratings for content, the only statistically significant difference was an increase in tutor questions in the DTVI groups.

We hypothesized that the technologically mediated interactions would cause the DTVI students to exhibit lower levels of group cohesion. But explicit measures of group cohesion showed no difference between TVI and DTVI groups. For instance, there was no difference in tutors' average group cohesion ratings of conversational exchanges. Nor did students report differences in cohesion in survey questions.

All in all, we conclude that the communication perturbations of video-mediated communication using DTVI are both less evident and less salient than might have been expected based on past research. Whatever "noise" DTVI added to the communication process was subtle enough to not be explicitly recognized by students and to have minimal impact on social discourse. But some process differences did show up in student behavior and attitudes DTVI students showed signs of a kind of disengagement, rating both facilitators' content knowledge and course difficulty lower than their TVI counterparts. Finally, DTVI groups showed evidence of a weakening of the discursive reciprocity covenant (one's obligation to sustain and attend to conversation). This may be because of factors such as gaze ambiguity in the DTVI condition. The impact of this reduced social presence and weakening of the reciprocity covenant is mixed. Lower reciprocity may have made it easier for DTVI students to hide or do other work instead of attending to the discussion. But it may also have made it easier for them to publicly take intellectual risks in front of their peers.

Enhanced DTVI

Towards the end of our main two-year experiment, our preliminary study of student grades made it clear that DTVI and TVI would probably be comparable and both would do better than lecture. Since computers can do more than ship video and audio among students, we naturally began to wonder how we might enhance the DTVI interface by supporting the students in other ways. The DTVI system can be immediately applied to many different courses in contrast to most educational software systems requiring specially authored software content for each course. We decided to build a generic shared note taking tool (Figure 2, 3), so that we could retain this advantage of generality.
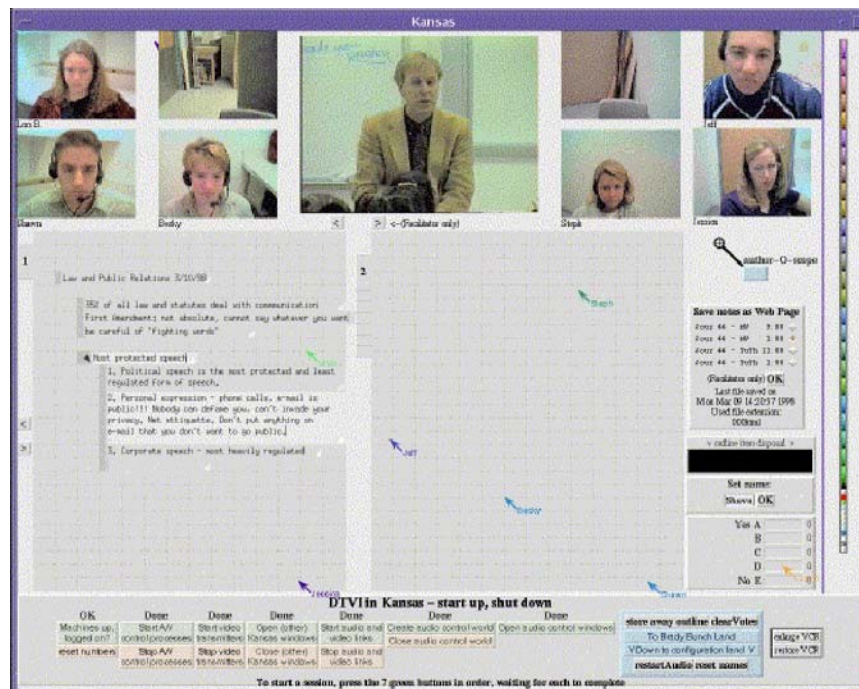

Figure 2. The Enhanced DTVI interface. In addition to the audio and video links among participants, the enhanced version includes a shared outlining tool (shown in close up in Figure 3) with which students can take notes and publish them to the web. The video windows for the students and facilitator are smaller than their DTVI counterparts, though the videotape window is comparable in size.

We used the Kansas system (Smith et. al., 1998; Smith et. al., 1997; Maloney & Smith, 1995), a toolkit for building shared 2-D worlds, to create the outlining facility. Both video and audio in Kansas are digital multicast technologies, whereas the main experiment used analog audio and video. (Students in these more informal digital studies were not part of the main experiment discussed above.) Students initially complained about the tendency for the digital system to crash, but with continued to improvements to the system, we were able to get it to run continuously for weeks and sometimes months without problems. Due to our use of compression, the digital image quality was slightly

lower than that of the analog system, but we believe such differences are small compared to the differences between a virtual system and co-present collaboration.

The outlining tool lets students take notes by typing onto a gridded page. The resulting text box can be manipulated to create a hierarchical outline. At the end of the session, the facilitator presses a button to generate a web page based on the notes. The students can then review the notes from their web browser at home, or can even view the notes of other sections to see how they differ. The instructor of the course can also visit the page, and make comments to be edited into the page in red text.

This was a small informal study, as our focus was on trying to tune the technology. We set up a preliminary version of the enhanced system mainly as an optional adjunct for one semester (with a total of 32 students in the Enhanced DTVI sections) to debug and assess usability. In a later semester we used the system as a full-blown basis for the sessions in a Journalism class at Chico. The Journalism class had 11 Enhanced DTVI students in 3 sections. Based on surveys and observations, we found that groups informally negotiated their own turn taking protocols, so that normally only one person was typing at a time. They sometimes used the note-taking tool to type comments to each other outside of the verbal channel, though this was uncommon. Interestingly, they generally did not see the group notes as a replacement for their own notes, preferring to have both. The instructor and students were generally enthusiastic about the system. The average of the Enhanced DTVI student grades was again higher than the average of the lecture student grades but statistical claims are weak with this small sample size. Unfortunately, students didn't often visit the web-page notes of other sections as we had hoped. But students particularly liked reading the instructor's comments scattered among their group's notes.

The study suggests a next step in which the Enhanced DTVI process and grades are compared with regular DTVI. Unfortunately, our time and resources were limited, so this system for now simply suggests an encouraging approach to the next generation of virtual collaborative learning.
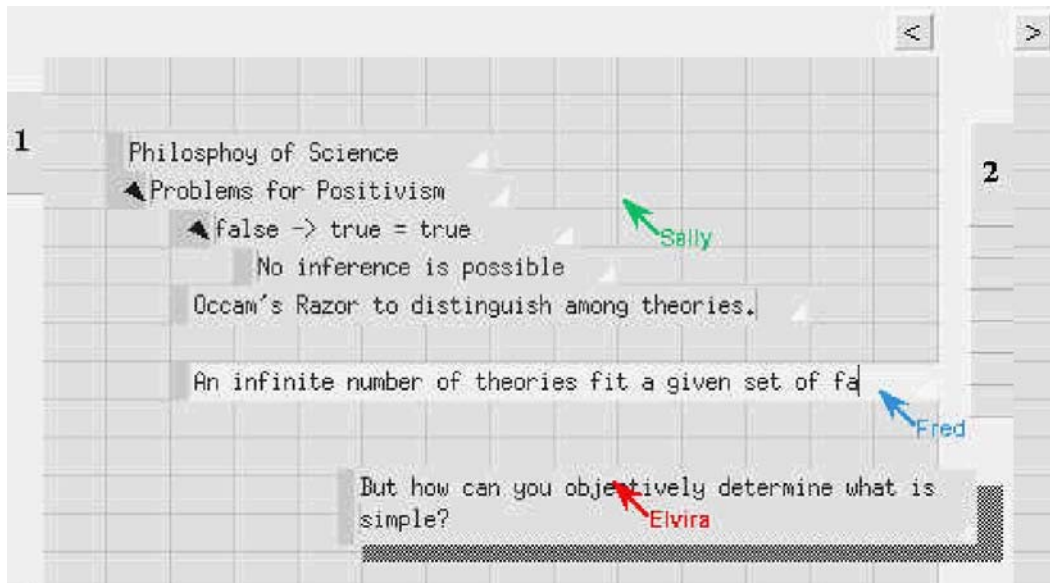
Figure 3. The shared outlining tool enables anyone to type anywhere, creating small text boxes on a gridded page. The text boxes can be picked up from one place, carried and dropped into another.

Conclusions

There is considerable prior research suggesting that communication technology can interfere with discourse. Since collaborative learning is dependent on discourse both for content and for group cohesion, we set out to determine whether the collaborative learning effect would be maintained in a virtual collaborative learning environment. Our analysis of over 700 student grades suggests that the collaborative learning effect is fully intact with DTVI virtual environment, opening the door to the widespread use of more effective distance learning models than the lecture-based model currently being used.

There are many other lessons to be learned from this research project. These lessons are particularly compelling because of the size, duration, and real-world nature of the experiment.

We cannot make a one-to-one match between elements of the DTVI interface and their impact in terms of social process. But we can say on the whole that the DTVI experience did not provide quite the same degree of "warm fuzziness" that face-to-face interaction provides. We do not know whether it is possible or—given the potential benefits of reduced social proximity (i.e., greater willingness to admit gaps in knowledge, etc.)—even desirable to create a virtual collaboration environment that does provide an equivalent feeling of human contact. But we can say that virtual collaboration maintains a surprising amount of the group cohesion of face-to-face groups and that the collaborative experience is still very rewarding for participants, albeit perhaps in a different way than face-to-face collaboration.

Most importantly, our research shows that video-mediated communication can in fact support both the content and relational components of discourse necessary for effective collaborative learning. In particular, the DTVI environment appears fully capable of reproducing the collaborative learning effect to the extent that learning is measured by university course grades.

Furthermore, we have demonstrated that video-mediated collaboration can generate high levels of user satisfaction. While the DTVI students reported enjoying their experience slightly less than the TVI students, they reported enjoying it much more than a typical classroom lecture.

With DTVI generating higher academic performance and more enjoyment than classroom lecture, distance learning no longer need be considered a poor cousin to face-to-face instruction.

This study opens up many areas for future research. We do not know, for instance, to what extent eliminating network transmission artifacts contributed to our finding the relatively small amount of communication disruption in the DTVI and Enhanced DTVI environments. While much previous research has been done on the effects of latency, audio quality, etc., most of this research has focused on explicit, observable communication breakdowns. The effect of these breakdowns on group cohesion and other social aspects of collaborative discourse deserves more study. We also should point out that, in this experiment, each DTVI student was in a fairly quiet and interruption-free small room or cubicle. Should DTVI be brought to the user's everyday desktop, the distractions of office or home life might diminish one's ability and willingness to attend to the group.

Lastly, the positive response to the enhanced note taking facility suggests that such features may help the collaborative learning process. Facilities such as shared notes provide a common focus that bears directly on the task of attending to and thinking about the videotaped material, and the notes can be referenced for future study so group members have an interest in seeing they are done well. Perhaps other features could be added to specifically target group cohesion. Such enhancements suggest how virtual collaborative learning may be able to produce a stronger collaborative learning effect than a face-to-face collaboration.

Bibliography

Cohen, E. (1994) Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research. 64*(1): 1-35.

Colston, H. L. & Shiano, D. J. (1995) Looking and lingering as conversational cues in video-mediated communication. In *CHI '95 conference companion*. 278-279. Denver, Colorado.

Fish, R. S., Kraut, R. E. & Chalfonte, B. L. (1990) The VideoWindow system in informal communications. In *Proceedings CSCW '90*. 1-11. Los Angeles, CA.

Fish, R. S., Kraut, R. E. & Root, R. W., & Rice, R. E.(1992) Evaluating video as a technology for informal communication. In *Proceedings CHI '92*. 37-48. Monterey, California.

Gale, G. (1998) The effects of gaze awareness on dialogue in a video-based collaborative manipulative task. In *Proceedings CHI '98 Summary*. 345-346. Los Angeles, California.

Gibbons, J. F., Kincheloe, W. R., & Down, K. S. (1977) Tutored videotape instruction: a new use of electronics media in education. *Science*. 195: 1139-1146.

Isaacs, E. A. & Tang, J. C. (1993) What video can and can't do for collaboration: A case study. *Proceedings of Multimedia*. 93: 199-206. New York: ACM Press.

Johnson, D.W. & Johnson, R. (1994) *Joining together: group theory and skills*. 5th ed., Englewood Cliffs, New Jersey: Prentice Hall.

Kies, J.K., Williges, R. C. & Rosson, M. B. (1996*) Controlled laboratory experimentation and field study evaluation of video conferencing for distance learning applications*. Hypermedia technical report, HCIL-96-02 (http://hci.ise.vt.edu/~hcil/htr/HCIL-96-02/HCIL-96-02.html).

Maloney, J. & Smith, R. B., (1995) Directness and Liveness in the Morphic User Interface. In *Proceedings of the UIST '95 Conference*. 21-28. Pittsburgh, Pennsylvania.

Monk, A. & Watts, L. (1995) A poor quality video link affects speech but not gaze. In *CHI '95 conference companion*. 274-275, Denver, Colorado.

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

Rocco, E. (1998) Trust breaks down in electronic contexts, but can be repaired by some initial face-to-face contact. In *Proceedings CHI '98*. 496-502, Los Angeles, California.

Sellen, A. (1992) Speech patterns in video-mediated conversations. In *Proceedings CHI '92*. 49-59. Monterey, California.

Sipusic, M., Pannoni, R.., Smith, R. B., Dutra, J., Gibbons, J. F., & Sutherland, W. R. (1999) *Virtual Collaborative Learning: A Comparison between Face-to-Face Tutored Video Instruction (TVI) and Distributed Tutored Video Instruction (DTVI)*. Sun Microsystems Laboratories Technical Report TR-99-72.

Smith, R. B., Hixon, R., & Horan, B. (1998) Supporting Flexible Roles in a Shared Space. In *Proceedings of Computer Supported Cooperative Work '98*. 197-206. Seattle, Washington.

Smith R. B., Wolczko, M. & Ungar, D. (1997) From Kansas to Oz: Collaborative Debugging when a Shred World Breaks. In *Communications of the ACM.* (40) 4: 72-78.

Storck, J. & Sproull, L. (1995) Through a glass darkly: what do people learn in videoconferences? *Human Communications Research.* 22(2): 197-219.

Authors' addresses

Randall B. Smith (randall.smith@sun.com)
Sun Microsystems Labs; 901San Antonio Rd.; Palo Alto, CA 94303. Tel. (650) 336-2620.
Michael J. Sipusic (sipusic@krdl.org.sg)
21, Heng Mui Keng Terrace; Singapore, 119613. Tel. +65 874 6242
Robert L. Pannoni (pannoni@raport-sys.com)
Rapport Systems; 160 W. Campbell Ave. #364; Campbell, CA 95008. Tel. (408) 871 251