Distance Learning Through Distributed Collaborative Video Viewing

JJ Cadiz, Anand Balachandran, Elizabeth Sanocki, Anoop Gupta, Jonathan Grudin, and Gavin Jancke

May 10th, 2000

Technical Report
MSR-TR-2000-42

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Distance Learning Through Distributed Collaborative Video Viewing

JJ Cadiz, Anand Balachandran, Elizabeth Sanocki,
Anoop Gupta, Jonathan Grudin, Gavin Jancke
Microsoft Research, Collaboration & Multimedia Group
One Microsoft Way, Redmond, WA 98052 USA
+1 425 705 4824
{jjcadiz; anoop; jgrudin}@microsoft.com

ABSTRACT

Previous research on Tutored Video Instruction (TVI) shows that learning is enhanced when small groups of students watch and discuss lecture videos together. Using specialized high-end videoconferencing systems, these improved results have been shown to apply even when the students are in different locations (Distributed TVI, or DTVI). In this paper, we explore two issues in making DTVI-like scenarios widely supported at low cost. First, we explore design of a system that allows distributed individuals to collectively watch video using shared VCR controls such as play, pause, seek, stop. We show how such a system can be built on top of existing commercial technologies. Second, we explore the impact of four alternative discussion channels on student learning and interaction behavior. The four channels—text chat, audioconferencing, videoconferencing, and face-to-face—have differing infrastructure requirements and costs. Our lab studies show that while text chat does not work, there is no significant difference in discussion behavior and learning between audioconferencing and videoconferencing. While lab studies have their limitations and long-term field studies need to be done, the preliminary results point to a low-cost way for a DTVI-like model to be deployed widely in the very near future.

Keywords

Distributed tutored video instruction (DTVI), Tutored video instruction (TVI), videoconferencing, distance learning

1 INTRODUCTION

With rapid changes in technology, what we learn today becomes obsolete or irrelevant within a short time span. We need to continuously update our skills, and cost-effective support for such lifelong learning is a key challenge facing our education systems. To address this problem, one solution adopted by many universities has been to broadcast classes via TV networks. For example, Stanford has pursued this strategy successfully for over 25 years. However, for many working students, this synchronous model of learning conflicts with their work commitments. The synchronous model is also not very scalable: even if the lecture could be broadcast to a million people concurrently, there is little chance that interesting interaction would happen between students and instructors in such an environment.

The flexibility and scalability issues can be resolved by using video streaming technologies and the Internet to create an on-demand, anytime, anywhere learning model. However, research shows that learning can suffer when students watch a lecture video individually. This concern is identified and addressed in a classic study reported in *Science* by Gibbons et al. [6]. As a solution, they propose the Tutored Video Instruction (TVI) model, wherein remote students watch lecture videos in small groups with a discussion facilitator. As they watch the video, they periodically pause and discuss it. The study shows that the TVI students outperform students who attended the live lectures in the classroom, as well as students who watched lectures live from remote locations and students who watched lecture videos individually (Table 1).

The advantage of TVI over classroom attendance was replicated at the University of Massachusetts, where some student groups met without tutors [17]. The advantage of TVI was also found by researchers at Sun Microsystems [15, 16]. Smith, Sipusic and their colleagues also tested an extension of TVI called Distributed Tutored Video Instruction (DTVI), which does not require students to meet in the same room. The researchers had groups of up to seven students watch and discuss pre-recorded lectures from multiple locations. In addition to the lecture video, students were connected with a high-quality, low-latency audio and video connection. In experiments involving six university courses and several hundred students, course grades of DTVI and TVI students were better than those of students who attended the live lectures and were indistinguishable from one another. The success of DTVI and TVI is attributed to the collaboration and discussion that occurs among the students watching the video.

Condition (# of students)	Average GPA (on 4.0 scale)
Campus (302)	3.4
Live Video (55)	3.2
Videotape: no tutor (6)	3.02
Videotape: with tutor (27)	3.7

Table 1. Results from Stanford's 1977 TVI Experiment. Students using TVI outperform live campus lecture attendance and other conditions (adapted from [6]).

DTVI is clearly a desirable model in many ways. It allows students to participate from anywhere, thus eliminating location constraints. It allows students to participate almost anytime, subject to finding a few partner students willing to participate at the same time. It is scalable, as hundreds or thousands of such small groups can exist and make progress independently. Best of all, it provides all these features while statistically showing better learning outcomes than being in the classroom.

Given the desirability of DTVI, this paper extends earlier work in two principal directions, allowing for wide-scale deployment of DTVI. First, a key component of DTVI is the ability to collaboratively view lecture video. For example, when one student presses the pause button, the video should pause for all remote students. The "what you see is what I see" principle needs to be preserved for all group members. Existing collaboration tools (e.g., NetMeeting) and streaming media players (e.g., Windows Media Player and Real Player) do not support such collaborative video viewing. We discuss why and show how existing commercial systems can be extended to support this functionality.

Second, and more importantly, DTVI results presented by Smith and Sipusic et al [15, 16] were based on the remote participants having a specialized, high-quality videoconferencing system, e.g., the system presented eight remote videos concurrently on a monitor, each at 30 frames per second, while also encoding the video of the local participant. Such infrastructure is unlikely to be available to most students, and past literature shows little benefit from including talking-head video in somewhat similar contexts [1, 2, 3, 4, 5, 7, 11, 12]. Consequently, in this paper we study the question of how DTVI is affected if the communication channel is less rich. Does it affect learning and interaction behavior? We present results from a lab study using four alternative channels: text chat, audio conferencing using telephone, video conferencing, and face-to-face discussion.

Our conjecture was that audio conferencing would provide most of the benefits, and expensive video conferencing support is not needed. Our lab study results confirm this hypothesis, thus encouraging larger-scale field trials of DTVI with simpler and highly available telephone channels for discussion.

1.1 Terminology

The acronyms TVI and DTVI, as the full form implies, assume the presence of a tutor to facilitate discussion among the students. We follow the Univ. of Massachusetts study model [17], where a tutor is not present and the students facilitate the discussion themselves. To avoid confusion about presence/absence of tutor, we use the following terminology in this paper. We use CVV (Collaborative Video Viewing) to be the same as TVI with the exception that no tutor is present. Similarly, we use DCVV to be same as DTVI sans tutor. We further refine DCVV into DCVV-chat, DCVV-audio, and DCVV-video depending on whether text chat, audioconferencing, or videoconferencing are used as the communication channel.

1.2 Paper Organization

In the next section we discuss why current technologies do not allow DCVV and how our system is implemented. Next, we outline the methodology and results for our first and second studies. Finally, we discuss the implications of our work for distance education in both companies and universities.

2 THE DISTRIBUTED COLLABORATIVE VIDEO VIEWING SYSTEM

Enabling the DCVV scenarios requires two components:

1. A distributed lecture video viewing system with shared VCR controls (play, stop, pause, seek).
2. A communication system for discussion around the video content.

Although the DCVV requirements are quite basic in many ways, today's streaming media products and application sharing products do not support it.

2.1 Distributed Video Viewing: Problem

Streaming media players (e.g., Microsoft Windows Media Player [8] and the Real Video Player [14]) allow multiple people to watch the same stored video from a shared video server. However, users do not share the VCR controls. Students could say "pausing video" on the conference call when pausing the player, but remote students will end up pausing their video players at slightly different times. In addition to being a cumbersome process, over multiple stop-starts, they can easily get out of sync.

One would expect application-sharing products such as Microsoft NetMeeting, Lotus Sametime, and WebEx.com to come to the rescue. Application sharing allows participants in a session to allow others to see and even control their desktop applications. For example, multiple students should be able to start NetMeeting and establish a common "call" to get into the same session. One student should be able to bring up the Windows Media Player, share the Media Player application to others (the others will now have the media player on their desktop), then load the URL for the lecture video and press play, and all others should see the video in their media players. Given application sharing, everyone should be able to pause, play, and seek the video with everyone remaining in sync.

Unfortunately, this almost works, but not adequately. The remote participants can see and control the video, but it is low quality, jittery, and unacceptable. Instead of the video data stream going from the video server to all the participants and being independently decoded and rendered on their PCs (shown in Figure 1), the video stream goes only to the student who starts the media player (say, student-1) (see Figure 2). It is decoded and rendered on student-1's machine, and then pixel-level screen changes are picked up by the application sharing system (e.g., NetMeeting). These are encoded using their own algorithm and sent to the other participating students' desktops. None of the application sharing products is optimized for this scenario (they are targeted towards transmitting infrequent changes to screen as may occur during web browsing), the performance is dismal compared to highly optimized video codecs, and the network bandwidth consumed is large.

Thus, although application sharing allows users to effectively share documents, presentation files, spreadsheets, and web pages, current systems do not work for dynamic media such as audio and video.

2.2 Distributed Video Viewing: Solution

To solve the above problem, what we need is *partial* sharing of the media player application. We want the video to stream independently to each students' media player and be decoded and rendered there (as in Figure 1), but we want to have the VCR control UI of the media player to work via NetMeeting's application sharing.

To implement the desired solution, we needed to combine the functionality of the Windows Media Player and NetMeeting. We did this in a way such that third parties outside of Microsoft can also achieve the same results.

2.2.1 User Interface

Before we discuss the implementation, we briefly discuss the UI of the resulting application (Figure 3). The system has an interface derived from its two parent applications.

The middle part of the DCVV window is composed of the standard Microsoft Windows Media Player UI. It displays the video image along with the basic VCR controls (play, pause, stop, and seek). All events generated by the VCR controls are captured by the UI shell surrounding the media player and transmitted to remote participants' media players using the T.120 data channel provided by NetMeeting.

The top slim menu bar is a new addition for conferencing. Under the *File* menu, the user finds all the typical functions relating to finding and playing a video. If one user opens a new video, the same video is opened on all remote players. The *Conference* menu provides methods to initiate a text chat, shared whiteboard, or file transfer with other participants. The portion below the standard media player is also added for conferencing. It provides UI to initiate a call to other participants and displays current attendees in the conference.

2.2.2 Implementation of the DCVV System

Our implementation combines the Windows Media Player and NetMeeting in a manner that can be duplicated by others.

While NetMeeting is a standard application provided by Microsoft [9], it is also available as a SDK (a Software Development Kit). The NetMeeting SDK [10] exposes a set of APIs that allow third-party programmers to design collaborative software. Using these interfaces, developers can create collaborative applications that involve multiple participants working within or across firewalls. Local instances of such an application, running on each participant's computer, can communicate using the NetMeeting data channel, thus making the windows application "conferencing-aware." The data channel is an abstraction of the IEEE T.120 data-conferencing interface standard. Data sent or received over the T.120 interface gets transmitted to other users over the Winsock API or any other transport utility.

The task for us then was to make the Windows Media

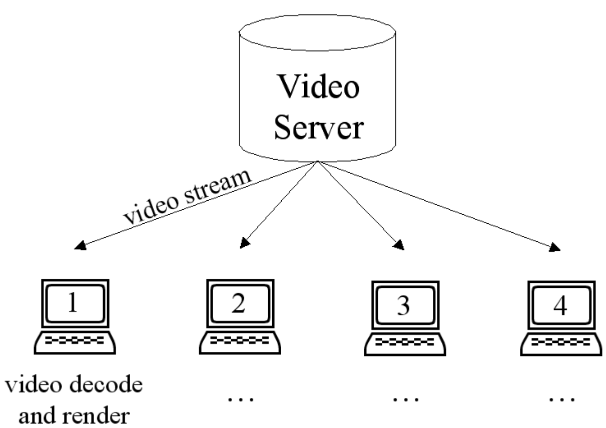

Figure 1: Conventional Video Flow. Flow of video when each Media Player independently connects to the video server.

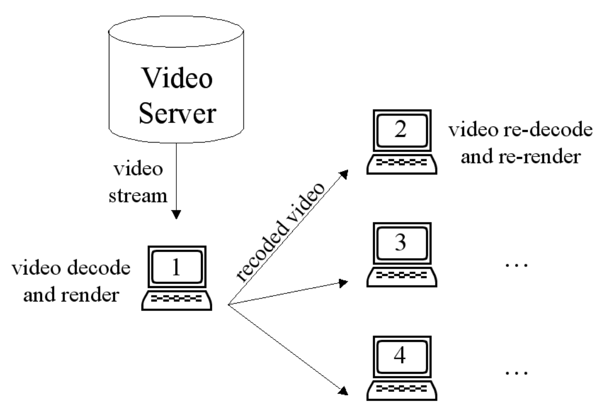

Figure 2: Video Flow under NetMeeting. Flow of video when media Players are shared under NetMeeting.

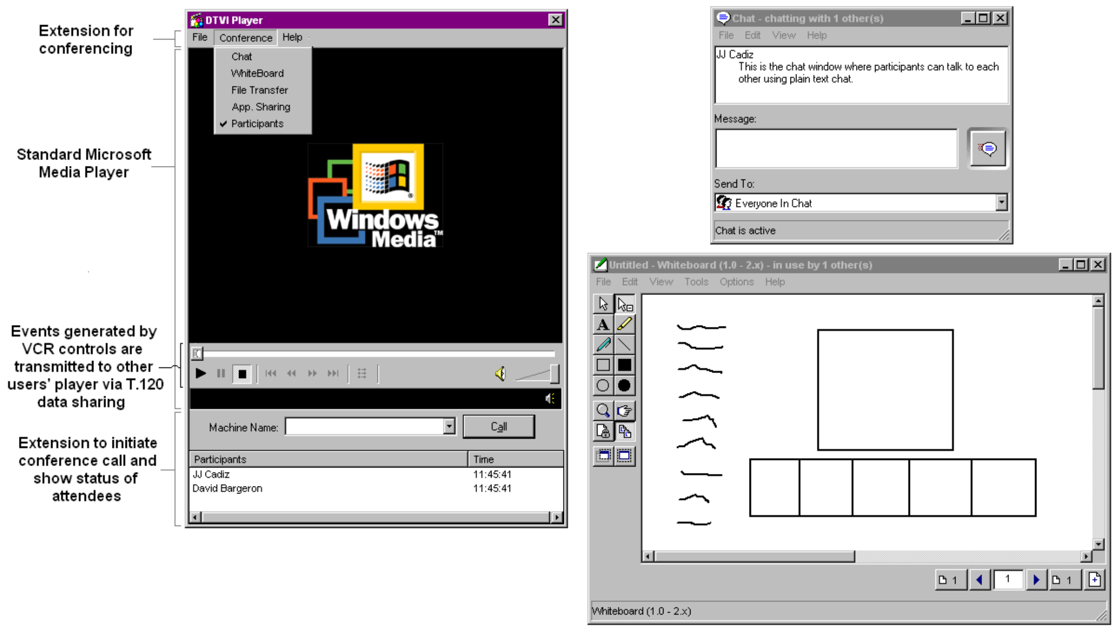

Figure 3: DCVV User Interface. The modified media player used by the participants is shown on the left. With NetMeeting, users can invoke conferencing utilities such as chat and a shared whiteboard, as shown on the right. (The shared whiteboard was not used by participants in our study.)

Player “conference aware”. The Media Player itself, in addition to being a free-standing application, is designed as an ActiveX control that can be embedded in other applications. The application shell that hosts it is referred to as the “container”. The Media Player exposes an automation API, which fires appropriate events that can be captured by the container application. Occurrence of events, e.g., use of play, pause, stop, fast-forward, seek buttons, is communicated to the container.

The “container” application we built is the DCVV media player (shown in Figure 3). It includes the Windows Media Player control inside it, and it also links to the NetMeeting SDK. It implements UI and code that allows users to open shared media files, initiate connections to remote participants, etc., as shown in the UI. In addition, it also captures all interesting local events and communicates them to the remote DCVV application. When these events are received by the remote DCVV application, they are parsed and then applied to the local media player or used to invoke the appropriate NetMeeting SDK function.

Thus when the first student opens a lecture video file, this action is communicated to all remote participants, and their media player also automatically opens the same lecture video. Similarly, when a student presses the “play” button, this event is communicated via the NetMeeting data channel to all participating students.

The packet format varies according to the nature of the event being transmitted. For example, a *play* event packet contains the name of the stream and the time relative to the start of the video at which the play occurred. In the case of a *seek*, the event records both the initial and final positions of the slider bar using timestamps relative to the beginning of the video.

One problem resulting from the distribution of control is that *race conditions* can occur when multiple participants concurrently press the VCR control buttons. As a concrete example, assume the video is playing at offset T=30sec, and simultaneously student-1 pushes pause and student-2 seeks to offset T=60sec. Given the delay that event messages take to propagate over network, student-1’s media player will first take pause action and then receive seek to T=60 sec, and start playing from that time offset. Student-2’s media player will first seek to T=60 sec and start playing from there, but very shortly later get the pause message, and then pause the media player. When this happens, the two media players are not in a consistent state—student-1’s player is continuing to play while student-2’s is paused.

These race conditions are inherent in distributed systems, and many classes of solutions are possible. It is necessary to give an *apparent* order to all commands being issued, but without first passing them through a central serializer to avoid making the user interface feel very sluggish. In our system, we introduced a notion of ordering of the transmitted messages using the timestamps of the sender’s video stream in each of the messages. Since the Media Player records the video timestamps relative to the start of the video stream, these timestamps are absolute and not influenced by the system clock. Any player that receives a message with a timestamp earlier than the one it had recorded on its last sent message ignores the received message. This ensures that all the players record the message that has the latest timestamp relative to the start of the video stream.

However, the *stop* message is an exception because stopping the Media Player takes the stream pointer to the start of the video, and a subsequent replay starts the video stream all over again. To prevent this from causing further asynchrony we handled the *stop* case differently. When a stop message is received, the player stops regardless of any messages that it broadcast to the other participants. This ensures that stopping one player causes everyone to stop, thereby resetting each player to the start position.

2.3 Supporting Discussion Around the Video

As mentioned in Section 1, we wish to explore the impact of three digital communication channels. The first, text chat, was supported using the standard chat available with NetMeeting. The second, audioconferencing, was supported using standard telephones with speakerphones.

The third, videoconferencing (transmission of live video of all participants) required extra work. A separate, special-purpose application was built to allow participants to see each other. This application only transmitted video; audio was provided by speakerphones or headsets (discussed below). This window displayed medium quality video (full color, 10 frames/second) of all group members. Each group member's image was approximately 2.25 inches (6 centimeters) wide and 1.75 inches (4.5 centimeters) tall. Video was captured using a camera on top of each participant's monitor.

3 STUDY FOCUS AND METHODOLOGY

A major goal of this research was to examine how various communication channels affected the DCVV experience. Specifically, we were interested in determining whether videoconferencing was worth the high cost relative to audioconferencing and text chat.

We examined three major sets of variables to determine the utility of these communication channels. First, because DCVV was designed for educational contexts, we measured participant **learning and comprehension**. Second, because of the tight relationship between learning and discussion in the TVI literature, we measured various aspects of the **group interaction and discussion**. Finally, we wanted to know if the DCVV experience was any less enjoyable than the CVV experience, thus we measured participant **satisfaction**.

3.1 Methodology

We conducted two lab studies to examine the effects of these communication channels. The first study served as a pretest to refine our methods prior to running the larger second study.

3.1.1 The Videos

For both studies, participants watched videos from the Harvard Business School Management Productions "People, Service, Success" video series (1993). This series of five videos discusses the importance of good customer service and its role in profitable, successful companies. These videos were of professional quality, unlike earlier DTVI studies based on videos taped directly from classroom lectures.

On average, the videos were originally 32 minutes long but were accelerated via time compression [13] and played at approximately 113% of their normal speed. No participants commented that they felt the video was playing too quickly; in fact, some commented that they liked the quick pace of the video.

3.1.2 Participants

Participants were recruited from the Seattle, Washington, community. All were intermediate computer users or better. Because we wanted to ensure that participants would enjoy the video and be able to discuss it, we required that all participants come from service-related jobs (for example, salespeople, customer service representatives, real estate agents, small business owners, etc.). Participants received a free software product for their time.

3.1.3 DCVV setup

For the DCVV conditions, each participant was seated alone in a room with a PC. For the conditions using audioconferencing, participants in the first study used speakerphones. Unfortunately, while the video was playing, sound from each computer was relayed through the speakerphones, causing an unacceptable distraction. This problem was resolved in the first study by having participants mute their speakerphones while watching the video. In the second study, participants used hands-free telephone headsets with directional microphones, which dramatically—but not completely—reduced the noise received from other participants' computers.

The desktop arrangement for DCVV-video participants is shown in Figure 4. The same setup was used for DCVV-audio participants, without the videoconferencing window.

3.1.4 CVV Setup

For the CVV condition, participants were seated together at a table. The video was played using the same software as the DCVV condition, except the computer's screen was displayed on a television at the head of the table. The person who sat closest to the PC used the mouse to pause and play the video during discussions. Group members not seated near the PC would ask this person to pause the video if they had something to say; often they would just begin talking and the person would automatically pause the video.

3.1.5 Scenario & Quizzes

Participants were asked to play the role of a student. Specifically, they were told that they had missed the last lecture of their course where the professor had played a video. To make up the missed course, they were watching the video together with other students who also missed the lecture. Participants were encouraged to pause the video and discuss it with each other. Participants were also provided with pen and paper to take notes.

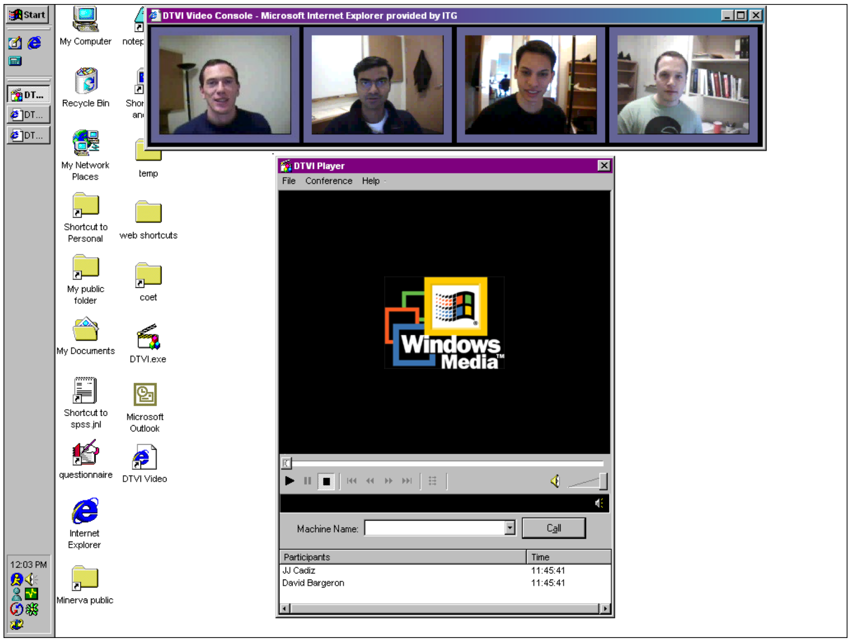

Figure 4: The desktop setup that DCVV participants used. Half of the DCVV participants were provided with the videoconferencing application shown at the top of the screen.

After watching the video, participants were given a multiple choice quiz of the video's content. Participants were not told ahead of time that they would be taking a quiz, but they were allowed to refer to the notes they took during the video while answering the questions.

4 THE FIRST STUDY (PRETEST)

Eleven participants were recruited for the first study. Participants were randomly divided into two groups of four people and one group of three people. A within-subjects design was employed where each group watched four videos, each time using a different experimental condition. The four conditions were:

TVI: all participants watched the video on a television in the same room.

DTVI-chat: all participants watched the video in different rooms and could communicate with each other using text chat.

DTVI-chat-audio: all participants watched the video in different rooms and could communicate with each other using text chat and audioconferencing.

DTVI-chat-video: all participants watched the video in different rooms and could communicate with each other using text chat and videoconferencing.

In each of these conditions, participants watched half of a video (about 15 minutes) and usually discussed it for about seven minutes. The entire experiment lasted about three hours for each group. A researcher participated in each discussion as a "tutor" to make the sessions similar in format to the methods of TVI and DTVI.

The central purpose of the first study was to refine our design prior to the larger second study. Three major observations informed the redesign. First, watching a 15 minute video while using the DCVV system did not provide enough content for the participants to have meaningful discussions about the video. Participants did pause the video to talk about it, but the conversations seemed to be of limited value given the brevity of the video. Thus, the second study used a between-subjects design where each group watched one entire video and experienced only one condition.

Second, the tutor did not seem to add much value to the discussions. In traditional TVI and DTVI sessions, tutors were typically graduate students whose purpose was to facilitate discussions among college students. However, in our pretest, the participants were typically older and did not need much encouragement to discuss the video (especially since they were recruited because they had backgrounds in customer service, which was the topic of the video). Thus, no tutor was used in the second study.

Finally, the most significant finding was that text chat was not a helpful medium to support the kind of interactive communication that makes DTVI successful. When given only plain text chat to communicate, none of our groups paused the video even once to discuss it. Videotape from these sessions provides a possible explanation. With audio or videoconferencing, it is not possible to talk while the video is playing because it is too difficult to listen to both sources at once. However, when using only text chat, participants multitasked: they felt that they could type and watch the video at the same time. But in the post-session survey, some participants commented on the problems with this strategy:

> *Although it was easy to use the chat window, it took away from what was going on in the video, thus it was too distracting.*
>
> *The chat function was difficult for me. I was caught between concentrating on the video and concentrating on a statement I was writing in the chat box.*
>
> *[I disliked] trying to read the comments and listen to the video. I missed some of the key points that were asked in the quiz … The tape should have been stopped so that more of a discussion could have happened and to express the thoughts completely.*
>
> *Typing through the chat window was a bit distracting and I didn't feel like I was able to pay as much attention to the video content as I would have liked and still keep up with the conversation.*
>
> *VERY HARD to pay attention to what was on the tape and reading or writing comments at the same time.*

Thus, even though participants recognized that chatting while watching the video was distracting, they never chose to pause the video, perhaps because of the awkwardness of pausing the video and then making everyone wait while you type your comment. For this reason, we eliminated this condition from the second study.

5 THE SECOND STUDY

After refining our methodology from the pretest, we recruited 90 participants for the second study. Participants were placed in groups of four and randomly assigned to one of three conditions:

> CVV: all participants watched the video on a television in one room.
>
> DCVV-audio: participants watched the video in different rooms using the DCVV software and could talk with each other via telephone.
>
> DCVV-video: participants watched the video in different rooms using the DCVV software, could talk with each other via telephone, and could see each other using a videoconferencing window.

Each group watched one 30 minute video, and each session lasted less than 90 minutes. Data from sixteen groups were analyzed: six CVV groups, six DCVV-video groups, and four DCVV-audio groups. Two DCVV-audio groups were dropped because of participant no-shows. All groups were comprised of either three or four participants (also due to no shows), with the exception of one CVV group that had six people due to a scheduling error.

Some of the following analyses are based on self-report measures. All of these measures were obtained using 7-point Likert scale questions. In many cases, variables were created by combining several questions using factor analysis. All analyses were done at the group level.

5.1 Comparing face-to-face and distributed groups

Our first set of analyses examined the differences between DCVV and CVV groups. The results, in Table 2, take into account the number of people and a pretest measure of involvement for each group. All analyses were done at the group level.

5.1.1 Learning and Comprehension

Both CVV and DCVV groups had essentially identical quiz scores (69% and 71%). In addition, the groups felt that they understood the video equally and found their teammate's comments to be equally helpful.

5.1.2 Discussing the Video

The number of times the groups paused the video was not significantly different, and both groups thought the number of times the video was paused was about right. Despite this, CVV groups reported being significantly more comfortable with pausing the video, ($t(15) = 3.6$, $p = 0.004$). Similarly, when asked about their comfort with asking questions and making comments, CVV groups were significantly more comfortable than DCVV groups ($t(15) = 2.6$, $p = 0.03$).

CVV groups also spent on average 10.9 minutes longer talking when the video was paused, which is significantly longer than DCVV groups ($t(15) = 3.2$, $p = 0.008$). Average length of discussion per pause was also examined, and it was found that the difference between CVV and DCVV groups was also significantly different ($t(15) = 3.2$, $p = 0.009$). Despite these differences, both groups of participants thought that the amount of time spent talking was about right. Furthermore, once the video had finished, both groups talked for about the same length of time.

5.1.3 Satisfaction with Experience

We examined participant satisfaction by using three measures: overall satisfaction, boredom, and satisfaction with the group discussion. DCVV groups were marginally more satisfied overall ($t(15) = -1.7$, $p = 0.11$) and significantly less bored ($t(15) = 2.2$, $p = 0.05$) than CVV groups. Both groups were equally satisfied with the discussion.

However, based on discussions with the participants and some free response survey data, we believe the satisfaction measures for the CVV groups are artificially low. Our intuition is that because the study was conducted by Microsoft Research, participants were expecting to see "futuristic" or "cool" technology. Instead, CVV participants sat around a table and watched a video together. In the post-experiment survey, one CVV participant wrote,

> *What we did was VERY contrary to what our expectations were for today. I came concerned about my level of knowledge in relationship to what I might be expected to do on the computer and ended up watching and discussing a video!*

Another CVV participant wrote,

	Variable, and number of questions used to measure variable	Score explanation	CVV (n = 6)	DCVV (n = 10)	p	t	df
Learning	Quiz score (ten questions)		69%	71%	0.61	0.52	15
	Feelings of video comprehension (two questions)	1=low, 7=high	5.1	5.6	0.31	-1.1	15
	Did group members think their teammates' comments were helpful? (two questions)	1=least helpful, 7=most helpful	5.1	5.5	0.22	-1.3	15
Discussion	Number of times video was paused		6.3	4.4	0.26	-1.2	15
	What did participants think about the number of times video was paused? (one question)	1=too few, 4=just right, 7=too many	3.9	3.8	0.15	1.6	15
	Overall comfort with pausing the tape (two questions)	1=uncomfortable, 7=comfortable	6.0	5.0	**0.004**	3.6	15
	Discussion time after the video was finished		15.3 minutes	13.9 minutes	0.39	0.89	15
	Discussion during the video (when video was paused)		17.5 minutes	6.6 minutes	**0.008**	3.2	15
	Average amount of discussion per pause		3.6 minutes	1.4 minutes	**0.009**	3.2	15
	What did participants think about the amount of discussion during the session? (one question)	1=too little, 4=just right, 7=too much	4.1	4.1	0.62	-0.51	15
	How comfortable were participants with asking questions and making comments? (four questions)	1=uncomfortable, 7=comfortable	5.9	5.3	**0.03**	2.6	15
Satisfaction	Overall level of satisfaction (two questions)	1=dissatisfied, 7=satisfied	5.4	6.2	0.11	-1.7	15
	Level of satisfaction with discussion (six questions)	1=dissatisfied, 7=satisfied	5.7	5.5	0.50	0.7	15
	Feelings of boredom (two questions)	1=low level of boredom, 7=high level of boredom	3.4	2.4	**0.05**	2.2	15

Table 2: Examining differences between CVV and DCVV groups. Results take into account pretest measures of involvement and number of group members. All analyses were done at the group level.

> *I think underlying our thoughts as a group is that we had expectations of testing Win2000 or something, and this is not what we had conceived in our minds.*

No DCVV participants made similar comments. To avoid this problem in future studies, we may return to within-subject instead of between-subject experimental designs when tasks in one condition are relatively plain.

5.2 Comparing audio- and videoconferencing

Our second set of analyses examined whether the DCVV experience was different for groups with video or audio-only interaction. The results (Table 3) consider the number of participants and a pretest measure of involvement for each group.

5.2.1 Learning and Comprehension

The groups had virtually identical quiz scores (70% and 71%). The groups also reported similar levels of comprehension after watching the video. However, the groups differed somewhat—but not in the expected direction—when assessing the perceived helpfulness of group member comments. Both groups reported that teammates' comments were helpful, but groups with audio only thought their teammates' comments were significantly more helpful ($t(15) = -2.7$, $p = 0.04$).

5.2.2 Discussing the Video

Although the groups with videoconferencing paused the video more often, the difference was not significant. In addition, both groups felt that the number of pauses was about right, and both groups were equally comfortable with pausing the video.

Both groups were also equivalent in terms of time spent discussing the video and comfort with asking questions and making comments. However, DCVV-audio groups felt that there was too much discussion, while DCVV-video groups felt that the amount of discussion was about right (a one-sample t-test comparing the DCVV-audio score to a "just right" score of 4 yielded $t(3) = 11$, $p = 0.002$).

5.2.3 Satisfaction with Experience

Contrary to our expectations, DCVV-audio groups were significantly more satisfied with the discussion than DCVV-video groups ($t(15) = -3.2$, $p = 0.02$). Aside from this

	Variable, and number of questions used to measure variable	Score explanation	DCVV audioconf (n = 4)	DCVV vidconf (n = 6)	p	t	df
Learning	Quiz score (ten questions)		70%	71%	0.81	-0.27	9
	Feelings of video comprehension (two questions)	1=low, 7=high	5.9	5.5	0.44	-0.83	9
	Did group members think their teammates' comments were helpful? (two questions)	1=least helpful, 7=most helpful	6.0	5.2	**0.04**	-2.7	9
Discussion	Number of times video was paused		3.3	5.2	0.55	0.64	9
	What did participants think about the number of times video was paused? (one question)	1=too few, 4=just right, 7=too many	3.7	3.8	0.80	0.27	9
	Overall comfort with pausing the tape (two questions)	1=uncomfortable, 7=comfortable	5.0	5.1	0.96	0.05	9
	Discussion time after the video was finished		15.3 minutes	13.0 minutes	0.15	-1.6	9
	Discussion during the video (when video was paused)		5.8 minutes	7.1 minutes	0.29	-1.2	9
	Average amount of discussion per pause		1.5 minutes	1.3 minutes	0.08	-2.1	9
	What did participants think about the amount of discussion during the session? (one question)	1=too little, 4=just right, 7=too much	4.7	3.7	**0.01**	-3.7	9
	How comfortable were participants with asking questions and making comments? (four questions)	1=uncomfortable, 7=comfortable	5.3	5.3	0.10	-2.0	9
Satisfaction	Overall level of satisfaction (two questions)	1=dissatisfied, 7=satisfied	6.4	6.0	0.17	-1.5	9
	Level of satisfaction with discussion (six questions)	1=dissatisfied, 7=satisfied	5.9	5.2	**0.02**	-3.2	9
	Level of boredom (two questions)	1=low level of boredom, 7=high level of boredom	1.9	2.7	0.39	0.9	9

Table 3: Examining differences between DCVV-audio and DCVV-video groups. Results take into account pretest measures of involvement and number of group members. All analyses were done at the group level.

difference, the groups did not differ on our satisfaction measures.

6 DISCUSSION

First we should note shortcomings of this study. First, the video we used for these studies was professionally made (not tapes of classroom lectures). However, if distance education takes hold it is likely that more resources will be invested in producing materials of higher quality than a recorded lecture. Second, this research has the generalizability issues common to lab studies. The majority of our participants were not students, and although our experiment setting was not very similar to a college classroom environment, it was similar to what is found in corporate training settings. As a result, this research could benefit from a field study replication.

6.1 DCVV Usage Experience

The first goal of this research was to build an effective DCVV system on top of existing, commercial technology. Ideally, we hoped to build a system that would recreate the CVV experience for students who could not meet in the same room. Our lab study found no significant differences between CVV and DCVV groups on most major measures, including our measures of learning.

However, our study did find one area in which the system could benefit from additional design. DCVV groups were not as comfortable with pausing and discussing the videotape, reflected by both the survey data and the reduced time spent discussing the video when it was paused. This result is worrisome given the tight assumed connection between discussion and learning in the TVI literature.

DCVV participant comments in the post-experiment survey suggested that they were reluctant to pause the video because of the effect on other participants and because they did not know the other participants. They wrote:

> *With this video I felt a bit hesitant to stop the video because I didn't want to interrupt my team members.*
>
> *Didn't want to hold the others back by stopping the video and reviewing it.*
>
> *I disliked not being able to view my teammates, body language, etc. I felt like if I stopped the video I would be*

holding up the group, who may or may not be able to learn/understand faster than me.

I felt everyone was a bit uncomfortable because we didn't know each other.

Since I didn't know my other classmate(s) very well, I was reluctant to stop the video and initiate discussion. I would have been less bashful if I had known the other people better who were sharing the experience with me.

Interestingly, in the CVV condition, pausing the video also affected other people, yet no CVV participants mentioned this as a problem.

This difference might disappear with extended use of the system. It does, however, suggest the possible usefulness of a designated discussion leader who takes the initiative to pause the tape for discussion. This was the method used by the CVV groups in our study.

It would also be possible to enhance the system to enable the course designer to build in discussion pauses at appropriate points, a possibility suggested by one participant in the survey. The video could be programmed to pause automatically at good places for discussion and resume once someone presses the play button. (Spontaneous pausing would of course remain an option.) This is consistent with the pausing behavior we witnessed during the study: most groups tended to pause the video during the transitions from one case study to another.

6.2 Utility of Communication Channels

The second goal of this research was to examine the effects of using different communication channels to link students together.

6.2.1 The Value of Text Chat

The most inexpensive and simple DTVI system would link students together using only plain text chat. However, observations from our first study suggested that text chat is not good enough for the type of communication that makes DTVI successful: none of our pretest groups paused the video even once when given only the text chat window to communicate. Thus, from these limited data, it seems that DCVV groups must be provided with at least audioconferencing to communicate.

6.2.2 The Value of Videoconferencing

Our hypothesis that the talking heads video channel provided little benefit was supported. In fact, on some measures, the audio-only groups did better than video-linked groups: The audio groups perceived their teammates' comments as being more helpful and were more satisfied with the discussions. However, they also thought that there was too much discussion, raising the possibility that the lack of visual contact with other group members made it difficult for people to gauge what others felt was an appropriate amount of talking. For example, one DCVV-video participant wrote:

I was able to see people's real feelings that they wouldn't normally 'betray'. Specifically, the last discussion I came to tell apparently bored two other participants.

Aside from this issue, our data suggest that providing video is not a helpful addition. If this finding is consistently replicated, it is good news for organizations that wish to implement DCVV given the cost of providing the video channel.

7 CONCLUDING REMARKS

The DTVI learning model provides significant benefits: scalability, good learning outcomes, and location and time flexibility for students. In this paper we have addressed two issues that together allow DTVI-like models to become more widespread. First, we have shown how a collaborative video viewing solution can be built on existing commercial technologies by any third party. Second, we have provided results on student learning, interaction, and satisfaction as a function of communication channels used for interaction. Our lab studies showed that audioconferencing performs as well as more expensive and inaccessible videoconferencing.

Both of these solutions illustrate that existing barriers to DTVI can be overcome without tremendous expense. However, this research still requires confirmation from field studies. In addition, future research should address the question of whether tutors are necessary, and the related design problem how to make people in distributed groups more comfortable with pausing the video, especially if no tutor is present.

ACKNOWLEDGMENTS

We thank Ken Bachman, Ed Barrett, and the rest of the Microsoft Usability Test Coordination team for recruiting and scheduling the study participants. We also thank Doug Tutmarc and the Microsoft Usability Setup Operations Support team for their assistance.

REFERENCES

1. Boyle, E., Anderson, A., and Newlands, A. (1994). The effects of visibility on dialogue performance in a cooperative problem solving task, *Language and Speech*, 37, 1, 1-20.
2. Chapanis, A. (1975). Interactive human communication. *Scientific American*, 232, 36-42.
3. Chapanis, A., Ochsman, R. B., Parrish, R. N., and Weeks, G. D. (1972). Studies in interactive communication: The effects of four communication modes on the behavior of teams during a co-operative problem solving. *Human Factors*, 14, 487-509.

4. Edigo, C. (1990). Teleconferencing as a technology to support co-operative work: Its possibilities and limitations. In J. Gallegher, R. E. Kraut, and C. Egido (eds.) *Intellectual teamwork: Social and technological foundations of cooperative work* (p. 351-371). Lawrence Erlbaum Associates.

5. Finn, K., Sellen, A., and Wilbur, S. (1997). *Video-mediated communication*. Lawrence Erlbaum Associates.

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

7. Green and Williges, R. C. (1975). Evaluation of alternative media used with a groupware editor in a simulated telecommunication environment. *Human Factors*, 37(3), 283-289.

8. The Microsoft Windows Media Player, http://www.microsoft.com/windows/windowsmedia

9. Microsoft Windows NetMeeting, http://www.microsoft.com/windows/netmeeting

10. The Microsoft Windows NetMeeting 3.0 SDK, http://www.microsoft.com/windows/NetMeeting/ Authors/SDK/default.ASP

11. O'Connaill, B., Whittaker, S., and Wilbur, S. (1993). Conversations over videoconferences: An evaluation of video mediated interaction. *Human Computer Interaction*, 8, 389-428.

12. Olson, J., Olson, G., and Meader, D. (1995). What mix of video and audio is useful for remote real-time work? *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '95).

13. Omoigui, N., He, L., Gupta, A., Grudin, J., and Sanocki, E. (1999). Time-Compression: System Concerns, Usage, and Benefits. *Proceedings of the ACM Conference on Human Factors in Computing* (CHI 99).

14. The Real Video Player, http://www.real.com

15. Sipusic, M., Pannoni, R., Smith, R., Dutra, J., Gibbons., J., and Sutherland, W. (1999). Virtual Collaborative Learning: A Comparison between Face-to-Face Tutored Video Instruction and Distributed Tutored Video Instruction (DTVI). Sun Microsystems Laboratories, Inc. TR-99-72. Available at http://www.sun.com/research/techrep/1999/ abstract-72.html

16. Smith, R., Sipusic, M., and Pannoni, R. (1999). Experiments Comparing Face-to-Face with Virtual Collaborative Learning. Sun Microsystems Laboratories, Inc. TR-99-0285.

17. Stone, H.R. (1990). Economic development and technology transfer: Implications for video-based distance education. In M. G. Moore (Ed.), *Contemporary issues in American distance education* (pp. 231-242). Oxford, England: Pergamon.