

Evolving Use of A System for Education at a Distance

Stephen A. White, Anoop Gupta, Jonathan Grudin, Harry Chesley, Greg Kimberly, Elizabeth Sanocki
Microsoft Research

{stevewh, anoop, jgrudin, harrych@microsoft.com, greg@gak.com; a-elisan@microsoft.com}

Abstract

Computers and networks are increasingly able to support distributed collaborative multimedia applications. In fact, the growing interest in distance learning reflects the awareness that these technologies could support the broad, complex interaction at the heart of instruction. However, we still have a way to go; users (instructors and students) and designers of such applications face many complex challenges. For example, social conventions governing use are needed, but for a given array of features in the application, we don't know what conventions are optimal or even adequate, or how they might vary with course content, class size, and instructional style. How will a flexible design that lets a class form its own practices fare? To what extent should conventions be designed in or promoted through training? Although researchers have begun to explore these issues, longitudinal studies are rare. In this paper, we look at these issues using Flatland, an extensible system that provides a wide range of interaction capabilities. We report on its use in three multi-session training courses. We comment on the overall reaction of students and instructors, changes in behavior and perception over sessions, and the formation of social conventions over sessions. We observed classes growing more comfortable with the technology (with exceptions) and developing conventions governing the use of features, but not always effectively. We discuss implications for adoption of such technology by organizations.

1. Introduction

Networked computers are increasingly able to support distributed, real-time multimedia presentations, including live audio, video, and feedback channels. A key application domain is distance education. Although controversial when seen as a replacement for standard classrooms, distance education can provide advantages when classroom attendance is not possible, or for students who wish to participate more casually. Not everyone who could benefit from training and other organizational learning activities can participate in person. At times we might like to see a presentation from our office, where we can timeshare with other activities or easily disengage. And mixtures of face-to-face and distance instruction are possible.

Many systems address intermixing of multiple streams of information in synchronous and asynchronous

environments. Our system, like the Xerox PARC Coral system [13], captures real-time interaction using a distributed object system in a tele-presentation setting. We differ by allowing a presenter to pre-author structure into the content. Our system has to address the timing variability of multiple streams of data [12]; in addition our system must handle synchronous collaboration with variable latency. We address the asynchronous variability problem by recording events directly into an ASF format. This compensation for latency allows our system to avoid the problems reported for the PERSYST system [6] in dealing with content change events.

Distance learning has a long history. "Correspondence schools" developed over time, culminating in the highly respected Open University courses, which often mix postal correspondence and live sessions. Stanford University has offered televised courses to Silicon Valley companies for 25 years [18]. The National Technical University (NTU) has for 15 years provided courses including live satellite-based audio and video lectures in conjunction with telephone, fax, and postal exchanges [4]. More recently, the Internet and Web have been the focus of experiments. For example, Carswell [3] contrasted postal and Internet communication in Open University courses. Lawhead et al. [11] summarize issues and findings on the Web as a medium for distance learning.

Most of this earlier work focuses on use of analog video transmitted to students, or simple communication channels such as use of bulletin boards and email for collaboration. Relatively little work has addressed newly emerging infrastructures supporting audio, video and synchronous collaboration over networked computers. Video-Mediated Communication [5] surveys the topic broadly. Sustained work on instructional use of multimedia over networks was conducted at Sun Microsystems [7-9].¹ Presenters and audience used Sun's Forum system from their desktops; live one-way video of the presenter, two-way audio, slides, and a number of feedback channels were supported. Audience members appreciated the convenience of attending via their desktop computers, but felt it was less effective than live attendance, as did the instructors, who found the lack of feedback disconcerting.

¹ Commercial products for telepresentation are appearing, such as Centra Symposium [16] and PlaceWare Auditorium [19].

Why the mixed response, given the benefits and convenience of distance learning? The Sun Forum designers built in several communication and feedback channels, but knowing which to employ at a given moment requires understanding their effects and agreeing on conventions for their use.

The Sun Forum experiments focused on single-session presentations (e.g., an invited researcher would come and give a talk to Sun employees using the system). Most technologies experience a learning curve as individuals develop an awareness of the range of features and how they can be used. It took time for social conventions to develop around the use of earlier technologies that support communication and coordination, such as the telephone and email. Such conventions can vary across groups. For example, who speaks first when a telephone is answered differs in different cultures. Such conventions must be established and agreed upon, and then learned. Not allowing for appropriate development and dissemination of conventions can significantly impact whether or not some of these technologies are adopted and the rate at which they penetrate their intended audience.

Previous research that focused on initial use of live multimedia presentation systems represents a first step. This study builds on it by looking at use of distance learning technologies over multiple sessions. Through such studies we obtain a better understanding of how groups start forming conventions around complex communication and awareness technologies. We can identify effective conventions in the environments studied, and less effective practices that might be avoided. We can select features to drop, add, or redesign, identify where more attention to the interface is required, and what to emphasize in training. It is safe to anticipate that in the beginning, designers, instructors, and students will seek to recreate classroom behaviors to a large degree, and well into the future will have arrived at a set of practices unimagined today. We hope to observe and assist the first steps in this journey.

Previous longitudinal studies of instructional technology include the Classroom 2000 project [1], a careful exploration of multimedia support for face-to-face instruction. Our contrasting desktop-to-desktop focus is similar to that in [6], an ambitious Internet-based multimedia system for short, multi-session courses. However, that report focuses on a proof of concept and does not describe participant behavior.

The paper is organized as follows. In the next section we describe the prototype system that we used in our studies, and the various interaction modalities that it supported. Following that we discuss the study environment and method used. Next we present our results, focusing on overall reaction of students and instructors, changes in behavior and perceptions over the multi-session classes, difficulties encountered due to uncertainty about others' context and experience, and finally the creation of social conventions. In the last section we conclude and present design implications.

2. System requirements and features

Systems that support distributed meetings or distance education force all awareness and communication to be mediated digitally. Users must find ways to compensate for lost information and develop social protocols to replace those disrupted by technology. Synchronous meeting support systems have been a research focus for thirty years [e.g., 2, 10], and a consistent conclusion is that systems must support diverse interaction channels.

2.1. Classroom Interactions

In the distance education context, the key focus is the awareness and communication that link instructor and students. In standard classroom instruction, a flexible range of communication channels is available—visual observation, voice, expression, gesture, passing notes, writing on a board, throwing an object for emphasis, walking about to view student work. Managing them isn't easy: Instructors profit from training, and even after years of experience, effective teaching is a demanding task. The way an instructor uses these channels to interact exudes a particular style. The list below shows a variety of commonly used interactions:

- Viewing and hearing the lecturer, including gestures
- Arrival and departure of participants
- Slides, with the ability to point or mark for emphasis
- Spontaneous writing and drawing (as on blackboard)
- Student questions on lecture content, including ability to support another's question (e.g., nodding in class)
- Spontaneous questioning of students by instructor
- Process-related issues, such as level of comprehension (in a class, communicated publicly with a comment or privately by facial expression)
- Discussions among students
- Demos or labs

2.2. System

The system used in this study consisted of two applications, Flatland and NetMeeting. Flatland is a flexible synchronous education environment designed for telepresentation. NetMeeting is a synchronous collaboration tool that supports application sharing.

2.2.1 Flatland architecture. Flatland is a multi-threaded, distributed, client-server application that must overcome network latencies in synchronizing audio, video, and other streams. It also must be a robust rapid-prototyping environment, to facilitate experimentation with features and interfaces. It is built on several existing technologies.

V-Worlds. V-Worlds [15] is a platform for distributed client-server-client applications. It provides automatic transfer of object properties, as well as remote procedure calls, among a set of communicating clients and a single

server. It also provides persistent storage of system objects.

DHTML. Dynamic HTML (DHTML) is composed of the Internet standards HyperText Markup Language (HTML) 4.0, Cascading Style Sheets (CSS) 1.0, and the associated document object model. Taken together, these provide a powerful system for quickly building and modifying an application user interface.

Scripting. Flatland uses JScript to implement the semantics of application elements within V-Worlds objects, and to implement the user interface within DHTML objects.

NetShow. Microsoft NetShow provides the streaming video facilities required by some Flatland applications.

2.2.2. Flatland user interface. Flatland combines Microsoft NetShow [17] streaming audio and video with a collection of feedback mechanisms that allow the presenter to receive both solicited and unsolicited responses from the viewers. Figure 1 shows the main Flatland screen layout as seen by a presenter. The audience sees a similar view, but without some of the controls. Figure 2 shows the Flatland components and their relationships. A presenter communicates with the audience using NetShow video and Flatland. The audience can pass questions, answers, and requests back to the presenter via Flatland.



Figure 1 – Flatland presenter layout.

In Figure 1, the middle left section of the layout contains the video of the presenter, provided using NetShow. Any Flatland participant with a video feed could present, but in these studies only the instructor did. The upper right section of the window contains slides and questions, as defined by the presenter. This area can include slides generated by PowerPoint, simple text slides, and audience Q&A slides that allow the audience to vote by selecting one of the answers to a multiple choice question. The presenter can also use a “pointer” to indicate specific sections of the slide during the presentation.

The presenter controls the selection of the currently displayed slide. However, a History button above the slide

area generates a separate window with the entire set of slides for the current presentation, allowing a viewer to browse slides other than the one currently being displayed.

Presenter controls in the slide area include facilities to select the slide to be displayed, using either the “next” and “previous” arrow buttons on the top right or the table-of-contents pop-up on the top left. There are buttons to edit or delete the current slide. A presenter can also create a new slide on the fly and insert it in the presentation.

Below the slides, on the right, is a text chat area. This allows free-form communication between audience members or between the audience and the presenter. Interactive chat gives audience members a strong feeling of the presence of other participants, and can be invaluable for resolving last minute technical problems that audience members may encounter. This window also reports when people join or leave a session.

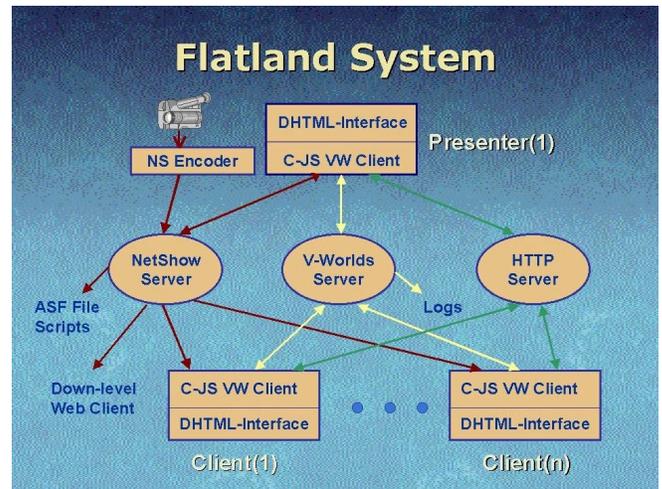


Figure 2 – Flatland system.

Although free-form chat is valuable in providing an open and unrestricted communications channel, it can easily become overwhelming. For questions specifically directed at the presenter, a separate question queue is provided to the presenter in the bottom left of the window. In this area (hereafter called the Q&A window), audience members can pose questions for the presenter. They can also add their support to questions posed by others by incrementing a counter. This voting capability could reduce duplicate questions and help a presenter decide which question to address next.

Finally, the upper left area of the window provides several lighter weight feedback mechanisms. On the right are two checkboxes that allow the audience to give continuous feedback on the speed and clarity of the presentation, displayed as a meter on the presenter window. On the left are buttons to leave the presentation, to show a list of audience members, and to raise a hand for impromptu audience polling. A pop-up, floating “tool tip” shows a list of members with their hands raised if the

cursor is left over the hand icon. The same information is also displayed in the pop-up audience member list.

2.2.3. NetMeeting. Demonstrations are a critical component of frequently-offered classroom training courses that is not supported by Flatland. We addressed this with the application sharing feature of NetMeeting, a freely available Microsoft software application. Instructor and students ran Flatland and NetMeeting sessions concurrently, first logging into Flatland and then joining a NetMeeting meeting.

NetMeeting application sharing allows everyone in a NetMeeting meeting to view any application running on a participant's machine. Viewers see dynamic screen changes and mouse pointer movement (with less delay than in Flatland). NetMeeting also supports point-to-point audio and shared floor control, but these were not used.

2.2.4. System supported interactions. To summarize, the interaction channels available to participants using both Flatland and NetMeeting are:

Flatland

- Synchronized audio/video window carrying the lecture
- Gesturing within range of the camera
- Slide window with pointer capability
- Text slides created dynamically (tools provided)
- Interactive query slides created dynamically (tools provided)
- Q&A window with question prioritizing capability
- Discussion or chat window with machine-generated announcements of participants coming and going
- Attendance window
- Slow/fast and confusing/clear checkbox features
- Hand-raising feature

NetMeeting

- NetMeeting application sharing
- NetMeeting chat window
- NetMeeting multi-user whiteboard (never used)

One might ask "Do systems such as Sun Forum and Flatland have *too many* interaction channels?" However, as we note below, instructors and students feel isolated and frequently request additional channels. Of course, the solution may not be in more channels but instead better protocols and/or fewer or more appropriate channels.

3. Method

Three technical courses usually presented in a classroom were taught desktop to desktop, with minimal modification, using Flatland. They were, respectively, two 3-hour sessions with 4 students, four 2-hour sessions with 10 students, and three 3-hour sessions with 7 students. All included live demonstrations via NetMeeting. The courses were conducted by professional instructors in a large corporate environment.

The students were volunteers from a list waiting to attend the classroom course. Their job functions and

technical expertise differed, but all had substantial computer experience. The population is thus not general, but perhaps representative of early adopters of technologies of this kind.

Each instructor was situated in a usability lab observation room, enabling us to videotape, log usage, observe, and take notes unobtrusively. One student in each of the first two courses also participated from a (different) usability lab. The other students attended from their offices.

Prior to teaching the first session, each instructor received about 15 minutes of instruction on the features of the system. We worked with the students primarily by email to insure they had the software installed. Throughout each session we had at least one observer and one person available for technical support. Following each session we asked instructor and students to fill out questionnaires (usually on-line), and interviewed the instructors for reactions and explanations of observed behavior. The questionnaires had around 15 measures, most of which asked for a rating on a scale of 1 to 5, and provided space for comments.

Prior to subsequent sessions taught by an instructor, we sometimes reminded the instructor of unused features, or demonstrated them. For example, we demonstrated the preparation of an interactive slide. Thus, we did not always leave instructors to explore by trial and error, but we chose to minimize our intervention, and as noted below, the suggestions we made were often not picked up. The instructors are professional teachers with personal styles, confident in their control of the class and material.

These short, ungraded courses do not allow the same growth of interaction that an academic course would, but the multiple sessions allowed us to see early exploration and articulation in the use of this complex technology.

4. Results and discussion

4.1 Overall reaction of students and instructors

"But I wouldn't have gotten this level of interactivity, and I think we used the remaining time more effectively than if we had been there in person."

—A student comparing traditional classrooms to Flatland

The three courses were held as scheduled and covered the planned material, overcoming minor technical glitches. Students are not tested following internal training courses, so assessment is based on questionnaires, debriefing, analysis of event logs, and observation.

Consistent with the results reported from Forum studies at Sun [7-9], students rated desktop-to-desktop instruction more highly than instructors. This is significant because scheduled training courses likely attract more focused audiences than the general informational talks broadcast over Sun Forum.

Students made use of the ability to timeshare and do other work while attending the lecture, a benefit of distance learning. One student in the observation lab, with

fewer personal materials and distractions than at his desk, checked email and carried out other tasks during slow periods, seeming to use verbal cues from the instructor to know when to return to the class. (When asked how he knew to return, he said "I'm not sure, I just did.") Despite this, students rated their overall level of "attention to class" consistently between 75% to 85%. This is closer to the level found at Sun among live attendees (84% vs. 65% for Forum attendees). This may reflect the greater focus of students registered for a class relative to remote attendees of an informational lecture.

Interestingly, the instructor was favorable when told of the student multitasking, noting that certain topics may be familiar or boring to some students. Furthermore, fidgety students in a live class distract other students and the instructor. In this environment they did not interrupt his focus, or that of other students who wanted to listen.

The fact that students particularly appreciated Flatland was recognized by Instructor 2, who wrote "That students wanted to take the follow-up course in a similar manner was evidence that they thought the course and the presentation technique were good."

Although Flatland instructors experienced the absence of customary feedback and student interaction as reported with Forum, two instructors grew more positive and later promoted the technology within the technical education group. The third instructor was increasingly negative.

Asked whether he would recommend Flatland or classroom participation, Instructor 1 wrote "I still think in person is best. If logistics do not allow in person then this is (better than instruction using large-room video teleconferences and PowerPoint.)" The second instructor also had a mixed final assessment: "I don't believe all courses would be equally adaptable to Flatland. Also, what makes many trainers good trainers is the classroom itself. We are performers, stage actors if you will, not radio voices." With this insight he was noting that there is a new skill to be learned. Many Flatland features were designed to provide feedback missing to "radio voices" but after 4 classes they had not compensated. Instructor 3 was more succinct: "I don't want to do this again."

Instructor 3 attributed his view in large part to disruptions when he switched between Flatland and NetMeeting for demonstrations. He experienced a few technical problems, but most disruptive was the lack of established protocols. For example, in one case he started a NetMeeting demonstration before all students had joined the class, confusing late arrivals. At another point, a student kept trying to join NetMeeting (and failing), popping a window on the instructor's desktop although no demonstration was being conducted. Other possible factors include attitude, teaching style, and comfort with novelty, dimensions on which this instructor appeared to differ from the previous two.

But the greatest factor may have been a chance occurrence. In this busy corporate environment, courses requiring 9 hours in a week often have student attrition rate of around 50%. Our first two courses did better and

the instructors were interested in learning to use the technology anyway. In the third course, the attendance in the three sessions went from 7 to 4 to 2 students. The instructor was less interested in the technology and keenly aware of the attrition, agonizing and speculating aloud that departures were due to him or the technology. We were able to contact four out of the five dropouts and found that they had conflicting commitments and seemed genuinely positive about the technology and class.

At a high level, from the perspective of eventual adoption and deployment of these technologies in corporations, the fact that student satisfaction was higher than instructor satisfaction may be significant. Tangible benefits to students were clear (they did not have to commute to the classroom, they could multi-task, etc.), but advantages for instructors were less clear. They had to learn new tasks and key, familiar feedback mechanisms were unavailable. Early adoption and recommendation of these technologies is likely to be driven by instructors more than students, so we must attend more to the instructors' experience with the system. In particular, our studies were conducted without guidelines for the instructors. Instructor 3's negative experience might have been avoided by using observations from the earlier trials.

4.2 Changes in behavior and perception over sessions

"Before it was just a feature. This time it was a tool."
—Instructor 1, discussing 'hand raising' after 2nd class

Flatland use changed considerably over sessions. Facing assorted communication and feedback channels to replace those of classroom instruction, participants tried key features, grew comfortable with some, then experimented with more effective uses and other features. Below we look at instruct and student experiences.

4.2.1. Instructors. Generally, instructors began stiffly as they learned to deal with the camera and slides. One instructor said of his first session "I spent most of my time making sure that I was in the center of the camera."

For students, audio, video, and slide changes were synchronized but delayed slightly. Instructors only saw the video delayed and initially found their own delayed video image disconcerting. Instructor 1 noted after his second session that he "felt more comfortable... was not as distracted by his own image." Following the first session, he had recommended dropping the "distracting" video feedback, but in the second session learned to use it occasionally to check his camera image and ignore it the rest of the time, and now favored retaining it. Instructor 2 also accommodated to the video, but Instructor 3 remained unhappy with it through 9 hours.

All instructors steadily increased pointer use. Instructor 1 said "I started using it as soon as I was aware of it" when asked how long it had taken to get familiar with the pointer. Instructor 2 increased use from 0 times the first session to 36 times the fourth.

After each session, instructors were asked how well they thought they handled student questions. Instructor 2

responded 3, 4, 4, 5 on a 1-5 scale over sessions. Asked to what extent Flatland interfered, his responses were (no response), “too early to say”, 4 (high), 2 (low). In an overall assessment, he wrote “Much more comfortable with the Flatland environment after 4 sessions.”

Responsiveness to multiple input channels built over time. In his second session, in response to a question Instructor 2 edited text on the fly while in NetMeeting for a demonstration. In his third session he began to make heavy use of dynamically created Flatland text slides to type in code examples and strongly endorsed this as a whiteboard substitute. It was also only in the third session that he responded promptly to Q&A window queries, and in the fourth that he monitored the chat traffic reliably.

As instructors’ comfort levels rose, they added characteristics of personal style. For example, Instructor 2 began greeting students in session 3, and at the end recommended clearer identification of students by name in the interface, a suggestion he did not make earlier.

Most, but not all, changes were positive. One instructor had previously taught using video teleconferencing centers where slides and audio are not in synch. He initially paused after every Flatland slide transition. Reminded that the students see everything in synch, he stopped compensating, but when using NetMeeting, where the display was not delayed and thus out of synch with Flatland audio, he did not relearn to compensate. Similarly, Instructor 3 learned to press the mouse button when using the pointer (required to display the pointer on student screens). Using ordinary PowerPoint in another class, he repeatedly pressed the mouse button when emphasizing a point on a slide, causing the slide to advance. “Flatland ruined me for life,” he wailed.

4.2.2. Students. The 4-session second course offered a clear view of changes in student behaviors and perceptions over time. (The first course was two sessions, and as noted above, attendance in the third course dropped sharply.)

When asked directly (in the questionnaire) how distracting Flatland technology was, student ratings went from 2.8 (neutral) to 1.7 (low) over the course. But in fact, it occupied a lot of their time at the outset. In the first session considerable time was spent dealing with technical startup problems and learning the interface. Three times as much non-technical communication occurred in the second class. In the final two sessions, the overall rate of chat and questions stayed relatively constant, but content discussion increased as technical Flatland discussion dropped. Over the last three sessions, exchanges went from 27% class-related, 11% social, and 62% technology-centered to 60% class-related, 26% social, and 14% on technology. Communication directed to the instructor doubled, with the number of responses to the instructor’s comments and questions rising from one in session 2 to 24 in session 4. This marked increase reflects verbal instructor questions such as asking students to identify the errors in some code.

The students were quick to notice the improvement in interaction. When asked their impression of the Flatland experience and how well Instructor 2 presented, typical comments were “Some difficulties with the interaction between new technology, students, and teachers” (session 1), “the interaction with the instructor was easier...his interaction was improved, and the course was likewise improved” (session 2). The last session “had the best interaction” but “still left room for improvement.” Among the contributing factors were increased familiarity with the technology, reduction in technical problems, improved technical support and more exchanges with the instructor.

Overall, for both instructors and students, we see a greater ability to handle multiple communication channels and an increased volume of communication over time. Thus, it is critical to evaluate use over more than one session. Of course, it is still critical for designers to ensure that the application is effective for first-time users, as they may never come back to it after an initial bad experience.

4.3. The creation of social conventions

The social conventions most often discussed in the context of group support technologies govern floor control or turn taking. In the absence of visual cues, the potential for uncertainty or confusion motivate adoption of an agreed-upon approach. Floor control is an issue for Flatland—students have some difficulty knowing when to interrupt and instructors are unsure how to pass control to a student—but there are broader sources of ambiguity and confusion that rob classes of efficiency and effectiveness.

The most pervasive is the tremendous uncertainty or incorrect assumptions by an instructor about what students are seeing, doing or thinking, and vice versa. Uncertainty can arise from lack of full understanding of features, differences in equipment configurations, and the lack of visual and auditory feedback channels relied on in classrooms. Whatever the causes, it leads to uncertainty as to what to do and when.

A related problem is deciding how to act, and which of the many interaction channels to use. This requires judgment about the attention of remote participants, their response options, and is also clearly influenced by the ease of acting given the real-time setting.

Conventions are critical to navigate through these potential sources of inefficiency. They are the primary means by which people avoid ambiguity and uncertainty. We saw instructors and students experiment with conventions, some effective and some not, some adopted and some abandoned. A few times we suggested conventions, and our suggestions were not always adopted. Before we work through some examples and draw conclusions, we will discuss further the problems we felt they were intended to address.

4.3.1. Uncertainty about others’ context and experience. A key to technology-mediated interaction is understanding the contexts of those with whom we

interact, the effects of our words and actions in their context and on them, and the sources of the words and actions we encounter in our environment.

This problem, and the lack of tools to support “mutual intelligibility,” has been pointed out repeatedly in the CSCW literature since the first conference [14]. The diverse communication channels found in many systems, including Flatland, are designed to alleviate the problem by providing greater awareness and means of communication, but at least initially they are not adequate. In this section we illustrate the extent of this issue.

Participants tended to assume that the other people shared their experience. This led to misunderstandings. When it became clear that others sometimes did not share experience, this created uncertainty or over-compensation, as when early on, instructors informed students of transitions that were obvious.

For example, we customized one instructor’s desktop to include a separate machine for NetMeeting. During an exercise, he put up an interactive query as a Flatland slide asking students to indicate their progress. He was irritated by the lack of compliance, unaware that the students had buried the Flatland window under the NetMeeting window on their single monitor in order to do the exercise.

Conversely, another instructor, who had a single machine, began the second class with a NetMeeting demo. Students arriving late had no way to deduce exactly what was going on. They typed questions into the Flatland chat and question windows, unaware that the instructor could not see their queries until he finished the demo and returned to Flatland.

Instructors can always see the slide pointer, but it only appears on students’ displays when the button is held down. Although informed of this, instructors frequently forgot and believed that students could see their gestures. Even when used properly, the erratic on-again off-again pointer prompted an inquiry during a class from a student who assumed it was a bug. When the same instructor selected a Q&A window question, expanding into the slide window, he was uncertain as to whether students could still see the pointer. These disruptions are minor but remind participants of their limited mutual awareness.

The clear/confusing and too-fast/too-slow indicators appeared as checkboxes on student monitors and as meters on the instructor’s. Students may have recalled or deduced that the instructor saw something different, but could not gauge the effect of their actions. They also did not know whether complaints were anonymous. When asked to try using it they did, but when the instructor did not respond in a detectable way, they stopped.

Students were uncertain whether handraising, Q&A queries, and interactive slide responses were anonymous. In fact, they are not, but it requires an effort by the instructor—bringing up the attendance window identifies hand-raisers, hovering the pointer over the meters or indicators reveals contributing students. One instructor was unaware of these techniques and assured students that

a vote would be confidential, only to have the students alerted by a peer who had discovered these methods.

Video is a common source of minor mutual unintelligibility. The camera field of view cannot include natural hand gestures and also provide a feeling of being close to the instructor. Although able to see what was being broadcast, two instructors made emphatic hand gestures off-camera. Similarly, the instructors, despite being aware that eye contact was an issue, did not frequently look at the camera.

The greatest category of problem may be the difficulty of interpreting no response—was your communication overlooked, or was it ignored? In face to face situations we can gauge the likelihood much more accurately.

Consider the chat window. Students frequently used it prior to the beginning of a lecture, with the instructor responding. Once the lecture began, the instructor shifted attention to the slide window, monitoring chat only occasionally. If a student posed a question that was not course-related in the chat, such as a problem joining the NetMeeting conference, how should no response be interpreted? Conversely, with students engaged in an exercise and an interactive slide posted on which students are to indicate their completion status, should a “still working” indication that persists be believed, or might the student have finished or even left the office without updating it? We actually encountered each of these problems and many more.

The hand icon, with its small size and subtle change indicator, led to missed signals. Some students assumed they should use it to get the instructor’s attention, as in a face to face class. When instructors did not monitor the hand while speaking and thus missed these signals, students might think they were intentionally ignored. Instructor 1 asked students to raise the hand to signal they had returned from break. At the time, a raised hand timed out after several seconds, so as students slowly raised their hands, the count remained low. When almost all students had in fact raised their hands, the instructor believed few had. (The timeout was later lengthened.)

Another key source of uncertainty for the instructor was the latency for their speech and image to reach the students (due to encoding and buffering delays in the encoder, video server, and playback client). After asking if students had any questions, the instructors often waited only a few seconds and then continued, assuming there were no questions. Some students heard the request for questions and started typing, but were confused and gave up when the instructor resumed lecturing soon after.

Instructor 3’s serious misreading of student reaction and mis-attribution of reasons for dropping indicates the cumulative effect of these phenomena.

Some comments from the instructors:

“Very little feedback from students... For some reason they seemed reluctant to give comments or participate. This should be explored as the teacher needs some kind of feedback...” NetMeeting: “worked very well from my end

but would be very interested how it worked for the students.” – Instructor 1

“In a conventional classroom I have lots of visual cues as to attention level and comprehension that are missing here. For instance, if I say something and get puzzled looks, I repeat myself using a different analogy or different words or a different approach. Not so easy to do that here.” - Instructor 2

“Lack of visible feedback makes it tough to gauge how well/badly things are going.” - Instructor 3

How serious is this, how can it be helped? The Flatland interface can be improved, but we did find significant changes as instructors grew more familiar with the system. They used more channels and developed novel approaches to exploit the technology they were given. One survey question was “To what extent did you get a sense of the students’ attitudes about the class.” After his first session, Instructor 2 skipped the question. When asked directly he said “I have no idea,” that he typically relies on audience feedback and was used to seeing the student. (He remarked, though, that “I like to talk” and thus was not too bothered.) After the next session he rated this ‘1,’ meaning he had the weakest possible sense of student attitudes. After the third session, he rated it ‘3’ and after the final session ‘5,’ the highest possible rating. He had developed approaches to eliciting student feedback.

4.3.2. Uncertainty about appropriate channels.

“Once the instructor had answered it (a question), I didn’t like the feeling of not being able to say ‘Thank you.’ ” —A student discussing the question window feature

In the ‘System Requirements and Features’ section we noted that normal classroom settings involve many direct and indirect means of obtaining or exchanging information. They are so familiar that we often handle them without conscious awareness. Thus, with practice we handle many channels, but have learned which are effective for which types of information. We can signal interest with our eyes, but probably not communicate “please slow down” or “please repeat that” on that ‘channel.’ Deciding whether to ask a particular question during a lecture or after class can involve a range of considerations including relevance, the demeanor of instructor and students, and the potential for embarrassment. We may rely on subtle cues without thinking about it, picking up information that on a system may be represented in different ways, require action to elicit, or be unavailable.

These systems have many interaction channels, each motivated by a perceived need. The channels and their interfaces can be improved, but it may always take time to learn to use them effectively. If they seem too numerous, bear in mind that our instructors and students requested additional channels, including more ways to ask questions, and audio- and possibly video-conferencing connecting all participants. Researchers have proposed yet others, such as emotion meters [2] and applause meters [7].

We observed a tendency to respond via the same channel that one is queried, even though that may not be optimal. For example, in his first class, Instructor 1 responded to early chat by typing in the chat window. In the second class he responded by speaking. When we inquired, he said he was unaware of making a shift. Similarly, one instructor’s first inclination was to respond to a Q&A window query through the same channel.

Of course, students could not respond by speaking. This created a number of dilemmas. When a student asks a formal Q&A question and the instructor verbally requests clarification, should the student respond by entering a non-question in the Q&A window, which everyone is attending to but which is set up for voting and prioritizing questions, or through the chat window, which people may not be attending to? In the same vein, following the instructor’s answer, students wrestled with where to say “Thanks.” Several chose the Q&A window.

Despite some training, use of the Q&A window was not evident to everyone. One student reported after his first class that he assumed this window was for the instructor to pose questions. This misperception was helped by particularly active student use of the less formal looking chat window. Virtually no use of Q&A was made in this class, suggesting a shared misperception. Similarly, as noted in the previous section, there was uncertainty about the role of the hand icon.

Communication possibilities could be clear but strike participants as awkward. Students did not adopt the checkboxes for identifying a lack of clarity or appropriate presentation speed. Normally, this is communicated tactfully and unobtrusively by puzzled or bored expressions. We found little use of our overt signals, as have previous researchers with similar features.

Each class wrestled with the role of the chat window. Students entering and leaving a class were heralded in the chat window, leading to heavy activity early in a session. Technical failures and occasional auditors injected noise into this channel. This activity also made it seem to students to be a logical place to report problems, and once attention was focused on it, to ask subject-matter related questions. Instructors, on the other hand, had heavy demands on their attention and generally stopped monitoring this noisy channel once a class was underway, except during exercises. When student questions went unanswered in the chat window, they would often be ‘escalated’ to the formal Q&A window, eventually leading to a protocol where the content questions were always asked there. For example, for the second course, the Q&A window was unused in session 2, reducing instructor attention to it; until a frustrated student escalated an ignored technical problem from chat to the Q&A window. By the third session it was used for content questions, with some student voting.

For the fourth session in the second course we asked that technical problems be reported in the NetMeeting chat or by phone, content questions in the Q&A window, and other discussion in the Flatland chat. This brought

some order to interaction, but was not always adhered to. Similarly, we asked the third course to direct technical problems to the NetMeeting chat, and some were, but most went to the Flatland chat, with a sizable number to the Flatland Q&A window. Thus, interventions were of limited success.

4.3.3. Conventions and protocols. Finding themselves adrift on this sea of opportunity and uncertainty, groups were observed trying to bring order to the chaos. (Although some uncertainty can be reduced by better design, substantial uncertainty is likely to accompany any complex system providing multiple communication channels.) The efforts had a trial-and-error appearance; some were abandoned, others were maintained longer, but there was rarely a sense of systematicity or exploration. The principal “meta-discussions” (other than ours with instructors between classes) were occasional student exchanges in the chat window.

Consider this journey by Instructor 1. When he first wanted to know whether the class was ready after an exercise, he asked students to “raise their hands.” As noted earlier, students raised their hands over a longer interval, the count remained low, and he incorrectly believe people had not responded. He switched to posting an interactive slide at the beginning of a break with entries such as “need more time” and “ready.” But students marked “Need more time” and later left their office or fail to change it, again misleading him. Finally he returned to hand raising, in conjunction with opening the attendance window, which allows him to see which students have hands up and thus to survey the entire class or particular students for responses.

In his first and second sessions, Instructor 3 made a few efforts to elicit student questions. First, he asked for questions, but did not give students enough time to type in a question. Then he tried asking questions by typing them into the Q&A window (which in other classes was used only by students) and selecting them to highlight/enlarge them in the slide window. The question on the slide got the students’ attention, but the slide was not interactive (i.e., the students could not vote on the multiple options), so students had to decide how to reply. (Although interactive slides were relatively easy to make and were made by the other instructors, this instructor felt they were too difficult to get to while lecturing.) Usually they replied in chat, sometimes in the Q&A window. This instructor opened the attendance window and went through it, exhorting people to participate. This instructor never adopted “Raise your hand” polling, which would have been far faster.

On the other hand, this class did develop a convention for using the hand icon. Because the instructor left so little time before resuming speaking after asking for questions, students who began typing questions (which are not distributed until finished) grew frustrated. Two of them began “raising their hand” to indicate a question was coming, but the instructor did not notice. One of them

noticed the other’s hand and they got into a chat discussion about it. Shortly afterward the instructor appeared to notice the discussion, and adopted the “raised hand means question coming” as a protocol.

Handling technical Flatland problems was also handled through conventions. Again, one student was seen straightening out another who had posted a message to chat, indicating that the questions should go to NetMeeting chat instead. However, in all classes a natural ‘escalation’ procedure appeared, wherein an ignored technical request in chat could be escalated to the Q&A window. Courses varied in how free they became in letting informal messages into this window.

We also saw rudimentary protocols develop for more informal interactions between the instructor and students. In regular classes, some instructors acknowledge students as they enter the class. Our instructors were pre-occupied in the first session and did not acknowledge the entry of the students (indicated in the chat window), but subsequently viewed the chat window, sought out full names from the attendance window, and in the two longer classes explicitly sought more informal contact by name.

These examples suggest that even after 6-9 hours (2-4 sessions) initial attempts were being made to exploit the technology in a more orderly fashion through establishing and agreeing on social protocols and conventions.

5. Conclusions and implications

In these three case studies of a prototype system used in a real educational setting, we observed a range of individual and social behaviors. The system was used successfully in multi-session courses; students were educated and liked the system, instructors missed the face to face experience more, but two of the three remained positive about the technology.

The technology is far from realizing its potential. We do not yet know which features and interfaces will be most useful. However, these systems will have to provide a variety of communication channels, and that uncertainty in how and which to use to achieve a goal will be an inherent challenge. Finding themselves amidst this sea of opportunity and uncertainty, groups must develop conventions and protocols to bring order to the chaos.

Designers and users face a stiff bootstrapping problem. It may not be enough to introduce a class to the features of the system and let it work out a way to use them. Trial and error is too slow. Groups may adopt severely suboptimal protocols, or protocols that conflict with those developed by others they must work with. Bad outcomes for early-adopters can prevent or delay adoption of a potentially useful system. Unfortunately, we have little understanding yet of the appropriate conventions for different kinds of classes, or how we would guide them to the right destination if we knew it. How much should be in the design, how much in training, how much in facilitation? Longitudinal studies are key to answering these questions. We have reached some tentative conclusions.

Increasing mutual awareness is a key to making progress. A prime directive is to find ways to bend all aspects of design toward increasing a level of mutual awareness that participants are comfortable with.

Previous researchers have supported maximizing awareness through "What-You-See-Is-What-I-See." However, there are difficulties. Differences in screen resolution are a notorious obstacle and are likely to get worse. Most laptops still support only 800x600 resolution, and designing for this least-common-denominator would be a major handicap. The ability of students to multitask (a key benefit of remote participation) creates another fundamental kind of non-uniformity, which will be almost impossible to eliminate by design. And a countervailing trend toward allowing individual customization of groupware is evident. The latter is understandable, but it will certainly undermine mutual intelligibility even if is handled carefully.

In a system this size, used under the pressures of real-time interaction, the human-computer interface is very significant. Our interface was not optimal, and we observed small deficiencies in the interface translating into significant distortions in the use of the system. (One can infer that small changes in the interface could strongly affect use, requiring further examination of a system.) Creating interface 'affordances' that suggest to users what particular actions will accomplish, and showing immediate and direct results of actions, are basic tenets of interface design. They are arguably more important in 'social interfaces' due to the potential for embarrassment. Yet flexibility and diversity must be supported.

The desire for greater feedback is manifest in frequent suggestions that audio or video from students to class be added. These and other improvements will create new issues, but are unlikely to change the outcome materially: Forum had similar responses despite using full audio, and the utility of video is uncertain.

We plan to follow this initial study with more extensive studies, repeated with different designs and user populations, to identify effective combinations of features, guidance, and flexibility.

Acknowledgments

We thank the Microsoft Technical Education group for their cooperation in this study. Special thanks to Ushani Nanayakkara for coordinating online classes and to the instructors for volunteering to teach on-line. We thank the Virtual Worlds Group developing the V-Worlds core on which Flatland is implemented. Special thanks to Peter Wong of the Virtual Worlds Group, whose tireless support made these studies possible.

References

- [1] Abowd, G.D., Atkeson, C.G., Brotherton, J., Enqvist, T., Gulley, P. & LeMon, J. (1998): "Investigating the capture, integration and access problem of ubiquitous computing in an educational setting", *Proc. CHI'98*, pp. 440-447, ACM.
- [2] Begeman, M., Cook, P., Ellis, C., Graf, M., Rein, G. & Smith, T. (1986): "Project Nick: Meetings augmentation and analysis", *Proc. CSCW'96*. Revised version in *Transactions on Office Information Systems*, vol. 5, no. 2, 1987.
- [3] Carswell, L. (1997): "Teaching via the Internet: The impact of the Internet as a communication medium on Distance Learning Introductory Computing Students", *Proc. ITiCSE'97*, pp. 1-5, ACM.
- [4] Dankel II, D.D. and Hearn, J. (1997): "The use of WWW to support distance learning through NTU", *Proc. ITiCSE'97*, pp. 8-11, ACM.
- [5] Finn, K.E., Sellen, A.J., & Wilbur, S.B. (Eds.) (1997): *Video-Mediated Communication*, Erlbaum, Mahwah, NJ.
- [6] Ginsberg, A., Hodge, P., Lindstrom, T., Sampieri, B., and Shiau, D. (1998): "The little Web schoolhouse: Using virtual rooms to create a multimedia distance learning environment", *Proc. Multimedia'98*, pp. 89-98, ACM.
- [7] Isaacs, E.A., Morris, T., and Rodriguez, T.K. (1994): "A Forum For Supporting Interactive Presentations to Distributed Audiences", *Proc. CSCW '94*, pp. 405-416
- [8] Isaacs, E.A., Morris, T., Rodriguez, T.K., and Tang, J.C. (1995): "A Comparison of Face-To-Face and Distributed Presentations", *Proc. CHI '95*, pp. 354-361, ACM.
- [9] Isaacs, E.A., and Tang, J.C (1997): "Studying Video-Based Collaboration in Context: From Small Workgroups to Large Organizations", in K.E. Finn, A.J. Sellen & S.B. Wilbur (Eds.): *Video-Mediated Communication*, Erlbaum, Mahwah, NJ, 1997, 173-197.
- [10] Kraemer, K. and King, J. (1988): "Computer-based systems for group decision support: Status of use and problems in development", *Computing Surveys*, 20, 2, 115-146.
- [11] Lawhead, P.B. (Chair) (1997): "The Web and distance learning: what is appropriate and what is not", *ITiCSE'97 Working Group Reports and Supplemental Proc.*, 27-37.
- [12] Manoharm N. and Prakash, A. (1995). Dealing with Synchronization and Timing Variability in the Playback of Interactive Session Recordings, *Proc. Multimedia '95*, 45-56.
- [13] Minneman, S., Harrison, S., Janssen, B., KurtenBach, G., Moran, T., Smith, I., and van Melle, B. (1995). A Confederation of Tools for Capturing and Accessing Collaborative Activity, *Proc. Multimedia '95*, 523-534.
- [14] Trigg, R., Suchman, L. & Halasz, F. (1986): "Supporting collaboration in Notecards", *Proc. CSCW'86*. ACM.
- [15] Vellon, M., Marple, K., Mitchell, D. & Drucker. S. (1998). The Architecture of a Distributed Virtual Worlds System. *Proc. 4th Object-Oriented Technologies and Systems Conf.*
- [16] Centra Symposium. <http://www.centra.com/>
- [17] Microsoft NetShow. <http://www.microsoft.com/>
- [18] Placeware Conference Center. /www.placeware.com/
- [19] Stanford Center for Professional Education. <http://scpd.stanford.edu/>