HDTV: A challenge to traditional video conferencing?

Gloria Mark[1] and Paul DeFlorio[2]
University of California, Irvine[1] and Jet Propulsion Lab, California Institute of Technology[2]
gmark@ics.uci.edu, deflorio@mail1.jpl.nasa.gov

The history of video conferencing has been one filled with promise. Yet from the origins of the idea in the 1930's, through the failure of the videophone in the 1970's, to the debate on its use today, we are still left with questions as to its value in supporting distance communication. Especially as research interest in video technologies begins to move beyond desktop conferencing, the support for conference room-conference room collaboration becomes a challenge. In this paper we present an experiment in which we examine how life-size HDTV as a "window" connecting two conference rooms might overcome some of the problems found with using traditional video conferencing in meeting rooms across distance. These problems include low resolution, narrow field-of-view, and lag. We are working with NASA[1] who is interested to use technology to support virtual collocation for project teams who are designing space mission proposals. We were interested to discover the impact that high telepresence (using life-size HDTV) in virtual collocation might have on team performance.

Video can overcome problems associated with audio-conferencing, such as not knowing who is present at the other end, or who is speaking, and can even convey basic nonverbal signals to govern speaking turns (Isaacs et al., 1995). However, low resolution and lag with current video systems hinder gestures from being naturalistically reproduced, leading people to resort to exaggerated hand gesturing (e.g. Heath and Luff, 1991). A well-known problem is that direct eye contact cannot be easily established. A video conference between two meeting rooms is a tradeoff between showing the breadth of the rooms or showing individuals but not the larger view with others. Either way, the remote viewer cannot clearly see all the detailed activity in the video image of the distant conference room.

We wanted to explore the consequences of using very high-end technologies for telepresence to overcome limitations of conventional systems in conference rooms. Traditional point-to-point commercial videoconferencing solutions are based on ISDN networking. A basic rate ISDN connection supports a data rate of 128kbps. Traditionally, using CIF (352x288) or QCIF (176x144) images, this will support an equivalent frame rate of roughly 5-10 frames per second (fps). We refer to this as "ISDN video". Standard television (or baseband video) signals provide a much higher resolution (around 440 x 480 in the American NTSC standard) and a frame rate of 25-30 fps. However, both of these suffer from limited field of view; without introducing distortion, standard video cameras have a field of view of 48.8 degrees. High-Definition Television, or HDTV, is a video standard that provides much higher resolution and a wider aspect ratio, so that the field of view is extended to 87 degrees. Because of its high frame rate, high resolution and wider field of view, HDTV would seem to provide a better basis for effective telepresence. We also chose HDTV that was life-size and wall-size, to create a sense of having a "window" into another meeting room.

Experimental set-up

We conducted the experiment using a specialized team of sixteen engineers that design space missions for NASA. Normally they meet face-to-face, but for this experiment, they were willing to divide the team into two separate conference rooms. Static video cameras were positioned at the front of each conference room, so that each room's respective camera-based coordinate systems would be mirrored by the other (figure 1). Both Room A and B cameras were located at the left-right center of the distant room projection. The camera lenses were positioned parallel to, and about 50" above the floor, to mimic the point-of-view of a seated person. The goal of this arrangement was to support consistent spatial or gaze geometry between the two rooms. In order to maintain a human scale to the projected images, the closest people were seated a minimum of six feet away from the camera, while people on the close periphery (which shows more aberration caused by lens curvature) were seated at least eight feet away. The closer objects (and people) are to the camera, the more apparent geometry mismatches become. For example, if you are standing seven meters from the camera and point at a person displayed on the screen showing the distant room, the remote person feels that you are pointing at them, even as they move about the room. If you are two meters from the camera, the remote person can readily tell that you are not looking or pointing at them if they are not near the room center line.

---

[1] National Aeronautics Space Administration Agency, with sites distributed around the U.S.

A wide angle lens with a focal length of five mm was used, providing a Field-of-view (FOV) of ~56 degrees vertical by ~87 degrees horizontal. The wide FOV was needed to capture the maximum amount of floor space (people) in each room. The short focal length lens also served to keep most of the room in good static focus. The minimum object distance for viewing was about 0.5 M., which could allow for the interactive high resolution display of material objects.

In Room A, a Panasonic high definition camcorder produced a 1920 x 1080I digital data output (60 interlaced frames a second). The camera in Room B produced a 1280 x 720P digital image at 60 full frames per second. Both displays used an image width to height ratio of 16:9. The size of both projections was the same. The 16:9 image was 128 inches wide by 72 inches high with the image beginning two feet above the floor and continuing to within 12 inches of the ceiling. This is significantly wider than the typical image display ratio of 4:3, and conveys a larger, more immersive image of the distant room without distortion. Telephones were provided in each room, with telephone numbers, to support *sidebar conversations*, i.e. small group discussions, between the sites.

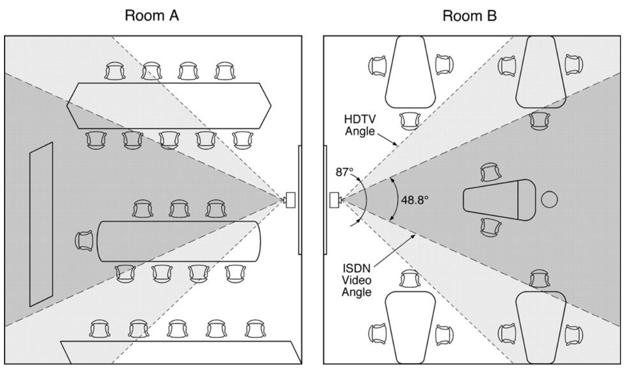

Figure 1. Field of view of HDTV cameras. People were only seated within the field of view.

On Day 1, a single customized binaural (stereo) microphone, placed around the video camera body, sent audio directly into the other room where it was reproduced via an amplifier and two tall floor-to-ceiling speakers. The video was compressed using MPEG2 and sent through a local Gigabit fiber network. There was a .8 second lag of video to audio. On Day 2, both the audio and video were digitized and transmitted as the video in Day 1. This resulted in synchronous audio and video, but with acoustic feedback, which was corrected mid-meeting.

The experiment was conducted over two sessions on Tuesday and Thursday of one week, three hours each session, and was an actual working session. In Room A were eight team members and the customers. In Room B were seven members and the team leader. Both authors observed the team, and activity was coded. Video-taping was not permitted on Day 1 by the customer, but was done on Day 2. Questionnaires were distributed to the team members after Day 2, and a one-hour group discussion was held with the team.

Results

The team leader sat in Room B and is shown in the HDTV image (fig. 2). In the authors' view, the life-size HDTV video produced an extremely compelling image in terms of high resolution, and showing breadth of the remote room.

Fig. 2. Room B in the HDTV experiment. The HDTV image appears small in the photo due to the wide-angle camera lens used for the photo. The team members in the video image in fact appeared life-size.

Perhaps one of the indications that video created a sense of telepresence was that team X members turned in their seats to face the video screen when speaking to remote team members, using social conventions as in face-to-face interaction. The audio created a spatial audio effect, i.e. it was easy to discern one's seating location in the remote room when a remote person spoke. However, when the team leader had his back to the camera, it was hard to hear him. Telephones were available (with phone numbers) to support inter-room sidebars, but were not used.

Questionnaire data. On a scale of 1 to 10 (high), the quality of the audio was judged by the team to be moderate (Mean=5.7, sd=2.2, Day 1 and 2 averaged, since there was no difference), and the HDTV video was judged higher, though not as high as we expected (Mean=8.0, sd=1.0, Days 1 and 2 averaged, since no difference). The engineers were asked how "present" they felt the team members in the other room were, and judged this to be moderately high (Mean=6.2, sd=3.1, with no difference on Day 1 and 2). Audio feedback was judged to be more of a problem on Day 2 (Mean=7.2, sd=1.9) than the .8 second lag on Day 1 (Mean=4.8, sd=2.6). Two members preferred the technology experience on Day 2, one member felt they were about the same, and the rest of the team preferred Day 1. Not unexpected, audio feedback was less tolerable than video that was out of synch with the audio.

Team members were far less satisfied with the distributed meetings (Mean=5.8, sd=2.1) than with a typical face-to-face meeting (Mean=9.4, sd=.9). There was no difference in satisfaction for the two experiment days, so results were combined and a paired-sample T-test showed the difference compared to face-to-face was highly significant $t(11)=-4.4$, $p<.004$.

Communication and sidebar conversations

The main problem that the team reported was that the video/audio did not support sidebar conversations across sites. Telephones were not used for these conversations. In the questionnaire, team members were asked what they did when they needed to engage in a discussion with another. Their choices were: *I did nothing, I walked over to a person in the same room to have a sidebar conversation, I walked over to a person in the other room to have a sidebar conversation, I spoke to the person through the technology connection, or other action*. Figure 3 shows the distribution of responses for Day 1 and Day 2. On Day 1, members claimed that on the average, 12% of the time they did nothing when they needed to have a sidebar, and perhaps through learning, this dropped to about 5% on Day 2. Both days, about 15% of the time the members reported that they walked into the other conference room. This behavior is further evidence that the technology did not support sidebars, as they had a real customer and did what they felt they needed to do to continue their work. Sidebars could be held using the HDTV and audio connection, and this was reported to have been done only about 12-15% of the time when sidebars were needed.

Sidebars were coded from the videotape of the Day 2 session, which clearly showed both rooms. In fact, compared to the 98 average number of sidebars coded during face-to-face meetings, only 74 sidebars occurred during the HDTV meeting. Fifty-six sidebars occurred in the same room, and 18 sidebars occurred when people crossed over into the other room to meet face-to-face.

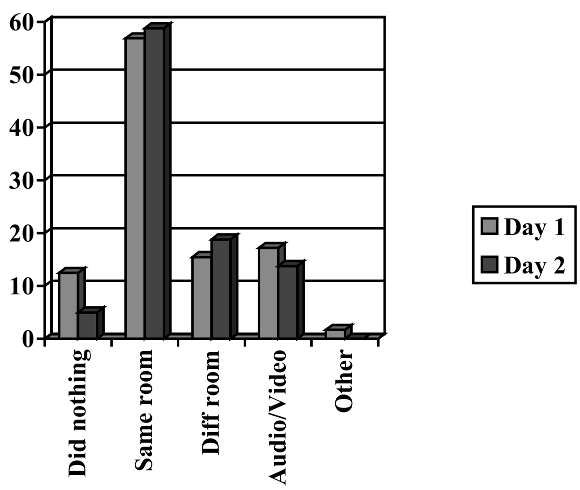

Figure 3. Estimates of sidebar activity during HDTV experiment .

If the HDTV/audio did not suitably support the team in terms of its sidebar conversations, what function did it have? The HDTV/audio appeared to have two primary functions for the team. First, it supported public conversation. As a single, directed channel, the HDTV/audio was very effective. Some sidebar conversations did occur through the HDTV/audio channel, but they became in fact, public conversations due to the nature of the audio channel. The team leader became a single primary channel of information for the group, as opposed to normal meetings, where he is wandering around the room, checking on progress and errors. The only information that he had as an overview of the team progress was the spreadsheet on a public display. He could see who was talking to whom, but he did not have detailed information, e.g. what was on people's screens, and sidebar conversations.

Second, the video image functioned as an awareness mechanism for activity in the other room. The HDTV showed all the action in the remote room. Perhaps one reason that the lag of video to audio was not judged very disturbing in Day 1, was that the HDTV image functioned primarily as visual awareness, rather than as a communication channel for the entire team. Team members saw who was speaking with whom through the HDTV image, and then walked into the other room (18 times) to join the conversation. However, activity through the HDTV image is not as salient as activity in the actual conference room, despite the best efforts to create high quality life-size video.

Discussion and implications

A virtual collocation environment as a window between conference rooms is a start, but is not enough. Neither telephones nor the single channel sufficiently supported sidebar interaction of members in different rooms. Interaction between sites requires far more effort than interaction within sites. The large ratio of within-room to between-room sidebar conversations support this, as well as the questionnaire responses. By missing out on between-room sidebars, the necessary human collaborative processing may not have been done during the experiment. This is a problem to be aware of in a virtual collocation setting. Within-room interaction will always be easier, and distributed work may suffer by losing out on between-room interaction.

It was our aim by using wall-size, life-size HDTV to overcome many of the problems found in interaction with regular video-conferencing systems. First, the HDTV conveyed a far wider angle of the remote room with less distortion, than normal video-conferencing. Members could move within a wide range in the room and still be captured by the video. Second, the resolution was a much higher quality than normal video. We did not observe any exaggerated gestures or movements to convey expression through the HDTV image, as Heath and Luff (1991) found with normal video. One engineer said that he saw the facial expression of a remote team member seated in the back of the remote room.

In the HDTV experiment, there were fewer sidebars overall (as a result of few between-room sidebars), which could have several implications for virtual collocation. First, the amount of automated information processing could be higher (e.g. email). Second, less oral processing of information could impact the quality of the results, as less options are explored, less assumptions are questioned, and less errors are caught and corrected. Third, members between rooms must rely on the single channel (team leader) for information on progress. There is the higher chance that the two meeting rooms can fall out of synch due to the limited information flow between rooms. The burden falls more on the team leader to keep both teams assessed of progress, design changes, and errors.

Stress has been reported as a problem by many of the team members. There is a large amount of information processing occurring during a session. By connecting teams across distance, e.g. through an HDTV/audio channel, it increases the amount of information processing even more, when it opens up a window into another room. It is more difficult to attend to the information on a screen, compared to physical activity in the same room.

It is a challenge in a virtual collocation environment to support sidebars that people intentionally initiate, but also sidebars that occur spontaneously. The HDTV video may show who is speaking, but there is so much audio information even in the same room that it is hard to discern content across distance. Even though the video shows who is talking with whom, the act of making the connection across distance loses the spontaneity found in a face-to-face environment. Technology design for virtual collocation needs to consider finer-grained support for interdependencies among group members, in terms of information exchange, and not just a broad single channel.

References

Heath, C. and Luff, P. (1991). Disembodied conduct: Communication through video in a multi-media office environment. *Proc. of CHI'91*, pp. 99-103.

Isaacs, E.A., Morris, T., Rodriguez, T.K. and Tang, J.C. (1995): “A comparison of face-to-face and distributed presentations”, *Proceedings CHI’95,* Denver, ACM Press, pp. 354-361.