Chapter 2: Telepresence Technology Elements

Videoconferencing has been around for two decades. During that time, many customers have focused on the potential for videoconferencing to reduce corporate travel, increase worker productivity, speed communications, reduce executive stress, and boost the effectiveness of dispersed teams. Yet, wide-scale adoption has been slow to materialize. The most common objections to videoconferencing include that the systems are too hard to use, too unreliable, and in the end, do not provide a viable or realistic alternative to in-person meetings.

These are the objectives that today's telepresence systems are designed to overcome. To achieve this goal and provide an experience that approximates a face-to-face meeting, all telepresence systems use a varying (and sometimes overlapping) mix of audio, video, and integration techniques and technologies. While each of these can be considered valuable by itself, the full effect and benefit of telepresence occurs when these items are combined to produce the desired illusion.

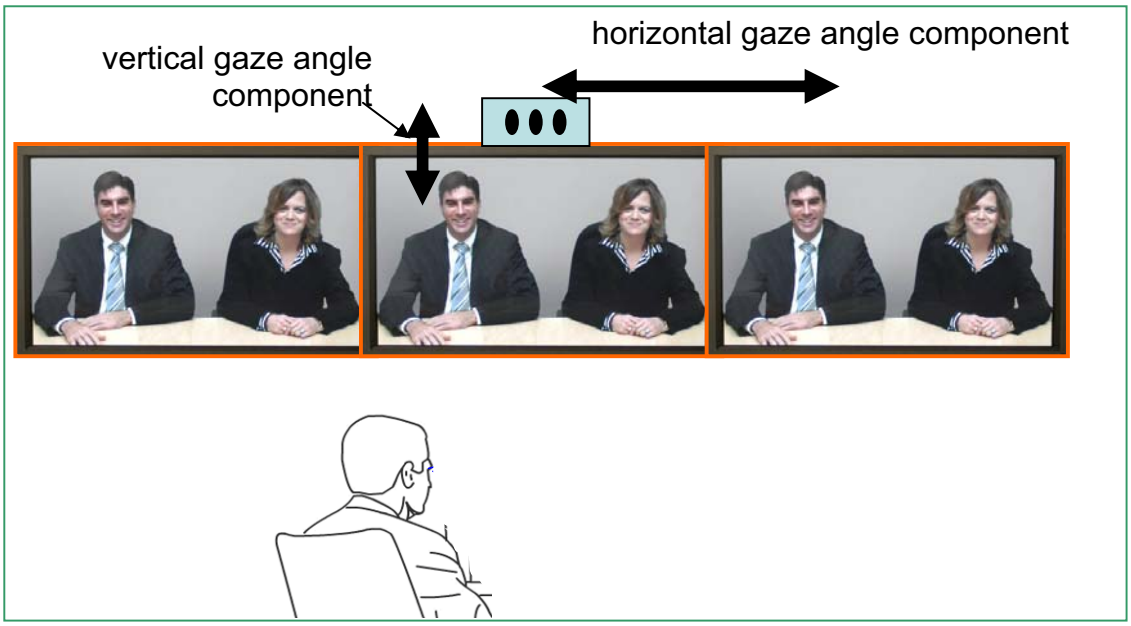
Telepresence Video Elements

Many elements go into the design of a telepresence video subsystem. To produce a pleasing meeting experience, telepresence video signals must be clear, provide life-like color, and support fluid motion without producing video artifacts. Meeting these goals requires high quality cameras and powerful codecs. To further enhance the meeting experience and differentiate a trelepresence session from a traditional videoconference, many systems have moved beyond the videoconferencing standard of CIF resolution to higher pixel counts ranging from 4CIF to 720p high definition (9 x CIF) to even 1080p high definition (18 x CIF). Higher resolution images enable viewers to sit closer to the display without perceiving pixelization while also providing a level of image detail that most people would associate with an in-person meeting.

A second video factor embraced by nearly all telepresence suppliers is the need for life-size images of remote meeting participants. Staring at a small face during a videoconference is a constant reminder that the meeting is taking place electronically. To provide life-size images, telepresence rooms balance the size of the display screens, the number of cameras, the focal length of the camera lens, the number of meeting participants, and the physical distance between the meeting participants and the displays. In order to eliminate the camera panning and zooming that are distracting to most videoconferencing users, telepresence systems typically employ multiple fixed cameras that are optimized to cover the entire meeting room space (without gaps), while still providing life-size images of individual faces.

Different telepresence vendors have paid varying degrees of attention to the issue of eye contact, or rather the lack thereof. Eye contact is one of the most important aspects of face-to-face communications. Eye contact instills trust and fosters an environment of cooperation and partnership; lack of eye contact, on the other hand, can generate feelings of negativity, discomfort, and sometimes even distrust. Providing natural feeling eye contact during a videoconference requires that the participants look directly into the camera. However, the design of most videoconferencing environments encourages users to instead look into the face of the remote participant on the display screen. If the camera and remote face are not in the same position, then eye contact cannot be established. This is known as the gaze angle problem, and it can exist in both the vertical and horizontal planes. A few vendors have solved this problem using beam/splitter mirrors that enable the camera to be positioned directly behind the display. Hence when the local participant looks at the display, he is also looking directly into the camera.

These solutions are effective at providing true eye contact, but are also costly and require some form of rear-projection.



Gaze angle is the angle between looking into the camera and looking at the remote participant's eyes.

Telepresence Audio Technologies

While most people focus on the "video" in videoconferencing and telepresence, audio is equally, if not more important. Studies have shown that poor audio quality measurably increases the stress level of meeting participants; this is a key consideration for telepresence solution vendors because customers are specifically paying for a low-stress meeting experience. There are four significant factors that contribute to the perceived audio quality during a conference: the clarity / quality / consistency of the sound, the "apparent" location of the sound source, the latency or lag time, and the synchronization between the sound and the video.

An important benefit of digital audio (all videoconferencing and telepresence systems are digital systems) is the ability to support wideband voice. Similar to the difference between AM and FM radio, wideband voice provides crucial improvements to the meeting experience. Ordinary telephones, which are limited by the bandwidth of the public phone network to the band from 300 Hz to 3.3 kHz, carry only 20 percent of the frequencies present in human speech. Telepresence systems today often support 7 kHz, 14 kHz, and even 20 or 22 kHz wideband audio. The range of audio frequencies transmitted is a strong determinant of conference quality. The lower frequencies in the voice range provide a sense of presence, while the higher frequencies provide much improved speech clarity and intelligibility. These improvements in turn lead to a significant reduction in "meeting fatigue."

A natural meeting experience requires that a person's voice appear to be coming from his or her mouth. While this may seem extremely obvious, many conference rooms utilize ceiling mounted or peripherally located speakers for conference audio. This approach makes the far-end speaker's audio seem to be coming either from above and behind the meeting participants, or in some cases from no perceivable location. To avoid this problem, the speakers in a telepresence system are typically located discretely near the displays showing the far-end participants. Some systems

have taken this one step further by introducing directional or spatial sound to the telepresence experience, similar to the consumer experience with DVDs, video games, and Dolby 5.1 music. With multiple microphones and multiple speakers, it is now possible to give telepresence users audio cues about the location of the remote participant. The audio from the participants shown on the left projection screen sounds like it's coming from the left side of the room. While the effects here are subtle, they further reduce meeting fatigue and contribute to the life-like feeling of the session. Implementing spatial audio requires careful alignment of the speaker, microphone, and camera subsystems, and can involve complex acoustic echo cancellation techniques.

To provide participants with a natural-feeling meeting experience, the remote participants' lips and the sounds one hears from that person must be in sync. When proper lip sync is in place, the electronic technology in use to transmit and reproduce the sound is somewhat transparent. Conversely, when significant lip sync problems occur, the video element becomes more of a distraction than an added benefit. Telepresence systems exhibit tight control of lip-sync.

In the conferencing world, one-way latency is defined as the time between when something is said by one participant and subsequently heard by a remote participant. Measured in milliseconds (ms), latency is introduced into a videoconference by the compression/decompression technology (codecs), by the performance of the network, and in some cases by the use of external multipoint bridges / MCUs. Interaction is the ability of two or more participants to interact naturally with each other in a remote conference. It is essential that one talker be able to interrupt the others without disturbing the flow of conversation. Otherwise, the dialogue feels stilted and unnatural. Latency below 50 ms is barely perceptible, while at 200 – 250 ms and above, latency can become annoying. As latency increases beyond this threshold, interactivity suffers, the ability to have a natural two-way conversation diminishes, and eventually a real-time two-way dialogue becomes virtually impossible. Telepresence systems typically utilize a variety of techniques, including the use of low-latency codecs and high QoS / high performance networks to minimize latency.

User Interface Technologies

Telepresence suppliers are keenly aware of a major complaint about videoconferencing – the systems are too hard to use. With an intended audience of senior executives, telepresence vendors have gone to extremes to make these systems suitable for non-techie users for whom time is their most valuable commodity and tolerance for delayed meetings is minimal. Most telepresence systems operate in a "closed" environment in which the number of user options and call settings is minimal; in fact, many systems can operate with just one button to push. Almost everything within the system is either automated or fixed. Many telepresence vendors also provide network services to ensure that the needed bandwidth and quality of service is always available and that users will never hve to deal with error codes, less than optimal audio/video quality, "call did not go through" messages, or unexpected disconnects.. Some telepresence systems come complete with a concierge service 24 / 7 support, call scheduling, call set up, and remote monitoring and management, as well as system maintenance. In this case, the concierge does everything on the user's behalf and there are zero buttons to push. Besides making the user interface extremely easy-to-use, such methods also make the telepresence system highly reliable and the user experience highly consistent.

Telepresence Environmental Designs

A key element in telepresence is the immersive and consistent environment; hence telepresence systems employ many facets of environmental design to foster an improved meeting experience.

Many systems intentionally make the technology invisible to the user: microphones, speakers, cameras, etc. are hidden to the extent possible following the belief that these items are not present in an in-person meeting and should therefore not be present in a telepresence meeting.

Most telepresence vendors and installers pay careful attention to room lighting and acoustics. For example, reverberation caused by the room echo that occurs between the talker and the microphone can be very annoying. Reverberation can be minimized by room design, microphone placement, wall coverings, and other things. Besides speech, microphones also pick up ambient noise from HVAC systems, elevator shafts, and even highway traffic. Ambient noise decreases the intelligibility of speech, so telepresence vendors typically specify the recommended noise criteria (NC-35 for example) or provide acoustic-dampening wall treatments.

In some telepresence systems, the tables and chairs are designed to "force" participants to sit in the location where lighting and sound are optimized, and where their images will appear life size to the remote side and not be split between two displays. Many vendors offer a mirrored environment – all rooms are equipped with identical furniture, features, and design elements - further adding to the sensation that the remote people are indeed in the same room. Some systems go a step further, with a table that is designed to "merge" into the display, giving the distinct impression that the remote table is actually connected to the local table, i.e…. the impression that the two tables are really one.

We have classified most systems discussed in this report into two categories.

1) Room replacement systems are those in which the vendor provides all the needed elements, including lighting, wall coverings, furniture, etc., to turn a base-building space into a telepresence suite. After one of these systems is installed, users would typically not recognize the room.

2) Room enhancement systems require the customer to provide a suitable (in terms of sound parameters, lighting, etc.) meeting space in which the vendor installs his telepresence system, ancillary electronics, and the basic "telepresence table." Such deployments tend to maintain the look and feel of the enterprise meeting room, and users entering the space are able to recognize elements of the previous conference room.

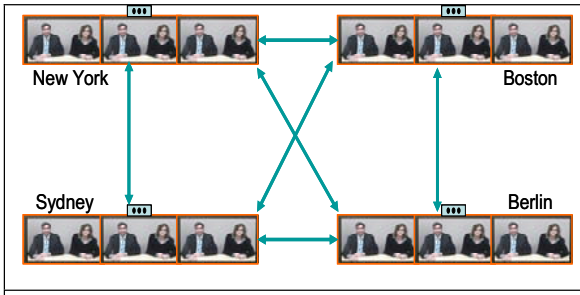
Support for Multipoint

Multipoint (3 or more location) meetings in the telepresence world are supported using one of two general approaches.

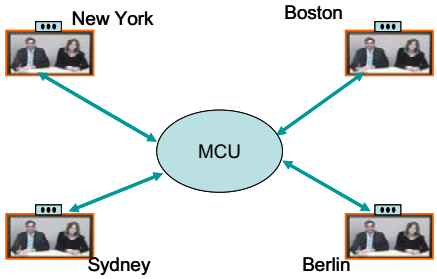
MCU-based: Traditional videoconferencing systems use a multipoint control unit (MCU), also known as a video bridge, to support multipoint meetings. With a multipoint bridge, the bridge receives the audio / video / content signals from each participating site, mixes the signals into the correct format, and then outputs a single signal (or stream) to each location. This stream could be a full-screen image of just the current speaker (a mode called voice-switched), or a "Hollywood squares" type of display (known as continuous presence) showing a number of participating locations simultaneously. The important concept is that each system needs to send only a single stream to the bridge and receive only a single stream from the bridge. In the case of a telepresence suite, in which three cameras (and video codecs) might be installed in each of four rooms, an MCU would treat this as a 12-way multipoint call. A great deal of switching intelligence would be relegated to the MCU that has to put images from Tokyo, Dallas, and

London on screens 1,2,3 in New York; images from Dallas, London, New York on screens 1,2,3 in Tokyo, etc. as well as handle the separate audio streams.

Non-MCU-based: Most telepresence systems today send and receive multiple streams directly between endpoints, which allows multipoint meetings to be supported without a centralized video bridge. In a four-way call, site A would send three video streams, one to site B, one to site C, and one to site D. At the same time site A would receive streams from each of those three sites and display those video images on its three screens. The same process would take place at each of the other sites.



Multipoint configuration typical for telepresence systems; each site has three connections to the other sites; six calls are involved



Traditional multipoint configuration for videoconferencing; each site sends and receives one stream to the MCU

The advantage of the centralized MCU method is lower bandwidth requirements (each site has to send and receive only a single stream) and as a result lower per-site costs (assuming an MCU is available for use). The advantage of the non-MCU-based (or peer to peer) method is decreased latency and improved interaction vs. the MCU approach; in addition there is no MCU to purchase.

Support for Data Collaboration

Data collaboration support in telepresence sessions varies by vendor in terms of technologies used, level of integration, and ability to involve non-telepresence sites / participants. Some vendors encourage users to conduct a separate, parallel web conference (typically using a 3rd party solution). Other vendors allow a user to connect the VGA signal from his laptop and hit a single button to begin data sharing the PC-image with the remote location(s). Some telepresence vendors also include additional data collaboration tools, such as a document camera or whiteboard, as a standard or optional part of the solution. This capability is well suited for certain vertical markets including manufacturing.

In terms of displaying the data images received from a remote site, different vendors have chosen different display approaches. Interestingly enough, the vendors all claim that their approach is based on the results of extensive user surveys and focus groups, and was selected because it was the favored approach. Cisco and Tandberg, for example, display remote data on a screen below the video screen (Cisco uses a projector, and as of this writing, does not support integrated data conferencing). HP's Halo design incorporates a data collaboration screen located above the main video screens. Teliris and others have data displays at the same eye level as the video displays, but located at the sides of the room. Polycom has taken a totally different approach using 15 inch LCD displays embedded in the meeting room tables.