Classification and Analysis of 3D Teleimmersive Activities

Ahsan Arefin, Zixia Huang, Raoul Rivas, Shu Shi, Pengye Xia, and Klara Nahrsted
University of Illinois at Urbana–Champaign

Wanmin Wu
University of California, San Diego

Gregorij Kurillo and Ruzena Bajcsy
University of California, Berkeley

To provide users with a high-quality experience, interactive telepresence system platforms must accommodate multiple performance profiles for diverse, shared cyberphysical activities.

With the increased availability of high-speed wired and wireless networks, current interactive telepresence systems such as Skype, Google mobile video chat, and Cisco telepresence are becoming an integral part of users' lives. To achieve optimal performance, such systems are optimized for a single activity (generally conversation). However, emerging 3D teleimmersive (TI) systems enable geographically distributed participants to engage in multiple, shared cyberphysical TI activities (such as conversation and collaborative dancing) with diverse physical characteristics and cyberperformance profiles using the same TI system platform

Our research team at the University of Illinois at Urbana–Champaign and University of California, Berkeley, has developed a multisite TI system called Teeve (Teleimmersion for Everybody). Using Teeve, we have explored various interactive activities ranging from conversational to collaborative fine- and gross-motor activities. For each new activity, we have qualitatively measured observable and controllable quality of service (QoS) parameters that influence the user's quality of experience (QoE).

Our qualitative analysis of TI activities revealed that it is not possible to design one performance profile that provides the best QoE for diverse TI activities over the same TI system platform. Thus, based on our analysis of Teeve activities and their corresponding QoS parameters and QoE responses, we propose a cyberphysical TI activity classification and a performance profile recommendation for each class of TI activities. These performance profiles, maintained within a TI system, ensure strong QoE depending on the activity types. Finally, we argue that to achieve these customizable performance profiles during runtime, we require highly configurable, programmable, and adaptable TI system platforms.

Teeve System Platform

The Teeve system is a multiparty, 3D TI system platform connecting multiple remote sites into one virtual shared space (see Figure 1). Figure 1a shows the Teeve system architecture, which consists of three architectural tiers: capturing, data dissemination, and rendering.

In the capturing tier, multiple capturing devices such as cameras, microphones, and other sensors (not shown in Figure 1a) capture each participant's cyberphysical multimodal information at his or her physical site. The captured multimodal data is synchronized and tagged with their temporal and spatial correlations, creating a bundle of audio, video, and sensory streams.[1] These cyberphysical, spatiotemporal correlated streams are called a *bundle of streams*.

The data dissemination tier consists of a peer-to-peer overlay network of gateways multiplexing a bundle of streams at each site. Each gateway is responsible for disseminating local bundles to other remote sites over Internet2.

In the rendering tier, multiple devices render different views of an ongoing activity and provide sensory feedback to users. Depending on the participants' views and their perceptions in the virtual space, the priority of the streams inside the bundle (called the *importance of modality*) can differ among viewers. Before the final rendering, streams in each bundle must be resynchronized due to various Internet dynamics (such as jitter and loss).

Figure 1b shows an example of a Teeve physical setup for a telehealth activity.

1070-986X/13/$31.00 © 2013 IEEE

Published by the IEEE Computer Society

Methodology

Our goal is to correlate performance profiles of a TI system to its cyberphysical TI activities. However, to achieve this goal efficiently, we first need to classify TI activities according to their physical characteristics and then measure their user-perceived QoE by varying underlying controllable QoS parameters. The steps are as follows:

1. Analyze physical-movement-based characteristics of TI activities and group them into TI activity classes.

2. Identify cyber QoS parameters to evaluate TI activities in virtual spaces. We observe and measure these parameters to determine performance profiles for each TI activity class.

3. Employ measurement and evaluation methods to obtain correlations between QoS and QoE values for each TI activity class.

Physical Characteristics for TI Activities

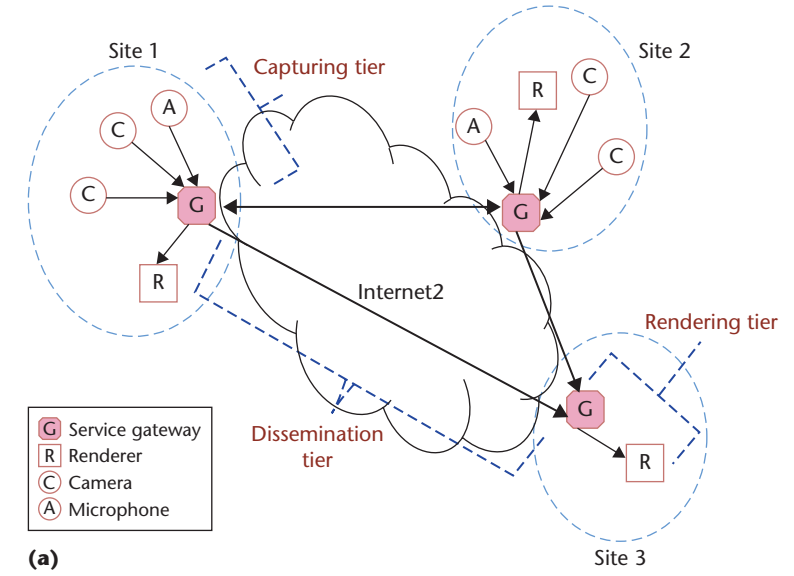
Because we are considering cyberphysical activities within the TI platform, we classify TI activities according to three movement-based characteristics: *space coverage* (the physical space coverage for each participant's individual movements), *speed* (the speed of the participant's movements), and the *interaction type* among participants in the virtual space.[2]

The space coverage depends on each participant's movement. Each participant's movement space can range from small (standing still) to large (jumping). For example, during *exer-gaming activities* (gaming activities involving physical exercises), participants might use the whole allocated physical space. However, during videoconferencing, participants only use a limited amount of physical space.

The movement speed can vary from slow to fast. For example, participants might move quickly during exer-gaming activities and slowly during the virtual teaching of an engine repair.

The interaction type among remote participants is highly influenced by the participants' intention. Depending on the muscles that participants use—for example, fine muscles control finger movements and gross muscles control leg movements—we can classify the interactions as fine- and gross-motor skill interactions. Fine-motor skills require a precise interaction among the participants, such as pointing out an engine's position. On the other hand, gross-motor skills require whole body coordination, such as arm and leg movement during virtual fencing.

Utilizing these movement-based characteristics, we can classify TI activities into three classes: conversational, fine-motor collaborative, and gross-motor collaborative activities. A *fine-motor collaborative activity* is usually defined as a collaborative activity that requires participants to use fine muscles along with different combinations of movement and



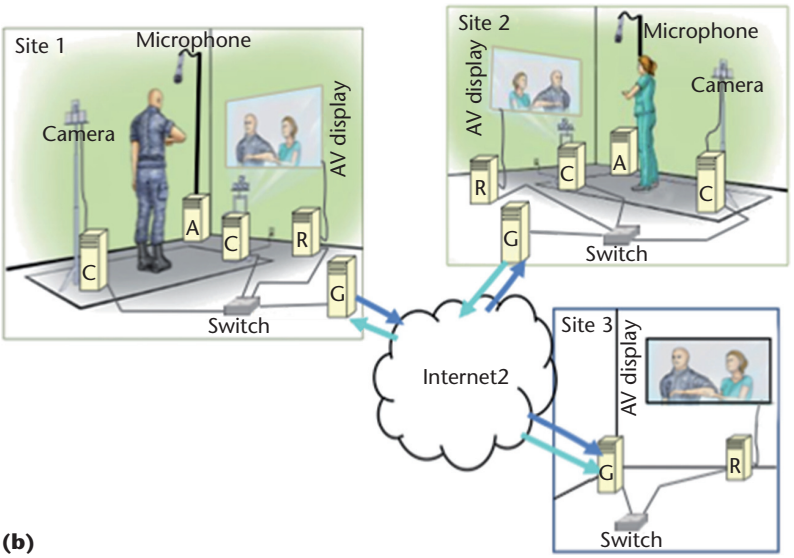


Figure 1. Teeve system. (a) The architecture consists of capturing, data dissemination, and rendering tiers. (b) Example physical setup for a telehealth TI activity.

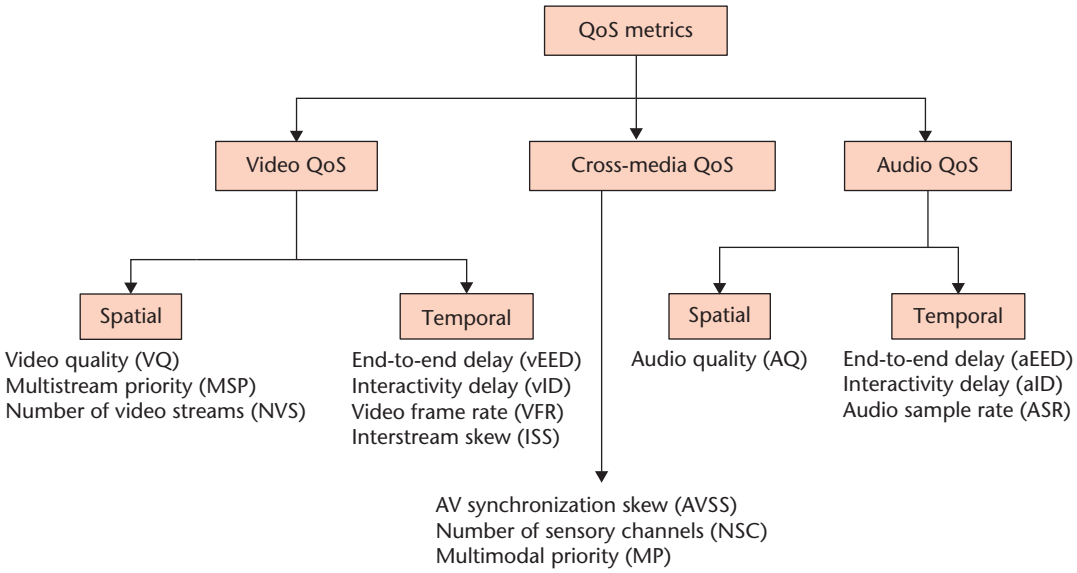


Figure 2. Activity QoS categories. The hierarchy of video QoS, audio QoS, and cross-media QoS categories and our selection of parameters for each.

Table 1. Definitions of activity QoS categories.

Cyber aQoS metrics	Definition
Video quality (VQ)	The spatial video frame resolution, measured in number of pixels per frame, bits per pixel, peak signal-to-noise ratio (PSNR), and color-to-depth level of detail (CZLoD).[3]
Multistream priority (MSP)	The priority among video streams in a bundle of streams.
Number of video streams (NVS)	The number of video streams in a bundle of streams.
End-to-end delay (EED)	The time interval between when a media frame is captured and when it is displayed for both audio (aEED) and video (vEED) media.
Interactivity delay (ID)	The round-trip delay, representing the length of time between the moment a user issues an interactive request and the moment the user receives a response for both audio (aID) and video (vID). The ID value can be more than two times the EED because ID also includes the time duration for updating and providing user feedback.
Video frame rate (VFR)	The application frame rate of a video stream.
Interstream skew (ISS)	The skew between streams of similar modality. Currently, we measure ISS among video streams.
Audio-visual synchronization skew (AVSS)	The perceptual skew between correlated audio and video frames.
Number of sensory channels (NSC)	The number of sensory devices used to construct immersive experiences.
Multimodal priority (MP)	The importance of a modality—that is, defining which modality is more important than other modalities.
Audio quality (AQ)	The quality of an audio signal as defined by the standard ITU-T P.862, Perceptual Evaluation of Speech Quality (PESQ).
Audio sample rate (ASR)	The application frame rate of an audio stream.

spatial coverage. A *gross-motor collaborative activity* is defined as an activity that requires participants to use gross muscles along with different combinations of movement and spatial coverage.

Activity Quality of Service

Variations in physical movement characteristics result in variations in activity QoS parameters and hence the user's perceived QoE. To organize the activity QoS (aQoS) parameters in our context, we divide them into three major categories: video QoS, audio QoS, and cross-media QoS. Figure 2 shows the hierarchy of QoS categories and the selection of parameters from each category in our analyses. Table 1 provides their definitions.

Evaluation Methods

The International Telecommunication Union (ITU) has prescribed several recommendations

Table 2. Considered activities for qualitative analysis.

Activity name	Space coverage	Movement speed	Interaction type	Activity class
TI conversation	Small	Slow	Fine	Conversation
TI archeology	Small	Slow to moderate	Fine	Fine-motor collaborative
Mobile block fencing	Small to moderate	Slow to moderate	Fine	Fine-motor collaborative
Virtual fencing	Large	Fast	Gross	Gross-motor collaborative
Collaborative dancing	Large	Fast	Gross	Gross-motor collaborative

for evaluating perceptual quality of videoconferencing systems that serve as useful TI guidelines. Unfortunately, little is understood about the impact of various QoS configurations on different TI activities in terms of QoE. This motivated us to perform our own evaluations. Our evaluation methodology involves three steps:

1. Perform objective evaluation of QoS.
2. Perform subjective evaluation of QoE.
3. Find correlations between QoS and QoE measurements.

Objective evaluation of QoS requires a service for active QoS monitoring. To measure and collect various QoS values in Teeve, we implemented a monitoring service that records the objective QoS values at runtime and allows distributed range queries from different TI sites.[4]

During and after activities, human participants perform a subjective evaluation of QoE by completing a questionnaire. This evaluation utilizes methods that record user responses, such as their perception of the video quality and their concentration level during activities.

Finally, we correlate objective QoS measurements with subjective QoE evaluation, for example, using comparative methods to find functional relations between user experiences and QoS configurations and resource allocation.

Here, we present two examples of our evaluation methods. In the first example, which considers conversational and fine-motor collaborative activities, we evaluated and compared activities by showing users different activity videos with diverse QoS configurations. We collected the users' perceived QoE in the form of comparative mean opinion scores (CMOS).[5] Using users' responses, we correlated the QoS parameters to the CMOS values. For example, the correlation mapping between the video frame rate (VFR) (r) and the associated CMOS value is represented by this exponential function:

$$\mathrm{CMOS} = Q - Q \times \frac{1 - e^{-c \times \frac{r}{r_{\max}}}}{1 - e^{-c}} \quad (1)$$

where $r_{\max}$ is the maximum achievable VFR, and Q and c are constants, highly dependent on the activity types.

In another example of fine-motor collaborative activities, we measured the perceptual thresholds of the visual quality. We used the Ascending Method of Limits from psychophysics to measure the *just noticeable degradation* and *just unacceptable degradation* thresholds on the color-plus-depth level of detail (CZLoD) spatial resolution video quality (VQ) parameter. We found that 70 percent of CZLoD degradation is imperceptible to the human eye.[3]

Qualitative Analysis of TI Activities

We performed a qualitative analysis of TI activities in Table 2 by running them within the Teeve testbed. We followed the methodology we discussed in the previous section and analyzed TI activities with respect to their objective QoS and subjective QoE assessment to gain an understanding of their possible correlation and interpretation.

Teleimmersive Conversation

As Figure 3 shows, the Teeve system enables geographically distributed users to walk into their individual TI physical spaces and engage in a conversation in a shared virtual space. In our TI conversational experiment with Teeve, users faced cameras with limited user movement. During this activity, participants concentrate on each other's faces, lips, and body language.

The Teeve system platform has enabled conversations among users located in Illinois, California, Texas, and Florida in the US as well as in Amsterdam. We configured each

Figure 3. Example teleimmersive conversation scene. The two remote participants share a virtual space.

site with a 61-inch NEC screen, offering 640 × 480 multiview video rendering of the immersive environment. To allow high-fidelity speech communication, we equipped users with wireless and wired headsets with a microphone input. The wideband speech signals were encoded with a 44 kilobit per second (Kbps) data rate, using the wideband Speex library. A four-channel microphone array was an add-on capability to capture ambient sounds, which was encoded using Advanced Audio Coding (AAC). We used a passive stereo[6] at the 3D cameras to capture participants' 3D images.

We analyzed the impact of a single-objective activity QoS parameter such as VFR, audio-visual synchronization skew (AVSS), and end-to-end delay (EED) on a subjective metric such as CMOS by keeping values of other QoS parameters optimal. In previous work,[5] we showed three findings. First, the correlation between VFR and CMOS parameters was exponential (see Figure 4a). We used an exponential model (Equation 1) to describe the resulting fitting curve, where $Q = 2.52$ and $c = 2.16$. Second, audio-visual synchronization skew (AVSS) was of great importance. Ralf Steinmetz stated that an audio-visual skew of more than 160 ms is noticeable in a videoconferencing system,[7] and our subjective study within Teeve confirmed this finding (see Figure 4b). Our study also confirmed that people were more tolerant of video ahead of audio than to audio ahead of video. Third, users experiencing long EED became impatient easily, and doubletalk occurred if users did not hear from the remote party within their maximum expected latency. Figure 4c shows the correlation mapping between the EED and the corresponding CMOS degradation. It shows that an EED larger than 400 ms leads to a poor interactive perception in the conversation.

Teleimmersive Archeology

We designed a TI archeology activity to enable archaeologists at different geographical locations to meet in the virtual space; examine, measure, and interact with digitized artifacts; explore a virtual site; and visualize maps and other data. Participants use hand, gesture, and finger fine-motor skills to interact with each other and the digital artifacts. Figure 5a shows two users collaborating in the virtual

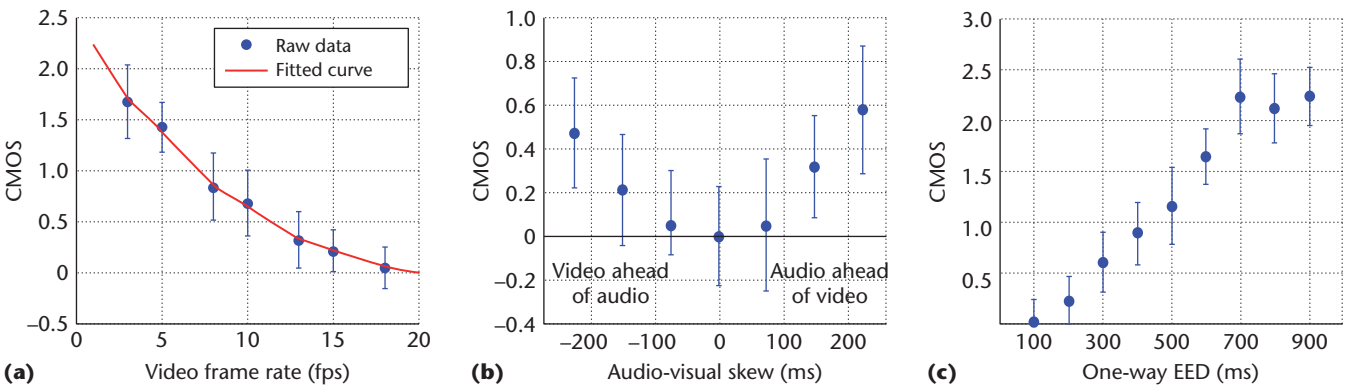
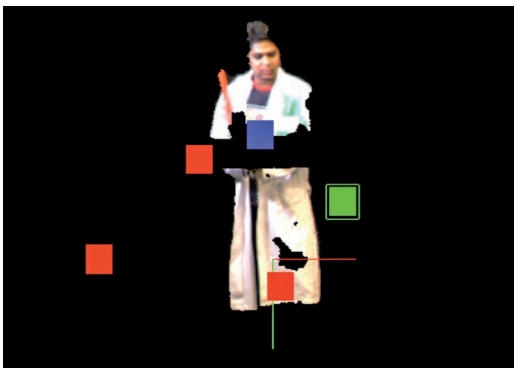


Figure 4. Comparative mean opinion score (CMOS) results as opposed to the optimal reference—4.5 Perceptual Evaluation of Speech Quality (PESQ), a 20 frame per second (fps) video frame rate, and zero audio-visual skew and interactive latency. (a) Impact of video frame rate deduction, (b) impact of audio-visual skew, and (c) impact of increased one-way, end-to-end delay.

(a)



(b)

Figure 5. Example collaborative TI activities using fine-motor skills. (a) TI archeology, and (b) mobile gaming.

reconstruction of a Mayan temple. The activity includes several collaborative elements:

- Each user observes the model from a first-person perspective and can thus navigate freely to various locations in the virtual space. For navigation, users change their position and orientation in the space using their 6 degree-of-freedom (DOF) input device, so the physical space coverage is small.
- When a new artifact object is loaded into the scene, both users can see the object in the shared 3D space.
- One users manipulates the artifact object while the other guides the first user by speaking and/or pointing to the object.
- The two users discuss whether the artifact's location seems appropriate with respect to the model and aim to interpret the meaning of the location.

We experimented with the TI archaeology activity at UC Berkeley.[8,9] To facilitate archeological interaction, independent of hardware, we implemented the collaborative architecture using the Vrui VR Toolkit at UC Davis.[10] With its collaborative extension, the toolkit provided an abstraction of display and input devices. We also performed remote experiments between UC Berkeley and UC Merced. At both locations, we set up five camera clusters (with passive stereo) focused primarily on the frontal part of the users, a 3D display with head tracking, and a 6-DOF input device (Wii remote with position and orientation tracking) for interaction. Each user was able to locate the other user's exact location based on the remote user's 3D avatar. Because we precalibrated the display, cameras, and tracking systems, users could point and interact with a specific part of the model on the 3D display and then see his or her avatar pointing at the same point in space. The video resolution was 320 × 240 pixels to provide a frame rate of 18 to 20 fps.

These experiments showed that being able to see details on remote users' faces was not as important as being able to see the locations of their hands when they gestured and pointed. One of the crucial elements of the interaction was the interactivity delay (ID) between the tracking system (which moves the objects) and the 3D video stream (which creates the remote avatar). With a significant delay, users could not interpret the other user's location with respect to the objects. Small interactivity delays (less than 200 ms) in response to rendering actions were tolerable, whereas large delays (more than 500 ms) caused a disconnect in the user interactions.

A high EED also disappointed users, and they were sensitive to the number of sensory channels (NSC). For example, we initially experimented with rendering on a 2D display, but we found that gestural interactions were difficult because the users could not tell which part of the 3D model they were pointing to. Using the 3D display with head tracking (which improved the NSC parameter), on the other hand, users could interact more naturally with the digital models and better interpret their dimensions and geometry.

Virtual Mobile Block Fencing

Virtual mobile gaming lets users watch a TI shared space on their mobile phones and interact with remote participants in the virtual world. The block-fencing game in Figure 5b is one such example.

This activity includes two different actions. First, the mobile user observes the game on

his or her mobile phone and changes the rendering viewpoint. The mobile phone sends the viewpoint update request to the TI rendering server, and the server starts to render the 3D video from the new viewpoint. Because the mobile device has limited network bandwidth and computing resources, the rendering of 3D video is performed on the TI rendering server and the rendering results are streamed to the mobile phone as 2D images.

Next, the mobile user can interact with a remote user by adding virtual blocks by touching the screen. The touch event is sent back to the rendering server, and a block element is drawn in the virtual world. When the new rendering result with this virtual block is displayed on the mobile device, the block position should be exactly where the user touched the screen. The remote TI user observes the added virtual blocks on the big TI display. The user can touch the block by moving his or her hands close to the block in the virtual world. Once a hand overlaps with the virtual block on the TI display, the block is considered touched and will be removed soon after.

We used one Teeve site and one iPhone in our virtual mobile block fencing activity. The TI system was connected to Internet2 through an Ethernet local area network (LAN), while the iPhone was wirelessly connected through a Wi-Fi network.

With this activity, we focused on the ID evaluation on the mobile side. In a remote rendering system like our experiment setup, ID was no less than the network round-trip time between the iPhone and the rendering server. ID was not always noticeable if both the iPhone and server were on the same LAN, but ID became unacceptable if the iPhone and rendering server were remotely connected through cellular networks (such as 3G). ID greater than 100 ms impaired the user QoE at different levels. To reduce ID, previous work enabled the mobile user to display the rendered avatar image at the new viewpoint before the server updates arrived, using warping and prediction techniques.[11] This approach reduced ID significantly. However, QoE was also influenced by the image's VQ, so such systems must consider both ID and VQ.[12]

Virtual Fencing

We have also implemented a virtual fencing activity with two players in geographically distributed locations using Teeve. The two players use physical swords to engage in a virtual duel. The use of 3D video improves the overall QoE[13] because players not only play the video games but also become part of them.

To engage in virtual fencing, participant 1 at site 1 puts on a lab coat with red patches and takes a green light saber, and participant 2 at site 2 puts on a coat with green patches and holds a red light saber. Next, each participant uses the light saber to hit the opponent's color patches in the virtual space as much and as fast as possible to earn points. When a hit occurs, the sword and the coat patches in the virtual space turn blue, and the participant making the hit gets points. The participant who gets hit received haptic feedback through vibration and lightning on his or her sword . The participants can move and even leave the space to avoid a hit.

Our Teeve experiment used a two-site TI system over a gigabit LAN. Each site was equipped with a 3D camera, microphone, speaker, and rendering display.

As previous work showed,[13] ID was severely impacted when EED was higher than 100 ms. In these cases, players could not hit their opponents accurately. Adults and children experienced problems with the system's consistency in the presence of such a noticeable latency. Increased EED led to confusion and reduced user concentration, causing some users to stop playing. Figure 6a shows the EED values (which consist of network dissemination, rendering, and 3D reconstruction delays) were bounded by 90 ms, and the most contributing factor was rendering delays at the output devices.

Due to the nature of gross-motor activities, VQ was only important in localized screen areas. For example, the VQ was only important in the area around the swords. Facial features were not important because the participants' main perceptual focus was on the swords. We found that users could successfully perform the activity with a video frame resolution of 320 $\times$ 240 pixels. Similarly, the VFR does not need to be very high. Participants did not feel any variation in the perceived QoE with the VFR variations between 14 and 20 fps, as Figure 6b shows. Therefore, the VFR value can be relaxed (as low as 14 fps) by sending fewer frames per second without lowering the QoE.

Collaborative Dancing

Collaborative dancing involves dancers at geographically distributed sites interacting with

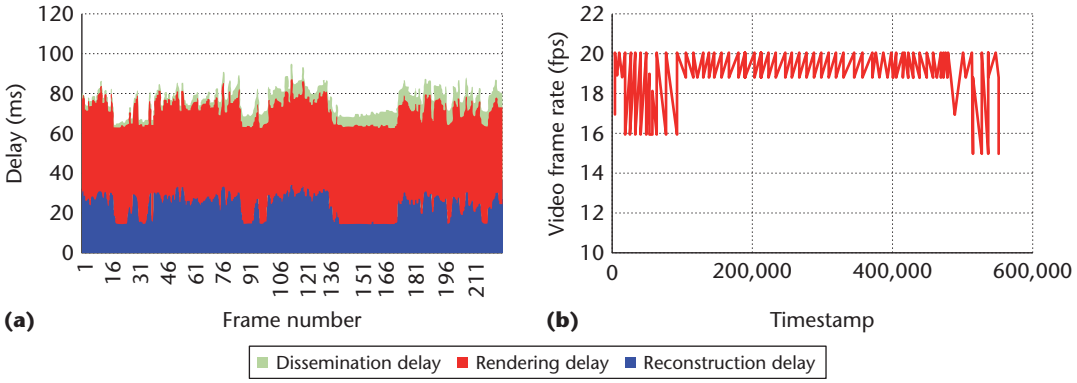


Figure 6. Virtual fencing QoS. (a) End-to-end delay results from network dissemination, rendering, and 3D reconstruction delays. (b) Video frame rate.

each other in the same virtual space.[14] In this experiment,[15] each dancer observed her mirror image and her remote partner(s). Figure 7 gives an example. Furthermore, the collaborative dancing includes another interesting feature, in which we render a prerecorded dancer into the virtual space and enable a live performance between a real dancer and the prerecorded dancer in the same virtual space. This feature shows great promise because it allows for enhanced training, self-assessment, and creativity possibilities.

We used the two Teeve sites at Illinois and UC Berkeley. Each site consisted of a physical dance space of approximately 6 × 6 square feet. Each dance space was surrounded by eight to 12 3D cameras installed in two rows to capture the dancers' upper and lower bodies from different angles. Each camera was configured at a 320 × 240 pixel resolution. We streamed the 3D video streams via a gigabit Ethernet LAN and via Internet2 to local and remote sites and displayed the rendered joint 3D images on 2D plasma and 3D stereoscopic video displays. To facilitate creative choreography, the 3D rendering component also provided digital options to dancers to explore novel virtual choreography design. That is, the system allowed us to

- change the scale, number, spatial placement, and appearance of dancers in the virtual space,
- load prerecorded 3D graphical worlds (such as a stage), and
- load prerecorded TI video (for example, allowing a dancer to dance with herself).

Figure 7. Example collaborative dancing scene. Each dancer observed her mirror image and her remote partner(s) in the virtual space.

We invited two professional dancers to assist us with both controlled and uncontrolled experiments.

The main goal of our controlled experiments was to characterize different performance metrics.[16] We let the dancers perform basic movements with varying movement speeds: slow, moving at a pace similar to Tai Chi; moderate, moving at a natural pace without the need to push for speed or consciously slow down; and fast, moving at a pace that is more driven and pushed beyond the level of comfort, such as playing competitive sports. We asked one dancer to lead and the other to follow in coordinated movements as if they were dancing a duet. We tried six combinations of movement speed between the leader and follower: (slow, slow), (medium, slow), (fast, slow), (medium, medium), (fast, medium), and (fast, fast).

After each of the six controlled experiment sessions, we asked the dancers to fill out questionnaires. We found that the dancers were

*Table 3. Performance QoS profiles for TI activity classes.**

Activity class	Activity QoS profiles											
	AQ	ASR	VQ	MSP	NVS	EED	ID	VFR	ISS	AVSS	NSC	MP
Conversational	H	H	M–H	Yes	L	M	L	L	L	H	L	Yes
Collaborative fine motor	M	M	M	Yes	H	H	H	H	H	M	H	Yes
Collaborative gross motor	L	L	L	Yes	H	H	H	H	H	L	H	Yes

* *L*, *M*, and *H* are the low, medium, and high importance levels for QoS parameters. See Table 1 for definitions of the activity QoS profiles. The MP and MSP columns specify if we need to differentiate the priority among modalities or among streams inside a modality, respectively, for a given activity class.

not satisfied with the visual resolution (320 × 240 pixels), EED (200 ms), and interactivity delay (nearly double the EED). The resolution was unsatisfactory mainly because it did not allow them to make eye contact. Even though 640 × 480 pixel resolution was technically possible with the camera hardware, it would have lowered the VFR. We also found that the subjective perception of intersite synchronization (ISS) and video continuity (VFR) depended on the movement speed. The average frame EED in the experiments was 200 ms with a standard deviation of 47 ms. When the dancers moved with slow-to-moderate physical motion, they were satisfied with the video's synchronization (ISS) and continuity (VFR). However, when they moved faster, they started to notice out-of-sync moments and video discontinuity.

The main goal of our uncontrolled experiments was to let the dancers move freely, explore possibilities in creative dancing, and understand the general usefulness of the TI technology for collaborative dancing.[15] We also interviewed the dancers after the uncontrolled experiments. In this case, the dancers actually considered the system's imperfections such as low VFRs, delays, ''ghosting'' (double images due to calibration errors), and jittering useful, creative compositional elements. As one of the dancers commented, ''From the artistic point of view, it has become apparent that retaining some imperfections is a highly desirable option.''[15]

Comparative Analysis and Lessons Learned

Our experiments demonstrate that the importance of QoS metrics varies across different TI activities. Table 3 summarizes our results, where *L*, *M*, and *H* are the low, medium, and high importance levels of QoS parameters that need to be considered for the conversational, fine-motor, and gross-motor TI activity classes. For multimodal priority (MP) and multistream priority (MSP), we specify whether we need to differentiate the priority among modalities or among streams inside a modality, respectively, for a given activity class. In the case of audio and video modalities, we present EED and ID together due to their intrinsic dependency in any activity.

For conversation activities, the audio quality (AQ) is important because of the activity's dominant auditory and conversational nature. Therefore, Perceptual Evaluation of Speech Quality (PESQ) and audio sample rate (ASR) are also important. However, because of no or limited movement during the conversation, motion jerkiness at a reduced VFR is less noticeable than in collaborative fine- or gross-motor skill activities. Another crucial QoS parameter for conversational activity is AVSS. A tight synchronization requires low skew of video ahead of audio and medium-to-high spatial resolution around the lips and face. The talk-spurt durations in the TI conversational activity are generally short, so lip skew at the end of an utterance is more noticeable. However, the EED tolerance level in conversation is medium, compared to gross-motor activities. The presence of other sensory information does not influence pure conversational activities, and the audio is considered more important than video.

For fine-motor activities, ID is arguably the most important metric that affects QoE. The TI archeology and the mobile block fencing activities both suffered from large interactivity delays. The VQ resolution of the whole video is not crucial, but fine-motor activities require a medium resolution for specific parts of the body (for example, hand resolution in archeology activity). EED and VFR are crucial because they impact the notion of telepresence. Incorporating multiple video streams and increasing

the number of other sensory streams (such as touch sensors) would improve system performance in terms of users' technology acceptance, so number of video streams (NVS) and NSC are both important QoS parameters for fine-motor activities. Although the audio-video synchronization skew is not crucial, the value of ISS for video streams is important because a large interstream skew might create inconsistent views (for example, the upper body can shift, compared to the lower body).

For gross-motor activities, the most important metric is the EED. Noticeable system delays interfered with both the virtual fencing and collaborative dancing activities. Without low and bounded EEDs, it is hard to create a synchronized performance. Similar reasoning applies ID. Unlike fine-motor activities, gross-motor activities do not require high VQ, but VFR must be high. The AQ and AVSS are less crucial because participants are closely engaged in visually dominant activities.

Conclusion

In the future, users will perform different activities on the same TI platform and during the same TI session. This will require session management to detect an ongoing TI activity and select the appropriate performance profile in real time. Hence, the future challenge will be to develop open session management architectures that adapt, are programmable in real time, and yield the best possible QoE for an ongoing activity under any given resource constraints. **MM**

In the future, users will perform different activities on the same TI platform and during the same TI session.

Acknowledgments

This research was funded by the US National Science Foundation under grants 0520182, 0549242, 0724464, 0720702, 0834480, 0840323, 0964081, and 1012194. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the views of the NSF.

References

1. P. Agarwal et al., ''Bundle of Streams: Concept and Evaluation in Distributed Interactive Multimedia Environments,'' *Proc. IEEE Int'l Symp. Multimedia* (ISM), IEEE CS, 2010, pp. 25–32.
2. J. Newlove and J. Dalby, *Laban for All,* Nick Hern Books, 2005.
3. W. Wu et al., ''Color-Plus-Depth Level-of-Detail in 3D Tele-immersive Video: A Psychophysical Approach,'' *Proc. 19th ACM Int'l Conf. Multimedia* (MM), ACM, 2011, pp. 13–22.
4. A. Arefin et al., ''Q-Tree: A Multi-attribute Based Query Solution for Tele-immersive Framework,'' *Proc. 29th IEEE Int'l Conf. Distributed Computing Systems* (ICDCS), IEEE CS, 2009, pp. 299–307.
5. Z. Huang et al., ''Towards the Understanding of Human Perceptual Quality In Tele-immersive Shared Activity,'' *Proc. ACM Multimedia Systems Conf.* (MMSys), ACM, 2012, pp. 29–34.
6. R. Vasudevan et al., ''High Quality Visualization for Geographically Distributed 3D Teleimmersive Applications,'' *IEEE Trans Multimedia,* vol. 13, no. 3, 2011, pp. 573–584.
7. R. Steinmetz, ''Human Perception of Jitter and Media Synchronization,'' *IEEE J. Selected Areas in Comm.,* vol. 14, no. 1, 1996, pp. 61–72.
8. M. Forte, G. Kurillo, and T. Matlock, ''Tele-immersive Archaeology: Simulation and Cognitive Impact,'' *Proc. EuroMed,* 2010, pp. 422–431;www.euromed2010.eu/e-proceedings/content/project/422.pdf.
9. G. Kurillo, M. Forte, and R. Bajcsy, ''Teleimmersive 3D Collaborative Environment for Cyberarchaeology,'' *Proc. IEEE CVPR Workshop on Applications of Computer Vision in Archaeology,* IEEE CS, 2010, pp. 23–28.
10. O. Kreylos, ''Environment-Independent VR Development,'' *Proc. 4th Int'l Symp. Advances in Visual Computing* (ISVC), 2008, pp. 901–912.
11. Shi et al., ''A High-Quality Low-Delay Remote Rendering System for 3D Video,'' *Proc. 18th ACM Int'l Conf. Multimedia* (MM), ACM, 2010, pp. 601–610.
12. S. Shi, K. Nahrstedt, and R.H. Campbell, ''Distortion Over Latency: Novel Metric for Measuring Interactive Performance in Remote Rendering Systems,'' *Proc. IEEE Int'l Conf. Multimedia and Expo* (ICME), IEEE CS, 2011, pp. 1–6.
13. W. Wu et al., ''I'm the Jedi! - A Case Study of User Experience in 3D Tele-immersive Gaming,'' *Proc. IEEE Int'l Symp. Multimedia* (ISM), IEEE CS, 2010, pp. 220–227.

14. K. Nahrstedt et al., ''Symbiosis of Tele-immersive Environments with Creative Choreography,'' *Proc. ACM Workshop on Supporting Creative Acts Beyond Dissemination, ACM Creativity and Cognition Conf.*, ACM, 2007.
15. R. Sheppard et al., ''Advancing Interactive Collaborative Mediums Through Tele-immersive Dance (TED): A Symbiotic Creativity and Design Environment for Art and Computer Science,'' *Proc. 16th ACM Int'l Conf. Multimedia* (MM), ACM, 2008, pp. 579–588.
16. Z. Yang et al., ''A Study of Collaborative Dancing in Tele-immersive Environment,'' *Proc. IEEE Int'l Symp. Multimedia* (ISM), IEEE CS, 2006, pp. 177–184.

Ahsan Arefin is a doctoral student in computer science at the University of Illinois at Urbana–Champaign. His research interests include multimedia systems, multimedia streaming, application and network quality of service management, and network measurement. Arefin has a BS in computer science and engineering from the Bangladesh University of Engineering and Technology. Contact him at marefin2@illinois.edu.

Zixia Huang performed this research while at the University of Illinois at Urbana–Champaign. His is now a software engineer in the Advanced Technology R&D Group at Google Fiber. His research interests include multimedia systems, computer networking, and distributed computing. Huang has a PhD in computer science from the University of Illinois at Urbana–Champaign. Contact him at zhuang21@illinois.edu.

Raoul Rivas is a doctoral student in computer science at the University of Illinois at Urbana–Champaign. His research interests include operating systems and multimedia systems including virtualization techniques, quality of service, soft real-time scheduling, power management, and heterogeneous architectures. Rivas has a BS in computer science from the University of Illinois at Urbana-Champaign. Contact him at trivas@illinois.edu.

Shu Shi performed this research while at the University of Illinois at Urbana–Champaign. He is now a research scientist at Ricoh Innovations. His research interests include multimedia systems, multimedia streaming, and mobile computing. Shi has a PhD in computer science from University of Illinois at Urbana–Champaign. Contact him at shushi@rii.ricoh.com.

Pengye Xia is a doctoral student in computer science at the University of Illinois at Urbana–Champaign. His research interests include multimedia systems, QoS, computer networking, and 3D video compression and transmission. Xia has an MS in computer science and engineering from the Hong Kong University of Science and Technology. Contact him at pxia3@illinois.edu.

Klara Nahrstedt is the Ralph and Catherine Fisher Professor in the Computer Science Department at the University of Illinois at Urbana–Champaign. Her research interests include teleimmersive and mobile systems. Nahrstedt has a PhD in computer and information science from the University of Pennsylvania. She is an IEEE Fellow, Humboldt Fellow, and ACM SIGMM Chair and received the IEEE Technical Achievement Award. Contact her at klara@illinois.edu.

Wanmin Wu is a postdoctoral researcher at the University of California, San Diego. Her research interests include multimedia systems, augmented reality, and mobile computing. Wu has a PhD in computer science from the University of Illinois at Urbana–Champaign. She is the recipient of the Best Student Paper Award at ACM Multimedia 2011 and the SIGMM Best PhD Thesis Award. Contact her at wwu@ucsd.edu.

Gregorij Kurillo is a research engineer on teleimmersion project at the University of California, Berkeley. His research interests include camera calibration, stereovision, robotics, and collaborative VR. Kurillo has a PhD from the School of Electrical Engineering at the University of Ljubljana, Slovenia. Contact him at gregorij@eecs.berkeley.edu.

Ruzena Bajcsy is a professor in the Electrical Engineering and Computer Science Department at the University of California, Berkeley. Her research interests include teleimmersion, computer vision, AI, robotics, and sensor networks. Bajcsy has a PhD in computer science from Stanford University. She is an ACM Fellow, AAAI Fellow, and IEEE Fellow and a member of both the National Academy of Engineering and the Institute of Medicine. Contact her at bajcsy@eecs.berkeley.edu.