Open access • Journal Article • DOI:10.1109/MMUL.2012.59

Classification and Analysis of 3D Teleimmersive Activities — Source link

Ahsan Arefin, Zixia Huang, Raoul Rivas, Shu Shi ...+5 more authors

Institutions: University of Illinois at Urbana–Champaign, University of California, San Diego, University of California, Berkeley

Published on: 01 Jan 2013 - IEEE MultiMedia (IEEE)

Topics: Quality of experience and Virtual reality

Classification and Analysis of 3D Tele-immersive Activities

Ahsan Arefin, Zixia Huang, Raoul Rivas, Shu Shi, Pengye Xia, Klara Nahrstedt
University of Illinois at Urbana-Champaign

Wanmin Wu
University of California, San Diego

Gregorij Kurillo, Ruzena Bajcsy
University of California, Berkeley

Abstract

Current interactive tele-presence systems are designed and optimized for one particular type of cyber-physical activity such as conversation, video chat, or gaming. However, with the emerging new 3D tele-immersive (TI) systems, such as our own TI system, called TEEVE (TEle-immersion for EVErybody), we observe that the same TI system platform is being used for very different activities. In this paper, we classify the TI activities with respect to their physical characteristics, qualitatively analyze the cyber side of TI activities, and argue that one needs to consider very different performance profiles of the same TI system platform in order to achieve high quality of experience (QoE) for different cyber-physical TI activities.

1. Introduction

With the increased high-speed wired and wireless network availability, current interactive tele-presence systems such as Skype, Google mobile video chat, and Cisco tele-presence are becoming an integral part of our lives. These systems are optimized for one particular activity (mostly conversation) to achieve the optimal performance for the intended activity. However, in the recent years, new 3D tele-immersive (TI) systems have emerged as shown in Figure 1, where geographically distributed participants can engage in different shared *cyber-physical TI activities* (e.g., conversation and collaborative dancing) with diverse *physical characteristics* and *cyber-performance profiles* using the same TI system platform.

(a) (b)

Figure 1. *Immersed scene of two remote participants*

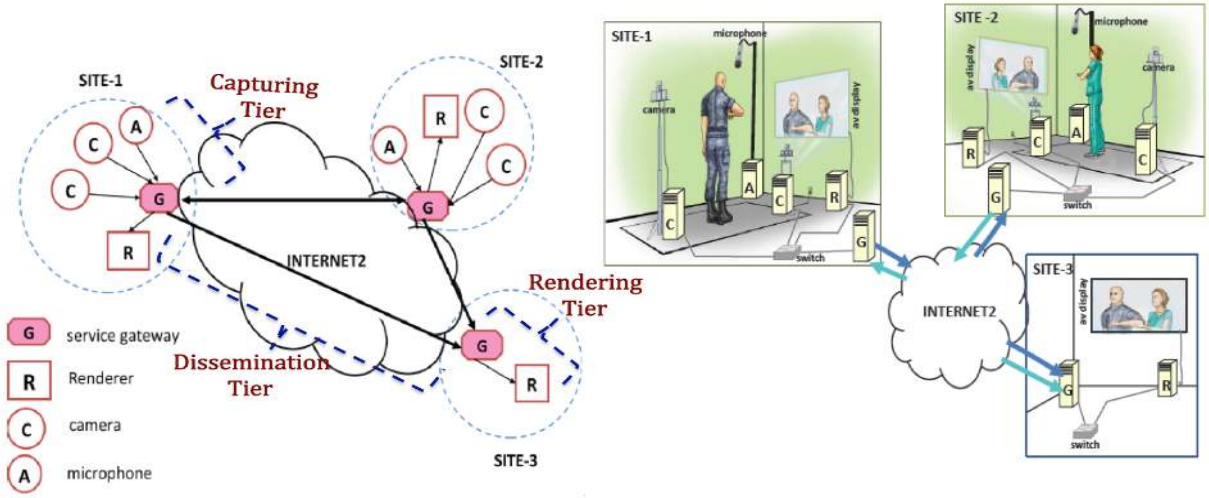
We have built a TI system with multiple sites across University of Illinois at Urbana-Champaign and University of California, Berkeley, called **TEEVE (TEle-immersion for EVErybody)**. Using TEEVE, we have explored a wide variety of interactive activities ranging from conversational to collaborative fine

and gross motor activities. For each new activity, we have *qualitatively* measured observable and controllable **Quality of Service (QoS)** parameters that influence the final user **Quality of Experience (QoE)**.

Based on our analysis of TEEVE activities and their corresponding QoS parameters and QoE responses, we present a cyber-physical **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 the run-time, we require highly configurable, programmable and adaptable TI system platforms.

2. TEEVE Tele-immersive System Platform

The TEEVE system is a multi-party 3D tele-immersive system platform connecting multiple remote sites into one virtual shared space as shown in Figure 2. The logical system architecture of TEEVE, shown in Figure 2(a), is composed of three architectural tiers: the **capturing tier**, the **data dissemination tier** and the **rendering tier**



(a) (b)

Figure 2. *TEEVE system*: *(a) logical architecture, and (b) physical setup for tele-health TI activity*

In the **capturing tier**, multiple capturing devices such as cameras (C), microphones (A) and other sensors (not shown in Figure 2(a)) capture the cyber-physical multi-modal information of each participant at his/her physical site. The captured multi-modal data is synchronized and tagged with their temporal and spatial correlations, creating a **bundle of audio, video,** and **sensory streams** [Agarwal2010]. Such cyber-physical spatio-temporal correlated streams are called a *bundle of streams*.

The **data dissemination tier** consists of a peer-to-peer overlay network of gateways (G) 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 rendering devices (R) render different views of an ongoing activity and provide sensory feedbacks to users. Depending on the participants' views and their perceptions in the virtual space, the priority of the streams inside the bundle (called *importance of modality*) can differ among the viewers. Before the final rendering, streams in each bundle must be re-synchronized

[Huang2011] due to various Internet dynamics (e.g., jitter and loss). An example of a TEEVE physical setup for the tele-health activity is shown in Figure 2(b).

3. Methodology

Our goal is to correlate performance profiles of a TI system to its cyber-physical TI activities. However, to achieve this goal in an efficient manner, first we 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:

- We analyze **physical** movement-based characteristics of TI activities and group them into TI activity classes.
- We identify **cyber** QoS parameters to evaluate TI activities in virtual spaces. These parameters will be observed and measured to determine **performance profiles** for each TI activity class.
- Finally, we employ **measurement and evaluation methods** to obtain correlations between QoS and QoE values for each TI activity class.

3.1 Physical Characteristics for TI Activities

Since we are considering cyber-physical activities within the TI platform, we classify TI activities according to *three movement-based characteristics* [Newlove2005]: **space coverage** (the physical space capacity of a participant for his/her individual movements in the physical space), **speed** (the speed of the participant's movements), and **interaction type among participants** in the virtual space.

The *space coverage* depends on the activity movement of each participant. Each participant's movement space can range from being ***small*** (standing still) to ***large*** (jumping). For example, during the exer-gaming activities (gaming activities involving physical exercises), participants cover the whole (large) allocated physical space. However, during the video conferencing, participants cover a very limited (small) physical space.

The *speed of movements* can vary from ***slow*** to ***fast***. During the exer-gaming activities, participants move fast with the intention of winning the game, while during the virtual teaching of an engine repair, participants move slowly.

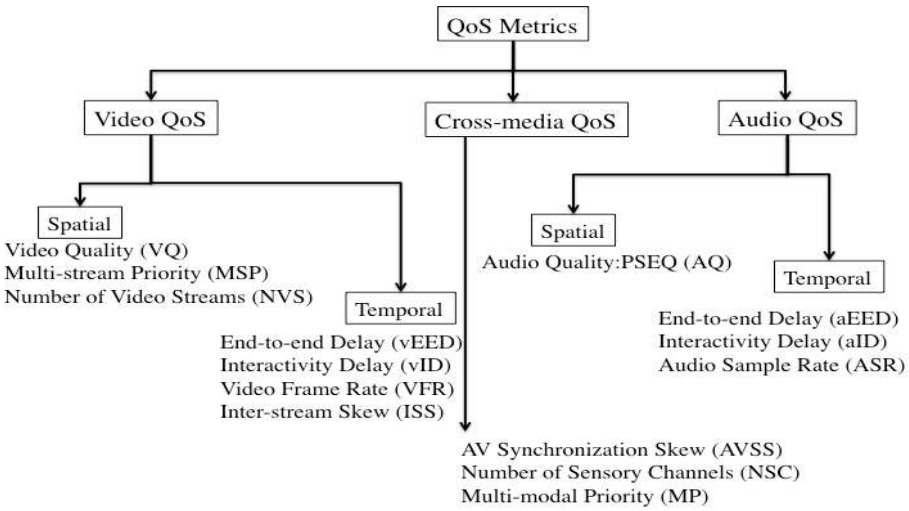


Figure 3. *Categories of activity QoS*

The *interaction type* among remote participants is highly influenced by the intention of the participants. Depending on the muscles (e.g., fine muscles that control finger movements or gross muscles that control leg movements) that participants use, the interactions can be classified as ***fine motor*** and ***gross motor*** skill interactions. The fine motor skill requires a very precise interaction among the participants, such as pointing out a position of an engine. On the other hand, the gross motor skill requires whole body coordination, such as arm and leg movement in a virtual fencing.

Utilizing the above movement-based characteristics, the TI activities can be classified into three classes: *conversational, fine motor collaborative* and *gross motor collaborative* activities. The fine motor collaborative activity is usually defined as the collaborative activity that requires the use of fine muscles from the participants along with different combinations of movement and spatial coverage. The gross motor collaborative activity is defined as the activity that requires the use of gross muscles from the participants along with different combinations of movement and spatial coverage. Examples are given in Section 4.

3.2. Activity Quality of Service

The variation in physical movement characteristics causes the variation of activity QoS parameters, and hence the user's perceived QoE. To organize the activity QoS parameters (aQoS) in our context, we divide them into three major categories: *video QoS, audio QoS* and *cross-media QoS*. Figure 3 shows the hierarchy of QoS categories and our selection of parameters from each category to use in our analyses. Table 1 shows their definitions.

Table 1. *Definitions of activity QoS*

Cyber aQoS Metrics	**Definition**
Video Quality (VQ)	The spatial video frame resolution, measured in number of pixels per frame, bits per pixel, PSNR (Peak Signal-to-Noise Ratio), and Color-to-Depth Level-of-Detail (CZLoD) [Wu2011].
Multi-Stream 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. It is considered 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 to the request. It is considered for both audio (aID) and video (vID). The value of ID may be larger than 2 x EED, since ID considers also the time duration for updating and providing user feedback.
Video Frame Rate (VFR)	The application frame rate of a video stream.
Inter-stream skew (ISS)	The skew between streams of similar modality. Currently, we measure ISS among video streams.
AV 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.
Multi-modal Priority (MP)	The importance of a modality, defining which modality is of greater importance than other modalities.

Audio Quality: PESQ (AQ)	The quality of an audio signal as defined by the standard ITU-T P.862.
Audio Sample Rate (ASR)	The application frame rate of an audio stream.

3. 3. Evaluation Methods

ITU (International Telecommunication Union) has prescribed several recommendations for evaluating perceptual quality of video-conferencing systems, which serve as useful guidelines for tele-immersion. But unfortunately, little is understood about the impact of different QoS configurations on different TI activities in terms of QoE. This motivates us to perform our own evaluations. The evaluation methodology involves three steps: (1) **objective evaluation** of QoS, (2) **subjective evaluation** of QoE and (3) **finding 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 have implemented a monitoring service that records the objective QoS values at run time and allows distributed range queries from different TI sites [Arefin2009].

During and/or after activities, subjective evaluation of QoE is performed with human participants. This evaluation utilizes methods that record **self-reported responses** of users, such as their perception of the video quality and their concentration level during activities. In TEEVE, we use questionnaires and comments from participants as subjective methods.

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.

We present two examples of our evaluation methods. In one example, considering the *conversational and fine-motor collaborative activities*, we evaluate and compare activities by showing users different activity videos with diverse QoS configurations. We collect users' perceived QoE in the form of *comparative mean opinion scores* (CMOS) [Huang2012]. Using users' responses, the QoS parameters are correlated to the CMOS values. For example, the correlation mapping between the VFR (r) and the associated CMOS value is represented by an exponential function in Equation (1), where r_{max} is the maximum achievable VFR, and Q and c are constants, highly dependent on the activity types:

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

In another example of *fine-motor collaborative activities*, we measure the perceptual thresholds of the visual quality. We employ the *Ascending Method of Limits* from psychophysics to measure the *Just Noticeable Degradation* and *Just Unacceptable Degradation* thresholds on one of the VQ parameters, the *Color-plus-Depth Level-of-Detail* (CZLoD) spatial resolution parameter. We find that 70% of CZLoD degradation is imperceptible to human eyes [Wu2011].

4. Qualitative Analysis of TI Activities

We perform a qualitative analysis of TI activities shown in Table 2 by running them within the TEEVE testbed. We will follow the methodology as discussed in Section 3 and analyze TI activities with respect to their objective QoS and subjective QoS evaluations to gain an understanding of their possible correlation and interpretation.

Table 2. *Consideration of 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

4.1. Tele-Immersive Conversation

Activity Description. 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 as Figure 1(a) shows. In our TI conversational experiment with TEEVE, users are facing cameras with a limited movement. During the activity, the participants concentrate on each other's faces, lips and body language.

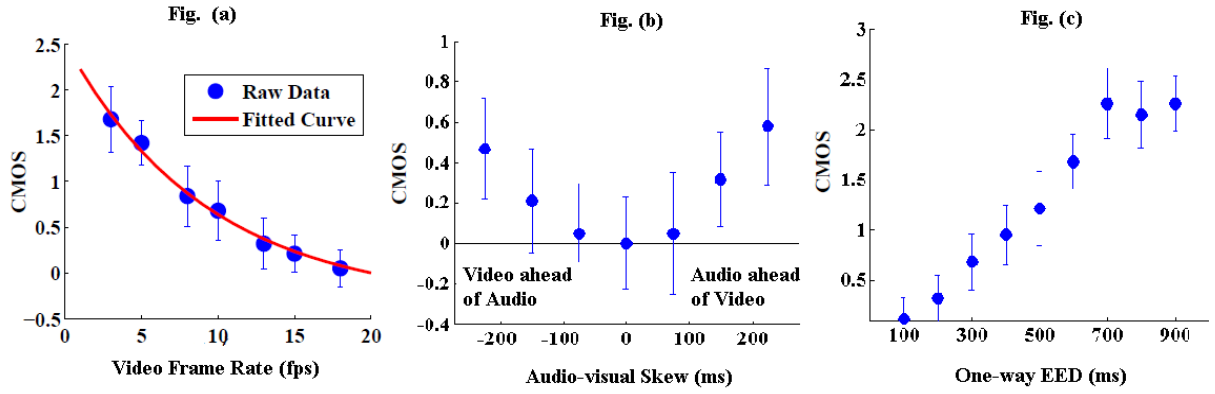


Figure 4. *CMOS results show resulting scores as opposed to optimal reference (PESQ of 4.5, video frame rate of 20 fps, and zero audio-visual skew and interactive latency). (a) Impact of video frame rate deduction, (b) impact of audio-visual skew, (c) impact of increased one-way end-to-end delay*

Activity Experimental Setup. Using the TEEVE system platform, we have enabled conversations among users located in Illinois, California, Texas, and Florida in US and Amsterdam in Europe. Each site was configured with a 61-inch NEC screen, offering 640x480 multi-view video rendering of the immersive environment. To allow high-fidelity speech communication, we equipped users with wireless/wired headsets with a microphone input. The wideband speech signals were encoded with 44 kbps data rate, using the wideband Speex library. A 4-channel microphone array was an add-on capability to capture the ambient sound, which was encoded using Advanced Audio Coding (AAC). A passive stereo [Vasudeven11] was used at the 3D cameras to capture participants' 3D images.

Activity Evaluation. We analyzed the impact of a single objective activity QoS parameter such as VFR, AVSS, and EED on the subjective metric such as CMOS by keeping values of other QoS parameters optimal. In [Huang2012], we showed three findings. (1) The correlation between VFR and CMOS parameters was exponential as shown in Figure 4(a). An exponential model (Equation 1) was used to describe the resulting fitting curve, where Q = 2.52 and c = 2.16. (2) Audio-visual synchronization skew (AVSS) was of great importance. Steinmetz et al. [Steinmetz96] stated that an audio-visual skew of more than 160 ms would be noticed in a video-conferencing system and our subjective study within TEEVE confirmed this finding as shown in Figure 4(b). Our study also confirmed that people were more tolerant of video ahead of audio than audio ahead of video. (3) Users with long EED easily became impatient, and if users did not hear from the remote party within his/her maximum expected latency, a doubletalk occurred. Figure 4(c) shows the correlation mapping between the EED and the corresponding CMOS degradation. It shows that EED larger than 400 ms leads to a very poor interactive perception in the conversation.

4.2 Tele-immersive Archeology

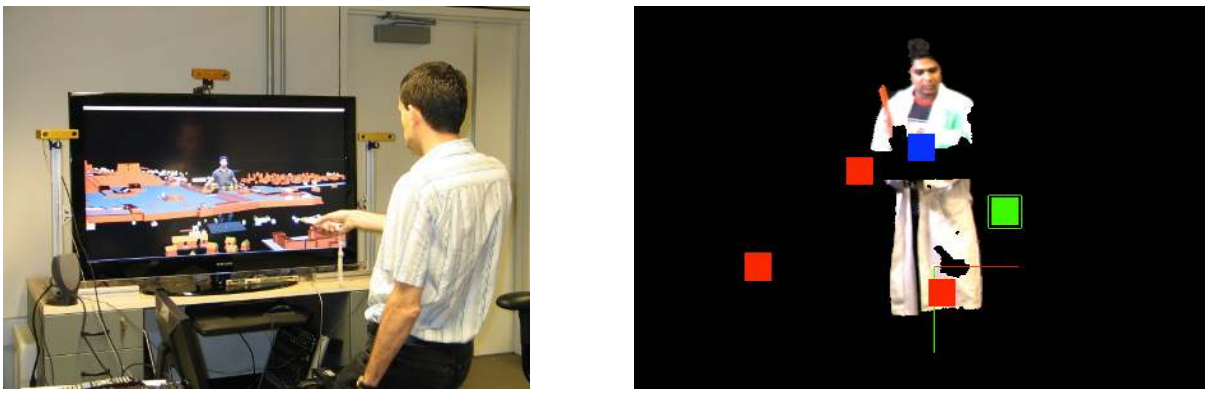
Activity Description. This TI activity is designed to enable multiple archaeologists at different geographical locations to meet in the virtual space, examine, measure and interact with digitized artifacts, explore a virtual site, visualize maps and other data. It uses fine motor skills of hands, gestures, and fingers to interact among participants and with digital artifacts as shown in Figure 5(a). In the figure**,** two users are collaborating in the virtual reconstruction of a Mayan temple. The activity includes the following steps:

(1) Each user observes the model in the first person perspective and is thus able to navigate freely to various locations in the virtual space. For navigation, the users use their 6 DOF (degree-of-freedom) input device to change their position and orientation in the space. So the physical space coverage is small.

(2) When a new artifact object is loaded into the scene, both users are able to see the object appearing in the shared 3D space.

(3) One of the users takes the role of manipulating the artifact object, while the other user guides the first user where the object should be positioned by speaking and/or pointing.

(4) The two users discuss whether the location of the artifact seems appropriate with respect to the model and aim to interpret the meaning of the location.



(a) (b)

Figure 5. *Collaborative TI activities with fine-motor skills (a) TI archeology, and (b) mobile gaming*

Activity Experimental Setup. At UC Berkeley, we experimented with the TI archaeology activity [Forte2010, Kurillo2010]. To facilitate archeological interaction, independent of hardware, we implemented the collaborative architecture, using Vrui VR Toolkit and developed at University of California, Davis [Kreylos2008]. The toolkit with its collaborative extension provided an abstraction of

display and input devices. We performed also remote experiments between UC Berkeley and UC Merced. At both locations, we set up 5 camera clusters (with passive stereo) focused primarily on the frontal part of the users, a 3D display with head tracking, and a six-degree-of-freedom input device (Wii remote with position/orientation tracking) for interaction. Each user was able to observe the exact location of the remote user through his/ her 3D avatar. Since the display, cameras and tracking systems were pre-calibrated, users were able to point and interact with a specific part of the model on the 3D display, and the remote user was able to see his/ her avatar pointing at the same point in space. The video resolution was 320 x 240 pixels to provide a frame rate of 18 - 20 fps.

Activity Evaluation. The experiments showed that being able to see details on users faces was not as important as being able to see the locations of their hands, used for gesturing and pointing. One of the crucial elements of the interaction was the interactivity delay (ID) between the tracking system (which consequently moves the objects) and the 3D video stream (which creates the remote avatar). If there was a significant delay between the two, the user was not able to interpret the location of the user with respect to the objects. Small interactivity delays (under 200 ms) in response of rendering actions were tolerable, while large delays (above 500 ms) caused a disconnection in the interaction between the users. The system was also very sensitive to EED. Users became disappointed if EED was higher. Users were also sensitive to NSC. For example, initially, we have experimented with rendering on a 2D display, and we found that the gestural interaction was difficult, since the users were unable to tell at which part of the 3D model they were pointing at. On the other hand, interaction through the 3D display with head tracking (i.e., improving NSC parameter in Table 1) allowed users to naturally interact with the digital models and helped them interpret their dimensions and geometry.

4.3. Virtual Mobile Block Fencing

Activity Description. The virtual mobile gaming is an activity that enables users to watch the TI shared space on their mobile phones and interact with remote participants in the virtual world. One such example is the *"block fencing"* game as shown in Figure 5(b). The activity includes two different actions as follows:

(1) The mobile user observes the game on the 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 at the new viewpoint. Since, the mobile phone does not have enough 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.

(2a) Mobile user interacts with a remote user by adding virtual blocks via 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 phone, the block position should be exactly where the user touched the screen.

(2b) Remote TI user observes the added virtual blocks on the big TI display. The user can touch the block by moving hands close to the block in the virtual world. Once the hand overlaps with the virtual block on the TI display, the block is considered as touched and will be removed soon after.

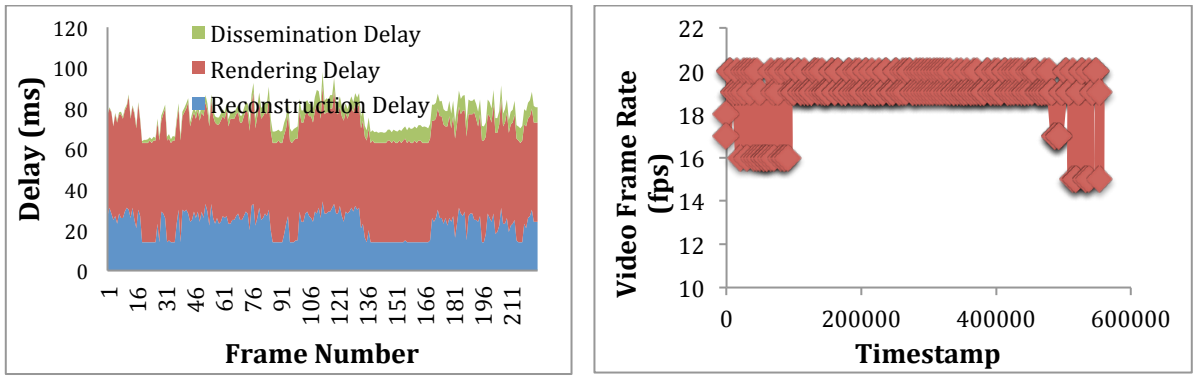
Activity Experimental Setup. One TEEVE site and one iPhone were used in our virtual mobile block fencing activity. The TI system was connected to the Internet2 through Ethernet LAN (Local Area Network), while the iPhone was wirelessly connected through Wi-Fi network.

Activity Evaluation. We focused on the interactive delay (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 iPhone and server were on the same LAN, but ID became unacceptable if the iPhone and rendering server were remotely connected through cellular networks (e.g., 3G). Large ID (greater than 100 ms) impaired the user QoE at

different levels. To reduce ID, [Shi2010] enabled the mobile user to display the rendered avatar image at the new viewpoint before the server updates arrived, using warping and prediction techniques. This approach reduced ID significantly. However, QoE was also influenced by the visual quality (VQ) of the image. Therefore, for such systems, both ID and VQ must be considered [Shi2011].

4.4 Virtual Fencing

Activity Description. In TEEVE, we have implemented the virtual fencing activity with two players in geographically distributed locations. The two players use physical swords to engage in a virtual duel. The use of 3D video improves the overall QoE [Wu2010], since players not only play the video games but also become part of them. To engage in virtual fencing, the following steps are taken: (1) Participant-1 in site-1 puts on a lab coat with red patches and takes a green light-saber, while the participant-2 in site-2 puts on a coat with green patches and holds a red light-saber. (2) Each participant uses the light-saber to hit the opponent's color patches in the virtual space as much and as fast as possible in order to gain points. (3) When a hit occurs, the sword and the coat patches in the virtual space turn blue and the participant, making the hit, gets points. Also the participant, who gets hit, feels a haptic feedback through vibration and lightning of his sword. (4) The participants are able to move and sometimes go out of the space to avoid the hit.



(a) (b)

Figure 6. *Virtual Fencing QoS: (a) end-to-end delay composed of network dissemination delay, rendering delay and 3D reconstruction delay, (b) video frame rate*

Activity Experimental Setup. Our TEEVE experiment used a two-site TI system over a Gigabit LAN, where each site was equipped with one 3D camera, one microphone, one speaker and one rendering display.

Activity Evaluation. As shown in [Wu2010], interactivity delay (ID) were severely impacted when EED became higher than 100 ms, which led to cases where players failed to perform the hitting. Adults and children experienced problems with the consistency of the system in the presence of such a noticeable latency. Increased EED led to confusion and decreased user concentration, causing some users to stop performing the activity. The EED values (composed of network dissemination delay, rendering delay and 3D reconstruction delay) were bounded by 90 ms as shown in Figure 6(a), and the most contributing factor was the rendering delays at the output devices. Due to the nature of gross motor activities, VQ was only important in localized areas of the screen. For example, the VQ was only important in the area around the swords. Facial features were not important because the main perceptual focus of the participants was on the swords. We found that users could successfully perform the activity with a video frame resolution of 320x240 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 to 20 fps as shown in Figure

6(b). Therefore, the value of VFR can be relaxed (as low as 14 fps) by sending less number of frames per second without lowering the QoE.

4.5 Collaborative Dancing

Activity Description. The collaborative dancing involves dancers at geographically distributed sites, interacting with each other in the same virtual space [Nahrstedt2007]. In the experiments done between Berkeley and Illinois [Sheppard2008], each dancer observes her mirrored image and her remote partner(s). An example is shown in Figure 1(b). Furthermore, the collaborative dancing also includes another interesting feature where we render a pre-recorded dancer into the virtual space, and hence enable a live performance between a real dancer and a stored, pre-recorded dancer in the same virtual space. This feature shows a great promise since it allows for enhanced training, self-assessment and creativity possibilities.

Activity Experimental Setup. We used two TEEVE sites, at Illinois and UC Berkeley. Each site consisted of a dancing physical space of about 6x6 square feet. Each dancing space was surrounded by 8-12 3D cameras that were installed in two rows to capture upper and lower bodies of dancers from different angles. The resolution of each camera was configured to 320x240 pixels. The 3D video streams were streamed via Gigabit Ethernet LAN and via Internet2 to local and remote sites, and the rendered joint 3D images were displayed 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. These included (a) abilities to change the scale, number, spatial placement and appearance of dancers in the virtual space, (b) loading of pre-recorded 3D graphical worlds (e.g., theatre stage), and (c) loading of pre-recorded tele-immersive video (dancing with yourself). Two professional dancers were invited for the experiments. We have conducted *controlled* and *uncontrolled* experiments with dancers.

- *Controlled experiments [Yang2006]*: The main goal was to characterize different performance metrics. We have let the dancers make 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, like playing competitive sports. One dancer (leader) was asked to lead, whereas the other (follower) was asked to make coordinated movement as if they were dancing a duet together. Depending on the movement speed between the leader and follower, we tried six combinations: (slow, slow), (medium, slow), (fast, slow), (medium, medium), (fast, medium), and (fast, fast). After each of the six sessions, we asked the dancers to fill up questionnaires.
- *Uncontrolled experiments [Sheppard2008]*: The main goal was to let the dancers move freely, explore possibilities in creative dancing, and understand the general usefulness of the TI technology for collaborative dancing. Interviews were conducted afterwards to collect opinions from the dancers.

Activity Evaluation: In the controlled experiments, we found that dancers were not quite satisfied with the visual resolution (320x240 pixels), EED (200 ms) and interactivity delay (nearly double of EED). The resolution was not satisfactory mainly because it did not allow them to make eye contact. Even though 640x480 pixel resolution was technically possible in the camera hardware, it would lower the video frame rate (VFR). We also found that the subjective perception of inter-site synchronization (ISS) and video continuity (VFR) actually depended on the movement speed. In the experiments, the average EED of frames was measured to be 200 ms with a standard deviation of 47 ms. When the dancers moved with slow to moderate physical motion, they were satisfied with the synchronization (ISS) and continuity (VFR) of the video; when moving at a fast speed, they started to notice out-of-sync moments and video discontinuity. In the uncontrolled experiments, the imperfections of the systems such as low video frame rates, delays, “ghosting” (double images due to calibration errors), and jittering were actually considered as useful, creative compositional elements by the dancers. 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" [Sheppard2008].

5. Comparative Analysis and Lessons Learned

From the above experiments, we conclude that the importance of QoS metrics across different TI activities varies. Our results are summarized in Table 3. The values ***L***, ***M*** and ***H*** mean the low, medium and high levels of importance of QoS parameters that need to be considered for different TI activity classes (conversational, fine motor and gross motor activities). For MP and MSP, we specify if we need to differentiate the priority among modalities or among streams inside a modality, respectively, for a given activity class (***Y***= yes and ***N***=No). In case of audio and video modalities, we present EED and ID together due to their intrinsic dependency in any activity.

For *conversation activities*, the **AQ** is very important due to the dominant auditory and conversational nature of the activity. Therefore, the importance of PESQ and ASR are high. However, due to no or limited movement during the conversation, motion jerkiness at a reduced VFR is less noticeable compared to 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 lips/face. The reason is that 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, however; the audio is considered more important than video.

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

Activity Class/QoS Profiles	AQ	ASR	VQ	MSP	NVS	EED	ID	VFR	ISS	AVSS	NSC	MP
Conversational	H	H	M-H	Y	L	M	L	L	L	H	L	Y
Collaborative Fine Motor	M	M	M	Y	H	H	H	H	H	M	H	Y
Collaborative Gross Motor	L	L	L	Y	H	H	H	H	H	L	H	Y

For *fine motor activities*, **ID** is arguably the most important metric that affects QoE of users. The TI archeology and the mobile block fencing activity both suffer in case of 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 (e.g., hand resolution in archeology activity). The consideration of **EED** and **VFR** is very crucial since it impacts the notion of tele-presence. Incorporating multiple video streams and increasing the number of other sensory streams (e.g., touch sensors) improve the system performance in terms of users' technology acceptance. Therefore, **NVS** and **NSC** both are important QoS parameters for fine motor activities. Though the audio-video synchronization skew is not crucial, the value of **ISS** for video streams is very important since a large inter-stream skew may create inconsistent views (e.g., upper body can shift, compared to the lower body).

For *gross motor activities*, the most important metric is the **EED**. No one wants to lose in virtual fencing due to the presence of noticeable delays in the system. This is also true for collaborative dancing. Without low and bounded EEDs, it is very hard to create a synchronized performance. Similar reasoning is applicable for **ID**. Unlike fine motor activities, gross-motor activities do not require high VQ. However, **VFR** must be high. The AQ is less crucial and so is the AVSS, since participants are closely engaged in visually dominant activities.

6. Conclusion

Our qualitative analysis of TI activities showed 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. We argued for customized performance QoS profiles for each activity class as shown in Table 3.

In the future, users will perform different activities not only on the same TI platform, but also during the same TI session. This is challenging since it will require a session management that (a) detects the ongoing TI activity, and (b) selects the appropriate performance profile in real-time. Hence, the future challenge will be to develop *open session management* architectures that will adapt, be programmable in real-time, and yield best possible QoE for an ongoing activity under given resource constraints.

Acknowledgement. This research was funded by the National Science Foundation (NSF) 0520182, 0549242, 0724464, 0720702, 0834480, 0840323, 0964081, 1012194. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[Agarwal2010] P. Agarwal, R. Rivas, W. Wu, K. Nahrstedt, A. Arefin, Bundle of streams: concept and evaluation in distributed interactive multimedia environments, IEEE ISM, 2010.

[Arefin2009] A. Arefin, Md Yusuf Sarwar Uddin, I. Gupta, K. Nahrstedt, Q-Tree: A multi-attribute based query solution for tele-immersive framework, IEEE ICDCS, 2009.

[Forte2010] M. Forte, G. Kurillo, T. Matlock, Tele-immersive archaeology: simulation and cognitive impact, EuroMed 2010.

[Huang2012] Z. Huang, A. Arefin, P. Agarwal, K. Nahrstedt, W. Wu, Towards the understanding of human perceptual quality in tele-immersive shared activity, ACM MMSys, 2012.

[Kreylos2008] O. Kreylos, Environment-independent VR development, ISVC, 2008.

[Kurillo2010] G. Kurillo, M. Forte, R. Bajcsy, Tele-immersive 3D collaborative environment for cyberarchaeology", IEEE CVPR workshop on Applications of Computer Vision in Archaeology, 2010.

[Nahrstedt2007] K. Nahrstedt, R. Bajcsy, L. Wymore, G. Kurillo, K. Mezur, R. Sheppard, Z. Yang, W. Wu, Symbiosis of tele-immersive environments with creative choreography, Workshop on Supporting Creative Acts Beyond Dissemination, ACM Creativity and Cognition Conference, 2007.

[Newlove2005] Newlove, J. & Dalby, J. Laban for All, Nick Hern Books, London. ISBN 978-1-85459-725-0, 2005.

[Sheppard2008] R. Sheppard, M. Kamali, R. Rivas, M. Tamai, Z. Yang, W. Wu, K. Nahrstedt, Advancing interactive collaborative mediums through tele-immersive dance (TED): a symbiotic creativity and design environment for art and computer Science, ACM MM, 2008.

[Shi2010] S. Shi, M. Kamali, K. Nahrstedt, J. C. Hart, R. H. Campbell, A high-quality low-delay remote rendering system for 3D video, ACM MM, 2010.

[Shi2011] S. Shi, K. Nahrstedt, R. H. Campbell, Distortion over latency: novel metric for measuring interactive performance in remote rendering systems, IEEE ICME, 2011.

[Steinmetz96] R. Steinmetz, Human perception of jitter and media synchronization, IEEE JSAC 1996.

[Vasudevan11] R. Vasudevan, G. Kurillo, E. Lobaton, T. Bernardin, O. Kreylos, R. Bajcsy, K. Nahrstedt, High Quality Visualization for Geographically Distributed 3D Teleimmersive Applications IEEE Transactions on Multimedia, 2011.

[Wu2010] W. Wu, A. Arefin, Z. Huang, P. Agarwal, S. Shi, R. Rivas, K. Nahrstedt, I'm the Jedi! - A Case study of user experience in 3D tele-immersive gaming, IEEE ISM, 2010.

[Wu2011] W. Wu, A. Arefin, G. Kurillo, P. Agarwal, K. Nahrstedt, R. Bajcsy, Color-plus-depth level-of-detail in 3D tele-immersive video: a psychophysical approach, ACM Multimedia 2011.

[Yang2006] Z. Yang, B. Yu, W. Wu, K. Nahrstedt, R. Diankov, R. Bajcsy, A study of collaborative dancing in tele-Immersive environment, IEEE ISM 2006.