Modeling QoE for Delay in 3D Tele-Immersion

1st Author Name
Affiliation
City, Country
e-mail address

2nd Author Name
Affiliation
City, Country
e-mail address

3rd Author Name
Affiliation
City, Country
e-mail address

ABSTRACT

Author Keywords
Telepresence, Delay, Network Performance, QOE.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; See http://acm.org/about/class/1998 for the full list of ACM classifiers. This section is required.

INTRODUCTION

Communications technology plays an important role in human development. The invention of telephone made most remote communications instantaneous. From than on, more and more physical meetings were replaced by phone calls, which saves a great deal of time and money. Nowadays, telepresence is becoming popular. It is the experience of presence in an environment by means of a communication medium [44]. For example, video-mediated telecommunication is providing convenience for teleconference [2, 3, 4], tele-collaboration [5, 6], presence remotely [7, 8, 9, 10], and so on.

Beyond that, researchers are also exploring telepresence with higher level of immersion. In the last decades, 3D tele-immersion (3DTI) developed rapidly. Several 3D-reconstruction-based systems were born [11, 12, 13, 14, 15]. They aim at making up for the lack of eye contact, body language and physical presence in video-mediated telecommunications. Microsoft Research' s Holoportation [15] was quite impressive. They presented an end-to-end 3DTI system with high-quality, real-time reconstructions of an entire space. Because of their promising quality of service (QoS) and the fact that hardware devices are getting cheaper and more powerful, we believe that these systems will become practical in the near future.

However, previous works about 3DTI bias to technical implementations. Only a few studies were conducted. Moreover, they either study on specific scenarios [18, 19, 20] or with pseudo-3D systems [2, 16, 17].

We argue that, fundamental studies on mapping quality of service (QoS) to quality of experience (QoE) in 3D tele-immersion is important. We have seen that the industrial standard of telephone contributes to its popularization, e.g. by avoiding network over-engineering [21]. In recent researches, the user experience (UX) studies of video-mediated telecommunications [2, 3, 4, 9, 17] are also helping its improvement. Similarly, an understanding of UX in 3DTI may well be helpful to both academic and industrial community.

In this paper, we focus on modeling the impact of delay, which is an important factor of QoS [5], in 3D tele-immersion. We first summarize suitable tasks from previous work. Then, we conducted a large online questionnaire (N=100) to introduce our systems, look for more candidate tasks and gather participants' expectation. Last, we selected typical applications for our user studies.

In implementation, we do not follow the highest quality technique [22, 1] (2016) proposed by Microsoft Research, but achieve a more responsive system. Our kernel is similar to Maimone et al. 's work [14] (2012). Supported by the recent progress of depth camera (RealSense-D435), GPU (Gtx1080 Ti) and VR device (HTC Vive), our frame rate reaches 40 FPS. Only one frame delay is necessary for transmission, so the delay is within 50ms. As several related works mentioned the importance of "shared objects" in 3DTI [19, ?], our system was designed to go around the common objects in both sides. Besides face-to-face telecommunications, our system provides an interactive process for non-professional users to easily set up objects-shared activities such as playing chess, piano duet and pair programming.

We have three main findings: first, some tasks with strong interaction, e.g. the finger-guessing game or piano duet, require low latency of 75ms. It breaks the "rule" in 2D telecommunication that 150ms is acceptable for most applications [5, 23]; second, participants' expected latency of tasks based on comparison can well predict the actual needs; third, we argue that the latency requirement of a task depends on its "bottleneck". For example, the bottleneck of most video-mediated telecommunications is *audio signals* [], which leads to an acceptable delay down to 150ms. A stronger bottleneck appears in our system as *synchronous gesture*, e.g. the gesture in the finger-guessing game. It requires a latency of 75ms.

Each submission will be assigned a DOI string to be included here.

RELATED WORK

3D Tele-Immersion

For the external validity of our fundamental QoE study, we had better implement a typical tele-immersion system. We conducted a review of 3DTI technologies in details. Basically, a 3DTI system requires three processes: reconstruction, transmission and rendering [24]. Finally, we developed our reconstruction algorithm based on TSDF Volume [25] and Marching Cubes [26]. We use network line between computers for high-bandwidth transmission, but do not focus on the transmission part as [27, 19] did. In the studies, we simulated various network performance through software methods. We use head-mounted display (HTC Vive) and Unity3D engine to render 3D scenes. Below are reviews of reconstruction and rendering technologies for 3DTI systems:

3D Reconstruction

In early works, researchers used an array of cameras to capture the dynamic scenes [28, 29]. For a given camera view, these systems create a polygonal model that will look correct. That is, they do not construct stand-alone 3D model from physical world.

TELEPORT [30] can composites video-textured surfaces within 3D geometric models. The only one camera limits its construction quality. In 2002 and 2003, researchers started to design immersive 3D video acquisition and rendering environment with multiple cameras [31, 32]. However, their 3D reconstruction output was only point cloud but not polygon mesh. In 2008, Kurillo et al. presented a framework for remote collaboration and training of physical activities [11]. This work tried a reconstruction method with triangulation, but only reached the frame rate of about 5-7 FPS. [12] and [33] for the first time presented compelling real-time reconstruction techniques with multiple cameras. However, the lack of depth dimension indicated their modeling with only silhouette boundaries.

Researchers have made great progress of 3DTI system in the last decade. Both the development of hardware and algorithm made contributions to the real-time performance of high-quality reconstruction. In October 2011, Maimone et al. presented a 3DTI system with Kinects [13]. They developed a pixel-based mesh generation algorithm and reached a frame rate of 30 FPS. This work was followed by Beck el al.'s group-to-group telepresence system [19]. In the same month, however, Microsoft introduced voxel-based [25, 26] system KinectFusion [34] and achieved a better reconstruction quality. Though the volumetric methods were invented about 30 years ago, the emerging depth cameras and GPUs made them practical. In the next year (2012), Maimone et al. also turned to the volumetric methods [14] for improved quality.

In 2016, Microsoft proposed reconstruction pipeline Fusion4D [22], which is highly robust to occlusions, large frame-to-frame motions and topology changes. "The fourth dimension" in this paper was the time dimension, indicating that it leverages the temporally coherence of physical scenes. In the same year, Microsoft integrated fusion4D into their 3DTI system Holoportation [1]. However, Fusion4D is extremely complex and not open-source. Even with expensive devices, Holoportation has an end-to-end latency of 60ms, which can not be ignored in our study. In this paper, we apply a 3D-reconstruction method similar to the one proposed by Maimone et al. [14]. It is a satisfactory system with high quality, responsive interaction and can be easily set up by commercial devices.

3D Rendering

Previous rendering techniques in 3DTI systems can be mainly divided into two categories: spatially immersive displays (SIDs) and head-mounted displays (HMDs). SIDs were earlier applied in 3DTI, while HMDs are becoming popular nowadays. These techniques meet the important need of conveying motion parallax and stereoscopy [?, 43] in telepresence.

Around year 2000, SIDs had become increasing significant [32]. CAVE [35] is a typical SIDs system, which bases on surround-screen projection. Users wear 3D glasses in a CAVE. Most 3DTI systems at that time applied rendering techniques similar to CAVE [30, 31, 32, 11, 18]. CAVE was design for one-to-many presentation. Latter researchers improved it for multi-user telepresence by polarization [36] and time sharing [37]. In 2013, Beck et al. proposed immersive group-to-group telepresence using multi-user SID [19]. There is also a simplified technique called head-tracked auto-stereo display [38, 39], which allows 3D feeling of view without glasses. Some 3DTI system [20, 13, 14] used it for rendering. However, these systems have to abandon the bonus of stereoscopy.

Recently, HMDs develop rapidly in industry. More 3DTI systems tend to apply HMDs for 3D rendering [1, 40, 41, 42]. HMDs are basically cheaper and easier to deploy compared to SIDs. Furthermore, only 3DTI systems with HMDs allow spaces to be shared and co-habited by remote and local users [1]. In 2018, Microsoft proposed Remixed Reality [42]. This approach combines the benefits of augmented reality and virtual reality using 3D reconstruction and VR HMD. Users can not only see their environment, but also apply spatial, appearance, temporal and viewpoint changes on it. For these reasons, we apply head-mounted VR (HTC Vive) to suppose the variety of our study tasks.

QoE of Delay in Telepresence

Quality of Experience (QoE) is defined as: the degree of delight or annoyance of the user of an application or service [46]. It is an integrative theory associated with user experience (UX), which has caused extensive concern in HCI. The bonus of studying QoE is two-fold: first, a QoE conclusion can help avoiding industrial over-engineering, e.g., the standard codec samples audio signals at 8kHz [47] to provide a good trade-off between quality and bandwidth;

second, studies of QoE provide guidelines for follow-up researches. For example, previous work found delay as one of the most crucial factors determining the QoE in telepresence [5, 48, 46, 49], which leads researchers to focus more on delay.

However, few works were conducted to study QoE in 3DTI systems. In 2009, Wu et al. described a user-centric QoE conceptual framework for distributed interactive multimedia environments (DIME) [52]. This framework took 3D Tele-immersion into account. Based on Wu’ s work, Pallot et al. conducted a study on user experience of 3DTI augmented sport [51]. This system was limited both in technical implementation and applications. They drew few conclusions on UX itself, but called for more comparative studies to build an integrative model.

We argue that a series of QoE studies in 3DTI system is required. QoE usually relates to Quality of Service (QoS), including delay, bandwidth, jitter and packet loss [5]. Previous works suggested that delay is one of the most critical QoS metrics in DIMEs [52, 56]. We also found that the impact of delay is mostly reported in 3DTI systems [19, 50, 13, 11, 30]. So in this paper, we tried to model QoE for delay in 3D tele-immersion.

Intuitively, we should take more situations into account in our 3DTI latency study. For audio-mediated telephone, a latency of 150ms is used as a rule of thumb [45, 54]. But in 2D telepresence, the impact of delay become complex. On the one hand, the combination of both audio and video channels makes delay of 80ms ~ 120ms noticeable [52]. This paper suggests that a delay of 120ms may be disruptive or distracting. On the other hand, Tam et al. suggested that delay has a weaker impact on perception of naturalness when both audio and video channels were available, up to 500ms, then when only the audio channel [53]. Furthermore, Schmitt et al. conducted an experiment with a video-mediated quiz task and found that even 500ms is not noticeable [55]. As Pallot el at. suggested [51], user experience related works in DIMEs often have some overlapping aspect and granularity inconsistencies. It may because of the variety of supported tasks in 2D telepresence. Similarly, the conclusion in 3DTI maybe more complex, reflecting more influence factors from physical world but not only the system itself. In this paper, we investigate the influence of delay in various tasks.

SYSTEM OVERVIEW

我们首先提出该系统支持的 applications。然后，我们将简单介绍我们的系统实现。最后，我们给该系统设计了一套交互流程。

Supported Applications

Tell a Lie
典型的 face-to-face telecommunication 任务[,,,,,]，一名用户说三句有关自己的话，其中两真一家，远端用户需要从中猜出撒谎的一句话。从以往论文看来，这个一个典型的，以声音为瓶颈的 task，估计延迟需要在 100~150 毫秒内。

搭积木
典型的 tele-collaboration 任务[,,]。

远程象棋
两端都有一副一模一样的物理象棋。白方棋手的物理世界中只有棋盘和白色棋子，黑方棋手的物理世界中只有棋盘和黑色棋子，我们的系统将合并两个棋盘，这种情况下双方都能在虚拟世界中看见一个完整的棋局。更有趣的是，他们都拥有几乎完美的触觉反馈。

结对编程
两端都一台计算机，但他们不一定需要是一模一样的，因为我们的系统将在虚拟世界中渲染一台两端公有的显示器，供两名用户结对编程。

四手联弹
两端的用户各坐在一个桌子旁，桌子上有一台 iPad，里面正运行着一个 2D 的弹钢琴 App。我们的系统可以 track iPad 的位置，将它渲染成 3D 钢琴上的一段键盘，并将身处两端的用户联结起来，实现四手联弹。

Implementation

这一部分说清楚就好，它本身是没有贡献的。可以强调一下，我们这个系统中的所有部件都是很容易买到的商用硬件。

Interaction

这一部分表明，我们的系统对存在 shared object 的应用场景做了很好的支持：用户可以自己 setup 具体的 application。主要涉及两端公用物体（如象棋），或者虚实物体（如 iPad 和 3D 虚拟钢琴）之间的校准问题。

STUDIES

在 lab study 之前，我们组织了大规模的 online questionnaire。问卷简单描述了我们的系统，介绍了我们系统已经支持的五个 applications，问卷要求参与者分别回答他们对这些任务所须网络质量的期望（即：你猜在这一任务中，你在网络延迟是多少毫秒时能够察觉？在察觉到延迟以后，能够接受多少毫秒以下的延迟？）。为了避免用户对延迟的具体数值缺少概念，问卷中提供了电话延迟至多 150ms，即时战略游戏延迟至多 200ms，这一先验标准。然后，我们请用户 brainstorming，他们认为我们这个系统还能支持哪些实用/有意思的任务。最后，我们还收集了用户的建议，以及让希望参与后续实验的联系方式。

在 lab study 中，由于我们的实验任务很多，耗时很长（大约 40 分钟/每人次），我们决定在校园中采取 volunteer sampling 的方式招募用户被试。每个用户可以只参加一个实验任务，或者参加多个不同的实验任务。为了平衡 learning effect，每个用户来到实验室以后，都会被随机分配到一个实验任务中，他们不能选

择自己特别想做的实验。最后，我们一共做了 5*16=80 人次的实验。

每一个实验的流程如下：首先，我们向两位用户介绍系统，用户可以针对实验任务，熟悉 5 到 10 分钟。接着，用户开始体验我们的系统，以 5 分钟一个 session，每个 session 中，我们会在后台修改整套系统的端到端延迟，每个 session 以后，用户将填写问卷评价刚刚 5 分钟的用户体验。评价体系中，最重要的两个指标是：1、用户是否感受到了延迟；2、如果有，是否严重影响了用户体验；除此之外，我们还问了若干个用于评价沉浸感、纽带性的小问题。每个用户实验共包含 6 个 session，每个 session 的延迟都是不一样的，顺序也随机。要测试的 6 个不同的延迟数值视任务而定，是根据 author 的经验来设计的，保证了用户体验的两个转折点（是否察觉到延迟；延迟是否影响体验）在测试的范围之内。

REFERENCES

1. Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., ... & Tankovich, V. (2016, October). Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 741-754). ACM.
2. Higuchi, K., Chen, Y., Chou, P. A., Zhang, Z., & Liu, Z. (2015, April). ImmerseBoard: Immersive telepresence experience using a digital whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*(pp. 2383-2392). ACM.
3. Nemiroff, G. (1989). Beyond" talking heads": Towards an empowering pedagogy of women's studies. *Atlantis: Critical Studies in Gender, Culture & Social Justice*, *15*(1).
4. Marlow, J., Van Everdingen, E., & Avrahami, D. (2016, February). Taking Notes or Playing Games?: Understanding Multitasking in Video Communication. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1726-1737). ACM.
5. Donovan, A., Alem, L., Huang, W., Liu, R., & Hedley, M. (2014, September). Understanding How Network Performance Affects User Experience of Remote Guidance. In *CYTED-RITOS International Workshop on Groupware* (pp. 1-12). Springer, Cham.
6. Avellino, I., Fleury, C., & Beaudouin-Lafon, M. (2015, April). Accuracy of deictic gestures to support telepresence on wall-sized displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2393-2396). ACM.
7. Nakanishi, H., Tanaka, K., & Wada, Y. (2014, April). Remote handshaking: touch enhances video-mediated social telepresence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2143-2152). ACM.
8. Misawa, K., & Rekimoto, J. (2015, April). ChameleonMask: Embodied physical and social telepresence using human surrogates. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 401-411). ACM.
9. Rae, I., & Neustaedter, C. (2017, May). Robotic Telepresence at Scale. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 313-324). ACM.
10. Neustaedter, C., Venolia, G., Procyk, J., & Hawkins, D. (2016, February). To Beam or not to Beam: A study of remote telepresence attendance at an academic conference. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 418-431). ACM.
11. Kurillo, G., Bajcsy, R., Nahrsted, K., & Kreylos, O. (2008, March). Immersive 3d environment for remote collaboration and training of physical activities. In *Virtual Reality Conference, 2008. VR'08. IEEE* (pp. 269-270). IEEE.
12. Petit, B., Lesage, J. D., Menier, C., Allard, J., Franco, J. S., Raffin, B., ... & Faure, F. (2010). Multicamera real-time 3d modeling for telepresence and remote collaboration. *International journal of digital multimedia broadcasting*, *2010*.
13. Maimone, A., & Fuchs, H. (2011, October). Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*(pp. 137-146). IEEE.
14. Maimone, A., & Fuchs, H. (2012, October). Real-time volumetric 3D capture of room-sized scenes for telepresence. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012* (pp. 1-4). IEEE.
15. Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., ... & Tankovich, V. (2016, October). Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 741-754). ACM.
16. Kuster, C., Ranieri, N., Zimmer, H., Bazin, J. C., Sun, C., Popa, T., & Gross, M. (2012, October). Towards next generation 3D teleconferencing systems. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012* (pp. 1-4). IEEE.
17. Boustila, S., Capobianco, A., & Bechmann, D. (2015, November). Evaluation of factors affecting distance

perception in architectural project review in immersive virtual environments. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology* (pp. 207-216). ACM.

18. Benko, H., Jota, R., & Wilson, A. (2012, May). MirageTable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 199-208). ACM.
19. Beck, S., Kunert, A., Kulik, A., & Froehlich, B. (2013). Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, *19*(4), 616-625.
20. Pejsa, T., Kantor, J., Benko, H., Ofek, E., & Wilson, A. (2016, February). Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing* (pp. 1716-1725). ACM.
21. Bergstra, J. A., & Middelburg, C. A. (2003). ITU-T Recommendation G. 107: The E-Model, a computational model for use in transmission planning.
22. Dou, M., Khamis, S., Degtyarev, Y., Davidson, P., Fanello, S. R., Kowdle, A., ... & Kohli, P. (2016). Fusion4d: Real-time performance capture of challenging scenes. *ACM Transactions on Graphics (TOG)*, *35*(4), 114.
23. ITU-T, I. T. U. T. (2003). Recommendation G. 114. *One-Way Transmission Time, Standard G*, *114*.
24. Fuchs, H., State, A., & Bazin, J. C. (2014). Immersive 3d telepresence. *Computer*, *47*(7), 46-52.
25. Curless, B., & Levoy, M. (1996, August). A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (pp. 303-312). ACM.
26. Lorensen, W. E., & Cline, H. E. (1987, August). Marching cubes: A high resolution 3D surface construction algorithm. In *ACM siggraph computer graphics* (Vol. 21, No. 4, pp. 163-169). ACM.
27. Pece, F., Kautz, J., & Weyrich, T. (2011, September). Adapting standard video codecs for depth streaming. In *Proceedings of the 17th Eurographics conference on Virtual Environments & Third Joint Virtual Reality* (pp. 59-66). Eurographics Association.
28. Fuchs, H., Bishop, G., Arthur, K., McMillan, L., Bajcsy, R., Lee, S., ... & Kanade, T. (1994, September). Virtual space teleconferencing using a sea of cameras. In *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery* (Vol. 26).
29. Kanade, T., Rander, P., & Narayanan, P. J. (1997). Virtualized reality: Constructing virtual worlds from real scenes. *IEEE multimedia*, *4*(1), 34-47.
30. Gibbs, S. J., Arapis, C., & Breiteneder, C. J. (1999). TELEPORT–Towards immersive copresence. *Multimedia Systems*, *7*(3), 214-221.
31. Towles, H., Chen, W. C., Yang, R., Kum, S. U., Kelshikar, H. F. N., Mulligan, J., ... & Holden, L. (2002). 3d tele-collaboration over internet2. In *In: International Workshop on Immersive Telepresence, Juan Les Pins*.
32. Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., ... & Strehlke, K. (2003, July). blue-c: a spatially immersive display and 3D video portal for telepresence. In *ACM Transactions on Graphics (TOG)* (Vol. 22, No. 3, pp. 819-827). ACM.
33. Loop, C., Zhang, C., & Zhang, Z. (2013, July). Real-time high-resolution sparse voxelization with application to image-based modeling. In *Proceedings of the 5th High-Performance Graphics Conference* (pp. 73-79). ACM.
34. Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., ... & Fitzgibbon, A. (2011, October). KinectFusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on* (pp. 127-136). IEEE.
35. Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993, September). Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques* (pp. 135-142). ACM.
36. Fröhlich, B., Hochstrate, J., Hoffmann, J., Klüger, K., Blach, R., Bues, M., & Stefani, O. (2005). Implementing multi-viewer stereo displays.
37. Kulik, A., Kunert, A., Beck, S., Reichel, R., Blach, R., Zink, A., & Froehlich, B. (2011, December). C1x6: a stereoscopic six-user display for co-located collaboration in shared virtual environments. In *ACM Transactions on Graphics (TOG)* (Vol. 30, No. 6, p. 188). ACM.
38. Benko, H., Wilson, A. D., & Zannier, F. (2014, October). Dyadic projected spatial augmented reality. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (pp. 645-655). ACM.
39. Jones, B., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A., ... & Shapira, L. (2014, October). RoomAlive: magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (pp. 637-644). ACM.

40. Smith, H. J., & Neff, M. (2018, April). Communication Behavior in Embodied Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 289). ACM.

41. Maimone, A., Yang, X., Dierk, N., State, A., Dou, M., & Fuchs, H. (2013, March). General-purpose telepresence with head-worn optical see-through displays and projector-based lighting. In *Virtual Reality (VR), 2013 IEEE* (pp. 23-26). IEEE.

42. Lindlbauer, D., & Wilson, A. D. (2018, April). Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 129). ACM.

43. Gotsch, D., Zhang, X., Merritt, T., & Vertegaal, R. (2018, April). TeleHuman2: A Cylindrical Light Field Teleconferencing System for Life-size 3D Human Telepresence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 522). ACM.

44. Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, *42*(4), 73-93.

45. Rec, I. T. U. T. (2003). G. 114,". *One-way transmission time*.

46. Brunnström, K., Beker, S. A., De Moor, K., Dooms, A., Egger, S., Garcia, M. N., ... & Lawlor, B. (2013). Qualinet white paper on definitions of quality of experience.

47. ITU-T, R. G., & Switzerland, I. (1972). 711, Pulse Code Modulation (PCM) of Voice Frequencies. *International Telecommunication Union*.

48. Schmitt, M., Gunkel, S., Cesar, P., & Bulterman, D. (2014, September). Asymmetric delay in video-mediated group discussions. In *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on* (pp. 19-24). IEEE.

49. Schmitt, M., Gunkel, S., Cesar, P., & Hughes, P. (2013, October). A QoE testbed for socially-aware video-mediated group communication. In *Proceedings of the 2nd international workshop on Socially-aware multimedia* (pp. 37-42). ACM.

50. Raghuraman, S., & Prabhakaran, B. (2015, November). Distortion score based pose selection for 3D tele-immersion. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology* (pp. 227-236). ACM.

51. Pallot, M., Eynard, R., Poussard, B., Christmann, O., & Richir, S. (2013, March). Augmented sport: exploring collective user experience. In *Proceedings of the Virtual Reality International Conference: Laval Virtual* (p. 4). ACM.

52. Wu, W., Arefin, A., Rivas, R., Nahrstedt, K., Sheppard, R., & Yang, Z. (2009, October). Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. In *Proceedings of the 17th ACM international conference on Multimedia* (pp. 481-490). ACM.

53. Tam, J., Carter, E., Kiesler, S., & Hodgins, J. (2012, May). Video increases the perception of naturalness during remote interactions with latency. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (pp. 2045-2050). ACM.

54. Percy, A. (1999). Understanding latency in IP telephony. *Brooktrout Technology, Needham, MA*.

55. Schmitt, M., Gunkel, S., Cesar, P., & Bulterman, D. (2014, November). The influence of interactivity patterns on the Quality of Experience in multi-party video-mediated conversations under symmetric delay conditions. In *Proceedings of the 3rd International Workshop on Socially-aware Multimedia* (pp. 13-16). ACM.

56. Vogel, A., Kerherve, B., von Bochmann, G., & Gecsei, J. (1995). Distributed multimedia and QoS: A survey. *IEEE multimedia*, *2*(2), 10-19.

57.