Understanding How Network Performance Affects User Experience of Remote Guidance

Angus Donovan[1], Leila Alem[1], Weidong Huang[2], Ren Liu[1], and Mark Hedley[1]

[1] CSIRO
[2] University of Tasmania

Abstract. Much research has been done to support remote collaboration on physical tasks. However, the focus of the research has been mainly on system and interface design and their impact on collaboration. Relatively less attention has been paid to investigating how network performance can affect user experience and task performance. In this paper, we present a preliminary user study on this issue in which participants were asked to work collaboratively in pair using a remote mobile tele-assistance system we developed. In this study, five network scenarios were examined and network performance (QoS) was measured using four metrics including delay, jitter, bandwidth and packet loss. User experience (QoE) was measured using both objective and subjective metrics. The formal included time taken and number of instructions repeated for task performance while the latter included user ratings of quality of audio experience, quality of video experience and overall quality of experience. The results indicated that the packet loss rate in QoS is the biggest contributor to loss in QoE. We also discuss implications of the study and possible directions of future work.

1 Introduction

Nowadays technologies are becoming increasingly ubiquitous and complex. As a result, expertise is often required for performing physical tasks on these technologies. Physical tasks are ones that require collaborators working on physical objects such as equipment maintenance. However, it is common that users do not have the required skills set and help is required to, for example, fix the technology when it breaks down. This can be potentially a big issue for many users as access to sound expertise and guidance is often lacking, particularly in rural and remote areas and in emergency situations. In response to the demand and in order to make expertise more accessible to users, a number of systems have been developed (e.g., [7, 12, 13]).

These systems typically have two network-connected units: one helper unit and one worker unit, and are constructed in a way that the remote helper is enabled to guide the local worker performing collaborative psychical tasks using both audio and visual communications just like they were co-located. Recently, we have developed a remote mobile tele-assistance tool called ReMoTe [9, 15]. The tool's worker unit is a wearable system that supports mobility of the worker

N. Baloian et al. (Eds.): CRIWG 2014, LNCS 8658, pp. 1–12, 2014.

Fig. 1. Helper Interface

Fig. 2. Worker Interface

and frees his two hands for manipulation of physical objects, while the helper unit is a tabletop system on which the helper can perform hand gestures for guiding purposes. ReMoTe allows two people to collaborate on a "shared visual space" [7] through audio and video communication over an internet connection in real time. The shared visual space is a main feature of the system interface: as depicted in Figure 1, the helper's hand movements are detected and augmented with the work-space of the worker, which is then displayed on a near-field display as visual aid to the worker who is completing the task as shown in Figure 2.

Much research has been done to investigate usability of these systems and their impact on users' collaboration behavior with an assumption of guaranteed network availability (e.g., [10,11,14,16]). However, little attention has been paid to the impact of varying network conditions on user experience and task performance [8,19]. ReMoTe is essentially a network based distributed system, in which the quality of network is a key to its usability. Therefore there is the value of exploring the use of ReMoTe in different network settings. This will provide a basis for us to understand when a drop of QoS (quality of service) may lead to a drop in QoE (quality of experience), which in turn may lead to a wish to drop the real time remote assistance session and reverse to an offline remote guiding session. As part of effort towards this end, we are in the process of conducing a series of studies. In this paper, we present a preliminary investigation of this issue to understand how network performance affects user experience of remote guidance.

2 Background

The specific goal of this particular investigation was to map the relationship of performance between QoS and QoE. To achieve the goal, we tested five network scenarios: Ethernet, Satellite, 3G mobile technology, Wifi and Fibre optic and determined how specific QoS measures within the scenario affected the QoE of an end user.

2.1 Network QoS Metrics

QoS is defined by the ITU (International Telecommunication Union) as the "Totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service" [3]. The QoS for this study is measured in four particular ways specified below. The performance limits for relevant applications are also given.

Delay. Delay is a measure of the interval between the system transmission and when the user receives the information. For typical personal computers or workstations as end-systems, the delay should not exceed 400ms for telephone quality speech. For interactive applications with real time sound transmission the maximum delay is of the order of 0.1 to 0.5 seconds [1,2]. Beyond 100 ms speech is still comprehensible although users may become irritated with the service.

Jitter. Jitter occurs when received signals have varying latency times on arrival. It is an error that occurs when transmitted data packets have variable queuing delays. Jitter requirements are essential to transmitting data at a constant, reliable rate. Real time sound is the most sensitive data transmission type to jitter [6]. As a result, the bounds for acceptable jitter are determined by audio quality. The jitter for videoconferencing applications should not exceed 400ms [1].

Bandwidth. Bandwidth is the average rate of successful message delivery over a communication channel measured in bits/s. The video codec selected for ReMoTe is the H.264 codec for video encoding and the Opus codec for audio encoding. Using the H.264 codec and ReMoTe's standard resolution of 640x480 (4:3), the associated bandwidth that is required is between 300kbps and 800kbps [4]. The Opus audio codec requires an average bandwidth of 25kbit/s [5].

Packet Loss. Packet loss is the amount of packets lost per time slot. This can be in seconds to mirror the bit/s throughput rate. Packet loss can occur because of signal degradation, multipath fading, channel congestion and faulty networks or hardware.

2.2 User QoE Meterics

QoE is defined as "the overall acceptability of an application or service, as perceived subjectively by the end-user [3]. QoE measures are subjective in nature, however they can be measured quantitatively by asking the following questions:

- Time taken to complete the task?
- How many instructions had to be repeated?
- Video quality rating?

– Audio quality rating?
– Overall quality rating?

Examining the time taken is a key indicator of how well the network is performing as it implies how clear the instructions and communications were to both the Helper and the Worker. A metric on the amount of instructions that have to be repeated gives an indication of where the ReMoTe system wasn't functioning in a seamless fashion due to network conditions or otherwise. The video, audio and overall quality ratings are determined using a scale from 1 (bad) to 5 (excellent).

3 Method

3.1 Design

Seven university students were recruited to participate in the study. Four of them were male and the rest are female. One of them was selected to play the role of Helper for all test sessions. The remaining six were asked to play the role of Worker. Both the Helper and the Worker were situated in the same room with both facing away from each other so that they cannot see any of the construction being undertaken on the other side. The room setup is shown in Figure 3. In this figure, the Helper is in the foreground with hands placed over the screen to convey gestures. During testing sessions, the Helper also had a headset (microphone and headphones) linked to the Helper computer that allows audio communication with the Worker. The Worker laptop is narrowly shown on the left side of the figure, in front of the Worker.

Fig. 3. Testing setup

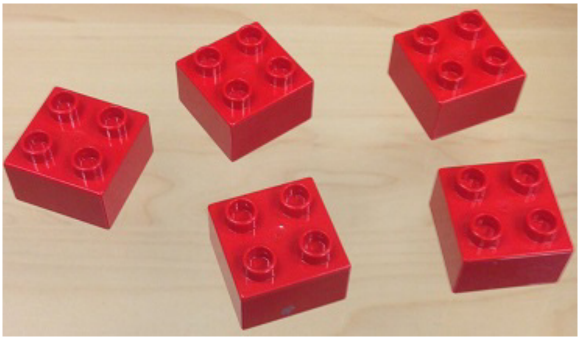

Fig. 4. Loose blocks

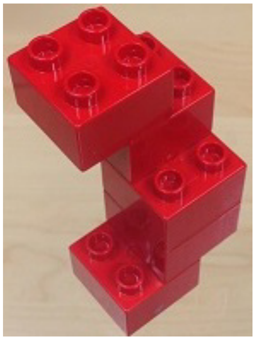

Fig. 5. Assembled blocks

A standard task was created involving large Lego ‘Duplo’ blocks for each of the test subjects. The Helper constructs a random arrangement of five Duplo blocks and keeps this arrangement in front of him. The Duplo blocks used are shown in Figures 4 and 5. The Worker sits at table and has five identical un-arranged Duplo blocks. The Helper must use ReMoTe to aid the Worker in copying the random construction.

Having the same Helper throughout the testing sessions allows consistency in process and instruction. Having blocks of the same color, shape and size increases the reliance of visual communication. Several practice runs were used in the beginning to ensure that both the Helper and the Worker were familiar with each other’s communication styles and the testing in general before the test was formally started.

The specifications of the QoS metrics we used for each of the five network conditions are shown in Table 1. In order to gain a baseline test of human performance with ReMoTe, an optimal network condition (no QoS impairments) was added. This means that there were six network conditions in total and each subject was required to completed the task for each condition. It should be noted that the order of the testing conditions was determined randomly.

Table 1. QoS Metrics for Network Scenarios

	Delay	Jitter	Bandwidth	Packet Loss
Ethernet	10ms	0.2ms	100Mbit/s	0.5%
Satellite	600ms	7ms	1Mbit/s	2%
3G	50ms	≈0	9Mbit/s	2%
Wi-Fi	10ms	30ms	30Mbit/s	5%
Fibre	190ms	40ms	1000Mbit/s	0.5%

3.2 Testing Protocol

The following test protocol was used:

1. Welcome the subject into the testing room and explain how testing will be completed.
2. Allow the Worker to set up the helmet as well as their microphone and speaker head set. Ensure that the Worker is as comfortable as possible before undertaking testing.
3. Complete 3 initial practice tests. The results of these tests are not recorded.
4. Complete the baseline test (no impairments). Ask for video, audio and overall quality ratings.
5. Complete the 5 different network scenario tests. Ask for video, audio and overall quality ratings at the end of each network scenario test.
6. Ask for additional comments and feedback in relation to the overall testing experience.

The vocal instructions used when undertaking testing were fairly consistent. Testing began with the Duplo blocks disassembled in front of the Worker. The set of instructions would usually follow in the style shown below:

– "Start with the Yellow block as the base"
– "Place the small green block on the 4 dots of the right hand side of the yellow block"
– "Place the blue block directly on top of the green block"
– etc.

4 Results and Discussion

Each of the six subjects completed six tasks (five network scenarios plus a baseline condition). In order to easily interpret the data obtained, the experimental data for each condition were aggregated and displayed in column graphs as shown in Figures 6 to 10. Note that the height of each column gives the combined total of the test scores attained by each subject and the score of each subject is differentiated by color.

4.1 Quality of Audio Experience

As can be seen from Figure 6, the audio quality remained largely the same quality throughout the testing. The audio experience is highly important to the QoE as the user becomes exponentially dissatisfied with phone calls or video conferencing when audio transmissions experience poor QoS [17]. There was significant delay of 600ms on the audio in the Satellite scenario however, as communication in the ReMoTe system is largely one-way (Helper instructing the Worker), there is minimal dialogue. It is predominantly in communications involving heavy dialogue (phone calls, videoconferencing etc.) that latency becomes a chief factor

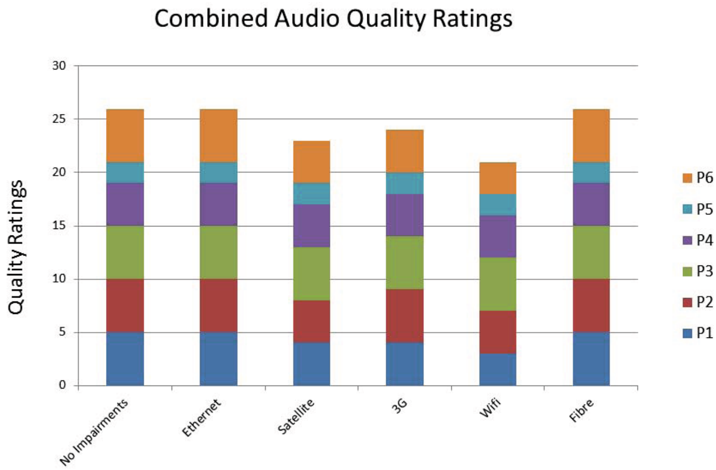

Fig. 6. Combined Audio Quality Ratings

in determining QoE. It is speculated that this is reason why QoE in the Satellite scenario was not at a lower score.

In the Wifi scenario, the high packet loss and noise created occasional 'breaking up' of the audio signal. This made the audio transmission prone to clicks and a small amount of distortion which affected the QoE ratings.

Jitter didn't seem to have any impact on the audio QoE within the experiment environment. This is confirmed by the high jitter in the fibre scenario but no loss of QoE ratings.

Bandwidth didn't seem to have an impact on the audio QoE within the experiment environment. This is supported by [18] which states that "for many applications high data rates do not translate to improved user experience unless the latency is low." However this isn't conclusive as the Satellite scenario had the lowest bandwidth of 1Mbit/s but the decrease in bandwidth may have been 'masked' by the high packet loss rate of 5%. Similarly for the 3G scenario, the comparatively low bandwidth of 9Mbit/s may have been 'masked' by the packet loss rate of 2%.

4.2 Quality of Video Experience

As can be seen from Figure 7, the quality of experience for video diminished with higher packet loss. This is clearly evident in the Satellite, 3G and Wifi scenarios where the packet loss was set to greater than 1%. This impacted video experience as the video feed would intermittently become 'garbled' where nothing was discernible on the screen for a period of approximately half of a second. For a 2% packet loss, this happened approximately every 18 seconds but for Wifi packet loss this happened more frequently at approximately once every 5 seconds. This has large implications for the video streaming in the ReMoTe system. If a hand

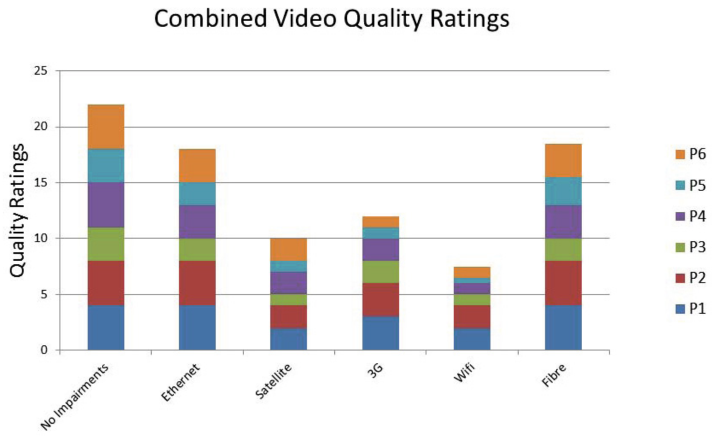

Fig. 7. Combined Video Quality Ratings

gesture is made and the video stream encounters noise at the same time, then the gesture will have to be repeated. It also reduces the QoE for the worker as it can be unclear exactly which information needs to be re-transmitted and the Worker must therefore try to describe what needs to be re-transmitted.

The delay of the Satellite at 600ms was detrimental to the video's quality of experience. When coupled with high noise, this made the ReMoTe system's operation highly unappealing to the user. It didn't however, have a particularly large effect on the ability of the Worker to complete the Duplo task. This indicates that whilst there was a poor quality of video, the audio instructions were the most effective means of communication to complete the task.

With the increased packet loss present in the Wifi scenario, it was also noted that increased pixilation and blurriness occurred in the video image. This has consequences for the quality of experience if ReMoTe is being used in a situation requiring higher precision visual information.

Jitter didn't seem to have any impact on the audio QoE within the experiment environment. This is confirmed by the high jitter in the Fibre scenario but no loss of QoE ratings.

The results suggest that the packet loss rate in a network QoS metrics is the biggest contributor to loss in QoE. Both the 3G and Wifi secnarios have a 2% and 5% packet loss rate and both these scenarios received the worst QoE scores. The Satellite scenario is not a reliable measure of the effect of packet loss on QoE because of it's high delay.

Bandwidth didn't seem to have an impact on the audio QoE within the experiment environment however this isn't conclusive for the reasons discussed in Subsection 4.1. The Satellite scenario had the lowest bandwidth of 1Mbit/s but the decrease in bandwidth may have been 'masked' by the high packet loss rate of 5%. Similarly for the 3G scenario, the comparatively low bandwidth of 9Mbit/s may have been 'masked' by the packet loss rate of 2%.

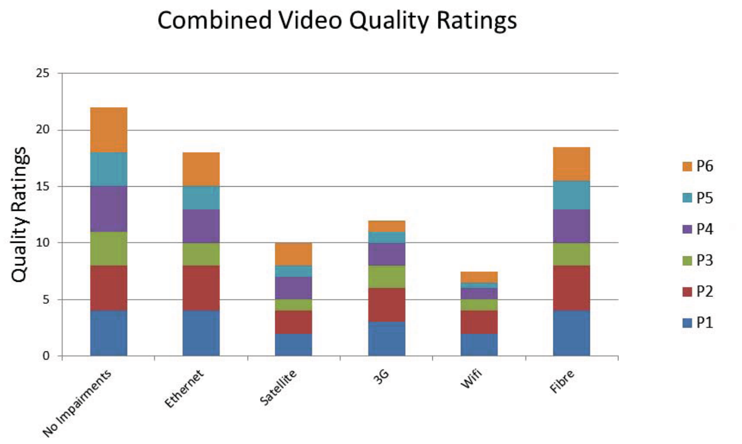

Fig. 8. Combined Overall Quality Ratings

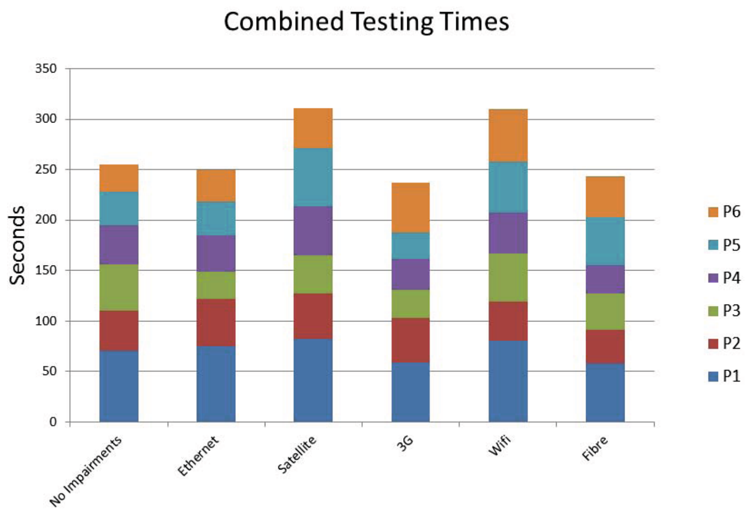

Fig. 9. Combined Testing Times

4.3 Overall Quality of Experience

As can be seen from Figure 8, the overall quality ratings show that 'Wifi' and 'Satellite' had the lowest QoE overall. This is primarily due to the larger differences in ratings that were scored in video quality ratings. The quality of audio experience contributes fairly evenly to the score of the overall experience.

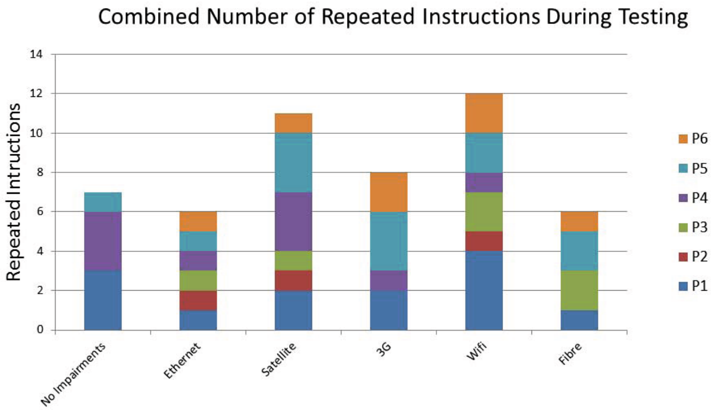

Fig. 10. Combined Number of Repeated Instructions During Testing

4.4 Testing Times

As can be seen from Figure 9, the combined testing times show that Satellite and Wifi network scenarios took the longest to complete by a small margin. The longer amount of time required is indicative of the QoS factors associated with each network scenario. Both Satellite and Wifi had poorer QoS than the other network scenarios. They also had comparatively poorer QoE ratings. The results suggest that a poorer QoS leads to a longer time required to complete tasks.

Further, it seems that the main two QoS factors that effected this timing were Satellite's 600ms delay and Wifi's 5% packet loss. Each had a heavier impact on timing and QoE than other QoS measures. This was confirmed in comments about the QoE by the vast majority of the subjects.

4.5 Repeated Instructions

The graph in Figure 10 shows that the QoS metrics was indicative of the number of mistakes made whilst carrying out testing. By comparison with similar results mentioned in previous sections, the Wifi and Satellite scenarios had the most mistakes as over 10 instructions needed to be repeated for these scenarios.

It can be argued that having to repeat instructions is the sole reason why the Satellite and Wifi network scenarios took a longer amount of time to complete than other scenarios. Therefore, the relationship between QoS and QoE is largely dependent on how many mistakes are made whilst using the ReMoTe system. Of course, additional testing with a high volume of subjects would be required in order to accurately examine this relationship.

Repeating instructions whilst carrying out a task also has implications for QoE as it can be irritating to repeat instructions.

4.6 Summary of Results

In summary, the results suggest that the packet loss rate in a network QoS metrics is the biggest contributor to loss in QoE. This has been determined by both the Wifi and Satellite scenarios which have a 2% and 5% packet loss rate and both these scenarios received the worst QoE scores. Satellite is highly likely to have performed poorly due to additional delay of 600ms. Both Satellite and Wifi also took longer times for task completion.

5 Conclusion

This paper presents a preliminary study investigating the impact of QoS on QoE using ReMoTe. In our study, QoS was measured in terms of delay, jitter, bandwidth and packet loss in five network scenarios including Ethernet, Satelite, 3G, Wifi and Fibre. QoE was measured in terms of time taken, number of instructions repeated, video quality rating, audio quality rating and overall quality rating.

ReMoTe was tested with the focus on gaining subjective results yielding the most useful information about ReMoTe and its QoE based on QoS. The results suggest that Satellite and Wifi network scenarios provide the worst QoE primarily due to packet loss but also due to signal delay.

The work presented in this paper provides the foundation for further study so that in future, the proposed measurements can be used to determine where QoS can be leveraged to provide the best QoE. More realistic testing can be completed which involves more reliance on the visual system of ReMoTe. High volume testing of numerous subjects would assist finding reliable results. These results would aid in optimizing ReMoTe's QoE over various network QoS conditions.

References

1. Fluckiger, F.: Understanding Networked Multimedia. Prentice Hall (1995)
2. Szuprowicz, B.: Multimedia Networking. McGraw-Hill Inc. (1995)
3. International Telecommunications Union, Definitions of terms related to quality of service. ITU Standardizatioin Sector (September 2008)
4. Microsoft, Network Bandwidth Requirements for Media Traffic, Lync TechNet Library (2013)
5. Sarwar, G., et al.: On the Quality of VoIP with DCCP for Satellite Communications. Internationl Journal of Satellite Communications (2000)
6. Chen, Y., Farley, T., Ye, N.: QoS Requirements of Network Applications on the Internet. Department of Industrial Engineering, Arizona State University (2004)
7. Fussell, S., Setlock, L., Yang, J., Ou, J., Mauer, E., Kramer, A.: Gestures over video streams to support remote collaboration on physical tasks. In: Hum.-Comput. Interact., vol. 19(3), pp. 273–309 (September 2004)
8. Gergle, D., Kraut, R., Fussell, S.: The impact of delayed visual feedback on collaborative performance. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006), pp. 1303–1312 (2006)

9. Huang, W., Alem, L.: HandsinAir: a wearable system for remote collaboration on physical tasks. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion (CSCW 2013), pp. 153–156 (2013)
10. Huang, W., Alem, L.: Supporting hand gestures in mobile remote collaboration: a usability evaluation. In: Proceedings of the 25th BCS Conference on Human-Computer Interaction (BCS-HCI 2011), pp. 211–216 (2011)
11. Kuzuoka, H., Yamazaki, K., Yamazaki, A., Kosaka, J., Suga, Y., Heath, C.: Dual ecologies of robot as communication media: thoughts on coordinating orientations and projectability. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2004), pp. 183–190 (2004)
12. Kuzuoka, H., Kosaka, J., Yamazaki, K., Suga, Y., Yamazaki, A., Luff, P., Heath, C.: Mediating dual ecologies. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW 2004), pp. 477–486 (2004)
13. Kuzuoka, H.: Spatial workspace collaboration: a SharedView video support system for remote collaboration capability. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1992), pp. 533–540 (1992)
14. Kirk, D., Rodden, T., Fraser, D.: Turn it this way: grounding collaborative action with remote gestures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2007), pp. 1039–1048 (2007)
15. Tecchia, F., Alem, L., Huang, W.: 3D helping hands: a gesture based MR system for remote collaboration. In: Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI 2012), pp. 323–328 (2012)
16. Yamashita, N., Kaji, K., Kuzuoka, H., Hirata, K.: Improving visibility of remote gestures in distributed tabletop collaboration. In: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW 2011), pp. 95–104 (2011)
17. Rajkumar, R., et al.: A resource allocation model for QoS management. In: IEEE Real Time Systems Symposium (1997)
18. Mohan, S., Kapoor, R., Mohanty, B.: Latency in HSPA Data Networks. Qualcomm Research Papers (2003)
19. Wu, W., Arefin, A., Rivas, R., Nahrstedt, K., Sheppard, R., Yang, Z.: Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. In: Proceedings of the 17th ACM International Conference on Multimedia (MM 2009), pp. 481–490 (2009)