The impact of bandwidth limitations and video resolution size on QoE for WebRTC-based mobile multi-party video conferencing

Dunja Vučić [1], Lea Skorin-Kapov [2], Mirko Sužnjević [2]

[1] Belmet 97

[2] Faculty of Electrical Engineering and Computing, University of Zagreb

dunja.vucic@belmet97.hr, {lea.skorin-kapov, mirko.suznjevic}@fer.hr

Abstract

Multi-party video conferencing is a highly delay-sensitive and resource-intensive service. A key challenge faced by service providers is meeting end user Quality of Experience (QoE) requirements given bandwidth limitations, especially in mobile networks. Service adaptation mechanisms may thus be employed to adapt the audiovisual quality so as to meet current network conditions and resource availability. In this paper we report on a subjective user study involving three-party audiovisual conversations established via mobile devices and aimed to investigate the impact of different video resolutions on QoE. We also investigate how different bandwidth limitations impact QoE for different video resolutions. The calls were established using a Web-based Real-Time Communication (WebRTC) application running on the Licode Multipoint Control Unit.

Index Terms: mobile multi-party video conferencing, QoE, WebRTC

1. Introduction

Advances in mobile network speeds, mobile device screen resolutions and processing power, and improved video coding standards, have given rise to increased audiovisual communication capabilities for mobile users. Going beyond two-way calls, there is an increase in multi-party conferencing services supporting three or more simultaneous participants in both business (e.g., telemeetings) and leisure (e.g., social interactions via Google+ hangouts) contexts. While a great deal of research has addressed the QoE modeling of two-way calls, multi-party scenarios have received less attention, in particular in the mobile context. Despite increases in available bandwidth, the high resource and strict latency requirements imposed in the context of multi-party mobile video conferencing continue to impose challenges in achieving a high level of end-user QoE. Coupled with addressing network challenges, service providers also face the challenge of delivering reliable, intuitive, and easy-to-use services.

Emerging technologies, such as Web Real-Time Communications (WebRTC), have made video conferencing free and available anywhere and on multiple devices. WebRTC is an open project that provides real-time audio and video communication within browsers without any plugins [1]. Supported interfaces built into the browser capture media streams from local devices (video cameras and microphones) and transfer media and data to other users.

Data transmission is enabled by either a peer-to-peer (P2P) or MCU (Multipoint Control Unit) architecture, with different infrastructure requirements on the client, server, and network (Figure 1). In the P2P case, both the number of network flows received and sent increases with the number of users. On the other hand, in the case of a deployed MCU (commonly a cloud-based solution), only the number of received flows increases. Previous research has shown that while current smartphones struggle to provide a sufficient level of QoE for a three-or-more party P2P video conferencing scenario [2, 3], improved performance may be achieved with the deployment of an MCU, given limited network resources (especially in the uplink direction) and highly demanding video streams.

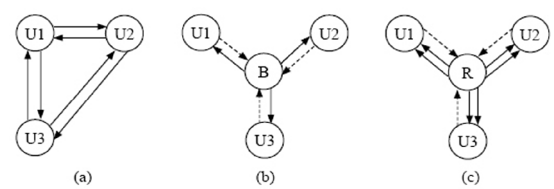

Figure 1: *Different video conferencing architectures: (a) fully meshed, (b) communication via a central bridge, (c) communication via a central media router*

Evaluation of the video conference requires the assessment of perceived quality by all participants. Unidirectional and bidirectional, dyadic standardized subjective quality assessment methods are well established by ITU Recommendations. While test methods for multi-party scenarios have also been proposed by the ITU-T [4], additional research is still needed to define relevant QoE models and provide the basis for improving service quality. Although there has been a significant amount of research on conversational audiovisual QoE, literature on multi-party video conferencing on mobile devices has not investigated multiple performance aspects relevant for achieving a satisfactory level of QoE. In [5, 6] the authors summarize QoE assessment challenges, and address special characteristics beside AV quality of telemeetings, such as mobility aspects, interoperability, additional functionalities, privacy, and secrecy concerns.

Video conferencing often results in an asymmetric set-up amongst participants, thus generating interoperability and resource issues. In [7], Lewcio et al. investigated speech and video telephony QoE for calls established via WiFi and HSDPA networks, and analyzed how the effect of switching between networks, codecs, and bit rates is perceived during an ongoing video call. Schmitt et al. conducted experiments in symmetric and asymmetric delay conditions in a controlled environment on a PC [11]. In similar studies [12, 13], the authors focused on symmetric delay and interactivity patterns. The impact of network QoS parameters was also evaluated in a mobile video chat experiment with "Vtok" between two end-users, one connected to 3G and the other connected to a WiFi network [5]. Video conferencing QoE is affected on several layers and often suffers because of congestions. To reduce loss and maintain a high level of bandwidth utilization, dynamic congestion control is proposed in [20]. In [21] authors investigated the efficiency of incentive mechanisms trading a higher QoE of video transmission for a user's consent to utilize their upload capacity. In [8], the authors performed subjective visual quality tests in an HD four-way desktop video conferencing system which requires audiovisual interaction. The authors investigated the impact of bitrate and packet-loss on overall, audio, and video quality under different Internet access technologies typical for domestic households: broadband, DSL, and mobile.

The layout used by a video conference application is important especially on small displays. In [16, 17], the authors explored user preferences for single and dual layouts for desktop video conferencing. To provide effective video conversation and task performance, orchestration could be used as a selection process of displayed information on each screen [18, 19].

Brendtsson et al. studied video quality for telemeeting scenarios with combinations of factors such as resolution, encoding bit rate, viewing distance and up scaling of video formats [15]. Experiments were conducted in different contexts (business meeting room, small office room, home living room, hotel room) with different types of connections and different end user devices (TV, PC, mobile phone). Wac et al. evaluated QoE for a set of widely used mobile applications on Android phones in natural environments and different contexts [14]. Thirty Android users participated in tests with three types of phones (Motorola, HTC and Samsung).

Building on previous work, in this paper we focus on examining which video resolutions are needed for achieving a satisfactory QoE, and how bitrate limitations impact QoE for different resolution settings in the context of multi-party mobile video conferencing. We report on a subjective user study involving 27 participants aimed at investigating the perceived quality of a three-party WebRTC-based video conference call with symmetric conditions. The two main research questions that we address are the following:

RQ1: When considering a three-way video call where each participant sees three video streams (the other two participants and their own video stream), what video resolutions are needed to achieve satisfactory QoE under different bandwidth constraints and taking into account device screen size?

RQ2: What is the impact of bandwidth restrictions on conversational interaction in a three-way video call established using smartphone devices?

2. Test methodology

Experiments included subjective end user assessments with the goal being to investigate the impact of different resolutions and bandwidth constraints on QoE. The subjective assessment followed the test procedure described in ITU-T Recommendation P.1301 [4]. Participants used the same multi-party video conferencing service and had the same smartphone configuration. The rational for using symmetric conditions was to eliminate the impact of different device and network settings between participants, as this will be addressed in future work. Tests were designed to mimic real world settings, hence experiments were conducted in a leisure context in a natural home environment.

The three-party video conference was set up using a WebRTC application running on the Licode MCU [10] installed in a local network, to avoid impairments caused by a commercial network, while still enabling us to control application configuration parameters, bandwidth, and video resolution (Figure 2).

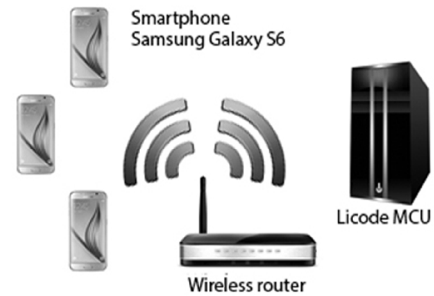

Figure 2: *System set-up over LAN*

Licode offers a client API - *Erizo* to handle connections to virtual meeting rooms and streams in Web applications, and a server API for communication with *Nuve*, a module that manages video conference rooms. To clarify, when referring to setting bandwidth constrains in all test cases, we refer to manipulation of the settings within Licode's Erizo API. Both average and maximum bandwidth values within the API have been set to the same value. For our testbed setup, Licode is installed on a laptop with Intel Core i5 Processor, 2.6 GHz, 8 GB RAM and Ubuntu 12.04 LTS. The LAN connection between end user devices and the media server is Wi-Fi 802.11n, on port 3001. Video conversation is initiated through the Samsung browser version 4.0.10-53. All participants used Samsung Galaxy S6, with display size 5.1", resolution 1440 x 2560 pixels, secondary camera 5 MP, f/1.9, 22mm and 1440p@30fps smartphones (Figure 3).

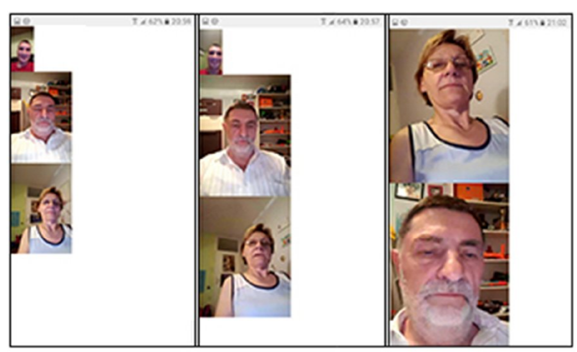

Figure 3: *Smartphone video conference over Licode, resolution 480x320, 640x480, 960x640*

Participants were located in three different rooms, with the following dimensions LxWxH (cm): room 1 - 385x327x260, room 2 - 385x250x260, room 3 - 385x320x260. During the experiments the maximum background light intensity was 21.3 lx with a maximum background noise level of 36.8 dB.

2.1. Participants

Twenty-seven participants took part in the study and were divided into **9 fixed groups** with **3 members each**. 14 male and 13 female participants took part in the studies, with an average age of 38 years (minimum 32 and maximum 65 years old). Considering acquaintances between users, free conversation was chosen to represent a natural interactive conversation [9]. The conversations were all conducted using the Croatian language, as this was the native language to all participants.

The selected participants had no special knowledge of AV technology nor were technical experts regarding the equipment and services to be tested. However, eight of them had participated previously in subjective assessments. Participants were comprised of volunteers, and all have normal or corrected vision and normal hearing.

2.2. Test conditions

To explore the effects of video resolutions and bandwidth limitations on perceived quality, and to avoid the impact of end user devices, all participants used the same high end smartphone configuration (in our previous work we studied the device requirements for various WebRTC applications in a 3-way mobile call scenario [3]). Overall 108 tests were performed. The test schedule consisted of each user group testing 12 conditions with different combinations of video resolutions and bandwidth, each lasting 3 minutes. As previously stated, both bandwidth and video resolution in the tests were controlled using settings in the Licode Erizo API. We first performed 3 tests in which only the video resolution was altered: 960x640, 640x480, 480x320 under no bandwidth constraints (i.e., a bandwidth setting of 50Mbit/s was set within Licode) so as to evaluate QoE differences under each resolution. We further performed tests in which groups were assigned three different bandwidth constraints per resolution, namely 300 kbps, 600 kbps and 1200 kbps. Although the process of setting up and session teardown has an impact on overall QoE, we wanted to avoid evaluation of their influence and thus had the test administrator establish the sessions prior to each test condition.

Table 1. *Highest measured values of packet loss and jitter per each condition*

Bitrate	Resolution	Packet loss %	Max jitter ms	Mean jitter ms
300 kbps	480x320	0,1	37,47	7,7
	640x480	0,02	27,27	7,75
	960x640	0,01	28,06	12,18
600 kbps	480x320	0,3	43,63	8,64
	640x480	0,1	34,34	9,16
	960x640	0,2	40,55	13,06
1200 kbps	480x320	0,3	41,23	13,65
	640x480	0,02	43,31	16,53
	960x640	0,32	40,92	15,45
50000 kbps	480x320	0,2	55,94	10,45
	640x480	0,6	55,94	12,82
	960x640	0,15	38,69	15,6

At the beginning of each session, a preliminary test was carried out to familiarize participants with the task and assessment questionnaire. Preliminary results are not taken into account. After the completion of each condition, subjects were asked to rate *overall quality* and *interaction quality* using a paper questionnaire and the 5-pt. Absolute Category Rating (ACR) scale: 1 "Bad", 2 "Poor", 3 "Fair", 4 "Good", 5 "Excellent".

The physical parameters during testing were slightly different, since each participant was located in separate room. The average RTT time from the MCU to all client devices was on average less than 50 ms. We further noted packet loss and jitter from analysis of the RTP stream as measured in Wireshark, and the highest measured values for each condition are shown in Table 1.

Although the set BW values on the Licode MCU were 300, 600, 1200 kbps, the actual measured bitrate values measured by Wireshark on the computer that hosted the Licode server are shown in Table 2.

Table 2. *Average measured throughput values for each test condition*

kbps / **resolution**	**300**	**600**	**1200**	**50000**
480x320	417,03	692,40	1096,46	1168,49
640x480	405,02	644,34	1009,51	1261,10
960x640	411,39	615,23	1073,74	1213,57

3. Discussion

Figure 4 depicts the dependency of overall quality ratings on different combinations of values for bandwidth and resolution parameters. It should be noted that the bandwidth of 50000 kbps represents unlimited bandwidth. Two main conclusions can be drawn from Figure 4: 1) the resolution 960x640 should not be set for any of the tested bandwidth limitations, as for that resolution MOS scores for overall quality are always below 4 and lower than other tested resolution settings; and 2) unlimited bandwidth setting results in significantly reduced user perceived overall quality for all resolutions above 480x320, meaning that the capabilities of the tested mobile phones had trouble processing multiple real-time videos with high bitrates and resolutions. These findings are in line with our previous findings [3] and may be considered generally applicable, as for the testing procedure very powerful Samsung Galaxy S6 mobile phones were used, which can be considered high end mobile devices available on today's market. In every test with unlimited bandwidth, participants reported picture freezing, although the speech was unimpaired, so communication was not completely interrupted. Consequently, the ratings were still fair, although significantly lower as compared to the other bandwidth limitations. The area with an optimal combination of parameters is clearly depicted in Figure 4. with MOS scores over 4.5. What is interesting that all combinations between 1200 and 300 kbps and both 480x320 and 640x480 resolutions are in this area. In experiments with bandwidth limitations of 300kbps overall quality gained the highest scores for all resolutions. The experiments with a set resolution of 640x480 gained the highest average scores (over 4.5) for overall quality.

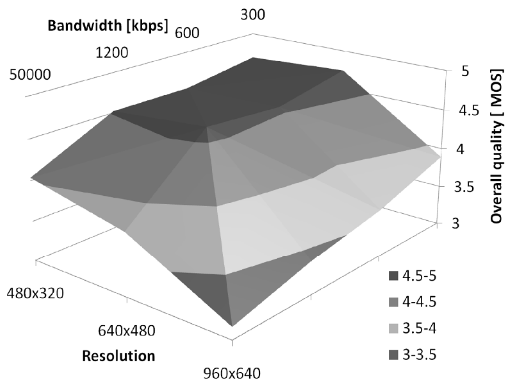

Figure 4: *Overall quality for each combination of bandwidth and resolution settings (note: bandwidth corresponds to settings in Licode and actual throughput was slightly higher, as shown in Table 2)*

Besides overall quality, we also measured interactivity perceived by the users. In Figure 5 we depict the MOS values for both overall quality and interactivity for resolution 640x480 across all bandwidth limitations. It can be noted that overall quality and interactivity are highly correlated and that their 95% confidence factors overlap for every experiment. This statement is valid for other resolutions as well (charts omitted due to space restrictions). Pearson's correlation coefficient between these two parameters for all test cases is 0.809.

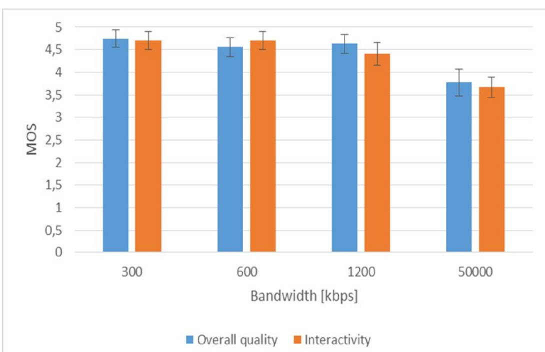

Figure 5: *Overall quality and interactivity for 640x480 resolution for all bandwidth conditions*

Instead of achieving the highest quality ratings, the largest test resolution as well as highest bitrate seem to have caused congestion on the smartphones, which ultimately affected the perceived quality. We analyzed the impact of the user throughput on QoE as shown in Figure 6. (single flow) and 7 (aggregated - all flows). It should be noted that for highest resolution, throughput of aggregated flows is lower due to frequent video freezing. The amount of generated traffic had a significant influence on the QoE, especially for bitrates higher then 1200kbps for each resolution tested, which may be attributed to high demands on smartphone processing power. However, despite the lack of smartphone processing power, 5.1" screen size remains as an argument that resolutions higher than 640x480 are unnecessary for three-party video conference calls.

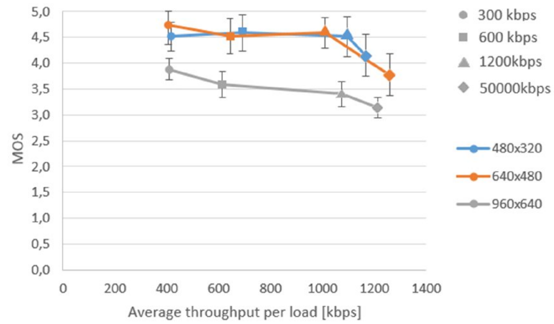

Figure 6: *Plot of average generated bitrate for different resolutions and impact on overall quality.*

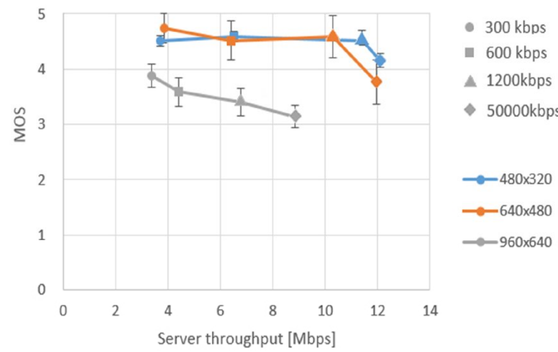

Figure 7: Plot of generated average throughput for different resolutions and impact on overall quality.

4. Conclusions

The goal of this paper has been to empirically study the impact of different video resolutions on QoE under bandwidth constraints in a multi-party mobile video conference call. Tests have been conducted in a realistic home environment setting under controlled network conditions and by employing the Licode MCU architecture. Results obtained in a subjective study showed that while higher video resolutions contribute to better video quality, they also impose higher processing requirements on the system, and lead to congestion under certain bandwidth limitations. The highest tested resolution yielded the lowest QoE both in cases of constrained bandwidth and unconstrained bandwidth, with the latter attributed to insufficient smartphone processing power. Differences in perceived QoE between 480x320 and 640x480 were not significant. Based on obtained results, the identified target resolution was 640x480 with a set bandwidth limitation of 300kbps (thus limiting the encoding bitrate). Future work will extend this investigation with respect to testing different mobile devices and the impact of realistic network bandwidth constraints on QoE, with the goal being to derive QoE-driven service adaptation strategies for multi-party mobile video conferencing. Moreover, clearer insight into bandwidth requirements for such services can provide input for network resource allocation mechanisms.

5. Acknowledgements

This work has been supported in part by the Croatian Science Foundation under the project UIP-2014-09-5605 (Q-MANIC).

6. References

[1] A. Bergkvist, D. C. Burnett, C. Jennings, "A.Narayanan, WebRTC 1.0: Real-time Communication Between Browsers," W3C Working Draft 10 September 2013.

[2] Y. Xu, C. Yu, J. Li, Y. Liu, "Video telephony for end consumers: measurement study of Google+, iChat, and Skype," Proceedings of the 2012 ACM conference on Internet measurement conference New York, 371–384. 2012.

[3] D. Vucic and L. Skorin-Kapov, "The impact of mobile device factors on QoE for multi-party video conferencing via WebRTC, " in Telecommunications (ConTEL), 2015 13th International Conference on, July 2015, pp. 1–8.

[4] Recommendation ITU-T P.1301 (2013), "Subjective quality evaluation of audio and audiovisual multiparty telemeetings"

[5] S. Jana, A. Pande, A. Chan, P. Mohapatra, Mobile video chat: issues and challenges," IEEE Communications Magazine, Year: 2013, Volume: 51, Issue: 6, Pages: 144 - 151

[6] J. Skowronek, K. Schoenenberg, and G. Berndtsson, "Multimedia Conferencing and Telemeetings, in Quality of Experience: Advanced Concepts, Applications, and Methods," Springer, 2014.

[7] B. Lewcio, "Management of Speech and Video Telephony Quality in Heterogeneous Wireless Networks," T-Labs Series in Telecommunication Services, DOI: 10.1007/978-3-319-02102-7_5, Springer International Publishing Switzerland 2014.

[8] M. Schmitt, J. Redi, P. Cesar, D. Bulterman, "1Mbps is enough: Video Quality and Individual Idiosyncrasies in Multiparty HD Video-Conferencing," Proceedings of the 8th International Workshop on Quality of Multimedia Experience (QoMEX), 2016.

[9] Recommendation ITU-T P.920 (2000), "Interactive test methods for audiovisual communications"

[10] Licode Open Source WebRTC Communications Platform http://lynckia.com/licode/

[11] M. Schmitt, S. Gunkel, P. Cesar, D. Bulterman, "Asymmetric Delay in Video Mediated Group Discussions," Proceedings of the 6th International Workshop on Quality of Multimedia Experience (QoMEX), 2014.

[12] M. Schmitt, S. Gunkel, P. Cesar, D. Bulterman, "The influence of interactivity patterns on the Quality of Experience in multi-party videomediated conversations under symmetric delay conditions," Proc. of the 3rd International Workshop on Socially-aware Multimedia, 2014.

[13] J. Xu, B. W. Wah, "Exploiting just-noticeable difference of delays for improving quality of experience in video conferencing," in Proceedings of the 4th ACM Multimedia Systems Conference, New York, NY, USA, pp. 238–248, 2013.

[14] K. Wac, S. Ickin, J. Hong, L. Janowski, M. Fiedler, A. Dey, "Studying the experience of mobile applications used in different contexts of daily life," In ACM SIGCOMM Workshop on Measurements Up the STack (WMUST), 2011.

[15] G. Berndtsson, M. Folkesson, V. Kulyk, "Subjective quality assessment of video conferences and telemeetings," 19th International Packet Video Workshop (PV), pp. 25–30, 2012., DOI: 10.1109/PV.2012.6229740

[16] S. Junuzovic, K. Inkpen, R. Hegde, Z. Zhang, "Towards ideal window layouts for multi-party, gaze-aware desktop videoconferencing," in Proceedings of Graphics Interface 2011, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2011, pp. 119–126.

[17] S. Gunkel, M. Schmitt, and P. Cesar, "A QoE study of different stream and layout configurations in video conferencing under limited network conditions," in Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on, May 2015, pp. 1–6.

[18] M. Ursu, P. Torres, V. Zsombori, M. Franztis, R. Kaiser, "Socialising through orchestrated video communication," in Proceedings of the 19th ACM international conference on Multimedia, pp. 981–984, 2011.

[19] M. Falelakis, M. Groen, M. Frantzis, R. Kaiser, M. Ursu, "Automatic orchestration of video streams to enhance group communication," in Proceedings of the 2012 international workshop on Socially-aware multimedia, New York, NY, USA, pp. 25–30., 2012., DOI: 10.1145/2390876.2390886

[20] G. Carlucci, L. De Cicco, S. Holmer, S. Mascolo, "Analysis and design of the google congestion control for web real-time communication (WebRTC)," MMSys '16 Proceedings of the 7th International Conference on Multimedia Systems Article No. 13

[21] M. Wichtlhuber, N. Aleksandrov, M. Franz, O. Hinz, D. Hausheer, "Are incentive schemes needed for WebRTC based distributed streaming?: a crowdsourced study on the relation of user motivation and quality of experience", MMSys '16 Proceedings of the 7th International Conference on Multimedia Systems