

The Impact of Network Impairments on the QoE of WebRTC applications: A Subjective study

Gulnaziye Bingol¹, Luigi Serreli¹, Simone Porcu^{1,2}, Alessandro Floris^{1,2}, and Luigi Atzori^{1,2}

¹ DIEE, University of Cagliari, 09123 Cagliari, Italy

² CNIT, University of Cagliari, 09123 Cagliari, Italy

{gulnaziye.bingol, luigi.serreli, simone.porcu, alessandro.floris84, l.atzori}@unica.it

Abstract—WebRTC-based applications allow for real-time communications that are subject to network impairments affecting the end user's Quality of Experience (QoE). In this paper, we conducted subjective tests involving 20 people to investigate the conversational quality of a two-party WebRTC-based audiovisual telemeeting service. A dedicated system was implemented to introduce controlled network impairments (delay, jitter, and packet loss) to impair the communication between the parties. In addition, test participants had to rate the perceived QoE for the audio, the video, and the overall service, as well as the three emotional dimensions, i.e., valence, arousal, and dominance. Extensive results were obtained regarding the impact of the network impairments on the multimedia quality, the emotional dimensions, and the communication feasibility.

Index Terms—Quality of Experience, WebRTC, Network impairments, Subjective assessment, Self-Assessment Manikin.

I. INTRODUCTION

The ongoing Covid-19 pandemic has contributed to a drastic increase in the use of videoconferencing tools for different purposes, such as work meetings, e-learning, and social interactions. Among common, dedicated tools (e.g., Cisco WebEx, Microsoft Teams, and Zoom), numerous applications based on WebRTC technology (e.g., Google Hangouts Meet, BlueJeans, Lifesize, Slack) have been widely used for audiovisual conversation over the Internet. One of the reasons is that WebRTC enables the easy embedding of real-time voice and video communications on Web browsers, thus avoiding the installation of dedicated tools on the user's device [1].

Being based on real-time communications, the WebRTC traffic is subject to network impairments (e.g., delay, jitter or packet loss), which affect the Quality of Service (QoS) and, in turn, the end-user's Quality of Experience (QoE). Thus, studies focused on the quality assessment of WebRTC-based applications are needed to gain insights into how user-perceived service quality can be measured and optimized within available resource constraints. In particular, the QoE assessment has

become of crucial importance for the successful deployment of multimedia services as the QoE reflects the subjective quality perceived by the user [2], [3]. The studies that have been conducted so far for the evaluation of WebRTC-based videoconferencing services have major limitations: most of them concern objective quality evaluation [4], some focus on passive viewing and listening to recorded audiovisual contents subjectively assessed by people [5], some others are based on subjective assessment where participants actively interact through audiovisual systems based on outdated setup [6]–[8]. To the best of the authors' knowledge, only two studies in the literature conducted subjective studies of interactive conversations using WebRTC-based applications [9], [10]. However, in [9], only four test conditions were considered, whereas in [10], QoE measurements were focused on single impairment scenarios.

The current state of art calls for further activities to explore the subject, especially when the service setting introduces some quality-related problems. Therefore, in this paper we conducted subjective tests to investigate the conversational quality of a two-party WebRTC-based audiovisual telemeeting service to answer the following questions:

- Q1: “Which are the relationships between major network impairments and the resulting quality?”
- Q2: “Which media component is the most sensitive to the network impairments and to which extent?”
- Q3: “Which is the impact of network impairments on the emotional dimensions and communication feasibility?”

The subjective quality assessment involved 20 participants, which were located in different rooms and had to communicate through the implemented WebRTC-based application. In between, a dedicated system introduced controlled impairments, i.e., network delay, jitter, and packet loss. The subjects had to rate the perceived QoE for the audio, for the video, and for the overall service as well as three emotional dimensions, i.e., valence, arousal, and dominance. Experiment results, in terms of the Mean Opinion Score (MOS), show the perceived audio QoE achieved better results than the video QoE and overall QoE. Also, the mean value computed between the audio QoE and video QoE seems to be a good indicator for estimating the overall QoE. Furthermore, the valence results suggest participants felt rather pleasant during the entire experiment, even when the conversation was exposed to quality degradations.

This work has been partially funded by the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) with the PON “Ricerca e Innovazione” 2014-2020 (PON R&I) “Azione IV.4 Dottorati e contratti di ricerca su tematiche dell'innovazione” and “Azione IV.5 Dottorati su tematiche green”, assigned with D.M. 1062 on 10.08.2021.

©2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. DOI: 10.1109/QoMEX55416.2022.9900882

The paper is structured as follows. Section II discusses the related work in this area. Section III presents the methodology followed by the proposed study, whereas Section IV discusses the experiment results. Finally, Section V concludes the paper.

II. RELATED WORK

Most of the studies available in the literature regarding the performance evaluation of audiovisual services concern objective quality evaluation [4] or passive viewing and listening of recorded audiovisual contents subjectively assessed by people [5]. Subjective assessment where participants actively interact through audiovisual systems are instead limited and require further study. To the best of the authors' knowledge, the only studies considering the interactive conversation scenario are the following. In [6], an open-source VoIP client, called PJSIP, was used to conduct experiments regarding interactive audiovisual conversations with realistic degradations, such as video/audio compression (bitrate and codec) and network packet loss rate (PLR). Three qualities were considered, namely, best (no PLR), medium (0.5% PLR), and low (5% PLR). The "building block scenario" described in the ITU-T Rec. P.920 was used as the conversational task [11]. Experimental results showed the audiovisual quality was more significantly influenced by the video than by the audio channel settings. The same authors compared in [7] the results of the passive and interactive assessment of audiovisual conversations, when impaired by media (codec and bitrate) and network (packet loss) distortions. It resulted that the perceived audio, video, and audiovisual quality were lower for the passive assessment, than the interactive assessment because the participants were more attentive to detecting visual impairments during the passive assessment, whereas during the interactive assessment they were more focused on the conversation with the partner. In [8], the Linphone software was used to create audiovisual communications that were distorted by network delay (6 values between 100 ms and 1800 ms) and packet loss (1%, 5%, and 15% for the audio and 0.1%, 0.2%, and 0.5% for the video, respectively). The "building block scenario" was also used in this case as the conversational task. When the delay was introduced in the communication, the perceived video quality was higher than the audio quality. However, even if 1800 ms of delay was applied, the audiovisual MOS only decreased from 4.50 to 3.46. On the other hand, when the packet loss was introduced, the video quality was lower than the audio quality, as in [6], and the audiovisual MOS decreased from 3.68 to 2.32.

It should be noted that these studies are based on a relatively outdated setup with low resolution video, low encoding bitrates, and outdated audio and video codecs. Also, none of them used WebRTC-based applications. Therefore, it is unknown whether these results are applicable to more recent settings. The only subjective studies of interactive conversations using WebRTC-based applications are [9] and [10]. In [9], four different technical conditions were considered, namely, no distortions, distorted audio (CPU usage limited to 70%), distorted video (PLR of 20%), and distorted audio and

video (delay of 500 ms and jitter of 300 ms). The celebrity name-guessing task was chosen as the conversational task. The obtained self-report results suggest that the overall quality and annoyance reflect the varying quality conditions. In particular, the distortion of both audio and video led to the lowest quality and highest annoyance ratings. However, this is not reflected when considering the emotional valence ratings. Four technical variables were considered in [10], namely, delay (from 150 to 1600 ms), jitter (from 0 to 400 ms), and packet loss (from 5 to 40%). The obtained subjective results were used to calibrate the proposed Deterministic QoE model (DQX), which outperformed the MOS prediction performance of the ITU E-model [12], in particular when the delay impairs the communication. This result may suggest that reference QoE models for VoIP, such as the E-model, may not be quite accurate when WebRTC technology is used. For this reason, in this paper, we conducted a subjective study to investigate the impact of network impairments on the QoE when using a WebRTC-based application for interactive conversations.

III. METHODOLOGY

A. Experiment design

The objective of the experiment is to investigate the conversational quality of a WebRTC-based audiovisual telemeeting service impaired by poor network conditions. The considered network parameters are the network delay, jitter, and packet loss. We combined the impact of these parameters for a total of 15 test conditions (TCs), which are summarized in Table I.

As highlighted by different research studies [13]–[15], delay, jitter, and packet loss are among the most influential factors for WebRTC video calling applications. A low delay in RTC applications is critical, particularly for interactive scenarios. The jitter is the variation in the delay on a packet flow and can lead to unintended deviations in audio and video that degrade the quality of communications. Finally, the packet loss regards the packets dropped to mitigate potential congestion issues, which contribute to the degradation of the audiovisual communication [1]. The values of delay, jitter, and packet loss selected for this experiment, shown in Table I, may seem quite higher than those normally considered for VoIP communications. For example, the ITU-T Rec. G.114 specifies that the one-way delay should preferably be kept below 150 ms for VoIP services [16]. Once the delay exceeds this value, exponential degradation of the perceived QoE of speech is observed, and delays above 400 ms are considered unacceptable [17]. The relationship between QoE and PLR is also exponential and depends on various parameters, such as codec resistance to packet loss and equipment impairment. Small values of PLR (e.g., 5%) for some cases may make the communication quality unacceptable [17]. However, WebRTC-based applications implement packet retransmissions and forward error correction (FEC) to handle network distortions. In our case, the Opus in-band FEC was used to protect the audio streaming, whereas redundant audio data (RED), FEC, and retransmission mechanisms were used by the RTCP (Real-time Transport Control Protocol) protocol to protect the media

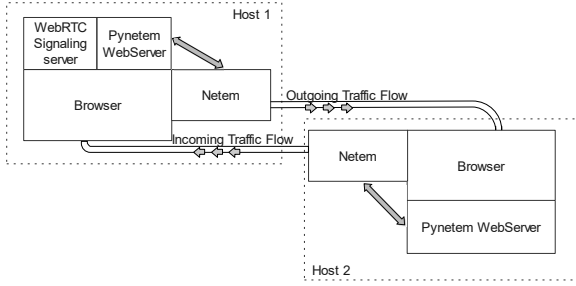


Fig. 1. The system architecture.

streaming. Furthermore, the Google Chrome browser was used for the experiment, which implements the Google Congestion Control (GCC) algorithm that adapts the media sending rate to the link capacity [18]. Thanks to the implemented packet retransmission and error correction techniques, WebRTC-based applications seem to be less sensitive to network impairments. It is then conceivable that stronger values of network distortions are needed if we want to be sure that the QoE of the test participants would be impacted during the conversation. Indeed, the studies in [9], [10] used values of delay, packet loss, and jitter comparable to those considered in this study.

B. The developed system

The architecture of the proposed system is shown in Fig. 1. The two hosts correspond to the two laptops used by the pair of participants for the telemeeting. The Google Chrome browser was used as the software to implement the WebRTC-based video call. The signalling server needed to synchronise the WebRTC session was only installed on one of the laptops. As shown in Fig. 2, a router was used to put the two hosts in communication in separate rooms by creating a dedicated and controlled network connection needed to avoid uncontrollable network distortions provided by the Internet traffic. It was necessary to impose network rules related to the traffic to perform the different test conditions by applying the network impairments. Although some proprietary network controlling tools are available, they are expensive hardware solutions. However, there are also open-source network emulation tools that are widely used for research initiatives. For instance, Dummynet was used to test and simulate network protocols [19], while the NetEm (Network Emulator) software [20] was used to correlate QoS parameters with the QoE by emulating network degradations for video streaming applications [21].

The proposed system used PyNetem, which establishes a Web server to perform dynamic and comfortable HTTP requests by embedding different NetEm rules to guarantee certain transparency for users who do not have to notice the employment of the impairments. However, NetEm enables the addition of traffic rules just to the outgoing traffic [22]. Therefore, as depicted in Fig.1, it was necessary to employ NetEm in both hosts to apply the rules to the incoming traffic. Afterwards, the iPerf tool was used to verify and measure the impairments carried out by NetEm in both directions, incoming and outgoing traffic. Once the system was ready to

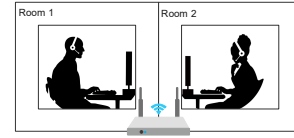


Fig. 2. The scenario of the experiment with the participants in different rooms.

introduce the network impairments for each WebRTC session, according to the ITU-T P.800, we set the system to randomly implement the considered TCs to the different test participants, so as to avoid connections among the parameters of interest.

C. Subjective quality assessment

In total, 20 participants (11 females and 9 males) between 23-36 years, who reported normal or corrected-to-normal vision, took part in the experiment. The ITU-T Rec. P. 920 suggests that lively audiovisual conversations can be stimulated if the participants in such a test know each other, and therefore familiarity between pairs of conversing participants is highly desirable. For these reasons, we selected students of the same course of study who have known each other for more than two years [11]. The assessment was conducted at the Department of Electrical and Electronic Engineering (DIEE) of the University of Cagliari, in Cagliari, Italy. The test desks were placed in two different rooms to prevent participants from seeing and hearing the person with whom they were making the Web call out of the call session. The participants were also asked to wear headphones during the assessment to prevent external disturbance. The “Who am I?” celebrity name-guessing task was chosen as the conversational task. As suggested by the ITU-T Rec. P.1305, the game character of this task lets people find themselves in a joyful state, have fun and interact naturally [23]. Each conversation partner had to guess which celebrity she/he was by asking yes/no questions to the conversation partner.

According to the ITU-T Rec. P.920, a scenario of the intended application of the system under test should be given to the subjects before starting the experiment to show the range and type of impairments. Therefore, before conducting the assessment, the participants were informed that the network conditions would be changed for each telemeeting and that they were asked at the end of each call to fill out a short questionnaire (using a web form) to report on their perceived quality and experienced emotion. To give participants an overview of the impairments involved in the telemeetings, three training telemeetings were conducted before the experiments: one with no network impairments, one with moderate impairments, and one setting the worst network conditions (as for TC15). However, during the experiments, all the TCs in Table I were applied in random order to the participants. During these training sessions, the participants also exercised with the celebrity name-guessing task. Before the assessment, the participants were also trained on the utilization of the rating methods. The single discrete Absolute Category Rating (ACR) scale with five category labels (Bad, Poor, Fair, Good, and

TABLE I
TEST CONDITIONS (TCs).

TC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Delay (ms)	0	500	1000	500	1000	0	500	1000	500	1000	0	500	1000	500	1000
Jitter (ms)	0	0	0	500	500	0	0	0	500	500	0	0	0	500	500
PLR (%)	0	0	0	0	0	15	15	15	15	15	30	30	30	30	30

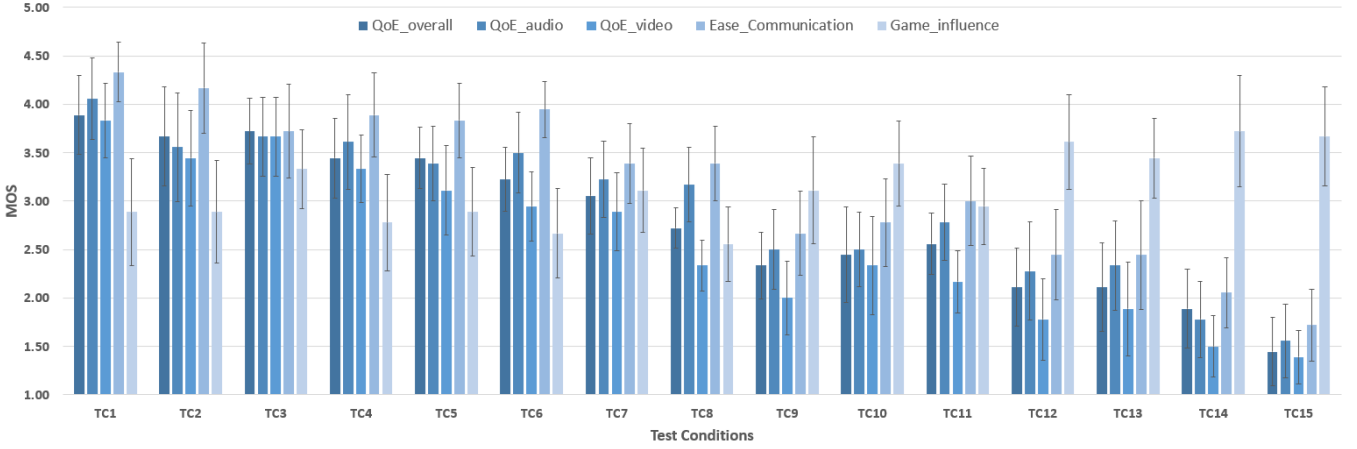


Fig. 3. Mean Opinion Score (MOS) with 95% confidence interval (CI).

Excellent) was used to rate the perceived overall QoE, the video QoE, and the audio QoE, according to the ITU-R Rec. P.800 [24]. Three discrete, 5-class graphical rating scales were used to rate the emotional dimensions (i.e., valence, arousal, and dominance) from the Self-Assessment Manikin (SAM) technique [25], which is a non-verbal pictorial assessment to evaluate emotions. Moreover, the participants were asked two further questions, i.e., “Has the call quality influenced the game?” (1-Not at all, 2-A little, 3-Slightly, 4-Significantly, 5-Extensively) and “How easy/difficult it was to communicate with the partner?” (1-Very difficult, 2-Difficult, 3-Neither easy nor difficult, 4-Easy, 5-Very easy). The questionnaire was then composed of a total of 8 questions. The overall time needed to complete the assessments was about 40 minutes. Indeed, each call was stopped after 2 minutes, even if participants did not guess the celebrity chosen by the partner. In that case, the participant could continue asking questions in the next TC. A break of 5 minutes was taken in the middle of the experiment to avoid test participants losing their attention.

IV. EXPERIMENT RESULTS

In this section, we present the experiment results, which were computed considering the data from 18 test participants. Indeed, 2 of the 20 participants were identified as outliers according to the “final rejection criteria for discarding an observer of a test” provided in the ITU-R BT.500-14 [26]. However, the data from 18 subjects is still enough to derive significant results since the ITU-T Rec. P.920 suggests that at least 16 subjects should participate in a test.

A. Impact of impairments on the QoE and on each communication component - Q1 and Q2

Fig. 3 shows the MOS with a 95% confidence interval (CI) computed for each of the 15 TCs in Table I and regarding the perceived overall QoE, audio QoE, video QoE, ease of communication, and influence of the audiovisual call quality on the game. From this, it can be noticed that the audio QoE is most of the time greater than the video QoE and the overall QoE. This is particularly true when the network impairments are more significant, which suggests the audio streaming is more resistant to poor network conditions. Indeed, a sufficient to good audio QoE (MOS between 3 and 4) is perceived when the PLR is not present (TCs 1-5) and when a PLR of 15% is present alone or combined with the delay (TCs 6-8). However, when also the jitter is added to the delay in combination with a PLR of 15% (TCs 9-10), the audio QoE decreases to the range of poor-sufficient quality (MOS between 2 and 3). The same quality is perceived when a PLR of 30% is present alone or combined with the delay only (TCs 11-13). The lowest audio QoE, poor to bad (MOS between 1 and 2), is perceived when the jitter is also added to the delay in combination with a 30% of PLR (TCs 14-15). On the other hand, the video QoE always achieved the lowest rates when compared with the overall QoE and, especially, the audio QoE. Again, this difference is clearer when the network impairments are substantial, particularly when the packet loss is present. Indeed, the video QoE is perceived as sufficient for good QoE (MOS between 3 and 4) only when the packet loss is not present (TCs 1-5). The quality decreases to the range poor-sufficient (MOS between 2 and 3) when a PLR of 15% is present alone or combined with delay and jitter (TCs

TABLE II
PEARSON CORRELATION COEFFICIENT (PCC) COMPUTED BETWEEN THE CONSIDERED EVALUATION METRICS.

PCC	Overall QoE	Audio QoE	Video QoE	Ease of Comm.	Game Influence	Valence	Arousal	Dominance
Overall QoE	1	0.979	0.988	0.977	-0.681	0.980	0.898	0.949
Audio QoE		1	0.962	0.986	-0.778	0.986	0.899	0.957
Video QoE			1	0.951	-0.617	0.961	0.913	0.951
Ease of Comm.				1	-0.801	0.984	0.886	0.949
Game Influence					1	-0.744	-0.610	-0.669
Valence						1	0.914	0.942
Arousal							1	0.930
Dominance								1

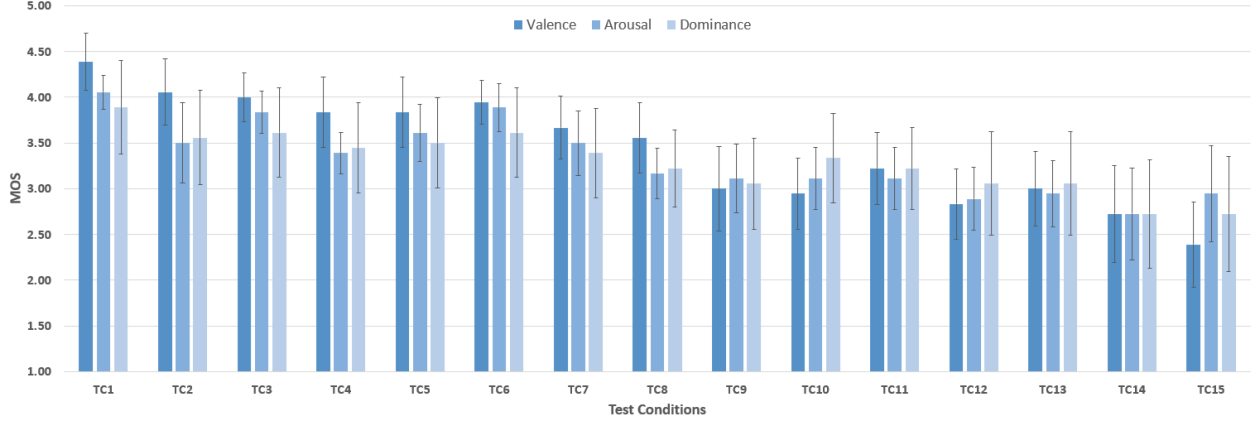


Fig. 4. MOS with 95% CI for emotions (valence, arousal, and dominance).

6-10), and when a PLR of 30% is present alone (TC 11). The lowest video QoE, poor to bad (MOS between 1 and 2), is perceived when a PLR of 30% is present in combination with delay and jitter (TCs 12-15). When considering single variable scenarios, delay and PLR achieved MOS values in line with those achieved in [10]. Multi-variable MOS results are in line with those achieved in [9], although the MOS difference among TCs is quite less evident in our case. In particular, in [9], the audiovideo-distorted TC (delay of 500 ms and jitter of 300 ms) was rated significantly worst (MOS around 2) than the video-only TC (PLR of 20%, MOS around 3). In our experiment, TC4 (delay of 500 ms and jitter of 500 ms) and TC6 (PLR of 15%) achieved a MOS of 3.44 and 3.22, respectively.

The mean MOS computed respectively for the overall MOS, audio MOS, and video MOS are 2.80, 2.93, and 2.57, which confirm the audio QoE was perceived much better than the video QoE. This can be due to the lower bandwidth needed by the audio data with respect to the video data, and by the concealment techniques implemented by the WebRTC application to mask the network impairments. Indeed, the audio stream is not subject to congestion control by the GCC due to its low data rate [27]. Interestingly, the overall MOS resulted very close to the mean between the audio MOS and the video MOS, which we computed for each TCs and referred to as MOS_{AV} . The Mean Squared Error (MSE) and the Pearson Correlation Coefficient (PCC) between the overall MOS and MOS_{AV} are respectively 0.01 and 0.993, which proves the mean MOS

between the audio MOS and video MOS to be a good indicator for estimating the overall MOS, for this case. The ease of communication achieved, in general, greater rating scores than the QoE. However, similarly to the QoE, these scores decrease as the network impairments increase. This result reveals that, as expected, the communication between the two parties was more difficult when the network impairments impacted more on the audiovisual signal. However, the last indicator in Fig. 3, i.e., the influence of the call quality on the game, indicates that even if the network was impaired, the two parties were somehow able to communicate and exchange useful insights to guess the celebrity associated with the conversation partner. The participants indicated, on average, a little or a slight influence of the audiovisual call quality on the game. Only when the PLR reached 30% in combination with delay and jitter (TCs 12-15) was the influence rated as significant.

Table II shows the PCC computed between the considered evaluation metrics. The overall QoE is highly correlated with all the evaluation metrics. Among these, the PCC with the video QoE is the greatest (0.988), whereas the lowest (and negative) is with the influence of the audiovisual call quality on the game (-0.681). The audio QoE is less correlated with the overall QoE (0.979) than the video QoE (0.988), but it is more correlated with the ease of communication (0.986) and the game influence (-0.778) than both the overall QoE (0.977 and -0.681, respectively) and the video QoE (0.951 and -0.617, respectively). These results strengthen the fact that the audio QoE was the most important quality indicator

for the considered experiment; thus, when the audio QoE was perceived as good, it was easier to communicate as well as to complete the game between the two parties.

B. Impact of network impairments on the different emotional dimensions and the communication feasibility - Q3

Fig. 4 shows the MOS with 95% CI computed for each of the 15 TCs in Table I and regarding the perceived emotional dimensions, i.e., valence, arousal, and dominance. The valence identifies the unpleasant to pleasant feelings of happiness. The mean valence is 3.43, which suggests participants felt rather pleasant during the entire experiment, even when the conversation was exposed to quality degradations. A similar result is achieved in [9]. Indeed, the valence MOS decreases with the increase of network impairments, but its value is always greater than 2.5, except for the worst network condition (TC15). According to this trend, the PCC between the valence and the QoE is very high, i.e., 0.980, 0.986, and 0.961 with the overall QoE, audio QoE, and video QoE, respectively. The valence also results highly correlated with the ease of communication (0.984) and the game influence (-0.744). The arousal reflects how excited or apathetic the emotion is, whereas the dominance takes into account the extent to which the emotion makes the subject feel they are in control of the situation. Unlike the valence, the arousal and dominance assumed a kind of *sinusoidal decreasing* trend with the increase of the network impairments. Because of that, these emotional dimensions are less correlated with the QoE, with the arousal achieving the lowest PCC results.

V. CONCLUSION

We investigated the impact of different network impairments (delay, jitter, packet loss) on the multimedia quality, the emotional dimensions, and the communication feasibility perceived when performing interactive audiovisual conversations through WebRTC-based applications. Experiment results show that the audio QoE achieved better results than the video QoE and overall QoE, with the video QoE achieving the lowest rates. Interestingly, the mean value computed between the audio QoE and video QoE seems to be a good indicator for estimating the overall QoE. The ease of communication results highlight that the communication between the two parties was more difficult when the network impairments impacted more on the audiovisual signal. Finally, with regard to the emotional dimensions, the valence results suggest participants felt rather pleasant during the entire experiment, even when the conversation was exposed to quality degradations.

Further studies are needed to investigate the presence of intrinsic biases that could have impacted the results, such as the positive mindset induced by the gaming-based conversational task or the fact that test participants know each other.

REFERENCES

- [1] B. Garcia, F. Gortazar, L. Lopez-Fernandez, M. Gallego, and M. Paris, "WebRTC Testing: Challenges and Practical Solutions," *IEEE Communications Standards Magazine*, vol. 1, no. 2, pp. 36–42, 2017.

- [2] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hofffeld, and P. Tran-Gia, "A Survey on Quality of Experience of HTTP Adaptive Streaming," *IEEE Comm. Surveys Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.
- [3] L. Jalal, M. Anedda, V. Popescu, and M. Murrone, "QoE Assessment for Broadcasting Multi Sensorial Media in Smart Home Scenario," in *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2018, pp. 1–5.
- [4] B. García, F. Gortázar, M. Gallego, and A. Hines, "Assessment of QoE for Video and Audio in WebRTC Applications Using Full-Reference Models," *Electronics*, vol. 9, no. 3, 2020.
- [5] M. Ndiaye, M.-C. Larabi, H. Saadane, G. Le Lay, C. Perrine, C. Quinquis, and L. Gros, "Subjective assessment of the perceived quality of video calling services over a real LTE/4G network," in *2015 7th Int. Workshop on Quality of Multim. Experience (QoMEX)*, 2015, pp. 1–6.
- [6] B. Belmudez, S. Moeller, B. Lewcio, A. Raake, and A. Mehmood, "Audio and video channel impact on perceived audio-visual quality in different interactive contexts," in *2009 IEEE International Workshop on Multimedia Signal Processing*, 2009, pp. 1–5.
- [7] B. Belmudez and S. Moeller, "Audiovisual quality integration for interactive communications," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013, no. 24, 2013.
- [8] F. Brauer, M. S. Ehsan, and G. Kubin, "Subjective evaluation of conversational multimedia quality in IP networks," in *2008 IEEE 10th Workshop on Multimedia Signal Processing*, 2008, pp. 872–876.
- [9] K. De Moor, S. Arndt, D. Ammar, J.-N. Voigt-Antons, A. Perkis, and P. E. Heegaard, "Exploring diverse measures for evaluating QoE in the context of WebRTC," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017, pp. 1–3.
- [10] C. Tsiaras, M. Röscher, and B. Stiller, "VoIP-based calibration of the DQX model," in *2015 IFIP Netw. Conf. (IFIP Networking)*, 2015, pp. 1–9.
- [11] ITU, "Interactive test methods for audiovisual communications." Recommendation ITU-T P.920, 2000.
- [12] —, "The E-model: a computational model for use in transmission planning." Recommendation ITU-T G.107, 2015.
- [13] J. Baraković Husić, S. Baraković, and A. Veispahić, "What factors influence the quality of experience for WebRTC video calls?" in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2017, pp. 428–433.
- [14] D. Ammar, K. De Moor, M. Xie, M. Fiedler, and P. Heegaard, "Video QoE killer and performance statistics in WebRTC-based video communication," in *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*, 2016, pp. 429–436.
- [15] B. Garcia, M. Gallego, F. Gortazar, and A. Bertolino, "Understanding and estimating quality of experience in WebRTC applications," *Computing*, vol. 101, no. 11, pp. 1585–1607, 2019.
- [16] ITU, "One-way transmission time." Rec. ITU-T G.114, 2003.
- [17] D. Tsolkas, E. Liotou, N. Passas, and L. Merakos, "A survey on parametric QoE estimation for popular services," *Journal of Network and Computer Applications*, vol. 77, no. 1, pp. 1–17, 2017.
- [18] G. Carlucci, L. De Cicco, S. Holmer, and S. Mascolo, "Analysis and Design of the Google Congestion Control for Web Real-Time Communication (WebRTC)," in *Proceedings of the 7th International Conference on Multimedia Systems*, ser. MMSys '16, 2016.
- [19] L. Rizzo, "Dummynet: a simple approach to the evaluation of network protocols," *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 1, pp. 31–41, 1997.
- [20] S. Hemminger, "Network emulation with NetEm," 2005.
- [21] N. Ahmad, A. Wahab, and J. Schormans, "Importance of Cross-Correlation of QoS metrics in Network Emulators to Evaluate QoE of Video Streaming Applications," in *2020 11th International Conference on Network of the Future (NoF)*, 2020, pp. 43–47.
- [22] I. Homoliak, M. Teknos, M. Barabas, and P. Hanacek, "Exploitation of netem utility for non-payload-based obfuscation techniques improving network anomaly detection," in *International Conference on Security and Privacy in Communication Systems*. Springer, 2016, pp. 770–773.
- [23] ITU, "Effect of delays on telemeeting quality." Recommendation ITU-T P.1305, 2016.
- [24] —, "Methods for subjective determination of transmission quality." Recommendation ITU-T P.800, 1996.
- [25] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49 – 59, 1994.
- [26] ITU, "Methodologies for the subjective assessment of the quality of television images." Recommendation ITU-R BT.500-14, 2019.

- [27] B. Jansen, T. Goodwin, V. Gupta, F. Kuipers, and G. Zussman, "Performance Evaluation of WebRTC-Based Video Conferencing," *SIGMETRICS Perform. Eval. Rev.*, vol. 45, no. 3, p. 56–68, Mar. 2018.