

“Hello everyone”: Realistic Crowdsourcing-based QoE Assessment of Multi-party Video Conferencing

Fabian Poignée*, Frank Loh*, Michael Seufert†, Tobias Hoßfeld*

* *University of Würzburg, Chair of Communication Networks,*

Würzburg, Germany, {fabian.poignee|frank.loh|tobias.hossfeld}@uni-wuerzburg.de

† *University of Augsburg, Chair of Networked Systems and Communication Networks,*
Augsburg, Germany, michael.seufert@uni-a.de

Abstract—Video conferencing has become indispensable in both professional and social contexts. From a networking point of view, timely distribution of media to all clients participating in a video conference is required. However, fluctuating network conditions pose a challenge, so that a high Quality of Experience (QoE) might not always be achieved. When assessing how network conditions impact the QoE of video conferencing, previous QoE studies focused on one-to-one settings and suffered from small numbers of participants due to their costly laboratory setups. In this work, we investigate crowdsourcing-based QoE studies for video conferencing in order to obtain a higher number and more globally distributed and diverse participants. For this, we consider multi-party video conferencing and design a web-based study framework which provides a high realism to online study participants by involving themselves in the presented video call. Finally, we conduct a QoE study with our framework to demonstrate the suitability of crowdsourcing for assessing the QoE impact of network-level and application-level degradations of video conferencing applications.

Index Terms—video conferencing, quality of experience, online user study, crowdsourcing, network-level impairment, application-level degradation

I. INTRODUCTION

Video conferencing has become an integral part of many people’s lives. Especially since the COVID-19 pandemic, it has proven to be an effective way to maintain connections in social settings and has become a regular feature in educational and work environments in a time where the concept of working from home is becoming ever more popular. The main feature of video conferencing tools, such as Zoom, Google Meet, Microsoft Teams, or Skype, is to enable audiovisual communication over the Internet. After the pandemic, Zoom is used by 300 million daily meeting participants [1], whereas the number of daily active Microsoft Teams users has more than doubled since 2021, reaching 320 million in 2024 [2].

As the user demand for video conferencing applications (VCAs) increases, both video conferencing service providers and network providers are keen on maintaining a satisfactory Quality of Experience (QoE) [3] for the users to avoid customer churn. However, delivering audio and video in time, while meeting increasing user demands on media quality, is challenging. For each user, content must be transmitted to every other member of a video conferencing session, burdening the network which, especially in mobile access networks,

often suffers from fluctuating network conditions [4]. There is already a corpus on video conferencing QoE assessment, e.g., investigating the impact of network parameters [5]. However, these studies have mainly investigated one-to-one setups, while existing tools also support group communication setups of multiple persons. This adds a new dimension, as group communication is different from traditional one-to-one communication [6] and delivering information to all participants burdens the network even more. A hindering effect of video conferencing QoE research is the time and cost of laboratory studies. Therefore, we investigate crowdsourcing as an option to obtain more general and representative results by increasing the number and diversity of participants via online studies.

A substantial challenge in this context is that video conferencing is interactive. In contrast, crowdsourcing study design requires a highly controlled and constrained environment which benefits from prerecorded videos. As such, one of our key considerations to ensure the validity of our results is realism. This means that participants should feel involved in the conversation, as if they were actually a part of it and are motivated to engage in the context of the conversation, even though it is a passive QoE assessment test. Therefore, we present a framework for generating authentic video conferencing stimuli by prerecording multi-party conversations, streaming them through a real video conferencing system under controlled network conditions, and recording the received feed. To enhance realism, we integrate the participant’s live webcam feed into the recorded conference using a modified web-based framework for behavioral experiments. To this end, we answer the following research questions in this work:

- RQ1:** Are online crowdsourcing tasks suited for video conferencing QoE assessment?
- RQ2:** How do different network-level and application-level impairments influence the video conferencing QoE?

For this, the remainder of this work is structured as follows. Section II covers QoE background, methodology on QoE studies, and related work on video conferencing QoE assessment. Section III describes our framework and study design. Results with respect to our research questions are presented in Section IV. We conclude in Section V and give an outlook on future work.

II. BACKGROUND AND RELATED WORK

In the following, we provide necessary technical background on crowdsourcing study design and video conferencing architectures. Furthermore, we discuss network-related video conferencing QoE factors and mitigation strategies before presenting existing works on video conferencing QoE assessment.

A. Crowdsourcing Study Design

For web-based multimedia services and applications, the subjective experience of the end user is highly important for network and service providers. As such, they are interested in quantifying the QoE of users. Guidelines for collecting QoE ratings via subjective user studies are available, e.g., in [7]–[10]. While lab studies are time and resource extensive tasks, crowdsourcing and remote studies provide the opportunity to investigate larger and more diverse participant groups on an international scale [11]–[13]. An essential challenge associated with these studies is the lack of supervision of online and remote participants, necessitating attention during the design phase and result evaluation [14,15]. Consequently, unreliable participants need to be identified and their data needs to be excluded from the results. Usually, crowdsourcing task design is limited to the rating of passive media consumption experiences or interaction with an application. Assessing video conferencing via crowdsourcing adds additional complexity since video conferences include interaction between multiple people, but is still deemed possible by the authors of [14].

1) *Video Conferencing Architecture*: This complexity carries over to the video conferencing infrastructure and application load, since more than one client and one application server are involved. While some applications adopt a peer-to-peer (P2P) setup to directly transmit media information among users, this approach has limitations. It can restrict the number of participants without additional infrastructure, as clients are responsible for delivering content to all others. For instance, Skype transitioned from a P2P architecture to a centralized server-client structure [16] to minimize client responsibilities and shift distribution load to cloud servers. Zoom, employs P2P connections for meetings with two participants and a client-server structure for larger meetings [17]. Resourceful multimedia servers are responsible for hosting the meetings and handling the distribution of voice and video content between all participants. Today, the open-source protocol WebRTC is widely used for video conferencing data transmission. However, proprietary software such as Zoom often relies on their own protocols [17]. An overview of real-time communication applications, including Microsoft Teams, Webex, GoTo Meeting, Zoom, Jitsi Meet, as well as their architectures and protocols is provided by [18].

2) *Influence of the Network on Video Conferencing QoE*: In comparison to, for example, video streaming, video conferencing platforms face distinct challenges. Each conference participant receives information from multiple sources, necessitating timely processing and replication of information to all clients. Due to the demand for interaction, low delay is crucial, leading to small media buffers that struggle with changing

network conditions [19]. Various strategies address different network degradations and their impact on QoE. For example, forward error correction (FEC) offers resilience against packet loss at the cost of additional bandwidth consumption [20].

However, sustained low bandwidth forces quality adjustments, as media encoding rates must stay below available network bandwidth. Adaptive methods lower video bit rate or frame rate [20], or adjust playback speed [19] to handle poor short-term network conditions. More drastic approaches prioritize audio over video due to its lower bandwidth requirements [21]. All of these methods try to ensure an uninterrupted communication feed for the user and are thus less detrimental to the QoE than running into a completely empty media buffer where both audio and video connection to the conference are lost. During such a stalling event, the user loses all information of the video conference and is reconnected to the live state of the video conference once the network conditions stabilize.

B. Studies on Video Conferencing QoE

There has been prior research on factors influencing video conferencing QoE and strategies for mitigation. A comprehensive survey about telemeetings and video conferencing has been conducted by [22], evaluating quality influence factors and QoE-relevant sensory and cognitive processes that go beyond perception and, e.g., include context influence factors, internal references, attention, and expectation. Moreover, the trade-off between CO₂ emission and video conferencing QoE has been explored in [23], emphasizing the need for greener networks while providing awareness and eco-friendly options for users. [24] developed a set of interactivity QoE metrics for video conferencing and a model which aims to map packet traces to those metrics. Furthermore, packet loss, jitter, and delay is investigated in a laboratory with 20 participants in [5], showing that high packet loss rates have a detrimental impact on QoE, especially when paired with delay or jitter. However, laboratory studies often suffer from a low number of participants and are conducted in a one-to-one conference setup. For example, [25] had 41 participants and achieved contrasting results while aiming to reproduce results from [26]. While the studies are years apart and [25] had a focus on the influence of the COVID pandemic, in general, study results benefit from more participants. Consequently, other works estimate video conferencing QoE via machine learning and facial expressions to allow for more scalable analyses [27]. Another option to increase the number of participants with manageable effort and expense is online crowdsourcing, as we explore in this work. A video data set for video conferencing quality assessment is presented in [28]. It includes videos for one-to-one conversation and follows a similar research direction as this work. However, we focus on setups with more than two persons which have different characteristics, e.g., [29] found the impact of delay is less strong in multi-party setups. Multi-party video conferencing via online crowdsourcing is investigated by [30], but show fully prerecorded conferences to the participants, so ratings are generated from an outside perspective. Since videoconferencing is an active experience,

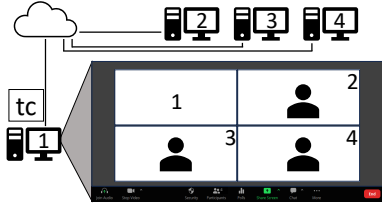


Fig. 1: Overview of video generation setup.

we need to assess the QoE of the involved participants rather than outside observers. Consequently, to the best of our knowledge, we are the first to do this with a realistic setup in which participants feel involved in the call. In the following, we present our study framework and investigate the influence of different application degradation scenarios in a multi-party setup via an online crowdsourcing study.

III. STUDY FRAMEWORK AND STUDY DESCRIPTION

To answer our research questions, we need a standardized experimental environment where every participant experiences the same video conferencing situation under fully controlled conditions. Thus, in the following, we describe our approach for generating realistic video conferencing QoE studies and provide details on the crowdsourcing study setup.

A. Framework for Realistic Video Conferencing QoE Studies

To ensure the same conditions for each participant, we prerecord four conversational scripts, each with three actors. To mitigate the lack of a real interaction with the participant, the scripts are designed to address the entire group during the conversation, including open questions and statements addressing everyone in the room, ensuring that the study participant, as the fourth person, would feel as part of the conversation without actively taking the word. The scripts cover exemplary social settings: a vacation trip, a work meeting, solving a technical issue, and planning a birthday party. We record video and audio of the actors via webcams and microphones during the conversations which last from 44 s to 67 s.

To generate authentic video conferencing experiences impacted by real network conditions, we use our recorded media as input to Zoom video conferencing sessions. We manipulate network parameters during the session while recording the resulting video conference. Note that, although we use Zoom in this work, our approach is agnostic to the VCA. More specifically, we employ four distinct PCs as shown in Figure 1. PCs 2–4 use the prerecorded conversations as input to the video and audio feed of their Zoom client, thereby impersonating one of the three actors each. PC 1 represents the future QoE study participant and feeds placeholder content to the conference. The main objective of PC 1 is to record the received Zoom conference. We crop the recording, which has a grid layout (see Figure 1), to obtain the received audio and video from PCs 2–4 from the perspective of PC 1.

We now manipulate the bandwidth of PC 1 via the traffic shaping tool tc [31] to trigger Zoom’s native mitigation be-

havior. This allows to record the resulting audiovisual quality of the received Zoom video conference for different network conditions, which we subsequently use in our QoE study. Additionally, or as an alternative to emulating network conditions, we can generate application-level impairments, such as, for example, stalling or resolution changes, directly into the recorded videos using video editing software.

To assess the resulting QoE of network- and application-level impairments in our resulting video conferencing sessions, we conduct a crowdsourcing QoE study. Therefore, we replicate the video conferencing layout in Figure 1 on a web page and show it to our study participants. We playback the recorded stimuli at the positions 2–4. The placeholder at position 1 in the top left is replaced with the live webcam feed of the study participant, so the participants can see themselves, to allow for a realistic video conferencing experience.

B. Study Description

The study was conducted on the Microworkers platform and accessible via a web browser. Our web application extends the JSPsych framework [32] for behavioral experiments using best practices from literature [13]–[15,33] and other features, such as a check for an enabled webcam and video playback manipulation. Throughout the study, user interaction with the website is logged via the framework to allow filtering of unreliable users which, for example, are too fast during the study and are assumed to not have read the instructions and questionnaires carefully. It consists of three major parts, as shown in Figure 2 and described in the following.

1) Study Preparation and Demographic Questionnaire:

When the study website is first accessed, all media is preloaded to prevent unplanned impairments during the execution of the experiment which could distort the results. Then, after an introduction and explanation of the rules of the experiment, users are asked to provide basic demographic information, i.e., gender, age, continent of origin, profession, and education. To ensure a reliable test environment, users need to enter the correct number of symbols from two contrast-evaluation images, verifying their screen settings. Additionally, participants solve an audio CAPTCHA to confirm audio is enabled. Furthermore, it is essential for a realistic experience that study participants enable their webcam. Consequently, there is a webcam check with instructions on how to enable the webcam in the browser. Before the final instructions on the main part of the study are given, users are informed that no footage of them is recorded.

2) Study Task and QoE Questionnaire:

The goal of the study is to investigate the impact of different network- and application-level impairments perceived by the end users. When Zoom adapts to poor network conditions, first the bit rate and playback rates are reduced, which results in audio glitches and slight video distortions. If the poor network conditions persist, ultimately stalling will occur. Note that in contrast to video streaming, the content during the playback interruption in a video conferencing is skipped, and thus, causing participants to miss a part of the conversation. For the QoE study, we selected a baseline scenario with a

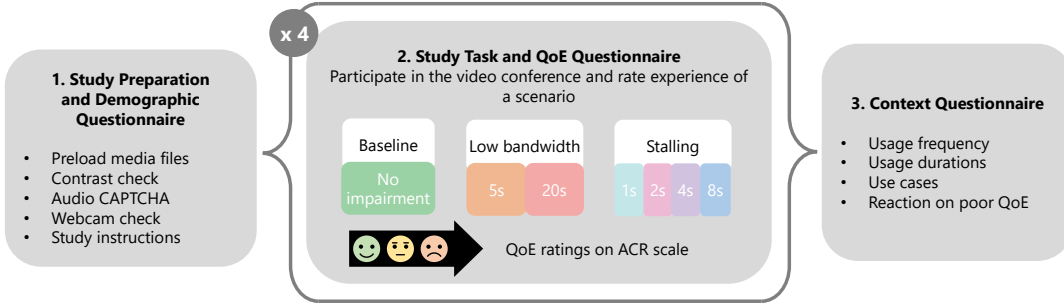


Fig. 2: Structure of the crowdsourcing study.

constant bandwidth of 50 Mbit/s, in which no impairment occurs, and compare it to two impairment types. First, we created a scenario with glitches, that could be obtained by a bandwidth drop from 50 Mbit/s to 1 Mbit/s after 20 s into the conference. Here, the bandwidth was reduced for 5 s before going back to 50 Mbit/s. Second, we trigger a stalling event where about 4 s of the conversation are lost, which could be obtained by limiting the bandwidth for 20 s instead. We name those scenarios BW5 and BW20 accordingly. Note, that these effects are the real result of the Zoom client in response to our bandwidth reductions. The selected bandwidth settings are a result of a prestudy aiming to force different impairments in order to investigate them in the study. We also investigate application-layer impairments, namely, scenarios with artificially introduced stalling events (incl. skipping of content) of different lengths of 1 s, 2 s, 4 s, and 8 s, which we name accordingly S1 to S8. The main part of the study includes four separate video conferencing sessions in which participants are asked to rate their experience. The scenarios are shuffled across sessions. While the baseline scenario occurs exactly once, other impairments occur at most once for each user. After each session, participants are asked to rate their overall experience and their immersion with the video conference on a 5-point Absolute Category Rating scale [34].

3) *Context Questionnaire*: Finally, participants answer context questions about the frequency, length, and setting (e.g., private or work environment) of their typical video conferencing use, as well as their behavior when experiencing poor QoE. At the end, they receive a Microworkers reimbursement code and a small monetary reward for their participation.

4) *Filtering of Unreliable Participants*: To determine participants who did not properly complete the task, the framework’s logs and reliability checks were assessed to exclude their results from the evaluation. If the average response time of a participant to questions was below 2 s we disregard their results since they did not read carefully. If more than four trials were necessary to pass the contrast checks, we exclude those participants from the results. We also included a study-related reliability check. After rating each session, a question about the conversation topic was asked to ensure the participant paid attention to the task. If two or more content answers were incorrect, the participant is removed from the result. If only one answer was incorrect, only this run is excluded from

TABLE I: Distribution of participants per setting.

	Baseline	BW 5s	20s	Stalling 1s	2s	4s	8s	Σ
#Participants	338	157	157	175	175	161	172	1,335

analysis. Finally, if a participant gave the same QoE rating across the whole study, their results are also excluded.

IV. RESULTS

In the following we analyze the collected answers to answer our research questions. First, the data set is described. Afterwards, we investigate the immersion ratings to answer **RQ1**. Then, to answer **RQ2**, we assess the video conferencing QoE.

A. Data Set Description

For our study, which took less than 10 minutes to complete, we enlisted a total of 483 participants. After filtering out unreliable participants, 374 participants remain. Of those, 78.88 % identify as male, the rest as female, none as other. The majority of our participants is young, 55.08 % are between 20 and 29 years old. 59.36 % live in Asia, and 18.18 % in Africa, which is in line with the global population distribution [35]. Europe is slightly overrepresented with 13.10 % compared to 9.29 % of the global population. North America and Oceania amount to 4.01 % of our data. Surprisingly, more than 80 % of the participants report that they have at least a Bachelor’s degree. Around three quarters (74.87 %) use video conferencing at least several times a month, and 37.17 % report spending on average more than 5 h per week in video conferences.

Asked for which purpose they mainly use VCAs (multiple options could be ticked), 64.44 % of the participants answered that they use it for work meetings. In addition, 44.92 % and 44.36 % use VCAs for virtual classrooms and meeting friends, respectively. Finally, 4.81 % use it for other purposes. When facing extremely poor QoE during a video conference, 57.22 % reported to try and restart the VCA, while 29.41 % also consider turning off the video camera, showing that they have a general understanding about VCA bandwidth requirements. Table I shows the number of participants for each scenario. As described in Section III, each participant experienced at least once a scenario without impairments, thus, this setting has significantly more data points.

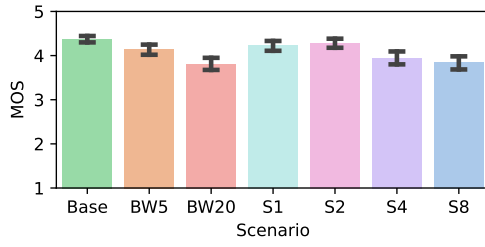


Fig. 3: Video conferencing MOS for each scenario.

B. Realism Analysis

Before we investigate the video conferencing QoE, we first have to validate that crowdsourcing is a sensible approach to assess video conferencing QoE. Therefore, we assess the baseline scenario where the participants were asked whether they found the video conference immersive. A MOS of 3.43 and a Good-or-Better (GoB) ratio of 52.37% is achieved. Thus, more than half of the participants were pleased with the realism of the setup. When considering all scenarios, the results are only slightly lower, with a MOS of 3.38 and a GoB of 49.89%. So while it appears that impairments may have a slightly negative impact on the immersion rating, none of these findings prove significant for any impairment. Consequently, we can answer our first research question positively, as with our study design which stresses on realism and conversational engagement, results show that online crowdsourcing tasks are suited for video conferencing QoE assessment.

C. Influence of Impairments on Video Conferencing QoE

To investigate the influence of the different impairments on video conferencing QoE, Figure 3 is presented. The figure shows the MOS values and the corresponding 95% confidence intervals for each scenario listed on the x-axis. The highest MOS value of 4.37 is achieved for the baseline scenario which is to be expected. Overall, participants were happy with their video conferencing experience. For the low bandwidth scenarios BW5 in orange and BW20 in red, we see a significant degradation of the QoE, since the confidence intervals neither overlap with the base scenario nor with each other. Whereas the 5s impairment results in a MOS of 4.14, the 20s impairment achieves the lowest result of all scenarios at a MOS of 3.81. For the stalling scenarios S1, shown in cyan, and S2 in pink, we obtain MOS values of 4.22 and 4.28, respectively. While they are slightly lower than the baseline, no significant differences in the MOS are observed. This and other results were confirmed with dedicated t-tests for all possible paired comparisons. We further apply Bonferroni correction to account for the multiple-comparisons problem and $p < 0.05$. We believe that contextual information of the video conference is important to understand the result, as we observe slight differences between the different video conferencing contents for the same impairment scenario. Human communication includes pauses, e.g., during speaker changes or mid-speech. So, while during one conversation, a short impairment might

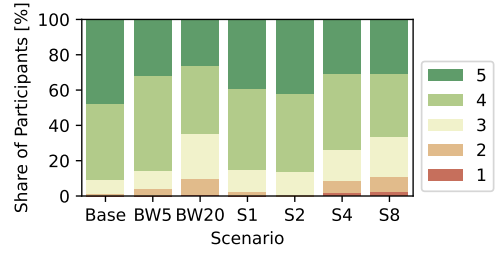


Fig. 4: Distribution of the video conferencing QoE ratings.

be hard to notice at all, in another conversation even a short impairment can make it difficult to understand the speaker and follow the conversation. A 4s or 8s stalling is guaranteed to be noticed by participants for all presented contents. Thus, the MOS for S4 (purple) at 3.94 and for S8 (blue) at 3.84 are both significantly lower than the baseline. However, there is no significant difference between S4 and S8. Interestingly, although the impairment duration of BW20 against S8 more than doubles (20s compared to 8s), they show comparable results. The reason is that during scenario S8, the conference connection suddenly stops and all information on the conversation is missing for a long time (8s), whereas in the BW20 scenario the quality constantly degrades while participants can still follow the speaker for a while before only some part (around 4s) of the conversation is lost. So the contribution of Zoom's low bandwidth mitigating strategy is two-fold. First, playback is stretched to allow for more conversational information to be transmitted to the receiver before a complete cut happens. Second, the interruption is not as abrupt, possibly decreasing the negative impact on the video conferencing QoE. The hypothesis is further supported by Figure 4 which shows the distribution of the QoE ratings for each scenario. As defined by the ACR rating scale [34] category 1 corresponds to bad experience, contrasted by category 5 which represents excellent experience. The only scenarios with ratings of category 1, shown in red, are S4 and S8. For the baseline scenario, most ratings are light or dark green, representing a good or excellent experience and a GoB of 90.83%. For BW5, this drops to a GoB of 85.99%. In the figure, we see less excellent ratings and more good ratings compared to the baseline which is further shifted towards fair ratings for BW20 which achieves a GoB of 64.97%. While ratings for S1 and S2 are similar, the difference in fair and poor ratings of which S8 has more.

To summarize, Zoom's mitigation strategy for low bandwidth can noticeably decrease the negative impact on QoE in poor network situations. Even our worst impairments yield a MOS sufficient for video conferencing. However, as our impairments are single, relatively short events compared to the session length, repeated or prolonged impairments are likely to further shift results towards an unacceptable experience.

D. Stalling in Video Conferencing and in Video Streaming

Finally, the question remains how our results on stalling in video conferencing align with existing crowdsourcing stalling

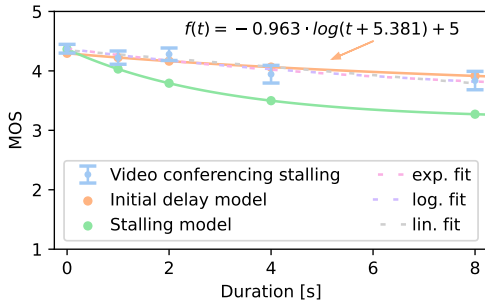


Fig. 5: Comparison of video conferencing stalling QoE to video streaming delay models from [36].

models. For this, we first fit models based on the presented video conferencing QoE ratings before comparing our results to existing crowdsourcing models.

Usually in QoE research about delays or interruptions, the relationships are modeled as logarithmic, according to the WQL hypothesis [37] based on the fundamental Weber-Fechner law from psychophysics, or as exponential, according to the IQX hypothesis [38]. Rarer are linear relationships [39]. We fit all function types on our data and plot them in Figure 5 in pink, purple and grey respectively. Our study results for the 1 s to 8 s scenarios S1 to S8, and additionally, for the baseline scenario with a stalling length of 0 s, are depicted by the error bars in blue. We obtain for the logarithmic fit $f_{log}(t) = a \cdot \log(t+b) + c$ the parameters $a = -1.28$, $b = 4.67$, and $c = 5.23$, and an R^2 of 0.903. For the exponential fit $f_{exp}(t) = a \cdot e^{b \cdot t} + c$, we obtain $a = 0.81$, $b = -0.14$, and $c = 3.56$, and an R^2 of 0.907. The linear fit $f_{lin}(t) = a \cdot t + b$ with $a = -0.07$, and $b = 4.35$ and an R^2 of 0.874 falls behind. While all fit our data well, $f_{exp}(t)$ fits best.

In Figure 5, we additionally compare our findings to the results of [36] because in comparison to others [40,41], their setup is closest to ours, i.e., crowdsourcing with 60 s session length with a single stalling event and multiple stalling lengths. However, in contrast to us, they investigated streaming, in particular stalling and initial delay. The video streaming stalling model from [36] is depicted in green, and their model for initial delay is shown in orange. While all models start at a similar MOS for no delay, the video streaming stalling model differs largely from our findings. Surprisingly, the figure shows a good fit of the initial delay model to our data. For a stalling length of 1 s, we obtain a MOS of 4.22 from the initial delay model, which is within the confidence intervals of scenario S1. For the other stalling lengths the model does not lie within the confidence intervals. Nonetheless, it is clear that the QoE is less affected by stalling events in video conferencing than in video streaming. A reason could be that for video streaming, users want to enjoy the audiovisual content, while in video conferencing nonfunctional and contextual factors regarding the ability to follow the conversation are more important.

In summary, we see that stalling in video conferencing does not have the exact same effect on the QoE as stalling in

video streaming, requiring dedicated QoE studies for video conferencing in the future. For this, our presented study framework can be used.

V. CONCLUSION

Since video conferencing has established itself in our daily lives, high Quality of Experience is a concern for network and application providers. While research exists, including research on the impact of different network conditions, there is a focus on one-to-one setups. Investigation of video conferencing QoE in group communication setups via laboratory studies is an even more time and cost consuming task, given the required participants. To address this gap and facilitate future video conferencing QoE research, we explored the potential of online crowdsourcing studies. For this, we developed a methodology to investigate the QoE impact of network- and application-level impairments with a focus on realism. We designed and implemented a framework for generating authentic video conferencing stimuli by prerecording conversations, streaming those conversations through a real video conferencing system while manipulating network conditions, and recording the received video conference. We further modified an existing web-based framework for behavioral experiments to display not only the recorded authentic video conference content, but also the live feed of the study participant's webcam to cater a realistic video conferencing experience.

Results showed that the participants of the conducted crowdsourcing study were involved in the video conference, i.e., felt as a real part of the conversation, allowing for valid insights from a large and diverse participant pool. Thus, we demonstrated that video conferencing QoE research can leverage crowdsourcing when following a sophisticated study design. Further, in our showcase study with network-level and application-level impairments we found that existing mitigation techniques of Zoom successfully decrease the negative impact on QoE during poor networking situations compared to artificially introduced stalling of similar lengths. Comparative analysis with video streaming investigations of [36] show that stalling during video conferences and in video streaming are not the same, while their model for initial delay in video streaming fits our data surprisingly well. Therefore, new dedicated studies to investigate video conferencing QoE are required, which can be developed using our presented framework. In future work, we will extend our study to include other aspects of video conferencing, such as usability, and to dive into different use-case scenarios such as screen sharing which allows for a more holistic understanding of the QoE of video conferencing.

ACKNOWLEDGEMENT

This work was partly funded by Deutsche Forschungsgemeinschaft (DFG) under grants HO 4770/7-1 and SE 3163/1-1, project number: 442413406, as well as SE 3163/3-1, project number: 500105691. The authors alone are responsible for the content.

REFERENCES

- [1] L. S. Vailshery, "Zoom Daily Meeting Participants Worldwide 2019-2020," 2022, accessed: 2025-03-05. [Online]. Available: <https://www.statista.com/statistics/1253972/zoom-daily-meeting-participants-global/>
- [2] —, "Number of Daily Active Users (DAU) of Microsoft Teams Worldwide as of 2022," 2023, accessed: 2025-03-05. [Online]. Available: <https://www.statista.com/statistics/1033742/worldwide-microsoft-teams-daily-and-monthly-users/>
- [3] *Qualinet White Paper on Definitions of Quality of Experience*. Lausanne, Switzerland: European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), 2012.
- [4] A. Bokani, M. Hassan, S. S. Kanhere, J. Yao, and G. Zhong, "Comprehensive mobile bandwidth traces from vehicular networks," in *Proceedings of the 7th International Conference on Multimedia Systems*, ser. MMSys '16. New York, NY, USA: Association for Computing Machinery, 2016. [Online]. Available: <https://doi.org/10.1145/2910017.2910618>
- [5] G. Bingöl, L. Serreli, S. Porcu, A. Floris, and L. Atzori, "The Impact of Network Impairments on the QoE of WebRTC applications: A Subjective study," in *2022 14th International Conference on Quality of Multimedia Experience (QoMEX)*, 2022, pp. 1–6.
- [6] M. Seufert, T. Höbfeld, A. Schwind, V. Burger, and P. Tran-Gia, "Group-based communication in WhatsApp," in *2016 IFIP Networking Conference (IFIP Networking) and Workshops*, 2016, pp. 536–541.
- [7] A. Raake, *Speech Quality of VoIP: Assessment and Prediction*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.
- [8] International Telecommunication Union, "ITU-T Recommendation G.1011: Reference Guide to Quality of Experience Assessment Methodologies," 2015.
- [9] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives," *Multimedia Syst.*, vol. 22, no. 2, p. 213–227, Mar. 2016. [Online]. Available: <https://doi.org/10.1007/s00530-014-0446-1>
- [10] M. Seufert, "Statistical Methods and Models based on Quality of Experience Distributions," *QUEX*, vol. 6, no. 1, 2021.
- [11] M. Hirth, T. Höbfeld, and P. Tran-Gia, "Anatomy of a Crowdsourcing Platform - Using the Example of Microworkers.com," 2011.
- [12] B. Gardlo, S. Egger, M. Seufert, and R. Schatz, "Crowdsourcing 2.0: Enhancing execution speed and reliability of web-based QoE testing," in *2014 IEEE International Conference on Communications (ICC)*, 2014, pp. 1070–1075.
- [13] S. Egger-Lampl, J. Redi, T. Höbfeld, M. Hirth, S. Möller, B. Naderi, C. Keimel, and D. Saupé, "Crowdsourcing Quality of Experience Experiments," in *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments*. Springer, 2017.
- [14] T. Höbfeld, M. Hirth, J. Redi, F. Mazza, P. Korshunov, B. Naderi, M. Seufert, B. Gardlo, S. Egger, and C. Keimel, "Best Practices and Recommendations for Crowdsourced QoE - Lessons Learned from the Qualinet Task Force Crowdsourcing," 2014.
- [15] F. Daniel, P. Kucherbaev, C. Cappelletto, B. Benatallah, and M. Allahbakhsh, "Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions," *ACM Computing Surveys*, vol. 51, no. 1, 2018.
- [16] S. Chan, "Skype will stop working on Windows Phone 8.1 (and older devices) in early 2017," August 2016, accessed: 2025-03-07. [Online]. Available: <https://mspoweruser.com/skype-will-stop-working-windows-phone-early-2017/>
- [17] O. Michel, S. Sengupta, H. Kim, R. Netravali, and J. Rexford, "Enabling passive measurement of Zoom performance in production networks," in *Proceedings of the 22nd ACM Internet Measurement Conference*, ser. IMC '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 244–260. [Online]. Available: <https://doi.org/10.1145/3517745.3561414>
- [18] A. Nistico, D. Markudova, M. Trevisan, M. Meo, and G. Carofoglio, "A comparative study of RTC applications," in *2020 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2020.
- [19] Z. Wang, Y. Cui, X. Hu, X. Wang, W. T. Ooi, Z. Cao, and Y. Li, "Multilive: Adaptive Bitrate Control for Low-delay Multi-party Interactive Live Streaming," *IEEE/ACM Transactions on Networking*, 2021.
- [20] I. Lee, J. Lee, K. Lee, D. Grunwald, and S. Ha, "Demystifying Commercial Video Conferencing Applications," in *Proceedings of the 29th ACM international conference on multimedia*, 2021.
- [21] "Optimizing Performance of Zoom in Low Bandwidth Environments," 2021, accessed: 2025-03-05. [Online]. Available: <https://assets.zoom.us/docs/user-guides/User%20Guide%20-%20Optimizing%20Performance%20in%20Low%20Bandwidth%20Environments.pdf>
- [22] J. Skowronek, A. Raake, G. H. Berndtsson, O. S. Rummukainen, P. Usai, S. N. B. Gunkel, M. Johanson, E. A. P. Habets, L. Malfait, D. Lindero, and A. Toet, "Quality of Experience in Telemeetings and Videoconferencing: A Comprehensive Survey," *IEEE Access*, 2022.
- [23] T. Höbfeld, M. Varela, L. Skorin-Kapov, and P. E. Heegaard, "What is the Trade-Off Between CO2 Emission and Video-Conferencing QoE?" *ACM SIGMultimedia Records*, vol. 14, no. 1, 2023.
- [24] J. He, M. Ammar, E. Zegura, E. Halepovic, and T. Karagioules, "QoE Metrics for Interactivity in Video Conferencing Applications: Definition and Evaluation Methodology," in *Proceedings of the 15th ACM Multimedia Systems Conference*, 2024.
- [25] C. Diao, L. Sinani, R. R. Ramachandra Rao, and A. Raake, "Revisiting Videoconferencing QoE: Impact of Network Delay and Resolution as Factors for Social Cue Perceptibility," in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, 2023.
- [26] K. Schoenenberg, A. Raake, and P. Lebreton, "Conversational Quality and Visual Interaction of Video-telephony under Synchronous and Asynchronous Transmission Delay," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2014.
- [27] G. Bingöl, S. Porcu, A. Floris, and L. Atzori, "QoE Estimation of WebRTC-based Audiovisual Conversations from Facial Expressions," in *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2022.
- [28] R. P. Spang, J.-N. Voigt-Antons, and S. Möller, "The Story time Dataset: Simulated Videotelephony Clips for Quality Perception Research," in *2022 14th International Conference on Quality of Multimedia Experience (QoMEX)*, 2022.
- [29] M. Schmitt, S. Gunkel, P. Cesar, and D. Bulterman, "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*, 2014, pp. 13–16.
- [30] M. Schmitt, D. C. Bulterman, and P. S. Cesar, "The Contrast Effect: QoE of Mixed Video-Qualities at the Same Time," *Quality and User Experience*, 2018.
- [31] M. Kerrisk, "tc — Linux Manual Page," Accessed: 2025-03-05. [Online]. Available: <https://man7.org/linux/man-pages/man8/tc.8.html>
- [32] J. R. De Leeuw, "jsPsych: A JavaScript library for creating behavioral experiments in a Web browser," *Behavior research methods*, 2015.
- [33] T. Höbfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best Practices for QoE Crowdttesting: QoE Assessment with Crowdsourcing," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 541–558, 2013.
- [34] International Telecommunication Union, "ITU-T Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications," 2008.
- [35] E. H. Dyvik, "Distribution of the Global Population 2023, by Continent," 2024, accessed: 2025-03-05. [Online]. Available: <https://www.statista.com/statistics/237584/distribution-of-the-world-population-by-continent/>
- [36] T. Höbfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea," in *2012 Fourth International Workshop on Quality of Multimedia Experience*. IEEE, 2012.
- [37] S. Egger, P. Reichl, T. Höbfeld, and R. Schatz, "Time is Bandwidth"? Narrowing the Gap Between Subjective Time Perception and Quality of Experience," in *2012 IEEE International Conference on Communications (ICC)*, 2012.
- [38] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A Generic Quantitative Relationship Between Quality of Experience and Quality of Service," *IEEE Network*, 2010.
- [39] N. Wehner, M. Seufert, R. Schatz, and T. Höbfeld, "Do you agree? Contrasting Google's Core Web Vitals and the impact of cookie consent banners with actual web QoE," *Quality and User Experience*, vol. 8, no. 1, p. 5, 2023.
- [40] T. Höbfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, "Quantification of YouTube QoE via Crowdsourcing," in *2011 IEEE International Symposium on Multimedia*. IEEE, 2011.
- [41] T. Höbfeld, R. Schatz, E. Biersack, and L. Plissonneau, "Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience," *Data Traffic Monitoring and Analysis: From Measurement, Classification, and Anomaly Detection to Quality of Experience*, 2013.