



Survey on user perceived system factors influencing the QoE of audiovisual calls on smartphones

Dunja Vučić¹ · Sabina Baraković² · Lea Skorin-Kapov³

Received: 31 March 2022 / Revised: 15 September 2022 / Accepted: 27 October 2022 /
Published online: 30 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

With the widespread use of applications and services supporting audiovisual calls via smartphones, both in business and leisure contexts, a key challenge for service providers is meeting end user Quality of Experience (QoE) expectations and requirements. To successfully meet this challenge, there is a need to identify and analyze the key system-related factors impacting user perceived quality. In this paper, we contribute beyond state-of-the-art by conducting a large scale web-based questionnaire survey to investigate the system-related factors that subjects identify as most influential in contributing to their overall experience and quality perception. We focus in particular on leisure audiovisual calls, established via mobile devices. Our initial survey (Phase 1) was conducted in Feb. 2020, just prior to the outbreak of the COVID-19 pandemic (272 participants). To investigate if the importance of factors has changed due to increased usage of the service caused by the pandemic among the general population, we conducted a second survey (Phase 2) in October 2021 with 249 participants. Based on obtained results, we identify key system-related QoE influence factors belonging to three categories: media quality, functional support, and usability and service design. We observe no significant differences in user opinions and expectations prior to and during the period of increased service usage, despite different participant demographics and study time frames, thus contributing to generalizability of obtained results. Study results contribute to providing insights for designing future user studies investigating QoE, in terms of key factors that should be considered.

Sabina Baraković and Lea Skorin-Kapov contributed equally to this work.

✉ Dunja Vučić
dunja.vucic@ericsson.com

Sabina Baraković
sabina.barakovic@fsk.unsa.ba

Lea Skorin-Kapov
Lea.Skorin-Kapov@fer.hr

¹ Ericsson Nikola Tesla d.d., Krapinska 45, Zagreb, 10000, Croatia

² Faculty of Traffic and Communications, University of Sarajevo, Zmaja od Bosne 8, Sarajevo, 71000, Bosnia and Herzegovina

³ University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, Zagreb, 10000, Croatia

Keywords QoE · Influence factors · Audiovisual calls · Smartphones · User perception

1 Introduction

In the past decade, video transmission over the Internet has experienced significant rise, enabled by technological advancements such as higher network transmission rates, improved video coding capabilities, and the widespread availability of high quality displays, cameras, speakers, and microphones on heterogeneous end user devices. Mobile devices, services, and applications have become an inseparable part of our daily lives, affecting relationships, social norms, communication, and interaction methods, even before global outbreak of the COVID-19 pandemic. Trends in the increasing use of audiovisual communication services, both in business and private contexts, have stemmed from evolving life dynamics and accelerated lifestyles, as well as recent distancing and lockdown measures [14]. A recent Sandvine report [39] highlighted the impact of the pandemic on dramatic increases in traffic corresponding to applications supporting video telephony such as Zoom and MS Teams from mid-March 2020 onward. While the global video conferencing market size in 2018 was USD 3.02 billion, estimated growth by 2026 was set to USD 6.37 billion [42].

Given end user needs and expectations, modern video conferencing (or telemeeting) services are expected to be reliable and available across heterogeneous access networks, devices, and usage contexts, with underlying platforms and protocols secure and easy to manage. The term *telemeeting* is defined by ITU-T Recommendation P.1301 as a meeting in which participants are located in at least two different locations and the communication takes place via a telecommunication system [21]. Telemeetings held in a private/leisure context have the primary objective of experiencing a sense of presence or social connection. Considering that the term *audiovisual call* is commonly used among the general population and is usually associated with the leisure context, we have opted to use the term audiovisual call instead of telemeeting in our study.

Technologies such as WebRTC (Web Real-Time Communications) have contributed to making many video conferencing services free and available to the wider public. Regardless of the system complexity, the service itself should be simple and participants should be able to use it without intense training. Features and functions must be useful and offer seamless access, both for when used in a business and leisure context. Video conferencing used in a business context generally has a specific objective, with a set of tasks that must be completed [36]. On the other hand, audiovisual calls used in the private/leisure context generally have the primary objective to experience a sense of presence or social connection. Due to the different objectives of the meeting or call, the quality expected by the participants may be different, with participants likely being less critical when it comes to the private context [18, 47].

Going beyond conversational services between two participants, users are increasingly using video conferencing/call services in both business and leisure contexts (e.g., social interactions via Skype, Viber, Whatsapp, Google Meet, Zoom, Microsoft Teams, Whereby, etc.). Such audiovisual settings impose a wide range of challenges with respect to identifying and quantifying the impact of various factors influencing end user Quality of Experience (QoE), in particular in the context of calls established via mobile devices. In [37], the authors summarize the challenges in properly assessing the QoE of such systems, and highlight mobility aspects, device and encoding interoperability, ease of use, and additional collaboration possibilities (e.g., exchanging pictures, files, chatting).

With the processing power of mobile devices such as smartphones and tablets becoming sufficient to simultaneously encode and decode video at a high spatial and temporal resolution during real-time communication, mobile video communication service use has grown rapidly [42]. A wide range of smartphone models available on the market, along with heterogeneous access networks, can create numerous different asymmetric scenarios, with video calls imposing strict low latency and high volume requirements on the underlying network. Service architectures, such as those relying on a centralized Selective Forwarding Unit (SFU) or Multipoint Control Unit (MCU) are thus commonly deployed to optimize resource utilization and ensure high service quality, in particular in situations with a large number of simultaneous users.

Designing and managing video conferencing services requires an understanding of the key underlying QoE influence factors. One of the key challenges faced by service providers lies in configuring the video encoding parameters so as to maximize participant QoE while meeting resource (network and mobile device) availability constraints. Currently developed QoE models can for the most part be applied to two interlocutors and in desktop environments. However, there is a lack of studies that focus on modeling and optimizing QoE for such services when using mobile devices. In our previous work, we have conducted numerous subjective user studies involving three-party video calls established via smartphone devices in both laboratory and field settings, with the aim being to study the impact of different video encoding parameters (encoding bitrate, resolution, and frame rate) on user perceived QoE [43–46].

In this paper, we aim to complement our earlier work and provide insights into designing future user studies by providing an in-depth investigation of users' opinions and expectations related to audiovisual calls on mobile devices, focusing on the leisure/private context. We conducted an extensive web-based questionnaire survey to investigate the system-related factors that subjects identify as most influential in contributing to their overall experience and quality perception. We highlight that we make complete survey results (anonymized) publicly available to the research community to foster reproducible research (link: <https://muexlab.fer.hr/muexlab/research/datasets>).

Our aim is not to quantify the impact of certain QoS factors on QoE (as has been done in a number of previous empirical studies), but rather to obtain up-to-date feedback from a large number of users on their opinions with respect to a wide range of potential QoE system influence factors. Therefore, our study is different in both focus and scope than previous dedicated studies focused on modeling QoE or MOS based on a limited set of chosen factors. We stress that our approach fills the gap of existing studies (especially those focusing on the mobile context) which do not explicitly ask users for their opinions on whether certain influence factors should be examined further, but rather focus on evaluating the impact of a limited set of previously chosen factors on QoE. Aiming to obtain insights into a wider range of potential factors that need to be addressed, we asked users which system-related factors are important to them in audiovisual calls on smartphones. Our aim is to obtain feedback that can be used as input for designing future user studies investigating QoE, whereby our results provide novel contributions in terms of input on what are the key QoE system influence factors that should be considered in future studies.

Our study was conducted in two phases, so as to further investigate differences in user opinions potentially triggered by the global outbreak of the COVID-19 pandemic. **Phase 1** included 272 participants and was conducted in February 2020, just prior to the global outbreak of the pandemic. Given the drastic increase in video communication services resulting from lockdown measures [39], we repeated the survey in **Phase 2 (October 2021)** which included 249 participants.

We address the following research questions:

- RQ1: What do users consider to be the most important system influence factors in terms of their importance and impact on QoE in the context of audiovisual calls established for leisure purposes on smartphones?
- RQ2: Are there significant differences in user opinions when comparing survey results reported prior to global outbreak of the pandemic and results obtained via an independent survey conducted 20 months into the pandemic?

The paper is organized as follows: Section 2 gives an overview of standards and related work, focusing on system-related QoE influence factors as pertaining to audiovisual calls established via smartphones. Our research methodology is described in Section 3, providing an overview of survey items and participant demographics. Results across both conducted surveys (Phase 1 and Phase 2) are analyzed in Section 4, summarizing opinions related to media quality, functional support of the service, and usability, service design, and resource consumption. The key influence factors, derived based on user opinions and analyzed results, are outlined in Section 5. Finally, Section 6 provides concluding remarks and an outlook for future research challenges in this area. The questionnaire used in both surveys is given in [Appendix](#).

2 Related work

International bodies including ITU-T and ETSI have specified definitions of QoE over the years. However, to establish definitions and methods for the quantitative assessment of QoE for multimedia content and services in a given situation and system configuration, such as audiovisual calls on smartphones, the COST IC 1003 Action Qualinet defines QoE as: *“the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user’s personality and current state”* [27].

The multidimensional nature of QoE stems from a number of different influence factors and perceived features comprising the overall QoE. The Qualinet white paper further defines an influence factor (IF) as *“any characteristic of a user, system, service, application, or context whose actual state or setting may have influence on the Quality of Experience for the user”* and groups them into three categories [23, 24, 27]: 1) Human IFs (HIF) as any variant or invariant property or characteristic of a human user. The characteristic can describe the demographic and socio-economic background, the physical and mental constitution, or the user’s emotional state; 2) Context IFs (CIF) factors that embrace any situational property to describe the user’s environment in terms of physical, temporal, social, economic, task, and technical characteristics; and 3) System IFs (SIF) referring to the properties and characteristics that determine the technically produced quality of an application or service. They are related to media capture, coding, transmission, storage, rendering, and reproduction/display, as well as to the communication of information itself from content production to user. Usually, they are grouped into network, media, content, and device categories.

Factors affecting the QoE of different components of audiovisual conferences/calls or the service as a whole have been an increasing topic of interest to the research community before pandemic. Those research studies report the existence of the impacts of various human, context, or system IFs on QoE. However, most studies have focused on the impact of various system factors, while context and human factors have been addressed to a limited extent.

Traditionally, addressed network related SIFs in terms of QoE are (wireless) channel characteristics and capacity (coverage, bandwidth, etc.), signal strength, and transmission impairments (delay, jitter, loss, etc.) [1, 4, 6, 9–13, 16, 26, 28, 29, 33, 34, 40, 41, 48]. Media and content related SIFs addressed so far cover content type and quality, synchronization, audiovisual features and quality, spatial and temporal artifacts [2, 7, 15, 19, 22, 30, 31, 35, 49]. In terms of application and device related SIFs, modern mobile users are looking to access and reliably utilize demanding services regardless of context or system influence factors, such as location, time, network conditions, service topology, or mobile device processing capabilities. Mobile end user devices such as smartphones used to take part in audiovisual calls often represent possible bottlenecks in the service delivery chain. Fortunately, each new generation of devices has brought more advanced hardware in terms of memory, processor power, camera, and battery cycle. Rapid development in the smartphone hardware industry in the last several years implies that the majority of recently released smartphones should be able to provide acceptable QoE for audiovisual calls with adapted video quality streams.

Trends have shown increases in smartphone screen sizes, aiming to accommodate higher screen resolutions. High resolution displays impose additional load on the processing unit, particularly on the graphics processor, needed to render high definition images faster. Smartphone screen sizes will most likely not be much bigger in the future, since carrying devices with displays larger than 6" and noticeable weight is not particularly convenient. Thus, an important feature of any service is the possibility to adapt the layout and content to viewing contexts and devices. Even though smartphone displays are relatively small, with limited options for manipulating the design layout, results from studies focusing on desktop video conferencing could be taken into consideration and further extended to mobile devices [17].

The authors in [37] identified mobility, device and encoding interoperability, ease of use, and additional collaboration possibilities (e.g., exchanging pictures, files, chatting) as the most important aspects for audiovisual call services. However, mobility and device capacity, alone and together, can create asymmetry, which is a realistic and common case. Thus, the impact of the mobility and device, depending on the number of the participants, can greatly differ due to the numerous possible combinations of connection type between locations and type of equipment being used.

In the context of mobile networks, characterized by variable resource availability, challenges arise with respect to meeting the QoE requirements of conversational real-time, media rich, and multiparty services [22]. In addition to network requirements, audiovisual calls impose high requirements on end user device processing capabilities, with the need for real-time encoding and decoding of multiple media streams.

Recently, after the COVID-19 outbreak, Skowronek et al. [38] has provided an extensive, detailed, and comprehensive study on factors affecting QoE of videoconferencing. This study, among others, systematically analyzed the impact of system, human, context, and mixed factors on QoE in fixed audiovisual calls context. Addressed SIFs were grouped into: ones related to signal transmission over the system (media richness, processing, network and topology, and time) and technical aspects related to user interaction with the system (setting up a call and management of one ongoing). Recognized IFs are: communication mode, audiovisual presentation of environment and participant, audio and visual mixing paradigm and signal processing, audiovisual synchronization, end-to-end delay, call setup and management of ongoing call, participation registration, network access behavior and computation distribution, installation complexity, user interface, etc. Although it offers a novel systematization and taxonomy approach to influence factors, this study together with several others from this period refer to older references.

There are also several studies that addressed the impact of various factors on QoE in services similar to audiovisual teleconferences, such as unified communications [5], video consultations [32], or video streaming [3].

While it is clear that a wide range of IFs affect QoE in audiovisual calls, questions remain as to the level and importance of the impact of particular factors, especially in a mobile context. For example, the question of whether certain impairments cause strong, noticeable, or imperceptible quality degradation commonly depends on the particular scenario, context, as well as the individual involved users.

However, although the existing studies analyzed these various impacts of IFs, what they lack is the user's opinion on whether those IFs should have been analyzed, i.e., do they matter to the end users of audiovisual calls on smartphones. That is why in this paper our primary focus has been on collecting user opinions pertaining to the impact of various system-related QoE influence factors, so that the most important ones (according to users' opinions) can be used afterwards as input in empirical studies examining QoE of audiovisual calls on smartphones.

Based on standards and related work, Table 1 provides a summary of QoE SIFs for audiovisual calls on smartphones to be considered according to our proposed categories. The categories are: *media quality*, *functional support*, and *usability*, *service design*, and *resource consumption*. These categories are used in the scope of our survey questionnaire which will be described in the next Section. It is important to emphasize that the *media quality* group is related to factors falling into network, media, and content SIFs. The *functional support* group as well as *usability*, *service design*, and *resource consumption* are related to device and application SIFs. They are divided into two groups, to ensure clearer distinction by participants.

In the following Section, we outline our survey methodology, focusing on contributing to state-of-the-art knowledge by collecting and analyzing user opinions and expectations related to various QoE SIFs, with a focus on audiovisual calls established on smartphones.

3 Methodology

Given the wide range of factors that may impact end user expectations and quality ratings, we conducted a web-based questionnaire survey with the goal being to investigate users' opinions and expectations related to audiovisual calls on mobile devices. Furthermore, we highlighted and clearly stated that reported answers should be considered in a leisure context, i.e., when communicating and interacting with family or friends. The aim of this questionnaire was to investigate the factors that users identify as most influential in contributing to their overall experience and quality perception.

As previously stated, we conducted our research in two phases. The first survey (S1) was conducted as a part of phase 1, just prior to the global outbreak of the COVID-19 pandemic, and reflects the views and opinions of users at that time. Due to the increase in video communication services resulting from lockdown measures, in phase 2 we repeated the survey 20 months later (Survey S2), so as to assess whether or not there are any differences in user opinions. The surveys were prepared using the Google Docs service and distributed via email to acquaintances, colleagues, and students. A total of 272 participants took part in S1, with responses collected over a period of thirteen days, from February 13 until February 26, 2020. The second survey, S2, was conducted during the period between October 5 and November 15, 2021, with 249 participants successfully having completed the questionnaire.

Table 1 System influence factors to be considered when assessing and modeling QoE for audiovisual calls on smartphones

Category	Influence factors
Media	Speech intelligibility Uninterrupted interaction Longer video freezes Perceptible audio delay Perceptible video delay Short video freezes Audio-video synchronization Image blurriness Voice naturalness Image sharpness Smooth movement Color accuracy
Functional support	Speaker identification Audio mute File transfer Texting Adaptive layout Audiovisual call recording Video pausing Applying filters
Usability, service design, resource consumption	Service reliability Security Device/browser interoperability Duration of call establishment time Service price Ease of use Installation complexity Noise free environment User interface aesthetics Low battery consumption Smooth simultaneous use of other apps

The majority of participants involved in the first study are from Croatia (88.97%), while 6.62% and 4.41% of participants are from Serbia and Bosnia and Herzegovina, respectively. More than half of the participants included in the second study are from Bosnia and Herzegovina (53.82%), followed by 42.97% of Croats, while the remaining 3.21% are from Serbia.

To gather user feedback on the perceived quality of audiovisual calls, two aspects of service delivery were considered: call initiation, and service operation once the audiovisual call is established. Both aspects are comprised of multiple dimensions that contribute to the

overall QoE: effort required by the user, responsiveness of the service, fidelity of information, security, and availability. The questionnaire covered ratings of the impacts and importance of considered factors referring to the application, resources, and context. Selected factors belong to the quality features that can be evaluated by a wider audience from a perceptual perspective.

Questions were divided into the following four groups which address SIFs from Table 1:

- *general information* - referring to the subject's demographic data and previous experiences with taking part in audiovisual calls;
- *media quality* - referring to the quality of the speech (audio) and the image (video) in terms of perceivable impairments (e.g., delay, blurriness);
- *functional support* - referring to the additional functionalities supported by audiovisual call/conferencing services, beyond only basic support for audiovisual call;
- *usability, service design, and resource consumption* - referring to the ease of use, aesthetic design, as well as various service design features, such as service reliability, security, price, and battery consumption.

In terms of network factors, participants are asked to rate their impact through quality and perceivable impairments.

We offered participants only closed-ended questions that provided a fixed set of options to choose from. Closed-ended response choices were comprised of yes/no options, multiple choice options, and rating scales. The impact of each factor was rated on a 5-point scale. One set of questions used the following rating scale to collect user opinions with respect to the importance of certain factors: 5 - "Very Important", 4 - "Important", 3 - "Moderately Important", 2 - "Slightly Important", 1 - "Not Important". The other set of questions used the following scale to collect feedback on the extent to which users considered certain factors to impact perceived quality: 5 - "To a great extent", 4 - "To a moderate extent", 3 - "To some extent", 2 - "To a small extent", 1 - "Not at All". Further details and concrete questionnaire items are given in [Appendix](#).

When creating the questionnaire, we adhered to the principle of simplicity. In other words, all questions were formulated in a simple and clear manner in order to avoid confusion. In cases where it was necessary, we added an explanation in the questionnaire for further clarification. Furthermore, prior to each question set, we provided a short explanation for the survey participants to know what they are rating. The short descriptions used in the questionnaire are included in the annexes.

The questionnaire was written in the Croatian language, while its English translation is provided in [Appendix](#). Given the similarities between the Croatian, Bosnian and Serbian languages, the choice of words and phrases used in the questionnaire ensured that all respondents (regardless of nationality) were able to interpret the questions in the correct manner. Furthermore, we conducted a pilot study prior to the two reported studies, involving 15 participants who were later on not included in the actual survey, in order to confirm that the questions were clearly formulated and comprehensible to people that are non experts in the AV field.

Given that we used no constructs in our evaluation questionnaire, there was no need to conduct the convergent and discriminant validity. Also, since the research does not include criterion, there was no need to assess criterion validity (concurrent and predictive). Finally, content validity has been addressed. The content validity assesses whether a test is representative of all aspects. There is no direct measure of content validity so it was tested by relevant ICT experts who reviewed and rated the survey questions, removing the ones that were marked as irrelevant and accepting and adjusting the ones that were relevant.

Reliability considers the extent to which the questions used in a survey instrument consistently elicit the same results each time it is asked in the same situation on repeated occasions. Reliability is a statistical measure of how reproducible the survey instrument's data is. A survey instrument is said to have high reliability if it produces similar results under consistent conditions, and any change would be due to a true change in the attitude, as opposed to changing interpretation (i.e., a measurement error). In our case, the survey instrument produces similar results, which proves survey reliability.

Participant demographics Information related to age, gender, and education was collected to identify participant demographics (Table 2). In the first survey (S1), the majority of users (49.6%) fit into the category 36–45 years old, while in S2 the largest percentage fit into the 18–25 years category with a share of 53.82%. Fairly equal gender representation was recorded in both surveys, with 51.1% females in S1 and 50.2% in S2. The highest response rate for educational level in S1 was a University degree (71%), in contrast to S2 where 61.05% were younger adults with only a high school diploma.

To investigate possible generational differences in perception we grouped participants into two categories (considering uneven age distribution in both surveys): young adults (covering the age from 18–35) and middle-aged adults and older (age 36 and older). In S1, the average calculated difference in mean values between the two age groups and considering all IFs was 4.16%, while the biggest difference (13.84%) in average ratings between the two age groups was observed for the importance of the functionality *video pausing* (we note that in the context of audiovisual calls, this refers to turning off the video while maintaining audio communication), where young adults rated the impact on average with 3.43, and adults (older than 35) with 2.95. In S2, the average calculated difference in mean values between the two age groups and considering all IFs was 4.8%, while the biggest difference (9.9%) between average ratings was observed for *video delay*, where young adults rated the impact with an average score of 4.16, and adults (older than 35) with an average score of 4.57. Given that in both surveys, average ratings for the majority of IFs did not significantly differ between young adults and middle-aged to older adults, we refrain from further analyzing the impact of age group in the scope of the results analysis given in the following section.

Table 2 Demographic information about participants that completed the survey in February 2020 and October 2021

Survey		February 2020 (S1)	October 2021 (S2)
No. of participants		272	249
Age group	18–25	11.00%	53.82%
	26–35	25.00%	5.62%
	36–45	49.60%	28.92%
	46–55	11.00%	10.84%
	> 55	3.40%	0.80%
Gender	Female	51.10%	50.20%
	Male	48.90%	49.80%
Educational level	High school degree	19.10%	61.05%
	University degree	71.00%	32.93%
	PhD degree	9.90%	6.02%

4 Analysis of survey results

Following the collection of demographic data, the remainder of the questionnaire focused on collecting information on service usage habits, followed by user opinions with respect to the importance of certain factors and the extent to which users considered certain factors to impact perceived quality. Details are given in the remainder of this section, focused also on comparing S1 and S2 results. The factors considered by users to have the greatest impact on QoE are further summarized and compared across both studies in Section 5.

4.1 Service usage

Following the collection of demographic data, the aim of the following set of questions was to identify users' habits associated with audiovisual calls. In terms of application frequency usage, we categorized participants per user type as: *very frequent user* (uses audiovisual call applications on a daily basis), *frequent user* (uses audiovisual call applications 2 to 3 times per week), *occasional user* (uses audiovisual call applications 4 to 7 times per month), and *light user* (uses audiovisual call applications rarely, 3 or less times per month). Of the 272 participants from S1, 16.2% reported participation in audiovisual calls in the last 30 days on a daily basis, while in S2 this increased to 21.3% (out of 249 participants) (Table 3). The biggest increase can be observed in the category of frequent usage (2-3 times per week), where the percentage of participants more than doubled and rose to 34.1% in October 2021. The percentage of participants that did not participate in any audiovisual call in the last 30 days dropped from 17.6% in 2020 to 3.6% in 2021.

The question addressing device usage was a multiple choice question type that allowed participants to select one or multiple answers from a defined list. Out of 272 participants from S1, 94.9% reported a smartphone as the device used to make audiovisual calls, 12.1% reported using a tablet, 52.2% a computer/laptop, while 1.5% responded they used some other device (Fig. 1). Participants in S2 reported similar usage of smartphones (94%) and tablets (10.1%), while the biggest difference can be observed in the computer/laptop category, where 81.9% participants reported using such devices for participating in audiovisual calls.

With respect to previous experiences with applications, participants were allowed to choose multiple predefined answers. In survey S1, Whatsapp and Skype were the two most commonly used applications in the audiovisual call context, with a share of 89.3% and 85.7%, respectively. This was followed by Viber (70.6%) and Google Hangouts (22.8%). Appair.in (renamed to Whereby in 2019) was used by 3.7%, while 26.5% of subjects used other apps (Fig. 2).

Table 3 Frequency of users that reported having participated in an audiovisual call during the last 30 days

Frequency of participation	February 2020 (S1)	October 2021 (S2)
Very frequently	16.2%	21.3%
Frequently	14.7%	34.1%
Occasionally	16.9%	20.9%
Rarely	34.6%	20.1%
Never	17.6%	3.6%

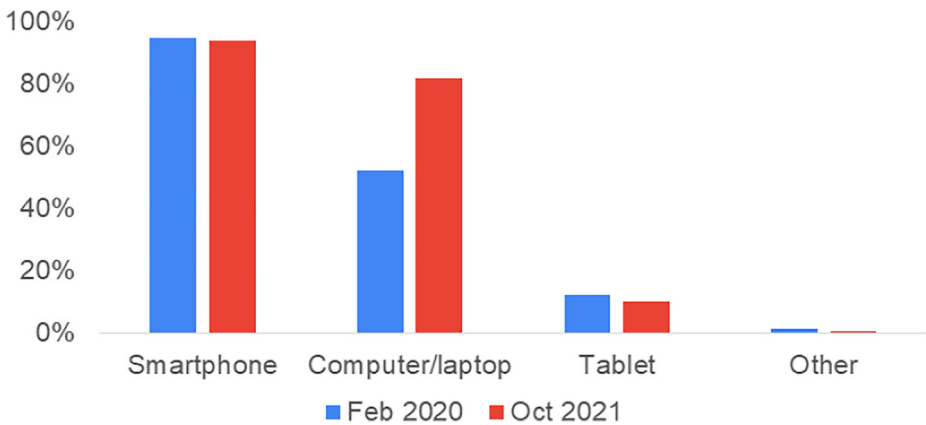


Fig. 1 Percentage of participants that reported using a given device when taking part in audiovisual calls

When conducting survey S2, we added additional popular applications, such as Microsoft Teams, Zoom, and FaceTime. In contrast to the first survey, Whatsapp (73.5%) reached fourth place, while Viber (83.5%) and Teams (75.9%) were the two most commonly used apps for establishing audiovisual calls as reported in our 2021 survey (Fig. 3).

4.2 Opinions related to media quality

Questions regarding the importance of media quality factors and impact on the overall perceived quality of audiovisual calls were comprised of questions including a predefined list of five answer options (Appendix). Questions were explicitly related to audiovisual calls established via a mobile smartphone in a leisure context. Rating distributions and descriptive statistics for both S1 (2020) and S2 (2021) are given in Fig. 4.

Ratings of the impact of *speech intelligibility* on overall audiovisual call quality had the highest mean value in both surveys, corresponding to 4.69 (S1) and 4.77 (S2). Corresponding values of standard deviation were the lowest within the *media quality* group. Speech intelligibility is a measure of the effectiveness of speech communication usually defined as

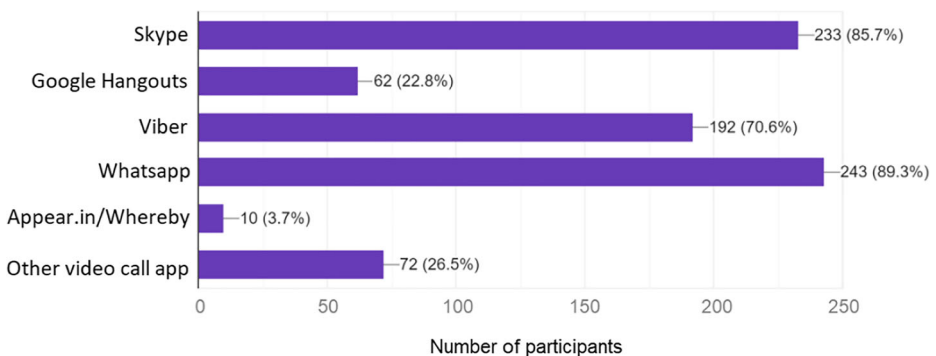


Fig. 2 Percentage of applications used when making an audiovisual call reported in 2020

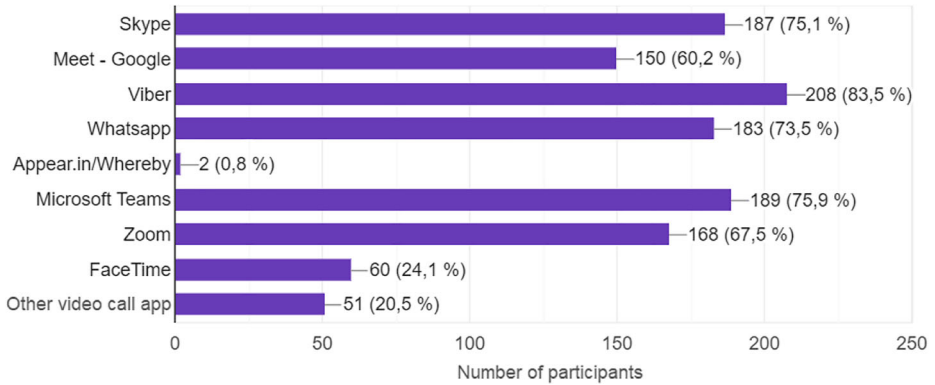


Fig. 3 Percentage of applications used when making an audiovisual call reported in 2021



Fig. 4 Distribution of ratings and descriptive statistics for IFs belonging to the *Media quality* group. For each factor, the first bar corresponds to 2020 and the second bar to 2021

the percentage of speech units (syllables, words, or sentences) correctly perceived by listeners. Reduced intelligibility occurs due to the nature of the spoken material (unfamiliarity with the speaker, possible abnormal speech characteristics, or unfamiliarity with the conversation topic) and the context of transmission [8]. It also depends on audio bandwidth, channel impairments, input (microphone) and output (speaker) of end user device characteristics and its placement in relation to the speaker/listener, acoustical properties of the room, sound pressure level, and background noise level [20]. If the cause of poor intelligibility does not lie in human characteristics, yet involves system components, there is a possibility to isolate the cause of the reduced quality and prevent further degradation.

The ability to interact in the presence of interruptions can be difficult even in face to face communication. Participants in video mediated communication can be severely affected by transmission delays, where comprehension can be distorted by mutual silence or double talk [25].

Further considering media quality, *image blurriness*, *image sharpness*, *voice naturalness* and *smooth movement in the video* showed high percentages of 4 - “Important” and 3 - “Moderately Important” ratings. *Color accuracy* was the influence factor that showed the highest dispersion among ratings, with the greatest standard deviation (0.99 for 2020 and 1.07 for 2021), as well as having the highest number of reported 1 - “Not Important” ratings in both surveys within the *media quality* group.

Seven out of twelve influence factors gained slight importance from February 2020 to October 2021, however there was no significant difference. In order to quantify the change from mean value (of factor importance reported in S1) to mean value (of that same factor reported in S2) and express the change as an increase or decrease we calculated the percentage change. Percentage change equals the change in mean value (S2-S1) divided by the value of the initial S1 mean value, multiplied by 100. The importance of *color accuracy* rose the most (on average by 3.66%), while the importance of *voice naturalness* decreased by 2.56%. Overall, a very strong negative correlation coefficient -0.9 (calculated for all factors belonging to the *media quality* group) between mean value and standard deviation shows that for factors rated as having higher importance in terms of impact on quality, there was less diversity in user ratings.

Summary of key findings *Speech intelligibility* and *uninterrupted interaction* were the two factors with the highest mean ratings in terms of importance. Factor analyses showed that the perceived importance of 8 IFs (*speech intelligibility*, *uninterrupted interaction*, *longer video freezes* (i.e., longer than 15 seconds), *perceptible audio delay*, *perceptible video delay*, *short video freezes* (lasting a few seconds), *audio-video synchronization*, *image blurriness*, *voice naturalness*, *image sharpness*, *smooth movement in the video* and *color accuracy*) on media quality did not significantly differ between S1 and S2, despite increased usage of audiovisual call services.

Furthermore, user opinions with respect to the importance of various media quality factors did not differ greatly with respect to the frequency of video conferencing service usage patterns. We thus conclude that even in cases of occasional service use, participants provided very similar opinions as did users who more frequently used such services. Due to space limitations, we refrain from further analysis, however interested researchers are referred to our publicly available survey results.

4.3 Opinions related to functional support

Questions regarding additional functionalities supported by conferencing services (beyond only audiovisual calls) and corresponding importance were comprised of closed-ended

questions including a predefined list of five answer options. Rating distributions and descriptive statistics for both S1 (2020) and S2 (2021) are given in Fig. 5. *Active speaker identification*, i.e., being able to identify the speaker who is currently talking, was the influence factor with the highest mean importance ratings in both surveys, 4.11 (S1) and 4.17 (S2), and lowest standard deviation values within the *functional support* group, namely 0.88 and 0.9, respectively.

On the other hand, being able to *apply make-up/filters/overlay items* was the functionality perceived as least important in both surveys, whereas importance in 2021 decreased to a mean value of 1.81, even though the majority of participants were of a younger age in S2. More than half of the participants perceived the possibility to apply filters as “*Not Important*”.

In case of additional functionalities, seven out of eight influence factors gained in mean importance ratings from February 2020 to October 2021, with mean ratings for *file transfer* and *audiovisual call recording* increasing by more than 0.5. Standard deviation values ranged from 0.96 (for *file transfer* in 2021) to 1.16 (for *audiovisual call recording* in 2020).

Overall results show a strong negative correlation (with Pearson correlation coefficient -0.66) between perceived importance of IFs and corresponding standard deviation values, meaning that with increased perceived importance, the dispersion of reported values does not decrease significantly, as was the case with the media quality factors.

Summary of key findings The most significant differences between S1 and S2 were observed in factors belonging to the *functional support* group. The increased importance of the IFs belonging to the functional support group is an important change indicating that pure audio and video communications is not necessarily sufficient anymore. Additional features are needed to enhance the meeting quality in terms of collaboration, engagement, and interaction, ultimately making the communication easier and more effective.

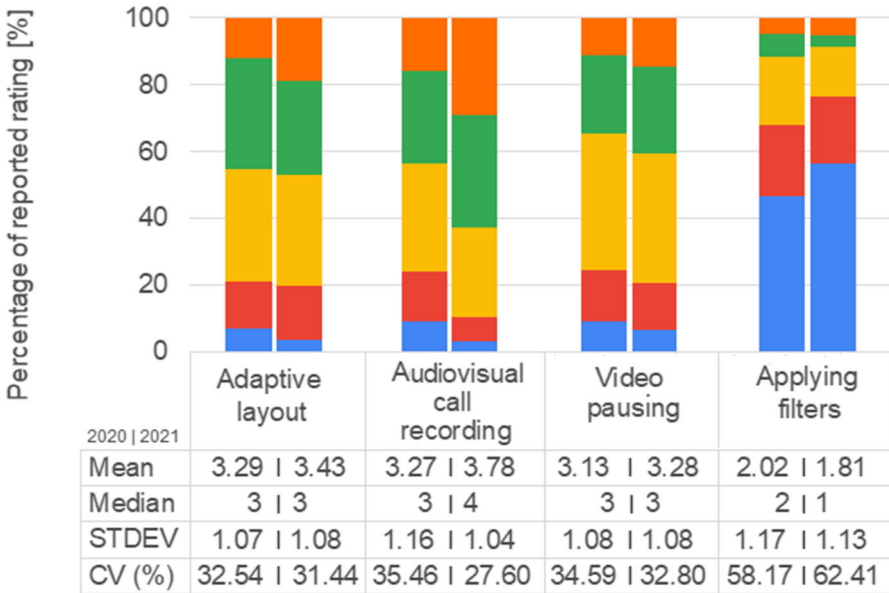
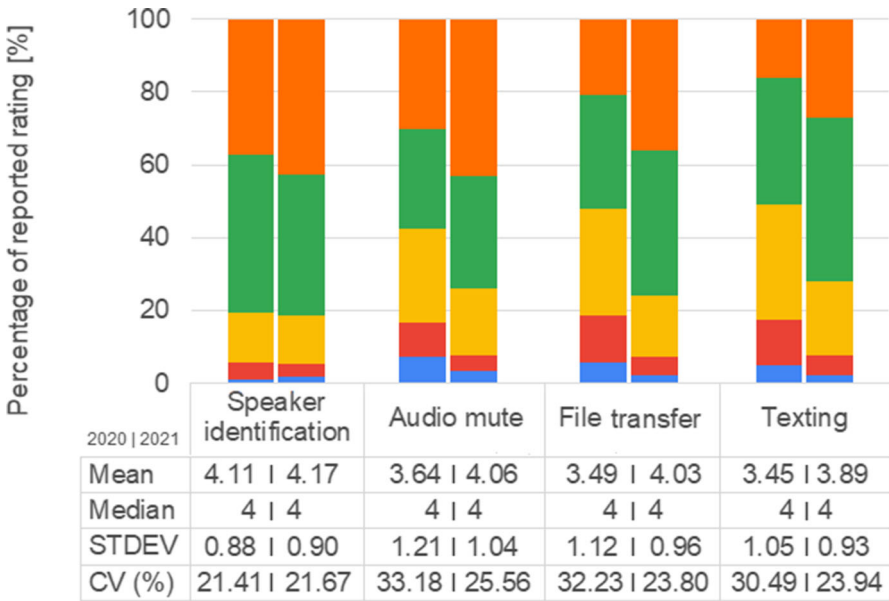
4.4 Opinions related to usability, service design, and resource consumption

Questions regarding usability, service design, and resource consumption referred to the ease of use of the application, the extent to which users feel they are able to conduct audiovisual calls, and to the mobile context, encompassing usability, portability (in terms of efficiency with which the audiovisual call application can be transferred from one operational or usage environment to another), and resource consumption (battery consumption and CPU utilization). The final set of questions included also a predefined list of five answer options. The descriptive statistics for the collected ratings are given in Fig. 6. With respect to usability, *ease of use* and *installation complexity* both had mean ratings greater than 4, indicating the high importance of such factors to end users.

On the other hand, *user interface aesthetics* were deemed less important, with an average rating of 3.42 (S1) and 3.53 (S2). With respect to resource consumption, and given that the questions were specified in the context of mobile device use, rating distributions clearly show the importance of *low battery consumption* and *smooth simultaneous use of other applications*.

Comparing the results between S1 and S2, we observe the biggest difference in *smooth simultaneous use of other applications*, with a higher mean score in 2021 (4.28) as compared to 2020 (3.79). This is followed by the importance of having a *noise free environment*, where the mean score increased by 7.42%.

Finally, rating distributions clearly show that users are highly concerned with service reliability and security (in this case referring to having an encrypted connection during the



■ % rating 1 ■ % rating 2 ■ % rating 3 ■ % rating 4 ■ % rating 5

Fig. 5 Distribution of ratings and descriptive statistics for IFs belonging to the *Functional support* group. For each factor, the first bar corresponds to 2020 and the second bar to 2021

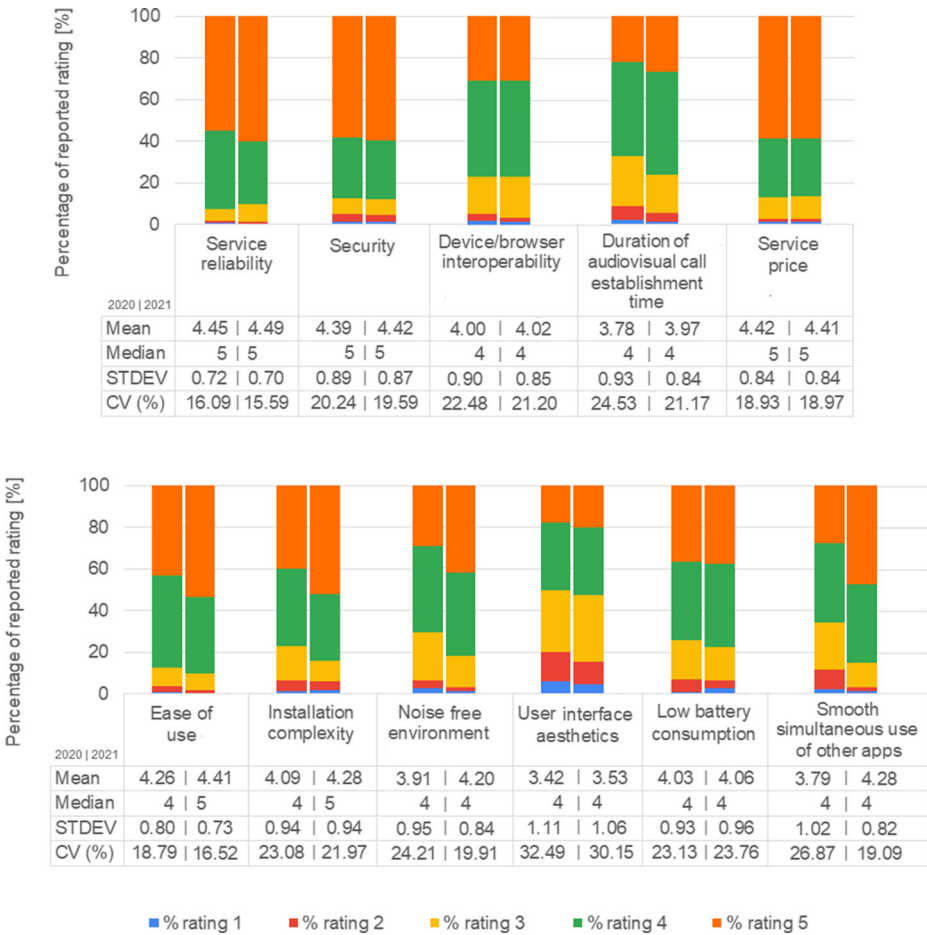


Fig. 6 Distribution of ratings and descriptive statistics for IFs belonging to the *Usability, service design, and resource consumption* group. For each factor, the first bar corresponds to 2020 and the second bar to 2021

audiovisual call). In total, even though ten out of eleven influence factors gained importance when comparing mean values of S1 and S2, a greater difference was found only for *smooth simultaneous use of other applications* and *noise free environment*. For this given group of influence factors, Pearson’s correlation coefficient between mean values and corresponding standard deviation showed a very strong negative correlation (− 0.84).

Summary of key findings Rating distributions showed high importance of *ease of use, resource consumption, security, service reliability, and price*. As expected, and potentially due to increased service usage, factors related to the usability were rated as being more important in S2 as compared to S1.

Impact of usage frequency on perceived importance of considered factors In Study 1, we compared the mean values with all participant ratings included, with the mean values obtained when excluding participants that had not used a video call service in the last 30

days (we note that this corresponds to 17.6% of participants in Study 1). Results showed that there is no significant difference between these mean values. The greatest difference (0.12) is identified for the functionality *audio muting*, where the reported mean value for all participants included is 3.64, and with excluded participants that had not used the service in the last 30 days 3.76. Most importantly, we found that there was no impact on our final conclusions in terms of factor importance, i.e., the list of factors identified as being the most important ones to consider remained the same, even when excluding ratings provided by participants with no recent experience in using audiovisual calls.

We performed the same analysis for Study 2 (where only 3.6% of participants had not taken part in a video call in the past 30 days), where the greatest difference in mean values was found for the factor *adaptive layout* where all participants combined rated *adaptive layout* importance as 3.43, while the group with excluded participants that did not use the video call service in the last 30 days rated with 3.45 in average. Based on the results, we can conclude that frequency of usage did not have a significant impact on perceived importance of considered factors.

Impact of educational level on perceived importance of considered factors To assess whether or not there are any differences in user opinions regarding educational level, we compare the mean ratings of both surveys, S1 and S2. In the first survey (S1), the majority of participants (71%) fit into the category *University degree*, while in S2 the largest percentage fit into the *High school degree* category with a share of 61.05%. The greatest change (in mean ratings) can be noticed within three factors belonging to the *Functional Support* group: *file transfer*, *texting* and *call recording*, and one factor *smooth simultaneous use of other applications* belonging to the *Usability, service design, and resource consumption* group. The importance of *file transfer* rose the most from 3.42 in S1 to 4.21 in S2, followed by *texting* 3.36 (S1) to 4.03 (S2), *audiovisual call recording* 3.32 (S1) to 3.89 (S2), and *smooth simultaneous use of other applications* from 3.81 (S1) to 4.36 (S2). All those factors help to elevate the audiovisual call experience, and the gained importance may possibly be attributed to increased usage of audiovisual calls. The importance of all other evaluated factors did not change greatly, the mean rating difference between two studies was lower than 0.29. For illustration purposes, we compare mean results (per factor) of participants who reported in study S1 level of education *University degree*, and in the second study *High school degree* (Fig. 7).

5 Factors considered by users as most important for audiovisual calls

Following the analysis of rating distributions across different groups of factors, we refer back to RQ1 and identify those factors rated by users as being the most important in terms of their importance for services offering audiovisual calls and in terms of their impact on QoE. The rationale for identifying such “key influence factors” lies in providing valuable input for service designers in terms of factors to consider and optimize so as to increase their customer base, prevent customer churn, and maintain high customer satisfaction. Referring further to RQ2, we compare key factors between our two studies conducted in 2020 and 2021.

To identify key factors, we consider IFs from all three groups, *media quality*, *functional support* and *usability, service design, and resource consumption*. We sorted IFs in descending order (by mean value) and selected the first ten factors that are considered to be most influential (Fig. 8). Selected key factors are bounded by coefficient of variation under 21%,

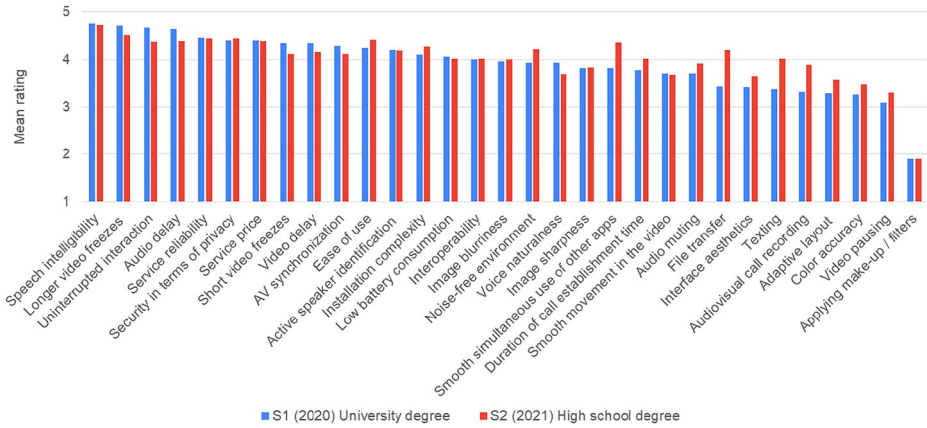


Fig. 7 Mean ratings (per factor) of participants who reported in study S1 level of education *University degree*, and in second study S2 *High school degree*

mean ratings higher than 4.2, and total percentage of given ratings 4 (*Important*) and 5 (*Very important*) combined representing over 85% of ratings. Mean IFs ratings from S1 in 2020 are color coded in gray, while mean factor ratings from S2 in 2021 are color coded in red and green. If the factor gained in importance in 2021 as compared to the previous year, bars are colored green. On the other hand, if importance was reduced, the bar is colored red.

The results of both surveys show that users perceived the same factors as being most important both before the pandemic and nearly two years after the outbreak, despite significantly increased usage of video conferencing services. Furthermore, the list of key factors is the same despite different participant demographics. The greatest positive change in 2021 in perceived factor importance (in terms of mean values) was found for *ease of use of the application*, while the factor *uninterrupted interaction during communication* showed the greatest decrease in importance. However, observed changes are not significant in terms of mean values, distribution, and variation of ratings. Reported data indicates that perceived importance of key influence factors did not change drastically, which might be expected due

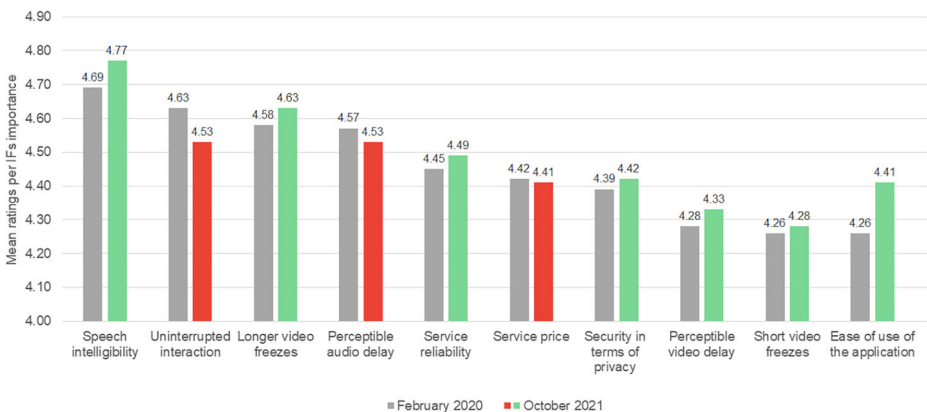


Fig. 8 Key influence factors, selected by the users, in February 2020 and October 2021

to the increased usage of audiovisual call services which could possibly lead to the higher or changed user perception and expectations.

Based on the surveys, we identify relevant areas impacting QoE as pertaining to both the application and network domains: quality of real-time media from the user perspective (*speech intelligibility, uninterrupted interaction, long (i.e., longer than 15 seconds) and short (lasting a few seconds) video freezes, perceptible audio and video delay*), quality of service in terms of *reliability*, and application management (*service price, security in terms of privacy, and ease of use of the application*).

Key influence factors that should be considered can be controlled, at least to some extent from an application point of view, by video encoding parameters, impacting the system as a whole. Namely, on application level it is possible to adapt video quality level (e.g., resolution, bitrate, and frame rate) to avoid CPU overload which can lead to congestion, prevent packet loss and delay, and save bandwidth needed for transmission, resulting at the end with acceptable QoE. Concrete recommendations in terms of video encoding parameters for three-party audiovisual calls established via mobile devices can be found in our earlier work [44, 45].

6 Conclusion

With increasing use of audiovisual communication services, in particular using mobile devices, this paper aims to contribute to insights with respect to user opinions on the impact of various system factors on QoE, in terms of their importance. The advantage of this study is that it fills the gaps of the existing research in this field and takes user-oriented approach in identifying the importance of different SIFs in terms of QoE for audiovisual calls on smartphones. We report on two large scale surveys conducted both before and during the COVID-19 pandemic. In total, 521 participants took part in an online questionnaire designed to investigate users' opinions and expectations related to audiovisual calls on mobile devices in the leisure/private context, with the main goal being to identify key system influence factors grouped in three categories.

The second survey results confirmed initial findings derived based on the first survey in terms of key influence factors, with just slight differences in user opinions reported, in order and mean value of importance. Given that the two surveys were conducted at different time frames and involving different participants, differing to a certain extent in terms of demographics, this contributes to the generalizability of obtained results. Furthermore, it can be concluded that increased and more intense usage of audiovisual calls did not impact greatly perceived importance of key influence factors.

Therefore, the contribution of the paper is three-fold. We have proposed a categorization of QoE SIFs for audiovisual calls on smartphones in the following groups: media quality, functional support, and usability, service design, and resource consumption. Further, we have identified the important system influence factors for QoE in this context. Finally, we have approached the topic by taking a complementary approach as compared to existing empirical user studies, i.e., we have determined the importance of SIFs by asking the users for their opinions.

Obtained results can provide valuable input for service providers in terms of factors to consider and optimize so as to increase their customer base and maintain high satisfaction. Selected factors can be controlled to some degree by adaptation (in accordance with available resources such as mobile devices processing power or network conditions) of video encoding parameters on application layer (video bitrate, resolution, and frame rate).

Additionally, identified factors can serve as input when deriving QoE models for audiovisual calls on mobile devices.

In future research, we aim to design and conduct ecologically valid studies further quantifying the impact of identified key influence factors in both leisure and business contexts.

Appendix: Questionnaire

The questionnaire collects general demographic information, users' habits, and ratings of the impacts and importance of considered factors referring to the application, resources, and context.

Table 4 General information and users' habits associated with audiovisual calls

How old are you?	18–25	26–35	36–45	46–55	More than 55
What is your gender?	Female	Male			
What is your country of origin?	a) Croatia b) Bosnia and Herzegovina c) Serbia d) Other				
What is your education level?	a) High school degree b) University degree c) Higher University degree (PhD)				
Please indicate which of the following applications you have used?	(Multiple choices are allowed)				
Feb. 2020:	a) Skype	b) G Hangouts	c) Viber	d) Whatsapp	e) Whereby f) Other
Oct. 2021:	a) Skype	b) Google Meet	c) Viber	d) Whatsapp	e) Whereby
	f) Zoom	g) Microsoft Teams	h) FaceTime	i) Other	
How often have you participated in the listed applications during the last 30 days?	a) Very Frequently (on a daily basis) b) Frequently (2–3 times per week)				
	c) Occasionally (4–7 time per month) d) Rarely (1–3 time per month) e) Never				
Which of the following devices have you used in the past to make audiovisual calls?	(Multiple choices are allowed)				
	a) Smartphone	b) Tablet	c) Computer/laptop	d) Other	

The second part of the questionnaire is focused on quality aspects (in terms of user opinions with respect to the importance of certain factors) of audiovisual calls established via smartphones in a leisure context. Two different rating questions were asked, with a scale of answer options where participants can select the number/word that represents their opinion.

Type 1 How important do you consider the following factor for overall audiovisual call quality?

This set of questions used the following rating scale to collect user opinions with respect to the importance of certain factors: 5 - “*Very Important*”, 4 - “*Important*”, 3 - “*Moderately Important*”, 2 - “*Slightly Important*”, 1 - “*Not Important*”.

Type 2 To what extent do you consider the following factor to impact overall audiovisual call quality?

Table 5 Questions related to *Media Quality* IFs. Media quality refers to the quality of the sound (audio) and the image (video) in terms of perceivable impairments (e.g., perceptible audio delay, image blurriness). Please evaluate how important you consider the following factors of audiovisual calls

Rating scale	5	4	3	2	1
How important do you consider the following factor for overall audiovisual call quality? ^a					
Speech intelligibility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Voice naturalness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Uninterrupted interaction during communication	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Audio-video synchronization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Image sharpness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Smooth movement in the video	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Color accuracy (colors do not differ significantly from real colors)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To what extent do you consider the following factor to impact overall audiovisual call quality? ^b					
Perceptible audio delay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Perceptible video delay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Short and occasional video freezes (lasting a few seconds)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Longer video freezes (i.e., longer than 15 seconds) impact overall audiovisual call quality, if the audio quality remains good for the duration of the call	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Image blurriness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

^a5 - “Very Important”, 4 - “Important”, 3 - “Moderately Important”, 2 - “Slightly Important”, 1 - “Not Important”

^b5 - “To a great extent”, 4 - “To a moderate extent”, 3 - “To some extent”, 2 - “To a small extent”, 1 - “Not at All”

This set of questions used the following scale to collect feedback on the extent to which users considered certain factors to impact perceived quality: 5 - “To a great extent”, 4 - “To a moderate extent”, 3 - “To some extent”, 2 - “To a small extent”, 1 - “Not at All”.

We specifically note before each group of factors that the following questions apply to calls established via smartphones in a private/leisure context (e.g., calls with friends, relatives, etc.) and not to business conference calls.

Table 6 Questions related to *Functional Support* IFs. Nowadays, many audiovisual call applications offer additional functionalities, beyond only audiovisual calls, such as image sharing or exchanging text messages. Please evaluate how important you consider the following factors of audiovisual calls

Rating scale	5	4	3	2	1
How important do you consider the following factor for overall audiovisual call quality? ^a					
File transfer (image, video, document sharing)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Texting (sending text messages)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Active speaker identification (i.e., the participant who is currently talking is highlighted/marked in some way)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Applying make-up filters/overlay items (e.g., hat, mask)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Adaptive layout (e.g., movable participant's preview window, display zooming)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Video pausing while in the call	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Audio muting while in the call	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Audiovisual call recording	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

^a5 - "Very Important", 4 - "Important", 3 - "Moderately Important", 2 - "Slightly Important", 1 - "Not Important"

Table 7 Questions related to *Usability, Service Design, and Resource Consumption* IFs. Usability, service design, and resource consumption factors are related to the primary service, such as extent to which you feel you are able to make and conduct audiovisual calls. Please answer the following questions regarding how important you consider the following factors

Rating scale	5	4	3	2	1
How important do you consider the following factor for overall audiovisual call quality? ^a					
Device/browser interoperability (meaning that participant can use audiovisual call application regardless of the participant's smartphone model or software installed)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Duration of audiovisual call establishment time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ease of use of the application (i.e., how easily you can use an application to communicate)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Installation complexity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
User interface aesthetics (visual appearance)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reliability of the service (i.e., being able to use the service - audiovisual call - correctly the first time)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Security in terms of privacy (i.e., information transmitted is encrypted)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Low battery consumption during the audiovisual call	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Smooth simultaneous use of other applications (enabled uninterrupted use of other applications at the same time)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Noise free environment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Service price	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

^a5 - "Very Important", 4 - "Important", 3 - "Moderately Important", 2 - "Slightly Important", 1 - "Not Important"

Author Contributions All authors contributed to the study conception and design equally. Material preparation, data collection and analysis were performed by Dunja Vučić, Sabina Baraković and Lea Skorin-Kapov. All authors read and approved the final manuscript. We confirm that the order of authors has been approved by all named authors.

Funding This work has been supported by the Croatian Science Foundation under the project IP-2019-04-9793 (Q-MERSIVE).

Availability of Data and Materials The anonymized datasets are publicly available via an open repository (link: <https://muexlab.fer.hr/muexlab/research/datasets>).

Declarations

Ethics Approval The survey was conducted in the scope of the Croatian Science Foundation project IP-2019-04-9793, which received approval of the Ethics Committee of the University of Zagreb Faculty of Electrical Engineering and Computing.

Consent to Participate The authors confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed.

Consent for Publication All authors agree with the content and give explicit consent to submit the paper. Consent was also obtained from the responsible authorities at the institute/organization where the work has been carried out, before the work was submitted.

Conflict of Interests The authors confirm there are no known conflicts of interest/competing interests associated with this paper that could inappropriately influence, or be perceived to influence, this work.

References

1. Ammar D, De Moor K, Heegaard P (2018) An experimental platform for qoe studies of webrtc-based multi-party video communication. *Int J New Comput Archit Appl* 8(2):89–95
2. Ammar D, De Moor K, Skorin-Kapov L, Fiedler M, Heegaard PE (2019) Exploring the usefulness of machine learning in the context of WebRTC performance estimation. In: 2019 IEEE 44th conference on local computer networks (LCN). IEEE, pp 406–413
3. Baraković Husić J, Baraković S (2022) Multidimensional modelling of quality of experience for video streaming. In: *Computers in human behavior*, vol 129
4. Balihodžić M, Husić JB, Baraković S (2020) The influence of system factors on QoE for WebRTC video communication. In: *International symposium on innovative and interdisciplinary applications of advanced technologies*. Springer, pp 255–267
5. Baraković Husić J, Baraković S, Cero E, Slamnik N, Oćuz M, Dedović A, Zupčić O (2020) Quality of experience for unified communications: a survey. *Int J Netw Manag* 30(3):2083
6. Bouraqlia K, Sabir E, Sadik M, Ladić L (2020) Quality of experience for streaming services: measurements, challenges and insights. *IEEE Access* 8
7. Carofigli G, Grassi G, Loparco E, Muscariello L, Papalini M, Samain J (2021) Characterizing the relationship between application QoE and network QoS for real-time services. In: *Proceedings of the ACM SIGCOMM 2021 workshop on network-application integration*, pp 20–25
8. Coppens-Hofman MC, Terband H, Snik AF, Maassen BA (2016) Speech characteristics and intelligibility in adults with mild and moderate intellectual disabilities. *Folia Phoniatr Logop* 68(4):175–182
9. Dasari M, Sanadhya S, Vlachou C, Kim K-H, Das SR (2018) Scalable ground-truth annotation for video qoe modeling in enterprise wifi. In: 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS). IEEE, pp 1–6
10. Dasari M, Vargas S, Bhattacharya A, Balasubramanian A, Das SR, Ferdman M (2018) Impact of device performance on mobile internet qoe. In: *Proceedings of the internet measurement conference 2018*, pp 1–7
11. De Masi A, Wac K (2022) Less annoying: quality of experience of commonly used mobile applications. In: *Proceedings of the 13th ACM multimedia systems conference*, pp 86–95

12. De Moor K, Arndt S, Ammar D, Voigt-Antons J-N, Perkis A, Heegaard PE (2017) Exploring diverse measures for evaluating QoE in the context of WebRTC. In: 2017 ninth international conference on quality of multimedia experience (QoMEX). IEEE, pp 1–3
13. Dobrian F, Awan A, Joseph D, Ganjam A, Zhan J, Sekar V, Stoica I, Zhang H (2013) Understanding the impact of video quality on user engagement. *Commun ACM* 56(3):91–99
14. Feldmann A et al (2021) Implications of the covid-19 pandemic on the internet traffic. In: Broadband coverage in Germany; 15th ITG-symposium. VDE, pp 1–5
15. García B, Gallego M, Gortázar F, Bertolino A (2019) Understanding and estimating quality of experience in WebRTC applications. *Computing* 101(11):1585–1607
16. Gulliver SR, Ghinea G (2006) Defining user perception of distributed multimedia quality. *ACM Trans Multimed Comput Commun Appl (TOMM)* 2(4):241–257
17. Holub J, Isabelle S, Krylová A, Avetisyan H (2022) Subjective influence of camera-gaze angular offset. *IEEE Access* 10:9321–9327
18. Höbfeld T, Biedermann S, Schatz R, Platzer A, Egger S, Fiedler M (2011) The memory effect and its implications on web QoE modeling. In: 2011 23rd international teletraffic congress (ITC). IEEE, pp 103–11
19. Husić JB, Baraković S, Veispahić A (2017) 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). IEEE, pp 428–433
20. ITU-T Rec (2016) P.807 Subjective test methodology for assessing speech intelligibility. Technical report, ITU-T
21. ITU-T Rec (2017) P.1301 Subjective quality evaluation of audio and audiovisual multiparty telemeetings. Technical report, ITU-T
22. Jana S, Chan A, Pande A, Mohapatra P (2016) QoE prediction model for mobile video telephony. *Multimed Tools Appl* 75(13):7957–7980
23. Jumisko-Pyykkö S (2011) User-centered quality of experience and its evaluation methods for mobile television. Tampere University of Technology 12
24. Jumisko-Pyykkö S, Vainio T (2010) Framing the context of use for mobile hci. *Int J Mob Human Comput Inter (IJMHCI)* 2(4):1–28
25. Kale V (2019) Digital transformation of enterprise architecture. CRC Press, Boca Raton
26. Laghari AA, Channam MI, He H (2018) Measuring effect of packet reordering on quality of experience (qoe) in video streaming. *3D Res* 9(3):30
27. Le Callet P, Möller S, Perkins A (2013) Qualinet white paper on definitions of quality of experience (2012) version 1.2. In: Proceeding European network quality experience multimedia system services (COST Action IC), pp 1–23
28. Lee I, Lee J, Lee K, Grunwald D, Ha S (2021) Demystifying commercial video conferencing applications. In: Proceedings of the 29th ACM international conference on multimedia, pp 3583–3591
29. Liu M, Joskowicz J, Sotelo R, Hu Y, Chen Z, Yang L (2022) Subjective quality assessment of one-to-one video-telephony services. In: 2022 IEEE international symposium on broadband multimedia systems and broadcasting (BMSB). IEEE, pp 1–6
30. Matulin M, Mrvelj Š, Abramović B, Šoštarić T, Čejvan M (2021) User quality of experience comparison between skype, microsoft teams and zoom videoconferencing tools. In: International conference on future access enablers of ubiquitous and intelligent infrastructures. Springer, pp 299–307
31. Mrvelj Š, Matulin M (2021) Modeling the level of user frustration for the impaired telemeeting service using user frustration susceptibility index (ufsi). *Electronics* 10(18):2202
32. Øie EB, Koniuch K, Cieplńska N, De Moor K (2021) Factors influencing qoe of video consultations. In: 2021 13th international conference on quality of multimedia experience (QoMEX). IEEE, pp 137–140
33. Rao N, Maleki A, Chen F, Chen W, Zhang C, Kaur N, Haque A (2019) Analysis of the effect of QoS on video conferencing QoE. In: 2019 15th international wireless communications & mobile computing conference (IWCMC). IEEE, pp 1267–1272
34. Schmitt M, Gunkel S, Cesar P, Bulterman D (2014) The influence of interactivity patterns on the quality of experience in multi-party video-mediated conversations under symmetric delay conditions. In: Proceedings of the 3rd international workshop on socially-aware multimedia, pp 13–16
35. Silva AFD, Mylène C (2018) Perceptual strengths of video impairments that combine blockiness, blurriness, and packet-loss artifacts. *Electronic Imaging* 2018(12):234–1
36. Skowronek J (2017) Quality of experience of multiparty conferencing and telemeeting systems. PhD thesis, Ph. D. thesis, Technical University of Berlin
37. Skowronek J, Schoenberg K, Berndtsson G (2014) Multimedia conferencing and telemeetings. In: Quality of experience. Springer, pp 213–228

38. Skowronek J, Raake A, Berndtsson G, Rummukainen OS, Usai P, Gunkel SN, Johanson M, Habets EA, Malfait L, Lindero D et al (2022) Quality of experience in telemeetings and videoconferencing: a comprehensive survey. IEEE Access
39. The global internet phenomena report covid-19 spotlight. Available at <https://www.sandvine.com/phenomena>. Accessed 15 Nov 2022
40. Usman MA, Shin SY, Shahid M, Lövsström B (2017) A no reference video quality metric based on jerkiness estimation focusing on multiple frame freezing in vide streaming. IETF Technical Review 54(3):309–320
41. Vega Torres MT, Perra C, Liotta A (2018) Resilience of video streaming services to network impairments. IEEE Trans Broadcast 64(2):220–234
42. Video conferencing market size, share and industry analysis. Available at <https://www.fortunebusinessinsights.com/industry-reports/video-conferencing-market-100293>. Accessed 15 Nov 2022
43. Vučić D, Skorin-Kapov L (2019) The impact of packet loss and google congestion control on QoE for WebRTC-based mobile multiparty audiovisual telemeetings. In: International conference on multimedia modeling. Springer, pp 459–470
44. Vučić D, Skorin-Kapov L (2019) QoE evaluation of WebRTC-based mobile multiparty video calls in light of different video codec settings. In: 2019 15th international conference on telecommunications (ConTEL). IEEE, pp 1–8
45. Vučić D, Skorin-Kapov L (2020) QoE assessment of mobile multiparty audiovisual telemeetings. IEEE Access 8:107669–107684
46. Vučić D, Skorin-Kapov L, Sužnjević M (2016) The impact of bandwidth limitations and video resolution size on qoe for WebRTC-based mobile multi-party video conferencing. Screen 18:19
47. Wac K, Ickin S, Hong J-H, Janowski L, Fiedler M, Dey AK (2011) Studying the experience of mobile applications used in different contexts of daily life. In: Proceedings of the first ACM SIGCOMM workshop on measurements up the stack, pp 7–12
48. Yu C, Xu Y, Liu B, Liu Y (2014) Can you see me now? A measurement study of mobile video calls. In: IEEE INFOCOM 2014-IEEE conference on computer communications. IEEE, pp 1456–1464
49. Zeng K, Zhao T, Rehman A, Wang Z (2014) Characterizing perceptual artifacts in compressed video streams. In: Human vision and electronic imaging XIX, vol 9014. International Society for Optics and Photonics, p 90140

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.