VLQoE : Video QoE instrumentation on the smartphone

Ickin, Selim; Fiedler, Markus; Wac, Katarzyna; Arlos, Patrik; Temiz, Canberk; Mkocha, Khadija

Multimed Tools Appl
DOI 10.1007/s11042-014-1919-0

VLQoE: Video QoE instrumentation on the smartphone

Selim Ickin · Markus Fiedler · Katarzyna Wac · Patrik Arlos · Canberk Temiz · Khadija Mkocha

Abstract The usage of network-demanding applications is growing rapidly such as video streaming on mobile terminals. However, network and/or service providers might not guarantee the perceived quality for video streaming that demands high packet transmission rate. In order to satisfy the user expectations and to minimize user churn, it is important for network operators to infer the end-user perceived quality in video streaming. Today, the most reliable method to obtain end-user perceived quality is through subjective tests, and the preferred location is the user interface as it is the closest point of application to the end-user. The end-user perceived quality on video streaming is highly influenced by occasional freezes; technically the extraordinary time gaps between two consecutive pictures that are displayed to the user, i.e., high inter-picture time. In this paper, we present a QoE instrumentation for video streaming, VLQoE. We added functionality to the VLC player to record

S. Ickin (✉) · M. Fiedler · P. Arlos
School of Computing, Blekinge Institute of Technology, Blekinge, Sweden
e-mail: Selim.Ickin@bth.se

M. Fiedler
e-mail: Markus.Fiedler@bth.se

P. Arlos
e-mail: Patrik.Arlos@bth.se

K. Wac
Institute of Services Science, University of Geneva, Geneva, Switzerland
e-mail: katarzyna.wac@unige.ch

C. Temiz
Blekinge Institute of Technology, Blekinge, Sweden
e-mail: canberk.tmz@gmail.com

K. Mkocha
Department of Electronics and Telecommunications Engineering, University of Dar es Salaam, Dar es Salaam, Tanzania
e-mail: pocadicci@yahoo.com

Published online: 04 April 2014

a set of metrics from the user interface, application-level, network-level, and from the available sensors of the device. To the best of our knowledge, VLQoE is the first tool of its kind that can be used in user experiments for video streaming. By using the tool, we present a two state model based on the inter-picture time, for the HTTP- and RTSP-based video streaming via 3.5G. Next, we studied the influence of inter-picture time on the user perceived quality through out a user study. We investigated the minimum user perceived inter-picture time, and the user response time.

Keywords QoE (Quality of Experience) · QoS (Quality of Service) · Smartphone · Video · User interface · Human Computer Interaction (HCI)

1 Introduction

Wireless communication data traffic surges as more mobile applications and services are being used in daily life for information sharing, communication and leisure [4, 22]. The mobile terminals run cloud-based applications and services that rely on the performance of Internet, and they are are sensitive to the variations of the wireless link capacity and the expected load [2], which eventually influences the end-user perceived quality of a service or an application. Mobile handheld terminals are one of the most challenging platforms for assessing the end-user perceived quality of services provided on them, as they can be used anytime, anywhere, and in any context [22]. Although a traditional network-centered Quality of Service (QoS) approach can identify the key influential factors, such as packet loss or delay, on the end-user perceived quality, the recent research suggests that QoS has to be complemented with a more user-centric approach [1], e.g., Quality of Experience (QoE), to satisfy end-user requirements and expectations. The most recent definition of QoE is "*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. In the context of communication services, QoE is influenced by service, content, device, application, and context of use*" [32]. QoE assessment is preferably conducted close to the end-user, e.g., at the user interface, and by this way, it is less complicated as compared to the network to interpret the relationship between the impairments at the user interface and the subjectively perceived QoE.

According to the world-wide smartphone sales share, Android OS smartphones are the most widely used terminals (more than 70 % of world sales share) [63], and they are capable of running diverse "life coach" applications/services. Moreover, amongst the available services and/or applications, the most bandwidth demanding ones are the video streaming applications [4] as those have large size end-to-end data delivery requirements. The end-user perceived QoE of the video applications is highly depending on, amongst others, the quantity and the duration of the freezes [16, 50] during the video streaming sessions. There are existing approaches (e.g., YouTube) to improve perceived quality such as "download-and-watch-later" with the cost of download time. However, those approaches do not suit for real-time video streaming (e.g., live-broadcast of a soccer match), as the user expects to receive the video content more or less at the same time with the service broadcast time. In order to study QoE on the smartphone-based real-time video applications, it is important to consider smartphone user interface, as it is the location of the concrete evidences, e.g., temporal impairments, which can directly be perceived by the users. The reason for occasional short-term or long-term temporal freezes [66] in the video playout might be due to the impairments at low layers in the network stack. In previous studies [11], the network-traffic is studied within the network-layer in two different states: ON during a packet flow (e.g., burst), and OFF when there is no packet flow, i.e., zero throughput. It is necessary to complement the network-based two-state ON/OFF model with the user interface measurements and then find out the distribution of the inter-picture times in the user interface, which later help to understand the relationship between different layers in the network stack. Yet, and to the best of our knowledge, there is no smartphone-based QoE tool that can simultaneously record potential QoE metrics at different layers of the network stack together with the sensor data. The contributions of this article are summarized as follows:

- We develop and present the VLQoE tool that enables video QoE user studies on Android-based smartphones. The collected data can help in design of future user-centric video streaming applications.
- We present an exponential ON/OFF state model based on temporal impairments of a video stream throughout measurements at the user interface during Enhanced 3rd Generation (3.5G) based real-time video streaming on the smartphone.
- We present the influence of the temporal impairments such as freezes that are detected objectively at the user interface on the end-user perceived QoE.
- We present the minimum perceived picture freeze duration for the users in the user study.
- We present the user response time with respect to the freezes during video streaming sessions. This helps to identify approximate user tolerance levels based on the temporal impairments.

In this paper, firstly, a QoE instrumentation of the smartphone-based VLC player [56], VLQoE, is introduced. Then, the studies conducted with VLQoE, which are based on the inter-picture time, are presented in two parts. The first part of the study involves an in-depth analysis of the inter-picture time metric, where the inter-picture times of the video is studied with a two-state (ON/OFF) model. We refer to ON state during a smooth video playout; and refer to OFF state during a video picture freeze. The second part of the study involves the subjective study with the focus on the inter-picture time metric quantifying a picture freeze. The influence of the inter-picture time on the end-user perceived quality of the video, measured with five-level Absolute Category Rating (ACR) scale (1-bad, 2-poor, 3-fair, 4-good, 5-excellent), is presented. The user response time, i.e., the time it takes for the user to react to a long picture freeze, are investigated in order to understand the user tolerance levels to the inter-picture times.

The remainder of this paper is structured as follows. The state of the art on the application layer video streaming protocols, and QoE assessment methods are recalled in Section 2. We also review the related work done on the video QoE in this section, and discuss a set of QoE metrics together with already existing tools and QoE management mechanisms. Section 3 presents the definition of proposed metrics as well as the detailed description of VLQoE tool. Section 4 describes the experiment testbed and the experiment methods of our study. Section 5 presents the results of the two-state ON/OFF modeling based on the inter-picture time, and also the results of the subjective study that identifies the relationship between the inter-picture time and the end-user perceived quality. The limitations in the study are given in Section 6. Section 7 concludes the paper, and it is followed by a set of items for future work in Section 8.

2 Background and related work

This section first gives background information on a set of application-layer video streaming protocols that are used in our work. Next, the definition and the application of the QoE assessment method called Experience Sampling Method (ESM), employed to acquire user-perceived quality (user ratings on the ACR scale), is given. Furthermore, the related work with respect to the factors influencing the perceived QoE, the corresponding existing QoE assessment tools, and QoE management mechanisms in video streaming are explored.

2.1 Video streaming protocols

Video clips on the mobile terminals can be streamed via two main application layer protocols: Real Time Streaming Protocol (RTSP) [45] or Hypertext Transfer Protocol (HTTP) [18]. The RTSP video streaming is mainly based on the unreliable User Datagram Protocol (UDP). The data packets are sent from the server to the smartphone as they are generated by the video/audio codec [2]. The flow control is done by the RTSP in the application layer with periodic control messages being sent from the client to the streaming server. RTSP is implemented for live multimedia streaming. On the other hand, the HTTP video stream is based on the reliable Transport Control Protocol (TCP) and the packets arrive at the client (e.g., sink) in segments. The size of the segments depends on the available bandwidth and vary with the window size of the TCP packets, which in turn influenced by Round Trip Time (RTT) and Bandwidth Delay Product (BDP) [28]. HTTP streaming is used for multimedia streaming and it commonly works based on progressive download, i.e., the video is downloaded from a webserver and temporarily stored at the cache of end-user device and then displayed on the screen. The periodicity of the control messages in RTSP streaming is longer than in the HTTP streaming. The characteristics of each protocol with respect to the network layer, e.g., burst size and duration, are hypothesized as different, and this difference is expected to influence the application performance, and eventually the end-user perceived QoE. Our aim in this paper is not to compare the RTSP and HTTP streaming, but to consider both protocols in order to intensify the representativeness of video streaming data.

2.2 Experience sampling method (ESM)

ESM is a research method that is used to assess phenomena at the time they occur, from the human perspective, in order to maximize the validity of the data [23]. Lutille et al. in [23] claimed that the best time to ask the user about the preferences and/or the user needs is in the midst of the actual activity being closely inspected. Moreover, an interface that could build up an awareness and understanding of the users' preferences is suggested. Previously in [22], we have employed ESM in order to identify the most influential factors on QoE of smartphone-based applications. Thereby, it is possible to consider the same method in widely used mobile applications, particularly video streaming. This helps to identify the end-user perceived quality on spot in ad-hoc manner, while delivering the video streaming service to the user, and to relate it to factors influencing it. We employed ESM to the VLC player in order to receive user feedback *in-situ*, while the users are watching the videos on the smartphone.

2.3 Related work on video QoE

We first recall the related work areas on the factors influencing video QoE.

Related work on influential metrics There are plenty of performance indicators to consider while assessing the QoE of a particular service. The influence factors on video streaming application in Internet are categorized in four categories: context (environment, social/cultural background, purpose of using the service), user (expectations of user, memory and recency effects, usage history), system (transmission network, devices, screens, video buffering strategies), and content level (video codex, format, resolution, duration, content, type, motion pattern) [17]. In our work, we obtained data regarding the environment and the user nationality to get cultural information within *context* category; service usage history within the *user* category; screen, device, and network type within the *system* category; and video resolution, duration, and content within the *content level* category.

The performance of 3.5G (also known as HSDPA) and Wireless Fidelity (WiFi) networks is studied by Junxianet et al. [19], and it is stated that the video download patterns depend on the link type, e.g., 3.5G or WiFi, the smartphone type, the protocol, and the video player. The different factors usually have different effects on the user perceived QoE, including, but not limited to, the type of end user device, the communications protocols behind the application as well as the transport methods in providing the video streaming service [8]. According to Bang et al. [61], TCP generally provides good streaming performance, when the achievable TCP throughput is roughly twice the media bitrate. Frame rate, the periodicity of the displayed video pictures on the screen, and the network bandwidth are important influential factors on the video QoE [33, 59]. However, it is also important to quantify a "good performance" from the user perspective based on measurement layers of the network stack. Yang et al. [64] and Quan et al. [44] studied the frame rate metric, i.e., the number of displayed per second, to quantify jerkiness during video streaming. VLQoE can also record the frame rate, yet we focussed only on the inter-picture time metric to be used in duration-based ON/OFF modeling. Hossfeld et al. [16] studied the influence of a stalling event, namely *picture freeze* (also called frame freeze) on the end-user perceived QoE by user ratings on the five-level ACR scale, while considering the duration and the number of occurrences. Based on the exponential QoE mapping functions to the stalling frequency and the durations, they stated that one freeze with two seconds duration within 30s-long video clip maps to a Mean Opinion Score (MOS) value 3 in desktop settings. Despite, further studies on smartphone settings and with longer video clips without any restrictions of video length need to be studied to obtain more reliable models. Van Kester et al. [52] have investigated the average end-user perceived QoE (via five-level MOS) on 20 seconds long video sources based on the number of freezes and concluded that the acceptable freezing time (MOS > 3.5) is 360 milliseconds, and they noted that the frequency of the freezes must also be taken into consideration in QoE assessment. In addition, no significant difference in the video QoE was observed between when the video is stopped with or without skipped frames.

The waiting times, i.e., the number and the duration of delays, for web browsing and HTTP video streaming have been studied in [9], and the authors have presented a logarithmic relationship between the waiting time and the QoE. They state that the longer the waiting time, the less satisfied the user becomes. The location of waiting time e.g., before/after the video has started displaying, is important as such. Hossfeld et al. [17] showed that initial delays before the video has started has no severe impact on QoE. Yet, we did not study the initial delays in our paper.

In addition, the difference between the QoE models based on the application-level and network-level page loading times (non-linear relationships and complex) are emphasized. The non-linear relationship between the video bitrate and the perceived quality by the user is stated in [36]. In [3], a No-Reference QoE assessment measure, A-PSQA, is proposed as a performance evaluation method, based on the analysis with respect to the packet-loss percentage. Zinner et al. [65] stated that the Structural Similarity Index Model (SSIM) can be used to quantify the influence of packet loss on the QoE; while the Peak Signal-Noise Ratio (PSNR) can not. Furthermore, by using SSIM, the authors also showed that the low-resolution videos obtain high QoE as compared to high resolution disturbed videos. In our work we conduct real-time measurements on the client device, thus, we do not consider SSIM as it measures the image similarity, while comparing with the undistorted version as a reference [60].

Related work on the tools PC-based based measurements on the VLC player were conducted in [58], where the video/audio buffer utilization based on the Realtime Transport Protocol (RTP) and UDP streaming is studied. The drawback of that study is that it does not consider the picture freezes and the corresponding end-user perceived quality. YoMo [49] is a java tool with a Firefox plugin and it estimates the video player buffer status on application layer. Pytomo [27] is a Python-based client-level tool that analyzes the playback performance of particular video sources on YouTube. It obtains the cache URL and IP addresses of the hosting video server, ping/download/playback statistics, initial buffer, stalling event, and buffering duration of the selected videos. These tools consider only YouTube videos and they are not suitable for smartphones. Vishwanath et al. [55] conducted QoE experiments using the ns-2 simulator and focused on application layer metrics such as the modulation coding schemes, the video bitrate on short duration video clips. Migliorini et al. [37] presented QoE results based on the experiments done with a packet-based simulator with the focus on PSNR and channel conditions. There exists the Perceptual Evaluation of Video Quality (PEVQ) tool [42] that estimates end-user perceived quality, by using averaged values of user ratings, i.e., MOS, with respect the spatial impairments. However that tool fails to deliver good estimations in case of picture freezes [38]. In contrast to the already existing tools, VLQoE can assess video QoE in real-time and on user's own smartphones in any context.

Related work on QoE management mechanisms Mekovski et al. [34] proposed a QoE control framework for adaptive video streaming that takes into account the temporal and spatial characteristics of the video. Similarly, Latre & Turck's work on QoE management in [31] used the already existing QoE metrics and techniques such as traffic flow adaptation (by changing transport configuration), admission control (by admitting/blocking new connections to avoid congestion), and video rate adaptation (by changing the actual video quality level). Their work focuses on the low layer design for autonomic management architecture, and emphasizes the need to design control loops at higher levels. In our work, although, we have not yet employed such QoE management mechanism, it can be employed in the future evolution of VLQoE as it is potentially possible by training it with the user data to automatically detect the future artifacts.

3 Inferring the video QoE on the smartphone

This section presents the studied metrics at the user interface in our work, and explains in details how they are related to QoE.

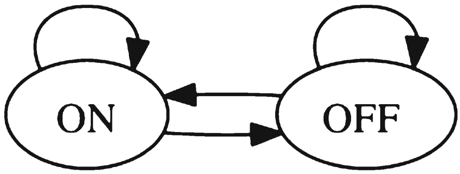


Fig. 1 Illustration of transitions between the two states

3.1 Acquisition of metrics

In this section, we studied the metrics that are measured at the user interface to investigate the performance of video streaming on the smartphone. The primary metric that we studied is the inter-picture time (D_p). Afterwards, on the basis of D_p, two state modeling is employed in the first part of our study. This is followed by a second part including a user study, where we investigate the minimum perceived D_p for each user. We conclude this section with a study on the user response time that reflects the time it takes for the user to react to a perceived video *freeze*.

3.1.1 Inter-picture time (D_p)

Video is delivered to the smartphone as a series of pictures and then displayed on the smartphone screen. The inter-picture time D_p, the time gap between two consecutive pictures displayed on the smartphone screen, are calculated as in Eq. 1. $T_\mathrm{p}(k)$ is the timestamp when the k_{th} picture is displayed on the smartphone screen.

$$D_\mathrm{p}(k) = T_\mathrm{p}(k) - T_\mathrm{p}(k-1) \tag{1}$$

3.1.2 Two-state (ON/OFF) modeling

In the first part of our study, we investigated the video streaming with the assumption that the video stream follows a two-state model, e.g., ON/OFF, throughout a video streaming session, as shown in Fig. 1. The *ON* (time interval of a smooth playout) and *OFF* (time interval of a freeze) states are defined based on the D_p metric. We considered an exponential two-state model for the ON and OFF durations as the modulating ON/OFF process can be assumed to be memoryless. In [46], 100 ms is defined as the maximum tolerance threshold for the user to feel that the system is reacting instantaneously. Thereby, we chose 100 ms as the state boundary between ON and OFF.

The Maximum Likelihood Estimation (MLE), developed by R.A. Fisher, is a standard approach to parameter estimation, i.e., to find the probability distribution that makes the observed data most likely [40]. It has optimal properties in estimation such as sufficiency, consistency, and efficiency. Thus, we found the MLE of the durations spent in each state (ON/OFF) is calculated as in Eq. 2–3. $L(\lambda)$ is the likelihood function for exponential distributed samples; n is the number of samples; $\bar{x}$ is the mean of the samples; and $\hat{\lambda}$ is the MLE.

$$L(\lambda) = \lambda^n \exp(-\lambda\, n\, \bar{x}) \tag{2}$$

$$\hat{\lambda} = \frac{1}{\bar{x}} \tag{3}$$

In our study, each sample x, represents a duration instance of a state. Thereby, two different sets of instances are obtained, i.e., ON and OFF.

3.1.3 Inferring QoE based on inter-picture time

There are objective and subjective QoE assessment metrics. Using objective metrics is less time consuming as they can assess the quality without the existence of actual users. However, as the objective metrics avoid the human subject expectations and context, they may be less accurate as compared to the subjective ones [25, 62].

One of the most commonly used subjective metrics to assess the subjectively perceived quality of a video is the five-level MOS scale [24]. R. Schatz et al. [48] recommended Absolute Category Rating with Hidden Reference (ACR-HR) test method with five-point quality scale as the mean value and the confidence intervals of the results from different methods do not deviate much. However, the averaged value of the user ratings gathered in five-level ACR scale has a set of drawbacks such as the weakness in the definition of quality as summarized in [20, 48]. In addition, it is influenced by the "forgiveness effect" [53]. As MOS might smooth out the low qualities, Hossfeld et al. [15] introduced Standard Deviation of Opinion Scores (SOS) as it can better reveal the fluctuations of users' perceived quality. The Maximum Likelihood Difference Scaling (MLDS) approach used by Menkovski et al. [36] relies on relative quality assessment that utilizes the direct comparison mechanisms rather than rating. It is indeed a good method to understand the perceived quality differences. However, we still believe that it is challenging to assess whether a "better" quality is "good-enough", or as compared to the quality at which point in time, especially knowing that we are running user experiments with a single long video that is streamed in real-time.

In our study, we used ACR while decreasing the degree of freedom of quality to the visual temporal impairments, i.e., stalling events. Firstly, we studied all the user indications including the user ratings on the five-level ACR scale and the freeze indications. We calculated the maximum inter-picture time values (D_{pmax}) amongst all D_{p} values during the time gap between two consecutive user indications, T_i and T_{i-1}. Thus T_i and D_{pmax} have a one-to-one relationship. We denote the display timestamp of the picture with the maximum display time (D_{pmax}) as $T_{D_{\text{pmax}}}$. In addition, we calculated the time between a user indication and $T_{D_{\text{pmax}}}$ as in Eq. 4.

$$\Delta(T_{i-D_{p\max}}) = T_i - T_{D_{\text{pmax}}} : T_{i-1} < T_{D_{\text{pmax}}} < T_i \tag{4}$$

First low user rating time ($T_{\text{LowRating}}$) and a number of alarms In the literature [17, 50], it is stated that the influence of stalling event on QoE shows significant differences with the duration of the video length. Therefore, we studied the influence of stalling events while considering the streaming duration until the first low rating is received. In the impairment definition of five-level ACR scale, the ratings below 4 reflects annoyance of the user [24]; and in the voice quality, MOS = 4 is still considered as acceptable [57]. In this work, we considered the *Low Rating* as the ratings less than 4 (UR 4) together with the underlying temporal reasons. We studied $T_{\text{LowRating}}$, i.e., the time it takes from the start of the video until the time of the first low user rating, e.g., UR $<$ 4, per video session in more details. Next, we focused on the low user ratings, i.e., UR $<$ 4, and their underlying temporal reasons in the application. We studied the frequency of the freezes and the corresponding subjectively perceived quality. Pictures with the display duration higher than 100 ms were considered as freezes. Thereby, the number of instance with the inter-picture (D_{p}) values higher than 100 ms are considered as freezes, namely the *alarms*. In parallel, we recorded the streaming duration until when the first low user rating ($T_{\text{LowRating}}$) has been received. We investigated those for each user in both RTSP and HTTP streaming experiments.

3.1.4 Minimum perceived inter-picture time

Pastrana et al. [54] stated that a single 200 ms long frame freeze was detected by all users in a 10 seconds-long sequence with CIF resolution. The perceived video freeze depends on the duration of the total duration of the video sequence. Staelens et al. [50] used short and long video sequences that comprise video freezes, which last up to 320 ms–400 ms range, and the authors claim that they have obtained different results with respect to the percentage of the detected video freezes when the subjective tests were done with videos with different durations. Therefore, the topic is still challenging, and it is important to understand the minimum perceived inter-picture time of the end-users during a video stream on the smartphone, which directly relates to the freezes and consequently to the end-user perceived QoE. Although, the visual perception depends on the context, e.g., time perception; memory effect [21]; illumination; content [36], we assumed that subjects indicate the freezes as they perceive them. We conducted a user study; collected user data from the subjects. The detailed procedure of the user study is given in Section 4.2. In summary, we asked the user to indicate a perceived freeze by using the "Freeze" button located on the user interface during the video playout.

Finding out the exact temporal impairment, i.e., pinpointing the exact D_{p} value for a given indicated freeze by a user, is challenging. In other words, there might be many high (higher than nominal 40 ms for 25 fps video) $D_{\mathrm{p}} > 40$ ms values which the users might react to. However, the user might not have reacted to all of them for some reason. An example scenario is illustrated in Fig. 2. In the sketched scenario, suppose $T_{\mathrm{p}}(k)$ is the timestamp of the last picture that is displayed; and $T_{i_{\mathrm{freeze}}}$ is the timestamp of the subject's i'th indication. The subscript "freeze" stands for the type of user indication. It is hard to interpret the reason for the freeze indication, i.e., upon which D_{p} values out of $\{D_{\mathrm{p}}(k-n-1), \ldots, D_{\mathrm{p}}(k-1), D_{\mathrm{p}}(k)\}$ was the freeze indicated for within some interval C_i'. Thereby, in order to guarantee the minimum perceived inter-picture time of a user, we analyzed the perceived user inter-picture time with the *MinOfMax* approach as explained via the following.

MinOfMax Firstly, for all freeze indications at $T_{i_{\mathrm{freeze}}}$ by a user, we calculated the maximum D_{p} value during $T_{i_{\mathrm{freeze}}}$ and T_{i-1}, i.e., C_i', as in Eq. 5. Note that T_{i-1} is not necessarily a freeze indication timestamp; it can be the timestamp of any indication amongst the six user feedback choices (one, two, three, four, five, and freeze). We assumed that the user perceives and indicates a new freeze after the previous indication. This procedure is repeated for each freeze indication. Secondly, the minimum of the maximum values were calculated in Eq. 6,

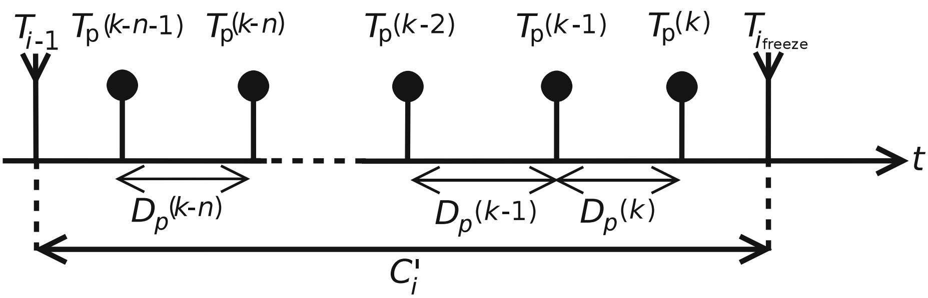


Fig. 2 Illustration of a scenario where it is hard to interpret what the user rates

which we called this value as the Minimum Perceived Inter-picture time ($D_{\mathrm{p}_S\mathrm{min}}$) of user S while considering two video streaming sessions (HTTP and RTSP). This way, the risk of wrong interpretation of the freeze indication by a user is reduced.

$$D_{\mathrm{pmax}_i} = \max\{D_\mathrm{p}(k)\},\ \forall k \text{ within the time interval } C_i' \tag{5}$$

$$D_{\mathrm{p}_S\mathrm{min}} = \min\{D_{\mathrm{pmax}_i}\}\ \forall i \text{ of subject } S \tag{6}$$

Yet, we defined how the minimum perceived inter-picture time for each user is calculated in this work. In addition, we believe that the time it takes for a user to react to the stalling event, referred to "User Response Time", needs to be investigated.

3.1.5 User response time

The influence of the waiting times to the end-user perceived quality is presented in previous work [9]. However, the duration until the user perceives a freeze and reacts to it has not been studied within the scope of QoE. This is important to know as QoE assessment depends on multidisciplinary parameters such as time perception and "inner-clock" [14] of the user. Musser [39] states that the human consciousness lags 80 ms behind the actual events. Studying the time it takes for a user to react (to press the "Freeze" button) to a freeze event, is complex and depends on many factors. The authors of [14] studied the time perception in depths, and state that the actual time is different than the subjectively perceived time. However, in this study, we assume that the users indicate a freeze as they perceive it on spot.

Therefore, in addition to the items discussed in Section 3.1.3 and Section 3.1.4, we investigated the user response time in two scenarios. The first scenario is when a user indicates a past freeze during a smooth playout that we will refer as *short*; and the second scenario is when the user indicates a freeze during a stalling event that we will refer as *long*.

First scenario – short Let's assume that $T_{D_{\mathrm{pmax}}}$ is the timestamp of the picture with the longest display time amongst all D_p values within the interval C_i' in Fig. 3. Assuming that the user reacted and pressed the freeze button with respect to $T_{D_{\mathrm{pmax}}}$, the user response time is calculated, as shown in Eq. 7.

$$D_{\mathrm{response}_{\mathrm{short}}} = T_{i_{\mathrm{freeze}}} - T_{D_{\mathrm{pmax}}} : T_{D_{\mathrm{pmax}}} \in \{T_\mathrm{p}(k)\} \forall k \text{ within } C_i' \tag{7}$$

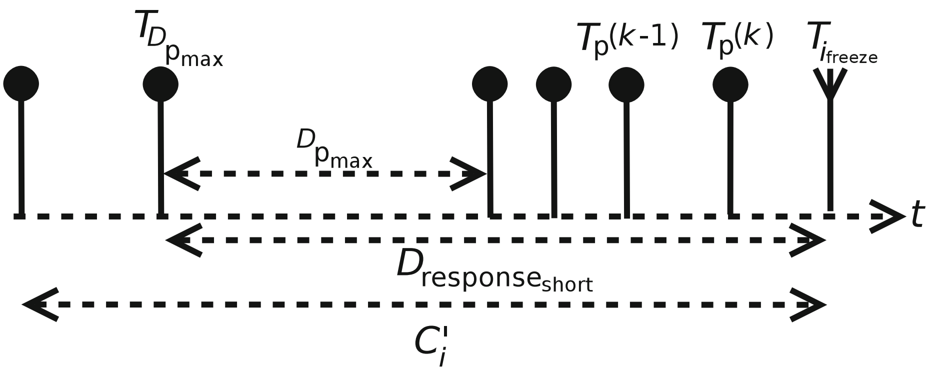


Fig. 3 Illustration for the calculation of $D_{\mathrm{response}_{\mathrm{short}}}$

Indeed, $D_{\text{response}_{\text{short}}}$ is considered as $\Delta(T_{i-D_{p\max}})$ as previously defined in Eq. 4, if the indication at T_i is a *freeze* indication.

Second scenario – long We study the freeze indications by the user for the *long* freezes, i.e., pauses [66] as this metric might signify the *end-users tolerance level* to the long-term freezes. This scenario is when the user intervenes the video stream by pressing the freeze button at $T_{i_{\text{freeze}}}$, before the next picture k is displayed at $T_{\text{p}}(k)$, as shown in Fig. 4. $D_{\text{response}_{\text{long}}}$ is calculated as the time gap between the display time of the frozen picture and the freeze indication time by the user as in Eq. 8.

$$D_{\text{response}_{\text{long}}} = T_{i_{\text{freeze}}} - T_{\text{p}}(k-1) \tag{8}$$

3.2 VLQoE tool description

We implemented the VLQoE tool, as a version of open-source VLC Media Player, while adding extra functionalities for video QoE assessment.

3.2.1 VLC media player

VLC is a packet-based media player [56] developed by Video LAN, first released in year 2000, available as open source, with GNU General Public License (GPL) and GNU Library General Public License (LGPL) software licenses. Amongst other available players, VLC is available for the most variety of Operating System (OS)'s, and also for Android-based terminals. According to the sourceforge.net, VLC has been downloaded more than 17 billion times [51]. VLC supports many video formats including *mp4*; and streaming protocols including HTTP and RTSP.

We added new functionalities to the VLC player source code in order to record picture display times and the user data with the context and the perceived quality ratings during video streaming. These functionalities are applied to different player components as detailed in Table 1, which are grouped in Application; User Interface; Network and Physical; and Other.

3.2.2 VLQoE tool

Although, the functionalities are implemented within VLQoE, in this study we do not investigate all of the metrics collected by VLQoE. We study the timestamp of the displayed

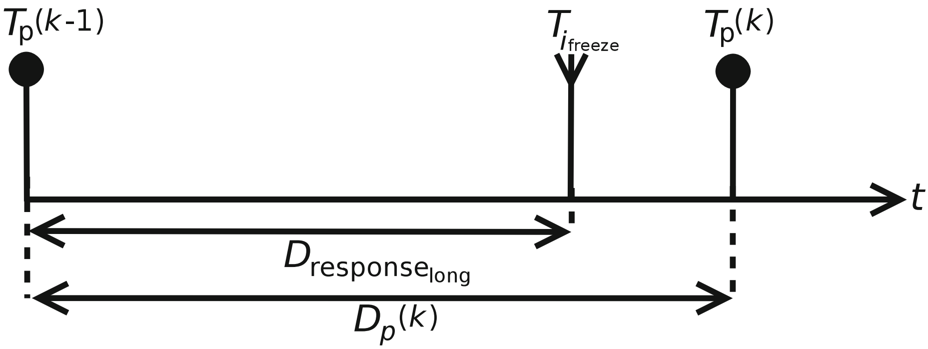


Fig. 4 Illustration for the calculation of $D_{\text{response}_{\text{long}}}$

Table 1 The VLQoE records the listed parameters with the corresponding timestamps

Components	Parameters
Application	re-buffering events
User Interface	displayed video picture(frame), user controls, freeze indication,
	user rating (UR), display orientation and brightness, screen touch events
Network & Physical	interface type, service provider, signal strength(RSSI), packets/sec
Other	GPS coordinates, battery level, unique device ID

video pictures; the timestamp of the user ratings; and the timestamp of the freeze indications by the users. The detailed description of the collected metrics is listed in the following subsections.

3.2.3 Application layer

The timestamp of the rebuffering events at the application are recorded within the VLQoE.

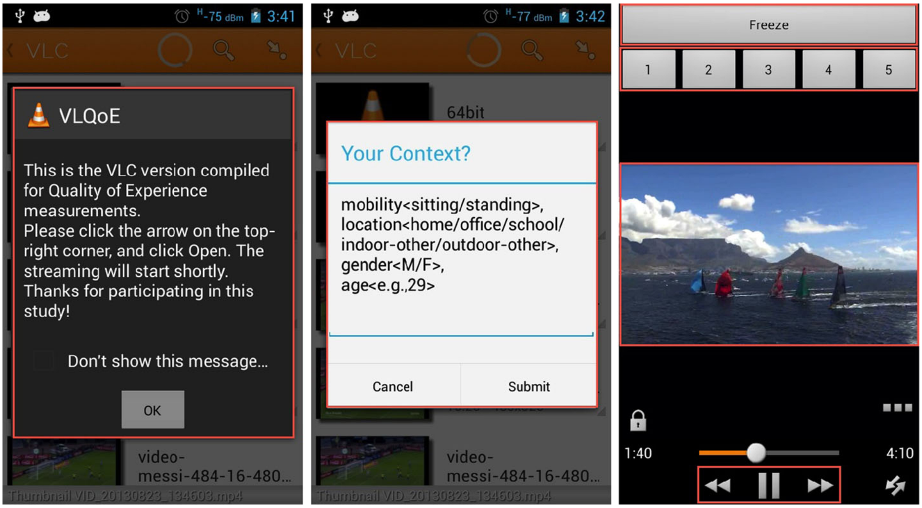
Rebuffering events The rebuffering of the video content causes undesired picture freezes in the midst of video streaming [9]. Thus, the rebuffering events upon the playback buffer starvation are recorded with the corresponding event timestamps.

3.2.4 User interface

The original beta version of the smartphone-based VLC player consists of the video pane that displays the video; a set of video control buttons such as play / pause / rewind / forward; and the additional buttons that we added in order to fetch the user feedback on spot. This is especially important from the ESM perspective, and minimizes the memory effect, i.e., the forgotten and/or neglected freezes as well as low QoE, during the video playout. The snapshots from the user interface of the player are presented in Fig. 5. When the user launches the player, s/he sees a welcome message (see Fig. 5a), and s/he is required to click the "OK" button. Afterwards, the user is asked to type in and submit information regarding her/his mobility; location; gender; and age (see Fig. 5b).

Video pane During a video stream, the timestamps of the displayed pictures on the device display (after decoding and rendering) are recorded. Then, the D_p values are calculated. The display is the video pane in Fig. 5c, and it is 196 × 117 pixels with a fixed vertical display view.

User rating buttons (1, 2, 3, 4, 5) The user is instructed to press one of the horizontally aligned five (based on ACR scale) user rating (UR) buttons (see Fig. 5c) whenever the user feels like rating the quality. This helps to match the underlying application metrics to the application QoE, however it does not give information whether a possible low QoE is caused by a temporal impairment. Therefore, we need an additional button that helps the user to indicate a freeze when required while the video is streaming. When any of the rating buttons is pressed by the user, the event is recorded together with the corresponding timestamp.



(a) Welcome message (b) Context, Demographics (c) Freeze, UR, Display, and Control

Fig. 5 Snapshots from the VLQoE player

Freeze button This is a button that is located on top of the user interface (see Fig. 5c) for the users to press whenever s/he perceives a picture freeze on the video display. When the freeze button is pressed by the user, the event is recorded together with the corresponding timestamp.

User control buttons There are three buttons (play/pause, rewind, forward) at the bottom of the video pane, as shown in Fig. 5c. When these buttons are pressed by the user, the events are recorded together with the corresponding timestamp.

Screen display The orientation (vertical/horizontal) and the brightness of the display are recorded with the corresponding timestamps.

Screen touch events The users' touch events on the display, while streaming video, are recorded with the corresponding timestamps. This is especially important to understand how the user reacts and/or interacts with smartphone upon a temporal impairment.

3.2.5 Network and physical layer

Packets The received and transmitted number of bytes to/from the smartphone are recorded every second with the corresponding timestamp.

Interface type The active wireless interface of the device (3.5G or WiFi) is recorded for once when the application is first launched.

Service provider The SIM card operator name is recorded when the application is first launched.

Signal strength The signal strength measured at the active wireless interface (3.5G or WiFi) is recorded for once, when the application is first launched. Next, the signal strength is recorded sporadically when the circumstances change, i.e., only when a different WiFi RSSI is triggered by the Android OS on the smartphone.

3.2.6 Other

Other recorded metrics belong to the following components:

GPS Users' GPS coordinates are recorded when the application launches for the first time. The minimum time gap between the GPS polling is set to one minute. In addition, the minimum distance change with respect to the previous location is set to 10 meters. Based on these conditions, the software records the new coordinates with the new timestamp. The user should manually enable the GPS and Network (via cell tower locations) to calculate the location on the Settings menu of the smartphone. If not enabled, the software pops up a notification and encourages enabling it when the application is launched for the first time. The type of the active GPS component (Network/GPS) is also recorded. This way, it is possible to get the mobility information of the user while video streaming.

Battery level The battery level is recorded on the device when the battery level changes. The battery change is triggered by the Android OS of the smartphone upon the battery change detection. The events are recorded with the corresponding timestamps.

Device ID The unique device ID of the smartphone is recorded for once when the application is launched for the first time. The names of all data files are labeled with the device ID. This is helpful to distinguish the users.

4 Experimental settings and methods

We have used the VLQoE tool primarily for performance analysis with respect to the D_{p} metric while considering the user feedback. We conducted the experiments in two parts. The first part consists of a study without user involvement to establish a ground truth and to understand the video streaming characteristics with respect to the ON/OFF modeling. The seconds part is the user study designed to capture the circumstances when the video QoE has occurred, i.e., UR and freeze indications with respect to the inter-picture time (D_{p}) values.

The first part of the experiments is the "ON/OFF Study" and it is introduced in Section 4.1; and the second part of the experiments is the "User Study" and it is introduced in Section 4.2.

Video content used in both studies We used the same video in both parts; the streamed video content was a 250.44 seconds-long water-sports theme video clip with 6251 pictures, nominal frame rate of 25 fps, and was encoded with a nominal bitrate of 1000 kbit/s. The video consists of various scenes comprising cheering fans, interviews with the sportsmen, racing scenes of the sailing boats. The Scene Complexity (SC) is a metric as the average of the spatial and the temporal details of all frames of a video clip, as shown in Eq. 9; where $F(n)$ is the luminance channel of the nth video frame; the temporal complexity (TI) is calculated as the root mean square (RMS) of frame-to-frame image changes as in Eq. 10;

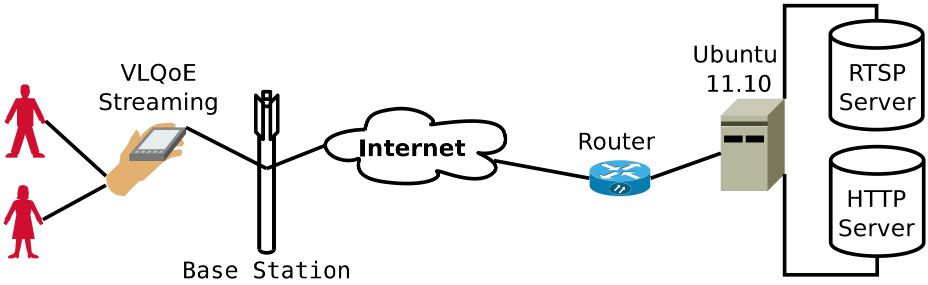


Fig. 6 Experiment testbed

and the spatial complexity (*SI*) is calculated as the RMS of the sobel filtered luminance channel of the frames, as shown in Eq. 11 [10, 26].

$$SC = \log_{10}(\text{mean}_n[SI(n) \cdot TI(n)]) \tag{9}$$

$$TI(n) = \text{RMS}_{\text{space}}[F(n) - F(n-1)] \tag{10}$$

$$SI(n) = \text{RMS}_{\text{space}}[\text{Sobel}(F(n))] \tag{11}$$

We calculated the *SC* of our video clip as 7.04 with a high *TI* value of 67, and *SI* value of 16. Based on the rough estimates of Ramos et al. [7], a video with *TI* greater than 38 could be considered as a video with high motion, but with some exceptions. In addition, we observed that our calculated *SC* value is comparably higher than all of 25 ANSI standard test scenes [10].

Common experiment settings The VLQoE software is installed to Samsung Galaxy S (Android 4.2.2). The video was located at a dedicated Ubuntu 11.10 streaming server within the BTH university campus (City of Karlskrona, Sweden), and it was streamed to the smartphone via RTSP and HTTP protocols (Fig. 6). The streamed videos were displayed on the smartphone screen with a resolution of 196 × 117 pixels.

4.1 ON/OFF study (Part 1)

In the ON/OFF study, our main goal is to investigate and obtain a model with respect to the D_{p} metric. During the first part of our experiments, we tested video streaming at different locations (comprising university library, apartment basement) in Karlskrona, and tried to find places where the video streaming quality was poor. Our purpose was to replicate the worst-case scenarios with distortions during a real-time video stream. We conducted this part without the user involvement (no user-interaction), and exactly the same video content was streamed consecutively using two streaming protocols, e.g., RTSP and HTTP via 3.5G. A test supervisor started the video stream and did not press any other control buttons, as this part of the experiments aims to model D_{p} at the user interface during network-based video playout. We performed two experiments, one for each protocol; and for each experiment we did 30 repetitions.

Based on the D_{p} values, the durations spent on the ON and OFF states are investigated. The state boundary between ON and OFF states is defined as 100 ms: the video is considered being in the ON state when the inter-picture time is less than 100 ms; and it is assumed in the OFF state when the inter-picture time is greater than or equal to 100 ms. We collected the ON and OFF durations in two different data sets called ON and OFF, and investigated the distribution as well as the Maximum Likelihood Estimation (MLE) of the durations within the states by using the Matlab. We did not apply extra artificial disturbances on the link between the streaming server and the smartphone, but instead collected data at a similar context that the users might experience in real life.

4.2 User study (Part 2)

In the second part of our study, we met with 30 subjects at various locations in Karlskrona, Sweden, and asked the users to watch the streamed video on the smartphone. A more detailed background description on the selected subjects is presented in Table 2 in Section 5.2. Each user first watched the RTSP-based video and then the HTTP-based video. In total, we conducted 60 (30 users $\times$ 2 protocols) user experiments.

During the experiments, we tried to stick to the ITU-T P.910 [25] recommendations for video quality in video conferencing and video-on-demand as much as possible. However, due to the inconsistency in the definition of experiment settings at different standards documents (ITU-R BT.500, ITU-T P.910/1/2, ITU-R BT.1438), this is still a challenge [48]. One of the strengths of our study is that we run the subjective tests in peoples' natural daily life settings. In our work, the test supervisor did not set any restrictions such as the distance between the phone and the user. Instead, the users were asked to hold the smartphone at a comfortable distance and take a comfortable position in a silence room with convenient illumination level, i.e., at a familiar physical space in daily life. The users were asked to press the "Freeze" button anytime to indicate an evidence of a visual freeze on the display, whenever s/he recognizes a freeze on the video. In addition each user was encouraged to rate the temporal quality based on the five-level MOS scale, while pressing one of the five user rating buttons at her/his own will during the playout.

As we study only the influence of temporal impairments on the users, not all QoE aspects were considered in this study. For example, we used an additional "Freeze" button on top of the ACR scale to collect the user reactions to the stalling events. ITU-P.910 [25] recommends that the source signal (recording environment/system, scene characteristics, spatial/temporal information of the video) needs to be taken into consideration to the configuration at the user experiments. The standard is of high importance in analysis of content level influential factors on QoE. However, in this study, our focus is more on the context, system, and the user level aspects and we consider the temporal impairments on a single video source with a length of more than four minutes long in realistic smartphone settings. The traditional methods ask for user feedback on the quality such as ACR, after the video is completed, however our aim in this work is to study and sample the experience on-the-fly as the video continues to stream in real-time. The users streamed exactly the same video content and the video is displayed on the video pane with 196 $\times$ 117 pixels resolution. We muted the videos to let the users focus only on the visual impairments. This way, the influence of other impairments such as the discontinuity of sound is minimized. For simplicity, the orientation (vertical) and the brightness of the display was kept constant for all users. Hossfeld et al. [16] stated that stalling of video stream disturbs the video experience of the user independent of the actual video characteristics. Yet, we have studied only one video content in this study, but we schedule to replicate the user experiments with different video characteristics.

Table 2 Participants demographics and statistics summary. Unavailable data is depicted with "-" sign

S	Nationality	Net. Stream per day	Stream via	OEF via	Interview MOS	VLQoE UR RTSP, HTTP
1	Pakistan	30 min	Both	3.5G	4	4(50%), 4(46%)
2	Pakistan	$>$ 1 hour	3.5G	3.5G	4	3(32%), 4(34%)
3	Iran	5 min	WiFi	3.5G	4	-, 4(63%)
4	Turkey	$<$ 2 min	WiFi	3.5G	4	4(50%), 4(28%)
5	China	Never	Local	-	-	4(33%),4(25%)
6	China	$>$ 1 hour	WiFi	WiFi	5	5(59%),-
7	Turkey	30 min	Both	3.5G	4	-, -
8	Sweden	$>$ 1 hour	3.5G	-	4	5(26%), 3(34%)
9	China	5 min	WiFi	3.5G	4	4(28%),2(25%)
10	India	30 min	WiFi	-	4	4(36%),4(33%)
11	Sweden	30 min	Both	3.5G	4	5(62%),4(42%)
12	Sweden	5 min	WiFi	3.5G	3	4(33%),5(38%)
13	Sweden	Never	-	3.5G	3	5(27%),-
14	India	$>$ 1 hour	Both	WiFi	4	4(46%),-
15	Pakistan	30 min	3.5G	3.5G	4	-,5(51%)
16	Sweden	5 min	3.5G	-	4	5(80%),5(66%)
17	Bangladesh	5 min	WiFi	-	4	4(33%),3(63%)
18	Bangladesh	$>$ 1 hour	3.5G	3.5G	3	4(27%),3(49%)
19	Pakistan	30 min	WiFi	3.5G	4	5(45%),4(58%)
20	Sweden	1 hour	Both	3.5G	4	5(36%),5(76%)
21	Pakistan	Never	Local	-	4	2(39%),4(35%)
22	Sweden	5 min	WiFi	3.5G	3	-,4(46%)
23	Pakistan	Never	Local	3.5G	4	5(53%),5(65%)
24	Pakistan	10 min	WiFi	WiFi	4	2(41%),4(47%)
25	India	10 min	WiFi	-	3	4(58%),4(57%)
26	Pakistan	5 min	WiFi	WiFi	4	4(39%),-
27	Pakistan	Never	Local	-	4	5(54%),3(33%)
28	Sweden	30 min	Both	3.5G	4	5(31%),4(37%)
29	Sweden	Never	-	-	4	-,3(25%)
30	Sweden	30 min	WiFi	3.5G	4	5(62%),4(33%)

Furthermore, we conducted a short interview with each subject just after the experiments. We collected demographics such as age, gender, and occupation. In addition we obtained information on: (i) how often they are streaming video over the network on their own mobile devices per day (less than 2 min; less than 5 min; less than 10 min; less than 30 min; less than 1 h; more than 1 h); (ii) which network interface (3.5G or WiFi) they are using during video streaming; (iii) the brand name and the operating system of their smartphone; (iv) which wireless network interface they are experiencing freezes the most, namely Often Experienced Freezes (OEF) on 3.5G or WiFi. After each user watched both videos, i.e., at the end of the experiments, the test supervisor asked to rate this time the overall quality based on both videos in the five-level MOS scale.

5 Results

The presentation of the results are structured in two parts. The results for the ON/OFF study, i.e., two-state modeling of the inter-picture times (D_p), are given in the Section 5.1. The results of the second part of the experiments, i.e., the user study, are presented in Section 5.2.

5.1 ON/OFF study (Part 1)

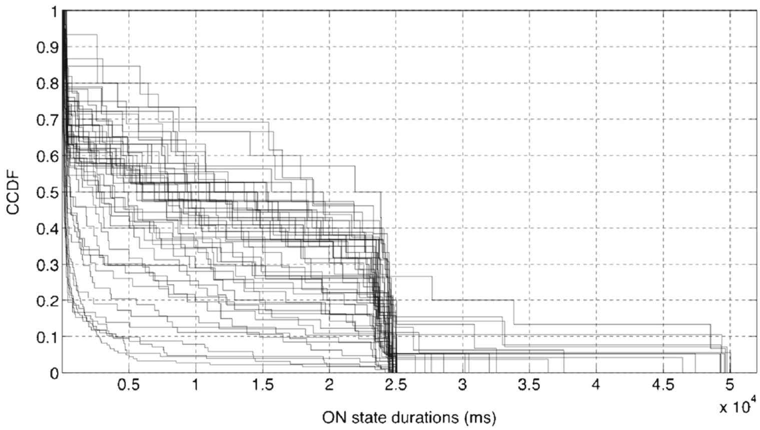
The durations of the ON and OFF states is calculated for the 58 iterations amongst the 60 iterations. During 2 iterations (RTSP run #29 and HTTP run #27), the video stopped in the very beginning due to very bad network coverage, which caused the video impossible to stream. We excluded the two iterations from our dataset. We studied the durations spent at the ON and OFF states. We first assured that the distribution of the ON and OFF state durations in all 58 runs could fit into an exponential curve by obtaining the Coefficient of Determination (R^2) values, as it can show how well our hypotheses is correct, i.e., how well the collected data points fit an exponential curve. We chose exponential modeling as we assumed that the ON and OFF durations is memoryless, i.e., the current state of the video streaming does not depend on its previous state. The mean R^2 values were calculated as 0.93 and 0.81 for OFF and ON states, respectively. Next, the MLEs for the durations of ON and OFF states for all 58 iterations are calculated, then the mean of all MLE values are obtained. The mean MLE of OFF and ON durations is calculated as 642 ms and 9.7 s, respectively. Complementary Cumulative Distribution Function (CCDF) plots are helpful in visualizing random variables that are rare and above particular levels. Our aim is to emphasize the high OFF durations that might cause low end-user perceived QoE. Thus we chose to visualize the ON and OFF durations for all runs separately via CCDF plots as given in Fig. 7a and b, respectively. We observe slight deviations between the CCDF plots, as the network conditions were not identical on all runs. However, the mean R^2 on exponential fits of all runs for ON and OFF scenarios were above 0.8.

We observed long OFF durations followed by very short inter-picture durations. This shows that the pictures were clumped and then displayed many back-to-back pictures to the smartphone screen at once in very short duration. This behavior eventually manifests itself on the user interface as fast-forward for a short period of time.

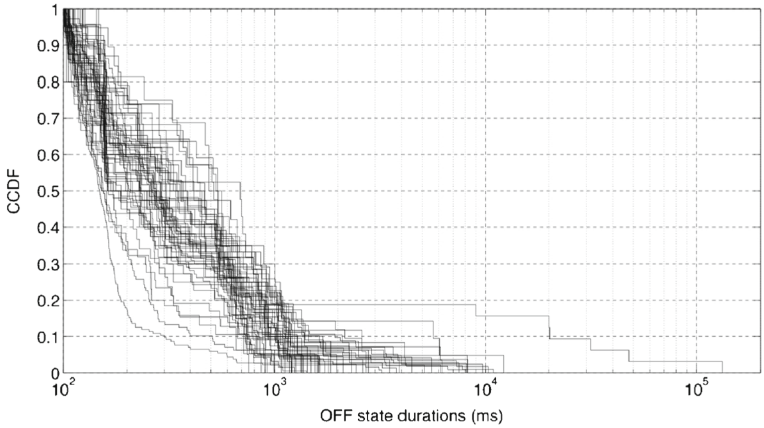
5.2 User study (Part 2)

The results with respect to the user study are presented in three parts. First the user feedbacks captured during the interview are presented in Section 5.2.1. Next, the results with respect to the data captured from the VLQoE is presented in Section 5.2.2. Afterwards, the minimum perceived inter-picture times per participant is presented in Section 5.2.3. Before presenting the results, we identified the corrupted (due to corrupted timestamps) or missing data.

Corrupted or missing data The freeze indications by the user are necessary to identify the minimum perceived inter-picture times. Some user data was corrupted, or did not help in identification of freezes due to the missing freeze indications. Amongst the 14 out of 60 user experiments, the users did not press the freeze button. We have not realized the corrupted/missing data during the experiments and thus could not prevent it. We suggest to check that starting from the early phases of the study as also suggested by the authors of [41], in further related user studies to increase both the quantity and the quality of the collected data.



(a)Screen ON state durations (ms). Linear x-axis.



(b) Screen OFF state durations (ms). Logarithmic x-axis.

Fig. 7 **CCDF** plots for the duration in ON and OFF states for 58 runs

5.2.1 Short interview with the participants

In total, there were 30 participants in the user study. 28 participants were within 21–30 years age range; *S*13 and *S*30 were within 31–40 years age range. Five subjects amongst all participants (*S*2, *S*5, *S*16, *S*17, *S*22) were female. Five subjects watched the videos at home, and 25 subjects at the university campus.

We collected background information regarding their experiences on their own smartphones, particularly during video streaming. This is helpful to select subjects from different nationalities, and gender. In addition, we aimed to select subjects that stream videos on their own smartphones in daily life and in their natural environments. The background information of the subjects are presented in columns 2–5 in Table 2. The first column is the subject ID of each user. The nationality of each selected user is listed in the second column; the subjects were from six different nationalities. The third column shows the duration that the users watch videos via wireless network interface on their own smartphones per day in daily life. 23 out of all subjects watch a video on the smartphone with a duration of at least 5 min per day. The fourth column shows how the users watch video on their own smartphones, e.g., streaming via only 3.5G; via only WiFi; via only *Local* storage of the smartphone; or via *Both* WiFi and 3.5G. 13 subjects stream video via only WiFi interface; five subjects stream via only 3.5G; six subjects stated that they use both interfaces occasionally to stream video; four subjects use only local streaming; and two subjects (*S*13, *S*29) never watch a video on the smartphone. Six subjects (*S*5, *S*13, *S*21, *S*23, *S*27, *S*29) claimed that they never watch a network-based video on the smartphone in daily life. Often Experienced Freezes (OEF) during video streaming is given in column 5 of Table 2 and it is categorized as either 3.5G or WiFi. In total, 17 subjects claimed that they often experienced freezes while streaming via 3.5G; and only four subjects stated that they experienced freezes while streaming via WiFi. *S*13 and *S*23 state that the reason for not watching network-streamed video is due to their previously perceived low quality of experience of the freezes via 3.5G streaming.

5.2.2 ESM results

On the sixth column of Table 2, the subjects' average evaluations (in MOS) on the temporal quality of the videos are presented in five-level ACR scale. Five subjects rated 3-Fair, 23 subjects rated 4-Good; one subject rated 5-Excellent for the average temporal quality of the streams. The average ACR received from one subject, i.e., *S*5, was lost. Although the subjects have experienced occasional freezes during the video sessions, which is detailed in Section 5.2.3, the overall UR values were higher than 2-Poor. This might be due to the cognitive factors such as the human memory effect [21, 34].

After studying the user-perceived quality based on the short interviews, we examined the user ratings in more details during the video stream, i.e., we analyzed the data collected via ESM. In the last column of Table 2, the dominant user rating (UR), i.e., the user rating that is rated with the highest amount, is given with the corresponding percentage amongst all other given ratings from that user. The most frequent user rating received during each streaming session is presented together with its percentage of its occurrence. For example, 4(50 %) means that the highest percentage (50 %) of all ratings is 4 of that particular subject. The values are stated for the two streaming protocols and for each user.

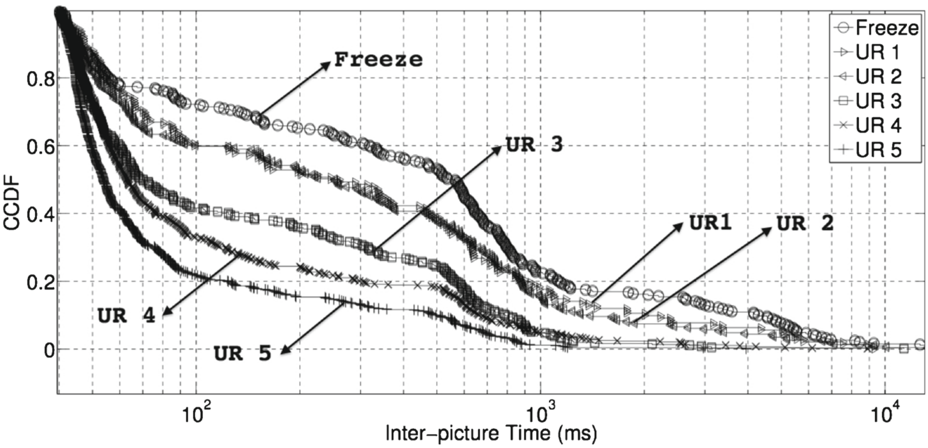


Fig. 8 Inter-picture time distribution of the corresponding user indications

The maximum inter-picture time value during the period between the current and previous user indication (user rating or a freeze) is studied. The CCDF plots of the maximum inter-picture time values for the user ratings 1–5, and freeze indications collected from all users are given in Fig. 8. The plot clearly shows that 60 % of the D_p values (where the y-axis is at 0.4) are less than or equal to approximately 65 ms, 80 ms, 150 ms, 500 ms, 500 ms, 700 ms for "UR 5", "UR 4", "UR 3", "UR 2", "UR 1", and "Freeze", respectively. Van Kester et al. [52] found the acceptable (UR $>$ 3.5) freezing time as 360 ms. According to our results, if we consider the acceptable boundary as UR $>$ 2, then the D_p should not exceed 500 ms. The CCDF plots of "UR 1" and "UR 2" are almost the same.

The more detailed statistics of the maximum D_p values before a user indication are given in Table 3. The inter-picture time values for $D_{\mathrm{p}_{\max}}$ values were exponential distributed with R^2 values given in Table 3. The dataset for the $D_{p_{\max}}$ values for all user indications were fit to an exponential distribution with R^2 value of higher than or equal to 0.84. The mean of the maximum inter-picture time values were calculated as 152 ms, 282 ms, 321 ms, 768 ms, 831 ms, 1289 ms for "UR 5", "UR 4", "UR 3", "UR 2", "UR 1", "Freeze", respectively.

We also calculated the time gap between the user indication time, T_i, and the timestamp of a displayed picture with the maximum display time, $T_{D_{\mathrm{p}\max}}$. The statistics on the latter are given in the last three rows of Table 3. The mean $\Delta(T_{i-D_{p\max}})$ values were calculated as 2109 ms, 2650 ms, 3267 ms, 3808 ms, 3809 ms, and 2603 ms for "Freeze", "UR 1", "UR 2", "UR 3", "UR 4", and "UR 5", respectively.

First low user rating time ($T_{\mathrm{LowRating}}$) *and number of alarms* In this part, we categorized the pictures that are displayed with more than 100 ms duration (D_p $>$ 100 ms) on the display as *alarms*. The number of *alarms* is visualised along with the corresponding low user ratings (UR $<$ 4) that are received, and this is done for all user experiments in Fig. 9. Each point on the plot represents the result of one video experiment session. $T_{\mathrm{LowRating}}$ for each user are shown on the y-axis; and total number of alarms that are detected until $T_{\mathrm{LowRating}}$, are shown on the x-axis. Some subjects were less picky, i.e., a low user rating at the 25$^{\text{th}}$ second of a video session is received with respect to the 37 alarms recorded by the application within this interval. In contrast, some people were rather picky, i.e., seven users rated the quality with UR $<$ 4 within the initial 30 seconds of the video, although there were no alarms with D_p $>$ 100 ms.

Table 3 Number of data points, goodness-of-fit, mean, median, standard deviation values are presented in rows 2-6 for $D_{\mathrm{p}_{\max}}$ together with the corresponding $\Delta(T_{i-D_{p\max}})$ values in rows 7–9

User indication	Freeze	UR=1	UR=2	UR=3	UR=4	UR=5	Unit
No. of Data Points	266	142	229	305	471	515	-
R^2(Exp. Distribution)	0.96	0.94	0.94	0.84	0.90	0.94	-
Mean$\{D_{\mathrm{p}_{\max}}\}$	1289	831	768	321	282	152	ms
Median$\{D_{\mathrm{p}_{\max}}\}$	550	253	253	70	63	54	ms
Std$\{D_{\mathrm{p}_{\max}}\}$	2495	1611	1666	758	706	295	ms
Mean$\{\Delta(T_{i-D_{p\max}})\}$	2109	2650	3267	3808	3809	2603	ms
Median$\{\Delta(T_{i-D_{p\max}})\}$	1332	1457	1686	2387	2527	1335	ms
Std$\{\Delta(T_{i-D_{p\max}})\}$	3147	3358	4220	4133	4016	3412	ms

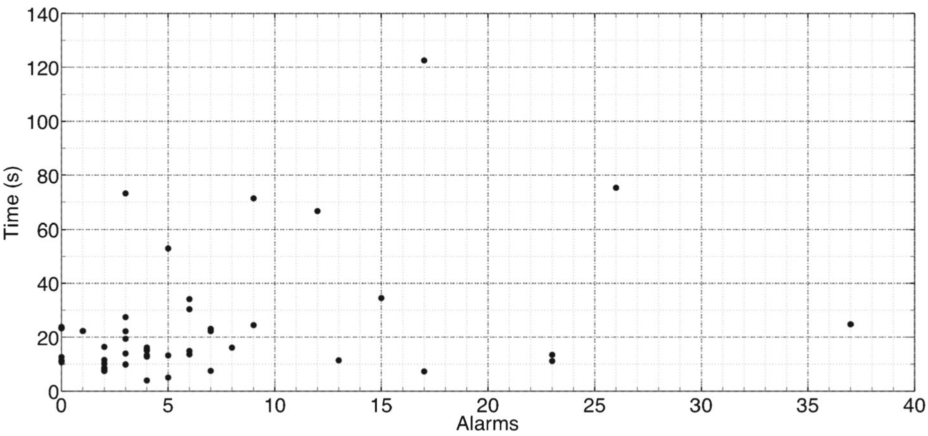


Fig. 9 The user ratings (UR < 4) are visualized with respect to the corresponding relative time from the start of the video, number of alarms raised by the application

5.2.3 Minimum perceived inter-picture time

The minimum perceived inter-picture time values for each subject (subject ID at the first column) are given in Table 4 at the columns 2 and 5 for HTTP and RTSP streaming, respectively. The total number of freeze indications (#i) by each user are given in columns 4 and 7, respectively. We considered the minimum perceived inter-picture times that were reacted by the users at least 80 ms after $T_{D_{\mathrm{pmax}}}$ [39]. There were two user experiments (RTSP S6, RTSP S13) out of 60, in which $D_{\mathrm{response_{short}}}$ were less than 80 ms, therefore from those two datasets, we selected the first $D_{\mathrm{response_{short}}}$ that is higher than 80 ms. The perceived inter-picture times for each user are presented in Table 4 ranging between 40 ms and 4542 ms. The distribution of the measured inter-picture time values is exponential with $R^2 = 0.82$. The exponential MLE of the perceived inter-picture time based on all subjects is 328 ms. Some subjects indicated many freezes such as $S6$ (with 39 freeze indications) during RTSP streaming, while some did not contribute in any freeze indication such as $S3$. The lack or low amount of freeze indications can be caused by many reasons, some factors are: (i) the subject forgets to indicate a freeze, (ii) the subject does not perceive the freezes, or (iii) there were not perceivable freezes in the video due to a good quality. The $D_{\mathrm{P}_S\mathrm{min}}$ results presented in Table 4 complement the results presented in Fig. 9 in the sense that the calculated D_{P_S} value is different for each user.

5.2.4 User response time

The user response time is calculated for each user while considering the first and the second scenario, separately.

First scenario – short In this scenario, the user indicates a perceived freeze by pressing the freeze button as the video plays out. The results with respect to the user response times for the short freezes as defined in Section 3.1.5 are given in Table 4. The third and the sixth column depicts the $D_{\mathrm{response_{short}}}$ values for all subjects and for both RTSP and HTTP streaming experiments, respectively.

Table 4 Minimum perceived inter-picture times for each user. Corrupted or missing data is indicated by "-"

	HTTP			RTSP		
S	$D_{\mathrm{ps}\,\mathrm{min}}$	$D_{\mathrm{response}_{\mathrm{short}}}$	#*i*	$D_{\mathrm{ps}\,\mathrm{min}}$	$D_{\mathrm{response}_{\mathrm{short}}}$	#*i*
1	49 ms	4451 ms	2	55 ms	4995 ms	4
2	53 ms	7075 ms	10	67 ms	318 ms	11
4	57 ms	10596 ms	8	43 ms	313 ms	7
5	41 ms	454 ms	5	43 ms	539 ms	5
6	-	-	-	41 ms	705 ms	39
7	51 ms	3538 ms	12	-	-	-
8	41 ms	218 ms	14	86 ms	111 ms	7
9	-	-	-	45 ms	1862 ms	2
12	103 ms	4401 ms	6	318 ms	2548 ms	7
13	-	-	-	515 ms	1399 ms	7
14	-	-	-	40 ms	385 ms	1
16	41 ms	275 ms	10	83 ms	534 ms	41
17	-	-	-	389 ms	6496 ms	8
18	50 ms	431 ms	4	50 ms	1124 ms	2
19	4542 ms	4969 ms	1	-	-	-
20	117 ms	4394 ms	9	45 ms	117 ms	9
21	43 ms	3367 ms	20	59 ms	666 ms	3
23	681 ms	1300 ms	3	2880 ms	2704 ms	2
24	53 ms	3959 ms	13	249 ms	1352 ms	1
27	474 ms	1493 ms	3	-	-	-
28	45 ms	2127 ms	4	49 ms	102 ms	15
29	2924 ms	319 ms	1	-	-	-
30	46 ms	369 ms	4	-	-	-

Second scenario – long In this scenario, the user indicates a freeze when the video pauses. This happens when the display duration of the stalled picture exceeds the difference between the timestamps of that displayed picture and the user's freeze indication. The $D_{\mathrm{response}_{\mathrm{long}}}$ values obtained from each subject are different. As the CCDF plot can reveal the high user response times in the data, we plotted the CCDF plot of $D_{\mathrm{response}_{\mathrm{long}}}$ values as given in Fig. 10. $D_{\mathrm{response}_{\mathrm{long}}}$ values are exponential distributed with R^2 of 0.98. The MLE of the $D_{\mathrm{response}_{\mathrm{short}}}$ values is calculated as 1533 ms. Again, based on the results plotted in Fig. 10, the user response times vary from user to user.

5.3 Summary of the results

We developed a tool (VLQoE) that can record all the metrics detailed in Section 3.2.2, and it is possible to analyze in detail all the collected parameters related to QoE either during or after the experiments. By using VLQoE, we conducted a study in two parts: the first part aims to model the inter-picture time at the user interface; while the second part aims to find out the video QoE via subjective study. The most important findings of our work are stated as follows:

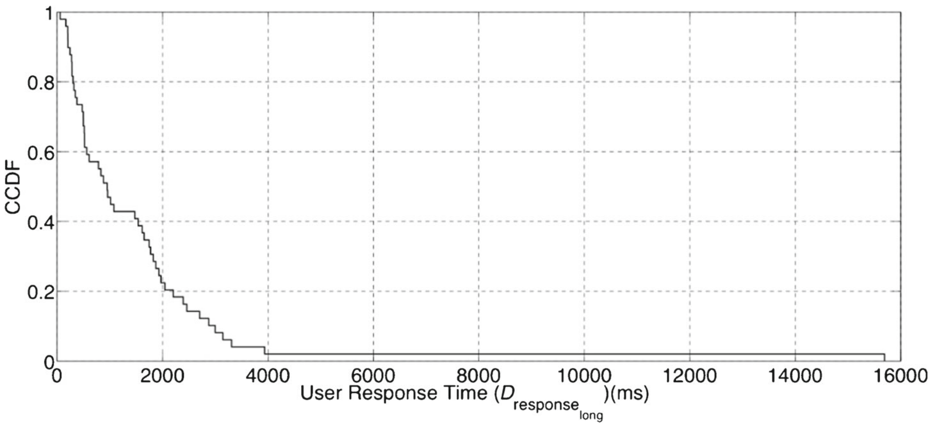


Fig. 10 **CCDF** plot of User Response Times, $D_{\mathrm{response}_{\mathrm{long}}}$, for the long freezes

– ON/OFF modeling based on the inter-picture time during video streaming is studied, and exponential models (with mean R^2 values 0.81 and 0.93 for ON and OFF states, respectively) are obtained. The mean ON duration is calculated as 9.7 s; and the mean OFF duration is calculated as 642 ms during the streaming experiments. (Part 1, Section 5.1)
– The minimum perceived inter-picture time is studied. The Exponential Maximum Likelihood Estimation (MLE) of the minimum perceived inter-picture time is calculated as 328 ms. (Part 2, Section 5.2)
– The mean of the highest acceptable (UR $>$ 3) inter-picture time values is calculated as 282 ms. (Part 2, Section 5.2)
– The MLE of User Response Time ($D_{\mathrm{response}_{\mathrm{long}}}$) during a long freeze scenario is calculated as 1533 ms. (Part 2, Section 5.2)
– The time it takes for the subjects to give a low user rating, i.e., UR $<$ 4, from the start of the video is studied. We found out that this is varying for each user. Some users are less picky, and do not respond immediately despite the high number of alarms (freezes) raised by the video player; while some users are rather picky and give a low user rating with low number of alarms. (Part 2, Section 5.2)

6 Limitations

The obtained numbers in the results of this paper are valid for the particular video, for the particular users involved in the study in particular conditions as stated. Further tests are under investigation for a wide variety of videos and with the participation of more users. The accuracy of the timestamping of the displayed pictures might be a few milliseconds varying from the actual values. This depends on the clock accuracy, and drifting as well as on the timestamping of the programming environment and operating system, and also to the implementation related factors, e.g., particular location of the timestamping function in the code. Moreover, the accuracy of the minimum perceived inter picture time values were based on the number of freeze indications per user as well as other factors related to human factors such as cognitive bias, user concentration level, etc. The improvement of the tool with respect to this aspect is ongoing.

During the experiments, we experienced corrupted or missing data for some subjects, e.g., some users did not press the freeze button. Based on this experience, we recommend a periodic check of the data during the study to detect and prevent similar anomalies in future studies.

In the current version of VLQoE, we have employed only a subjective method with the five-level ACR scale as it is a commonly used. On the other hand there are other recommended assessment methods as such some focuses on the relative quality [36]. We leave the task of studying other QoE methods on the user interface for future work, and in parallel we encourage QoE researchers to modify the VLQoE for the target studies.

7 Conclusions

In this article, we presented the extended version of the VLC-media player, i.e., VLQoE that is aimed to help researchers to conduct smartphone-based video QoE experiments. We did this by adding an extra functionality to an open source VLC media player. VLQoE records a set of metrics from the user interface, network-layer, and the available sensor metrics from the Android OS. The network- and physical-level measurements enable the collection of lower layer QoS related metrics such as signal strength transmitted/received packets. The measurements at the user interface enable the collection of direct user feedback with respect to the temporal impairments; and the sensor/GPS components enable the collection of user context during video streaming. The VLQoE tool can be used for user-centric QoE modeling as a function of collected metrics such as the smartphone battery level; application rebuffering events; numbers of received and transmitted video packets; smartphone screen orientation/brightness level; smartphone network interface signal strength; and the location/context of the end-user.

We used VLQoE to further study the temporal impairments of a network-based video stream. We presented our approach in two parts. In the first part, we modeled the video streaming session in two states with the assumption that 100 ms is the boundary inter-picture time for a video playout to be considered either in ON (smooth playout), or in OFF (video picture freeze) state. We presented the duration of ON and OFF states with exponential models, and presented the inferential statistics for both states. In part 2, we focused on the QoE centric modeling and conducted a user study to find out the boundary inter-picture times between the ON and OFF states. To do that, we studied the minimum perceived inter picture times, i.e., the minimum display duration of a video picture that is perceived by a user, and we concluded that this metric varies from user to user (max. 2880 ms, min. 40 ms).

Next, we showed the influence of the inter-picture times on the end-user perceived QoE as measured at the user interface. Based on the results obtained from the user study, the mean of the maximum inter picture time values were 152 ms, 282 ms, 321 ms, 768 ms, 831 ms, and 1289 ms for “UR 5”, “UR 4”, “UR 3”, “UR 2”, “UR 1”, “Freeze”, respectively. The maximum inter-picture time increases as the user rating decreases; and the highest mean of maximum inter-picture times have matched with the “Freeze” indications, as expected. We observed almost similar maximum inter-picture time distributions for “UR 1” and “UR 2”. Moreover, we studied the first low (“UR < 4”) ratings that were received by the users during the video streaming sessions. The occurrences of the high (greater than 100 ms) inter-picture times, i.e., alarms, are studied together with the streaming time until the reception of the first low user rating. In parallel, we analyzed the overall user response time of the subjects in two scenarios (short/long freezes). The distribution of the user response time with respect to the long freezes is found to be exponential distributed, and the MLE of the user response time is calculated roughly to 1.5 s. We presented the demographics and background information of each user based on users' daily life video streaming experiences on the smartphones.

The ultimate goal of video QoE experiments is to detect degradation in the quality of video stream, and to react in order to minimize the influence of it on the end-user perceived quality. We believe that VLQoE tool can provide metrics with high information gain to the machine learning mechanisms that will benefit in terms of improvement of the future streaming quality. The content presented in this paper can be extended with more number of test subjects (e.g., via Google Play), with long-duration studies, and could shed light on construction of new methodologies for smartphone-based subjective studies and eventually QoE modeling on the smartphone.

8 Future work areas

We believe that the VLQoE tool has great potential for further improvements for mobile-based QoE assessments. It can further be enhanced with more variety of QoE assessment methods, video content, and eventually to be put in the Android application store in order to reach a larger number of subjects. The tool can be improved by adding the options that helps the researchers to choose the desired QoE methodology (e.g., % slider instead of five buttons), choose/upload new test videos (e.g., VQEG video clips), to be used in the video experiments. This could be done by adding a separate drop-down interface that lets the researchers to choose. The outcomes of testing different existing assessment methodologies and videos will help in standardizing smartphone-based QoE assessment.

Assessment of QoE in real life user studies while being minimally obtrusive to the user is challenging. Obtaining the user data non-obtrusively, i.e., without explicitly asking the user to press the button, can be done by the accelerometer data of the device during the video streaming. This might help to identify the user behavior (e.g., jiggling the mobile device) with respect to the video freeze. Similarly, the obtained user data on the screen touch events may potentially reflect the user perception in a less obtrusive and more objective way. These items can be studied in future work. On the other hand, continuous recording of accelerometer data might be resource consuming. Therefore, one recommendation is an automatic recording of the accelerometer data once a high inter-picture time at the user interface has been detected.

Some further future works could be applied; in fact be re-designed in order to increase the engagement of the subjects to the study, e.g., tournament to detect freezes. The VLQoE tool can be turned into a game platform where the user feedback with respect to the quality of the video would be collected in minimally intrusive away. This way, the number of participants, the variety of context, and consequently the quality of the data might be improved. The score of each user based on the game can be presented as public and the subjects will be able to compete with each other to detect and indicate the highest number of freezes with the cost of uploading their data to a public QoE database. The collected data can then be fed into machine learning mechanisms so that the application itself will be trained and then suggest the "good- enough" levels of the parameters for the video source, which could then be fed into the control loop.

We have so far studied inter-picture time with the focus on network-based video streaming, and the collected data depends on the real-time condition (available bandwidth) of the network link during the experiments. Therefore, we plan to complement our work with controlled experiments, e.g., predefined disturbances, to have ground truth, and stream videos that are pre-recorded from the local storage of the smartphone.

Acknowledgments The work of Katarzyna Wac has been supported by AAL-MyGuardian and WayFIS projects. We would like to thank anonymous reviewers for the valuable comments and suggestions to improve our paper.

References

1. Barakovic S, Kapov LS (2013) Survey and challenges of QoE management issues in wireless networks. J Comput Netw Commun 2013, Article ID 165146, 28 pages
2. Bonald T, Roberts JW (2012) Internet and the Erlang formula. SIGCOMM Comput Commun Rev 42(1):23–30
3. Cherif W, Ksentini A, Negru D, Sidibe M (2011) APSQA: efficient real time video streaming QoE tool in a future media internet context. In: IEEE Conference in Multimedia and Expo (ICME), pp 1–6
4. Cisco Visual Networking Index: Global mobile data traffic forecast update 2012–2017. [Online], Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-.520862.html. Accessed Sept 2013
5. Dalal AC, Bouchard AK, Cantor S, Guo Y, Johnson A (2012) Assessing QoE of on-demand TCP video streams in real time. In: IEEE International Conference on Communications (ICC), pp 1165–1170
6. Darwin Streaming Server. [Online], Available: http://developer.apple.com/opensource/server/streaming. Accessed Mar 2013
7. De la Cruz Ramos P, Vidal FG, Leal RP (2010) Perceived video quality estimation from spatial and temporal information contents and network performance parameters in IPTV. In: Fifth International Conference on Digital Telecommunications (ICDT), vol., no., June
8. Dialogic Corporation Quality of Experience for Mobile Video Users. White Paper, Canada, 2009
9. Egger S, Hossfeld T, Schatz R, Fiedler M (2012) Waiting times for quality of experience for web based services. In: Proceedings of the 4th international workshop on quality of multimedia experience
10. Fenimore C, Libert J (2013) Perceptual effects of noise in digital video compression. [Online], Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.196.1791&rep=rep1&type=pdf. Accessed Sept 2013
11. Fiedler M (2011) Traffic models for quality of experience assessment, tutorial at Proceedings of the 23rd international Teletraffic Congress. San Francisco
12. French H, Lin J, Phan T, Dala AC (2011) Real time video QoE analysis of RTMP streams. 30th International IEEE Performance Computing and Communications Conference (IPCCC), pp 1–2
13. Gardlo B, Ries M, Rupp M, Jarina R (2011) A QoE evaluation methodology for HD video streaming using social networking. In: IEEE International Symposium on Multimedia (ISM), pp 222–227
14. Grondin S (2010) Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. Atten Percept Psychophys 72(3):561–582
15. Hossfeld T, Schatz R, Egger S (2011) SOS: the MOS is not enough. In: Proceedings of the third international workshop on quality of multimedia experience, pp 131–136
16. Hossfeld T, Schatz R, Seufert M, Hirth M, Zinner T, Tran-Gia P (2011) Quantification of YouTube QoE via crowdsourcing. In: IEEE international workshop on multimedia quality of experience-modeling, evaluation, and directions (MQoE 2011). USA
17. Hossfeld T, Schatz R, Biersack E, Plissonneau L (2013) Internet video delivery in YouTube: from traffic measurements to quality of experience. Data Traffic Monit Anal Lect Notes Comput Sci 7754:264–301
18. HTTP Live Streaming draft-pantos-http-live-streaming-02. [Online], Available: http://tools.ietf.org/pdf/draft-pantos-http-live-streaming-02.pdf. Accessed Sept 2013
19. Huang J, Xu Q, Tiwana B, Mao ZM, Zhang M, Bahl P (2010) Anatomizing application performance differences on smartphones. In: Proceedings of the 8th international conference on mobile systems, applications, and services (MobiSys '10). ACM, New York, pp 165–178
20. Ickin S (2013) Identification of influential factors on Android Smartphone-based video quality of experience. Licentiate Dissertation, Sweden
21. Ickin S, Janowski L, Wac K, Fiedler M (2012) Studying the challenges in assessing the perceived quality of mobile-phone based video. In: Fourth international workshop on Quality of Multimedia Experience (QoMEX), pp 164,169
22. Ickin S, Wac K, Fiedler M, Janowski L, Hong JH, Dey AK (2012) Factors influencing quality of experience of commonly-used mobile applications. IEEE Commun Mag 50(4):48–56

23. Intille S, Kukla C, Ma X (2002) Eliciting user preferences using image-based experience sampling and reflection. In: Proceedings of human factors in computing systems CHI
24. ITU-T Rec. P. 800 (1996) Methods for subjective determination of transmission quality
25. ITU-T Recommendation P.910 Subjective video quality assessment methods for multimedia applications, Recommendations of the ITU (Telecommunication Standardization Sector), [Online], Available: http://www.itu.int/rec/T-REC-P.910-200804-I/en. Accessed Sept 2013
26. Janowski L, Romaniak P, Papir Z (2011) Content driven QoE assessment for video frame rate and frame resolution reduction. In: Multimedia tools and applications, pp 1–18
27. Juluri P, Plissonneau L, Medhi D (2011) Pytomo: a tool for analyzing playback quality of YouTube videos. In: 23rd International Teletraffic Congress (ITC), pp 304,305, 6–9
28. Katabi D, Handley M, Rohrs C (2002) Congestion control for high bandwidth-delay product networks. In: Proceedings of the conference on applications, technologies, architectures, and protocols for computer communications (SIGCOMM '02). ACM, New York, pp 89–102
29. Khan A, Sun L, Jammeh E, Ifeachor E (2010) Quality of experience-driven adaptation scheme for video applications over wireless networks. IET Commun 4(11):1337–1347
30. Laghari KR, Connelly K (2012) Toward total quality of experience: a QoE model in a communication ecosystem. IEEE Commun Mag 50(4):58–65
31. Latre S, De Turck F (2012) Autonomic quality of experience management of multimedia networks. In: IEEE IFIP network operations and management symposium, pp 872–879
32. Le Callet P, Moller S, Perkis A (eds) Qualinet white paper on definitions of quality of experience (QoE). [Online], Available: http://www.qualinet.eu/. Accessed Mar 2013
33. Li M, Li F, Claypool M, Kinicki R (2005) Weather forecasting - predicting performance for streaming video over wireless LANs. In: Proceeedings of NOSSDAV. Stevenson, Washington
34. Menkovski V (2013) Intelligent control for adaptive video streaming. In: Proceedings of the international conference on consumer electronics. Las Vegas
35. Menkovski V, Liotta A (2013) QoE for mobile streaming, mobile multimedia - user and technology perspectives, InTech Europe, (2012), [Online], Available: http://cdn.intechopen.com/pdfs/26696/InTech-Qoe_for_mobile_streaming.pdf. Accessed: September
36. Menkovski V, Exarchakos G, Liotta A (2011) The value of relative quality in video delivery. J Mob Multimed 7(3):151–162
37. Migliorini D, Mingozzi E, Vallati C (2010) QoE oriented performance evaluation of video streaming over WiMAX. In: Eighth international conference on wired/wireless Internet communications. Sweden
38. Minhas T (2012) Network impact on quality of experience of mobile video, Licentiate dissertation, Blekinge Institute of Technology. Sweden
39. Musser G (2013) Time on the brain: how you are always living in the past, and other quirks of perception. [Online], Available: http://blogs.scientificamerican.com/observations/2011/09/15/time-on-the-brain-how-you-are-always-living-in-the-past-and-other-quirks-of-perception/. Accessed Apr
40. Myung IJ (2003) Tutorial on maximum likelihood estimation. J Math Psychol 47(1):90–100
41. Oliver E (2010) The challenges in large-scale smartphone user studies. In: Proceedings of the 2nd ACM international workshop on hot topics in planet-scale measurement. ACM, New York, Article 5, 5 p
42. PEVQ Perceptual Evaluation of Video Quality. [Online], Available: http://www.pevq.org/. Accessed Mar 2013
43. Qiao Z (2011) Smarter phone based live QoE measurement. In: 15th international conference on intelligence in next generation networks
44. Quan HT, Ghanbari M (2008) Temporal aspect of perceived quality of mobile video broadcasting. IEEE Trans Broadcast 54(3):641–651
45. RFC2326: Real Time Streaming Protocol (RTSP). [Online], Available: http://www.ietf.org/rfc/rfc2326.txt. Accessed Sept 2013
46. Seow SC (2008) Designing and engineering time: the psychology of time perception in software. Addison-Wesley Professional
47. Singh KD, HadjadjAoul Y, Rubino G (2012) Quality of experience estimation for adoptive HTTP/TCP video streaming using H.264/AVC. In: IEEE Consumer Communications & Networking Conference, (CCNC), pp 127–131
48. Schatz R, Hossfeld T, Janowski L, Egger S (2013) From packets to people: quality of experience as a new measurement challenge. Data Traffic Monit Anal Lect Notes Comput Sci 7754:219–263
49. Staehle B, Hirth M, Pries R, Wamser F, Staehle D (2010) YoMo: a Youtube application comfort monitoring tool. In: Proceedings of EuroITV Workshop QoE for Multimedia Content Sharing (QoEMCS'10). Tampere

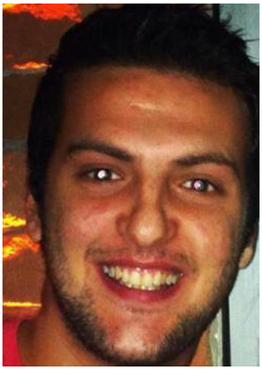
50. Staelens N, Moens S, Van den Broeck W, Marien I, Vermeulen B, Lambert P, Van De Walle R, Demeester P (2010) Assessing quality of experience of IPTV and video on demand services in real-life environments. IEEE Trans Broadcast 56(4):458–466
51. Top Project Listings. VideoLAN, Retrieved: 2013-01-22
52. Van Kester S, Xiao T, Kooij RE, Brunnstrm K, Ahmed OK (2011) Estimating the impact of single and multiple freezes on video quality. In: Proceedings of SPIE, vol 7865
53. Venkataraman M, Chatterja M (2009) Evaluating quality of experience for streaming video in real time. In: IEEE GLOBECOM
54. Vidal RP, Gicquel J, Colomes C, Cherifi H (2004) Sporadic frame dropping impact on quality perception. Human Vis Electron Imaging IX, vol 5292
55. Vishwanath A, Dutta P, Chetlur M, Gupta P, Kalyanaraman S, Ghosh A (2010) Perspectives on quality of experience for video streaming over WiMAX. ACM SIGMOBILE Mobile Comput Commun Rev 13(4):15–25
56. VLC Media Player for Android (2013) [Online], Available, http://www.videolan.org/vlc/download-android.html. Accessed Mar 2013
57. Voice quality and MOS. [Online], Available: http://searchunifiedcommunications.techtarget.com/tip/Voice-quality-and-MOS. Accessed Apr 2013
58. Vuppala A, Sriram LN (2011) Measurement of user related performance problems of live video streaming in user interface. Masters Thesis, Blekinge Institute of Technology
59. Wang Y, Claypool M (2005) RealTracer - tools for measuring the performance of RealVideo on the internet. In: Kluwer multimedia tools and applications, vol 27, issue no 3
60. Wang Z, Lu L, Bovik AC (2004) Video quality assessment based on structural distortion measurement. Proc Sig Process Image Commun Spec Issue Objective Video Qual Metrics 19(2):121–132
61. Wang B, Kurose J, Shenoy P, Towsley D (2008) Multimedia streaming via TCP: an analytic performance study. ACM Trans Multimedia Comput Commun Appl 4, 2, Article 16, 22 p
62. Winkler S, Mohandas P (2008) The evaluation of video quality measurement: from PSNR to hybrid metrics. IEEE Trans Broadcast 54(3):660–668
63. World Wide Smartphone Sales Share. [Online], Available: http://en.wikipedia.org/wiki/File:World_Wide_Smartphone_Sales_Share.png. Accessed Apr 2013
64. Yang KC, Guest CC, El-Maleh K, Das PK (2007) Perceptual temporal quality metric for compressed video. IEEE Trans Multimedia 9
65. Zinner T, Abboud O, Hohlfeld O, Hossfeld T, Tran Gia P (2010) Towards QoE management for scalable video streaming. In: 21st ITC specialist seminar on multimedia applications, traffic, performance and QoE. Japan
66. Zupernick H, et al. (2008) Quality of experience based cross-layer design of mobile video systems. In: Proceedings of the 18th ITC specialist seminar on quality of experience

Selim Ickin is an Electrical and Electronics engineer and currently working as a Ph.D. candidate at Blekinge Institute of Technology (BTH), Sweden. He was involved in the EU STREP FP7 PERIMETER project. He has his B.Sc. degree from the Department of Electrical and Electronics Engineering, Bilkent University, and M.Sc. degree from BTH. He has worked in Nortel Networks and completed his internships in Siemens AB and Schneider Electric. His current research is on the end-user perceived Quality of Experience (QoE) vs. Energy Consumption of the Android OS wireless network-based applications. His major work involves implementation of software tools for Android-based platforms to understand the underlying QoE metrics of wireless network-based applications and to understand the relation between user-perceived QoE and the energy consumption.

Markus Fiedler is holding a Ph.D. degree in Electrical Engineering with focus on ICT and a Docent degree in Telecommunication Systems. Starting from modeling and analysis of traffic in telecommunication and computer networks, he has been focusing on Quality of Experience (QoE) during the recent years. His interests and experience include various topics in the area of the performance of networks and network applications. These topics comprise amongst others the areas of user perception of waiting times, QoE models, performance and security, and in particular instrumented user interfaces. Until 2012, he was member of the Steering Board of the NoE Euro-NF and also co-chair of Future Internet Cluster in the ICT domain of the European Commission.

Katarzyna Wac is a senior scientist at the Institute of Services Science of University of Geneva and leader of the Quality of Life research area since 2010. She holds a BSc and MSc degree in Computer Science from Wroclaw University of Technology (Poland), an MSc in Telematics from University of Twente (the Netherlands), as well as a PhD in Information Systems from University of Geneva (Switzerland). In 2009–2010, Ms. Wac was on leave for a one-year SNSF Fellowship at Carnegie Mellon University (USA), Human-Computer Interaction Institute. Ms. Wac's research interests include Quality of Service-aware mobile systems and services with special emphasis on support of adaptive multimedia protocols and Quality of Service mechanisms, especially in the mobile healthcare domain. Ms. Wac is an Associate Expert of the International Telecommunication Union (ITU) European Regional Initiative for mHealth.

Patrik Arlos is Assistant Professor at the School of Computing, Blekinge Institute of Technology. He holds a M.Sc. degree in Electrical Engineering from Hogskolan i Karlskrona/Ronneby, and in 2005 he defended his PhD in Telecommunications Systems at Blekinge Institute of Technology. His research interests lies in network measurements, network performance, Quality of Service and Experience.



Canberk Temiz is a B.Sc student in the Blekinge Institute of Technology (BTH), Sweden. He had graduated from the science oriented department of Avni Akyol High School. He was involved in Boat Project and LED Cube project in BTH. He has worked in a trainee program at Netport Information and Technology.

Khadija Mkocha received a B.Sc. degree and a M.Sc. degree in Telecommunications Engineering from the University of Dar es Salaam, Tanzania, in 2005 and 2009, respectively. She joined the academic staff of the University of Dar es Salaam as a Tutorial Assistant in 2007. Currently, she is an Assistant Lecturer in the Department of Electronics and Telecommunication Engineering of the University of Dar es Salaam. Her research interests include delivery of educational materials in distance learning, next generation networks, communication theory, and mobile programming. She is a member of the Institution of Engineers Tanzania (IET) and IEEE.