TELECOM INFRA PROJECT®

MRN PG Report

QoE Framework Application to Short Form Video

Table of Contents

Authors and Contributors

Xinli Hou

Connectivity Technologies and Ecosystems Manager Meta Platforms Inc.,
xinlihou@meta.com

François Blouin

QoE Engineering, Meta Platforms Inc,
francoisb@meta.com

Minqi Wang

Research Engineer, Orange
minqi.wang@orange.com

Xinli Hou

Connectivity Technologies and Ecosystems Manager, Meta Platforms Inc.
xinlihou@meta.com

Kevin Smith

Distinguished Engineer, Vodafone Networks,
kevin.smith@vodafone.com

Javed Rahman

Technology Development Strategist, T-Mobile USA
Javed.2.rahman@t-mobile.com

Chris Murphy

Regional CTO, EMEA, VIAVI Solutions
chris.murphy@viavisolutions.com

Kafi Hassan

Technology Development Strategist, T-Mobile USA.
kafi.hassan@t-mobile.com

Ricardo Serrano Gutierrez

Core & service platforms engineer, Telefónica,
ricardo.serranogutierrez@telefonica.com

Gavin Young

Head Of Fixed Access Centre Of Excellence, Vodafone Networks
gavin.young2@vodafone.com

Mayur Channegowda
Broadband Architect, Vodafone Networks
mayur.channegowda@vodafone.com

Durga Satapathy
Director, Advanced Technologies & Innovation, T-Mobile USA
Durga.Satapathy@t-mobile.com

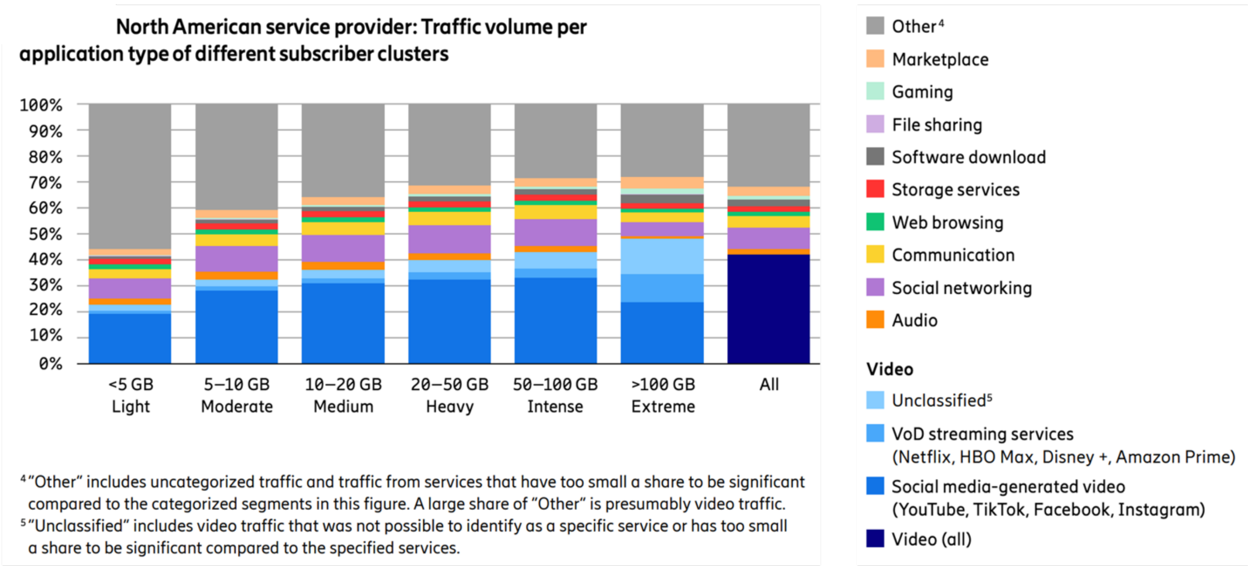
Executive Summary

The momentum of short form video (SFV) has been growing in recent years, making it one of the most important apps that drives (mobile) Internet traffic growth. Communication Service Providers (CSPs), in collaboration with Content Application Providers (CAPs), are looking for ways to support SFV to provide best user experience while minimizing use of network resources. This effort will have a great impact on network evolution.

Applying the QOE/QoS Measurement framework developed in the Telecom Infra Project (TIP) Metaverse Ready Network (MRN) Project Group (see here for more information), this document identifies metrics fundamental to estimating quality of experience (QoE). It intends to lay a foundation for future work to determine values of QoE metrics that define a target QoE, and to determine quality of service (QoS) requirements on end to end networks in order to deliver the target QoE.

Introduction

Short Form Video (SFV) has seen tremendous growth in the last few years, quickly becoming the dominant traffic globally. According to the Ericsson Mobility Report (June 2023 edition), short form video, here called social media generated video, is the single most important traffic-driving app and generates 20 - 30% of total mobile broadband traffic in North America.



It is therefore important to understand what the network QoS requirements would be to support a target level of QoE for SFV, because it will shape the evolution of cellular networks and build a foundation for future more sophisticated metaverse apps. To some extent, SFV may also drive the evolution of Wi-Fi/fixed networks as most people view SFV on smartphones connected to Wi-Fi.

By applying the QOE/QoS Measurement framework developed in the Telecom Infra Project (TIP) Metaverse Ready Network (MRN) Project Group, QoE metrics are identified and defined as being most important to estimate the user experience of short form video. The document will also explain where and how to measure QoE. The document is structured as follows: it starts with a description of characteristics of SFV apps and a hypothetical reference architecture to set context for the QoE and

QoS definition, followed by a brief survey of work done by the research, standardization and industry community. It then describes the QoE metrics that are most important to SFV, where and how they can be measured. Finally, the document presents future work such as identifying QoS metrics relevant to deliver the targeted QoE and studying the relationship between QoE/QoS for SFV apps based on measurements and parametric models. The goal of the future work is to provide answers to the following questions: What are the QoS requirements to deliver a target QoE? What is the expected QoE corresponding to a given QoS support by a network?

Short Form Video

Use Case Description

Short-form video, also called social media produced video, is basically video streaming but the following characteristics distinguish it from other forms of video services.

A. It is typically less than 60 - 90 seconds in length, though some content and applications providers (CAPs) consider content up to three (or even ten) minutes long to be short-form.

   There is no standard agreement as to the duration of SFV. While the maximum video length permitted by a SFV content application provider (CAP) has become longer in the past few years, for example, in 2022,

   - TikTok allowed video content going up to 10 minutes.
   - Instagram Reels are up to 90 s.
   - YouTube shorts are up to 60 s.
   - Facebook reels are up to 90 s.
   - Kwai is15s -5 mins.
   - Thriller is up to 60 s.

   While the permitted duration of a short form video becomes longer, it does not mean video on SFV platforms are getting longer. According to Statista, from August 2022 to January 2023, TikTok accounts with up to 500 views produced videos of approximately 32.4 seconds on average. Small accounts produced videos of around 36.4 seconds in length. Huge accounts, which presented over 25,001 views, produced content with a duration of around 42 seconds on average as of the examined period.

   In the context of QoS, both short and longer videos share many similar KPIs, but these impact the QoE with different weights. Apart from the short form videos buffer being shorter (limited to the whole length of video e.g. 10s), intuitively, the impact of loading time at the beginning is much more important than for a long video: since the video is short and users are quickly scrolling through social feeds, within less than 3 seconds, a user will decide to continue watching a video or not. If not, the user will scroll to the next one. To satisfy users, the user-expected initial loading time for SFV would be much smaller than that for streaming TV.

B. SFVs are mostly uploaded to social media platforms for viewing on a smartphone (vs. on a large screen in a living room)

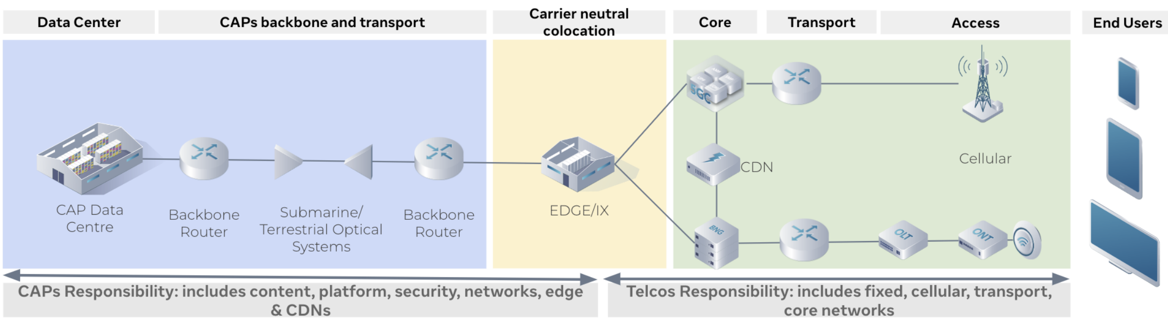
   Users know or assume that SFVs are mostly produced by non-professional cameras by non-professional photographers. The User's expectation on the quality of the video played on their (smartphone) screen is not as high as for movies streamed to their big screen in the living room. This means that video quality in SFV is not as important as in long form video streaming of professionally produced video such as IPTV/mobile TV.

C. Contents are auto-fed (downloaded) to viewers based on algorithmic recommendations.
D. SFV is meant to be easily digestible and quickly consumed, allowing viewers to scroll through and view multiple pieces of content in a short period of time.

End to End Delivery of short form video

Both CAPs and Telcos have a role to play in delivering short form video to end users

- CAPs invest in content, platform, security, their internal networks connecting data centers to Edge locations, and Content Delivery Networks (CDN) residing in Telcos networks to bring the video content as close as possible to the end user.
- Telcos invest in access, transport, security, core networks and peering connections to deliver the video - among other types of data - from the edges of the CAP networks to the end user.



This set up means it is in the joint interest of CAPs and Telcos to find the optimal points of Video Multi Method Assessment Fusion (VMAF) / Universal Video Quality (UVQ) metrics vs bitrates to deliver the best balance between video quality and network efficiency.

Related work

Early in 2017, ITU-T published the MOS model for adaptive audiovisual streaming services [ITU-TP1203]. This recommendation focuses on 1 - 5 minutes video with some ranges for initial loading delay and stalling events as shown below [Table1 from ITU-T P.1203.3]:

- Initial loading delay and stalling: 0-10 seconds
- Maximum number of stalling events: 5
- Maximum length of a single stalling event: 15 seconds
- Total stalling duration: 30 seconds
- Other details: No stalling within 5 seconds of the start of the video playing.

Useful as a reference to understand work in the related area, the above target values are applicable to IPTV streaming and more general streaming, not for short form videos. However, SFV has the "quickly scrolling" feature where new parameters should be taken into consideration for QoE modeling as well. Related work [1, 5, 7] concentrate on SFV QoE with some concrete testing scenarios with the most popular SFV platforms (TikTok, Instagram Reels, YouTube shorts and Facebook Watch). According to these documents, key indicators for video (or short-term video) that impact QoE are identified as video quality, audio quality, quality variations, buffer event, start delay and video duration with additional variables such as encoding and rendering capacities, packet size and scrolling speed.

The full-referenced approach is often adopted by lab studies for video quality and recently used for SFV. For example, authors in [1] used the Video Multimethod Assessment Fusion (VMAF) method to reveal an objective QoE score ranging from 0 to 100 (highest score when video quality equals to original content). In the real world, it is hard to get a full-referenced measurement since service providers have no visibility of network infrastructure where each telecom operator has its own QoS management strategy and vice versa.

Moreover, the recent study [9] focuses on video streaming applications (i.e. Tiktok, Youtube Live, Zoom) and tries to draw some conclusions between objective QoE and QoS. It points out that the average stall ratio (the studied objective QoE) is four times higher than that of the simulated testbed which brings more complexity for QoE and QoS studies.

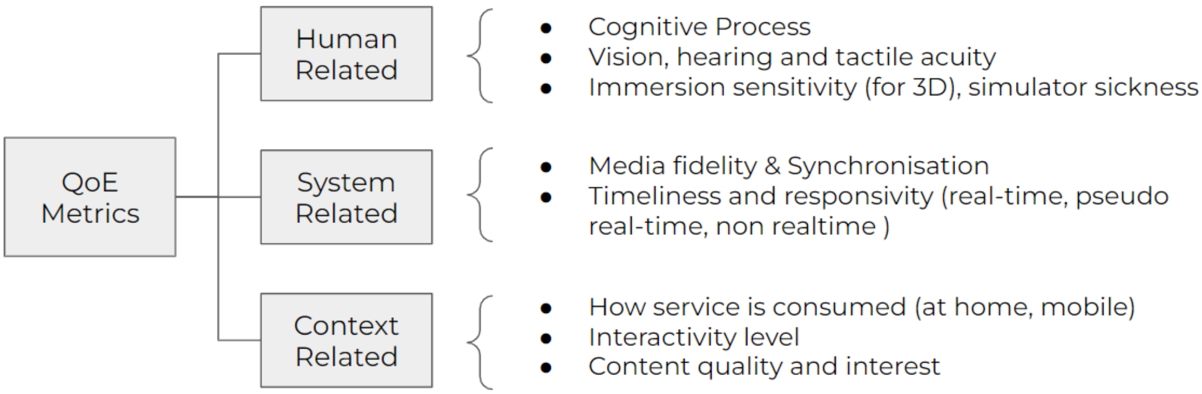
It is also important to note that using HTTP adaptive mechanism (i.e. DASH) for streaming delivery and pre-loading contents could improve QoE for SFV. However, as mentioned in [1, 6, 8], over-buffering (preloading) drains the network bandwidth and causes poor QoE for other services in the network. Thus, a trade-off exists between QoE for the whole network and energy or resource consumption.

[2] presents the concept of Network SLO (service level objective) developed at Meta with the objective to readily determine if network conditions are responsible for instances of poor quality of experience (QoE) such as images loading slowly or video stalling during playback. Network SLO can be thought of as an App's "minimum network requirements" for good QoE. If the network between Meta and a user does not meet the App's SLO requirements, QoE will be degraded.

[3] presented how Meta measures visual quality and then how it uses these metrics to make smarter trade-offs in the player when optimizing for the users' quality of experience. Quality metrics are used to decide the best ceiling to use when playing videos on cellular networks. Quality metrics are also leveraged when making tradeoffs between visual quality and risk of in-play stalls in Meta's adaptive streaming algorithms.

Quality of Experience (QoE) Metrics

Quality of Experience (QoE) describes how consumers judge the quality of a service. In order to quantify video experience, we need a perceptual model that covers the key factors that impact the end-user appreciation and satisfaction with a service. We generally characterize video Quality of Experience (QoE) according to three factors:

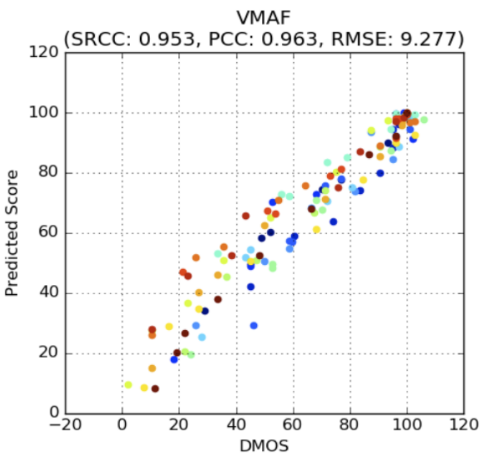


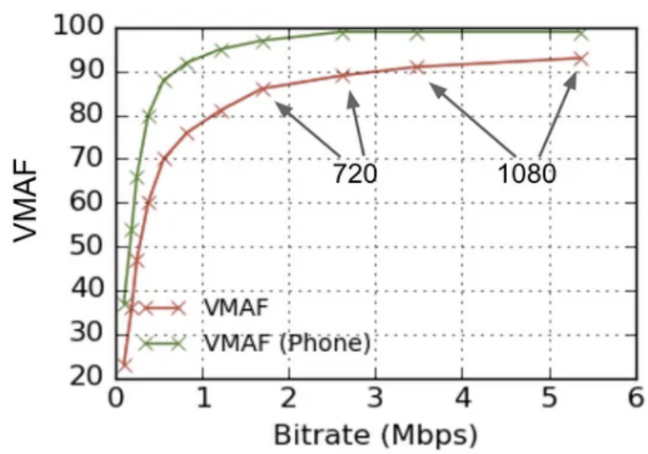
Sources:
European Network on Quality of Experience in Multimedia Systems and Services, Qualinet , 2013
ITU-T G.1032 (2017), ITU-T G.1035 (2021), ITU-T G.1320 (2022), ITU-T G.1036 (

- Human-related influencing factors: vary from person to person and can be physiological (ex. hearing acuity, visual acuity)
- System-related influencing factors: objective and quantifiable metrics associated with media capture, transmission, networking, coding, storage and rendering, display.
- Context related influencing factors: describes the user's surrounding environment, in terms of physical location and space, mobile vs fixed access network, temporal, social (people present or involved in the experience)

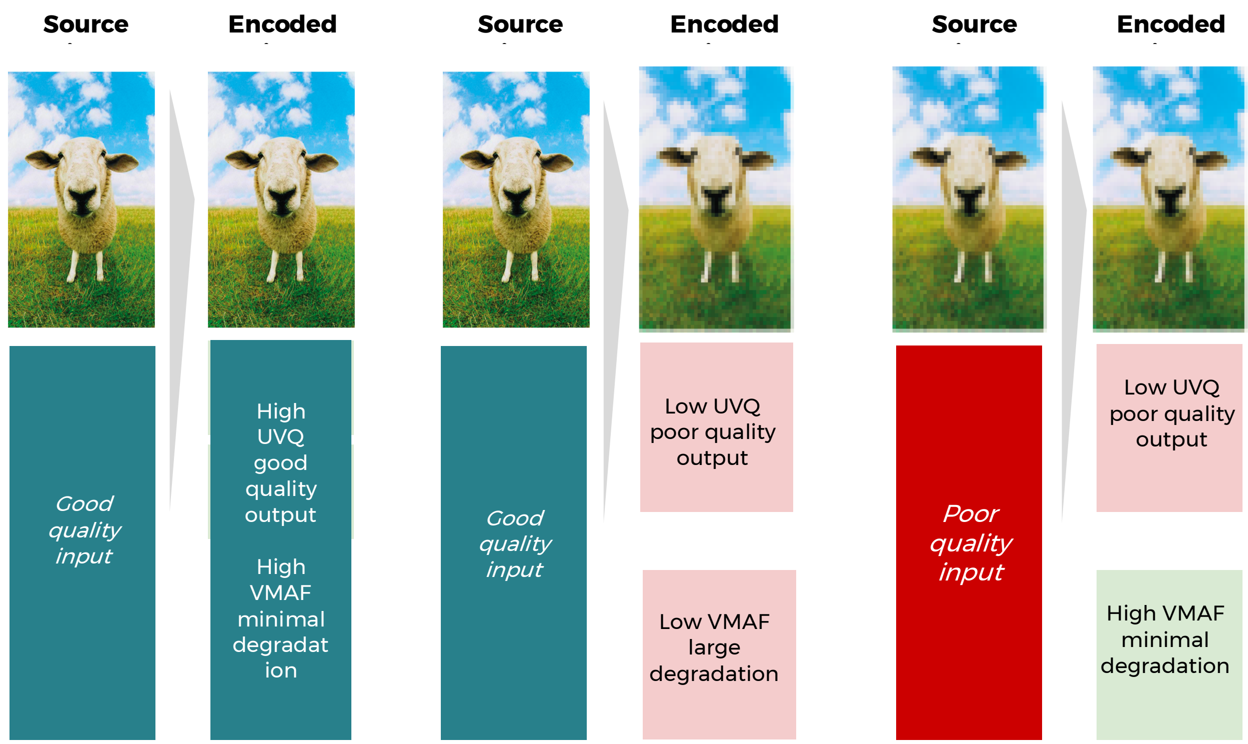
Gathering subjective user opinion scores in response to visual stimuli has been the cornerstone in establishing a ground truth that objective metrics attempt to automate. Statistical aggregation of opinion scores results in a Mean Opinion Score (MOS), which considers a mix of human-related and system-related factors in a single measure. After MOS is collected across a large sample of users it can be converted to a Differential MOS (DMOS) by subtracting it from the score obtained for the reference, unprocessed visual stimuli, thus expressing the level of quality drop that compression and other processes introduce to videos. While DMOS provides a good indication of user experience, it's impractical to ask the end user to rate every piece of content they watch, hence CAPs have developed a set of metrics that can quantify video experience at scale. Video quality metrics such as Video Multi Method Assessment Fusion (VMAF)[2] (first pioneered by [Netflix]) and the Universal Video Quality (UVQ)[3] (developed by [Google]) provide two such solutions.

VMAF provides a combined human vision modeling with machine learning that fuses a few primary metrics for a better correlation between DMOS to achieve a scaled, repeatable way of measuring Video experience. VMAF is core to most video processing platforms, and is used to measure the quality of the encoded bitstreams as compared to the source. VMAF or (Google relevant Metric) can then be used to extrapolate bitrates required over networks.



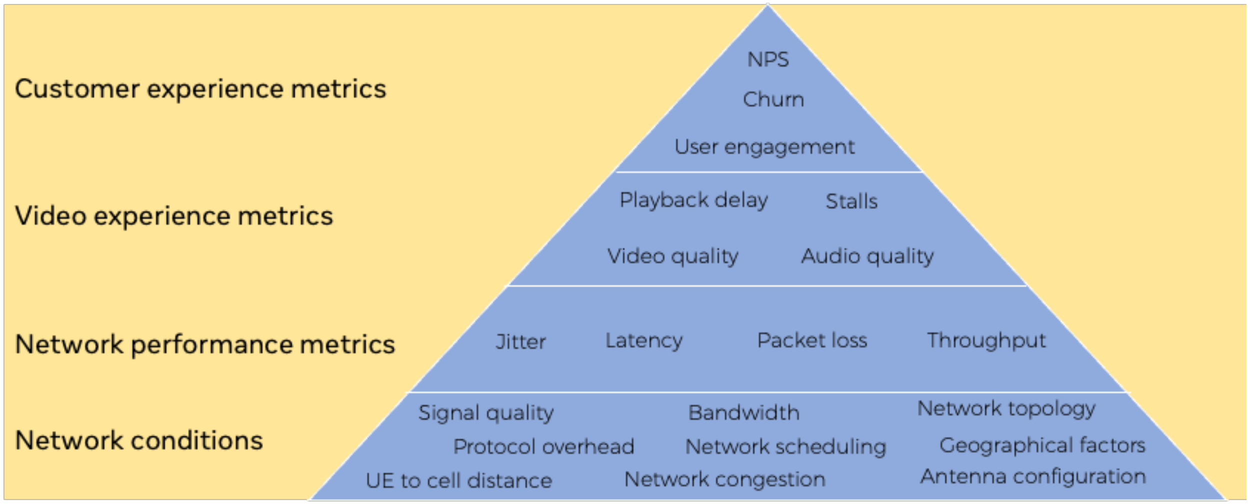


Reference-based QoE metrics such as VMAF measure the quality of the encoded bitstreams as compared to the source. Meanwhile, no-reference metrics such as UVQ[3] provide an objective quality score for a given video, and thus incorporate source quality as a factor:



When the source is of very high quality, the VMAF score approaches an absolute quality score. However, when the source video is of lower quality, VMAF would no longer be an accurate measure of the quality of that video. Also note that, no-reference metrics are computationally expensive, and thus scale is a challenge[4]. A hybrid approach can be used where the no-reference metric is run once to establish a baseline quality score for a given video, and that no-reference score is then used to adjust the VMAF scores for the encoded streams.

Beyond video quality metrics, additional metrics such as Play Delay and Stalls are also highly relevant. These metrics focus on other system-related aspects of service delivery, and measure the timeliness and responsiveness of the video service. Together, video experience metrics comprise the quality of experience (QoE) for Video watching.



In turn, these QoE metrics correlate with higher order user experiences (e.g. churn, watch time, Net Promoter Score/NPS) and lower order network behaviors (e.g. latency, signal strength) as depicted above.

It should be pointed out that short form video requires a fairly heavy backend workload, both from a streaming perspective but also from the AI inference engine, to ensure timeliness for fluidity of playback as well as the right content users like to maintain user attention and retention. The network is a critical contributor to those factors, Telemetry at the application and network level is becoming very strategic and an important component metrics for traffic, customer retention, operation, troubleshooting as well as capacity planning. Transport protocols such as QUIC/BBR impacts traffic volume efficiency and thus should also be considered in the delivery of such services.

To summarize, QoE metrics relevant to short form video are summarized below.

QoE Aspects	QoE metrics	How to measure
Audio quality	Audio MOS	ViSQOL, PESQ, POLQA, Codec type, Bit rate
Timeliness	Initial loading time	Measured on the client, the interval between the time when a user swipes a video and the time when the video starts to play on the screen
	Play success rate n (PSRn)	Percentage of SFV views which has a CTPT less than n seconds
	Stalls	Measured on the client side per viewing session by some or a combination of 1. number of stalls (longer than a to be specified number of milliseconds) per time unit (e.g., per hour) 2. total time of stalls (milliseconds) in a given duration (e.g., an hour) 3. meantime between stalls during a session
Temporal quality	Fluidity	Measure on client by number of frames per second
	Synchronicity	Measured on the client side per viewing session by some/combination of 1. numbers of audio/video out-of-synch 2. total time of audio/video out-of-synch 3. meantime between audio/video out-of-synch
Spatial quality	Video fidelity	No Reference: under development Full Reference: PSNR, SSIM, or VMAF
Context	Client Device, Location	Display resolution and audio fidelity, mobile or stationary, network type
Human factors	User rating	Users are asked, during or after their viewing, to rate their satisfaction in scale of 1 - 5

References

[1] Shangyue Zhu et al, "Swipe Along: A Measurement Study of Short Video Services", MMSys'22: Proceedings of the 13th ACM Multimedia Systems Conference, June 2022, Pages 123–135

[2]] Brandon Schlinker et al, *"Network SLOs: Knowing When The Network Is The Barrier To Application Quality Of Experience*", 2022

[3] Denise Noyes, *Using Video Quality Metrics for Smarter Client Side Decisions*, presentation in "@Scale" web-conference, 2023

[4] ITU-T recommendations on quality assessment of streaming services

https://www.itu.int/rec/T-REC-P.1203

https://www.itu.int/rec/T-REC-P.1203.1/en

https://www.itu.int/rec/T-REC-P.1203.2-201710-I

https://www.itu.int/rec/T-REC-P.1203.3/en

[5] Haodan Zhang et al, "QUTY: Towards Better Understanding and Optimization of Short Video Quality", MMSys '23: Proceedings of the 14th Conference on ACM Multimedia Systems, June 2023, Pages 173–182

[6] Guanghui Zhang et al, "DUASVS: A Mobile Data Saving Strategy in Short-Form Video Streaming", IEEE Transactions on Services Computing, Vol. 16, Issue 2, March-April 2023

[7] Xiaoqi Yin et al, "A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP", SIGCOMM '15: Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, August 2015, Pages 325–338

[8] Ximing Wu et al, "QoE-aware Download Control and Bitrate Adaptation for Short Video Streaming", MM '22: Proceedings of the 30th ACM International Conference on Multimedia, October 2022, Pages 7115–7119

[9] Yanan Li et al, "Demystifying the QoS and QoE of Edge-hosted Video Streaming Applications in the Wild with SNESet", Proc. ACM Manag. Data, Vol. 1, No. 4 (SIGMOD), Article 236. December 2023

TELECOM INFRA PROJECT®