Real-Time QoE Assessment of Video Streaming based on ITU-T P.1203

Valerio Triolo[1,*,†], Marco Scarpa[1,†], Salvatore Serrano[1,†] and Salvatore Distefano[2,†]

[1] *Department of Engineering, University of Messina, C.da Di Dio (Villaggio S. Agata), Messina, 98166, ME, Italy*

[2] *Department of Mathematics and Computer Sciences, Physical Sciences and Earth Sciences, University of Messina, Viale Ferdinando Stagno d'Alcontres 31 - 98166 Messina, Italy*

Abstract

In this paper, a system for the real-time evaluation of video streaming Quality of Experience (QoE) based on the ITU-T P.1203 standard is presented. The proposed approach aims at estimating the Mean Opinion Score (MOS) in live streaming scenarios. It analyzes incoming media segments in real-time, leveraging a sliding window mechanism to assess the quality on a configurable number of segments. The system manages playback interruptions, such as buffering events, by integrating into the assessment pipeline a module to monitor player stalls and quality level switches. Experimental results demonstrate the system's effectiveness in different streaming conditions and highlight its ability to capture perceptual quality fluctuations, making it a valuable tool for content providers aiming to optimize user experience in adaptive streaming environments.

Keywords

MOS, ITU-T P.1203, video and audio streaming, Quality of Experience, quality assessment

1. Introduction

In recent years, the rapid growth of multimedia-centric services and applications has significantly grown, affecting the network traffic and becoming a major challenge for video streaming service providers. Assessing the Quality of Experience (QoE) of video content streaming provides multimedia service providers valuable insights on how their networks are performing in delivering the content. As users become more accustomed to high-quality streaming services, their expectations grow accordingly. To deliver services that meet these expectations, providers must develop a shared understanding of the issues affecting users and how these factors influence their perception of service quality.

Unlike Quality of Service (QoS) metrics, which focus on network-level parameters and disregard human perception, QoE captures the subjective user experience by directly assessing the perceived visual and auditory quality of a video sequence during playback. Specifically, real time monitoring of video QoE allows operators to proactively adjust bandwidth allocation and optimize traffic routing to enhance the quality of real-time television broadcasting services over IP networks.

The implementation of the ITU-T P.1203 standard [1], currently available on GitHub [2], enables the extraction of the Mean Opinion Score (MOS) for a given video sequence based on a combination of audiovisual quality metrics. However, this implementation is primarily designed for the assessment of pre-recorded content and is not designed for real-time, continuous evaluation of live streaming scenarios, in case of adaptive streaming protocols such as HLS (HTTP Live Streaming) and MPEG-DASH (Dynamic Adaptive Streaming over HTTP) [3]. In this paper, we report experiments on an implementation of ITU-T P.1203 customized for integrating real-time capabilities. We measured its behavior changing characteristic parameter values to derive their impact on the final MOS value.

QualITA 2025: The Fourth Conference on System and Service Quality, June 25 and 27, 2025, Catania, Italy

[*] Corresponding author.

[†] These authors contributed equally.

valerio.triolo@studenti.unime.it (V. Triolo); mscarpa@unime.it (M. Scarpa); sserrano@unime.it (S. Serrano); salvatore.distefano@unime.it (S. Distefano)

valeriot30.github.io (V. Triolo)

0009-0001-5825-8894 (V. Triolo); 0000-0002-9560-7504 (M. Scarpa); 0000-0003-0507-5186 (S. Serrano); 0000-0002-2752-626X (S. Distefano)

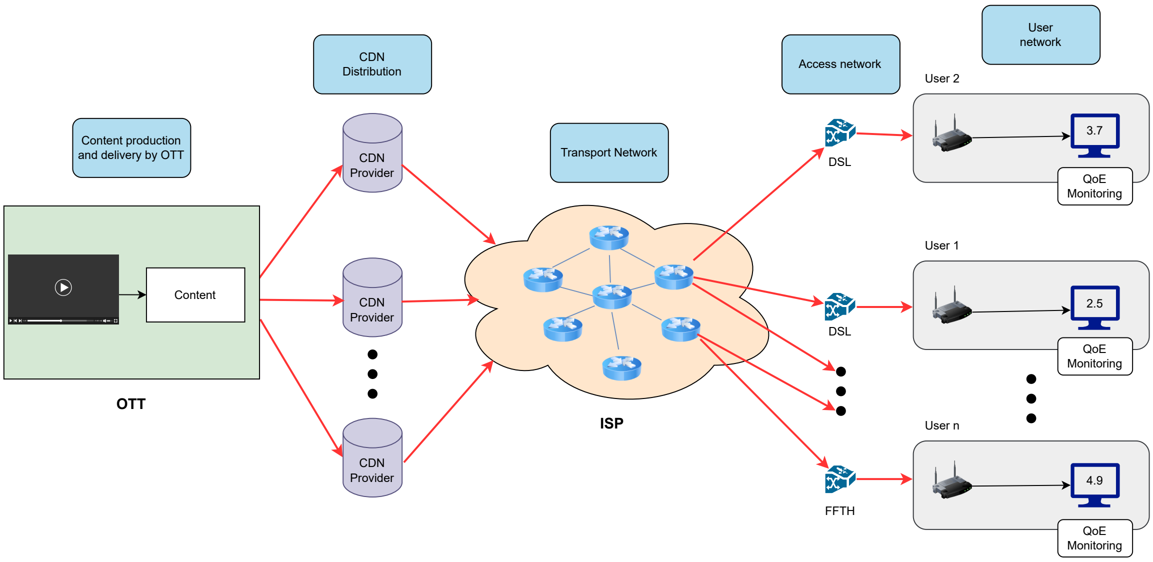

Figure 1: End-to-end content delivery chain from OTT production to user devices, highlighting the role of CDN providers, transport and access networks, and the importance of QoE monitoring through MOS, which may vary significantly even when network metrics appear satisfactory.

2. Preliminary concepts

2.1. Problem description

The digital transformation of media provisioning and fruition has ushered in an era where multimedia content delivery dominates global network traffic, with video streaming accounting for more than half of downstream internet traffic worldwide. This paradigm shift has created a complex ecosystem where traditional telecommunications infrastructure must seamlessly support diverse, bandwidth-intensive applications while meeting increasingly sophisticated user expectations for uninterrupted, high-definition content delivery. The evolution from broadcast television to live streaming by the so-called *Over The Top* (OTT) services has fundamentally altered the relationship between content providers and consumers. Unlike traditional broadcasting models where technical limitations were accepted as inherent constraints, modern streaming services operate under implicit service-level agreements where users expect consistent, premium-quality experiences commensurate with their subscription fees. This expectation creates a critical business challenge: technical failures during content delivery are no longer mere inconveniences but represent potential contractual breaches that can trigger financial liabilities and permanent subscriber loss.

In this context, two interrelated concepts, Quality of Service (QoS) and Quality of Experience (QoE), play a pivotal role in defining the success of content delivery. Quality of service refers to the objective and measurable performance of the underlying network and delivery infrastructure, which includes parameters such as latency, jitter, throughput, and packet loss. It represents the technical foundation upon which reliable and efficient multimedia transmission is built. QoE, by contrast, captures the end-user's subjective perception of service quality, integrating both the technical performance and the user's expectations, context, and satisfaction. While high QoS is a necessary condition for ensuring smooth streaming, it is not sufficient on its own; true competitive advantage lies in optimizing QoE, where the ultimate goal is to deliver an experience that users perceive as seamless, engaging, and worth the subscription.

Figure 1 depicts the context where monitoring the streaming video QoE is relevant. The multimedia content is delivered through a chain of interconnected elements, by the OTT provider, across CDN (Content Delivery Network) distribution, through the transport and access networks, and finally to the user device. Each of such stages may introduce noise and potential degradation of the multimedia

Table 1
The five-grade ACR (Absolute Category Rating) MOS scale.

MOS	Quality	Impairment
1	Bad	Very annoying
2	Poor	Annoying
3	Fair	Slightly annoying
4	Good	Perceptible but not annoying
5	Excellent	Imperceptible

content. Although operators have traditionally relied on objective QoS indicators such as latency, jitter, or packet loss to assess network health, these metrics alone often fail to capture the true quality perceived by the end user. This is where QoE monitoring becomes essential.

A widely used metric to quantify the QoE is the Mean Opinion Score (MOS), originally proposed for subjective assessment of audio quality [4], has been widely adopted and later extended to evaluate video quality as well [5]. It reflects the average rating given by viewers - typically on a scale from 1 to 5 - based on their perception of a video stream. providing a standardized way to quantify human perception of visual distortion by translating subjective judgments into a numerical quality scale. The typical five-grade MOS scale is presented in Table 1.

By monitoring MOS alongside QoS metrics, operators can bridge the gap between technical performance and subjective perception, identifying cases where buffering, CDN inefficiencies, or application-level issues degrade the experience despite acceptable network conditions. In this way, MOS monitoring provides a more complete understanding of service quality, ensuring that operational priorities align not only with infrastructure performance but also with user satisfaction and business outcomes.

2.2. Related work

The rapid growth of video traffic over IP networks, as predicted by Cisco [6], has led to a strong interest in monitoring and improving the QoE for end users. Traditional objective metrics such as Peak Signal to Noise Ratio (PSNR) [7] and SSIM (Structural Similarity Index) [8] offer basic image-based quality estimates and have the needs of comparing the received content with the original one, since they are full reference algorithms, but lack correlation with user perception, especially in streaming contexts where temporal effects (e.g., stalling, quality switches) dominate. VMAF (Video Multi-Method Assessment Fusion) [9], introduced by Netflix, improves alignment with human judgments by combining perceptual features via machine learning, but remains a reference metric and many techniques of temporal pooling are currently being studied. Additional frameworks have explored QoE estimation in operational contexts. For example, Alvarez et al. proposed a flexible QoE framework adaptable to different service requirements and client conditions [10]. Further studies have validated P.1203 in real-world contexts. Bermudez et al. [11] evaluated the model performance under LTE network conditions, Robitza et al. [12] also applied the model to YouTube streaming under constrained bandwidth, finding that P.1203 could accurately reflect QoE degradation due to reduced throughput. More recently, Viola et al. [13] proposed an edge computing architecture for distributed QoE analytics using P.1203 in dense client environments. Their system integrates subjective quality estimates with network-level monitoring, suggesting a promising path toward hybrid QoE/QoS models.

2.3. The ITU-T P.1203 Standard

The starting point of this work is a recent standard for QoE estimation in HTTP-based streaming, namely the ITU-T P.1203 model [14]. This standard is specifically designed for HTTP Adaptive Streaming (HAS) of video sequences encoded in H.264, supporting resolutions up to Full HD (1920 × 1080) and sequence durations ranging from 30 seconds to 5 minutes. The model addresses quality impairments caused by representation switching, such as changes in bitrate, resolution, and frame rate, as well as initial loading

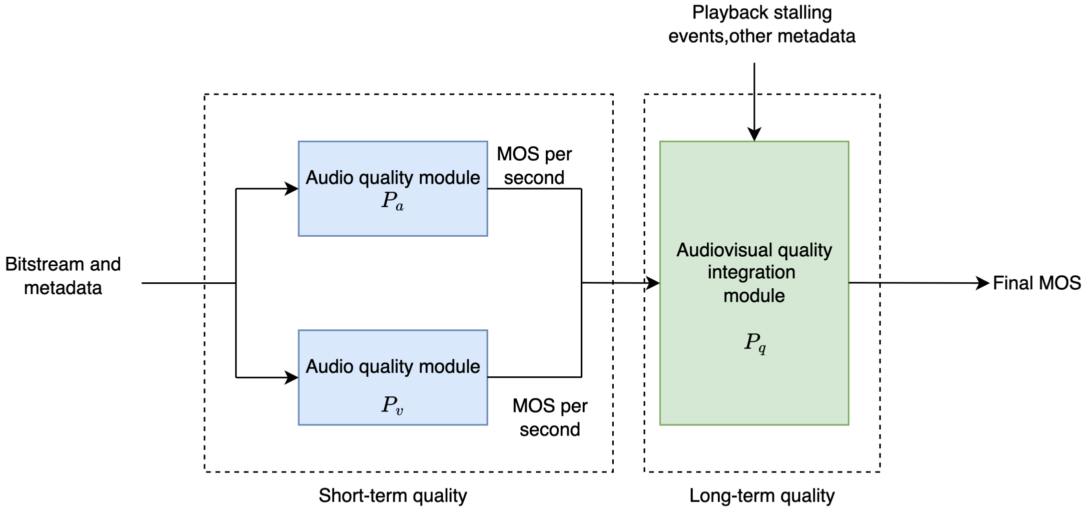

Figure 2: ITU-T P.1203 model scheme.

delays and playback interruptions (stalling).

The architecture of the ITU-T P.1203 standard is shown in Figure 2. The P.1203 model is composed of three modules devoted to: i) video quality estimation (P.1203.1, Pv), ii) audio quality assessment (P.1203.2, Pa), and iii) overall audiovisual quality computation (P.1203.3, Pq), respectively. The first two modules, Pa and Pv, require bitstreams data and specific metadata in input to produce per-second MOS lists reflecting the perceived quality over time. The last module, Pq, integrates MOS lists with additional playback-related information, such as video player behavior and playout buffer status, to generate a single final MOS score representing the overall quality of the streaming session.

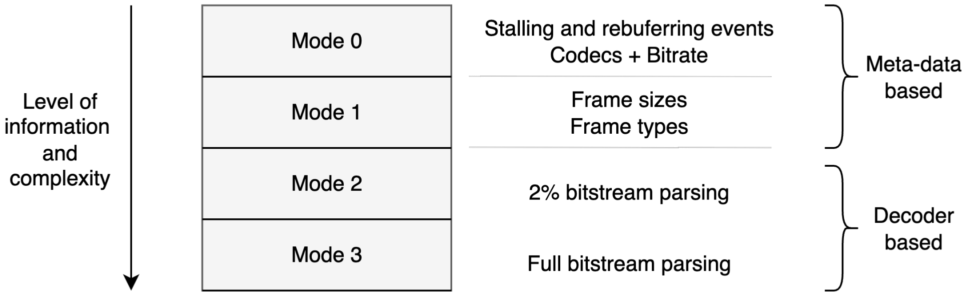

Figure 3: Different modes for estimating video quality. Mode $n + 1$ has access to the information of all mode n

To deal with data availability issues, the modules can operate in four different modes. These modes, represented in Figure 3, allow for progressively richer input, ranging from metadata-only configurations to full access to the complete bitstream, thereby supporting a flexible quality estimation framework based on available information.

The P.1203 model has been shown to provide accurate predictions of user QoE even in evaluations carried out on data collected from real-world scenarios, however, it lacks of a real-time implementation.

The ITU-T P.1203 standard has been validated across a wide range of conditions relevant to adaptive streaming. Video compression degradations were tested using H.264/AVC (High Profile) with bitrates ranging from 75 kbit/s to 12.5 Mbit/s, while audio compression degradations were evaluated using AAC-LC ($32 - 196$ kbit/s). The audio quality module (P_a) is also assumed valid for other codecs, such as HE-AACv2, AC3, and MPEG-LII, based on the previous testing in P.1201 for bitrates from $24 - 196$ kbit/s.

The validation included video content with varying spatiotemporal complexity, display resolutions up to Full HD (1920×1080), and playback on different devices (PC/TV monitors and smartphones, for

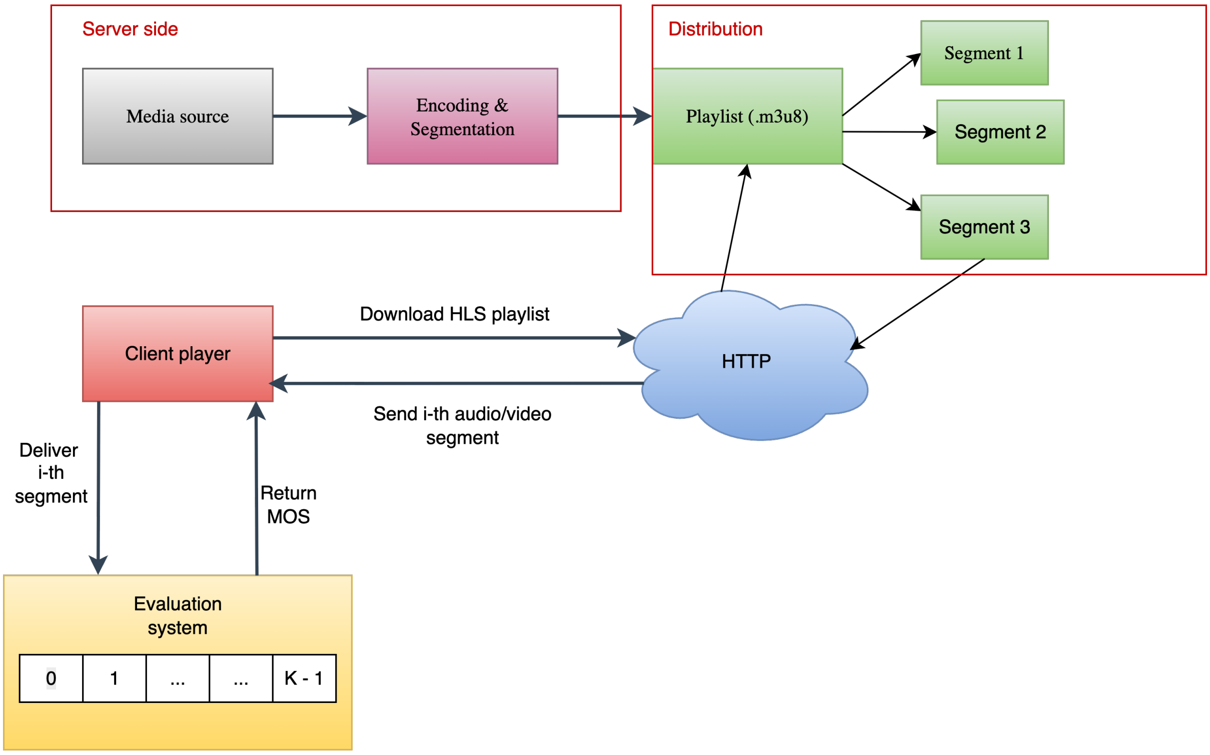

Figure 4: The proposed architecture for evaluating the MOS. Segments are progressively downloaded through HTTP and delivered to the evaluation system, which may reside on a separate machine, even outside the local network.

example, Samsung Galaxy S5). Quality variations due to media adaptation—such as switching between different bitrate or resolution layers—were considered, along with frame rates between 8 and 30 fps. Initial loading delays and stalling events were also part of the evaluation. The P_q model has been validated only for inputs of up to 5 min length — that is the duration of the test sequences that have been shown to the test subjects during subjective evaluation.

3. Proposed solution

To enable real-time QoE evaluation, the ITU-T P.1203 standard has been integrated into a local system to emulate streaming by the HLS protocol. In this setup, video and audio segments are progressively delivered to the player and, as they are displayed, simultaneously forwarded to the evaluation system. The latter, implemented as a C application, shown in Figure 4, listens to a TCP port, uses a low-latency protocol based on WebSocket for communication, using raw binary data, and is designed to be multi-threaded. Specifically, it employs a timer mechanism with two threads synchronized using a *semaphore* as a synchronization primitive, to ensure that the QoE evaluation is performed synchronously at fixed time intervals, while keeping a dedicated thread responsible for accepting incoming segments. It processes each received segment and returns the corresponding MOS to the player. For live and timely assessment, the segments are analyzed by a sliding window composed of n segments, and a result is triggered any time p segments, i.e. the sliding window step, are received. Given that each segment typically has a duration d ranging from 2 to 4 seconds, the resulting MOS is shown to the user approximately every $p \cdot d$ seconds. Therefore, the total processing time required by the evaluation system must be less than or equal to the interval between the evaluations, that is, $p \cdot d$. This condition guarantees that the evaluation process remains synchronized with the media playback timeline and avoids any latency accumulation.

For practical usage of the application, it is possible to provide a JSON-like configuration file to the evaluation system to specify parameters like n, p and the mode of operation of the standard,

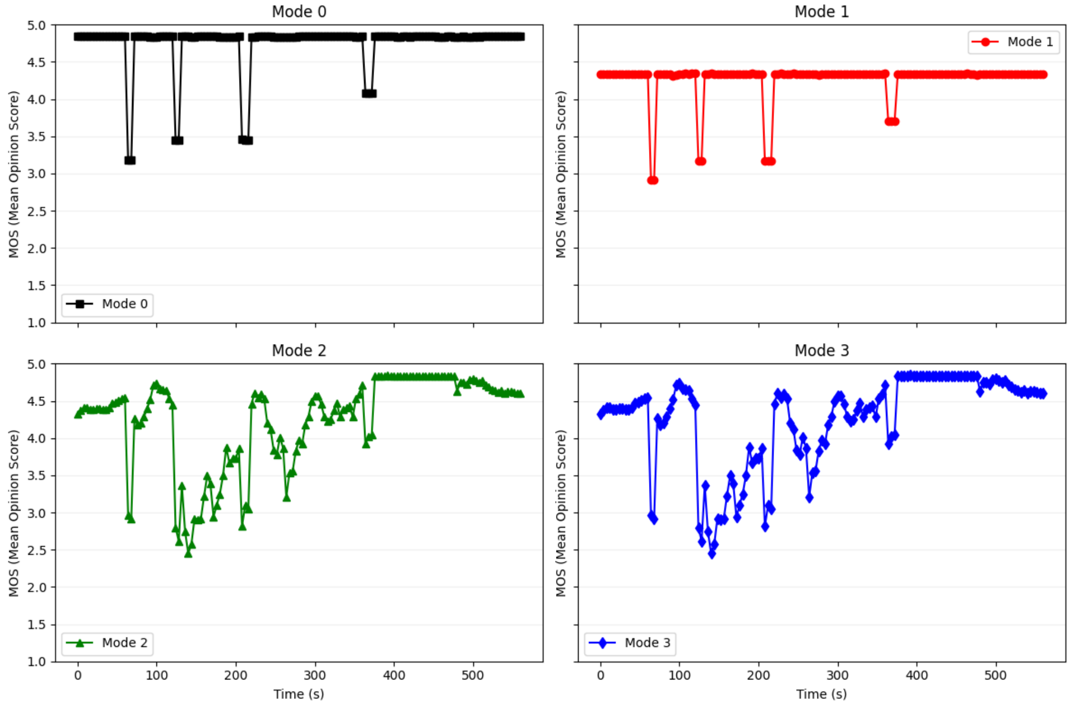

Figure 5: Temporal evolution of the MOS values over time. The data is based on a total of 285 segments, approximately 10 minutes ($600\ s$) of live streaming. $n = 5, p = 2$.

allowing flexible assessment based on different streaming requirements or experimental setups. Such configuration capabilities also enable the deployment of distributed measurements across multiple instances of the system, making it possible to evaluate the stream under various configurations and network conditions simultaneously.

4. Experiments and Results

4.1. Experimental setup

All experiments have been carried out on a single machine equipped with a multicore 1.2 GHz CPU and 10 MB of L3 cache, hosting both the player and the evaluation system. A full-HD (1920x1080) screen has been used in the tests and the ITU-T P.1203 resolution parameter has been set accordingly. The open source *HLS.js* media player version 1.6.0 has been used to play the stream, using the player API to enable logging. A fine-tuning of the playback behavior was also performed, as configuration parameters can be provided to *hls.js* upon the instantiation of the `Hls` object, to adjust the buffering settings. Although the HLS playlist has been deployed on a separate server, no delays due to routing protocols or network congestion have been introduced, since playback and evaluation are performed on the local machine.

To test the real-time performance of ITU-T P.1203, which is originally designed for sequences of around 5 minutes, longer live streaming segments (e.g. 10 and 15 minutes) are exploited. Stalling events have been also randomly simulated by limiting the network throughput between the player and the server hosting the HLS, by exploiting Google Chrome network throttler feature. A sample stream with a segment duration of $d = 2\ s$ has been used throughout the experiments.

4.2. Tests and results

The first set of tests focuses on evaluating live streaming performance across the four defined modes of operation, with particular attention to the handling of playback stalling events. These stalls occurred as

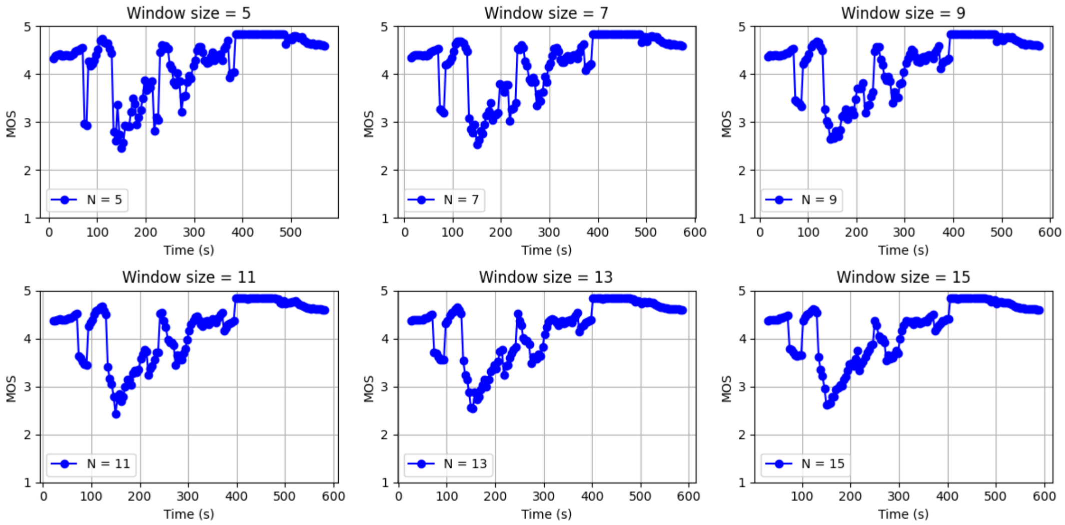

Figure 6: MOS value for different window sizes n and $p = 2$ in Mode 3.

a direct consequence of the quality degradation introduced by simulating a 3G network profile, where the available bitrate was limited to approximately 1 Mbps to emulate low user download bandwidth. The playback stalls occurred at some points during the streaming session, directly impacting the user experience. Considering that a stream with $d = 2$ s was used, the first stall appeared at second 70 (segment 35) and lasted about 8 seconds, equivalent to four segments. A second stall happened at second 130 with the same duration of 8 seconds. Later, at second 216, the stream paused again for roughly 6 seconds, corresponding to three segments. Finally, a shorter interruption took place at second 372, lasting 2 seconds, or one segment duration.

The goal is to assess how each mode manages real-time MOS computation under different playback conditions. As shown in Figure 5, modes 2 and 3 exhibit higher accuracy in assessing the perceived QoE, especially in presence of playback interruptions and quality switches.

Additional tests have been carried out to evaluate the stream in Mode 3 by varying the window size n from 5 to 15, while keeping fixed the step size $p = 2$; results are shown in Figure 6. From our experimental attempts, we can deduce that the sliding window size does not significantly affect the final MOS trend, as long as n remains within a reasonable range. However, increasing the window size could produce a slight delay in the responsiveness of the MOS updates since more segments need to be processed before a new score is computed, even though this delay is not perceptible from the user's perspective.

5. Conclusions

This work presented the integration of the ITU-T P.1203 standard into a system for real-time QoE evaluation of HTTP Live Streaming content. By simulating realistic conditions, the system enabled detailed analysis of perceived quality. Results showed that Modes 2 and 3 produced more accurate and stable MOS trends, especially during stalling events.

Future work will aim to support evaluation in Modes 0 and 1, even in encrypted or DRM-protected scenarios where bitstream-level access is restricted, as it would be sufficient to extract only the segment metadata and construct the input for the model accordingly.

Additionally, integrating the ITU-T P.1204 standard [15] would significantly extend the system's capabilities, as it supports a broader range of codecs (i.e., H.265/HEVC, VP9), resolutions up to 4K, and frame rates up to 60 fps, in contrast with the ITU-T P.1203 P_v which only addresses H.264 Full-HD 30 FPS segments.

Most importantly, the integration should aim to combine both QoE and QoS metrics to ensure a comprehensive assessment of streaming performance from both network and user perspectives, making it more suitable for modern adaptive streaming environments.

Declarations on Generative AI

During the preparation of this work, the authors used Grammarly to check grammar and spelling. After using this service, the authors reviewed and edited the content as needed and took full responsibility for the published content.

References

[1] W. Robitza, S. Göring, A. Raake, D. Lindegren, G. Heikkilä, J. Gustafsson, P. List, B. Feiten, U. Wüstenhagen, M.-N. Garcia, K. Yamagishi, S. Broom, HTTP Adaptive Streaming QoE Estimation with ITU-T Rec. P.1203 – Open Databases and Software, in: 9th ACM Multimedia Systems Conference, Amsterdam, 2018. doi:10.1145/3204949.3208124.

[2] W. Robitza, S. Göring, A. Raake, D. Lindegren, G. Heikkilä, J. Gustafsson, P. List, B. Feiten, U. Wüstenhagen, M.-N. Garcia, K. Yamagishi, S. Broom, Itu-t rec. p.1203 standalone implementation, https://github.com/itu-p1203/itu-p1203, 2018.

[3] M. Michalos, S. Kessanidis, S. Nalmpantis, Dynamic adaptive streaming over http, Journal of Engineering Science and Technology Review 5 (2012) 30–34. doi:10.25103/jestr.052.06.

[4] ITU-T, Methods for subjective determination of transmission quality, Recommendation P.800, International Telecommunication Union, Geneva, Switzerland, 1996. Series P: Telephone Transmission Quality, ITU-T Recommendation P.800.

[5] International Telecommunication Union, ITU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications, Tech. Rep. P.910, International Telecommunication Union, 2021. https://www.itu.int/rec/T-REC-P.910-202102-I/en.

[6] T. Barnett, S. Jain, U. Andra, T. Khurana, Cisco visual networking index (vni) complete forecast update, 2017–2022, Americas/EMEAR Cisco Knowledge Network (CKN) Presentation 1 (2018).

[7] A. M. Eskicioglu, P. S. Fisher, Image quality measures and their performance, IEEE Trans. Commun. 43 (1995) 2959–2965.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.

[9] T.-J. Liu, Y.-C. Lin, W. Lin, C.-C. J. Kuo, Visual quality assessment: recent developments, coding applications and future trends, APSIPA Trans. Signal Inf. Process. 2 (2013).

[10] A. Alvarez, S. Cabrero, X. G. Pañeda, R. Garcia, D. Melendi, R. Orea, A flexible qoe framework for video streaming services, in: 2011 IEEE GLOBECOM Workshops (GC Wkshps), IEEE, 2011, pp. 1226–1230.

[11] H. Bermúdez-Orozco, J.-M. Martinez-Caro, R. Sanchez-Iborra, J. Arciniegas, M.-D. Cano, Live video-streaming evaluation using the itu-t p.1203 qoe model in lte networks, Computer Networks 165 (2019) 106967. doi:10.1016/j.comnet.2019.106967.

[12] W. Robitza, D. G. Kittur, A. M. Dethof, S. Görin, B. Feiten, A. Raake, Measuring youtube qoe with itu-t p. 1203 under constrained bandwidth conditions, in: 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2018, pp. 1–6.

[13] R. Viola, M. Zorrilla, P. Angueira, J. Montalban, Multi-access edge computing video analytics of itu-t p. 1203 quality of experience for streaming monitoring in dense client cells, Multimedia Tools and Applications 81 (2022) 12387–12403.

[14] A. Raake, M.-N. Garcia, W. Robitza, P. List, S. Göring, B. Feiten, A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1, in: Ninth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, Erfurt, 2017. URL: http://ieeexplore.ieee.org/document/7965631/. doi:10.1109/QoMEX.2017.7965631.

[15] R. R. R. Rao, S. Goring, P. List, W. Robitza, B. Feiten, U. Wustenhagen, A. Raake, Bitstream-based

model standard for 4K/UHD: ITU-T p.1204.3 — model details, evaluation, analysis and open source implementation, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2020.