IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 5, 2019, accepted February 21, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901778

QoE Modeling for HTTP Adaptive Video Streaming–A Survey and Open Challenges

NABAJEET BARMAN, (Member, IEEE), AND MARIA G. MARTINI, (Senior Member, IEEE)
Wireless and Multimedia Networking Research Group, Faculty of Science, Engineering and Computing, School of Computer Science and Mathematics, Kingston University, London KT1 2EE, U.K.

Corresponding author: Maria Martini (m.martini@kingston.ac.uk)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement 643072.

ABSTRACT With the recent increased usage of video services, the focus has recently shifted from the traditional quality of service-based video delivery to quality of experience (QoE)-based video delivery. Over the past 15 years, many video quality assessment metrics have been proposed with the goal to predict the video quality as perceived by the end user. HTTP adaptive streaming (HAS) has recently gained much attention and is currently used by the majority of video streaming services, such as Netflix and YouTube. HAS, using reliable transport protocols, such as TCP, does not suffer from image artifacts due to packet losses, which are common in traditional streaming technologies. Hence, the QoE models developed for other streaming technologies alone are not sufficient. Recently, many works have focused on developing QoE models targeting HAS-based applications. Also, the recently published ITU-T Recommendation series P.1203 proposes a parametric bitstream-based model for the quality assessment of progressive download and adaptive audiovisual streaming services over a reliable transport. The main contribution of this paper is to present a comprehensive overview of recent and currently undergoing works in the field of QoE modeling for HAS. The HAS QoE models, influence factors, and subjective test methodologies are discussed, as well as existing challenges and shortcomings. The survey can serve as a guideline for researchers interested in QoE modeling for HAS and also discusses possible future work.

INDEX TERMS HTTP adaptive streaming, QoE modeling, TCP, video quality assessment.

I. INTRODUCTION

The Cisco Visual Networking Index forecasts an increase of Internet traffic, with video alone being 82% of the net consumer Internet traffic by 2021 [1]. There has been a considerable amount of work on video delivery over the Internet to meet this increased demand. With the deployment of new wireless technologies such as 4G LTE-Advanced, the available end-user bandwidth has increased considerably over the recent years and it will further increase with 5G wireless systems. However, with the emerging video formats (e.g., Ultra High Definition (UHD), High Dynamic Range (HDR), Light Field) and new services such as Virtual Reality, Social-TV, Cloud Gaming, the available network technology will not be able to meet the increased demand for high bandwidth for all the users and to satisfy users' expectations for any content, any place, any time. The new video formats such as 4K and HDR result in files of enormous size and hence call for modern video compression standards. The effort in this direction resulted in the recently introduced new video compression standard H.265/MPEG-HEVC, which on an average, for the tested sequences, is shown to achieve 50% higher compression efficiency than its predecessor H.264/MPEG-AVC [2]–[4]. VP9, a royalty-free encoder developed by Google as a competitor of the H.265/HEVC encoder, has gained much popularity and is supported by almost all browsers except for Safari. Licensing issues with H.265/HEVC and the aim to develop a more futuristic royalty-free video codec led to the creation of a consortium of industry partners called Alliance for Open Media (AOM).[1] The joint efforts of the members of AOM have since then drove to the development of the AV1 codec[2] with the final bitstream specification frozen in early 2018.

The associate editor coordinating the review of this manuscript and approving it for publication was Martin Reisslein.

[1] http://aomedia.org/

[2] https://aomedia.googlesource.com/aom/

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Recent studies comparing the performance of AV1 with x265, x264 and libvpx considering on-demand adaptive streaming applications have found it to result in the highest bitrate savings but at the cost of huge encoding times [5], [6]. The applicability of such encoders for live streaming applications remains an open question.

The advancements in the field of video streaming have recently resulted in the rise of both Video-On-Demand (VOD) (YouTube, Netflix, Amazon Video, Hulu, etc.) and Live (Twitch.Tv, YouTubeGaming) streaming services. As evident, video streaming is not a niche market anymore, and there exist a wide range of options for the consumers to choose from. Hence, as a service provider, it is no more sufficient just to provide a service, but it is equally important to make sure that the needs and expectations of the end user of the offered services are met. This has led to the shift from traditional technical Quality of Service (QoS) based assessment (see, e.g., [7]) to Quality of Experience (QoE) based assessment (see, e.g., [8], [9]).

To correctly determine the end user QoE and subsequently move towards QoE based control and management, there exists a need for the development of reliable and accurate QoE models. Such models usually take into account various network and application level factors (including several QoS factors) and aim at predicting the QoE as experienced by the end user.

Having established the importance of QoE modeling and considering that HTTP Adaptive Streaming (HAS) is the preferred video streaming technology, we present in this paper a review of existing QoE models for HAS applications. While there exist previous surveys, such as by Seufert *et al.* [10], which discuss HAS and related influence factors, and by Juluri *et al.* [11], which discuss tools and measurement methodologies for predicting QoE of online video streaming services, a survey of QoE models for HAS applications is still missing. Towards this end we present in this paper a review of the proposed QoE models for HAS applications. The major objectives of this review are:

- To classify the existing models and provide the reader with an overview of different works so far in the field of QoE modeling for HAS applications (Section V).
- To identify the different influence factors as considered by the model proponents and discuss their impact on the model design and performance (Section VI).
- To present the different subjective test methodologies used for model design and validation. We discuss how such information can favor reproducible research and steer the development of models valid in different settings and conditions (Section VII).
- To present a list of publicly available open source datasets for HAS QoE model design and/or validation (Section VIII).
- To identify existing research gaps and provide a set of recommendations for future model design and validation (Section IX).

The rest of this paper is organized as follows. We start with a brief introduction to QoE, QoE assessment methodologies and the various influence factors which need to be taken into account for QoE model design in Section II. In Section III we discuss QoE modeling and how QoE models can be classified based on the type of input information they require. Then we briefly introduce in Section IV the HAS technology. Section V reviews the existing work in the field of HAS modeling and provides a detailed discussion of the proposed models. In Section VI a detailed discussion on the effect of various influence factors is presented and in Section VII subjective test methodologies as used for model validation and/or testing by the model proponents is discussed along with their importance, advantage and shortcomings. Section VIII presents a discussion on publicly available HAS based datasets which can act as a valuable resource for model design and validation by future researchers. Finally, in Section IX we summarize our observations and findings and point out some existing gaps and challenges for future work.

II. QoE: DEFINITION AND ASSESSMENT METHODOLOGIES

A. QoE DEFINITION

The EU Qualinet community (COST Action IC1003: ''European Network on Quality of Experience in Multimedia Systems and Services'') defines QoE as: ''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'' [12], [13]. QoE takes into account the end user's experience and level of satisfaction and is of much interest to both academic and industrial players in the field of multimedia. Understanding the end users' expectations and experience is paramount to the development of future services as well as improvement of the existing technologies and services. While traditionally QoS has been used to measure the effectiveness of a service, it fails to take into account end user related factors (user expectation, environmental factors, etc.). Also, QoS is limited to telecommunication services and relies only on technical measurements. QoE on the other hand covers domains beyond telecommunications and is multidisciplinary in nature, including domains such as psychology, business, technical, environmental, etc. Figure 1 illustrates the encapsulation of QoS and QoE.

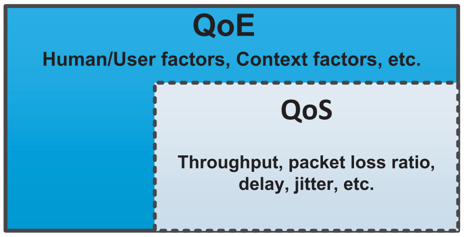
QoE
Human/User factors, Context factors, etc.
QoS
Throughput, packet loss ratio, delay, jitter, etc.

FIGURE 1. QoS and QoE encapsulation.

B. QoE ASSESSMENT

ITU-T Recommendation P.10/G.100 Amendment 5 defines QoE assessment as the process of measuring or estimating the QoE for a set of users of an application or a service with a dedicated procedure, and considering the influencing factors (possibly controlled, measured, or simply collected and reported) [13]. The main objective of QoE assessment is the design of a system which can identify the various factors and their influence on the end user QoE. Such information can then be used by the various stakeholders for optimization along the process of service delivery (encoding pipeline, load balancing, resource allocation, etc.) to provide a reasonable QoE to the end user while making optimized usage of the available resources. Lossy compression is usually required for multimedia data which need to be transported over the Internet, to decrease the required bandwidth and transport costs. During lossy compression, information is lost, with higher compression ratios resulting in a higher amount of information loss. Also, in traditional streaming technologies, transmission errors such as jitter, delay, packet loss, etc., lead to further artifacts which are annoying to the end user. Since it is almost impossible for most practical applications to provide a service without any artifact, a proper QoE model/metric can help quantifying the amount and kind of distortions and the magnitude of their effect on the end user QoE, which can then lead to the design of proper strategies to help overcoming such artifacts.

C. VIDEO QUALITY ASSESSMENT (VQA) METHODOLOGIES

VQA approaches can be categorized into two main categories: objective and subjective. Objective VQA methods are mathematical models that aim at providing a quality score which closely resembles the perceived image/video quality. Subjective VQA, on the other hand, tries to take into account the user feedback in the form of ratings and targets to estimate the video quality as perceived by the end user.

Subjective assessment scores are typically reported as Mean Opinion Score (MOS) which is the average of the opinion scores collected from the assessors. For repeatability and validation purpose, common guidelines for conducting subjective tests are issued in ITU-T Rec BT.500 and ITU-T Rec P.910 [14], [15]. These recommendations include a detailed description of the test settings, methodology and procedures that need to be followed, including data processing guidelines, such as outlier detection, etc.

The common approach to evaluate an objective quality metric's performance is to calculate the correlation coefficients and MSE values between the MOS scores estimated via the objective VQA metrics and the actual MOS scores from subjective assessment, for the same set of test sequences.

Both objective and subjective VQA approaches have inherent drawbacks. While subjective VQA provides information on the actual quality experienced by the users, it is not suitable for real-world applications. Also, conducting subjective tests incurs costs and time, and only a small number of influence factors can be evaluated due to constraints in test duration and assessors. Objective VQA using metrics such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) index, while fast and comparatively easier to implement, do not always correlate well with the end user quality [16], [17]. For two videos of different (perceivable) quality, the objective metric may provide a similar score and hence does not necessarily reflect the end user's perceived quality. Also, many objective metrics require the source sequences, which is not practical in most of the real-world quality estimation scenarios.

Quality metrics such as PSNR and SSIM were initially developed and used for Image Quality Assessment (IQA). For Video Quality Assessment (VQA), they are calculated on a frame-by-frame basis and then the final score is reported as the average of the individual scores over the full duration of the video sequence. There also exist different pooling methods to combine the scores such as Minkowski summation, exponential weighting, etc. A discussion of temporal pooling strategies is out of the scope of this paper and interested readers can refer to [18] for an interesting comparison of the pooling mechanisms and their performance in HAS applications.

Traditional models used for VQA, such as PSNR, SSIM, VQM [19], etc., are not designed for long-term quality predictions. Also, most of the traditional objective VQA metrics were designed for quality estimation of impairments due to compression and/or due to packet losses etc., during the transmission process. They do not take into account impairments such as rebuffering, quality switches etc., which are present in HAS applications. Therefore, new approaches for QoE estimation model design are required for HAS applications which take into account IFs such as rebuffering and quality switching along with impairments due to lossy encoding.

D. QoE INFLUENCE FACTORS

A QoE influence factor is ‘‘any characteristic of a user, system, service, application, or context whose actual state or setting may have an influence on the Quality of Experience for the user’’ [12]. As defined in ITU-T Rec. P.10/G.100 Amendment 5, QoE influence factors include the type and characteristics of the application or service, context of use, the user's expectations with respect to the application or service and their fulfillment, the user's cultural background, socio-economic issues, psychological profiles, emotional state of the user, and other factors whose number will likely expand with further research [13]. Influence factors on QoE can be grouped into the following four categories as described by Skorin-Kapov and Varela [20].

1) SYSTEM IFs

System IFs mostly consist of the technical aspects of quality, for example, the ones which can be measured using QoS based measurement approaches. They cover a wide range of aspects such as media related (quality switching events), network related (wired/wireless/mobile, bandwidth, delay, jitter, packet loss, etc., resulting in impairments such as temporal interruptions/pauses) or end-user device related (display resolution, playback capabilities such as supported codecs, formats, etc.).

2) HUMAN IFs

Human or User IFs include aspects which refer to the information about the end-user and related aspects. These include individual characteristics of a user such as expectations from the service, memory and recency effects, usage history of the application (e.g., browsing history, frequently played video), demographic and socio-economic background, physical and mental constitution (users' emotional state), memory, categorization and attention among many others.

3) CONTEXT IFs

Context IFs deal with factors such as location, end user environment (viewing environment, acoustic conditions, etc.), time of the day, type of usage (e.g., just casual browsing, newly released episode of favorite TV show), time of service consumption (peak time, offload time, etc.)

4) CONTENT IFs

One of the most important is the content IFs which addresses the characteristics of the content. The aspects in this category include information about the content being offered by the service/application under consideration. For example, for video, the content level IFs are duration, video type and content complexity (spatial and temporal complexity).

III. QoE MODELING

Managing Quality of Experience (QoE) in a communication system is a complex task, primarily consisting of three steps, as shown in Figure 2 and discussed in [21] and [22]. A key step in QoE management is the design of QoE models. ITU-T Recommendation P.1201 defines a QoE model as "An algorithm with the purpose of estimating the subjective (perceived) quality of a media sequence" [8]. QoE models take into account various influence factors and try to estimate the end user QoE. QoE monitoring and measurement(s) can be done by any stakeholder and the parameters measured will depend on the application and the interests of the stakeholder [23], [24]. The final step in QoE management includes QoE optimization and control, typically performed based on models or measurements. Again, the optimization process and the parameters controlled will depend on the stakeholder and the application type. In this paper, we limit our discussion to the first step, focusing on QoE Modeling for HAS applications using reliable transport protocols such as TCP or Quick UDP Internet Connections (QUIC) [25].

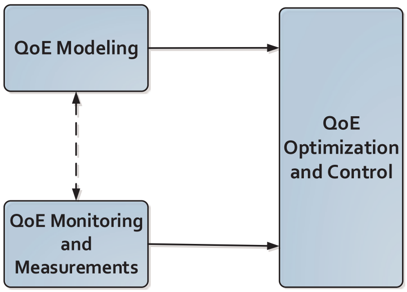

FIGURE 2. QoE management process.

A. IMPORTANCE OF QoE MODELING FOR DIFFERENT STAKEHOLDERS

QoE modeling is one of the critical steps in the QoE management process chain, as the performance of the QoE model will decide the reliability and accuracy of the next steps along QoE based management. We discuss next the importance of QoE modeling from the point of view of various stakeholders in the multimedia streaming process chain.

1) NETWORK PROVIDER

With increasing demand for OTT services, both VOD and live, there is a tremendous pressure on the network operators to provide seamless connectivity and high QoE to the end users. QoE models can help network operators identifying the various IFs and their respective impact on the end user QoE and hence allow the network operators to take necessary actions (resource allocation such as network throttling, load balancing, caching and network provisioning) to prevent user churn.

2) SERVICE PROVIDER

In today's highly competitive environment with almost similar pricing schemes, the service provider cannot rely on profit generation based solely on the provision of a service, but should also take into account different factors which may shift the user base to the competitors. For example, for a service provider measurable QoE factors such as viewing duration are of huge interest [26]. For advertisement based services, longer viewing duration implies more advertisement. On the other hand, for subscription based services, shift of even a smaller percentage of viewer base can result in significant effect on revenues. One of the disadvantages of HAS services is the requirement of additional storage space, as multiple copies of the same file are stored in the server. In such cases, optimized encoding bitrates can lead to huge storage space savings for the OTT provider while also reducing the demand for required bandwidth. Hence, proper QoE models can provide an insight into the IFs and their impact on the service, and in turn allow the service provider to take appropriate decisions/measures to ensure high end user QoE.

3) DEVICE MANUFACTURER

Nowadays, most of the device manufacturers, such as Samsung, LG, Sony, etc., are involved in manufacturing of both small screen devices (mobiles, tablets) and big screen devices (PC/TV). Different devices have different capabilities and the perceived quality depends on various factors, one of which is the device screen size. Also, small screen devices have different processing capabilities compared to large screen devices. Hence, good QoE models can provide insight to the device manufacturers, considering the device features (display size, display resolution, CPU, ram, etc.), on what settings to use such that the QoE of the end user can be maximized. Also, media-layer models (see Section III-C.1) can be used for codec comparison and hence allow device manufacturers to provide optimized encoding and decoding support so as to support the latest codecs in the shortest possible time. Many device manufacturers are also interested in QoE modeling for production of QoE monitoring solutions such as probes, QoE estimation modules etc.

4) END USER

In the end, the user is the king or queen. The success of a service will depend on the acceptance of the same by users. As mentioned in [22], successful QoE management will lead to satisfied end users as their requirements and/or expectations will be met and hence they may be further open to adopt new and complex services, leading to growth of more advanced technologies.

To summarize, QoE modeling can help us identify the various Key Performance Indicators (KPIs). The actual applicability and performance of the model will vary depending on the stakeholder as different actors involved will focus on different aspects (mostly the ones they can control). For example, in the case of HAS, a network provider may be interested in rebuffering, quality switches, etc. and their corresponding effect on QoE as they are directly or indirectly related to the network QoS parameters such as delay, jitter, packet loss, etc. A content provider may be interested more in the effect of average bitrate, segment size, video popularity, etc., for example, to save storage costs, optimized video caching, etc. At the application layer, the service provider may be interested in IFs such as adaptation frequency, adaptation magnitude, etc. to take these into account for the design of the client's adaptation algorithm.

B. QoE MODEL PERFORMANCE EVALUATION

The criteria for the evaluation of the performance of an objective QoE model, as mentioned initially in Video Quality Experts Group (VQEG) FRTV Phase I and later in VQEG FRTV Phase II [27], [28], are:

- *Prediction Accuracy* It refers to the ability of a model to predict the subjective rating scores with low error. The accuracy of the QoE model will affect the applicability and effectiveness of the QoE management process.
- *Prediction Monotonicity* It refers to the degree of model's prediction agreement with the relative magnitudes of the subjective rating scores.
- *Prediction Consistency* It refers to the ability of a model to maintain prediction accuracy over a wide range of test sequences with a variety of video impairments.

The prediction accuracy of a model can be evaluated by using the Pearson Linear Correlation Coefficient (PLCC) between the predicted and actual subjective rating scores. Similarly, the prediction monotonicity of a model can be evaluated using the Spearman's Rank Correlation Coefficient (SROCC) between the predicted and actual subjective rating scores. Finally, the prediction consistency of the model can be evaluated using measurements such as the Outlier Ratio (OR). A low OR value indicates a high consistency of prediction, with $OR = 0$ implying that the model will be stable to predict the QoE. A good QoE model should provide insight on how the IFs affect the QoE of the end user. Such insight can help various stakeholders in a more efficient and optimized system design.

C. QoE MODEL CLASSIFICATION

Depending on the application area or range of system or service the model applies to, there exist many ways to classify models such as based on model input parameters, application scope, measurement scope, etc. [22]. While there exist many approaches for classification of models, we use the approach presented by Takahashi *et al.* [29], similar to the one presented by Raake *et al.* [30] as shown in Figure 3.

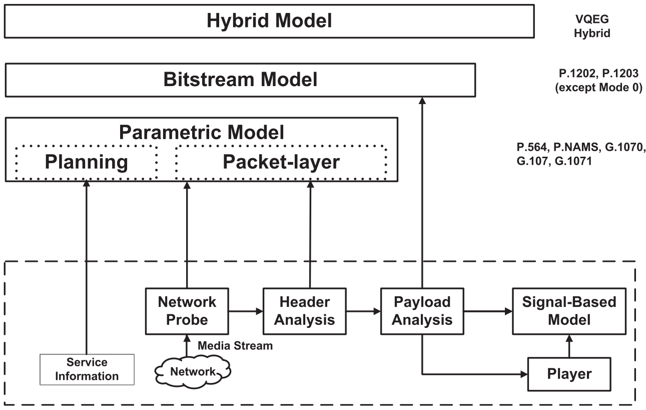

FIGURE 3. QoE model classification for streaming applications (adapted based on input from [30]).

1) SIGNAL-BASED MODELS

Signal-based models, also known as pixel-based models or media-layer models, utilize the decoded audio/video signal to estimate the video quality. Since such models do not use any codec specific information, they are widely used in codec comparison and optimization of unknown systems.

Based on the relationship between the input and output of the system, i.e., depending on the amount of source (reference) information required, VQA metrics can be classified as Full Reference (FR), Reduced Reference (RR) and No Reference (NR).

(a) *FR:* As the name suggests, FR metrics require the availability of full information of the source video. They are computed based on a frame-by-frame comparison between the reference and the distorted image/video. The source video should be available in pristine quality (unimpaired and uncompressed) so that there can be a direct comparison (e.g., pixel by pixel) between the reference and distorted image/video. Due to the availability of full source information, these metrics are usually more accurate than their counterpart (RR or NR metrics) but as such are not suitable for most real-world applications. Some of the most widely used quality metrics in the field of image and VQA are FR metrics such as MSE, PSNR and SSIM [16] and ITU-T Recommendations [31]–[33].

(b) *RR:* RR metrics have access to limited source information. Due to partial source information, they are usually less accurate than the FR metrics. Some of the RR metrics are [34]–[41].

(c) *NR:* No reference quality metrics do no use any source/reference information and try to predict the quality based on the received signal. Commonly used NR metrics include DIIVINE, BRISQUE, BLIINDS and NIQE [42]–[45]. In the absence of source information, such metrics are usually less accurate than their counterparts, FR and RR metrics.

2) PARAMETRIC MODELS

Parametric models use measured or expected packet/network related parameters to estimate the quality. These can be further classified in packet-layer models and planning models, described below.

(a) *Packet-layer models:* Parametric packet-layer models utilize only information that can be extracted from packet headers, such as bitrate, packet loss rate (PLR), frame rate, frame type, etc., and no media signal information is required. Such models are hence non-intrusive in nature and are easily deployable and computationally very inexpensive (e.g., ITU-T Rec. P.564 for speech and ITU-T Rec. P.NAMS [8], [46]). Due to the absence of any payload information, such models are not suitable for individual QoE monitoring solutions such as determination of effect of content dependence on end-user QoE.

(a) *Planning Models:* Unlike other models, planning models do not require input information from an existing service. Such models estimate the quality based on the quality planning information available during the planning phase from the networks and terminals. Information such as expected bitrate, PLR, codec type, etc. are used as input in this kind of models. Such model type includes some of the most widely used model in the field of videophone services (ITU-T Rec. G.1070 [47]), E-model (ITU-T Rec. G.107, widely used network tool for public switched telephone network (PSTN) and Voice over Internet Protocol (VoIP) [48]) and for video and audio streaming applications [49].

3) BITSTREAM MODELS

Bitstream models take into account the encoded bitstream and packet layer information. Features such as bitrate, frame rate, Quantization Parameter (QP), PLR, motion vector, macroblock size (MBS), DCT coefficients, etc. are extracted and used as input to the model. Such models are also relatively computationally inexpensive and can be used for real-time QoE monitoring. Bitstream based models have recently found application in the field of multimedia streaming services such as ITU-T Rec. P.1202, with ITU-T Rec. P.1203 being the most recently approved recommendation for adaptive audio-visual streaming services over reliable transport [9]. While bitstream based models show comparatively higher correlation with subjective quality scores, they suffer from the drawback that they are suitable for a specific codec. Bitstream models which can minimize their performance reliance on codec specific parameters such as size of MB, motion vector size, etc. will prove to be more useful and find wider acceptance.

4) HYBRID MODELS

Hybrid models are usually the most effective ones as they combine two or more of the previously described models and hence can use much more information as input compared to any of the standalone models discussed previously.

IV. HTTP ADAPTIVE VIDEO STREAMING

In this paper we focus exclusively on HTTP Adaptive Streaming (HAS) applications using reliable delivery mechanisms such as TCP and QUIC. Reliable transport protocols such as TCP make sure that all data will be delivered correctly to the destination process without any errors. This is usually achieved by a connection oriented approach between the sender and the receiver with the receiver acknowledging the receipt of packets and retransmission of lost or erroneous packets. Some of the most widely used implementations of HAS include:

- Adobe HTTP Dynamic Streaming (HDS) [50]
- Apple HTTP Live Streaming (HLS) [51]
- Microsoft Smooth Streaming [52]
- Dynamic Adaptive Streaming over HTTP (DASH) [53].

The first three are proprietary and vendor specific HAS implementations while DASH, also commonly known as MPEG-DASH, is an open source international standard developed by MPEG [54]. The underlying logic is common in all these implementations with some differences in the manifest file, recommended segment size, etc.

A. CONCEPT OVERVIEW

Figure 4 illustrates the basic concept behind HAS applications. The video file is encoded at different representation levels (spatial/temporal/quality, see Section IV-B) and then divided into chunks (also referred to as segments) of equal durations (often 2, 4 or 10 seconds, but depends on the standard/implementation) which are then stored on a server. The reverse process of first segmenting and then encoding can also be used, as currently done by most of the Over-the-top (OTT) providers to speed up the encoding process. When a first request for the video file is made by the client, the server sends the corresponding manifest file (e.g., .mpd for DASH, .m3u8 for HLS) which consists of the

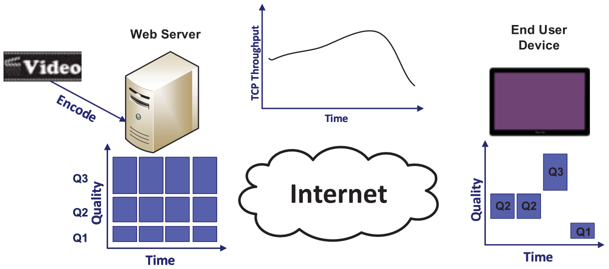

FIGURE 4. HAS Schematic (Q3, Q2 and Q1 denote high, medium and low quality level respectively).

details about the video file such as video duration, segment size, available representation levels, codec, etc. The client then requests for video chunks based on its rate adaptation logic. The client's rate adaptation logic can be broadly categorized into throughput-based, buffer-based and hybrid approach. For a comprehensive survey of the rate adaptation methods for HAS, we refer the readers to the survey paper of Kua *et al.* [55]. Figure 4 illustrates the concept of streaming assuming a throughput-based rate adaptation method. It can be observed that the client, based on its network condition, adapts the quality of the video to provide a smooth streaming experience to the end user.

B. QUALITY SWITCHING DIMENSIONS

Videos can be encoded at different bitrates (quality levels) by adjusting any/two/all of the following parameters: spatial resolution, frame rate and QP. A bitrate decrease usually indicates lower quality but the reverse does not necessarily holds true, i.e., increasing the bitrate after a certain threshold (which depends on the video content type) does not necessarily result in higher (perceived) quality videos. Figure 5 illustrates the adaptation dimensions for video encoding, described in the following:

1) *Spatial Adaptation*: The videos are encoded at different resolutions, hence decreasing the number of pixels in the vertical and/or horizontal dimensions.
2) *Temporal Adaptation*: The temporal resolution of the video is decreased by dropping some of the frames, i.e., encoding a lower number of frames per second, hence reducing the encoded bitrate.
3) *Compression Quality Adaptation (Switching)*: Increasing (decreasing) QP values results in an allocation of less (more) bits per pixel, hence resulting in lower (higher) bitrate values.

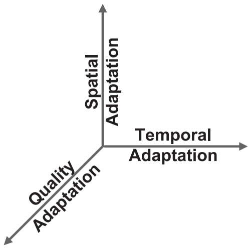

FIGURE 5. Video quality switching dimensions.

The actual dimensions of adaptation will depend on the application type and also on the content type. For most content types, compression based quality is considered the most important dimension. For similar bitrate values, spatial resolution reduction is perceived better than frame rate reduction (the actual impact of upscaling depends on the specific player used for video playback at the end user device), hence resolution is one of the most widely used adaptation dimensions [56]. For smaller screen sized devices such as mobile, tablets, etc., spatial resolution plays an important role in QoE. In general, in HAS, adaptation in multiple dimensions is perceived better than a single dimension adaptation [57] and hence is widely used by major OTT providers.

HAS is one of the most popular streaming technologies for video delivery over the Internet, currently used by the primary OTT providers such as Netflix and YouTube, with both together consisting of more than 50% of the total peak Internet traffic for fixed access networks in North America and Latin America [58]. The success of HAS can be attributed to the following advantages it offers over traditional streaming technologies:

1) *Scalability*: Since HTTP based progressive download solutions already existed, no special streaming server infrastructure is required allowing for the reuse of existing infrastructure.
2) *Reliability*: HAS uses reliable transport protocols (mostly TCP, recently QUIC) with guaranteed packet delivery and congestion control mechanisms. Hence network impairments such as packet loss do not cause any artifacts such as blurring, motion jerkiness, etc., as the lost/corrupted packets are retransmitted.
3) *Runs natively over HTTP*: HAS uses HTTP, which is firewall friendly and avoids Network Address Translation (NAT), leading to easier access to HAS services to the end users.
4) *Stateless protocol*: In HAS, the server does not store any information related to the client and/or the requests. This is useful from a network point of view (e.g., load balancing) as now each request is treated individually, hence can be handled by any of the servers, without keeping track of which server is serving which request.

Some of the challenges in the implementation of HAS include:

1) *Increased overhead*: In general, for a good streaming performance, TCP throughput of approximately twice of the video bitrate is required, which points to a major drawback of HAS applications [59].
2) *Increased storage and encoding costs*: Due to the creation of multiple quality representations for the same video/audio content, HAS solutions need much higher storage requirements compared to other traditional streaming solutions. While the costs of storage have considerably decreased over the recent years, new

video formats such as 4k and HDR results in huge file sizes. Hence, the high storage costs are still a concern for OTT providers, especially because a typical OTT provider includes millions of video contents.

3) *Quality switching*: The rate adaptation algorithm switches video quality depending on the network condition and/or buffer status. While quality switching is an important feature of HAS which helps in minimizing the number of stalling events, frequent quality switching might result in increased user annoyance.
4) *Live streaming*: During the initial years, HAS was exclusively used for VOD/Offline streaming applications. While many services currently use HAS for real-time applications, encoding videos in multiple representations in real-time remains a big challenge.
5) *Full segment download*: For most of the HAS applications, full segment download is required before playback of the segment can start. Such requirement can lead to increased cases of stalling events during video playback.

V. HAS QoE MODELING

In this section, we review the work related to models which predict the subjective quality (e.g., MOS) for HAS applications. Table 2 presents a comprehensive overview of all the models (26 models in total) reviewed in this work. The models are classified into three categories depending on their type. The table describes the various IFs considered by the models, along with the modeling method and the main observations as reported by the model proponents. It is important to note that in this review we limit the scope only to models proposed for HAS applications. For a more generic overview of models for QoE prediction, we refer the reader to the survey paper by Juluri *et al.* [11].

We start in Section V-A with a discussion of definitions and terminology along with a common set of symbols so as to have a more comprehensive understanding of the models discussed later in Section V-B. The models are presented and discussed based on their classification as described in Table 2.

A. SYMBOLS AND TERMINOLOGY

We introduce here the terminology we use for the description of the models: for simplicity and easier comparison of the models later, our goal is to use consistent terminology and symbols for all the models described.

- *Media Session*: Media session indicates video/audiovisual playback from the start till the end of the video and includes the effects of initial loading delay, rebuffering events and quality switching if any. Hence, in the presence of any of these events, the media session length will be longer than that of total video/audiovisual playback length.
- *Rebuffering*: Rebuffering refers to the event when there is no data in buffer, hence video playback is stalled (frame freezing occurs). Such events in video streaming are usually represented by a loading sign or a spinning wheel, or sometimes just the current frozen frame, and occur because of the video packets arriving late.
- *Total duration of rebuffering*: It refers to the combined length of all rebuffering events in a single media session.
- *Frequency of rebuffering*: Frequency of rebuffering refers to the number of rebuffering events per unit of time.
- *Temporal location of rebuffering*: Temporal location of rebuffering indicates the time instant when a rebuffering event starts.
- *Quality switching*: Quality switching, also referred to as rate adaptation or quality adaptation, refers to the change of quality over the duration of the media playback.
- *Quality switching frequency*: It refers to the rate of change of the quality during the media playback.
- *Quality switching magnitude*: It refers to the ‘‘gap’’ between the levels of quality switching.
- *Down-switching*: Quality switching from a higher quality level to a lower quality level.
- *Up-switching*: Quality switching from a lower quality level to a higher quality level.
- *Time on the highest layer*: Time on the highest layer indicates the percentage of time the media playback is at the highest quality.
- *Initial Loading Delay*: Also known as initial buffering, initial loading delay is the time duration between the request for video playback by the client and the actual start of the video playback.
- *Encoding Quality*: It refers to the quality of the compressed video/audio sequence due to loss of data following the encoding process. This is typically expressed in terms of an objective quality metric (e.g., PSNR, SSIM, VMAF). Some authors characterize the encoding quality in terms of bit-rate or QP value.
- *Primacy and Recency Effects*: The psychological phenomena according to which experiences which occurred recently (recency), and experiences that occurred at the very start of the session (primacy) affect more the experience quality.

Table 1 describes the parameters and corresponding symbols used in this review. In addition we use I_{QS}, I_{ILD} & I_{RB} to denote the impairment due to quality switching, initial loading delay and rebuffering respectively.[3]

B. HAS QoE MODELS

Here we present and discuss the QoE models in detail. We start with a discussion of the proposed parametric models, followed by a discussion of bitstream and hybrid models. We classify the models based on the discussion in Section III-C.

[3] I_{ILD}, I_{RB} and I_{QS} refer only to the respective type of impairment and not necessarily to how they are actually calculated

TABLE 1. Summary of symbols used in this review.

Symbol	Description
R_{AVG}	Average duration of the rebuffering events during a media session (s)
R_{MAX}	Maximum duration of a rebuffering event considering all rebuffering events for a media session (s)
R_{ALL}	Combined length of all rebuffering events in a single media session (s)
L_i	Average length of rebuffering occurring in the same temporal segment (s)
R_N	Total number of rebuffering events (pauses) during video playback)
R_F	Rebuffering frequency (number of rebuffering events per unit of time)
A	Average interval between rebuffering events (s)
R_{N_s}	Average number of rebuffering occurring in the same temporal segment
S_n	Impact of rebuffering at individual frames
T_R	Time elapsed since last rebuffering event
V_{MS}	Length of the total media session (s)
V_{LS}	Length of a single segment (s)
L_{ILD}	Length of the initial loading delay (s)
t_h	Time on the highest layer (s)
BR	Bitrate (kbps)
FR	Frame rate
QP	Quantization Parameter
N	Number of frames
MB	Macroblock Size (pixels)
μ	The standard deviation of quality information
σ	Average of the quality info
P_n	Instantaneous video presentation quality
Q_{median}	Median of the average quality
Q_{min}	Minimum of the average quality
Q_i^S	Intrinsic segment spatial quality
Q_i^T	Intrinsic segment temporal quality
m	Number of spatial resolution switching types
n	Number of temporal resolution switching types
j	Temporal Switching Type (=1 to n)
l	Spatial switching type (=1 to m)
$R_{ji}\,(S_{li})$	Number of switching type j (l) during temporal segment i
W_i	Weight factor representing the amount of degradation each temporal segment adds to the total video degradation
$P_{ji}\,(Q_{li})$	Weight factor associated with the temporal switching type j (spatial switching type l) during temporal segment i
N_{QS}	Number of quality switches
N_{SQ}	Number of segment quality bins
F_{Qn}	Frequency of segment quality bins
$B_{\bigtriangledown Q_m}$	Quality gradient bin
$F_{\bigtriangledown Q_m}$	Frequency of $B_{\bigtriangledown Q_m}$
β_R	Impact of recency effect
α_P	Impact of primacy effect
M	Memory retention
T_M	Relative strength of memory (s)
Q_t	(Instantaneous) quality at time instant t
Q_{ST}	Short-term quality
$\widehat{q}$	(predicted) Time Varying Subjective Quality (TVSQ)
q^{st}	short-term (subjective) quality per-frame
Q_{LT}	Long-term session (audiovisual) quality
T	Total duration over which MOS is evaluated (s)
B, C, α_n, β_m, α, ϵ, β, γ, δ, s_1, s_2 and s_3, a, b, c, d, e	Constant*.
$F_\tau, F_\beta, F_\delta$	Functions representing change in quality switching, amount of rebuffering and frame rate respectively
α_d	Exponential decay factor
k	Total number of segments in a media session.
t	Current time instant (s)
N_S	Total number of segments
Q_{Seg}	Quality of video segment
$Q_{Overall}$	(predicted) Overall QoE score

Note: *: The symbols just represent that the parameter is a constant and the actual value may vary from model to model.

1) PARAMETRIC MODELS

One of the earliest works towards building a QoE model for HAS applications was presented by Mok *et al.* [60]. This model quantifies QoE for HAS applications using network and application layer QoS parameters. Based on analytical models, empirical evaluation, and (subsequent) subjective tests, Mok *et al.* quantified the predicted MOS as a simple equation as:

$$MOS = 4.23 - 0.0672L_{ti} - 0.742L_{fr} - 0.106L_{tr} \qquad (1)$$

where L_{ti}, L_{fr} and L_{tr} are the levels (1, 2 or 3 corresponding to low, medium and high levels) of initial loading delay (L_{ILD}), rebuffering frequency (R_N) and rebuffering duration (R_{AVG}) respectively. The rebuffering frequency is found to be the main IF. While this work has the advantage of proposing a simple linear equation mapping application QoS metrics to QoE, the subjective assessment used to perform the regression analysis to obtain the proposed model was limited to only a single video (single content type) rated by 10 users and limited to a single resolution, which is not realistic for most HAS applications. An evaluation of the proposed model on a subjective database of new data is missing. Also, the work assumes constant network bandwidth, Round Trip Time (RTT) and Packet Loss Rate (PLR), which is not always true for the real networks and also leaves out one of the major IFs of HAS: quality switching. The authors conducted further studies to correlate QoE with network QoS, and it is observed that the rebuffering frequency increases due to decreased network throughput by packet loss and RTT. One of the major advantages of this model is the fact that content-related information is not used, hence the model can be used for encrypted traffic quality estimation by stakeholders such as network provider of third-party OTTs.

An extended version of this model is presented in [61] which takes into account user actions such as pausing and forward/backwards seeking, leading to a better model fit and an increase in its explanatory power. Video impairments may lead to various user reactions such as pausing the video, resizing, etc. and hence such factors need to be considered in the model design for a more realistic QoE model. Among all the models reviewed in this paper, this is the only work which considers user action. Based on the model, it is found that while some user actions such as pause show a marginal effect on the final QoE, other user actions such as switching the screen size have no significant impact on the final QoE score. While the proposed model is an improvement over the previous model [60] taking into account more content types, more test subjects and multiple resolutions, it is still limited by the network parameters taken into consideration and also does not take into account quality switching related impairments. Also, the performance evaluation of the model is missing.

Rodríguez *et al.* [62] model the effect of location of pauses depending on their position in the video. They propose video

TABLE 2. Overview of the reviewed models.

Models	Year	Influencing Factors	Modeling Methodology	Prediction	Major Observation(s)
Parametric Models					
Mok *et al.* [60]	2011	Initial loading delay, Mean rebuffering duration, rebuffering frequency	Regression Analysis	Overall	Rebuffering frequency is the major IF.
Mok *et al.* [61]	2011	Extension of [60] to include user-viewing actions (pausing and reducing screen size)	(Ordinal Logistic) Regression	Overall	User action pause and rebuffering have marginal significance on QoE. Other user-actions do not affect QoE significantly.
Rodriquez *et al.* [62]	2012	Number of rebuffering, rebuffering duration, Weight of temporal segment during which the rebuffering occurs	Exponential Mapping of weighted sum of segment qualities	Overall	Pauses at the beginning have higher impact than that at the middle and end.
Alberti *et al.* [63]	2013	Rebuffering frequency, rebuffering average duration, quality switching, rate, bitrate, QP and framerate	Psychometric Model (Polynomial function with rational exponent power)	Overall	QoE degradation due to encoding is on a shorter time interval compared to degradation due to quality switches and rebuffering.
Hoßfeld *et al.* [64]	2014	Quality switch amplitude, time on the highest quality level	Exponential Decay	Overall	Time on each quality level has a higher impact than frequency of rebuffering. Quality switching shadows the effect of recency and recency time.
Lievens *et al.* [65]	2015	Changes in quality, amount and total time of rebuffering and frame-rate	Linear equation	Overall	Below a certain bitrate, upscaled video results in higher quality sequences than downscaled sequences. Effect of rebuffering is non-linear.
Rodriguez *et al.* [66]	2016	Initial loading delay, number, length and temporal location of rebuffering events, number, type and temporal location of quality switching events	Exponential Model	Overall	Quality of initial segment has a greater effect on QoE. For quality switching, spatial resolution has greater impact than temporal resolutions.
Yamagishi and Hayashi [67]	2017	Video and audio bitrate, video resolution, rebuffering, content length, audiovisual interaction	Parametric quality estimation model based on modular approach	Overall (using per-sec scores)	Content length plays an important role in quality degradation due to rebuffering.
Media-layer Models					
Wang *et al.* [68]	2016	Encoding quality and quality switching	Classification and regression with Data Mining	Overall	Last two segments are found to have greater influence on the QoE.
Bitstream Models					
Singh *et al.* [69]	2012	QP, Rebuffering	Random Neural Networks	Overall	Users are more sensitive to rebuffering than encoding quality degradation QoE scores decreases faster after a certain QoE value.
Xue *et al.* [70]	2014	Initial loading delay, quality of video segments, quality switching, rebuffering	Exponential Equation	Cts + Overall	Cumulative QoE value can be obtained by weighting summation of instantaneous quality using forgetting curve of human memory
Guo *et al.* [71]	2015	Non-periodic temporal variation of quantization step	Exponential model	Cts + Overall	Composing frequency components can be used to estimate the overall quality of a non-periodic QP varying video session.
Tran *et al.* [72]	2016	Encoding quality, quality switching	Linear equation	Overall	Up-switching has a negligible effect on QoE compared to down-switching.
Tran *et al.* [73]	2016	Quality switching: amplitude and starting switching quality, rebuffering duration, initial delay	Individual expression for IFs, Curve fitting for parameter estimation	Overall	Impact of switching amplitude also depends on the starting quality. Rebuffering duration of 0.25 sec or less have negligible impact on QoE.
Robitza *et al.* [74]	2017	Initial loading delay, length and location of rebuferring events, quality switching, recency, encoding quality	Modular framework with learning based parametric model for individual equations	Overall (using per-sec scores)	Simple averaging results in almost similar performance when compared to more complicated temporal pooling strategies.
Hybrid Models					
Vriendt *et al.* [75]	2013	Per-segment metrics (PSNR, SSIM) over all frames, bitrate, quality level, segment quality	Linear equation	Overall	Number of quality switches is not important in case std dev. is taken into account. Models using PSNR, SSIM and Quality level are highly sensitive to model parameters.
Chen *et al.* [76]	2014	Short-term subjective quality, quality switching	Hammerstein-Wiener Model	Cts	Users have higher sensitivity towards quality variation in lower quality region than in higher quality region. Current TVSQ can affect the TVSQ for next 15 s.
Shen *et al.* [77]	2014	Bitrate distribution, primacy and recency effect, segment bitrate, content type	Non-linear regression for individual segments followed by weighted averaging	Overall	Bitrate is best indicator of quality. Bitrate distribution has a bigger impact on QoE than average bitrate. Also bitrate switching frequency affect QoE significantly. High initial and end quality leads to higher QoE.
Liu *et al.* [78]	2015	Initial loading delay, rebuffering, bitrate	Individual expression for IFs, Final model based on ITU-T E model for audio	Overall	Rebuffering related impairments are difficult to model and have significant effect on the user QoE.
Garcia *et al.* [79]	2015	Encoding bitrate, frame-rate, resolution, GOP, quality switches, average and total number of rebuffering events	Different short term quality models, temporal pooling for long term models	Cts + Overall	Best short-term quality model also results in best long-term quality model. Different pooling strategies results in almost similar performance.
Duanmu *et al.* [80]	2017	Encoding quality, initial loading delay, rebuffering and memory	FR VQA and Piecewise model for initial loading delay and rebuffering	Overall	For a rebuffering event of equal duration and at identical temporal location, QoE is inversely related to quality of the frame at the given time instant.
Bampis *et al.* [81]	2017	Encoding quality, length of each rebuffering event, Number of rebuffering events, Memory effect (Time since last rebuffering/rate drop)	Machine Learning	Overall	Encoding video quality is found to have a considerable on QoE while rebuffering duration is found to have small effect.
Bampis *et al.* [82]	2017	Encoding quality, rebuffering, memory effect, quality switching	Machine Learning (NARX)	Cts	QoE prediction increases when VQA are supplemented by rebuffering and quality switching related information.
Bampis *et al.* [83]	2017	Encoding quality, quality switching	Machine Learning (NARX) with multiple VQA metrics	Cts	Combining different VQA metrics can results in higher quality prediction.
Eswara *et al.* [84]	2017	Rebuffering Frequency (per minute) and duration (seconds), recency, encoding quality, quality switching	Learning based (SVR) during playback, parametric (exponential model) during rebuffering	Cts + Overall	Most recent experience has a much larger effect on the instantaneous QoE and also have a significant effect on the final QoE
Ghadiyaram *et al.* [85]	2018	Rebuffering length, total number and frequency; time since the previous rebuffering, client-side buffer model, scene criticality, perceptual quality	Hammerstein-Wiener Model for IFs; Wiener and SVR for individual prediction fusion	Cts + Overall	Using objective metric such as NIQE [43] results in increased performance when estimating overall QoE but degrades performance during continuous quality evaluation.
Eswara *et al.* [86]	2018	Short-term subjective quality, playback indicator, time elapsed since the last rebuffering event	Long Short-Term Memory Network	Cts + Overall	LSTM networks are capable of capturing the complex temporal dependencies of the non-Markovian dynamics of the QoE process. Also, mean and median pooling of continuous QoE scores are a goo indicator of overall QoE.
Duanmu *et al.* [87]	2018	Encoding quality, quality switching (quality intensity change, type of switching and quality when the switching occurred), content type.	Expectation Confirmation Theory (ECT)	Cts + Overall	For a rebuffering event of equal duration and at identical temporal location, QoE is inversely related to quality of the frame at the given time instant.

Streaming Quality Metrics ($VsQM$) as:

$$VsQM = \sum_{i=1}^{k} \frac{R_N \, L_i \, W_i}{V_{LS}} \tag{2}$$

where k, R_N, L_i, W_i and V_{LS} are the number of temporal segments of a video, number of rebuffering events, average length of the pauses, weight factor representing the degree of degradation and length of each segment respectively. Based on the subjective scores, this is then mapped into 5-point MOS scale as:

$$VsQM_{MOS} = C \, exp\left(\sum_{i=1}^{k} \frac{R_N \, L_i \, W_i}{V_{LS}} \right) \tag{3}$$

where C is a constant and all other factors are as defined in (2). Based on the subjective assessment results, it was found that the first segment has higher impairment weight compared to middle or end segments, based on which the authors conclude that the pauses, in the beginning, are more important and hence will have a higher impact on the final QoE value for streaming scenarios. This is in contradiction to other works which consider the *recency* effect to have a high impact on the QoE. The authors also propose some guidelines for subjective test assessment methodologies such as considering longer duration sequences which is more typical of HAS applications and to allow multiple viewing of the test sequences as desired by the test subjects.

An extension of this model is presented by Rodríguez *et al.* [66]. Here temporal interruptions (number, location and length of the rebuffering events) during a video session, initial loading delay and quality switching (number and location) are considered to propose a new quality metric, $VsQM_{DASH}$. The effect of initial loading delay is modeled as:

$$I_{ILD} = 5 - B \, exp(\alpha_d \, L_{ILD}/V_L) \tag{4}$$

where L_{ILD}, α, V_L and B are initial buffering delay (seconds), exponential decay factor, total video length and constant respectively. For quality switching events, the authors observe that for the same frequency of rebuffering, compared to temporal resolution changes, spatial resolution changes have a more significant effect on users' QoE. The final QoE model, $VsQM_{DASH}$, modeled using 5-point ACR MOS scores is:

$$\begin{aligned} VsQM_{DASH} = C \, exp\Bigg[\sum_{i=1}^{k} \frac{W_i}{V_{LS}} \bigg(R_{N_s} L_i \\ + \sum_{j=1}^{n} P_{ji} R_{ji} + \sum_{l=1}^{m} Q_{li} S_{li} \bigg) \Bigg] - I_{ILD} \end{aligned} \tag{5}$$

where C is a constant, i, j and l indicates the current segment, temporal switching type and spatial switching type respectively, k is the total number of segments in a media session, R_{N_s} and L_i are number and average length of pauses in the same temporal segment, m and n are number of spatial and temporal resolution switching types respectively, W_i, P_{ji} and Qli are weight factors and S_{li} is the number of switching type and I_{ILD} is the effect of initial loading delay as defined in (4). It was observed that the quality of the initial temporal segment has a greater influence on the QoE and for switching events, the spatial resolution affects the quality more than the temporal resolutions. The model is shown to be of low complexity in terms of processing and energy consumption and hence suitable for devices such as mobile phones and tablets which have limited power and processing capabilities. The proposed parametric model uses only application-level parameters and hence is suitable for QoE monitoring of encrypted traffic, specifically at the network side. The model validation is done using similar types of patterns as used for model design, and also considers a fixed number (four) of segments, hence leaving an open question about the performance of the model on unknown dataset employing different playout patterns and of different video length.

Alberti *et al.* [63] present a parametric QoE model which maps the QoS parameters to estimate QoE as:

$$eMOS = \sum_{i=0}^{N-1} a_i \, x_i^{k_i} \tag{6}$$

where $x_0 \ldots x_{N-1}$ are measured values of parameters such as video bitrate, frame rate, QP, rebuffering frequency, average rebuffering duration and quality switching rate, whereas $a_0 \ldots a_{N-1}$ and $k_0 \ldots k_{N-1}$ are tunable parameters. The authors report that QoE degradation due to encoding quality is on a shorter time interval compared to QoE degradation due to IFs such as rebuffering and quality switching. The model parameter estimation and design are done using subjective tests consisting of two video sequences and taking into account various QP, rebuffering and quality switching factors. The authors report high prediction accuracy with 0.5 MOS difference for the worst case when compared to MOS scores obtained by subjective tests. In the absence of the model validation and performance estimation (e.g., regarding the correlation of the predicted MOS with the actual MOS), the actual performance of the model remains an open question.

Hoßfeld *et al.* [64] investigate the effect of five IFs: quality switching amplitude, last quality level, recency time for the different number of switches, the frequency of quality switching and time on the highest quality level. The authors found that quality switching shadows the effect of recency and also recency time (total duration of high-quality playback after the last quality switch) does not affect the QoE. Also, it was observed that the time on each quality level has a more significant impact than that of the frequency of rebuffering. Discarding other IFs (based on statistical analysis), the authors propose a simple QoE model, considering only two IFs, which take into account the effect of amplitude (the difference between the two quality levels) and time on the highest level using an exponential relationship as:

$$y(t_h) = 0.003 e^{0.06 \, t_h} + 2.498 \tag{7}$$

where $y(t_h)$ is the predicted MOS, and t_h is the time on the highest level. The effect of switching amplitude is quantified by bounding the MOS values to the quality levels. The proposed model only proposes a parametric equation using subjective test results using a single content type and considers only two quality levels and lacks performance validation.

Lievens *et al.* [65] propose a MOS predictor, PQM, based on user evaluations as:

$$PQM(T) = \frac{1}{T+\gamma R_{ALL}} \sum_{T} Q\left[fidelity\big(t - F_\tau(\tfrac{\partial fidelity(t)}{\partial t})\big)\right] - \varepsilon\alpha^{F_\beta\left(\frac{\partial freezes(t)}{\partial t}\right)} - F_\delta\left(\frac{\partial framerate(t)}{\partial t}\right) \quad (8)$$

where F_τ, F_β, F_δ, T and R_{ALL} are functions which represent quality switching, amount of rebuffering events, frame rate, total duration over which MOS is evaluated and total time of rebuffering event, respectively. α, γ and ε are constants and Q is the encoder-side MOS for a given *fidelity* (quality level). Based on the subjective assessment using three Full HD (FHD) video sequences and various encoding and rebuffering conditions (not described in the paper) the authors observe an increase of MOS with an increase in resolution or bitrate. Below a specific bitrate, upscaled lower resolution video is found to be of higher quality compared to higher resolution video encoded at the same bitrate. On the temporal scale, no significant difference was found in between 50fps and 25 fps video while lower frame rate video (below 25fps) was rated lower with the video having quality changes rated lower than that of constant quality. Effect of rebuffering was observed to be non-linear depending on the individual duration of each event and frequency of rebuffering. The work presents only a parametric equation taking into account the various IFs but does not report the performance of the model using subjective assessment.

Yamagashi and Hayashi [67] present a quality model which was submitted as part of the competition for the ITU-T Rec. 1203. The model follows the framework used in Parametric Non-intrusive Assessment of TCP-based multimedia Streaming quality (P.NATS) consisting of an audio quality estimation module and video quality estimation module which output per-second respective quality scores which are then integrated into per-second audiovisual coding quality scores in the audiovisual-integration/temporal module. The overall QoE is defined as:

$$Q_{Overall} = 1 + (Q_{ST} - 1)S \quad (9)$$

which integrates the short term (per-second) audio-visual coding quality, Q_{ST}, with other IFs factors as:

$$S = exp(\frac{-R_N}{s_1})exp(-\frac{R_{ALL}/V_L}{s_2})exp(-\frac{A/V_L}{s_3}) \quad (10)$$

where R_N is the number of rebuffering events, R_{ALL} is the total length of rebuffering events, A is the average interval between rebuffering events, V_L is the length of the content and s_1, s_2 and s_3 are constants with positive values.

The MSQ is modeled and evaluated in terms of 5-point ACR. The proposed model parameter selection and validation are performed by using well designed and defined subjective assessment using a total of thirty 1-min audiovisual SRCs and eleven 3-minute audiovisual source sequences. While, as discussed by the authors, the test design "hides" the effect of source quality on the QoE, in terms of the reported RMSE and PLCC values, the overall model performance still looks quite promising, especially considering the fact that the model does not use any media bitstream information, resulting in a low complexity model which is suitable for encrypted QoE monitoring. The authors report that the model performs quite well for video sequences without rebuffering and also with some specific sequences with rebuffering (where the rebuffering occurs at the point where the compression quality is worse). This leads to the observation that the amount of QoE degradation due to rebuffering is dependent on the quality of the video frame where the rebuffering occurs. Hence results from other works which take into account the temporal location of pauses (e.g., [62]) can be used to further improve upon this work. Unlike most of the other works, Yamagashi and Hayashi discuss the limitations of their work such as verification of the model for the H.264 high profile (which is still the preferred and widely used profile for TV sets), validation of the model for small screen devices, performance evaluation of individual quality estimation modules, etc. Future work in this direction may include addressing these shortcomings and also the possible inclusion of other IFs such as initial loading delay, etc.

2) MEDIA-LAYER MODELS

While the most used video quality metrics (e.g., Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), Video Multimethod Assessment Fusion (VMAF)) are in this category, we focus here only on the metrics specifically developed for adaptive streaming over HTTP.

Taking into account the multi-segment and multi-rate features of HAS applications, Wang *et al.* [68] present two QoE models based on regression and classification. Using regression they propose an evolved PSNR (*ePSNR*) model based on average, maximum, minimum and standard deviation of differential PSNR (*dPSNR*), where *dPSNR* is defined as:

$$dPSNR = PSNR - PSNR_{ref} \quad (11)$$

where $PSNR_{ref}$ is the PSNR of the available highest rate segment and $PSNR$ is the PSNR of the segment under consideration. *ePSNR* is then defined as:

$$ePSNR = [a \quad b \quad c \quad d] \times \widetilde{\mathbf{Q}} + e \quad (12)$$

where a, b, c, d, e are constant values and $\widetilde{\mathbf{Q}}$ is the vector defined as:

$$\widetilde{\mathbf{Q}} = \begin{bmatrix} \underset{j}{mean}(q_{ij}) & \underset{j}{max}(q_{ij}) & \underset{j}{min}(q_{ij}) & \underset{j}{std}(q_{ij}) \end{bmatrix}^T \quad (13)$$

where q_{ij} represents the $dPSNR$ of the i^{th} video scene and j^{th} video segment. Please note that T here refers to the transpose operation. The classification method model uses weighted k-nearest neighbor (*WkNN*) based on segment bitrate and video segment position to predict QoE. Both models are evaluated using subjective tests consisting of two videos using a real-world LTE network testbed. Both regression and classification based methods are shown to provide high correlation with subjective MOS. Based on the correlation results, the last two segments have been found to have more effect than the other segments. In terms of PLCC results, the classification based model is found to have higher performance compared to the regression method, but in terms of complexity the *ePSNR* model is found to be of lower complexity.

3) BITSTREAM MODELS

Singh *et al.* [69] propose a bitstream model for QoE prediction by considering QP and frequency (R_N), average (R_{AVG}) and maximum duration (R_{MAX}) of rebuffering events. Considering H.264/AVC as the encoder, for QP estimation, the authors use the average of QP values over all macroblocks in all video frames. The playout interruptions are modeled as a function of R_N, R_{AVG} and R_{MAX} using the cumulative distribution function, $F(x)$, of the delay as:

$$F(x) = \begin{cases} \dfrac{\alpha x}{R_{AVG}}, & \text{if } x \leq R_{AVG} \\ (1-\alpha)\dfrac{x - R_{AVG}}{R_{MAX} - R_{AVG}}, & \text{if } x \in \left[R_{AVG}, R_{MAX}\right] \\ 1, & \text{otherwise} \end{cases} \tag{14}$$

where $\alpha = 1 - \frac{R_{AVG}}{R_{MAX}}$ and R_{MAX} and R_{AVG} are maximum and average values of the individual rebuffering events during the video playback. Pesudo-random values distributed uniformly on [0, 1] and the inverse function of $F(x)$ are used to obtain the playout interruption duration values based on which pauses of that duration are then inserted in the videos. The authors observe that compared to video quality due to higher QP values, users are more sensitive to rebuffering events with higher rate of drop of QoE with increase in R_{MAX}, which saturates after a certain value (6-8 seconds). In contrast, initial increase in QP results in slower QoE degradation with rapid fall in QoE at higher QP values. The 3-layer RNN model is validated using RMSE using subjective test scores. Since the model uses bitstream level information, the model suffers from inherent drawbacks of bitstream models such as limited scope of applications and also limited applicability to single codec. The proposed model was evaluated using only four content types of short duration (16 secs).

Xue *et al.* [70] propose a QoE model which combines instantaneous qualities and cumulative quality taking into account video segment quality, quality switching and rebuffering events. The instantaneous perceptual quality is evaluated using a linear model using QP values, and instantaneous rebuffering related degradation is modeled as the opposite of the weighted intensity of the interrupted frame. Initial loading delay related degradation is assumed to be constant and is modeled using the initial QP value which approximately represents the average quality of the video. The instantaneous qualities are then pooled using exponential decay temporal pooling (which takes into account the end user attention memory) to obtain the final QoE estimation. The model is shown to be of low complexity and stable with reasonable performance results. Since the subjective tests for model parameter estimation and subsequent validation are done using only two QP values, we will see later that, in the presence of multiple resolutions and QP values, the model performance is not that satisfactory.

Guo *et al.* [71] propose a model which estimates the overall quality using a linear combination of median and minimum of the instantaneous quality as:

$$Q_{Overall} = \alpha Q_{median} + \beta Q_{min} \tag{15}$$

where α and β are constants (0.68 and 0.33 respectively), and Q_{median} and Q_{min} are the median and minimum of the average quality. The instantaneous quality is obtained from QP values using the normalized quality vs. inverted normalized quantization stepsize (NQQ) model in [88]. Based on this work, the authors also observe that the qualities of the composing frequency components of a non-periodic QP varying video session can be used to estimate the overall quality of the video. Among all these frequency components (of the instantaneous qualities), the one with the worst quality has the highest impact on the final quality.

Tran *et al.* [72] present a QoE estimation model considering encoded video quality and quality variation as the IFs. The quality of the encoded video is calculated for each segment considering the average QP which is then used to model the effect of encoding quality and quality variation using the histogram of bins of segment qualities and segment quality gradients respectively. The overall session quality is modeled as:

$$Q_{Overall} = \sum_{n=1}^{N_{SQ}} \alpha_n F_{Q_n} + \sum_{m=-M}^{1} \beta_m F_{\nabla Q_m} \tag{16}$$

where α_n and β_m are model parameters, $N_{SQ}(= 5$ in this work), F_{Q_n} and $F_{\nabla Q_m}$ are number of segment quality bins, frequency of segment quality bins and frequency of quality gradient bin respectively. Segment quality bins represent the encoded video quality while quality gradient bins represent quality variations. Model parameter estimation and validation are done using subjective assessment for three videos of 74 seconds consisting of 2-second length segments and nine quality levels. A comparison with previously discussed models [71] and [75] for the given dataset shows a superior performance of the proposed model in terms of PLCC and Root Mean Square Error (RMSE). As in [78], the authors conclude that the effect of quality up-switching has a negligible impact on the overall QoE compared to that of quality down-switching. IFs such as rebuffering events, initial loading delay and quality switching of starting quality values are not taken into account in their model. The authors also assume that various representations are of the same resolution and frame-rate which is the case in many popular HAS applications which use multi-resolution video representation in their applications.

An extension of the previous model [72] is presented in [73], where the authors, in addition to quality degradation due to encoding and quality switching, also consider the effect of different initial quality, initial loading delay and rebuffering related impairments. The overall QoE is estimated as:

$$QoE_{Overall} = I_{QS} - I_{RB} - I_{ILD} \tag{17}$$

where I_{QS} is the impairment factor due to varying quality modeled using the switching amplitude and the initial quality value, I_{RB} is the impairment factor due to rebuffering duration, and I_{ILD} is the impairment factor due to initial delay modeled using a logarithmic function. The authors find that the impact of switching amplitude depends not only on switching amplitude but also on the starting quality. For example, for equal switching amplitude, down-switching in a low-quality region is worse than down-switching in the higher quality region. Also, rebuffering duration of 0.25 seconds or less have a negligible effect on the final QoE value, while rebuffering durations of more than 2 seconds can lead to extreme QoE degradation.

Robitza *et al.* [74] describe another candidate model for ITU-T Rec P.1203 competition. It follows a similar modular approach where the pooled audiovisual per second scores, representing the media quality (Q_{LT}) and degradation due to initial loading delay (I_{ILD}) and rebuffering events (I_{RB}), are combined to obtain the final Audiovisual MOS (MOSAVFinal) value as:

$$MOSAVFinal = Q_{LT} - (I_{ILD} + I_{RB}). \tag{18}$$

The model considers quality variations over time, recency effect, length and location of rebuffering events and encoding quality and is designed for sequences up to 5 minutes in length. The authors use simple averaging of the per-second scores into the final session quality score as other temporal pooling methods did not seem to provide increased performance gains. A similar observation was also reported in [18]. While the authors claim the model to be video or audio codec agnostic, the performance results for the proposed model is reported only for the mode using full bitstream information (Mode 3), hence leaving an open question about its performance for other modes (Mode 0, Mode 1 and Mode 2). Parameter selection based on the manual count of quality changes and exhaustive brute-force optimization procedure, as used by the authors, may lead to an over-fitting of the model parameters for the given test conditions and hence the performance of the same for other datasets can help in the evaluation of the actual performance gains of the model for possible real-world applications. Also, the model performance was only evaluated on PC/TV databases and its performance for mobile video streaming scenario still remains an open question.

4) HYBRID MODELS

Vriendt *et al.* [75] propose the following relationship for MOS prediction

$$M_{pred} = \alpha\mu - \beta\sigma - \gamma R_{QS} + \delta \tag{19}$$

where α, β, γ and δ are tunable parameters, and μ, σ and R_{QS} represent the average of the quality of the chunks, the standard deviation of quality information and frequency of switches respectively. Depending on how the parameter values are estimated, equation (19) can be used to obtain four different models (bitrate, objective quality (PSNR/SSIM), chunk-MOS and quality level). The chunk-MOS model uses MOS values associated with each quality level which can be estimated during the parameter tuning process, as is done for other parameters, or can be assumed to be uniformly spaced between a maximum and minimum value (which is equivalent to the quality level model). The parameter estimation is performed based on RMSE values using subjective MOS scores. Based on the results obtained in terms of RMSE, PLCC and SROCC values considering mobile phone and tablet devices, the general chunk-MOS model was found to perform better than others. As discussed by the authors, the results are limited to a single content type and a particular rate decision algorithm.

Chen *et al.* [76] model the Time Varying Subjective Quality (TVSQ) of HAS rate-adaptive video streams using a Hammerstein-Wiener (H-W) model with input and output functions as:

$$u[t] = \beta_3 + \beta_4 \frac{1}{1 + exp(-(\beta_1 q^{st}[t] + \beta_2))} \tag{20}$$

and

$$\widehat{q}[t] = \gamma_3 + \gamma_4 \frac{1}{1 + exp(-(\gamma_1 v[t] + \gamma_2))} \tag{21}$$

where $\widehat{q}$ is the predicted TVSQ, β and γ are model parameters, q^{st} is the Short Term Subjective Quality (STSQ) and $v[t]$ is the output of the linear filter of the form

$$v[t] = b^T \left(u\right)_{t-r:t} + f^T \left(v\right)_{t-r:t-1'} \tag{22}$$

where $b = (b_0, \ldots b_r)^T$ and $f = (f_0, \ldots f_r)^T$ are model parameters. Temporal distortions such as mosquito effects, jerkiness, etc., are captured using Video-RRED STSQ predictor [89]. The proposed model, while achieving good performance and providing valuable insights into the TVSQ optimization problem, does not take into account playback interruptions such as rebuffering, which limits the model application for more realistic cases. Also, the H-W model implementation as used by the authors is not suitable for videos of different durations [82].

Shen *et al.* [77] present a QoE model which takes into account segment quality, primacy and recency effects and quality switching (using bitrate distribution) as IFs. Each segment of the video is assumed to be of Constant Bitrate (CBR) and the respective encoded video quality of each segment is calculated as:

$$Q_{Seg} = \gamma \frac{BR}{MV + \delta} \tag{23}$$

where BR is the bitrate, γ, δ are constants, and MV is the motion parameter calculated as:

$$MV = \frac{1}{N-1} \sum_{f=2}^{N} std_{space}|y(f, w, h) - y(f-1, w, h)| \tag{24}$$

where $y(f, w, h)$ is the pixel value at position (w, h) of the $f - th$ frame. The primacy and recency effects are modeled as:

$$f(t) = \frac{\alpha_P}{1 + \alpha_P^2 t^2} + \frac{\beta_R}{1 + \beta_R^2 (t-T)^2}, \quad 0 \leq t \leq T \tag{25}$$

where α_P and β_R correspond to the effect of primacy and recency respectively. The overall adaptive streaming QoE is given by:

$$Q_{Overall} = I_{QS} \overrightarrow{S} \overrightarrow{W}^T \tag{26}$$

where I_{QS} represents the impact of quality switching, $\overrightarrow{S}$ is a vector consisting of the QoE of each segment as estimated using (23) and $\overrightarrow{W}$ is the weight vector for taking into consideration memory related factors (primacy and recency) using (25). The authors observe that at a particular average bitrate, down-switching achieves higher QoE than up-switching. Also, video sequences with high startup and end quality receive higher ratings due to primacy and recency effect, with the primacy effect decreasing for long video sequences. Bitrate distribution is found to be the major IF. The model was evaluated using only a single content type and also limited to the test conditions with different average bitrates. Hence the performance of the model for real-world applications remains an open question, mainly because the model does not take into account rebuffering related impairments.

Liu *et al.* [78] propose a no-reference QoE model considering both temporal and spatial quality and taking into account IFs such as initial delay, rebuffering and quality switching. The proposed overall QoE model is adapted from the ITU-T E-model [48] as:

$$DASH - MOS = 1 + 0.035R + 7 \times 10^{-6} R(R-60)(100-R) \tag{27}$$

where R is estimated based on impairment due to initial delay (I_{ILD}), stalling (I_{RB}) and quality switching (I_{QS}) as:

$$\begin{aligned} R = 100 - I_{ILD} - I_{RB} - I_{QS} \\ + \alpha\, I_{ILD} \sqrt{I_{RB} + I_{QS}} + \beta \sqrt{I_{RB} * I_{QS}}. \end{aligned} \tag{28}$$

Here α and β are estimated using subjective assessment (as 0.15 and 0.82 respectively). Based on the subjective assessment, the authors find that the initial loading delay related impairment is linear and hence is modeled using a linear equation. Impairments due to rebuffering, which are more complicated to estimate and have more dependent variables, are modeled using a combination of a number of rebuffering events, total rebuffering duration, and video motion content of the video. Quality switching related impairments are modeled using the VQM [19] metric by taking into account both encoding related impairments and impairments due to quality switching. Based on their tests, the authors observe that, for a fixed number of rebuffering events, the impairment increases monotonically with the rebuffering duration, while for a fixed rebuffering duration, the impairment due to rebuffering frequency does not increase monotonically. Also, higher frequency of rebuffering leads to higher impairment. While the model was designed and evaluated using 1-minute long video sequences, a preliminary investigation by the authors shows that it performs quite well for video sequences of up to 10 minutes duration.

Garcia *et al.* [79] present an interesting modular approach of pooling short-term quality models for long-term quality estimation which then are combined with rebuffering related information to obtain the overall media session quality. Such a modular approach leaves out the interdependencies, leading easier integration and development. The proposed model can be summarized as:

$$Q = Q_{LT} - I_{RB} \tag{29}$$

where Q_{LT} is obtained by pooling short-term audiovisual quality scores and I_{RB} is the quality degradation due to rebuffering. Six different models are used to estimate the short-term audiovisual quality scores: VQM_{AV} is the general VQM model, $PSNR_{AV}$ is the PSNR averaged per segment, $DT0$ is the frame-based model based on ITU-T Rec series [8], $DT1$ and $DT2$ are variants of $DT0$ and Dummy is 5-point scale quality levels. *degStal* is calculated as defined in ITU-T Rec series [90]. Irrespective of the pooling method used, the performance of short-term quality models is found to be a good representative of the long-term quality model performance. It is observed that the best short-term quality models also perform best for long-term models, with $DT2$ resulting in the best performance in terms of RMSE values.

Duanmu *et al.* [80] present a QoE model (referred to as Streaming Quality Index (SQI)) considering the combined effect of initial loading delay, rebuffering and encoding quality. The overall quality is computed from the instantaneous quality in a moving average fashion where the instantaneous quality at each time unit, Q_n, is considered to be a linear combination of instantaneous video presentation quality P_n estimated at the server side by frame-level VQA model and impact of rebuffering at individual frames S_n as:

$$Q_n = P_n + S_n. \tag{30}$$

Based on the assumption that each rebuffering event is additive and independent, the authors model the memory decline of memory retention due to rebuffering (based on Hermann Ebbinghaus forgetting curve [91]) as:

$$M = exp(-\frac{t}{T_M}) \tag{31}$$

where M, t and T_M represent memory retention, the current time instant and relative strength of memory respectively, which are then used in a piecewise model to get the collective effect of rebuffering on QoE degradation. The authors find that for a given rebuffering event at the same temporal location and of the same duration, the QoE is inversely related to the quality of the frame at that same temporal instant. The overall QoE value is calculated as the average of the predicted individual QoE scores. An evaluation of the existing models (PSNR, SSIM, MS-SSIM, SSIMplus [92], FTW [93], Mok *et al.* [60], VsQM [62] and Xue *et al.* [70]) and the proposed SQI using PSNR, SSIM, SSIMplus, MS-SSIM on the designed database shows that the proposed SQI model, when used with SSIMplus as the VQA model, has the highest performance, with other SQI models (SQI with PSNR, SSIM and MS-SSIM as VQA) performing better than the other compared models. The presented model is a big step forward towards QoE modeling considering both encoded video quality and rebuffering related information with reasonable performance on the given dataset. Given that the database and IFs considered in this work are somewhat limited due to the short duration of the sequences (only 10 second videos, fixed duration rebuffering events and just two rebuffering events at fixed location (start and middle)) which is not realistic, the performance of the model on more practical datasets remains an open question. We will discuss later how the model, when evaluated by other authors, does not result in high performance. The authors publicly released one of the first subjective databases for HAS application scenarios which considers rebuffering.

Bampis and Bovik [81] propose a machine learning-based framework, Video ATLAS, which combines QoE related features such as objective quality metrics, rebuffering related factors and memory-related functions to predict the end user QoE. Simple regressors combined with main IFs such as video quality, rebuffering and memory-related effects are found to provide good results. The video quality is evaluated using well-known image and video quality metrics and other IFs, such as length of each rebuffering event normalized to the duration of each video, the number of rebuffering events, number of seconds with normal playback at the maximum possible bitrate until the end of the video and time per video over which a bitrate drop took place, both normalized to the duration of individual video. The calculated features are then combined using various learning-based algorithms (Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting (GB), Extra Trees (ET) and Ridge and Lasso regression [94]) to provide a single final overall QoE score. The authors evaluate 6 objective IQA metrics (PSNR, PSNRHVS [95], SSIM, MS-SSIM [96], NIQE [43] and GMSD [97]) and two VQA metrics (VMAF [98] and STRRED [89]) on the subjective dataset and it is observed that STRRED gives the highest performance in terms of SROCC considering both a subset of the database with no rebuffering and considering the whole dataset. Based on this observation the authors conclude that IFs such as rebuffering and bitrate changes should be considered jointly and not separately which contradicts the approach of many other models discussed here (e.g., [67], [74]). In terms of content independence, MS-SSIM using ET was found to perform the best in terms of SROCC while STRRED using SVR performed best in terms of PLCC. Based on the results, it is observed that the video quality model used for the prediction of compressed video quality plays a very important role in the QoE prediction quality. Also, rebuffering duration is shown to have a small effect with a possible explanation of the duration neglect effect [99]. Using STRRED as the objective video quality metric, it was observed that for various combinations of IFs considered in this study, linear regressors Ridge and Lasso performed best in terms of SROCC and PLCC. In terms of prediction monotonicity (median SROCC) and performance (median PLCC), for a different amount of training-test data split, MS-SSIM performed the best (considering ET as the learning algorithm). Compared to other models (FTW, VsQM, PSNR, SSIM, MS-SSIM and SQI), the proposed model is shown to have superior performance when using the SSIM and MS-SSIM for all regression models.

Similar to their previous work, in [82], Bampis *et al.* present a machine learning based Nonlinear Autoregressive Network with Exogenous Inputs (NARX) model which uses objective metrics for video quality prediction, rebuffering related information and memory related features for QoE prediction. NARX is a nolinear-autoregressive model which assumes a non linear relationship between its output and inputs (delayed versions of its output, y_{t-1}, y_{t-2} and so on which helps in modeling the memory effect) along with exogenous inputs given by the vector, u_t (e.g., video encoding quality, rebuffering information) which can be defined approximately as:

$$y_t = F(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, u_t, u_{t-1}, u_{t-1}, \ldots). \quad (32)$$

As discussed by the authors, the usage of such autoregressive models for real-time QoE prediction may result in erroneous QoE prediction results due to prediction error propagation/amplification (as the prediction scores are fed back to the prediction engine). The proposed model is trained using the Levenberg-Marquardt algorithm, and QoE prediction is performed on a continuous time scale and hence can be used for continuous QoE monitoring solutions. Based on the model evaluation on LIVE-NFLX database, it is observed that the model performance varies across different playout patterns which point towards the instability of the model. Considering only objective VQA metrics, STRRED results in the best performance compared to PSNR, SSIM, MS-SSIM, NIQE and VMAF while, if rebuffering and memory effects are taken into account, both SSIM and STRRED give the best prediction results. When compared to the earlier proposed continuous QoE prediction model by Chen *et al.* [76], considering only bitrate related impaired sequences, the proposed model is shown to have better RMSE and outage rate but worse dynamic time warping (DTW) [100] distance. A possible extension of the proposed model can be to evaluate its performance for retrospective QoE prediction using various temporal pooling strategies.

Bampis and Bovik present another model in [83] which builds upon the previous two models [81], [82] addressing one of the significant shortcomings of the two earlier discussed continuous-time quality prediction models [76], [82]: instability. The authors propose a new model based on an augmented NARX approach for continuous QoE prediction taking into account degradation due to compression and rate adaptation. In contrast to the previous models where a single objective quality metric was used for encoded video quality estimation, here multiple VQA metric outputs are used as inputs for quality prediction which results in superior performance in comparison to [82]. It is observed that when VQA models are used together, the prediction quality improves significantly. This is based on the observation that while a single VQA metric alone may not be designed to take into account all types of quality impairments, multiple VQA metrics collectively can better model the distortions, which results in significant increase in prediction accuracy. The model performance evaluation is done using the same database as used by Chen *et al.* [76]. While the performance for the proposed NARX model with multiple VQA inputs is quite promising, the model complexity is quite high and is not suitable for practical applications as it does not take into account QoE degradation due to rebuffering. Model performance evaluation and possible enhancements taking into account rebuffering related impairments could be an impressive future work.

All the three models discussed above [81]–[83] are designed and evaluated using the partly public LIVE-NFLX database [101]. The LIVE-NFLX database consists of 14 source videos at FHD resolution encoded using H.264 using 8 different playout patterns (constant encoding at 250 and 500 kbps, adaptive rate drops at 66 and 100 kbps, two patterns of constant encoding with one rebuffering event, constant encoding with two rebuffering events and one with adaptive rate drops with rebuffering) rated by 56 test subjects. For a more detailed description of the database, we refer the reader to the related publication [101]. One of the major shortcomings of the previous three models is that they are all evaluated using the same database which is designed for low-bitrate applications (considering videos of max 250 kbps bitrate and min 100 kbps) such as video streaming over mobile networks and hence the performance efficiency and applicability of such models for larger displays and higher bitrate applications (PC/TV) using networks with higher throughput remains an open question. Also, it can be observed that the number of stall patterns and rate adaptation conditions are quite limited and fixed. Also, in the absence of the full database (only three out of total 14 source and respective HRCs are made public), a comparative study and further model improvement remain challenging.

Eswara *et al.* [84] present a QoE evaluation framework and a model for continuous time QoE prediction taking into account rebuffering frequency (per minute), rebuffering duration (in seconds), memory effects (recency) and objective video quality metric. Based on the premise that quality degradation due to encoding and rebuffering are mutually exclusive, the model is divided into two parts: QoE during regular playback and QoE during the rebuffering. The authors employ SVR for QoE estimation of the video during normal playback which is trained using Reduced Reference (RR) metric STRRED [89] and previous time instant QoE value. The QoE degradation due to rebuffering is modeled using the IQX hypothesis (exponential Interdependency of QoS and QoE [102]) as:

$$Q(t) = e^{-\lambda} Q(t-1) \tag{33}$$

where λ depends on the QoE value just before the onset of rebuffering and QoE value at the end of rebuffering. The proposed model is designed and validated using well designed subjective assessment. A total of 18 uncompressed reference videos covering a wide range of genres and 36 distorted videos are used in the subjective assessment. Based on the results of the model performance in terms of PLCC of the recency effect on overall QoE, the authors conclude that both instantaneous QoE and overall QoE values depend to a great extent on the most recent experience of the user. One of the significant advantages of the proposed model is that, among all reviewed works, this is the only one which considers UHD videos. Also, this is one of the first publicly available database consisting of FHD and UHD video sequences which jointly considers both quality switching and rebuffering distortion on a continuous time scale. While the authors used learning based QoE estimation using Video-RRED for standard video playback quality and exponential model based on IQX hypothesis for QoE during rebuffering state, they acknowledge that there does not exist any particular reason for their selection which can easily be replaced by other VQA and learning algorithm and parametric model respectively. The performance of such model indeed will need to be evaluated on the given dataset which can be an exciting future work. Some of the limitations of this work include usage of limited test conditions such as only two quality switching patterns, which leaves an open question about the performance of the model in real-world scenarios.

Ghadiyaram *et al.* [85] build upon the work of Chen *et al.* in [76] and their previous work in [103] which uses the Hammerstein-Wiener (H-W) model for QoE modeling as discussed previously in the discussion of the work of Chen *et al.* [76]. In addition to the rebuffering related impairments (see Table 2), client-side buffer model, scene criticality and perceptual quality IFs are first modeled mathematically. Each of these mathematical models is then used to train a Single Input Single Output (SISO) H-W model with memory, thus capturing the hysteresis effects and non-linearity of the human behaviour. Depending on the methodology used to combine the individual H-W model outputs, two variants of the continuous-time QoE prediction model are proposed. The first continuous model, TV-QoE2 uses the model outputs of the individual H-W model as input to train a Multiple Input Single Output (MISO) Wiener model (a variant of the Hammerstein-Wiener model without an input non-linearity block). The second variant of the continuous QoE prediction model, TV-QoE1, uses SVR instead of the Wiener model. The various model parameters are estimated using training data. In addition to the continuous QoE model, an overall QoE model is also proposed which takes into account number, total duration, frequency and rate of rebuffering events along with time since the last rebuffering event and perceptual quality score. The proposed model is modular in nature as additional/existing inputs can be added/removed without changes to the model structure. Also, the model is found to be computationally efficient for both training and real-time calculations. Both continuous QoE prediction model and the overall QoE prediction model are trained and evaluated using three different publicly available QoE databases ([80], [101], [103], see Table 5). In terms of the median of the per-frame correlation and RMSE between actual and predicted QoE score on a continuous time scale, the proposed model is found to outperform the SQI model in [80]. Also among the two proposed continuous-time QoE models, TV-QoE-2 performs slightly better than TV-QoE-1. In general the global QoE model providing the overall estimation of quality, while in terms of correlation and RMSE values is found to perform quite well on all the three databases, fails to provide superior performance when compared to SQI [80] and Video ATLAS [81] models. Also in the absence of taking into account quality switching as an IF, the performance of the model on real-world use cases remains an open question.

Eswara *et al.* [86] propose a recurrent neural network (Long Short-Term Memory (LSTM) network) based QoE prediction model, LSTM-QoE, to predict the time varying QoE. The authors argue that the continuous QoE is a nonlinear stochastic process which exhibits non-Markovian temporal dynamics due to the hysteresis effect which can be modeled using a network of multi-layered, multi-unit LSTMs. The predicted instantaneous QoE, $Q(t)$ is modeled as:

$$Q(t) = LSTM^{o}_{l,d}(\mathbf{x}(t), \mathbf{c}(t-1)) \tag{34}$$

where $\mathbf{x}(t)$ is the input feature vector, $\mathbf{c}(t)$ represent the set of LSTM cell states in the network, l and d are the number of LSTM layers and number of LSTM units respectively. $LSTM_{l,d}$ provides two functionalities: $LSTM^{o}_{l,d}$ for output QoE prediction and $LSTM^{c}_{l,d}$ for cell state update which is defined as:

$$\mathbf{c}(t) = LSTM^{c}_{l,d}(\mathbf{c}(1:t-1), Q(1:t-1)), \quad \forall\, t > 1. \tag{35}$$

Three IFs are considered for QoE prediction: STSQ, current playback status and total time since the last rebuffering event. STSQ, which takes into account the perceptual quality of a video segment, is calculated using traditional VQA metrics such as STRRED, NIQE, etc., as was also used in previously discussed models [82]–[84]. The proposed model is evaluated using four publicly available HAS datasets: LIVE QoE Dataset for HTTP based Video Streaming, LIVE Netflix Video QoE Database, LFOVIA Video QoE Database and LIVE Mobile Stall Video Database (see Table 5 and Section VIII for more details about the databases). Model design and evaluation over the four publicly available datasets and performance comparison against different state-of-the-art continuous quality prediction models [82], [84], [103] demonstrates a superior performance of the proposed model. The authors also report that mean and media QoE score obtained by pooling the continuous QoE scores correlates well with the reported overall QoE scores.

Duanmu *et al.* [87] investigate a novel approach where they consider an Expectation Confirmation Theory (ECT) based model design to predict the end-user QoE. The proposed model primarily takes into account the effect of adaptation intensity, adaptation type, intrinsic quality and content type IFs on the end user QoE. For a methodological study and investigation into the effect of quality adaptations (compression, spatial and temporal) on end user QoE they designed a new and now publicly available dataset which is then used for model design and evaluation (see Waterloo QoE Database (ECT) in Section VIII for more details on the dataset). The post-hoc quality of the n^{th} segment Q^{n}_{p} is defined as a function of intrinsic spatial quality (Q^{S}_{i}) and intrinsic temporal quality (Q^{T}_{i}) feature representation as:

$$\begin{aligned} Q_{Seg}(n) = f(Q^{S}_{i}(n) - Q^{S}_{i}(n-1), Q^{S}_{i}(n), Q^{T}_{i}(n) \\ -Q^{T}_{i}(n-1), Q^{T}_{i}(n)). \end{aligned} \tag{36}$$

The authors observe that the average pooling of the segment-level post-hoc quality scores correlate well with the overall QoE scores and hence the overall QoE is given by:

$$QoE_{Overall} = \sum_{n=1}^{N_s} Q_{Seg}(n) \tag{37}$$

where N_s is the total number of video segments. A comparison of the model performance with other state-of-the-art QoE models such as [9], [76], [78] etc. indicates superior performance of the proposed ECT-QoE model on the subjective test dataset. While the investigation and possible use of ECT for QoE prediction with promising results are quite impressive, the current work is limited in that the dataset used for its evaluation consisted of videos of only 8 seconds duration and one quality adaptation and some important factors (such as rebuffering events) are not considered. Future assessment on a more exhaustive dataset considering more realistic streaming scenarios can help better understand the applicability of such model for QoE evaluation.

As briefly mentioned earlier, towards building a model for adaptive audiovisual streaming services, ITU-T Rec. P.1203, also known as P.NATS was approved and finalized in Nov. 2016 [9]. The ITU-T Rec. P.1203 series describes model algorithms to predict the audiovisual quality of progressive download and adaptive streaming based applications considering reliable transport protocols such as TCP. The model proposed in this recommendation series follows a modular approach which consists of a short-term audio-video quality model providing per-one-second output scores which are then integrated along with initial loading delay and rebuffering events IFs, to give an estimate of quality for HAS media session between 10 secs to 5 minutes. The model consists of three modules, a video module Pv, an audio module Pa and an audio-visual integration module, Pq. The short-term scores from Pa and Pv are integrated into the Pq module along with rebuffering related information. Depending on the amount of required input information to Pv module, the model provides four different modes of operation: Mode 0, Mode 1, Mode 2 and Mode 3 (in increasing order of complexity). Mode 0 includes display resolution, frame rate and target and real bitrate, Mode 1 consists of all of Mode 0 and frame related information such as frame type and frame size and Mode 2 includes all of Mode 1 and partial bitstream information. Mode 3 consists of Mode 1 along with complete bitstream information. For detailed information about the models and the integration module, we refer the user to the recommendation series, P.1203. The recommendation, while indeed a significant step towards building a QoE model for HAS application, in its current form suffers from many drawbacks. For example, it assumes a perfect knowledge of buffering duration, number of re-buffering events, etc., which is not always practical. The model has been developed and validated using a fixed set of encoding settings using a single codec. While adaptive streaming applications such as HAS are codec agnostic, the P.NATS model is bitstream-based (except for mode 0), which makes the proposed model's performance codec dependent. Satti *et al.* performed a preliminary real-streaming application analysis of the P.1203 model for YouTube, Vimeo, Amazon Instant Video and proprietary DASH-based streaming framework [104]. The authors found the overall performance of Mode 0 and Mode 1 to be quite accurate for H.264 codec configuration, except for the lower quality range where the predictions were found not to be so precise. More tests in real-world applications are required to better understand the performance of the model and future development of more accurate and reliable models. A software implementation of the P.1203 ITU Rec has been made publicly available by the authors in [105] which also includes subjective ratings, per condition metadata (bitrates, resolutions, initial loading delay and rebuffering events), per-frame statistics (frame types, sizes) and bitstream level statistics (QP values and macroblock types) from four out of the total 30 datasets used in the design and validation of the recommendation. Due to the absence of the video sequences (reference as well as distorted videos), such database is of very limited use for model design and/or validation. For a more exhaustive model, joint work by ITU-T Study Group 12 and VQEG known as AVHD-AS/P.NATS Phase II is ongoing which aims towards building a comprehensive model considering a higher number of codecs (AVC, HEVC and VP9), higher frame rate (up to 60 fps), higher resolution videos (up to UHD) and a wider range of encoding settings.

C. SUMMARY

On a very abstract level, QoE can be described as:

$$QoE = f(x_1, x_2, \ldots, x_n) \tag{38}$$

where $x_1, x_2, \ldots, x_n$ are the various IFs [106]. There exist lots of IFs [10], each leading to increased complexity of the model design.

Based on the list of models in Table 2, we can observe that the focus recently has shifted from initial parametric models (which usually tried to map QoS based IFs to QoE) towards hybrid models which take into account media signals as well as impairments such as quality switching and rebuffering. Regarding IFs, we observed that while rebuffering, quality switching and encoding related factors are taken into account by most of the models, other IFs such as initial loading delay, recency and primacy effects and user engagement are considered by only a few of the models. While not all IFs has a significant impact on the final QoE, there are still many IFs whose effects are not investigated or have not been taken into consideration for model design. While Mok *et al.* [61] found user action such as pause to have a marginal effect on QoE, there may exists other user factors, which when considered together, may result in a significant effect on the end user QoE. We also observe that with the recent trend towards the design of hybrid models, the focus has shifted towards additive models where impairments due to various IFs are calculated separately and are then combined to obtain the combined effect of all the impairments as done in [74], [79], [84], among others. Such additive models have also been used in ITU-T Rec. P.1201 (Amd. 2) [90] and more recently in [9]. Hoßfeld *et al.* [106] discuss how an additive or a multiplicative model, combining existing single-parameter QoE models into a multidimensional QoE model, may lead to different results. Hence, such models need to be verified using independent subjective databases.

Regarding the model type, parametric models are not that accurate but are ideal for encrypted traffic monitoring applications. Also, such models can be used at the client-side because of low-complexity. On the other hand, bitstream models suffer from the limitation that they are specific to one codec and hence cannot generalize well, but are usually more accurate than parametric models. Usually, hybrid models are more precise than parametric, and bitstream models, but are of higher complexity and also need access to media-signals, thus limiting their application to client or server-side monitoring. Hence depending on the stakeholders involved and the desired complexity, different model type needs to be developed.

Also, while most of the models provide only an overall quality estimation for a media session, some of the models provide the prediction on a continuous-time scale. Some of the continuous-time (usually per-sec) models also provide final session quality which is usually the temporal averaging of continuous-time scores. Both approaches have their advantages and disadvantages. Continuous-time models are more useful in applications where it is possible to take corrective actions based on the estimated instantaneous quality, such as in real-time streaming application where the encoding settings and or transmission parameters may be adjusted based on the estimated QoE of the user. Some continuous-time prediction QoE models can also be used for rate adaptation in HAS applications, but such models are usually more complex as they need to be calculated in real-time. On the other hand, models providing overall QoE estimation are more suited for applications where the prediction values can be used retrospectively to design better systems, encoding strategies, network planning etc. They are usually computationally inexpensive as the parameters gathered and prediction values can usually be gathered and processed separately and not necessarily at the server/client/network side.

VI. DISCUSSION ON THE IMPACT OF INFLUENCE FACTORS

In Section V we presented the models along with the description of the IFs considered and how they were taken into account in the model design which is summarized in Table 3. Here we present a discussion of the IFs and general observations about their effect on QoE as described by the models. We not only limit the discussion to the reviewed models but also take into account the observations reported by other works, so as to get a complete understanding of the influence factors and their effects. Here we discuss the various IFs considered by the models and their respective importance in the QoE prediction. Since the IFs as considered by models and their respective observations were already discussed in Section V, here we limit our discussion only to effects of the IFs and we do not describe them for each model separately. For a more detailed discussion of how the effects of various IFs are being proposed and considered by other related works, we guide the reader to a comprehensive survey by Seufert *et al.* [10].

A. QUALITY SWITCHING

This is one of the main differentiating features of HAS compared to other traditional streaming technologies and is commonly used by HAS clients to adapt the media playback to the anticipated/experienced network conditions and/or buffer status. As the rate adaptation algorithm is not standardized as part of the MPEG-DASH standard, it varies depending on the client's rate adaptation logic. While most of the rate adaptation techniques aim at minimizing rebuffering events, frequent quality switches may lead to annoyance and hence need to be minimized.

- *Quality switching frequency*: Too frequent quality changes leads to end-user annoyance. Some of the models such as [72] and [79] consider adaptation frequency as one of the IFs for their model design.
- *Quality switching magnitude*: It refers to the "gap" between the levels of quality switching. In general, for down-switching, quality switching of lower magnitudes, i.e., in gradual steps (high → medium or medium → low) is considered to be less annoying than that of high magnitudes (abrupt high-low) [107].
- *Quality switching direction* In terms of the effect of switching direction and their relative importance, there does not seem to exist a conclusive agreement. While some observe no significant affect of up-switching when compared to down-switching [72], [77], others, like [108], find that both switching directions have a considerable impact on user QoE.

TABLE 3. HAS models and corresponding IFs.

IFs Models	Rebuffering Events	Quality Switching	Initial Loading Delay	Encoding Quality	Memory Factors	User Engagement
Mok *et al.* [60]	✗		✗			
Mok *et al.* [61]	✗		✗			✗
Rodriquez *et al.* [62]	✗					
Singh *et al.* [69]	✗			✗		
Vriendt *et al.* [75]		✗		✗		
Alberti *et al.* [63]	✗	✗		✗		
Hoßfeld *et al.* [64]		✗				
Xue *et al.* [70]	✗	✗	✗		✗	
Shen *et al.* [77]		✗		✗	✗	
Chen *et al.* [76]		✗		✗		
Lievens *et al.* [65]	✗	✗		✗		
Garcia *et al.* [79]	✗	✗		✗		
Guo *et al.* [71]		✗		✗		
Liu *et al.* [78]	✗		✗	✗		
Tran *et al.* [72]		✗		✗		
Tran *et al.* [73]	✗	✗	✗	✗		
Rodriguez *et al.* [66]	✗	✗	✗			
Wang *et al.* [68]		✗		✗		
Duanmu *et al.* [80]	✗		✗	✗	✗	
Bampis *et al.* [81]	✗	✗		✗	✗	
Bampis *et al.* [82]	✗	✗			✗	
Bampis *et al.* [83]		✗		✗		
Yamagishi and Hayashi [67]	✗	✗	✗			
Robitza *et al.* [74]	✗	✗	✗	✗	✗	
Eswara *et al.* [84]	✗	✗		✗	✗	
Ghadiyaram *et al.* [85]	✗			✗	✗	
Eswara *et al.* [86]	✗	✗		✗		
Duanmu *et al.* [87]	✗	✗		✗		

- *Time on the highest layer*: Time on the highest layer indicates the percentage of time the media playback was at the highest quality. High values of time on the highest layer indicate that the media playback was of high quality for a high percentage of media playback and hence can be used as an IF for model design as done in [64].

B. REBUFFERING

Rebuffering has long been considered as one of major IF in streaming applications and should be avoided or minimized as much as possible. The rate adaptation (quality switching) feature of HAS applications was actually designed with the major goal of minimizing rebuffering events during media playback. All models except for [64], [68], [71], [72], [75]–[77], [83] take into account one or more rebuffering related impairments as an IF in their model design.

- *Duration of rebuffering*: While the general agreement among researchers is that longer rebuffering duration leads to increased annoyance of the end user, there exists some disagreement when it comes to acceptable level of rebuffering duration. While some researchers say that rebuffering should be avoided at all costs, there exists some who say that in general rebuffering events of shorter durations (e.g., of 0.25 seconds [73]) are not noticeable and hence do not lead to QoE degradation.

Duration of rebuffering can be taken into account as considering the average duration of all rebuffering events as done in [60] and [67] or taking into account individual rebuffering event length as done in [74] and [81].

- *Frequency of rebuffering*: Highly frequent interruptions are considered annoying and can result in a very non-pleasant experience for the end user. Some of the models such as [66], [68], [80], [81] among others, consider the frequency of rebuffering as an IF.
- *Temporal Location of Rebuffering*: Temporal location of pauses, while not as important as frequency of rebuffering and duration of rebuffering, certainly plays a role in the end user QoE as a pause during an interesting scene is considered to be more annoying than one just before a scene change. The models in [62], [66], [74], [80], etc., take into account location of rebuffering as an IF in their model.

C. ENCODING QUALITY

Encoded quality plays an important role in the end user QoE. For example, higher compression may result in noticeable artifacts in the encoded video which results in decreased end-user QoE. There exist many different approaches which can be used to estimate the video quality, such as QP, bitrate, framerate, resolution etc. Many earlier works have focused on the design of QoE models and objective metrics to estimate the encoded video quality. In particular, for HAS applications, the segment quality (in terms of bitrate/QP values) can also be used to represent the encoded video quality. The type of content plays a vital role in the perceived end-user quality. The actual effect of the various quality switchings (see Section IV-B) depend on the content type. For example, dropping frames will have a less noticeable effect on a video with high motion content compared to a video with less motion content. Also, content complexity will decide the quality of the encoded media. Few models such as [77] directly consider content type information as an IF in their model design. Other use parameters such as bitrate, QP etc. or existing QA metrics as discussed below.

- *Bitrate*: Bitrate is one of the most commonly used parameters to estimate the encoded audio/video quality. Higher bitrate values usually indicate higher quality videos. The media quality can be approximated by using the downloaded bitrate values for a given session. The models in [63], [67], and [77]–[79] use bitrate values as an IF in their model.
- *QP*: QP is another commonly used factor to estimate encoded audio/video quality. Higher QP values result in higher compression and vice versa, and hence QP values can be used to determine the quality of the encoded media representation. The models proposed in [63], [65], and [69]–[71] use QP values as one of the IFs.
- *Objective Metrics*: Many models such as those proposed in [68], [75], [81]–[84] among others use already existing or modified IQA or VQA metrics to estimate the encoded video quality. Using such well established and widely used metrics benefits from the previous research work in the field of quality assessment. One of the shortcomings of such models is the need of such models to have access to the media signals, hence making them less suitable for applications where the traffic is encrypted.

D. INITIAL LOADING DELAY

Initial loading delay is usually present in all streaming applications and is used by the applications to buffer some video bits to minimize rebuffering related impairments. The general agreement is that while shorter initial loading delays do not have a significant impact on QoE, with some users actually preferring higher initial loading delay than rebuffering [109], very long initial loading delays may lead to user dis-satisfaction which depends on application type and usage scenario. Initial loading delay is used in models such as [60], [61], [66], [67], [70], [72], [74], [78], and [80].

E. MEMORY RELATED FEATURES

Memory effects such as primacy and recency have recently found application in the field of quality assessment. In video streaming applications, primacy related factors may refer to experience due to initial loading delay, starting quality etc. while recency related factors may refer to effects due to quality level, rebuffering events etc. towards the end of video playback. Only a few of the models directly use memory related factors. In general, primacy effects are considered not that important, especially when considering long video sequences as it is believed other factors will shadow the effect towards the end [64]. Shen *et al.* [77] use primacy in their model with the observation that higher quality at the start leads to higher experience quality ratings but since they use short-duration sequences in their tests, this observation validity for longer duration sequences remains questionable. The recency effect is more widely used memory-related factor with many studies reporting a high correlation between the quality towards the end and the score provided by the end user [68], [70], [74], [77], [80], [81].

F. USER ENGAGEMENT

User engagement refers to user actions during the media playback, such as pause, seek forward/backward, aspect ratio change (full-screen, etc.) which also influence the final end user QoE. In the absence of recommended practices for such user behavior related measurements, such factors are not considered in the models reviewed in this paper with the exception of Mok *et al.* [61], where the authors take into account end user actions for the design of their model. Only few works so far have investigated the user behavior and its effect on the end-user QoE [110].

VII. HAS QoE MODELS: SUBJECTIVE TEST METHODOLOGIES

Table 4 summarizes the subjective assessment methodologies as used by the model proponents for their model design and/or

TABLE 4. Summary of subjective evaluation methodologies used by the models (D: Duration (seconds), V/AV: Video/Audiovisual), N_V: Number of videos, NA: Not available.

Models	Test Environment	Network	Device	Test Methodology	Number of Test Subjects	N_V	Resolution	D (s)	V/AV	Codec
Mok *et al.* [60]	NA	Fixed Network (simulated)	NA	NA	10	1	864 × 480	87	AV	H.264/ ACC
Mok *et al.* [61]	*Real world*	Internet	NA	7-point Likert scale	22	3	240p, 360p, 480p, 720p	400	AV	H.264
Rodríguez et al. [62]	Real world	Offline	NA	ACR	96, at-least 15 scores per PVS	3	360p	120, 240	AV	H.264/ ACC
Singh *et al.* [69]	NA	Offline	NA	SS	15	4	720p	16	NA	H.264
Vriendt *et al.* [75]	Crowdsourcing	LTE + Offline	Tablet, Mobile	NA	500 (16-26 scores per PVS)	2	NA	120	AV	H.264
Alberti *et al.* [63]	Lab	WiFi, 3G	NA	DSIS	10	2	720p	NA	NA	MPEG-4 AVC
Hoßfeld *et al.* [64]	Crowdsourcing	Mobile traces (controlled)	Various devices	ACR	710 (at-least 82 scores per PVS)	1	360p, 180p, 90p	14	AV	H.264
Xue *et al.* [70]	Controlled	Offline	NA	SS (0-10)	30	10	NA	10-15	NA	H.264
Shen *et al.* [77]	*Controlled*	Offline	NA	SS (1-5)	141	*1*	720p	S	NA	NA
Chen *et al.** [76], Bampis *et al.* [83]	Lab	Offline	58" HDTV	Continuous Scale 1-5	25	3	720p	300	NA	H.264
Lievens *et al.* [65]	Lab	Offline	Screen	SS	17, 18, 18	3	1080p	10, 8, 15	NA	H.264
Garcia *et al.* [79]	Lab	Offline	42" display	ACR	30	8	1080p, 720p, 360p	10, 42-71, 180	AV	H.264/ AAC
Guo *et al.* [71]	Crowdsourcing	Offline	Various devices		20	4	720p	10	*V*	NA
Liu *et al.* [78]	Lab	Offline	Tablet	100 point discrete scale	47	4	*1280 × 768*	60	V	H.264
Tran *et al.* [72]	NA	Mobile Bandwidth traces + Offline	NA	ACR	25	3	720p	74	V	H.264
Tran *et al.* [73]	NA	Mobile Network	14" screen	ACR	66	3	720p	74	V	H.264
Rodríguez *et al.* [66]	(partial) Lab	Offline	21.5" display	ACR	119, at-least 15 scores per PVS	3	854 × 480, 360p, 400 × 224, 320 × 180	60-240	AV	H.264/ ACC
Wang *et al.* [68]	Lab	LTE	NA	ACR	90, at-least 6 scores per PVS	2	988 × 420	NA	AV	H.264/ ACC
Duanmu *et al.** [80]	Home setting	Offline	LCD	100 point cts scale	25	20	1080p	10	V	H.264
Bampis *et al.** [81,82]	Offline	Offline	Mobile (Samsung S5)	Continuous (Likert) Scale 0-5	55	14	1080p	≥ 60	V	H.264
Yamagishi and Hayashi [67]	*Lab*	Offline	Mobile, TV/PC	ACR	24	30	1080p	up to 5 mins	AV	H.264
Robitza *et al.* [74]	*Lab*	Offline	Mobile, TV/PC	NA	NA	NA	*up to 1080p*	up to 5 mins	AV	H.264
Eswara *et al.** [84]	Lab	Offline	TV	SSCQE-HR	21	18	2160p, 1080p, 720p, 360p	120	V	H.264
Duanmu *et al.** [87]	Home setting	Offline	LCD Monitor	11-point quality scale	36	18	1080p, 480 × 270, 768 × 432	4, 8	*V*	H.264

Note: The values in italics indicate that the corresponding parameter was not explicitly mentioned but can be deduced based on the discussion in the paper. The model in [83] was validated using the database from [76], hence placed together in the table. Models [81], [82] were designed and/or validated using the LIVE-NFLX database. Models with * are associated to publicly available databases. The model in Ghadiyaram *et al.* [85] and the model in Eswara *et al.* [86] were designed and tested using publicly available databases [111], [112] [113] and [112], [113], [114], [115] respectively (see Table 5 and Section VIII for more details about these datasets).

validation. It can be observed that for many models there are certain fields with missing information (marked by NA in the table). The lack of such information might leave the reader with a gap in understanding the actual applicability and validity of the proposed model(s) for specific application scenario and also limit their reproducibility and comparability with other existing models. Hence new works which propose a model should provide as much information as possible about the considered conditions for the reader to understand both its advantages as well as limitations and also the applicability of the model in real-world applications for QoE estimation. Next, the different individual fields are discussed in detail.

A. DISPLAY DEVICE INFORMATION

Many studies in the past have found a strong correlation between the device and QoE, with some even reporting high correlation between the type of display and QoE (for the same display size) [116]. Also, display size is shown to have a great effect on QoE, with impact of higher resolution becoming more prominent in displays of larger size. As evident from Table 4, most of the models do not mention the display type (mobile/tablet/PC/TV, etc.) and size of the display. Without such validation of the models for different display size and display types, their applicability and performance remain questionable for real-world applications.

B. TEST SEQUENCE DURATION

Until recently, model design and validation was performed using test sequences of 10-15 seconds duration which is also recommended by ITU-T Recommendations [14], [15]. Short duration sequences for such model design were sufficient as they mostly only dealt with perceptual video quality due to loss of information due to compression, packet loss, errors during transmission etc. On the contrary, short duration sequences are not sufficient for effective consideration of IFs such as rebuffering, quality switching, primacy and recency effects, etc. For proper modeling of these effects the sequence duration should be longer, possibly between 3 and 5 minutes, which is the common viewing duration for most watched videos streamed over the Internet [66], as considered by some models in [67], [74], [76], among others.

C. NUMBER OF SOURCE VIDEOS

As discussed earlier in Section VI-C, the effect of compression for a given parameter (e.g., bitrate, QP, framerate) depends to a great extent on the content complexity. For a model to give a stable performance and to be applicable to more practical scenarios, it needs to be validated for different content types. As is evident from the table, some of the models were designed and validated using few source sequences and hence their effectiveness for other content types remain questionable.

D. VIDEO RESOLUTION

Most of the earlier works were limited to low source resolution such as CIF [77], and SD [60], [62], [64], [66]. Some more recent works have considered higher source resolution formats such as HD [63], [71], [72], [76], FHD [65], [67], [74], [79] and only one work has considered UHD sequences [84]. Also, spatial resolution adaptation, which consist of encoding the video at a lower resolution (called as encoding resolution), is one of the most commonly used strategies for quality adaptation by almost all major OTT service providers such as YouTube, Netflix, Amazon Prime, etc. While some of the works such as [66], [67], [74], and [79] have considered such multiple resolution-bitrate pair encoding conditions, many others only consider quality adaptation at a single resolution (by using different bitrates/QP settings) and hence those might not lead to satisfactory performance when used for quality evaluation of such applications.

E. MODEL PERFORMANCE EVALUATION

As discussed in Section III-B, a performance evaluation of a model for consistency, generality and prediction accuracy can be done using Outlier Ratio (OR), Spearman's Rank Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC) respectively. Some of the models lack a complete validation (e.g., [60], [61], [64], [65]), which leaves an open question about the performance of the models on unknown datasets and/or real-world applications. Also, a comparison study of the proposed models with other existing models is absent in most cases except for a few like [81] and [82].

F. VIDEO/AUDIOVISUAL SEQUENCES

Some of the proposed models are limited to video only (e.g., [72], [73], [78], [80], [81]), without considering audio in their test sequences. This is not typical of real world scenarios where most of the media consumed is audiovisual. Also, none of the studies so far have included non-synchronized audio-video playback at the end user device. It has been found that audiovisual quality estimation is more challenging than video alone due to the complex nature of HVS with cross-modal interactions measured on an average of 0.5 on a 5-point MOS scale [117].

G. CODEC

While currently H.264 remains one of the most widely used codecs, the limitation of the proposed models to one codec makes one question the future applicability of such models; similarly some models only refer to a particular application. For example, many applications like YouTube, etc., support multiple encoders. Hence a proposed model dependent on codec related parameters (e.g., bitstream based models) may result in good performance but will fail for videos encoded using another codec but streamed using the same application. A possible solution for such applications will be designing models which take into account the type of codec used and then accordingly changing the parameters to compensate for the differences between the codec performance or bitstream syntax. An interesting work currently in this direction is on-going under the joint collaboration of VQEG and ITU

TABLE 5. Publicly available HAS databases.

Database Name	Video Resolution	Framerate (fps)	Device Type	Test Methodology	Stalling	Quality Switches	Duration	Prediction Scores	Access Link
LIVE QoE Database for HTTP-based Video Streaming [76]	720p	30	TV	Continuous (ACR) Scale 1-5	No	Yes	300 secs	Cts + Overall	[115]
Waterloo QoE Database [80]	1080p	24, 25, 30, 50	TV	100 point continuous scale	yes	yes	10 secs	Overall	[111]
LIVE Netflix Video QoE Database [101]	1080p	24, 25, 30	Mobile	Continuous (Likert) Scale 0-5	yes	yes	at least 1 minute	Cts + Overall	[112]
LFOVIA Video QoE Database [84]	(up to) 2160p	24, 25, 30, 60	TV	SSCQE-HR	yes	yes	120 secs	Cts + Overall	[114]
LIVE Mobile Stall Video Database II [103]	(up to) 1024x576	30	Mobile	SSCQE (ACR) Scale 1-5	yes	no	29-134 secs (including stalling)	Cts + Overall	[113]
Waterloo QoE Database (ECT) [87]	(up to) 1080p	30, 5, 10	LCD monitor	ACR 11 point scale	no	yes	4 and 8 secs	Cts (per 4 sec segment) + Overall	[119]
LIVE-NFLX-II Subjective Video QoE Database [120]	1080p	up to 60	Computer Monitor	Single-stimulus continuous quality	yes	yes	≈ 25 secs	Cts + Overall	[121]

project called AVHD/P.NATS Phase 2 which includes bitstream and pixel based models considering three encoders (h.264, h.265 and vp9).

VIII. PUBLICLY AVAILABLE HAS DATASETS

Based on our discussion so far, it is clear that very few of the works have made their implementation and/or the dataset public. Recent years have seen tremendous growth in the field of VQA, one of the main reasons behind which was the availability of open source databases such as LIVE Video Quality Database [118]. The availability of such open source datasets allows researchers to gain comparable and more generalizable results for VQA, QoE prediction modeling, etc. by providing a baseline for comparing the performance of newly proposed models and metrics against the existing state-of-the-art metrics. We discuss briefly in this section the seven currently publicly available datasets and their advantages and limitations in terms of their suitability for being used as a benchmark for HAS models design and/or validation and comparison. These are reported in Table 5 and discussed in the following.

1) *LIVE QoE Database for HTTP based Video Streaming* is one of the first publicly available dataset for modeling continuous time-varying subjective quality. The available videos are of 720p resolution and 300 seconds duration, obtained by concatenating smaller duration videos. The quality switching is performed only using the quality (compression) adaptation dimension and does not include multiple resolution-bitrate pairs, which is more realistic of the real-world applications. While this dataset is very useful for studying and/or modeling the continuous time quality varying prediction models, in the absence of other impairments as commonly observed in real-world HAS based applications (rebuffering events, etc.) it is quite limited in scope for the design and validation of a comprehensive HAS QoE model.
2) *Waterloo QoE Database* consists of 20 uncompressed HD videos and 60 compressed videos obtained by encoding the videos at three different bitrate levels (500 kbps, 1500 kbps and 3000 kbps) and 60 each by introducing a 5 second stalling event at the start and middle of the video playback resulting in a total of 180 distorted video sequences. While this dataset includes both stalling and quality switching, as discussed previously, this is fully realistic as the stalling events are of fixed duration as well as at fixed locations (start and middle of video playback). Also, quality adaptation is considered based on only one dimension (compression) not taking into account other adaptation dimensions (spatial and temporal).
3) *LIVE Netflix Video Quality of Experience Database* consists of subjective ratings considering 14 source video contents and 112 distorted video sequences obtained by compressing the videos using the H.264 encoder and eight different playout patterns (including rebuffering events). The video dataset is limited to a single resolution of 1080p and of different frame rates (24, 25 and 30 fps). One of the notable shortcomings of this dataset is that since it uses eleven copyright-protected videos out of a total of 14, only three source videos and the corresponding distorted videos are provided in this dataset. While most of the commonly used FR and RR metric values are already

provided, such a dataset is not suitable to evaluate custom QoE models.

4) *LFOVIA Video QoE Database* consists of 18 uncompressed reference videos and 36 distorted video sequences of 120 seconds duration and is the only dataset so far which includes videos of resolution up to 4K. The dataset considers both rebuffering events (rebuffering frequency and rebuffering duration) and quality switching (multiple resolution-bitrate pairs) which are representative of real-world conditions (though the ideal fixed duration up and down switching may not be too realistic). Such a dataset, which includes both continuous and overall scores, is comprehensive enough for design/validation of real-world applications.
5) *Live Mobile Stall Video Database II*, which focuses only on stalling events, consists of 24 reference videos and 174 distorted videos of 720p resolution generated using 26 different stalling patterns. The dataset provides both continuous as well as retrospective scores. Such a dataset can be used to study and probably model the effect of stalling on user QoE, but, in the absence of sequences and corresponding subjective ratings taking into account other IFs which may affect end user QoE in typical HAS based applications, it is not exhaustive enough for design and/or validation of QoE models.
6) *Waterloo QoE Database (ECT)* consists of 12 source videos 8s long, which are then further segmented to 4s segments (referred to as short segments). The short segments are then encoded into seven different representation sets obtained by encoding them at different quality, frame rates and resolution. By concatenating the 4s segments, 8s segments are obtained to represent different adaptation types (quality/spatial/temporal). A total of 168 4s short segments and 588 eight sec segments and their corresponding subjective ratings (overall and continuous (per segment)) are made available in the dataset. The dataset can be used as a baseline towards studying the effects of quality adaptation but is limited in many aspects, such as single adaptation event only and missing impact of other IFs, hence is not comprehensive enough for design and/or validation of HAS models.
7) The latest, newly designed, *LIVE-NFLX-II Subjective Video QoE Database* is one of the most comprehensive databases available till date. The database consists of 15 source videos and a total of 420 distorted sequences (using seven mobile network traces and considering four client adaptation algorithms) but is limited in that it considers only one resolution. The encoding bitrates are obtained using the recently proposed Dynamic Optimizer [122]. The use of four different adaptation algorithms in the database is useful to investigate the effect of such client-side adaptation on end user QoE and hence, in the design of a more exhaustive QoE model. The database includes both continuous as well as retrospective prediction scores. Additionally, an open-source Python based tool called Psychopy, to generate and display visual stimuli and collect continuous per-frame subjective ratings, is made available.

IX. CONCLUSION, CHALLENGES AND FUTURE WORK

In this paper, we surveyed the key QoE models for HAS applications. It was observed that rebuffering, quality switching and encoding related impairments are the most widely considered IFs. It is interesting to note that context IFs such as viewing environment, video popularity, type of usage, etc. are still not considered by any model except for one by Mok *et al.* [61]. It is also observed that most of the proposed models are limited in several aspects (considered IFs, performance evaluation, modeling of IFs/model, etc.), with a general comprehensive QoE model still far away from being ready. Regarding the effect of various IFs on the end user QoE, there remains a disagreement in the research community on the relationship and importance of a particular IF on the end user QoE. For example, some of the models advocate the usage of memory-related features, while others ignore them with the reasoning that such factors do not have a significant effect on the final QoE. More systematic, well-designed, large-scale subjective tests are required to quantify the impact of various IFs, as done in [108] for quantifying the effect of resolution switching on QoE.

One of the biggest challenge currently faced by the research community involved in QoE modeling for HAS applications is that it is almost impossible for a single proponent to design and conduct all-inclusive and comprehensive subjective test(s) due to high costs and time constraints, especially when considering that, when considering multiple IFs, the number of possible test conditions is enormous. This has been further hindered by the lack of open source databases. One of the primary reasons behind the progress in the field of quality assessment for image and VQA can be attributed to the open source databases such as LIVE Video Quality Assessment Database [118] which facilitated design and comparison of many quality metrics. For HAS applications related models, out of the 28 reviewed models, only few have made their databases public (see Section VIII).

This calls for a need for the research community to move towards reproducible research by making work available in the form of open source databases. As evaluated and discussed by Tavakoli *et al.* [107], subjective data gathered across different lab contexts provide comparable results. Therefore, there is a need for the design of a methodological approach for subjective test assessment procedures for HAS applications so that the results and observations can be reproduced, reused and compared. There exist several methods to perform subjective assessment, but, to make the results reproducible, a set of standardized methods are published by ITU-T in the form of recommendations, which discuss the methodology for deciding various test conditions such as the selection of proper test sequences, display settings, test environment, etc. [14], [15]. While such strict adherence to lab-based conditions are not imperative for HAS related subjective tests, following proper subjective test methodologies and, even more importantly, reporting the conditions, can lead to easier understanding and reuse of results by other researchers. For a more detailed analysis of subjective test assessment methodologies and some related open questions, we refer the reader to the work of Tavakoli *et al.* [107] and Garcia *et al.* [123].

QoE modeling, due to its multi-disciplinary and highly subjective nature, is a challenging topic, especially for HAS applications where there are many IFs that need to be considered. Even though QoE modeling for HAS applications has recently gained the attention of the research community, there remain several open challenges and issues such as:

1) *Multi-factor QoE model design*: As discussed in [10], there exist lots of influence factors which need to be taken into consideration for the design of a comprehensive QoE model. Some IFs, especially context related ones, such as the effect of environment, purpose of watching the service, etc., are still not considered by any model. As discussed in [124], to truly understand the user's QoE, a complete understanding of both streaming technique and implementation details of each application is needed. Such detailed information can then be used for the design of a more ''realistic'' QoE model which can also take into account user initiated actions such as play, pause, seeking (forward/backwards), etc. Future models should consider taking into account such IFs in their model design.
2) *Model Complexity*: Most of the works reviewed in this paper, with the exception of two (Xue *et al.* [70] and Rodríguez *et al.* [66]), do not provide any discussion on the model complexity and/or energy consumption associated to the quality evaluation based on the model. The use of high complexity QoE models at the client device can lead to reduced performance of the application due to increased consumption of power and computing resources. Similarly, for server-based models, IFs measurement information (rebuffering duration, number of quality switches etc.) needs to be sent from the client to the server to be considered by the model. Hence, we argue that studies on the complexity of the existing models are needed to help understand their real-world applicability; similarly, relevant discussions should be provided by the proponents of new models.
3) *Subjective test methodology*: As discussed in Section VII, there still exists a need for proper subjective assessment methodology for HAS applications, hence research on this aspect is encouraged, for more scientific and reproducible research.
4) *Privacy Issues*: Another challenge is the decision of where (client/server/network) to deploy the monitoring tool to acquire the measurements of the IFs considered by the model. Client-side monitoring and management are an invasion of privacy and also suffer from shortcomings such as possible cheating by end-user to receive better service, etc. [21]. Network-side monitoring, while overcoming these issues, is not that effective regarding insight into the influence of factors on QoE [22].
5) *Stakeholder*: Depending on the amount and type of information required as input to the model, its measurement can be intrusive or non-intrusive. Also, some models are designed to work with encrypted data, while others require access to bitstream or media signals. Depending on the stakeholder, the requirements will vary. For example, a network provider, to monitor third-party OTT traffic, may prefer a QoE model that works with encrypted video, as sooner or later, all video streaming traffic will be encrypted [124]. Such factors need to be taken into consideration during model design.
6) *QoE based management*: Individual and joint effect of the various IFs need to be evaluated for the design of appropriate QoE control and management strategies. Such insight can then be used for other applications. For example, the knowledge that frequent quality switching can lead to a decrease of QoE can lead to the design of better rate adaptation algorithms by the application provider while a network operator can compensate for quality fluctuations by throttling the network throughput, to limit the bandwidth fluctuation.

The research community has some exciting challenges ahead of them. Faster and better results can be achieved by collaborative efforts and by moving further towards reproducible research.

REFERENCES

[1] Cisco. (Jun. 2017). *Cisco Visual Networking Index: Forecast and Methodology, 2016–2021*. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, ''Overview of the high efficiency video coding (HEVC) standard,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[3] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, ''Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC),'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[4] F. Bossen, *Common Test Conditions and Software Reference Configurations*, document JCTVC-L1100, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), Geneva, Switzerland, Jan. 2013.

[5] L. Guo, J. De Cock, and A. Aaron, ''Compression performance comparison of x264, x265, LIBVPX and AOMENC for on-demand adaptive streaming applications,'' in *Proc. Picture Coding Symp. (PCS)*, San Francisco, CA, USA, Jun. 2018, pp. 26–30.

[6] A. Zabrovskiy, C. Feldmann, and C. Timmerer, ''A practical evaluation of video codecs for large-scale HTTP adaptive streaming services,'' in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 998–1002.

[7] *A Proposed Media Delivery Index (MDI)*, Standard RFC 4445, Apr. 2006.

[8] *Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality*, document ITU-T P.1201 Recommendation, Oct. 2012.

[9] *Parametric Bitstream-Based Quality Assessment of Progressive Download and Adaptive Audiovisual Streaming Services Over Reliable Transport*, document P.1203 ITU-T Recommendation, Nov. 2016.

[10] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, ''A survey on quality of experience of HTTP adaptive streaming,'' *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 469–492, 1st Quart., 2015.

[11] P. Juluri, V. Tamarapalli, and D. Medhi, ''Measurement of quality of experience of video-on-demand services: A survey,'' *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 401–418, 1st Quart., 2016.

[12] P. Le Callet, S. Möller, and A. Perkis, ''Qualinet white paper on definitions of quality of experience (2012),'' in *Proc. Eur. Netw. Qual. Exper. Multimedia Syst. Services (COST Action IC)*, 2012.

[13] *Vocabulary for Performance and Quality of Service. Amendment 5: New Definitions for Inclusion in Recommendation*, document ITU-T P.10/G.100 Recommendation, Jul. 2016.

[14] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document ITU-T P.910 Recommendation, Apr. 2008.

[15] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-T BT.500 Recommendation, Jan. 2012.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, ''Image quality assessment: From error visibility to structural similarity,'' *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[17] *Reference Algorithm for Computing Peak Signal to Noise Ratio of a Processed Video Sequence With Compensation for Constant Spatial Shifts, Constant Temporal Shift, and Constant Luminance Gain and Offset*, document ITU-T J.340 Recommendation, Jun. 2010.

[18] M. Seufert, M. Slanina, S. Egger, and M. Kottkamp, ''‘To pool or not to pool’: A comparison of temporal pooling methods for HTTP adaptive video streaming,'' in *Proc. 5th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Klagenfurt, Austria, Jul. 2013, pp. 52–57

[19] M. H. Pinson and S. Wolf, ''A new standardized method for objectively measuring video quality,'' *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[20] L. Skorin-Kapov and M. Varela, ''A multi-dimensional view of QoE: the ARCU model,'' in *Proc. 35th Int. Conv. MIPRO*, Opatija, Croatia, May 2012, pp. 662–666.

[21] T. Hobfeld, R. Schatz, M. Varela, and C. Timmerer, ''Challenges of QoE management for cloud applications,'' *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 28–36, Apr. 2012.

[22] S. Baraković and L. Skorin-Kapov, ''Survey and challenges of QoE management issues in wireless networks,'' *J. Comput. Netw. Commun.*, vol. 2013, pp. 165146:1–165146:28, Dec. 2013.

[23] W. Robitza *et al.*, ''Challenges of future multimedia QoE monitoring for Internet service providers,'' *Multimedia Tools Appl.*, vol. 76, pp. 22243–22266, Nov. 2017.

[24] A. Ahmad, L. Atzori, and M. G. Martini, ''Qualia: A multilayer solution for QoE passive monitoring at the user terminal,'' in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[25] (2016). *QUIC: A UDP-Based Secure and Reliable Transport for HTTP/2*. Accessed: Nov. 27, 2018. [Online]. Available: https://tools.ietf.org/html/draft-tsvwg-quic-protocol-02

[26] Y. Chen, K. Wu, and Q. Zhang, ''From QoS to QoE: A tutorial on video quality assessment,'' *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 1126–1165, 2nd Quart., 2015.

[27] VQEG. (2000). *VQEG FRTV Phase I Final Report*. [Online]. Available: https://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-i/frtv-phase-i.aspx

[28] VQEG. (2003). *VQEG FRTV Phase II Final Report*. [Online]. Available: https://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-ii/frtv-phase-ii.aspx

[29] A. Takahashi, D. Hands, and V. Barriac, ''Standardization activities in the ITU for a QoE assessment of IPTV,'' *IEEE Commun. Mag.*, vol. 46, no. 2, pp. 78–84, Feb. 2008.

[30] A. Raake *et al.*, ''IP-based mobile and fixed network audiovisual media services,'' *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 68–79, Nov. 2011.

[31] *Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference*, document J.144 ITU-T Recommendation, Mar. 2001.

[32] *Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference*, document ITU-T J.247 Recommendation, Aug. 2008.

[33] *Objective Perceptual Multimedia Video Quality Measurement of HDTV for Digital Cable Television in the Presence of a Full Reference*, document ITU-T J.341 Recommendation, Mar. 2016.

[34] M. G. Martini, B. Villarini, and F. Fiorucci, ''A reduced-reference perceptual image and video quality metric based on edge preservation,'' *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 66, 2012. doi: 10.1186/1687-6180-2012-66.

[35] C. T. E. R. Hewage and M. G. Martini, ''Reduced-reference quality assessment for 3D video compression and transmission,'' *IEEE Trans. Consum. Electron.*, vol. 57, no. 3, pp. 1185–1193, Aug. 2011.

[36] C. T. E. R. Hewage and M. G. Martini, ''Edge-based reduced-reference quality metric for 3-D video compression and transmission,'' *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 5, pp. 471–482, Sep. 2012.

[37] Z. Wang and E. P. Simoncelli, ''Reduced-reference image quality assessment using a wavelet-domain natural image statistic model,'' *Proc. SPIE Conf. Hum. Vis. Electron. Imag.*, vol. 5666, pp. 149–159, Jan. 2005.

[38] Q. Li and Z. Wang, ''Reduced-reference image quality assessment using divisive normalization-based image representation,'' *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[39] L. Ma, S. Li, F. Zhang, and K. N. Ngan, ''Reduced-reference image quality assessment using reorganized DCT-based image representation,'' *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824–829, Aug. 2011.

[40] M. Carnec, P. Le Callet, and D. Barba, ''Objective quality assessment of color images based on a generic perceptual reduced reference,'' *Signal Process., Image Commun.*, vol. 23, no. 4, pp. 239–256, Apr. 2008.

[41] *Perceptual Visual Quality Measurement Techniques for Multimedia Services Over Digital Cable Television Networks in the Presence of a Reduced Bandwidth Reference*, document ITU-T J.246 Recommendation, Aug. 2008.

[42] A. Mittal, A. K. Moorthy, and A. C. Bovik, ''No-reference image quality assessment in the spatial domain,'' *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[43] A. Mittal, R. Soundararajan, and A. C. Bovik, ''Making a ‘Completely Blind’ Image Quality Analyzer,'' *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[44] M. A. Saad, A. C. Bovik, and C. Charrier, ''Blind image quality assessment: A natural scene statistics approach in the DCT domain,'' *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[45] A. K. Moorthy and A. C. Bovik, ''A two-step framework for constructing blind image quality indices,'' *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

[46] *Conformance Testing for Voice Over IP Transmission Quality Assessment Models*, document ITU-T P.564 Recommendation, Nov. 2007.

[47] *Opinion Model for Video-Telephony Applications*, document G.1070 ITU-T Recommendation, Jul. 2012.

[48] *The E-Model: A Computational Model for Use in Transmission Planning*, document G.107 ITU-T Recommendation, Jun. 2015.

[49] *Opinion Model for Network Planning of video and Audio Streaming Applications*, document G.1071 ITU-T Recommendation, Nov. 2016.

[50] (2017). *Adobe HTTP Dynamic Streaming (HDS)*. Accessed: Nov. 17, 2018. [Online]. Available: https://www.adobe.com/devnet/hds.html

[51] Apple. *HTTP Live Streaming*. Accessed: Nov. 17, 2018. [Online]. Available: https://developer.apple.com/streaming/

[52] (2017). *Microsoft Silverlight Smooth Streaming*. Accessed: Nov. 17, 2018. [Online]. Available: https://www.microsoft.com/silverlight/smoothstreaming/

[53] *Information Technology–Dynamic Adaptive Streaming Over HTTP (DASH)—Part 1: Media Presentation Description and Segment Formats*, Standard ISO/IEC 23009-1:2014, 2017. Accessed: Nov. 17, 2018. [Online]. Available: https://www.iso.org/standard/65274.html

[54] I. Sodagar, ''The MPEG-DASH standard for multimedia streaming over the Internet,'' *IEEE Multimedia*, vol. 18, no. 4, pp. 62–67, Apr. 2011.

[55] J. Kua, G. Armitage, and P. Branch, ''A survey of rate adaptation techniques for dynamic adaptive streaming over HTTP,'' *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1842–1866, 3rd Quart., 2017.

[56] N. Cranley, P. Perry, and L. Murphy, ''User perception of adapting video quality,'' *Int. J. Hum.-Comput. Stud.*, vol. 64, no. 8, pp. 637–647, 2006.

[57] S. Egger, B. Gardlo, M. Seufert, and R. Schatz, ''The impact of adaptation strategies on perceived quality of HTTP adaptive streaming,'' in *Proc. Workshop Design, Qual. Deployment Adapt. Video Streaming*, Sydney, NSW, Australia, 2014, pp. 31–36. [Online]. Available: http://doi.acm.org/10.1145/2676652.2676658

[58] (2016). *Global Internet Phenomena Report: North America and Latin America*. Accessed: Nov. 14, 2018. [Online]. Available: https://www.sandvine.com/resources/global-internet-phenomena/2016/north-america-and-latin-america.html

[59] B. Wang, J. Kurose, P. Shenoy, and D. Towsley, ''Multimedia streaming via TCP: An analytic performance study,'' *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 4, no. 2, pp. 16:1–16:22, May 2008. [Online]. Available: http://doi.acm.org/10.1145/1352012.1352020

[60] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, ''Measuring the quality of experience of HTTP video streaming,'' in *Proc. 12th IFIP/IEEE Int. Symp. Integr. Netw. Manage. (IM) Workshops*, Dublin, Ireland, May 2011, pp. 485–492.

[61] R. K. P. Mok, E. W. W. Chan, X. Luo, and R. K. C. Chang, ''Inferring the QoE of HTTP video streaming from user-viewing activities,'' in *Proc. 1st ACM SIGCOMM Workshop Meas. Up Stack*, Toronto, ON, Canada, 2011, pp. 31–36.

[62] D. Z. Rodríguez, J. Abrahao, D. C. Begazo, R. L. Rosa, and G. Bressan, ''Quality metric to assess video streaming service over TCP considering temporal location of pauses,'' *IEEE Trans. Consumer Electron.*, vol. 58, no. 3, pp. 985–992, Aug. 2012.

[63] C. Alberti *et al.*, ''Automated QoE evaluation of dynamic adaptive streaming over HTTP,'' in *Proc. 5th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Klagenfurt, Austria, Jul. 2013, pp. 58–63.

[64] T. Hoßfeld, M. Seufert, C. Sieber, and T. Zinner, ''Assessing effect sizes of influence factors towards a QoE model for HTTP adaptive streaming,'' in *Proc. 6th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Singapore, Sep. 2014, pp. 111–116.

[65] J. Lievens, A. Munteanu, D. De Vleeschauwer, and W. Van Leekwijck, ''Perceptual video quality assessment in HTTP adaptive streaming,'' in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2015, pp. 72–73.

[66] D. Z. Rodríguez, R. L. Rosa, E. C. Alfaia, J. I. Abrahão, and G. Bressan, ''Video quality metric for streaming service using DASH standard,'' *IEEE Trans. Broadcasting*, vol. 62, no. 3, pp. 628–639, Sep. 2016.

[67] K. Yamagishi and T. Hayashi, ''Parametric quality-estimation model for adaptive-bitrate-streaming services,'' *IEEE Trans. Multimedia*, vol. 19, no. 7, pp. 1545–1557, Jul. 2017.

[68] F. Wang, Z. Fei, J. Wang, Y. Liu, and Z. Wu, ''HAS QoE prediction based on dynamic video features with data mining in LTE network,'' *Sci. China Inf. Sci.*, vol. 60, no. 4, pp. 042404:1–042404:14, Apr. 2017.

[69] K. D. Singh, Y. Hadjadj-Aoul, and G. Rubino, ''Quality of experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC,'' in *Proc. IEEE Consumer Commun. Netw. Conf. (CCNC)*, Las Vegas, NV, USA, Jan. 2012, pp. 127–131

[70] J. Xue, D.-Q. Zhang, H. Yu, and C. W. Chen, ''Assessing quality of experience for adaptive HTTP video streaming,'' in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Chengdu, China, Jul. 2014, pp. 1–6.

[71] Z. Guo, Y. Wang, and X. Zhu, ''Assessing the visual effect of non-periodic temporal variation of quantization stepsize in compressed video,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sep. 2015, pp. 3121–3125.

[72] H. T. T. Tran, T. Vu, N. P. Ngoc, and T. C. Thang, ''A novel quality model for HTTP adaptive streaming,'' in *Proc. IEEE 6th Int. Conf. Commun. Electron. (ICCE)*, Ha Long, Vietnam, Jul. 2016, pp. 423–428.

[73] H. T. T. Tran, N. P. Ngoc, A. T. Pham, and T. C. Thang, ''A multi-factor QoE model for adaptive streaming over mobile networks,'' in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[74] W. Robitza, M.-N. Garcia, and A. Raake, ''A modular HTTP adaptive streaming QoE model—Candidate for ITU-T P.1203 ('P.NATS'),'' in *Proc. 9th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Erfurt, Germany, May/Jun. 2017, pp. 1–6.

[75] J. De Vriendt, D. De Vleeschauwer, and D. Robinson, ''Model for estimating QoE of video delivered using HTTP adaptive streaming,'' in *Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manage. (IM)*, Ghent, Belgium, May 2013, pp. 1288–1293.

[76] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath, and A. C. Bovik, ''Modeling the time–varying subjective quality of HTTP video streams with rate adaptations,'' *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2206–2221, May 2014.

[77] Y. Shen, Y. Liu, Q. Liu, and D. Yang, ''A method of QoE evaluation for adaptive streaming based on bitrate distribution,'' in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 551–556.

[78] Y. Liu, S. Dey, F. Ulupinar, M. Luby, and Y. Mao, ''Deriving and validating user experience model for DASH video streaming,'' *IEEE Trans. Broadcast.*, vol. 61, no. 4, pp. 651–665, Dec. 2015.

[79] M. N. Garcia, W. Robitza, and A. Raake, ''On the accuracy of short-term quality models for long-term quality prediction,'' in *Proc. 7th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Pylos, Greece, May 2015, pp. 1–6.

[80] Z. Duanmu, K. Zeng, K. Ma, A. Rehman, and Z. Wang, ''A quality-of-experience index for streaming video,'' *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 154–166, Feb. 2017.

[81] C. G. Bampis and A. C. Bovik, ''Learning to predict streaming video QoE: Distortions, rebuffering and memory,'' *CoRR*, vol. abs/1703.00633, Mar. 2017. [Online]. Available: http://arxiv.org/abs/1703.00633

[82] C. G. Bampis, Z. Li, and A. C. Bovik, ''Continuous prediction of streaming video QoE using dynamic networks,'' *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 1083–1087, Jul. 2017.

[83] C. G. Bampis and A. C. Bovik. (2017). ''An augmented autoregressive approach to HTTP video stream quality prediction.'' [Online]. Available: https://arxiv.org/abs/1707.02709

[84] N. Eswara *et al.*, ''A continuous QoE evaluation framework for video streaming over HTTP,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3236–3250, Nov. 2018.

[85] D. Ghadiyaram, J. Pan, and A. C. Bovik, ''Learning a continuous-time streaming video QoE model,'' *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2257–2271, May 2018.

[86] N. Eswara *et al.*, ''Streaming video QoE modeling and prediction: A long short-term memory approach,'' *CoRR*, vol. abs/1807.07126, Jul. 2018. [Online]. Available: https://arxiv.org/abs/1807.07126

[87] Z. Duanmu, K. Ma, and Z. Wang, ''Quality-of-experience for adaptive streaming videos: An expectation confirmation theory motivated approach,'' *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 6135–6146, Dec. 2018.

[88] Y.-F. Ou, Y. Xue, and Y. Wang, ''Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions,'' *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2473–2486, Jun. 2014.

[89] R. Soundararajan and A. C. Bovik, ''Video quality assessment by reduced reference spatio-temporal entropic differencing,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2013.

[90] *Amendment 2: New Appendix III—Use of P.1201 for Non-Adaptive, Progressive Download Type Media Streaming*, document ITU-T P.1201 Recommendation, Dec. 2013.

[91] H. Ebbinghaus, *Memory: A Contribution to Experimental Psychology* (H. A. Ruger & C. E. Bussenius, Trans.) New York, NY, USA: Teachers College Press, 1913. doi: 10.1037/10011-000.

[92] A. Rehman, K. Zeng, and Z. Wang, ''Display device-adapted video quality-of-experience assessment,'' *Proc. SPIE*, vol. 9394, pp. 9394-1–9394-11, Mar. 2015.

[93] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, ''Quantification of YouTube QoE via crowdsourcing,'' in *Proc. IEEE Int. Symp. Multimedia*, Dana Point, CA, USA, Dec. 2011, pp. 494–499.

[94] *Scikit-Learn: Machine Learning in Python*. Accessed: Dec. 17, 2018. [Online]. Available: http://scikit-learn.org/stable/

[95] N. Ponomarenko, F. Silvestri, K. Egiazarian, J. A. M. Carli, and V. Lukin, ''On between-coefficient contrast masking of DCT basis functions,'' in *Proc. Int. Workshop Video Process. Qual. Metrics*, 2007, pp. 1–4.

[96] Z. Wang, E. P. Simoncelli, and A. C. Bovik, ''Multiscale structural similarity for image quality assessment,'' in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2003, pp. 1398–1402.

[97] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, ''Gradient magnitude similarity deviation: A highly efficient perceptual image quality index,'' *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.

[98] Netflix. *VMAF Development Kit (VDK 1.0.0)*. Accessed: Nov. 12, 2018. [Online]. Available: https://github.com/Netflix/vmaf

[99] D. S. Hands and S. E. Avons, ''Recency and duration neglect in subjective assessment of television picture quality,'' *Appl. Cognit. Psychol.*, vol. 15, no. 6, pp. 639–657, 2001.

[100] D. J. Berndt and J. Clifford, ''Using dynamic time warping to find patterns in time series,'' in *Proc. 3rd Int. Conf. Knowl. Discovery Data Mining*, Seattle, WA, USA: AAAI Press, 1994, pp. 359–370. [Online]. Available: http://dl.acm.org/citation.cfm?id=3000850.3000887

[101] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, ''Study of temporal effects on subjective video quality of experience,'' *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5217–5231, Nov. 2017.

[102] M. Fiedler, T. Hoßfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Netw.*, vol. 24, no. 2, pp. 36–41, Mar./Apr. 2010.

[103] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A subjective and objective study of stalling events in mobile streaming videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 183–197, Jan. 2019.

[104] S. Satti, C. Schmidmer, M. Obermann, R. Bitto, L. Agarwal, and M. Keyhl, "P.1203 evaluation of real OTT video services," in *Proc. 9th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Erfurt, Germany, May/Jun. 2017, pp. 1–3.

[105] W. Robitza *et al.*, "HTTP adaptive streaming QoE estimation with ITU-T Rec. P. 1203: Open databases and software," in *Proc. 9th ACM Multimedia Syst. Conf.*, Amsterdam, The Netherlands, 2018, pp. 466–471. [Online]. Available: http://doi.acm.org/10.1145/3204949.3208124

[106] T. Hoßfeld, L. Skorin-Kapov, P. E. Heegaard, M. Varela, and K.-T. Chen, "On additive and multiplicative QoS-QoE models for multiple QoS parameters," in *Proc. 5th ISCA/DEGA Workshop Perceptual Qual. Syst.*, Berlin, Germany, 2016, pp. 44–48.

[107] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, and N. García, "Perceptual quality of HTTP adaptive streaming strategies: Cross-experimental analysis of multi-laboratory and crowdsourced subjective studies," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2141–2153, Aug. 2016.

[108] A. Asan, W. Robitza, I. H. Mkwawa, L. Sun, E. Ifeachor, and A. Raake, "Impact of video resolution changes on QoE for adaptive video streaming," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Hong Kong, Jul. 2017, pp. 499–504.

[109] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial delay vs. Interruptions: Between the devil and the deep blue sea," in *Proc. Int. Workshop Quality Multimedia Exper.*, Jul. 2012, pp. 1–6.

[110] W. Robitza, P. A. Kara, M. G. Martini, and A. Raake, "On the experimental biases in user behavior and QoE assessment in the lab," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[111] *Waterloo QoE Database*. Accessed: Nov. 5, 2018. [Online]. Available: https://ece.uwaterloo.ca/~zduanmu/jstsp16qoe/

[112] *LIVE Netflix Video Quality of Experience Database*. Accessed: Nov. 5, 2018. [Online]. Available: http://live.ece.utexas.edu/research/LIVE_NFLXStudy/nflx_index.html

[113] *Live Mobile Stall Video Database II*. Accessed: Nov. 15, 2018. [Online]. Available: http://live.ece.utexas.edu/research/LIVEStallStudy/liveMobile.html

[114] *LFOVIA Video QoE Database*. Accessed: Nov. 5, 2018. [Online]. Available: https://www.iith.ac.in/~lfovia/downloads.html

[115] *LIVE QoE Database for HTTP Based Video Streaming*. Accessed: Nov. 15, 2018. [Online]. Available: http://live.ece.utexas.edu/research/Quality/TVSQ_VQA_database.html

[116] R. Schatz and S. Egger, "On the impact of terminal performance and screen size on QoE," in *Proc. ETSI Workshop Sel. Items Telecommun. Qual. Matters*, Vienna, Austria, Nov. 2012, pp. 1–26.

[117] B. Belmudez and S. Möller, "Audiovisual quality integration for interactive communications," *EURASIP J. Audio, Speech, Music Process.*, vol. 2013, no. 1, p. 24, 2013.

[118] *LIVE Video Quality Assessment Database*. Accessed: Nov. 15, 2018. [Online]. Available: http://live.ece.utexas.edu/research/Quality/live_video.html

[119] *Waterloo QoE Database (ECT)*. Accessed: Nov. 15, 2018. [Online]. Available: https://ece.uwaterloo.ca/zduanmu/tip2018ectqoe/

[120] C. G. Bampis, Z. Li, I. Katsavounidis, T.-Y. Huang, C. Ekanadham, and A. C. Bovik, "Towards perceptually optimized end-to-end adaptive video streaming," *CoRR*, vol. abs/1808.03898, Aug. 2018. [Online]. Available: https://arxiv.org/abs/1808.03898

[121] *LIVE-NFLX-II Subjective Video QoE Database*. Accessed: Nov. 5, 2018. [Online]. Available: http://live.ece.utexas.edu/research/LIVE_NFLX_II/live_nflx_plus.html

[122] Netflix. (Mar. 2018). *Dynamic Optimizer—A Perceptual Video Encoding Optimization Framework*. Accessed: Nov. 15, 2018. [Online]. Available: https://medium.com/netflix-techblog/dynamic-optimizer-a-perceptual-video-encoding-optimization-framework-e19f1e3a277f

[123] M.-N. Garcia *et al.*, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *Proc. 6th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Singapore, Sep. 2014, pp. 141–146

[124] (2016). *Video Quality of Experience: Requirements and Considerations for Meaningful Insight*. Accessed: Nov. 14, 2018. [Online]. Available: https://www.sandvine.com/resources/whitepapers/video-quality-of-experience.html

NABAJEET BARMAN received the B.Tech. degree in electronics engineering from the National Institute of Technology, Surat, India, with a focus on wireless networks, and the M.Sc. degree in information technology with specialization in communication engineering and media technology from Universität Stuttgart, Germany. He is currently pursuing the Ph.D. degree in quality of experience of gaming video streaming applications with Kingston University. He was with Bell Labs, Stuttgart, Germany, as a part of his internship and master's thesis. He is currently a Research Associate with the Wireless Multimedia and Networking Research Group, Kingston University, where he is working on QoE-aware video coding strategies as a part of MSCA ITN QoE-Net. He is currently a Video Quality Expert Group Board Member as a part of the Computer Graphics Imagery Project and is also involved in ITU-T standardization activities. His research interests include wireless networking, multimedia communications, and machine learning.

MARIA G. MARTINI (SM'07) received the Laurea degree *(summa cum laude)* in electronic engineering from the University of Perugia, Italy, in 1998, and the Ph.D. degree in electronics and computer science from the University of Bologna, Italy, in 2002. She is a Professor with the Faculty of Science, Engineering and Computing, Kingston University, London, U.K., where she also leads the Wireless Multimedia Networking Research Group. She has led the KU Team in a number of national and international research projects, funded by the European Commission (e.g., OPTIMIX, CONCERTO, QoE-NET, and Qualinet), U.K. research councils, U.K. Technology Strategy Board / InnovateUK, and international industries. She has authored about 150 scientific articles, contributions to standardization groups (IEEE, ITU), and several patents on wireless video. Her research interests include QoE-driven wireless multimedia communications, decision theory, video quality assessment, and medical applications. She chaired/organized a number of conferences and workshops. She is a member of international committees and expert groups, including the NetWorld2020 European Technology Platform Expert Advisory Group, the Video Quality Expert Group, and the IEEE Multimedia Communications Technical Committee, where she has served as the Vice-Chair (2014–2016), as the Chair (2012–2014) of the 3D Rendering, Processing, and Communications Interest Group, and as a Key Member of the QoE and Multimedia Streaming IG. She is an Expert Evaluator for the European Commission, EPSRC, and other research funding bodies. She was an Associate Editor of the IEEE Transactions on Multimedia (2014–2018). She has also been a Lead Guest Editor of the IEEE JSAC special issue on QoE-aware wireless multimedia systems and a Guest Editor of the IEEE Journal of Biomedical and Health Informatics, the IEEE Multimedia, and the *International Journal of Telemedicine and Applications*, among others. She is currently an Associate Editor of the *IEEE Signal Processing Magazine*.