Hindawi
International Journal of Digital Multimedia Broadcasting
Volume 2018, Article ID 1391724, 25 pages
https://doi.org/10.1155/2018/1391724

Review Article

A Survey of Standardized Approaches towards the Quality of Experience Evaluation for Video Services: An ITU Perspective

Debajyoti Pal and Tuul Triyason

IP Communications Laboratory, School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

Correspondence should be addressed to Debajyoti Pal; debajyoti.pal@gmail.com

Received 29 January 2018; Revised 8 March 2018; Accepted 21 March 2018; Published 27 May 2018

Academic Editor: Homero Toral Cruz

Over the past few years there has been an exponential increase in the amount of multimedia data being streamed over the Internet. At the same time, we are also witnessing a change in the way quality of any particular service is interpreted, with more emphasis being given to the end-users. Thus, silently there has been a paradigm shift from the traditional Quality of Service approach (QoS) towards a Quality of Experience (QoE) model while evaluating the service quality. A lot of work that tries to evaluate the quality of audio, video, and multimedia services over the Internet has been done. At the same time, research is also going on trying to map the two different domains of quality metrics, i.e., the QoS and QoE domain. Apart from the work done by individual researchers, the International Telecommunications Union (ITU) has been quite active in this area of quality assessment. This is obvious from the large number of ITU standards that are available for different application types. The sheer variety of techniques being employed by ITU as well as other researchers sometimes tends to be too complex and diversified. Although there are survey papers that try to present the current state of the art methodologies for video quality evaluation, none has focused on the ITU perspective. In this work, we try to fill up this void by presenting up-to-date information on the different measurement methods that are currently being employed by ITU for a video streaming scenario. We highlight the outline of each method with sufficient detail and try to analyze the challenges being faced along with the direction of future research.

1. Introduction

There has been a rapid advance in various video services and its applications, like video telephony, High-Definition (HD) and Ultrahigh-Definition (UHD) television, Internet protocol television (IPTV), and mobile multimedia streaming in recent years. Thus, quality assessment of videos that are being streamed and watched online has become an area of active research. As per a report published in [1–3], video streaming over the Internet is becoming increasingly popular and accounts for more than 55% of the overall traffic. A lot of work has been done by several researchers towards the quality assessment of streaming multimedia services [4–8]. At the same time, organizations like the International Telecommunication Union (ITU) also have in place different models and standardization efforts towards the perceived video quality evaluation under a variety of application scenarios. The main objective of this paper is to provide an up-to-date review of this research field from a standard ITU perspective.

Figure 1 shows a typical video streaming scenario over the Internet. Broadly, three distinct regions are identified as the production network (head-end), the distribution network (carrier), and the consumer network (tail end). Relevant contents are created, edited, encoded, and stored in suitable multimedia databases ready to be transported to the end-users (consumer network) over the Internet with the help of streaming servers. This multimedia traffic has to pass through the unreliable Internet (distribution network) where they are fragmented into various IP segments and ultimately delivered to the consumer end where they are displayed on a variety of devices like television, computers, or mobile phones. The inherent unreliable service provided by the Internet necessitates the use of perceptual quality evaluation schemes for such video traffic.

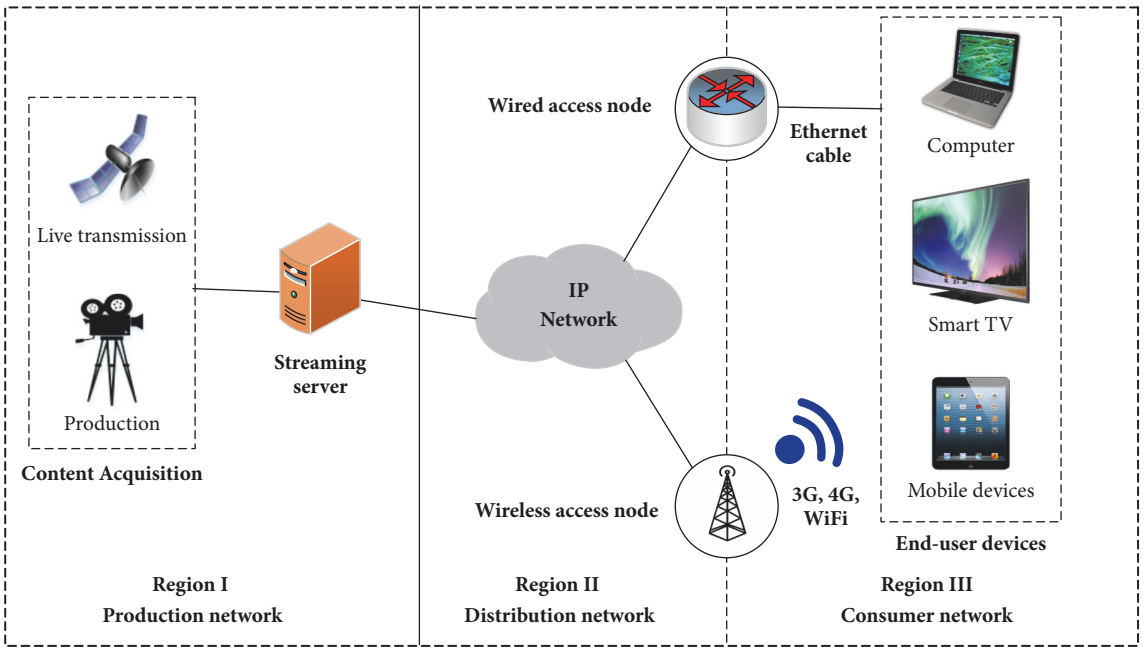

Figure 1: Typical video streaming scenario.

We segregate the multimedia streaming scenario presented in Figure 1 to two different types based upon the ownership use case of the Internet as the Internet protocol television (IPTV) service and over-the-top (OTT) streaming service. YouTube, Netflix, Hulu, etc. are prime examples of the OTT service. IPTV runs on a private, fully controlled network and hence has the advantage of tight control and guaranteed (overprovisioned) bandwidth [9]. IPTV typically uses the User Datagram Protocol (UDP) at the transport layer, and hence in case of any packet loss retransmission does not happen. Still, the reliability of IPTV service is generally high because the video traffic is being carried over a fully controlled network (usually private). On the contrary, in case of OTT services, the contents are streamed over the open and unmanaged public Internet. Thus, IPTV services utilize a network that guarantees a Quality of Service (QoS), which differentiates them from the other OTT services. Quality of Experience (QoE) provisioning for OTT services is a far more challenging job as compared to IPTV services. Hence, for this work we focus only on those ITU standards that do not include IPTV services. More specifically, we focus on video streaming over the public Internet only.

The main goal of this article is to summarize the current and other emerging approaches of video quality evaluation of a streaming service within the scope of ITU. Often due to the sheer variety of the different ITU standards, it becomes difficult for a new researcher to select a suitable method. This work aims to bridge the aforementioned gap by carefully analyzing the relevant ITU standards in detail and giving suitable recommendations as to which standard to choose for a specific context.

We begin by presenting the concepts related to QoS and QoE in Section 2 along with the interrelationship between them. Sections 3 and 4 present the review of subjective and objective methods, respectively. In Section 5, we discuss the current challenges in video quality measurement and the future trends. Finally, Section 6 provides the conclusion.

2. QoS and QoE

We begin the survey process by explaining the key concepts of QoS and QoE explicitly highlighting their differences.

2.1. QoS Concepts

2.1.1. QoS Definition. QoS has been defined by ITU-T as *"totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service"* [15]. This definition of QoS is extremely generic in nature and needs to be reapplied in a specific application context. Figure 2 shows the concept of end-to-end QoS that is commonly prevalent in almost all scenarios. Terminal equipment refers to the devices that are used either by the service provider or by the consumer in order to provide/avail a particular service. Access network is a combination of the access medium and technology used for a particular service (e.g., wireless, cable, ADSL). Access network generally belongs to a specific service provider. Core network refers to the IP backbone network, which is usually controlled by different stakeholders. The QoS contribution of the core network is governed by the technology used (digital multiplexing, IP, etc.) and transmission media (air, cable, optical, etc.) along with other factors. While specifying the end-to-end QoS, it is necessary to state the specified operating conditions in which a service is supported over a connection (connectionless or connection-oriented) scheme. QoS is also affected by factors like traffic and routing [16]. Each of the elements presented in Figure 2 affects the QoS in

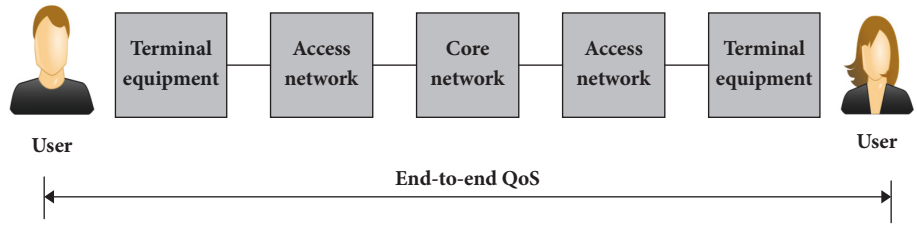

Figure 2: The concept of end-to-end QoS.

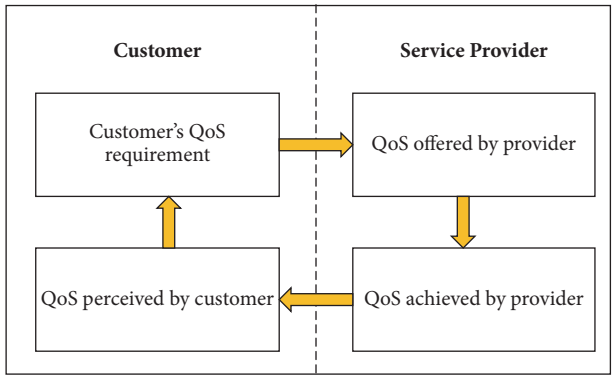

Figure 3: Four Viewpoints of QoS [10].

its own way. In addition, it is evident that QoS comprises both network performance (NP) and non-network-related factors. Bit-error rate, latency, and jitter are some of the NP related factors, while tariff levels, service-repair time, etc. are the non-network parameters. Four different angles from which QoS can be viewed are discussed next.

2.1.2. Viewpoints of QoS and Their Interrelationship. We can classify the different perspectives of QoS into four different types as shown in Figure 3.

(i) *Customer's QoS requirement* refers to the quality level of any application that is expected by the end-users and expressed in nontechnical terms. The customer is not bothered about how a service is offered or about the internals of the network/application design; rather the focus is on the overall end-to-end quality.

(ii) *QoS offered by the provider* refers to the level of service quality that the provider is expected to provide to the customers. The level of quality is expressed by values assigned to QoS parameters. Primarily this is used for planning purpose and framing of Service Level Agreements (SLA) between the provider and the customer.

(iii) *QoS achieved by the provider* refers to the quality level of the service that the provider actually delivers to the customer, which ideally should be the same as the QoS offered by the provider. In reality, the values are different and the performance is compared across the two groups over a certain period.

(iv) *QoS perceived by the customer* refers to the satisfaction level that the customer "*believes*" to have experienced. This is usually assessed from data gathered through customer surveys or individual assessment by a customer for the service.

The four viewpoints are interconnected as shown in Figure 3. Logically, the process starts at the customer's QoS requirement stage. These requirements act as input suggestions to the service provider who plans to offer the desired level of quality. Most of the time, the planned level of service quality is not met due to several factors. As discussed before, these factors are primarily NP related ones like packet loss, jitter, latency, and throughput. A tradeoff between the cost incurred to deliver the ideal quality and the viability of the overall business model has to be done, which affects the service quality in general. The service is ultimately delivered to the customers who perceive the real quality that is achieved by the provider.

From the above discussion, it is clear that the customer viewpoint is the most important one for any service to be successful. This is exactly the reason why ITU has a separate recommendation in [11] that defines a model for multimedia QoS categories from an end-user viewpoint. Next, a brief overview of this recommendation is provided.

2.1.3. QoS Requirements of Different Application Types. Different types of applications are identified like voice, video, and web browsing, with each having different performance requirements for achieving a good perceived quality. Figure 4 shows a classification based upon the overall requirements of the applications in terms of two important QoS parameters, namely, packet loss and one-way delay.

The applications have been classified into eight distinct groups. Some applications such as conversational voice and video are sensitive to delay, but can tolerate a certain extent of packet loss. On the other hand, applications like Fax are sensitive to packet loss, but can withstand delay to a certain extent. Other interactive applications like online gaming are extremely sensitive to both packet loss and delay. These facts are presented in a more clear fashion in Figure 5. The figure shows four distinct delay types depending upon the extent of user interaction involved.

The recommended range of QoS values for some important applications have been provided in Table 1 [11]. The target values of certain applications like audio streaming, videophone, and video streaming are outdated as of 2018. For example, in case of video streaming the typical data rates

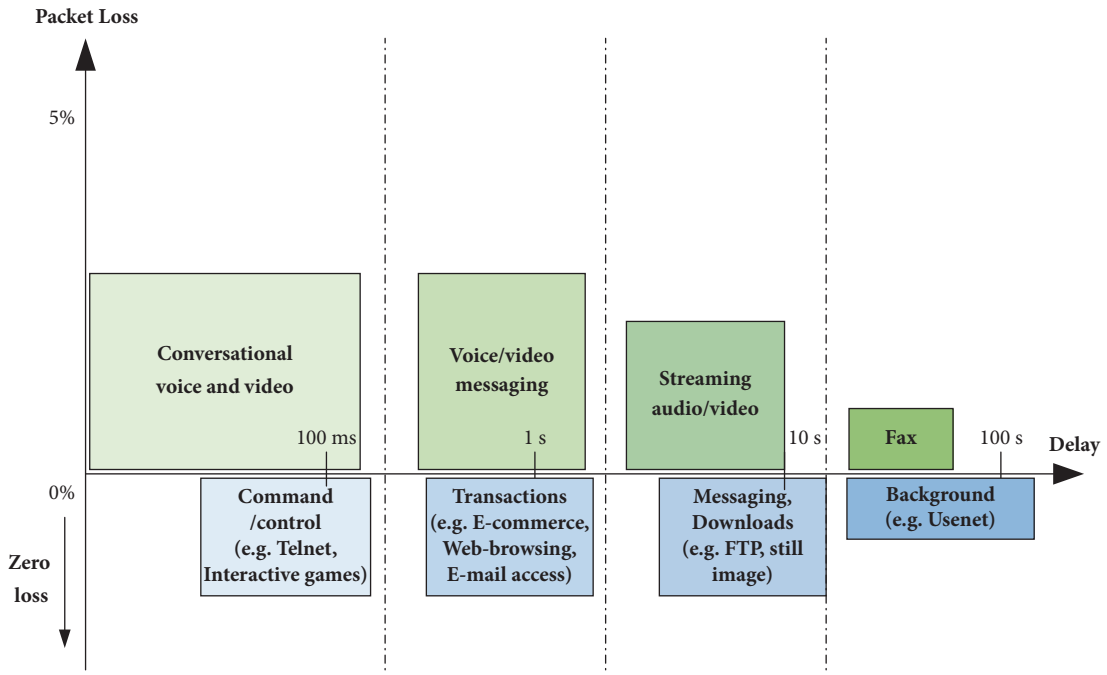

Figure 4: QoS requirements for different applications.

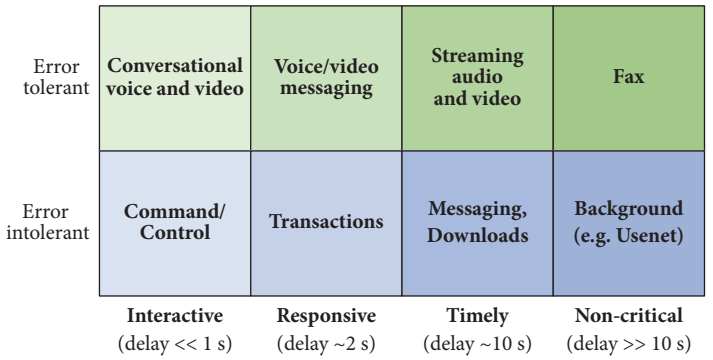

Figure 5: QoS requirements for different applications [11].

Table 1: Performance target for different applications.

Application	Typical Data Rates	Performance Parameters and Target Values		
		One-way Delay	Jitter	Packet Loss
Conversational Voice	4–64 kbps	<150 ms (preferred) <400 ms (limit)	<1 ms	<3%
Voice Messaging	4–32 kbps	<1 s (playback) <2 s (record)	<1 ms	<3%
Audio Streaming	16–128 kbps	<10 s	<1 ms	<1%
Videophone	16–384 kbps	<150 ms (preferred) <400 ms (limit)	-	<1%
Video Streaming	16–384 kbps	<10 s	-	<1%
Web Browsing and HTML	NA	<2 s	NA	Zero
E-commerce Services	NA	<2 s	NA	Zero
Interactive Games	NA	<200 ms	NA	Zero

NA: not applicable.

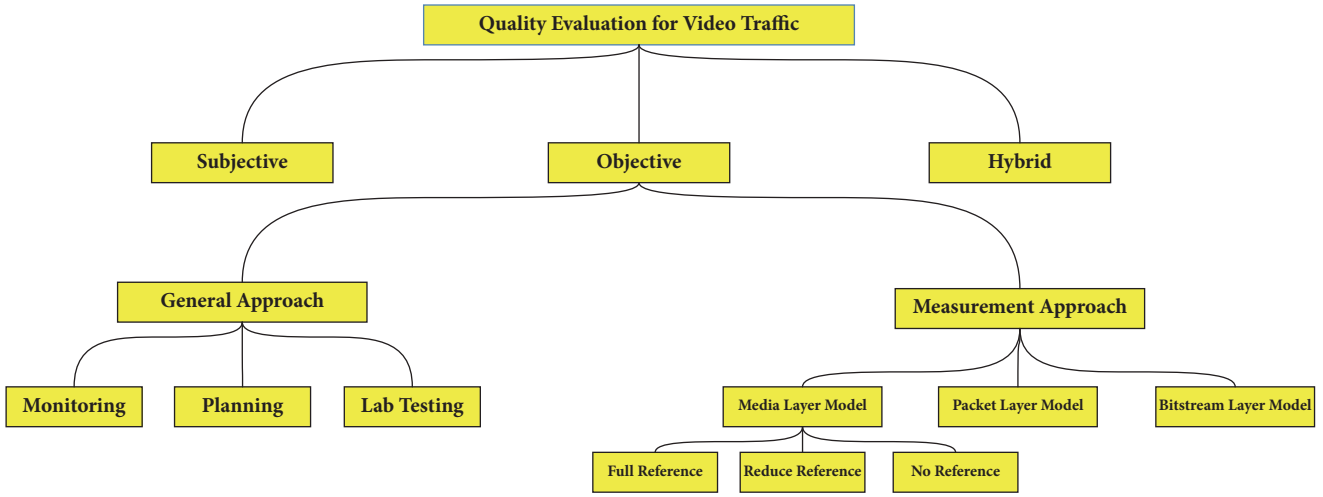

Figure 6: Categorization of different quality assessment methodologies by ITU for video applications.

can easily shoot up to the order of tens of Mbps instead of 384 kbps due to an increase in the network throughput as well as the video resolutions [17]. Similarly, with the advent of modern techniques like dynamic adaptive streaming over HTTP (DASH based streaming), the upper and lower bounds of the other QoS parameters like jitter, one-way delay, and packet loss also need to be updated.

2.2. QoE Concepts

2.2.1. QoE Definition. QoE is defined as the degree of delight or annoyance of the user of an application or service [18, 19]. The concept of QoE is closely related to the human auditory and visual system (HAS and HVS, respectively) and the overall satisfaction that the end-user has in using such a service. Thus, QoE also refers to a complete end-to-end experience that has been shown previously in Figure 2. It is obvious that for any service to succeed, it must provide a good experience to the end-users. A lot of work is being done by ITU towards the quality assessment of various application types. For this article, however, we concentrate only on the video streaming applications. Next, a general overview of the different QoE assessment methodologies being employed by ITU has been provided.

2.2.2. QoE Assessment Methodologies. Confining the scope of this work to video streaming only, Figure 6 shows an overview of the different QoE assessment methodologies being currently used by ITU. Irrespective of the methodology used, the QoE assessment technique must be valid and reliable. The concept of validity versus reliability has been shown in Figure 7. Validity describes how well a method measures what it is intended to measure, while reliability refers to the accuracy of a method in terms of scattering of results (for example, when a test assessment is repeated) [20].

The end-user experience can be measured using two broad techniques: subjective and objective tests [12]. Subjective tests that involve human subjects are considered the most accurate means of quality estimation. Objective tests on the other hand use some mathematical formulae or algorithms to predict the quality. Despite the accuracy of the objective methods being lesser than the subjective ones, they are preferred in many situations as they are automatic, i.e., easy and faster to be carried out and much cheaper than the subjective tests.

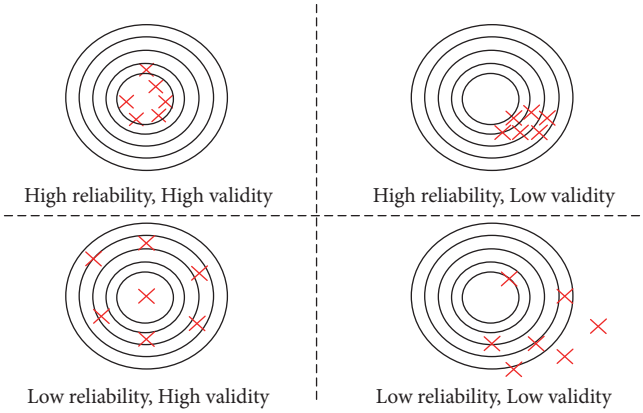

Figure 7: Concept of validity versus reliability [12].

One way to categorize the objective methods is by a general approach, which lists down the different application scenarios in which a particular objective model can be used. There are three specific use-cases as mentioned below:

(i) *Monitoring*: in which a particular objective model is used for live quality assessment of a video application. This is a real time usage scenario that assesses the video quality, e.g., ITU-T P.1202.

(ii) *Planning*: in which an objective model can be used for network planning before an actual service startup. Mainly these models are used as network planning tools in which they help in selecting IP-network transmission settings such as the video format, video codec, and video bitrates with the assumption that the

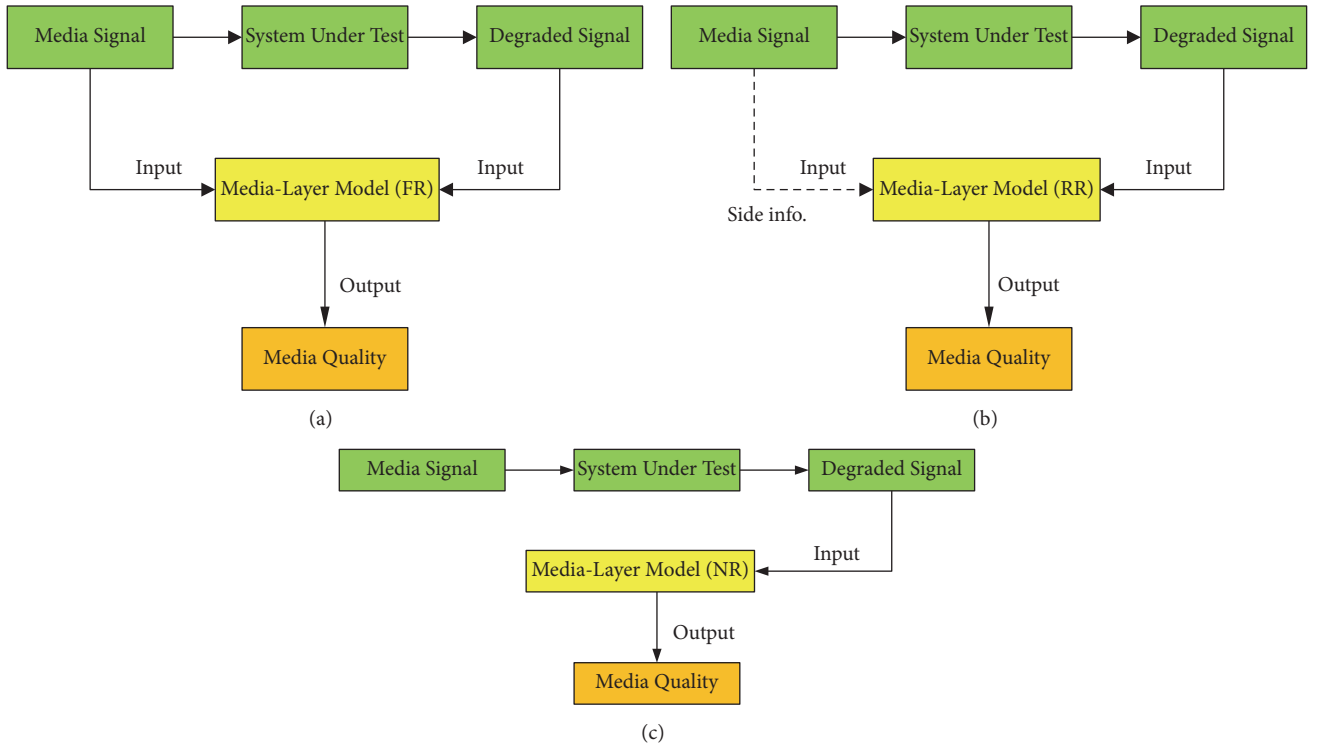

Figure 8: Conceptual view of media layer model: (a) FR methods, (b) RR methods, (c) NR methods.

underlying network is subjected to packet loss, e.g., ITU-T G.1071.

(iii) *Lab testing*: in which quality assessment is done in a typical laboratory setup. This type of approach is used when commercially it is not feasible to assess the quality or in certain situations that require the presence of the original source signal for the purpose of quality measurement or during the development and testing of particular equipment, e.g., ITU-T J.341.

In the second approach, the objective methods are classified based upon the type of measurement used as follows:

(i) *Media layer model*: this uses actual audio/video signals as their input. They also take into account codec compression and the channel characteristics. These types of models can further be subdivided into three different types depending upon the extent of the original reference signal that they have for quality assessment:

(1) *Full reference (FR) methods* in which a reference video is compared frame-by-frame with a distorted video sequence in order to obtain the quality. The comparison can be from many aspects like color processing, spatial and temporal features, contrast features, etc. These methods are generally used in lab-testing environments, e.g., ITU-T J.247.

(2) *Reduced reference (RR) methods* in which certain characteristics/features of the reference signal are extracted out and used for the quality evaluation of the distorted signal. Hence, instead of the entire reference signal, only subsets of its features are used for quality assessment, e.g., ITU-T J.246.

(3) *No reference (NR) methods* are those that do not require the reference video to be present while assessing the quality of the distorted video sequences. These methods are generally used for real time quality assessment of videos, e.g., ITU-T P.1201. Both the RR and NR methods can be applied to either the mid-points or the end-points of the network.

Figures 8(a), 8(b), and 8(c) show the conceptual view of the FR, RR, and NR type media layer models just discussed.

(ii) *Packet layer model*: this utilizes only the packet header information for the purpose of QoE prediction. These models do not have the ability to check the payload information. Therefore, they are not suitable for situations that require the presence of media contents. Generally, such model types are used as network-probes at the mid-points or end-points of the network. Figure 9 shows the conceptual view of a packet layer model, e.g., ITU-T P.1201.

(iii) *Bitstream layer model*: this type takes into account not only the encoded bitstream information, but also the packet header information while assessing the video quality. They are actually a combination of the media

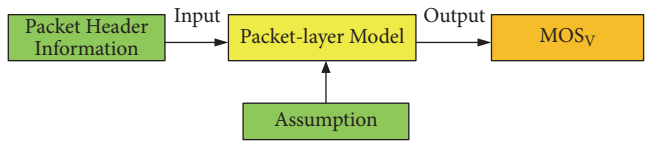

Figure 9: Conceptual view of packet layer model.

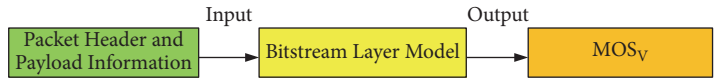

Figure 10: Conceptual view of bitstream layer model.

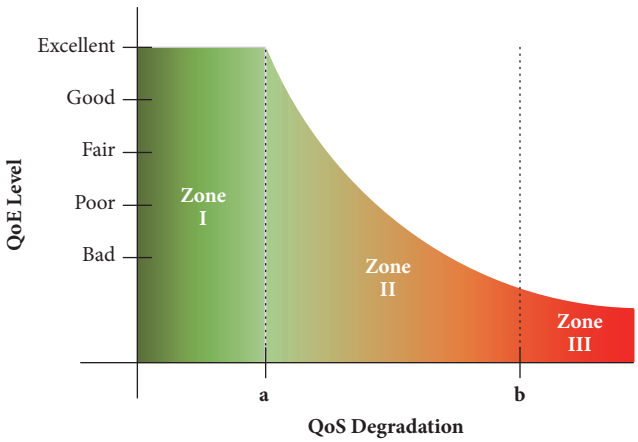

Figure 11: QoS-QoE relationship [13].

layer and packet layer models. Figure 10 shows the conceptual view of a bitstream layer model. These are ideal for live quality monitoring purpose, e.g., ITU-T P.1202.

A third approach to QoE assessment known as the hybrid method uses a combination of the subjective and objective techniques [21, 22]. In this method, typically at the beginning a subjective test is carried out to gather the opinion from the people regarding the quality of the test video sequences under consideration. These test-videos are impaired by one or more QoS factors (NP or non-NP related) depending upon the experimental scenario and requirements. Thereafter, mathematical techniques like linear or nonlinear regression, different types of neural networks, or other machine learning algorithms are used for creating a quality prediction model based upon the subjective scores. This approach tries to take into account the advantages of both the subjective and objective techniques [23], e.g., ITU-T G.1070.

2.3. The Relationship between QoS and QoE. After an elaborate explanation of QoS and QoE from an ITU perspective, now we present the interdependence between them. A possible relationship between the two has been shown in Figure 11. The QoS-QoE relationship has been separated into three distinct zones. Zone I (marked in green) shows the ideal region where the perceived video QoE should be. The users experience an excellent viewing quality. A certain QoS level needs to be maintained (corresponding to point *"a"* on the graph) in order to achieve this QoE. This point *"a"* represents the ideal threshold QoS level (in terms of packet loss, jitter, network throughput, or other factors) that should be maintained theoretically by all the concerned stakeholders. Zone II shows a diminishing QoE region where further deterioration in the QoS values results in a sharp drop in QoE. The point *"b"* on the graph represents the actual threshold value below which the user will probably stop using the service. There is no exact relationship that models this region of diminishing QoE [24, 25]. However, a number of ITU recommendations like ITU-T G.1070, ITU-T G.1071, and ITU-T P.1201 attempt to model this scenario. Zone 3 (marked in red) shows the region where the QoE is extremely poor and should be avoided under all circumstances.

The taxonomy of all the ITU recommendations related to video streaming that have been covered in this survey is shown in Figure 12.

3. Subjective Methodologies

In this section, we present the relevant subjective methods that are used for video streaming applications.

3.1. ITU-T Recommendation P.910. Noninteractive subjective assessment methods for evaluating the one-way overall video quality of multimedia applications such as videoconferencing, storage, and retrieval applications have been covered in ITU Recommendation P.910 [26]. The number of subjects in the tests varies from 4 to 40.

3.1.1. Overall Experiment Design. The test is usually carried out in a recording environment that has sufficient lighting. The lighting conditions should be representative of a typical office scenario rather than studio lighting. Specifically, the ambient lighting of the room should be between 100 lux and 10,000 lux.

The reference video sequences that are used for showing to the human subjects are extremely important. Perceived video quality depends largely on the type of video content [27–30]. Hence, while selecting the reference sequences, spatial information (SI) and temporal information (TI) are two critical factors that must be taken into account. SI gives an indication to the amount of spatial details that each frame

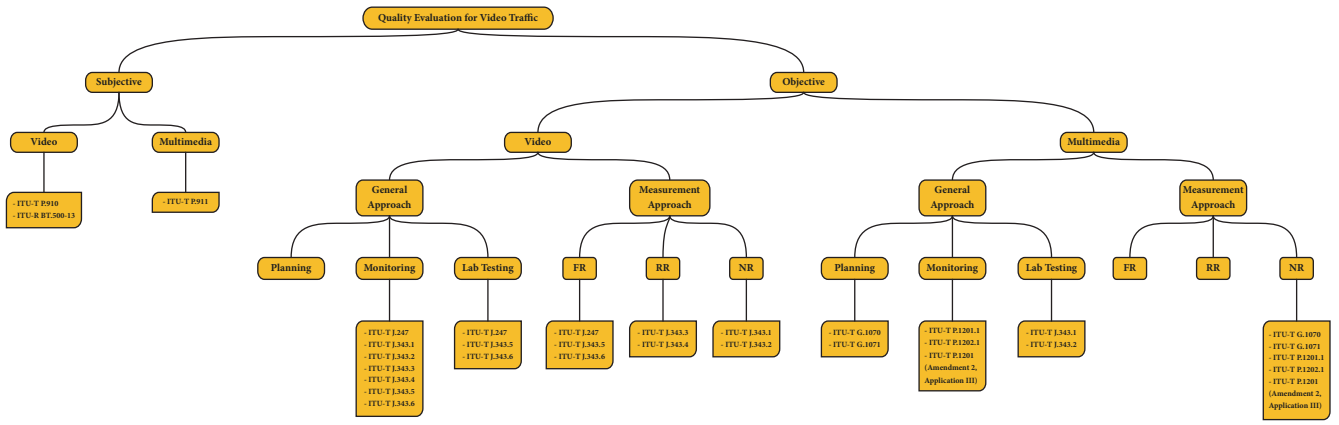

Figure 12: Taxonomy of ITU recommendations related to video streaming.

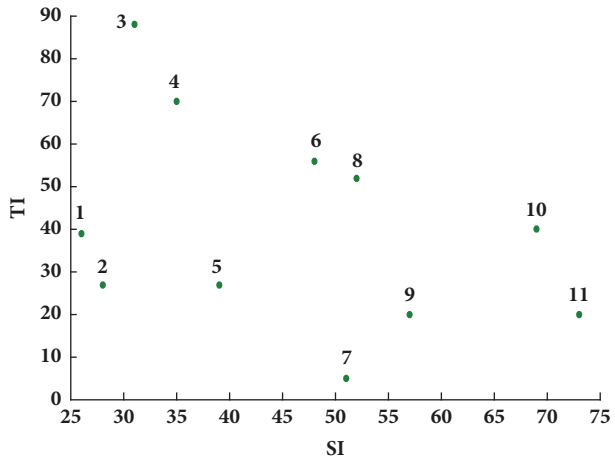

Figure 13: SI/TI values of some commonly used video sequences by ITU [14].

has and it has a higher value for more spatially complex scenes. The SI value for every video frame is calculated by filtering each one of them using the Sobel filter followed by computing the standard deviation. The maximum value in the frame represents the SI content of the scene. Similarly, TI values give an indication of the amount of temporal changes in a particular video sequence and it has a higher value for sequences having greater amount of motion. Equations (1) and (2) show the calculation of the SI and TI values, respectively:

$$\mathrm{SI} = \max_{\mathrm{time}} \left\{ \mathrm{std}_{\mathrm{space}} \left[\mathrm{Sobel} \left(F_n \right) \right] \right\} \tag{1}$$

$$\mathrm{TI} = \max_{\mathrm{time}} \left\{ \mathrm{std}_{\mathrm{space}} \left[F_n \left(i, j \right) - F_{n-1} \left(i, j \right) \right] \right\}, \tag{2}$$

where F_n is the video frame at time n, $\mathrm{std}_{\mathrm{space}}$ the standard deviation across all the pixels for each filtered frame, and $\max_{\mathrm{time}}$ the corresponding maximum value in the considered time interval.

Figure 13 shows the SI and TI values of some commonly used video sequences [14]. The publicly available video database of VQEG is used most frequently while selecting the reference videos [31]. The relevant video details are given in Table 2. Table 3 summarizes the viewing conditions that must be satisfied. Normally, at-least 4 different types of video sequences should be used in a particular test.

Next, we present a brief overview of the different methods that are used by this recommendation.

3.1.2. Different Test Methods. Four different types of methods are used in this recommendation and they are classified as Absolute Category Rating (ACR), Absolute Category Rating with Hidden Reference (ACR-HR), Degradation Category Rating (DCR), and Pair Comparison (PC) method. Each of these techniques is discussed next.

(i) *ACR method*: here the distorted test sequences are presented one at a time and the users give opinion scores (typically on a scale of 1 to 5), which are averaged into a Mean Opinion Score (MOS) [32]. Table 4 shows the MOS scale. The timing diagram of the stimulus presentation has been shown in Figure 14(a). The users are shown video sequences, which typically last for 10 seconds followed by a voting time interval of 10 seconds approximately, wherein the subjects need to enter their opinion in the form of MOS scores. The video presentation time can be increased or decreased depending on the test sequences.

(ii) *ACR-HR method*: it is similar to the ACR method, with an exception that the reference version of each presented distorted test sequence is also shown to the subjects. This is referred to as the hidden reference condition. The subjects give their opinion in the form of MOS scores. However, for final quality assessment a differential quality score (DMOS) is computed for each distorted sequence and its corresponding reference one as per the following equation:

$$\mathrm{DMOS} = \mathrm{MOS}_S - \mathrm{MOS}_R + 5, \tag{3}$$

where MOS_S represents the MOS of a particular distorted video sequence and MOS_R represents the

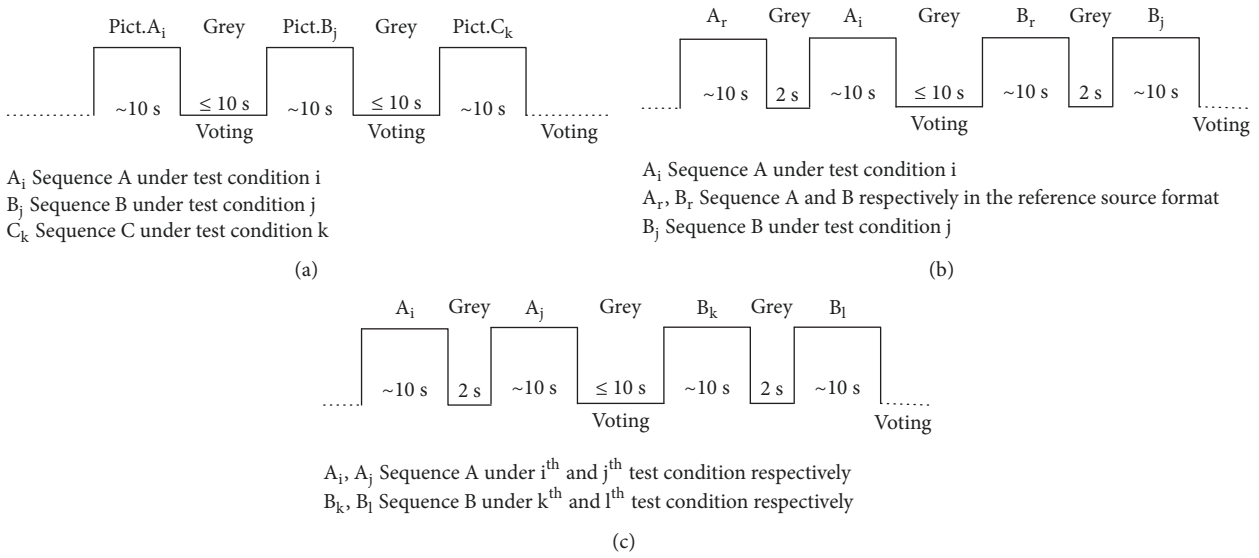

Figure 14: Timing diagram for stimulus presentation: (a) ACR/ACR-HR, (b) DCR, (c) PC.

Table 2: Relevant video details.

Seq No.	Seq Name	Frame Rate	Chroma Format	Content Complexity
1	Harbor	60 fps	4.2.0	1014
2	Ice	60 fps	4.2.0	756
3	DucksTakeOff	50 fps	4.2.0	2728
4	ParkJoy	50 fps	4.2.0	2450
5	Crew	60 fps	4.2.0	1053
6	CrowdRun	50 fps	4.2.0	2688
7	Akiyo	30 fps	4.2.0	255
8	Soccer	60 fps	4.2.0	2704
9	Foreman	30 fps	4.2.0	1140
10	Football	30 fps	4.2.0	2760
11	News	30 fps	4.2.0	1470

Table 3: Summary of viewing conditions.

Parameter	Settings
Viewing distance	1–8 times picture height
Peak luminance of screen	100–200 cd/m
Ratio of luminance of inactive screen to peak luminance	≤0.05
Ratio of luminance of screen, when displaying only black level in a complete dark room to a peak white	≤0.1
Background room illumination	≤20 lux

Table 4: MOS scale.

Rating	Meaning
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

MOS of its corresponding reference sequence. DMOS is also measured on a scale of 1 to 5 identical to MOS. If the distorted video sequence has a better quality than its corresponding reference one, the DMOS value will be greater than 5, which is valid and indicative of an excellent quality (better than the reference one).

Similarly, when the values of MOS_S and MOS_R are the same, the DMOS value is maximum, i.e., 5, indicating no perceptual difference in quality between the distorted and the reference video sequences. The timing diagram is the same as the ACR method.

(iii) *DCR method*: in this type, the test sequences are presented in pairs. In a pair, the reference sequence is always shown first followed by the distorted sequence. The timing diagram for this type of method has been shown in Figure 14(b). The two sequences should be perfectly synchronized; i.e., both of them must start

Table 5: DCR 5 level opinion scale.

Rating	Meaning
5	Imperceptible
4	Perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

and stop at the same frame. In this case, the subjects are asked to rate the distorted sequences with respect to the reference on a 5-point scale. Table 5 presents the 5-level opinion scale.

(iv) *PC method*: in this method, the test sequences are presented in pairs like DCR. However, none of the sequences in the pair is a reference sequence. Instead, all the distorted sequences are combined in all possible combinations and then presented in pairs to the subjects. After each presentation, a judgment is made by the subject on which is the preferred sequence in the pair. The timing diagram has been shown in Figure 14(c).

3.1.3. Comparison of the Test Methods. The most crucial decision is to choose the right technique for a particular application. Normally, the choice is between applications that require or do not require the presence of the reference sequences. The DCR method should be chosen when testing the fidelity of transmission with respect to the reference signal. ACR is easy, fast to implement, and hence commonly used. The basic advantage of ACR-HR over ACR is that the memory effect of the reference sequences can be removed from the subjective scores. PC method should be used when a high discriminatory power is required on the subjective scores.

3.2. ITU-R Recommendation BT.500-13. This recommendation gives different methodologies for assessing the picture and video quality for any generic application scenario, not only restricting to a video streaming case [33]. Considering the popularity of the methods that have been outlined in this recommendation, we chose to include them as a part of this survey. The subjects can be experts or nonexperts depending upon the objectives of the assessment. Minimum 15 observers must be present with no limits on the upper bound. Next, the different test methodologies that are enumerated in this recommendation are presented.

Different Test Methods. Five different types of test procedures are described. They are the Single Stimulus Continuous Quality Evaluation (SSCQE) method, Double Stimulus Continuous Quality Scale (DSCQS) method, Double Stimulus Impairment Scale (DSIS) method, Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) method, and the Stimulus Comparison Adjectival Categorical Judgment (SCACJ) method. The first one is an example of a single stimulus technique, while all the remaining four are examples of double stimulus methods, wherein both the reference and distorted video sequences must be presented simultaneously.

Table 6: Continuous quality scale.

Rating	Meaning
80–100	Excellent
60–80	Good
40–60	Fair
20–40	Poor
0–20	Bad

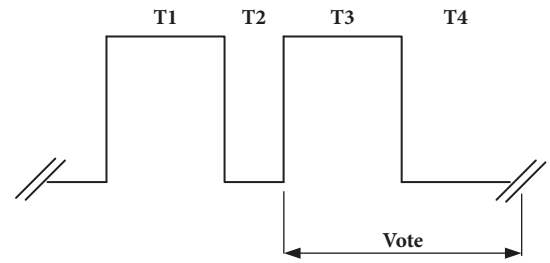

Figure 15: Timing diagram for stimulus presentation (variant 1).

(i) *SSCQE method*: this is a single stimulus method that enables a continuous evaluation of the distorted video sequences on a scale that has been shown in Table 6. The items are normalized in a range of 0 to 100. Generally, each video sequence lasts for at-least 5 minutes.

(ii) *DSIS method*: this is a type of cyclic method in the sense that the subject is at first presented with the original sequence and then with the same impaired sequence. Each sequence is generally reproduced either one (variant 1) or two times (variant 2), after which the subject evaluates the distorted video sequence using an opinion scale that has been shown in Table 5. Interpretations for both the DCR and DSIS methods are the same. The timing diagrams for variants 1 and 2 are shown in Figures 15 and 16, respectively. For both the variants, the subjects need to watch the video sequence during the time slots T_1 and T_3 and voting is permitted only in T_4. Time slots T_1 and T_3 are approximately of 10 seconds duration each, with T_2 being around 3-second pause/gap period and T_4 lasting for 5–11 seconds. T_1 time slot shows the reference sequence, followed by the distorted sequence in T_3.

(iii) *DSCQS method*: this is also a type of cyclic method in which the subject is asked to view a pair of video sequences consecutively, with both of them being from the same source, but one being the original reference sequence and the other one the distorted version of the same source. The subjects assess the quality of both the sequences on a continuous scale that has been shown in Table 6. In this case, the subjects do not know that whether a particular sequence is a reference one or the distorted version. The

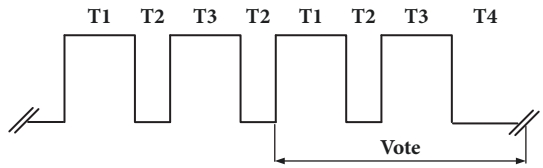

Figure 16: Timing diagram for stimulus presentation (variant 2).

Table 7: SCACJ quality scale.

Rating	Meaning
−3	Much worse
−2	Worse
−1	Slightly worse
0	The same
+1	Slightly better
+2	Better
+3	Much better

general timing diagram of the stimulus presentation for DSCQS method is the same as the second variant of the DSIS method (shown in Figure 16). However, the interpretations of the time slots T_1, T_2, T_3, and T_4 are different. In time slots T_1 and T_3 the test sequences are presented (in no particular order and generally changed across different sequences in a pseudorandom fashion) while in slot T_4 the voting is done. T_2 represents a short gap period between T_1 and T_3. Recommended values for the four time slots are the same as those in the case of the DSIS method.

(iv) *SDSCE method*: in this procedure, the subjects are allowed to watch two video sequences simultaneously, where one sequence is the reference and the other one its distorted counterpart. Generally, both the sequences are shown side by side and the subjects know which is the reference sequence and which is its distorted version. This method is generally used for judging the fidelity of the video information. It is recommended when the video sequences are of longer duration (at-least 5 minutes) and uses the same scale that has been presented in Table 6.

(v) *SCACJ method*: this is an example of a stimulus comparison method and similar to the double stimulus methods discussed above. However, the only difference is that the reference sequences are not shown in this case and only the distorted sequences are presented to the subjects. The subject has to rate the quality of the second video in comparison to the first one based upon the scale which has been shown in Table 7.

3.3. ITU-T Recommendation P.911. This recommendation presents the different subjective quality assessment methods for multimedia applications [34]. The number of subjects varies from 6 to 40. It uses four different techniques, namely, ACR, DCR, PC, and SSCQE. All these techniques have already been discussed in the previous sections. The only difference is in the stimulus type that is shown to the users. In this case video sequences are shown which have an audio counterpart. Therefore, the subjects evaluate the overall multimedia quality. However, in case of the previous recommendations, the videos normally do not have any audio portion. Next, a brief summary of the subjective methods discussed above and their shortcomings is presented.

3.4. Summary of Subjective Methods. Subjective methods are more accurate in gauging the user opinion when compared to the objective ones. A variety of techniques is available and a proper one should be chosen based upon the time available and application requirement. If time is not a constraint, then any of the methods discussed above can be used. For time critical conditions, generally ACR or ACR-HR method is preferred. Similarly, presence or absence of reference content also affects the choice of a particular technique. Sometimes, the duration of the video sequence that needs to be evaluated also plays a judgmental role in deciding which technique is to be chosen. For longer video sequences, normally SSCQE or SDSCE is used. Requirements related to certain specific quality aspects can also sometimes dictate a specific choice.

Reliability of subjects is one of the crucial factors that affect the quality of the results obtained from these subjective techniques. Human perception is often influence by factors like ambient room conditions, emotional and mental state of the subjects, personal profile (age, gender, etc.) that can affect the results obtained [35, 36].

It is obvious from the above discussion that a number of different subjective techniques are available. Hence, for a new researcher it becomes rather confusing which method to select out of the numerous alternatives. In Table 8 we try to provide a guideline to the best subjective technique that should be considered depending upon certain requirements like video duration, presence/absence of reference videos, and need for video repetition.

4. Objective Methodologies

In this section, we provide an overview of the objective models that are used for video streaming and listed in Figure 12. For each model, the overall methodology is discussed along with the mathematical relationships and algorithms wherever necessary.

4.1. ITU-T Recommendation G.1070. This recommendation proposes an algorithm that estimates the videophone quality and is specifically useful for the QoS/QoE planners [37]. This multimedia model takes input from the network and application layers of the TCP/IP protocol stack.

4.1.1. Overall Model Framework. The overall framework of the model has been shown in Figure 17. Certain video and speech quality parameters are given as inputs to the model and there are three main outputs: $V_q(S_q)$, $S_q(V_q)$, and MM_q. $V_q(S_q)$ refers to the video quality influenced by the speech quality, $S_q(V_q)$ refers to the speech quality influenced by the

Table 8: Guidelines to choose a proper subjective approach.

Parameter	Technique								
	ACR	ACR-HR	DCR	PC	SSCQE	DSIS	DSCQS	SDSCE	SCACJ
Stimulus type	Single	Single	Double	Comparison	Single	Double	Double	Double	Comparison
Video duration	10 s	10 s	10 s	-	5 m	10 s	10 s	5 m	-
Explicit video reference	No	No	Yes	No	No	Yes	No	Yes	No
Hidden video reference	No	Yes	No	No	No	No	Yes	No	No
Video repetition	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes
Quality evaluation scale	Table 4	Table 4	Table 5	-	Table 6	Table 5	Table 6	Table 6	Table 7

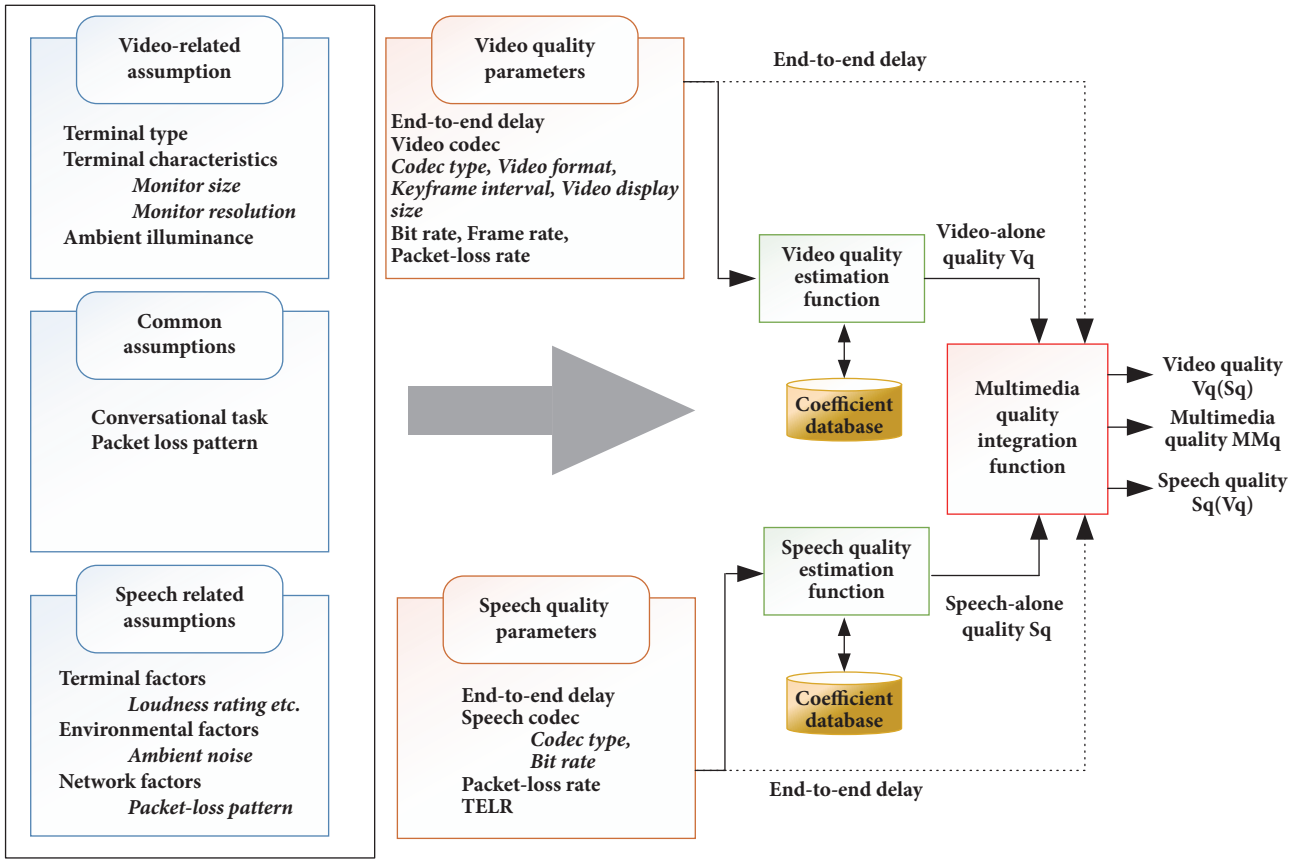

Figure 17: G.1070 model framework.

video quality, and MM_q refers to the overall multimedia quality outputted by the model. In this survey for every recommendation, which produces a multimedia quality as output, we concentrate only on the video quality evaluation part. Therefore, our discussion will focus only on the video quality V_q. Packet loss rate and jitter are the factors considered from the network layer, while bitrate, frame rate, codec type, and video format are the application layer factors.

4.1.2. General Model Equations. The overall video quality predicted by the model is given by

$$V_q = 1 + I_{\text{Coding}} \exp\left(-\frac{P_{pl_V}}{D_{Ppl_V}}\right), \quad (4)$$

where I_{Coding} represents the basic video quality affected by the coding distortion, D_{Ppl_V} expresses the degree of video quality robustness due to packet loss, and P_{pl_V} denotes the packet loss percentage. I_{Coding} is further expressed as

$$I_{\text{Coding}} = I_{O_{fr}} \exp\left\{-\frac{\left(\ln\left(F_{rV}\right) - \ln\left(O_{fr}\right)\right)^2}{2D_{FrV}{}^2}\right\}, \quad (5)$$

where O_{fr} is an optimal frame rate that maximizes the video quality at each video bitrate B_{rV} and is expressed as

$$O_{fr} = \nu_1 + \nu_2 B_{rV}, \quad 1 \le O_{fr} \le 30, \ \nu_1 \text{ and } \nu_2 \text{ constants}, \quad (6)$$

where $F_{rV} = O_{fr}$, $I_{\text{Coding}} = I_{Ofr}$, I_{Ofr} represents maximum video quality at each video bitrate B_{rV} and is expressed as

$$I_{Ofr} = v_3 - \frac{v_3}{1 + \left(B_{rV}/v_4\right)^{v_5}}, \quad 0 \le I_{Ofr} \le 4,\ v_3, v_4 \text{ and } v_5 \text{constants.} \tag{7}$$

D_{FrV} represents the degree of video quality robustness due to frame rate F_{rV} and is expressed as

$$D_{FrV} = v_6 + v_7 B_{rV}, \quad 0 < D_{FrV},\ v_6 \text{ and } v_7 \text{ constants.} \tag{8}$$

The packet loss robustness factor D_{Ppl_V} introduced in (4) is expressed as

$$D_{Ppl_V} = v_{10} + v_{11} \exp\left(-\frac{F_{rV}}{v_8}\right) + v_{12} \exp\left(-\frac{B_{rV}}{v_9}\right), \quad 0 < D_{Ppl_V}. \tag{9}$$

All the coefficients v_1 to v_{12} are dependent on the codec type, the video format, and the video display size and need to be found out by carrying suitable subjective tests.

Equation (4) highlights the fact that ITU-T G.1070 takes into account factors from the network as well as the application layer when evaluating the video quality. Therefore, this method is suitable when any new codec is to be tested for judging their performance. All the equations from (4) to (9) are generic in nature and show how this technique can be ported to a specific context (like evaluating the performance of a new codec along with the network QoS factors) by evaluating the coefficients v_1 to v_{12}. ITU has validated this model only for a limited number of codecs (MPEG-2 and MPEG-4) across VGA, QVGA, and QQVGA resolutions [38, 39]. However, following the procedure that has been outlined through (4)–(9), this model has been extended to other recent codecs like H.265/HEVC and VP9 also [40, 41].

4.2. ITU-T Recommendation G.1071. This recommendation provides an opinion model for network planning of video and audio streaming applications [42]. Two application areas are addressed by this objective technique: a high-resolution area including IPTV and a low-resolution area including services like mobile TV. For reasons that we discussed previously, this survey presents only the mobile streaming application that is an IP based service. This algorithmic model tries to estimate the impact of typical IP layer impairments on the end-user QoE over transport formats such as Real Time Transport Protocol (RTP) over User Datagram Protocol (UDP), Motion Picture Experts Group-2 Transport Stream (MPEG2-TS) over UDP or RTP/UDP, and 3rd Generation Partnership Project Packet-Switched Steaming Service (3GPP-PSS) over RTP. Dynamic adaptive streaming over HTTP or DASH streaming that is currently being used by commercial services like YouTube and Netflix is not taken into account by this model.

4.2.1. Overall Model Framework. The overall model framework has been shown in Figure 18. The general way by which this model works is similar to [43] with an exception in the input that it takes. While as input this model takes into account different network planning parameters like the video bitrate, video codec type, video resolution, and the packet loss rate, the one described in [43] uses the IP packet header information to extract relevant parameters for predicting the video quality. Since the primary video quality estimation block is the same for both models, a conversion rule is applied for those planning parameters that are not taken into account by [43] in order to make it compatible. As output, this model provides three parameters:

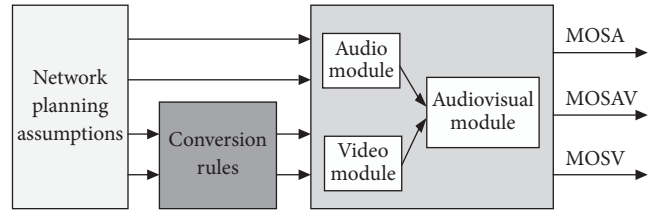

Figure 18: Overall G.1071 model framework.

(i) *audiovisual Quality (MOS_{AV})* on a scale of 1–5,

(ii) *video only MOS (MOS_V)* on a scale of 1–5 (without audio stream),

(iii) *audio only MOS (MOS_A)* on scale of 1–5 (without video stream).

Here we discuss only MOS_V. When compared against similar subjective tests, this model attains a Root Mean Square Error (RMSE) value of 0.60 and a Pearson Correlation Coefficient (PCC) value of 0.78 across 1430 different sample video types.

4.2.2. General Model Algorithm. The overall video quality MOS_V can be classified into three different types, MOS_{VC}, MOS_{VP}, and MOS_{VR}, where

MOS_{VC} is video MOS in case of no packet loss and no rebuffering *(video quality due to compression)*,

MOS_{VP} is video MOS in case of packet loss but no rebuffering *(video quality due to packet loss)*,

MOS_{VR} is video MOS in case of no packet loss but only rebuffering *(video quality due to rebuffering)*.

An elaborate methodology to calculate the three different types of MOS_V has been provided in [42]. However, in order to highlight the factors that this model takes into consideration and motivate the readers to port this for codecs that have not been tested by ITU yet, we present a snapshot of the calculation process by introducing three different algorithms. The same procedure can be applied to any other codec for evaluating the video quality. This ITU model is primarily used for planning purposes only. Since it does not take into account any reference video, it is an example of a NR scheme.

MOS_{VC} is calculated as per Algorithm 1. For this algorithm, V_{CCF} represents the video content complexity factor, i.e., the spatiotemporal complexity of the video sequence and

(1) set $MOS_{MAX} = 5$
(2) set $MOS_{MIN} = 1$
(3) set $V_{DC} = 0$
(4) **if** ($videoFrameRate \geq 24$)
(5) compute $V_{DC} = \dfrac{MOS_{MAX} - MOS_{MIN}}{1 + \left(V_{NBR}/v_3 \times V_{CCF} + v_4\right)^{(v_5 \times V_{CCF} + v_6)}}$
(6) compute $MOS_{VC} = MOS_{MAX} - V_{DC}$
(7) **else**
(8) compute $V_{\mathrm{DC}} = \dfrac{MOS_{MAX} - MOS_{MIN}}{1 + \left(V_{NBR}/v_3 \times V_{CCF} + v_4\right)^{(v_5 \times V_{CCF} + v_6)}}$
(9) compute $MOS_{VC} = (MOS_{MAX} - V_{DC}) \times \left(1 + v_1 \times V_{CCF} - v_2 \times V_{CCF} \times \log\left(\dfrac{1000}{videoFrameRate}\right)\right)$
(10) **end if**

Algorithm 1: Calculation of $\mathrm{MOS_{VC}}$.

(1) set $MOS_{MIN} = 1$
(2) set $V_{DP} = 0$
(3) denote $scene = (slicing\ OR\ freezing)$
(4) **if** ($scene = slicing$)
(5) compute $V_{DP} = (MOS_{VC} - MOS_{MIN}) \times \dfrac{\left(V_{AIRF \times V_{IR}}/\left(v_7 \times V_{CCF} + v_8\right)\right)^{v_9} \times \left(V_{PLEF}/\left(v_{10} \times V_{CCF} + v_{11}\right)\right)^{v_{12}}}{1 + \left(V_{AIRF \times V_{IR}}/\left(v_7 \times V_{CCF} + v_8\right)\right)^{v_9} \times \left(V_{PLEF}/\left(v_{10} \times V_{CCF} + v_{11}\right)\right)^{v_{12}}}$
(6) **else**
(7) compute $V_{DP} = (MOS_{VC} - MOS_{MIN}) \times \dfrac{\left(V_{IR}/\left(v_7 \times V_{CCF} + v_8\right)\right)^{v_9} \times \left(V_{PLEF}/\left(v_{10} \times V_{CCF} + v_{11}\right)\right)^{v_{12}}}{1 + \left(V_{IR}/\left(v_7 \times V_{CCF} + v_8\right)\right)^{v_9} \times \left(V_{PLEF}/\left(v_{10} \times V_{CCF} + v_{11}\right)\right)^{v_{12}}}$
(8) **end if**
(9) compute $MOS_{VP} = MOS_{VC} - V_{DP}$

Algorithm 2: Calculation of $\mathrm{MOS_{VP}}$.

it can vary from an initial default value of 0.5 to a maximum value of 1. It has to be calculated for every sequence used. V_{NBR} represents the normalized video bitrate in kbps and depends upon the video frame rate. The coefficients v_1 to v_6 are provided by ITU for H.264 and MPEG4 encoded video sequences at QCIF, QVGA, and HVGA resolutions only.

The procedure for calculating $\mathrm{MOS_{VP}}$ is given in Algorithm 2. V_{DP} represents the video quality distortion due to packet loss, which can lead to either a slicing or video freezing scenario. Depending upon the scenario V_{DP} is calculated appropriately. V_{AIRF} represents the average impairment rate of the video frames whereas V_{IR} represents the impairment rate of the entire video stream itself. Both of these values lie between 0 and 1, with 0 depicting the best and 1 the worst case. V_{PLEF} represents the video packet loss event frequency, which is incremented by 1 each time a slicing or freezing event occurs. v_7 to v_{12} are the coefficients provided by ITU for the same set of conditions as discussed before.

Algorithm 3 summarizes the procedure for calculation of $\mathrm{MOS_{VR}}$. NRE represents the number of rebuffering events, ARL represents the average rebuffering length, and MREEF represents the multiple rebuffering events effect factor. The coefficients v_{13} to v_{18} are obtained in the same fashion as discussed before for the other coefficients.

4.3. ITU-T Recommendation P.1201/P.1201.1. This recommendation provides a parametric nonintrusive assessment of audiovisual media streaming quality [43]. This is a nonintrusive model based upon the packet header information, which provides certain algorithms for evaluating the audiovisual quality of IP based video services. The packet header information is fed to the algorithm in a Packet Capture Format (PCAP).

This model has 2 subparts: ITU-T P.1201.1 and ITU-T P.1201.2 [44, 45]. While the first one is intended for low-resolution application areas like mobile TV, the second one targets a high-resolution IPTV service. As output, the algorithm estimates the audio, video, and combined audiovisual quality in terms of the 5-point MOS scale.

Primarily, these models are used for in-service monitoring of perceived transmission quality or for maintenance purpose. As such they can be deployed either at the endpoints of the transmission system, i.e., the service provider or customers premises, or in the middle of the network as monitoring points. This model works only for a UDP based streaming service. An alternative version has been proposed in [46] that uses TCP for a nonadaptive and progressive download type media streaming. Table 9 summarizes the

(1) set $MOS_{MIN} = 1$
(2) set $V_{DR} = 0$
(3) set $Video_{Quality} = 0$
(4) denote $scene = (rebuff\ AND\ packet\ loss)\ OR\ rebuff$
(5) **if** $(scene = rebuff\ AND\ packet\ loss)$
(6) set $Video_{Quality} = MOS_{VP}$
(7) **else**
(8) set $Video_{Quality} = MOS_{VC}$
(9) **end if**
(10) compute $$V_{DR} = (Video_{Quality} - MOS_{MIN}) \times \frac{(NRE/v_{13})^{v_{14}} \times (ARL/v_{15})^{v_{16}} \times (MREEF/v_{17})^{v_{18}}}{1 + (NRE/v_{13})^{v_{14}} \times (ARL/v_{15})^{v_{16}} \times (MREEF/v_{17})^{v_{18}}}$$
(11) compute $MOS_{VR} = Video_{Quality} - V_{DR}$

Algorithm 3: Calculation of MOS_{VR}.

Table 9: Application areas, test factors, and technology used by the ITU-T P.1201.1 model.

Type	Description
Application intended	In-service monitoring of audio, video and audio-visual streaming quality
Transport protocol	UDP
Model input (primary)	Packet header information (PCAP files)
Model input (out-of-band)	Codec related factors like bit-rate, frame rate, GOP structure, resolution, etc.
Input video bit-rate range	40–6000 kbps
Input packet loss type	Random and bursty
Input packet loss range	0–10% (random loss) 0–10% (4-state Markov model-bursty loss)
Input frame rate	5–30 fps
Input video resolution	HVGA, QVGA and QCIF
Video codecs	MPEG4 Part 2, H.264 (MPEG4 Part 10)

main input types and scope of this model. For any other specific factor or technology used that has not been mentioned in Table 9, the model needs to be retrained and revalidated. An overview of the model inputs and outputs has been shown in Figure 19.

The packet header information is obtained dynamically from the transport layer in a PCAP file format (interface I.2). Since this model is used for monitoring the video quality in real time, the transport layer input information (in the form of transport header) is dynamic by nature. Relevant information from this PCAP file is filtered out by interface I.3 and fed to the core MOS estimation module. Additional information about the media stream and the decoder behavior is taken out of band in a static manner with certain predefined values. This is the function of the interface I.1. Interface I.4 provides information about the rebuffering information that is extracted and measured at the end-points and provided as an input to the core MOS estimation module.

Three model outputs are provided: MOS_A, MOS_V, and MOS_{AV} referring to the audio only, video only, and combined audiovisual quality all in a MOS scale of 1–5. The overall block diagram of the ITU-T P.1201.1 model has been shown in Figure 20.

The parameter extraction modules for audio, video, and audiovisual scenarios are labeled as PEA, PEV, and PER, respectively. The procedure for calculating the overall video quality MOS_V is the same that has been presented previously in Algorithms 1, 2, and 3. For video, only MOS_V of the model attains a RMSE value of 0.535 (based on 1430 samples) and PCC value of 0.830.

4.4. ITU-T Recommendation P.1202/P.1202.1. This recommendation is similar to the ITU-T P.1201 discussed above. However, in order to evaluate the perceived quality, this algorithm takes into account the bitstream information also, as well as the packet header information that has been used in the previous case [47]. Similar to the previous algorithm, in this case also the model can be subdivided into two parts: ITU-T P.1202.1, which is targeted towards low-resolution areas like mobile video streaming, and ITU-T P.1202.2, which is targeted towards high-resolution IPTV application [48, 49]. Since this model parses information from both the IP header and the payload, it is more accurate when compared to the previous algorithm but requires more computational effort. Also for this model to work, the payload data must be in an unencrypted form. There is another striking difference between this model and ITU-T P.1201 with reference to the number of outputs. P.1202 provides only 1 video MOS as the output, whereas P.1201 provides 3 outputs (audio only, video only, and audiovisual MOS). A summary of the application

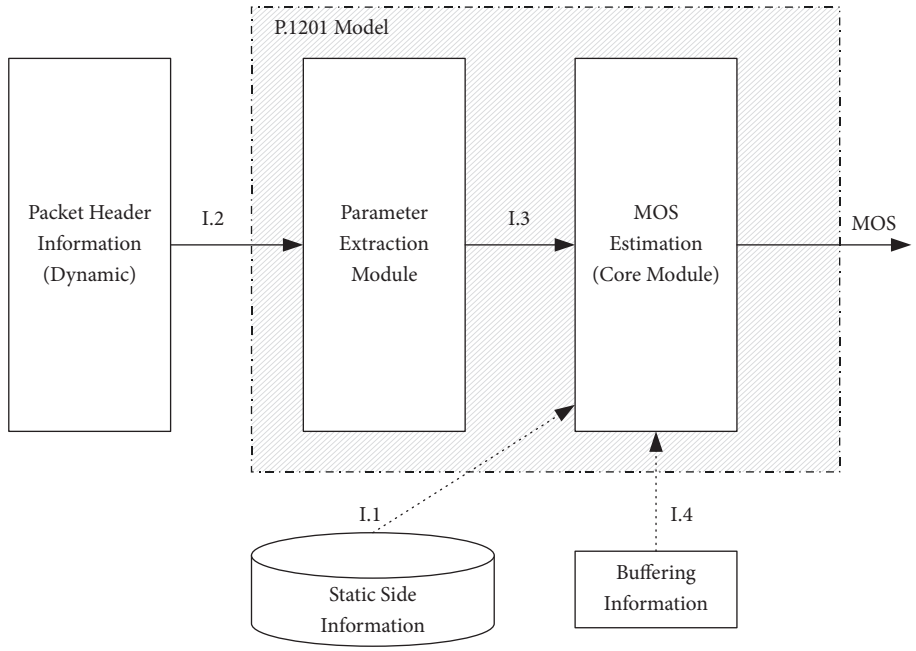

Figure 19: Overview of model inputs and output.

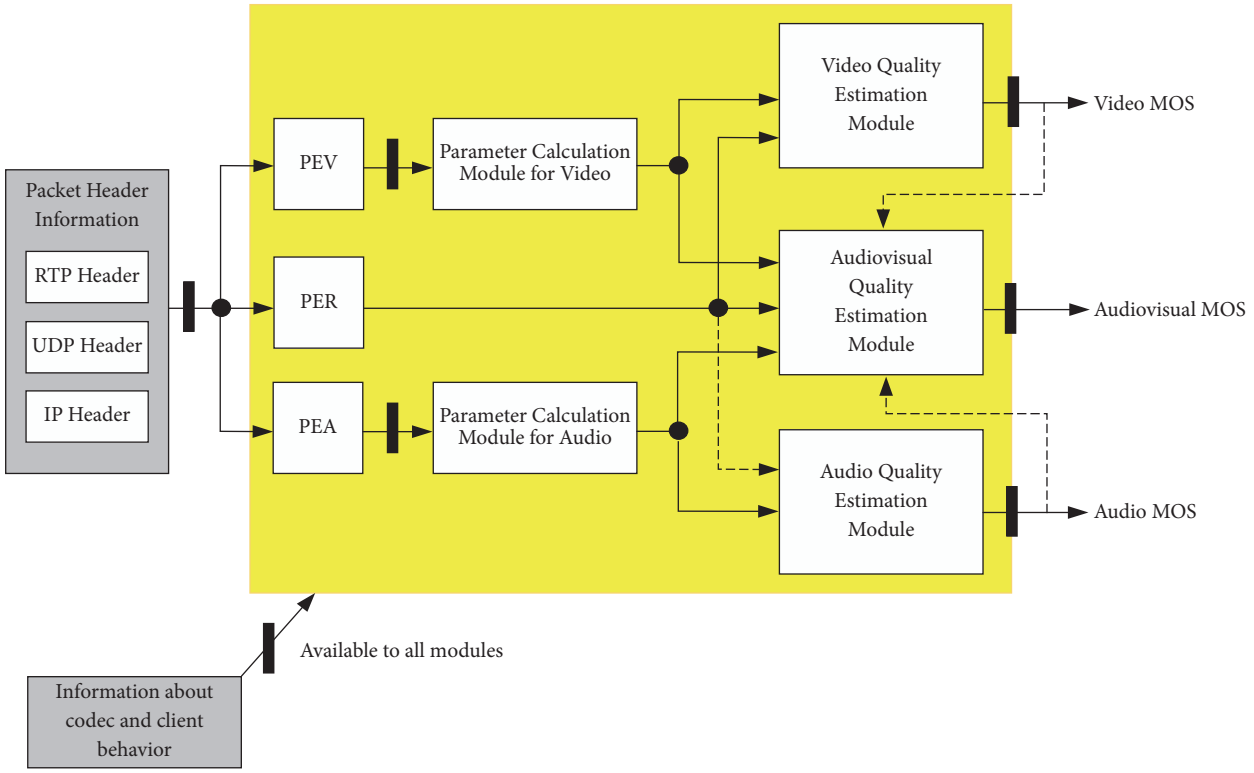

Figure 20: Overall block diagram of ITU-T P.1201.1 model.

areas, test factors, and technology used by this model has been presented in Table 10. An overview of the model interfaces is shown in Figure 21. Interface I.1 provides the static information about the media stream and the decoder. These have certain predefined values and obtained from packet information or player application program interface (API). Interface I.2 provides the detailed packet layer header and payload data information in the form of a PCAP file. Relevant parameters are extracted from the PCAP file by the interface I.3. The model outputs a video only MOS.

Table 10: Application areas, test factors, and technology used by the ITU-T P.1202.1 model.

Type	Description
Application intended	In service monitoring of video streaming quality and quality assessment of live networks including transmission and encoding related errors
Transport protocol	UDP
Model input	Packet header and payload information (unencrypted)
Input video bit-rate range	50–6000 kbps
Input packet loss type	Random and bursty
Input packet loss range	0–6% (random loss) 0–6% (4-state Markov model-bursty loss)
Input frame rate	12.5–30 fps
Input video resolution	HVGA, QVGA and QCIF
Video codecs	H.264/AVC (baseline profile)

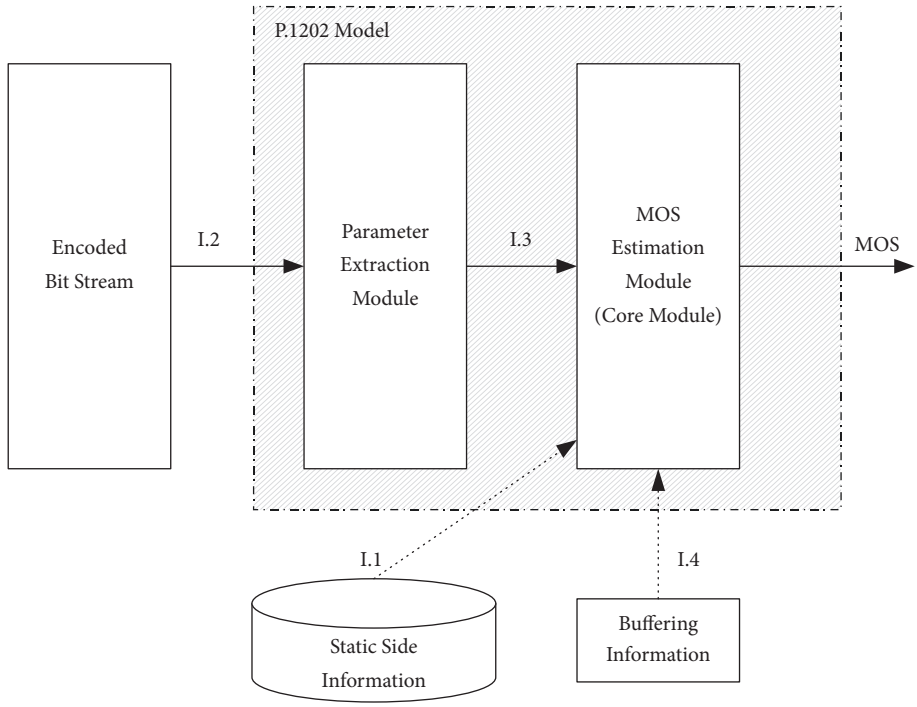

Figure 21: Overview of ITU-P.1202 model interfaces.

The model description in a block diagram format has been shown in Figure 22. H.264 encoded video bitstream, along with other side information (error concealment type, rebuffering, etc.), is taken as input; relevant parameters are extracted out and then aggregated, which are then used to predict the video QoE.

Compression, slicing, freezing, and rebuffering are the four different types of artifacts considered by this model and included in the final video MOS. Each of them is calculated separately and they are finally aligned together to the same level (MOS) by using suitable mapping functions. This model attains a RMSE value of 0.357 (across 982 sequences) and a PCC value of 0.918.

4.5. ITU-T Recommendation J.247. This recommendation provides guidelines on the selection of an appropriate video quality measurement method when a full reference is available [50]. Presently this model has 4 different flavors: Video Quality Expert Group (VQEG) Proponent A (NTT, Japan), VQEG Proponent B (OPTICOM, Germany), VQEG Proponent C (Psytechnics, UK), and VQEG Proponent D (Yonsei University, South Korea). All these 4 models have been tested across video sequences having resolution of VGA, CIF, and QCIF only. All of them take the same inputs and provide the same output in terms of the video MOS (outperforming the commonly used Peak Signal to Noise Ratio (PSNR) model) [51]. Depending upon the operational requirement, these models can predict the quality of videos that have been impaired by codec related factors only, network transmission related factors, or a combination of both. Table 11 lists down the factors for which this model has been evaluated.

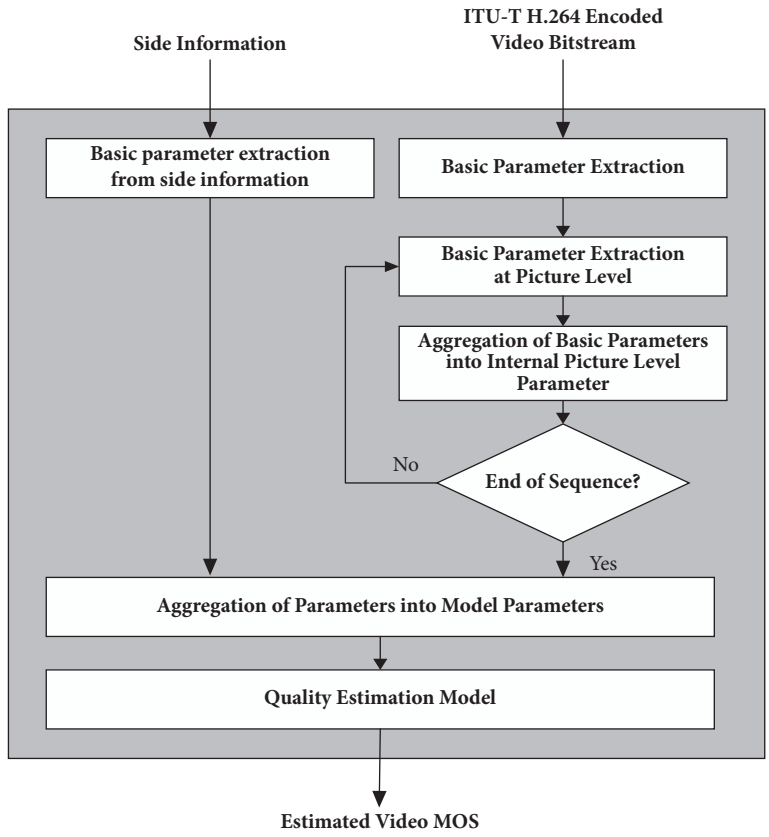

Figure 22: Block diagram of ITU-P.1202.1 model.

Table 11: Application areas, test factors, and technology used by the ITU-T J.247 model.

Type	Description
Application intended	Real time in-service quality monitoring at source and at remote destination (when copy of the source is available) and lab testing
Model input	Reference video and impaired video (transmission errors with packet loss and temporal errors of pausing with skipping)
Input video bit-rate range	16 kbps–4 mbps
Input packet loss type	Random
Input frame rate	5–30 fps
Input video resolution	QCIF, CIF and VGA
Video codecs	H.264/AVC, VC-1, Windows Media 9, Real Video (RV 10), MPEG-4 (Part 2), DivX, Cinepak, H.261, H.263, H.263+, Sorenson and Theora

The performance overview for the 4 different models across the 3 different resolutions has been shown in Table 12. The PCC values are obtained by comparing the objective scores across the three different resolutions against the subjective data from 984 end-users. Figure 23 shows the comparison of the model performances (in terms of PCC values only). The outlier ratio is obtained by using the standard error of the mean as per the formulae given in

$$\text{outlier ratio (OR)} = \frac{(\text{total no of outliers})}{N}, \tag{10}$$

where an outlier is a point for which

$$\left|\left(\text{MOS}_{\text{Subjective}} - \text{MOS}_{\text{Objective}}\right)\right| > C_1 \times \frac{\sigma\left(\text{DMOS}\left(i\right)\right)}{\sqrt{N_{\text{Subjects}}}}. \tag{11}$$

In (11), C_1 is a constant that depends on the nature of the score distribution (Gaussian, exponential, etc.), $\sigma(\text{DMOS}(i))$ represents the standard deviation of the individual scores associated with the ith video clip, and N_{Subjects} is the number of viewers per video clip i.

Table 12: Model performance overview.

Model	Resolution	PCC	RMSE	Outlier Ratio
NTT	VGA	0.786	0.621	0.523
OPTICOM		0.825	0.571	0.502
Psytechnics		0.822	0.566	0.524
Yonsei		0.805	0.593	0.542
PSNR		0.713	0.714	0.615
NTT	CIF	0.777	0.604	0.538
OPTICOM		0.808	0.562	0.513
Psytechnics		0.836	0.526	0.507
Yonsei		0.785	0.594	0.522
PSNR		0.656	0.720	0.632
NTT	QCIF	0.819	0.551	0.497
OPTICOM		0.841	0.516	0.461
Psytechnics		0.830	0.517	0.458
Yonsei		0.756	0.617	0.523
PSNR		0.662	0.721	0.596

Table 13: Model input across different variants.

Model Type	Model Name	Required Inputs
Hybrid NR (encrypted)	RST-V model	Processed video sequence (PVS) and encrypted bitstream
	YHyNRe model	PVS and encrypted bitstream
Hybrid NR	YHyNR model	PVS and non-encrypted bitstream
Hybrid RR (encrypted)	YHyRRe model	PVS, extracted features from source reference channel (SRC) and encrypted bitstream
Hybrid RR	YHyRR model	PVS, features extracted from SRC and non-encrypted bitstream
Hybrid FR (encrypted)	PEVQ-S (e)	PVS, SRC and encrypted bitstream
	YHyFRe model	PVS, SRC and encrypted bitstream
Hybrid FR	PEVQ-S	PVS, SRC and non-encrypted bitstream
	YHyFR model	PVS, SRC and non-encrypted bitstream

4.6. ITU-T Recommendation J.343. This recommendation specifies objective methods that use bitstream data in addition to the processed video sequences [52]. As this is a bitstream model, it has additional information about the payload data like codec type, bitrate, frame rate, spatial, and temporal shifts apart from the transmission errors like delay and packet loss. Six different application areas are addressed by it through [53–58]. This model can work in FR, RR, and NR modes for both encrypted and unencrypted video payload data. Table 13 shows a summary of the inputs that this model can take across its different variants.

Figures 24–26 show the hybrid NR, RR, and FR models (for both encrypted and nonencrypted video data). While the NR models have access to the bitstream and the PVS data, the RR models have access to the bitstream data and the source video sequences having some reduced set of features, and the FR models have full access to the bitstream data along with the entire source video sequences. For all the versions, the encrypted model does not have access to the video payload data and operates without parsing the packet payload.

Table 14 enlists the various parameters for which the models have been tested. The model performance summary has been shown in Table 15. PCC and RMSE values have been used for calculating the model performance statistics. For each of the models, relevant subjective tests are carried out, the results of which are fitted using a third order monotonic polynomial function. In case of the NR models, MOS values are used (obtained from the ACR subjective technique), while for the RR and FR models DMOS values are used (obtained from the ACR-HR subjective technique) for evaluating the model accuracy.

From the above discussion it is clear that a variety of objective techniques that can be used in a number of different scenarios are available. For evaluating the video quality, while some models take in account the presence of reference video signals (FR and RR methods), others do not have this requirement. Similarly, each of the models has been tested for specific codecs only corresponding to specific resolutions. In order to generalize them for different codecs and other factors like resolution and content complexity, we have provided a snapshot of the relevant methodologies. The video sequences that are used for testing the models also vary in terms of the video duration, content complexity, etc. Some of the ITU models are best suited for network monitoring (ITU-T P.1201 series), whereas some are used for network or QoS/QoE planning (ITU-T G.1070, ITU-T G.1071), while the others are used for laboratory testing purpose (ITU-T J.247). Due to this wide variety of objective ITU techniques, it becomes

Table 14: Application areas, test factors, and technology used by the ITU-T J.343 model series.

Type	Description
Application intended	Monitoring the quality of deployed networks and lab testing purpose
Input video bit-rate range	1–30 mbps (HD) and 100 kbps–3 mbps (VGA/WVGA)
Input frame rate	25 and 29.97 fps (FHD)/25 and 30 fps (VGA/WVGA)
Input video resolution	FHD, VGA and WVGA
Video codecs	MPEG-4 Part 10
Length of test sequences	10 s and 15 s (in case of buffering)

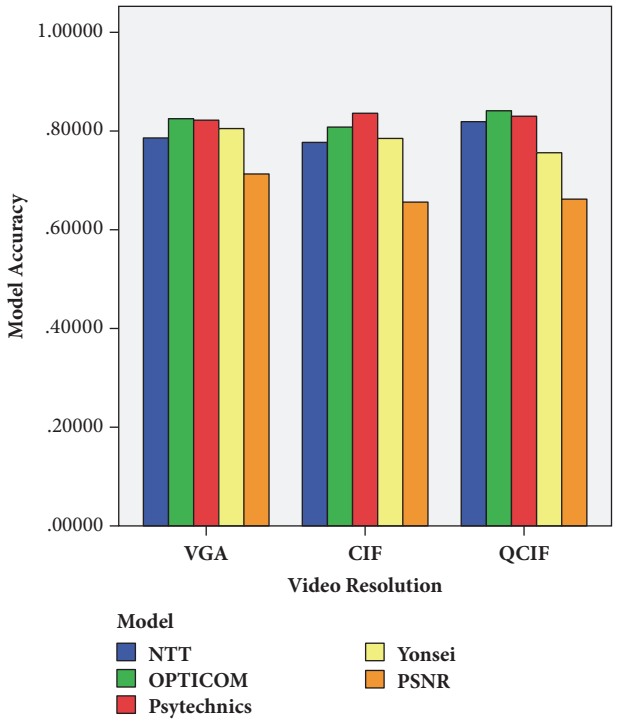

Figure 23: Comparison of model performances.

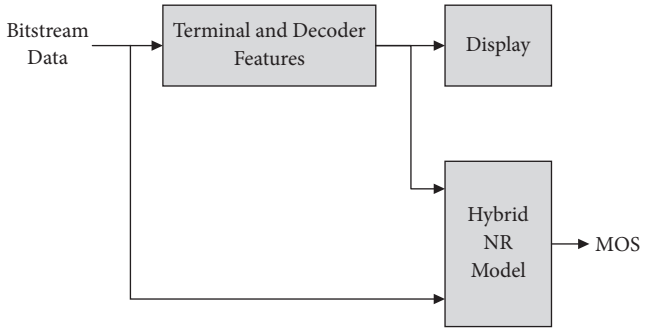

Figure 24: Block diagram of hybrid NR models.

confusing for a researcher to select an appropriate method depending upon the requirements. Therefore, in order to make the model selection process easier, we list down certain factors in Table 16 that can serve as the baseline for selecting the most appropriate model under a specific circumstance. Once a particular model is selected, the necessary changes

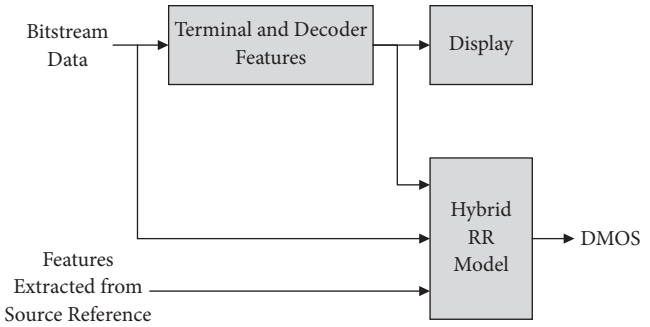

Figure 25: Block diagram of hybrid RR models.

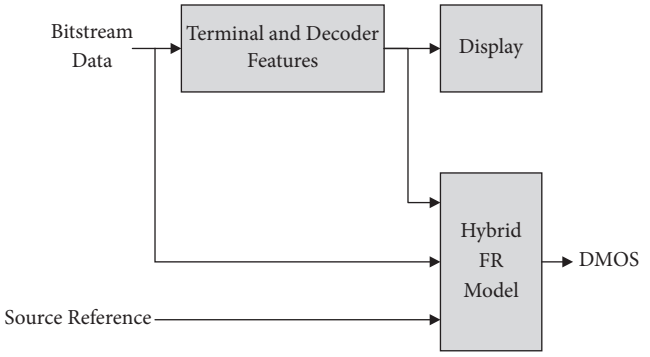

Figure 26: Block diagram of hybrid FR models.

can be made depending upon the research context. Table 16 highlights the network and application parameters that each of the models takes into account along with their intended purpose. Therefore, based upon the parameters of interest for quality prediction and the application scenario, it will be easy to choose a particular reference model.

5. Current Limitations and Challenges

A lot of work is going on within ITU to assess the quality of video streaming services. However, a number of shortcomings exist especially for the quality evaluation of videos that are streamed to mobile devices. We enlist here the challenges that are being faced and should be addressed.

The primary dilemma is in the existence of numerous models, the basic aim of which is to measure the video QoE and the varied type of inputs that they take in predicting the quality. Each model takes a different input based upon either

Table 15: Model performance summary.

Model	Resolution	PCC	RMSE
Hybrid NR YHyNR model	VGA	0.78	0.59
	WVGA	0.81	0.56
	FHD	0.85	0.52
Hybrid NR (encrypted) RST-V model	VGA	0.76	0.61
	WVGA	0.79	0.59
	FHD	0.77	0.64
Hybrid NR (encrypted) YHyNRe model	VGA	0.72	0.66
	WVGA	0.77	0.62
	FHD	0.78	0.61
Hybrid RR YHyRR model	VGA	0.80	0.57
	WVGA	0.84	0.52
	FHD	0.86	0.52
Hybrid RR (encrypted) YHyRRe model	VGA	0.79	0.58
	WVGA	0.84	0.53
	FHD	0.84	0.55
Hybrid FR PEVQ-S	VGA	0.81	0.57
	WVGA	0.83	0.55
	FHD	0.88	0.48
Hybrid FR YHyFR model	VGA	0.80	0.66
	WVGA	0.84	0.61
	FHD	0.86	0.52
Hybrid FR (encrypted) PEVQ-S (e)	VGA	0.81	0.57
	WVGA	0.83	0.55
	FHD	0.88	0.48
Hybrid FR YHyFRe model	VGA	0.72	0.58
	WVGA	0.79	0.52
	FHD	0.84	0.55

network parameters (packet loss, delay, jitter, etc.) or video characteristics (bitrate, frame rate, resolution, content type, etc.) or a combination of both. There can be variations among the network parameters itself. For example, the packet loss pattern may be random or bursty by nature. Similar situations can arise in case of delay also.

The assessment methodologies are also different in terms of the subjective, objective, and hybrid methods. To make the situation even more complex, the different QoS factors (network or application level) as outlined in this survey are not sufficient in predicting the QoE accurately. QoE is strongly influenced by external factors like the type of device used in viewing, the surrounding environmental conditions, and other factors. For majority of the models, the video sequences that are selected from the VQEG database are very short in duration (roughly 10 s only) and hence their ability to portray a real life-streaming scenario is questionable. In addition, the effect of using videos lesser or greater than 10 seconds on the subjective quality assessment has not been accounted for [59, 60].

When streaming is done on mobile devices, the characteristics of the device itself should be taken into account because the viewing experience is quite different on small form factor mobile screens and conventional televisions [61, 62]. There are several limitations to the mobile devices in terms of the variety of screen sizes, display resolution, limited battery backup, limited storage, and other connectivity problems [63]. Currently, none of the existing ITU models considers the peculiarities that are unique to a mobile streaming environment. Despite the fact that more than 55% of the overall Internet traffic is generated by some form of multimedia streaming over a mobile device, lack of a model that particularly addresses this scenario leaves a great void and a lot of scope for further research into this aspect [3].

In a mobile video streaming environment, the inherent unreliable nature of the wireless networks should also be kept in mind. A detailed analysis of the video QoE over a WiFi network and other mobile networks like 2G, 3G, and 4G should be carried out with sufficient detail. Often the low speeds that are associated with mobile networks result in a poor video QoE, which has prompted companies like Google to release a new version of the most popular YouTube application named as YouTube Go that is supposed to work in low speed networks [64]. Thus, ITU should have in place

Table 16: Objective model selection criterion.

Model	Network Factors	Application Factors	Video Sequence Duration	Video Codec	Video Resolution	Packet Loss Type	Model Type	Service Category	Intended Purpose
ITU-T G.1070	PL, J	BR, FR, CT, VR	10 seconds	MPEG-2, MPEG-4	VGA, QVGA and QQVGA	Random	NR	Video Telephony	QoS/QoE Planning
ITU-T G.1071	PL, J	BR, FR, CT, VR	8–24 seconds	MPEG-4, ITU-T H.264	QCIF, QVGA, HVGA	Uniform, Bursty	NR	Video Streaming	Network Planning
ITU-T P.1201 Series	PL, J, T	×	8–24 seconds	MPEG-4, ITU-T H.264	QCIF, QVGA, HVGA	Random, Bursty	NR	Video Streaming	Network Monitoring
ITU-T P.1202 Series	PL, J, T	BR, FR, CT, VR	10–16 seconds	ITU-T H.264	QCIF, QVGA, HVGA	Random, Bursty	NR	Video Streaming	Network Monitoring
ITU-T J.247 Series	PL	BR, FR, CT	×	H.264/AVC, RV 10, WM 9, MPEG-4 (Part 2)	QCIF, CIF, VGA	×	FR	Video Telephony, Video Streaming	Laboratory Testing, Network Monitoring
ITU-T J.343 Series	PL, J, T	BR, FR, CT, VR	×	H.264/AVC	VGA, WVGA, HD	Random, Bursty	NR, RR, FR	Video Streaming	Network Monitoring

PL: packet loss, J: jitter; T: throughput, BR: bitrate; FR: frame rate, CT: video content type; VR: video resolution.

models that simulate these wireless environments in detail and are targeted towards mobile devices considering the recent trend of watching videos online.

Most of the ITU models use videos having low resolutions of VGA, HVGA, CIF, and QCIF only. Practically, only the J.343 series take into account HD resolution. This is in sharp contrast to the current trend where 4K is gaining in popularity. In fact streaming services like YouTube and Netflix have contents that can be streamed in 4K. However, ITU does not provide any model that is dedicated towards such high-resolution videos. Recent advances in virtual reality (VR) and augmented reality (AR) platforms coupled with the availability of mobile devices that can support these have carved out a new way in which videos are being watched by the users. These recent trends and changing viewing habits should be incorporated into future ITU models.

6. Conclusion

Video streaming has become extremely popular these days, which allows the users to watch videos anytime and anywhere. However, for the success of such a service, the quality provided to the end-users must be excellent. There are a number of challenges being faced particularly in a mobile streaming environment. The QoE should be calculated keeping in mind not only the network QoS factors like packet loss, jitter, delay, and throughput and application QoS factors like bitrate, frame rate, and content complexity, but also the nature and characteristics of the mobile devices being used together with the surrounding environment.

In this article, we have presented an in-depth review of the standardized approaches being followed by ITU towards the video quality evaluation. Proper definitions of QoS and QoE have been provided along with the interrelationship between the two. Taxonomy of all the ITU models has been provided based on a general approach and the measurement methodology used. The basic overview and working of all the objective models are provided with suitable diagrams and algorithms/mathematical formulae. Finally, the current drawbacks are discussed along with the scope of future work.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank Dr. Borworn Papasratorn from the School of Information Technology, KMUTT, for sharing his long expertise in telecommunications research and providing the guidelines for writing an effective review paper.

References

[1] "2013Video Index-TV is no longer a single screen in your Living Room," Ooyala Corp, USA, 2013.

[2] G. O. Young, "Synthetic structure of industrial," in *Plastics*, J. Peters, Ed., vol. 3, pp. 15–64, McGraw-Hill, New York, NY, USA, 2nd edition, 1964.

[3] *Cisco Global Mobile Data Traffic Forecast Update Report 2014-2019*, Cisco Corp, USA, 2016.

[4] Z. Cheng, L. Ding, W. Huang, F. Yang, and L. Qian, "A unified QoE prediction framework for HEVC encoded video streaming over wireless networks," in *Proceedings of the 12th IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, BMSB 2017*, Cagliari, Italy, June 2017.

[5] S. Mori and M. Bandai, "QoE-aware quality selection method for adaptive video streaming with scalable video coding," in *Proceedings of the 2018 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–4, Las Vegas, NV, USA, January 2018.

[6] L. Yu, T. Tillo, and J. Xiao, "QoE-Driven dynamic adaptive video streaming strategy with future information," *IEEE Transactions on Broadcasting*, vol. 63, no. 3, pp. 523–534, 2017.

[7] M. García-Pineda, J. Segura-García, and S. Felici-Castell, "A holistic modeling for QoE estimation in live video streaming applications over LTE Advanced technologies with Full and Non Reference approaches," *Computer Communications*, vol. 117, pp. 13–23, 2018.

[8] J. Nightingale, Q. Wang, C. Grecos, and S. Goma, "The impact of network impairment on quality of experience (QoE) in H.265/HEVC video streaming," *IEEE Transactions on Consumer Electronics*, vol. 60, no. 2, pp. 242–250, 2014.

[9] H. J. Kim and S. G. Choi, "A study on a QoS/QoE correlation model for QoE evaluation on IPTV service," in *Proceedings of the 12th International Conference on Advanced Communication Technology: ICT for Green Growth and Sustainable Development, ICACT 2010*, pp. 1377–1382, Phoenix Park, Korea, February 2010.

[10] "Communication Quality of Service: A Framework and Definition," ITU-T Recommendation G.1000, November 2001.

[11] "End-user Multimedia QoS Categories," ITU-T Recommendation G.1010, November 2001.

[12] "Reference Guide to Quality of Experience Assessment Methodologies," ITU-T Recommendation G.1011, July 2016.

[13] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, 2010.

[14] D. Pal and V. Vanijja, "A no-reference modular video quality prediction model for H.265/HEVC and VP9 codecs on a mobile device," *Advances in Multimedia*, vol. 2017, Article ID 8317590, pp. 1–19, 2017.

[15] "Definitions of terms related to Quality of Service," ITU-T Recommendation E.800, September 2008.

[16] C. Xu, P. Zhang, S. Jia, M. Wang, and G.-M. Muntean, "Video streaming in content-centric mobile networks: challenges and solutions," *IEEE Wireless Communications Magazine*, vol. 24, no. 5, pp. 157–165, 2017.

[17] C. Ge, N. Wang, G. Foster, and M. Wilson, "Toward QoE-Assured 4K Video-on-Demand Delivery Through Mobile Edge Virtualization with Adaptive Prefetching," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2222–2237, 2017.

[18] "Amendment 5: New Definitions for inclusion in Recommendation ITU-T P.10/G.100," ITU-T Recommendation P.10/G.100 Amendment 5, July 2016.

[19] *Definitions on Quality of Experience*, Qualinet White Paper from the 5th Qualinet Meeting, March 2013.

[20] W. Robitza, A. Ahmad, P. A. Kara et al., "Challenges of future multimedia QoE monitoring for internet service providers," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 22243–22266, 2017.

[21] S. Winkler and P. Mohandas, "The evolution of video quality measurement: from PSNR to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, 2008.

[22] O. B. Maia, H. C. Yehia, and L. de Errico, "A concise review of the quality of experience assessment for video streaming," *Computer Communications*, vol. 57, pp. 1–12, 2015.

[23] M. Ghareeb and C. Viho, "Hybrid QoE assessment is well-suited for Multiple Description Coding video streaming in overlay networks," in *Proceedings of the 8th Annual Conference on Communication Networks and Services Research, CNSR 2010*, pp. 327–333, Montreal, Canada, May 2010.

[24] H. Rifaï, S. Mohammed, and A. Mellouk, "A brief synthesis of QoS-QoE methodologies," in *Proceedings of the 10th International Symposium on Programming and Systems, ISPS' 2011*, pp. 32–38, Algiers, Algeria, April 2011.

[25] H. J. Kim and S. G. Choi, "QoE assessment model for multimedia streaming services using QoS parameters," *Multimedia Tools and Applications*, pp. 1–13, 2013.

[26] "Subjective Video Quality Assessment Methods for Multimedia Applications," ITU-T Recommendation P.910, April 2008.

[27] K. Gu, J. Zhou, J.-F. Qiao, G. Zhai, W. Lin, and A. C. Bovik, "No-reference quality assessment of screen content pictures," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 4005–4018, 2017.

[28] A. Khan, L. Sun, and E. Ifeachor, "Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks," in *Proceedings of the 2009 IEEE International Conference on Communications, ICC 2009*, Germany, June 2009.

[29] H. Malekmohamadi, W. A. C. Fernando, and A. M. Kondoz, "Content-based subjective quality prediction in stereoscopic videos with machine learning," *IEEE Electronics Letters*, vol. 48, no. 21, pp. 1344–1346, 2012.

[30] T. Ghalut, H. Larijani, and A. Shahrabi, "Content-Based Video Quality Prediction Using Random Neural Networks for Video Streaming over LTE Networks," in *Procceedings of the IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable*, pp. 1626–1631, Liverpool, England, 2015.

[31] "VQEG Standard Database," https://www.its.bldrdoc.gov/vqeg/downloads.aspx.

[32] "Methods for Objective and Subjective Assessment of Speech and Video Quality," ITU-T Recommendation P.800.1, July 2016.

[33] "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU-R Recommendation BT.500-13, January 2012.

[34] "Subjective Audiovisual Quality Assessment Methods for Multimedia Applications," ITU-T Recommendation P.911, December 1998.

[35] R. Stankiewicz and A. Jajszczyk, "A survey of QoE assurance in converged networks," *Computer Networks*, vol. 55, no. 7, pp. 1459–1473, 2011.

[36] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, 2010.

[37] "Opinion Model for Video-telephony Applications," ITU-T Recommendation G.1070, July 2012.

[38] N. D. Narvekar, T. Liu, D. Zou, and J. A. Bloom, "Extending G.1070 for video quality monitoring," in *Proceedings of the 2011 12th IEEE International Conference on Multimedia and Expo, ICME 2011*, pp. 1–4, Barcelona, Spain, July 2011.

[39] B. Belmudez and S. Möller, "Extension of the G.1070 video quality function for the MPEG2 video codec," in *Proceedings of the 2010 2nd International Workshop on Quality of Multimedia Experience, QoMEX 2010*, pp. 7–10, IEEE, Trondheim, Norway, June 2010.

[40] D. Pal, T. Triyason, and V. Vanijja, "Extending the ITU-T G.1070 opinion model to support current generation H.265/HEVC video codec," in *Proceedings of the International Conference on Computational Science and Its Applications*, vol. 9787, pp. 106–116, Beijing, China, 2016.

[41] D. Pal and V. Vanijja, "G.1070 model extension at full HD resolution for VP9/HEVC codec," *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 8, no. 9, pp. 139–147, 2016.

[42] "Opinion Model for Network Planning of Video and Audio Streaming Applications," ITU-T Recommendation G.1071, November 2016.

[43] "Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality," ITU-T Recommendation P.1201, October 2012.

[44] "Parametric non-intrusive Assessment of Audiovisual Media Streaming Quality: Lower Resolution Application Area," ITU-T Recommendation P.1201.1, October 2012.

[45] "Parametric non-intrusive Assessment of Audiovisual Media Streaming Quality: Higher Resolution Application Area," ITU-T Recommendation P.1201.2, October 2012.

[46] "Use of ITU-T P.1201 for Non-adaptive, Progressive Download type Media Streaming," ITU-T Recommendation P.1201 Amendment 2:New Appendix III, December 2013.

[47] "Parametric non-intrusive Bitstream Assessment of Video Media Streaming Quality," ITU-T Recommendation P.1202, October 2012.

[48] "Parametric non-intrusive Bitstream Assessment of Video Media Streaming Quality-Lower Resolution Application Area," ITU-T Recommendation P.1202.1, October 2012.

[49] "Parametric non-intrusive Bitstream Assessment of Video Media Streaming Quality-Higher Resolution Application Area," ITU-T Recommendation P.1202.2, May 2013.

[50] "Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference," ITU-T Recommendation J.247, August 2008.

[51] "Perceptual Visual Quality Measurement Techniques for Multimedia Services over Digital Cable Television Networks in the Presence of a Reduced Bandwidth Reference," ITU-T J.246, August 2008.

[52] "Hybrid Perceptual Bitstream Models for Objective Video Quality Measurements," ITU-T Recommendation J.343, November 2014.

[53] "Hybrid-NRe Objective Perceptual Video Quality Measurement for HDTV and Multimedia IP-based Video Services in the Presence of Encrypted Bitstream Data," ITU-T Recommendation J.343.1, November 2014.

[54] "Hybrid-NR Objective Perceptual Video Quality Measurement for HDTV and Multimedia IP-based Video Services in the Presence of Non-encrypted Bitstream Data," ITU-T Recommendation J.343.2, November 2014.

[55] "Hybrid-RRe Objective Perceptual Video Quality Measurement for HDTV and Multimedia IP-based Video Services in the Presence of a Reduced Reference Signal and Encrypted Bitstream Data," ITU-T Recommendation J.343.3, November 2014.

[56] "Hybrid-RR Objective Perceptual Video Quality Measurement for HDTV and Multimedia IP-based Video Services in the Presence of a Reduced Reference Signal and Non-encrypted Bitstream Data," ITU-T Recommendation J.343.4, November 2014.

[57] "Hybrid-FRe Objective Perceptual Video Quality Measurement for HDTV and Multimedia IP-based Video Services in the Presence of a Full Reference Signal and Encrypted Bitstream Data," ITU-T Recommendation J.343.5, November 2014.

[58] "Hybrid-FR Objective Perceptual Video Quality Measurement for HDTV and Multimedia IP-based Video Services in the Presence of a Full Reference Signal and Non-encrypted Bitstream Data," ITU-T Recommendation J.343.6, November 2014.

[59] F. M. Moss, K. Wang, F. Zhang, R. Baddeley, and D. R. Bull, "On the optimal presentation duration for subjective video quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 11, pp. 1977–1987, 2016.

[60] C. G. Bampis, Z. Li, A. K. Moorthy, and I. Katsavounidis, "Study of temporal effects on subjective video quality of experience," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5217–5231, 2017.

[61] W. Song, D. Tjondronegoro, and M. Docherty, "Exploration and optimization of user experience in viewing videos on a mobile phone," *International Journal of Software Engineering and Knowledge Engineering*, vol. 20, no. 8, pp. 1045–1075, 2010.

[62] H. Knoche, J. D. McCarthy, and M. A. Sasse, "Can small be beautiful? assessing image resolution requirements for mobile TV," in *Proceedings of the 13th ACM International Conference on Multimedia, MM 2005*, pp. 829–838, Singapore, November 2005.

[63] S. Park and S.-H. Jeong, "Mobile IPTV: Approaches, challenges, standards, and QoS support," *IEEE Internet Computing*, vol. 13, no. 3, pp. 23–31, 2009.

[64] "YouTube Go application," Available at https://play.google.com/store/apps/details?id=com.google.android.apps.youtube.mango&hl=en, Last accessed 14th September, 2017.