Quality of Experience Predictive Models for Video on Demand Services: A Review

Paul Wahome Kariuki[a], Ephantus Karari Kinyanjui[b], Samuel Maina Muriuki[b], Dr. Gabriel Kamau[b]

[a]School of Pure and Applied Science, Karatina University, P. O Box 1957-10101, Karatina, Kenya, E-mail: kariukip@karu.ac.ke

[b]School of Computer Science and Information Technology, Dedan Kimathi University of Technology, Private Bag, Nyeri, 7381, Kenya.

Abstract

The number of services available on the internet has grown exponentially in the last decade. Internet dependent services like on demand video services (streaming), online gaming, online data management and processing have gained more prominence than ever before. Users are increasingly shifting away from the classical way of accessing services and embracing modern approaches. At the heart of offering services online is the Quality of Experience (QoE) phenomenon especially for video on demand services which requires a holistic understanding of the Human Factors coupled with Quality of Service elements to predict QoE. This paper takes its point of departure from the presenting the theoretical formulation of QoE phenomenon, which clearly captures dynamic uncertainty in user behavior and sits at the intersections of quantitative and qualitative parameters. We further describe the taxonomy of QoE, present the state of the art and related literature in QoE for VoD predictive models.

Keywords: Quality of Experience, Video on Demand, Video Quality Assessment

1. INTRODUCTION

Video On Demand (VoD) is an interactive media distribution technology that al- lows users to access multimedia content irrespective of time or their geographical location. Owing to the readily available miniaturized high computing devices and access to broadband internet connectivity, inter- net services have become diverse and rich in content [80]. Among the dynamic services offered through the Real-Time Streaming Protocol are multimedia services such as VoD, news, Internet Protocol Television (IPTV), social networks, video Teleconferencing, immersive and virtual reality games, multimedia conferencing, live sports, movies among others [2]. Real-Time Streaming Protocols which mediate signal transmission in a VoD ecosystem were established through a collaborative effort be- tween the World Wide Web Consortium (W3C) and the Internet Engineering Task Force (IETF).

W3C prescribes the JavaScript APIs and the standard HTML5 tags for peer-to-peer (P2P) connections and the IETF which prescribes the protocols for setting up and management of a communication channel [28]. In a VoD environment, the abstraction is created by storing the video content as a com- pressed digital file in a centralized server environment. This allows a viewer to put in a re- quest for an on demand video which is then compressed and transmitted to an internet enabled device either at a cost or freely [46]. Nowadays, the most popular interactive multimedia service is video on demand and as projected by Cisco, video traffic has now sur- passed 82 percent of the total web consumer traffic [75].

However, even with the traction VoD has gained, it is still prone to a myriad of factors which can result in degradation of the agreed Service Level Agreement (SLA). These factors could range from lossy signal encoding or decoding, fluctuation of throughput, net- work overload, video freezes, delay, increase in turnaround time and jitter [25] which in-turn results to decreased user satisfaction which eventually leads to low Quality of Experience (QoE). This also implies that a positive change in the factors mentioned above toward attainment of the required threshold leads to high QoE perception. As a result, the research and industry community, through numerous extant studies have invented various schemes to better understand how consumers perceive and experience service degradation, and to investigate mechanisms of providing them with better experiences [63] [2] as well as predicting the optimum Quality of Experience levels.

The construct of quality of a service cannot be ignored in contemporary world. Modern day users can be perceived as [73] quality meters who are not just service consumers but evaluators who conduct quality assessments on the video and audio content they receive. Countless studies [45] [24] [80] [74] [11] have demonstrated that while traditional quality assessments focused mainly on Quality of Ser- vice (QoS), QoE has been proven to be a more effective way of evaluating user perceptions. Quality of Experience is post exposure holistic phenomenon that depicts user's perception of a product or a service. It seeks to measure the user's satisfaction through objective or subjective measures. Provision of above board QoE for VoD is a herculean task for many service providers. This is due to the fluid nature of the requirements which are defined from an end user perceptive and due to the fact that they are agnostic to the network architecture as well as the communication protocols in use. However, despite the strong arguments exuded in the studies mentioned above in favour of QoE, the existing literature provides no adequate instruments to quantify the service quality of video-on-demand services as its a novel concept. As such, the objective of this paper will be to conduct a review of related literature on QoE predictive models for VoD. The methodology proposed by Kitchenman [44] for conducting rigorous literature review was adopted.

2. THEORETICAL BACKGROUND

This section presents the context of Quality of Experience in prediction and estimation of Video on Demand. It further presents a review of recent related studies with a focus on the taxonomy proposed for each model. The section also highlights a critique and open issues for research.

2.1 Quality of Experience and Quality of Service

Over the last two decades, Quality of Experience research has been very active leading to even broader computer science allied tenets like emotional and psychological experiences, aesthetics among others. Various researchers have provided different QoE definitions, International Telecommunication Union (ITU-T) for example defines QoE as "the overall acceptability of an application or service, as perceived subjectively by the end-user". This implies that since QoE is an end to end assessment of the system, the overall experience and perspective may be prejudiced due to context and user expectations. Therefore, this makes QoE prediction and estimation a complex task that can only be addressed comprehensively though qualitative (subjective)and quantitative (objective) techniques.

Subjective QoE measures can be derived from conducting interviews with users of a VoD services, web surveys, questionnaires and suggestion boxes. On the other hand, objective QoE can be derived from network-based Quality of Service (QoS) parameters which quantify the conditions set out in the SLA, cognitive parameters, physiological and psychological parameters [28].

In an contemporary multimedia environment where VoD is offered as a service, QoE subjective tests can be categorized as either passive assessments or active tests (conversational assessments). In passive assessment, users are exposed to a series of original and distorted audio visual frames for scoring using Mean Opinion Scores (MOS) [42] [39]. While true MoS can only be derived from subjective tests, objective assessments can also be carried out to reproduce human perception. In the case of VoD, objective tests can be categorized as either (i) Parametric Packet layer models that estimate QoE based on packet header information rather than analyzing media signals are known as parametric packet-layer models. ii) Parametric planning models which predict QoE using quality planning criteria for networks and associated terminals. iii) Bit-stream-layer models which estimate QoE using encoded bit-stream and packet-layer information. iv) Hybrid models which takes data from the signal, bit stream, and/or packet headers as inputs. v) Media layer model which utilizes certain reference data and degraded audio as well video signals for robust QoE estimation [28]. Acknowledging that subjective measures provided better accuracy in prediction of QoE, they are characterized by major downside based on the following arguments. firstly, subjective tests ought to be conducted in a controlled environment and under limited conditions.

Figure 1: *Quality of Experience Influence Factors*

Categories		**Typical QoE IFs**
System IFs	Content-related	Spatial-temporal requirements, colour depth, 2D/3D, texture, content reliability, etc.
	Media-related	Encoding, resolution, sampling rate, media synchronization, etc.
	Network-related	Bandwidth, delay, jitter, loss, error rate, throughput, etc.
	Device-related	System specification, equipment specifications, device capabilities, provider specification & capabilities, etc.
Context IFs	Physical	Location, space, movement, etc.
	Temporal	Playing time, duration, frequency, etc.
	Economic	Cost, brand, subscription type, etc.
	Task/Social	Multitasking, social occasion, etc.
	Tech & Info	Relationship between the system of interest and other relevant system.
Human IFs	Low-level processing	Physical, emotional and mental constitution of human etc.
	High-level processing	Demographic and socio-economic background, etc.

Secondly, subjective tests cannot be done in real time and require a large group people which not only makes it expensive but also time consuming. Notably, subjective assessments may also be affected by the user's opinions, preferences, prior and posterior state of the mind, past perceptions and user understanding.

2.2 Quality of Experience Taxonomy and Measurement

The main debate in prediction and estimation of QoE for Video on Demand services is determination of the critical parameters that have a significant impact. Researchers have demonstrated a divergent perspective on the how-to arrival at the taxonomy for QoE. In their white paper, [12] argue that the system, context, and human/user factors as the critical performance indicators when deriving QoE taxonomy. System Influence Factors (IFs)lays its main focus on parametric technical measures of a system which can be expressed using QoS parameters. The user IFs refer to observable human aspects and behaviour while interacting with a system. Context IFs focuses on the ecosystem of the user like their environment, location, time among others while content IFs speak to the quality of the content accessed by the user [80]. These performance indicators created a basis for further analysis and improvements as opined by [36] [86] [57] [35].

Various IFs in the end-to-end chain of video communication, which includes video capture, coding, storage, delivery, decoding, rendering, display, context of use, and user factors, are known to influence QoE. However, monitoring QoE with all the IFs is impractical due to the fact that some of these Influence Factors are difficult to observe and monitor, and again these Influence Factors may be correlated. Given the aforementioned, a comprehensive QoE Model is one that measures both subjective and objective parameters. The assessments must be carefully planned and the subjects prepared for the experiments as per the ITU-T P.800 recommendations [29].

QoE was represented as a linear equation by [62] as shown below;

$$\sqrt{QoE = \mu.QoS + 1\mu.UX} \quad \textbf{(1)}$$

Where

$$QoS = \frac{\sum_i niSi}{\sum_i ni\neq} \quad \text{and} \quad \textbf{(2)}$$

$$\sqrt{UX = \frac{\sum_i aipi}{\sum_i ni \neq} + \frac{\sum_i \beta iRi}{\sum_i \beta i} + C\frac{\sum_k YkUk}{\sum_k \beta k}} \quad \textbf{(3)}$$

where:

The variable$\boldsymbol{\mu}$ controls the proportional weight related to the quality of service parameters as compared to the user experience (UX). At the same time, the variables S_l, P_i,R_j,U_k, represent the quality values that are provided to the individual parameters of quality of service measurements (S_l), perception measurements (P_i), rendering quality measurements (R_i), and user state measurements (U_k) respectively.

The variables *A*, *B*, *C* are computed empirically with weighting constants for the respective perception measurements, rendering quality measurements, and user state measurements. The variables η_l, α_i, β_j, γ_k are weighing factors that rely on the proportionate individual user experience's quality value parameters underneath quality of service measures, that of perception measurements, rendering quality measurements, and user state measurements respectively.

2.3 Quality of Experience Assessment for Video on Demand

A: Subjective QoE Assessment

Subjective QoE assessment aims at deriving quantitative data that depicts the perceived Quality of Experience by allowing humans to score video frames after watching them in a controlled environment as prescribed ITU-T and ITU-R in [40] [41] [38]. The scores are represented by Mean Opinion Scores whose approaches are shown below in Table 1 [17].

Table 1: *Summary of MOS Scoring Approaches*

Sr No.	MOS Scoring Approach	ITU Recommendation	Frames Evaluated	Grading Scale
1	SS/ACR	ITU-R Rec.BT.500, ITU-T Rec.P.910	One	ITU-R Quality Scale
2	ACR/HR	ITU-T Rec. P.910	One	ITU-R Quality Scale and DMOS
3	DSIS/DCR	ITU-R Rec.BT.500, ITU-T Rec. p. 910	Two	ITUR-R Impairment Scale
4	DSCQS	ITU-R Rec.BT.500	Two	ITU-R Quality Scale
5	SC	ITU-R Rec.BT.500	Two	ITU-R comparison Scale.
6	PC	ITU-T Rec. P.910	Two	ITU-R Quality or Comparison Scale.
7	SSCQE	ITU-R Rec.BT.1788	Multiple	Continuous Rating Scale between 0 and 100.
8	SDSCE	ITU-R Rec.BT.500	One, Long Sequences	Continuous Rating Scale between 0 and 100.
9	SDSCE	ITU-R Rec.BT.500	Two, Long Sequences	Continuous Rating Scale between 0 and 100

From the summary table, mean opinion scores are categorized as Single, Double, Multiple or Long Sequence Tests.

The Single Sequence Tests can be carried as either Single Stimuli (SS), Absolute Category Rating (ACR) or Absolute Category Rating with Hidden Reference (ACR-HR). In the first two tests categories, user is required to watch one video frame/sequence a time then score it based on either the ACR Five, Nine or Eleven grade. In the ACR-HR, the process is identical to ACR, except that all sequences of the unaltered references are displayed and rated separately. Then, between each sequence's MOS values and its hidden reference, a Differential MOS (DMOS) is calculated.

The Double Sequence Tests are executed as either Double Stimuli Impairment Scale (DSIS), Degradation Category Rating (DCR), Double Stimuli Continuous Quality Scale (DSCQS) or Stimulus Comparison. In first two i.e., DSIS and DCR, a user watches an unimpaired sequence of frames after which they are asked to score an impaired sequence with the same source based five grade ITU-R impairment scale. For the DSCQS, users are asked to score the quality of two sequences from the same source using the ITU-R five grade quality scale, with at least one sequence being unimpaired. In the SC, a user compares two video frame sequences and scores them using the ITU-R comparison scale. This method of scoring offers an index of the relationship between two presentations that are being compared. The last category of Double Sequence Tests is the PC where a user is shown a pair of sequences for each test. These two sequences are chosen from a pool of N sequences (for a total of N(N-1) potential pairs). The individual is asked to choose which of these two sequences is more preferable.

Multiple Sequence Test mainly refers to the Subjective Assessment of Multimedia Video Quality (SAMVQ) which is entails a participant seeing an explicit and unaltered reference video frames sequence as well as various different variations of the same video content. The respondent is then asked to grade all of the sequences on a scale of 0 to 100. After careful comparisons, the subject can evaluate and alter the scores of all variations during the test. A hidden reference is sometimes used to assess intrinsically, the quality of a given reference video sequence.

Long Sequence Tests include Single Stimulus Continuous Quality Evaluation (SSCQE) and the Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) methods. In the first one, a subject continuously watches and grade sequences of video frames using a 0 to 100 quality scale. Since the scoring is done continuously and with the option of play back, then a continuous video quality curve is obtained. The SDSCE has similarities with the SSCQE, however, the main distinction between these two methods is that in SDSCE, an unaffected reference sequence is played alongside the assessed video sequence. Eventually, only the impaired/tested sequence is graded by the subject.

B: Objective QoE Assessment:

While Subjective Quality of Experience assessments are the most reliable [86], for real time and continuous assessment of video sequences, objective parameters would be more efficient. This is because unlike other assessments, video content is characterized by unique and varied dynamic requirements which are best addressed through objective measures.

Highly dependent on the degree of information available for a video signal, objective video quality assessments can be categorized into; Full Reference (FR), Reduced Reference (RR) and No Reference (NR) [9] [8]. In the full reference method, reference multimedia materials are fully available for comparison with the received distorted/impaired contents in order to evaluate visual quality. The original signal is not available at the client or receiver end in most real situations. Studies by [49] [58] [66] depict Structural Similarity Image Index (SSIM), Multi-Scale Structural Similarity Index (MS-SSIM), Feature Similarity Index (FSIM), Gradient Magnitude Similarity Deviation (GMSD), and Perceptual Similarity Index (PSIM) as available metrics for objective similarity assessment which have all been developed using the FR method.

$$\mathrm{SSIM}\ (x,y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad \textbf{(4)}$$

where the first, second, and third terms in the first line of measuring image luminance, contrast and structure similarity, respectively; μ, σ, σ_{xy} are local mean, variance, and covariance; x, y indicate two images; c_1, c_2 are small two stabilization constants.

$$\mathrm{PLCC} = \frac{\sum_i (q_i - \bar{q}) \cdot (o_i - \bar{o})}{\sqrt{\sum_i (q_i - \bar{q})^2 \cdot \sum_i (o_i - \bar{o})^2}} \quad \textbf{(5)}$$

where s_i and p_i indicate the i-th image's subjective score and converted objective rating after nonlinear mapping, s^- and p^- are the mean of all s_i and p_i.

$$MS - \mathrm{SSIM}\ (x,y) = \frac{2f_x f_y + c}{f_x^2 + f_y^2 + c} \quad \textbf{(6)}$$

where f indicates the feature Phase Congruency (PC) or Gradient Magnitude (GM), x, y indicate two image while c is a stabilization constant. PC similarity and GM similarity are then multiplied to the FSIM.

$$\mathrm{FSIM} = \frac{\sum_{\mathbf{x}\in\Omega} S_L(\mathbf{x}) \cdot PC_m(\mathbf{x})}{\sum_{\mathbf{x}\in\Omega} PC_m(\mathbf{x})} \quad \textbf{(7)}$$

where we use $PC_m(\mathbf{x}) = \max(PC_1(\mathbf{x}), PC_2(\mathbf{x}))$ to weight the importance of $S_L(\mathbf{x})$ in the overall similarity between f_1 and f_2, and accordingly the FSIM index between f_1 and f_2 as defined above and where Ω means the whole image spatial domain.

$$GMS(i) = (2m_r(i)m_d(i) + c) \,/(m_r^2(i) + m_d^2(i)\text{+c}) \quad \textbf{(8)}$$

where c is a positive constant that supplies numerical stability, L is the range of the image intensity. The GMS index is computed in a pixel-wise manner; nonetheless, important to note that a value $\mathbf{m}_r(i)$ or $\mathbf{m}_d(i)$ in the gradient magnitude image is computed from a small local patch in the original image $\mathbf{r}$ or $\mathbf{d}$.

Figure 2: *Summary of Subjective QoE Assessment Techniques [2]*

Parameter	ACR	ACR-HR	DRC	PC	SSQE	DSIS	DSCQS	SDSCE	SCACJ
Stimulus type	Single	Single	Double	Comparison	Single	Double	Double	Double	Comparison
Video duration	10s	10s	10s	.	5m	10s	10s	5m	.
Explicit Video reference	No	No	Yes	No	No	Yes	Yes	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

In the reduced reference method, the need to offer particular properties like texture, edges among others of the original image or video for quality assessment instead of supplying the original image or video as a reference is underscored. With the proliferation of social networking, online games, and video streaming, RR-based quality evaluation has been a hot study field as observed by [21]. Therefore, RR based methods can be termed as intermediate between FR and NR since it only requires partial information on the multimedia content only on the receiver's end. In the no reference method, because there is no reference, the quality assessment of an image or video in this category is done blindly using extracted information from the multimedia content. However, evaluating the picture and video quality with NR is difficult since the recovered features may only provide a limited amount of information. As such, [66] proposes an NR picture quality measurement metric based on the free energy-based brain theory and features influenced by the Human Visual System (HVS). In addition, [31] suggested a different technique based on a Free Energy-based Distortion Measure (FEDM) and a structural degradation model. Due to the lack of a reference, no reference method is more difficult compared to the other two methods explained above.

In summary, subjective studies as discussed above have been proven to be very demanding and must comply with established rules like number of subjects, age, noise free/dedicated environment among others. The more number of participants in a subjective study, the higher the accuracy of of the overall QoE rating. For objective assessments, tools like google microwork site and Amazon Mechanical Trunk are now being used due to their ability to allow an internet service consumer to carry out formal tests. Recently, machine learning based methodologies like Pseudo Subjective Quality Assessment (PSQA) method based on a Random Neural Network (RNN) has also been used as a learning tool. in RNN, assessment results are used to capture the link and relationship between the distortion parameters and the perceived consistency.

3. REVIEW OF RELATED LITERATURE

This section presents an overview of related literature on Quality of Experience predictive models for Video on Demand. Specifically, the section presents state of the art and a review of techniques, methods and models published in period between 2016 - 2021.

3.1 State of the Art

Given the recent trends in QoE prediction, the state of the art is presented based of the following foremost fronts; (a) Over The Top (OTT) broadcasting of VoD content, (b) Adaptive Bitrate Streaming, (c) High Speed Broadband Requirements, (d) MULSEMEDIA QoE, (e) Machine Learning for QoE Prediction.

Over The Top (OTT) broadcasting of VoD content OTT is an emerging media distribution platform which uses internet over a public network to distribute digital media [3]. This paradigm shift has resulted in many companies creating individual media distributing applications called OTT Media Platforms (OMP). Compared to the classical media distribution techniques, OTT provides consumers with flexibility and ability to access the multimedia content through multiplicity of ways like smartphone apps, social media and websites. The advent and proliferation of OMP shows a new trend of perceiving entertainment as a bundle of several multimedia packages. In addition, this new paradigm is customer centric and with competition forcing industry player to remain creative and innovative [14].

HTTP adaptive streaming (HAS) also called MPEG-DASH is an adaptive bitrate streaming method that has in the recent past attracted a lot of attention. It is the technology behind most of the common VoD services like YouTube, Vimeo LLC, Show Max Amazon Video (Amazon VoD), Hulu and Netflix due to its ability to support transmission of high-quality multimedia content from the classical HTTP severs. Naturally, HTTP DASH initiates pertinent questions on how the varying quality levels should be treated with the main debate being instantaneous QoE during transition from different service levels [30].

High Speed Broadband Requirements At the backdrop of Video on Demand services is a requirement for fast and reliable internet connection. Unlike many online services, for optimal experience, VoD requires high download speed and network reliability. Many countries today offer 5G which is a new network design meant to deliver multi-GPS peak data speeds, dismal latency levels, high availability characterized by better user experience. As opined by [2], "5G networks will have to deal with dynamic, diverse, fast, and multi-tier networks, but they also should meet the desired QoE". As such, proposals have been put forth towards developing intelligent neural networks with potent to learn interrelation between resulting QoE, devices and network parameters [51].

MULtiple SEnsorial MEDIA (MULSEMEDIA QoE) is a modern advancement of multimedia which complements traditional multimedia with novelties like thermal, haptics and olfactory objects characteristics. This is increasingly becoming a common requirement especially for immersive, virtual and 3D based multimedia applications [81]. Ongoing research geared towards improving the QoE in Mulsemedia include; advancements and refinement of haptics interactions especially in kinesthetics and tactile perceptions, special mulsemedia effects like wind illumination, motion among others and how the they affect sensor-based data, cross modal interactions and communication of sensory effects in heterogeneous ecosystems [55] [65].

Machine Learning for QoE Prediction Majority of the non-AI/ML based QoE prediction models are reliant on the explicit modelling of highly non-linear behaviour of users and their perception. Comparing this classical approach to contemporary AI/ML based models, traditional models normally results into over fitting or uncertain overall reliability. It is these cons that are currently being addressed using deep learning. More attention is required to demonstrate how feature extraction and selection, classification, mapping, correlations and representational learning on QoE for VoD models [2].

Edge Assisted Video Distribution
Method. The exponential growth of rapid video apps poses significant technological challenges to current processing and connectivity networks. All videos are stored in the cloud and distributed to viewers via rival distribution networks or the Internet in standard techniques. They do, however, generate a lot of network traffic and have a long wait time, which has a detrimental influence on the consumer viewing experience. Towards this end, [16] proposes an edge-assisted rapid video distribution method to address these issues, which involves storing selected videos that consumers enjoy on edge servers, which individuals can obtain via increased network connections. Due to the limited computing and storage capacity of edge servers, they propose creation of an online algorithm with a provable approximation ratio to assess whether videos should be archived at edge servers, regardless of prospective network efficiency or watching patterns shifts. The opine that this can increase efficiency by taking video fetching and user-edge association into consideration simultaneously.

3.2 Related Work

This section presents recent QoE predictive models for Video on Demand published within the period 2015 to 2021. Service Quality assessment can be approached from various perspectives. for instance, network engineers will focus on quality of service and general network aspects, in the business discourse, they will focus on the customer churn, revenues per user while behaviour scientists will be concerned with experiences like happiness, sadness, enjoyment among others [54]. As such, various studies analysing diverse QoE concepts have been studied as discussed below.

Early studies like the one done by [61] used the regression method called Least Absolute Shrinkage and Selection Operator (LASSO) to evaluate the bit-stream parameters accuracy while using Full Reference. In their study, they conducted both objective and subjective measures on the original and impaired frames. The impaired video frames were either compressed and synthetic impairments added to them. The study opines in conclusion that visible losses have a heavy impact on video quality degradation and ultimately, the overall experience.

[69] proposed an ensemble model which brought together various bit-stream layer features using machine learning. Specifically, the study used Artificial Neural Network (ANN) to estimate the quality of video frames provided to users. This method was tested and compared with PSNR and it proved to provide better performance.

In their work, [89] developed tools to measure user perceptions when interacting with varying multimedia content. The work culminated in development of a Facebook application called Youo for automatic collection of user QoE data as Facebook users watch videos. This application has since been made open to researches for customization based on various stimuli [60] [88] [52] [87] [13].

[89] went ahead to model QoE based on user factors. They classified the user factors as either physiological, demographic, psychological and social-cultural factors. The physiological factors are aimed at collecting information about the sensory system of the user to build objective VoD quality assessment metrics. Demographic factors are mainly attributing that can be used to define a particular group of users. These may include their gender, age, nationality, education level among others which have been proven to have a correlation on the requirements for certain levels of QoE [80]. Psychological factors focus on the mood and state of the user.

[83] developed a psychological mapping model based on the Weber–Fechner law [64], and also leveraged on previous research [26] and [7] assessment models targeting the elements and using the logarithmic function as the primary mapping connection to build a viable QoE model by combining QoE and video characteristics. [85] in their work, they have consistently shown that the mood of a user has a major significance on the quality preferences and ultimately, the QoE rating. Social Cultural factors address the education, social, cultural. Generally speaking, social-cultural factor relates to the exposure of a user and how it affects their expectations while conducting QoE assessments [33].

In another study by [27], they presented a layered approach towards QoE assessment where various Influence Factors are aggregated and an overall QoE parameter derived. Using this approach, the output of upper and lower layers can be aggregated and enhanced to act as input for other layers. The five-layer considered were physical, network, virtualization, combination and application layers. The results from the paper show that while users expect to have accurate and synchronized data in terms of video quality and its presentations, these factors do not have a major impact on the user assessment of QoE compared to other IFs.

[37] also proposed a scalable layered QoE evaluation methodology that takes into consideration numerous quality parameters. Physical and metaphysical measurements were presented as new concepts for consideration. Physical metrics are connected to the technology architectural metrics like network, sensor, and computational quality, whereas metaphysical metrics are involved quality measures that end users frequently desire or expect to find in a system. Based on their relationship to the surrounding situation, the physical metrics were further divided into static and dynamic metrics. The metaphysical metrics are closely linked to the notions of Quality of Context QoC and Quality of Information QoI, which have been suggested in earlier studies such as [70].

[43] established a QoE prediction model which was based on fine-grained analysis of user perceptual motion characteristics and statistical content features that have potent in influencing user motion perception. The features were from user physiological characteristics, with the goal of predicting the degree of VR illness when watching VR videos to ensure a comfortable

viewing experience. After the experiments, it they concluded that the correctional between the proposed QoE model and subjective scores was found to be 72 percent in experimental results.

[32] Investigated proved the need to simultaneously ensure overall QoE maximization during adaptive video streaming modelling in a network with high variability requirements. The authors used QoE impact factors from historical literature, where QoE was expressed a linear equation derived from various taxonomic parameters whose summation leads to QoE estimation.

[34] employed a QoE model for rate adaptation, weighting the QoE parameters to optimize the utility of the QoE model under restrictions. Experiments reveal that VAS360 increases the user experience, as well as the user satisfaction. The viewport-adaptive solution was rate at between 23–45 percent better than the no viewport adaptive solution in terms of quality. The authors used a Quality of Experience (QoE) model for their research which entailed rate adaptation aimed at maximizing utility if the QoE model under constraints by weighting the taxonomic parameters. In their results, they observe that use of the VAS360 increases user experience, as well as user satisfaction.

The study by [23] proposed a instant QoE index based model called Streaming QoE Index (SQI). They used parameters like video stalling, effect of initial buffering and video compression to build a video database to be used for the experiments. The main objective of the objective was to investigate the interactions between the video quality and playback stalling with an aim deriving the QoE and determining optimization points from a user perspective. The results of the study depicted a good performance of SQI in determining optimization of media. Notably, the SQI model has limited motoring parameters for service degradation. streaming services.

In another study by [84], an Adaptive Video Streaming Evaluation framework (AdViSE) was proposed. This model has support for a variety of media types, networking parameters, and adaptation algorithm implementations. AdViSE includes a set of quality-of-service and quality-of-experience measurements collected and evaluated during adaptive streaming evaluation review as well as a segment log requests, which were used to create the impaired media for evaluation and judgment using subjective measures. Nonetheless, the study did not provide a source code level analysis of well-known HTTP Dynamic Adaptive Streaming (DASH) as well as famous commercial streaming players.

In another follow up study, the authors in [84] developed an end to end QoE evaluation model to collect and analyses user experiences both objectively and subjectively. For subjective measures, the authors used Web based Subjective Evaluation Platform (WESP) while for the objective measures, AdViSE was used [71].

[53] presented VideoNOC which is a video quality of experience monitoring prototype platform for mobile network operators. It uses specific video quality of experience parameters into account like bit-rate, rebuffering among others. VideoNOC enables network impact analysis, as well as demands for network conditions all through the experiments. The derived Quality of Experience (QoE) metric reflects video demand across the entire network and this can be used to develop and improve networks as well as streaming services.

In a similar spirit, ViCrypt model was presented in [67]. ViCrypt is an online Machine Learning (ML) model which was developed to predict re-buffering occurrences from encrypted video streaming traffic in real-time. After subdividing the video streaming session into a number of time periods of equal length, a fine-grained time slot length of 1 second to extract the characteristics (for a proper compromise between precision and stalling delay detection) is applied. After that, they're fed into an ML model to anticipate when stalling will happen. It's worth noting that the initial delay and duration of stalling events can be retrieved as well.

As a follow-up to their earlier work, the authors showed that ViCrypt can reliably estimate video resolution and average video bit rate. This was well articulated by the authors in [76] that ViCrypt can reliably estimate video resolution and average video bit rate.

In another study by [72], the authors also chose to use a machine learning model to estimate the rebuffering ratio based on hidden and context information in order to improve the precision of prediction using logistic regression.

[48] used supervised machine learning and a support vector machine to predict users' QoE based on the number of active users and channel conditions. Based on cell-related information acquired at the start of a video session, they categorize a session into two categories (i.e., with or without stall incidents). Mobile users are more likely to face starvation events than users in the adaptive streaming and static ecosystems. In the ir findings, they opine that the predictive model was accurate and convenient in predicting the starvation events and the impact on QoE.

In the same vein, [18] have also created a multistage ML cognitive technique for QoE prediction. In this study, the proposed model combines unsupervised learning of video attributes with a supervised classifier trained to automatically extract quality-rate features. The results depict that this model outperformed other previous ML based approaches to video QoE prediction

Similarly, [59] presents YouQ, which is an android app that uses machine learning to predict QoE based on taxonomic features like delays, video stalls, quality of playout, and its variations using objective metrics collected from a stream of encrypted packets. Despite the encouraging outcome, the bulk of the characteristics are dependent on TCP, implying that these strategies will most likely fail in the case of UDP.

In another related study, the authors [50] proposed a deep learning-based QoE detector dependent on data extracted from network packets. An RNN, Convolutional Neural Network, and Gaussian Process (GP) classifier are used in the model. This classifier can detect and forecast video problems (such as black pixels, ghosts, blockness, columns, chrominance, colour bleeding, and blur) at the present time interval (in one second). The model is designed to predict video quality of experience in real time, but it may run into problems due to a lack of training data.

Based on the expectation-confirmation theory and the video dataset constructed by [22], the ECT-QoE framework was developed with the ability to predict the QoE of streaming over DASH at any given time. The findings show that model outperformed a number of other models, particularly when paired with the SSIMplus model. Notably, ECT-QoE, can only be used on videos with view parts.

Unlike earlier proposals, [78] model examines the global intensity & local texture metrics taken from a decoded video to forecast video stalls and evaluate the user's quality assessments. The approach uses linear combinations to map the normalized number and length of stalls. When compared with other models, their proposition appears to be more consistent especially on the subjective perspective front.

Trend analysis of literature depict an shift from the classical objective and subjective based QoE models to hybrid QoE models that combines subjective and objective taxonomic measures using ML algorithms, statistics and other fields. This has been shown to be the most accurate way of measuring and predicting QoE as it can be used in real-time and eliminates the weaknesses of the traditional approaches [82] [79] [10].

Trend analysis through literature depicts an increasing shift from traditional measurement and prediction of QoE which relied on objective and subjective taxonomic parameters towards use of hybrid methods. Hybrid methods takes the approach of combining subjective and objective measurements using ML, statistical methods and other fields. It has been shown through numerous studies that the hybrid approach is more accurate than traditional ways because it not only eliminates the shortcomings of traditional approaches but also enables real time measurement and prediction of QoE [82] [79] [10]. A case in point the PSQA assessment method whose proponents argued that the QoE measures it produced were similar to human results based on accuracy. PSQA is based on the training of a statistical learning approach known as a Random Neural Network (RNN). The influence factors on the quality are chosen to generate numerous distorted video samples in order to evaluate the video quality. Following that, the samples are subjectively evaluated. The RNN is then trained using the observations to understand the relationship between the components that create the distortion as well as overall perception by real humans. After completing the training technique, the trained neural network model can be used in real time.

In another study [1], four machine learning algorithms were combined namely Decision Tree, Neural Network, K-Nearest Neighbour and Random Forest. The aim was to evaluate Mean Opinion Score based on the VQM and SSIM methods. The Pearson correlation coefficient and the Root Mean Square Error are used to evaluate the performance of these algorithms. The Random Forest algorithm performed the best in terms of anticipating user impression, according to the findings. Network factors such as transmission latency and reaction time, on the other hand, are not taken into account.

Further, [15] created MLQoE, a modular user centred approach based on supervised learning to connect QoE and network statistics like average latency, packet loss, and average jitter. Multiple machine learning algorithms are used in the framework (i.e., Artificial Neural Networks, Support Vector Regression Machines, Decision Trees, and Gaussian Naive Bayes.). The one that outperforms the others will be chosen automatically, along with its parameters, based on the dataset used as input. In comparison to other known machine learning models, MLQoE can accurately predict the QoE score, according to their findings. In addition, the authors of [47] went ahead to propose a trained machine learning model that has the ability to predict the mean opinion score value in software defined networks using the following network parameters; bandwidth, jitter, and delay.

In a related paper, [68] YoMoApp (YouTube Monitoring App) which is a monitoring tool developed as an android application for both mobile and WIFI networks streaming settings. It monitors the application and the network layer (i.e., the total quantity of uploaded and downloaded data is logged frequently). It does this by obtaining subjective QoE assessments from end-users which are expressed in form of mean opinion scores. The information is anonymously uploaded to a third-party database after which using all of the users provided data, a map is created that shows how each network operator works and how they can be benchmarked. YoMoApp delivers precise measurements on a sufficiently short time period (1 second). They suggested that QoE metrics take into account longer video segments and in order to test the precision of alternative methodologies, the latter was used, along with another Android-based passive monitoring application. As a result, streaming characteristics showed higher correlations with subjective quality than with objective quality, indicating that it is more suited for QoE monitoring. The downside of the tool is that it makes use of JavaScript, which might occasionally lead to inconsistencies and problems.

YoMoApp was also utilized by [77] to monitor video sessions and extract many parameters from end-user cell phones such as signal quality and the quantity of incoming and outgoing bytes subject to the device of choice. They offer a lightweight approach to forecast Video streaming QoE parameters like initial latency, number, and ratio of stalls, as well as user engagement, using machine learning. According to their assessment, network layer features are sufficient for obtaining reliable findings.

[5] proposed a machine learning technique dubbed Video Assessment of Temporal Artifacts and Stalls (ATLAS). ATLAS predicts QoE using VQA method that combines QoE related variables and memory features. In their work, the LIVE-Netflix Video QoE Database was used

to evaluate their model. The algorithm, however, is only capable of delivering overall QoE scores and cannot be utilized to make real-time bit-rate decisions [6].

In more recent studies related work like, [4] the approach involves a server-side assessment which determines a bit-rate for each user connected to the bottleneck connection using reinforcement learning (RL), but in a way that is fair to all concurrent DASH users. In their future work, they suggested boosting their research work in order to reduce the quantity of process required to keep the process balanced. They're adding an LSTM model (particularly, an LSTM recurrent neural network) to the server's learning process to improve the consistency and quantity of behaviour it takes to better direct the user.

Another recent paper by, [19] They aimed examining QoE as it relates to smartphone applications running over Wi-Fi or cellular networks, with the goal of establishing predictive QoE models and developing guidelines for app developers. to collect and use data aggregations and collecting procedures and processes such as physical behaviour, network usage, and battery knowledge in order to create numerous QoE models. Although, according to several surveys, the user's intent of the software (i.e., WhatsApp's intention to achieve anything, such as listening to music or using Spotify), and finally, network QoS, device usage on a smartphone is the most important of these three desired features.

On the same verge, [20] laid a lot of emphasis on network traffic, which is regularly disseminated to users via Wi-Fi connection points.
When the Wi-Fi connection is down, the quality of the service is prone to deterioration, which may generate unfairness for those who are using the same network. They offer a strategy for improving QoE fairness by taking into account the particular video and resource demand features of Wi-Fi streams that redistribute certain slices over DASH flows in this work.

Finally, the work by [56] used the widely known Internet Control Message Protocol probes. In their work, they offer a realistic approach for estimating quality if experience for video on demand. Based on the ITU-T P.1203 Recommendation, measured network circumstances are used as input to a ml model that calculates QoE in terms of mos. Switches in video quality and playback delays are included in the estimate. They estimate MOS for a catalog of 25 different videos using an average Root Mean Square Error of 1.05, training a model using sessions from the shortest video, and evaluate generalization to the entire catalog.

The surveyed literature depicts Root Mean Square Error (RSME), Spearman Rank-order Correlation Coefficient (SRCC) and Pearson Linear Correlation Coefficient (PLCC) as the most common performance evaluation and quality indicators. While the SRCC index assesses the consistency of subjective and objective ratings predicted by IQA measures, the PLCC and RMSE indicators assess the accuracy of image quality algorithms' anticipated scores. The higher the SRCC and PLCC values, and the lower the RMSE value, the better the performance of the associated IQA metric.
RMSE is given by the equation;

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - s_i)^2} \qquad \textbf{(9)}$$

where s_i and p_i indicate the i-th image's subjective score and converted objective rating after nonlinear mapping, $\bar{s}$ and $\bar{p}$ are the mean of all s_i and p_i. The PLCC is computed as;

$$PLCC = \frac{\sum_t (a_i - \bar{a})(l_t - \bar{l})}{\sqrt{\sum_i (a_i - \bar{a})^2 \sum_i (l_i - \bar{l})^2}} \tag{10}$$

where a_i and $\bar{a}$ are the estimated quality score of the i-th synthesized image and the average value of all a_i, respectively. l_i and $\bar{l}$ are the subjective quality label of the i-th synthesized image and the average value of all l_i, respectively. The SRCC is defined as:

$$SRCC = 1 - \frac{6\sum_{q-1}^{Q} d_q^2}{Q(Q^2 - 1)} \tag{11}$$

where Q is the number of pairs of predicted quality scores and subjective quality labels. d_q represents the ranking difference between the predicted quality scores and the subjective quality labels in each group.

4. METHODOLOGY

In the sections above, we have provided a theoretical background and a survey of related literature on quality of experience models for predicting video on demand. The surveyed literature has provided a snapshot of the state of the art as well as existing work in this domain. The literature review was guided by the original methodology for conducting literature review as proposed by [44]. Kitchenham proposed the following stages;

1. Formulate the research questions (RQs).
2. Define the search process (what are the source selection and search keywords?).
3. Define the exclusion/inclusion criteria and the quality assessment to be followed.
4. Conduct data collection and extraction.
5. Analyse the results.

Stage One: The first step was formulation of the research questions to guide the process of identification of related work. The following Research Questions were formulated;

i) RQ1 - Which QoE prediction models for Video on Demand have been employed between 2016 & 2021?
ii) RQ2 - Can the models be categorized based on the taxonomic parameters used i.e., subjective, objective, machine learning or hybrid parameters?

Table 2: *Selection Criteria Depicting EC, IC & QA*

ID	Exclusion Criteria (EC)
EC1	Papers that are still in the early phases of development or are not fully developed.
EC2	Workshop and tutorial summaries, editorials that do not give sufficient information.
EC3	Duplicate entries.
EC4	Paper content is not related to QoE prediction.

EC5	Opinion papers and discussion papers which do not propose a a QoE prediction model.
EC6	Paper whose full content is not accessible or are published in journals which the author has not access to.
EC7	Papers published in a language other than English.
ID	**Inclusion Criteria (IC)**
IC1	Journal and conference publications that report on, discuss, or explore QoS/QoE assessment in its entirety.
IC2	Papers published between 2015 to 2021.
IC3	Papers published in English language.
IC4	Papers that propose prediction models based on subjective, objective, ML or hybrid taxonomic parameters.
ID	**Quality Assessment (QA)**
QA1	The paper is based on research.
QA2	The paper has a clear statement of the objective.
QA3	The paper provides a clear description of the context under which the study was executed.
QA4	The QoE assessment methods and approached are well defined.
QA5	The paper provides a clear statement and discussion of the findings.
QA6	The paper provides a discussion on how the results were validated.

Stage Two: The search process was mainly conducted using the google scholar search engine. The target was papers published in SCI, SCIE and Scopus journals and conference proceedings.
Stage Three: The process followed to determine the criteria for inclusion or exclusion is summarized in table 2 below. In order for a paper to be included in the final selection, it must meet all of the inclusion criteria listed in this table. Each exclusion criterion, on the other hand, is sufficient to eliminate a paper. Finally, we employ six quality evaluation (QA) questions to grade each paper's quality. Each QA question is responded with a score of Y(Yes), P (Partial), or N (No), with Y = 1, P = 0.5, and N = 0. A paper is only included in the final selection if the summation of the QA scores across all six questions is more than or equal to four.
Stage Four: The fourth step is data collection and extraction. A generic search lead us to 89 papers. Strictly following the framework depicted above in table 2 by considering the inclusion, Exclusion and Content Criteria, we were left with 43 papers which we reviewed and provided a critique.
Stage Five: The last stage in the literature review ought to provide an analysis of the results. In this paper, to examine the findings, we first look at the various types of assessment methods used in the selected publications, using the classification described in earlier sections. Then, because the number of articles that were ultimately chosen is small, we present a brief overview organized by the various sorts of assessment methodologies.

Figure 3: *Categorization of Taxonomic Parameters & Approaches in Selected Papers*

Id	QoS	QoE
S1	Packet delay, end-to end delay, bandwidth, throughput, packet loss, jitter, bitrate.	X
S2	X	Objective (VQM)
S3	Jitter, Latency, packet loss, bandwidth	Subjective (MOS)
S4	X	Subjective (MOS)
S5	Call setup time, end-to-end delay, jitter	X
S6	End-to-end delay, bitrate, jitter	X

S7	Throughput, jitter, packet loss	X
S8	Throughput, bandwidth, packet loss, bitrate, picture loss indication, bucket delay	Subjective
S9	Call setup time, end-to-end delay	X
S10		Objective(PESQ, PEVQ)
S11	Throughput, end-to-end delay, error rate	X
S12	End-to-end delay	
S13	End-to-end delay	Objective (PESQ, SSIM, PSNR)
S14	Bandwidth	Subjective
S15	Bandwidth, jitter, bitrate, framerate, packet rate, packet loss, latency, packet delay.	X

From the figure above, Serial number 1 to15 shows the methods and taxonomy used to evaluate and predict the overall QoE by the selected papers. According to the findings, End-to-end delay, jitter, packet loss, bandwidth, throughput, and bitrate are the preferable characteristics. Subjective MOS is the recommended technique of assessment for QoE, followed by numerous objective measurement methods namely PESQ, PEVQ, VQM, PSNR, and SSIM.

5. OPEN ISSUES AND RESEARCH GAPS

Despite the fact that QoE modelling has recently received a lot of attention, it is still a difficult topic to tackle because it is a trans-disciplinary and subjective domain. As such, given the reviewed literature, and to evolve the presented state if the art, the following are presented as open issues with potent to shape future research direction.

1. **QoE for future display technologies**. Display technology has been advancing in recent years to allow users to enjoy high-quality content. With larger screens and more complex user interaction, these technologies continue to improve consumer happiness. The augmented resolution
   of 2D displays to S3D displays creates room for a more engaging experience, all of which can be extrapolated from this ongoing technological advancement.

2. **Deep Learning for HVS based QoE Modelling.** As previously observed in the sections above, deep CNN has emerged as a core technology capable of breaking most performance records in the domain of QoE through intense dataset-based training. As a result, several attempts have been made to use the deep-learning technique to discover new factors without needing prior data. New QoE measurements are expected to be generated by modelling the Human Visual System. As a result, the motion component of an image and the weighting process of the human visual mechanism retrieved from the deep model could be projected as useful metrics in calculating the visually perceived QoE.

3. **Standardization of QoE Taxonomy for QoE Assessment & Prediction.** Existing literature shows the absence of a standard taxonomy of objective or subjective QoE factors that are associated with human judgment and perception. While there exist some efforts and propositions on taxonomic parameters, non-can pass for a industry-wide standard. As a result, it's critical to establish appropriate taxonomy standards, which will aid not only in quality measurement but also in standardizing the framework of commercial QoE systems.

4. **QoE Based Physiological Parameters Using Less Obtrusive Sensors.** The time is ripe for establishment of the groundwork for future multimedia services and technologies by evaluating physiological measures in connection to QoE Taxonomy. Understanding a user's physiology for instance their eyesight and hearing ability, cognitive and emotional state as well as behavioural phenomena could potentially lead to a better understanding of many influence factors and their relationship to QoE. Physiological observable parameters include electroencephalography (EEG), functional magnetic resonance imaging (fMRI), skin temperature, brain activity among others.

   One critical concern with physiological assessments is intrusiveness nature of how they are collected, particularly those that require subjects to wear sensors, potentially causing the user to disrupt their typical behaviour and to some extent experience pain. There exist an eminent need and research gap in devising less intrusive methods in gathering physiological parameters.

5. **Trusted Crowdsourcing.** The absence of publicly available large datasets with higher subjective evaluations is a persistent problem in QoE evaluation. While existing Crowdsourcing methods such as Amazon's Mechanical Turk, Facebook, and Micro workers may be a viable answer to this problem, there is need to address the question of trust. Crowdsourcing methods bring the evaluation out of the lab and into the Internet, allowing for a larger and more diversified pool of subjects, as well as the inclusion of real-world locations in the assessment task and a faster turnaround time. However, because of the anonymity of users on crowdsourcing platforms and the loss of total control and the trustworthiness of ratings may be questionable.

6. CONCLUSION

In this paper, we have examined the QoE prediction models for video on Demand services and classification various assessment methods from a comprehensive review viewpoint. An analysis of existing literature on QoE models has been presented based on the taxonomic parameters and the assessment approach used to predict QoE. Finally, novel open issues within this domain were presented. Despite significant advances, we observe that more precise measurement of QoE continue to be a difficult task due to the fact that QoE is user centred, individual and multidimensional.

Given the aforementioned, we opine that there is a great need for significant research efforts aimed at finding efficient and optimal algorithms as well as models for assessment and prediction of QoE for Video on Demand services. Goes without saying, these demands a huge level of effort to unearth the novel potential of QoE and ultimately improve the end user experiences and satisfaction.

REFERENCES

[1] Tasnim Abar, Asma Ben Letaifa, and Sadok El Asmi. Machine learning based qoe prediction in sdn networks. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1395–1400. IEEE, 2017.

[2] Zahid Akhtar, Kamran Siddique, Ajita Rattani, Syaheerah Lebai Lutfi, and Tiago H Falk. Why is multimedia quality of experience assessment a challenging problem? *IEEE Access*, 7:117897–117915, 2019.

[3] Evangelos Alexiou, Nanyang Yang, and Touradj Ebrahimi. Pointxr: A toolbox for visualization and subjective evaluation of point clouds in virtual reality. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2020.

[4] Sa'di Altamimi and Shervin Shirmohammadi. Qoe-fair dash video streaming using server-side reinforcement learning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(2s):1–21, 2020.

[5] Christos G Bampis and Alan C Bovik. Learning to predict streaming video qoe: Distortions, rebuffering and memory. *arXiv preprint arXiv:1703.00633*, 2017.

[6] Christos George Bampis, Zhi Li, Anush Krishna Moorthy, Ioannis Katsavounidis, Anne Aaron, and Alan Conrad Bovik. Study of temporal effects on subjective video quality of experience. *IEEE Transactions on Image Processing*, 26(11):5217–5231, 2017.

[7] Benjamin Belmudez and S Moller. An approach for modeling the effects of video resolution and size on the perceived visual quality. In *2011 IEEE International Symposium on Multimedia*, pages 464–469. IEEE, 2011.

[8] Abdelkader Benmir, Ahmed Korichi, Abdelhabib Bourouis, and Mohammed Alreshoodi. Survey on qoe/qos correlation models for video streaming over vehicular ad-hoc networks. *Journal of computing and information technology*, 26(4):267–287, 2018.

[9] Abdelkader Benmir, Ahmed Korichi, Abdelhabib Bourouis, Mohammed Alreshoodi, and Laith Al-Jobouri. Geoqoevanet: Qoe-aware geographic routing protocol for video streaming over vehicular ad-hoc networks. *Computers*, 9(2):45, 2020.

[10] Ahmed Binsahaq, Tarek R Sheltami, and Khaled Salah. A survey on autonomic provisioning and management of qos in sdn networks. *IEEE Access*, 7:73384–73435, 2019.

[11] Khadija Bouraqia, Essaid Sabir, Mohamed Sadik, and Latif Ladid. Quality of experience for streaming services: measurements, challenges and insights. *IEEE Access*, 8:13341–13361, 2020.

[12] Kjell Brunnström, Sergio Ariel Beker, Katrien De Moor, Ann Dooms, Sebastian Egger, Marie-Neige Garcia, Tobias Hossfeld, Satu Jumisko-Pyykkö, Christian Keimel, Mohamed-Chaker Larabi, et al. Qualinet white paper on definitions of quality of experience. 2013.

[13] Özge Celenk, Thomas Bauschert, and Marcus Eckert. Machine learning based kpi monitoring of video streaming traffic for qoe estimation. *ACM SIGMETRICS Performance Evaluation Review*, 48(4):33–36, 2021.

[14] Imon Chakraborty, Ravinder Kumar Verma, Mohit Hiteshkumar Joshi, and P Vigneswara Ilavarasan. Customers' willingness to use new ott media platforms in india. In *ICT Systems and Sustainability*, pages 751–760. Springer, 2022.

[15] Paulos Charonyktakis, Maria Plakia, Ioannis Tsamardinos, and Maria Papadopouli. On user-centric modular qoe prediction for voip based on machine-learning algorithms. *IEEE Transactions on mobile computing*, 15(6):1443–1456, 2015.

[16] Fahao Chen, Peng Li, Deze Zeng, and Song Guo. Edge-assisted short video sharing with guaranteed quality-of-experience. *IEEE Transactions on Cloud Computing*, 2021.

[17] Satyaki Datta et al. A qualitative study proposing service quality dimensions for video-on-demand services through overthe-top medium. *International Journal of Pervasive Computing and Communications*, 2022.

[18] Michele De Filippo De Grazia, Daniel Zucchetto, Alberto Testolin, Andrea Zanella, Marco Zorzi, and Michele Zorzi. Qoe multi-stage machine learning for dynamic video streaming. *IEEE Transactions on Cognitive Communications and Networking*, 4(1):146–161, 2017.

[19] Alexandre De Masi and Katarzyna Wac. Towards accurate models for predicting smartphone applications' qoe with data from a living lab study. *Quality and user experience*, 5(1):1–18, 2020.

[20] Mauricio de Oliveira and Daniel F Macedo. Slicing wi-fi links based on qoe video streaming fairness. *International Journal of Network Management*, 31(6):e2155, 2021.

[21] Shahi Dost, Faryal Saud, Maham Shabbir, Muhammad Gufran Khan, Muhammad Shahid, and Benny Lovstrom. Reduced reference image and video quality assessments: review of methods. *EURASIP Journal on Image and Video Processing*, 2022(1):1– 31, 2022.

[22] Zhengfang Duanmu, Kede Ma, and Zhou Wang. Quality-of-experience for adaptive streaming videos: An expectation confirmation theory motivated approach. *IEEE Transactions on Image Processing*, 27(12):6135–6146, 2018.

[23] Zhengfang Duanmu, Kai Zeng, Kede Ma, Abdul Rehman, and Zhou Wang. A quality-of-experience index for streaming video. *IEEE Journal of Selected Topics in Signal Processing*, 11(1):154–166, 2016.

[24] Ching-Ling Fan, Tse-Hou Hung, and Cheng-Hsin Hsu. Modeling the user experience of watching 360 videos with head-mounted displays. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 18(1):1–23, 2022.

[25] Markus Fiedler, Usha Chapala, and Sridhar Peteti. Modeling instantaneous quality of experience using machine learning of model trees. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2019.

[26] Markus Fiedler, Tobias Hossfeld, and Phuoc Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *IEEE Network*, 24(2):36–41, 2010.

[27] Alessandro Floris and Luigi Atzori. Managing the quality of experience in the multimedia internet of things: A layeredbased approach. *Sensors*, 16(12):2057, 2016.

[28] Boni García, Micael Gallego, Francisco Gortázar, and Antonia Bertolino. Understanding and estimating quality of experience in webrtc applications. *Computing*, 101(11):1585–1607, 2019.

[29] Ahmed Sileh Ghafil and Israa Hadi Ali. State-of-the-art in qoe for video streaming. In *2021 International Conference on Advanced Computer Applications (ACA)*, pages 206–210. IEEE, 2021.

[30] Monalisa Ghosh, Dr Chetna Singhal, and Rushikesh Wayal. Desvq: Deep learning based streaming video qoe estimation. In *23rd International Conference on Distributed Computing and Networking*, pages 19–25, 2022.

[31] Ke Gu, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, and Longfei Liang. Noreference image quality assessment metric by combining free energy theory and structural degradation model. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2013.

[32] Yi Han, Yafeng Ma, Yangzhe Liao, and Gabriel-Miro Muntean. Qoe oriented adaptive streaming method for 360° virtual reality videos. In *2019 IEEE SmartWorld, Ubiquitous*

Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), pages 1655–1659. IEEE, 2019.

[33] Brett Holfeld, Kyle Hill, and F Richard Ferraro. Cognitive psychology. *The Encyclopedia of Cross-Cultural Psychology*, 1:189–192, 2013.

[34] Yuxiang Hu, Yu Liu, and Yumei Wang. Vas360: Qoe-driven viewport adaptive streaming for 360 video. In *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 324–329. IEEE, 2019.

[35] J Barakovic´ Husic´, S Barakovic´, and A Veispahic´. What factors influence the quality of experience for webrtc video calls? In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 428–433. IEEE, 2017.

[36] Selim Ickin, Katarzyna Wac, Markus Fiedler, Lucjan Janowski, Jin-Hyuk Hong, and Anind K Dey. Factors influencing quality of experience of commonly used mobile applications. *IEEE Communications Magazine*, 50(4):48–56, 2012.

[37] Yasuhiro Ikeda, Shouji Kouno, Akihiro Shiozu, and Katsunori Noritake. A framework of scalable qoe modeling for application explosion in the internet of things. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pages 425–429. IEEE, 2016.

[38] ITU-R. Itu-r rec. bt.1788 methodology for the subjective assessment of video quality in multimedia applications," int. telecommun. union, geneva, switzerland. 2007.

[39] ITU-R. Recommendation bt.1788. methodology for the subjective assessment of video quality in multimedia applications. 2007.

[40] ITU-R. Itu-r rec. bt.500-13 methodology for the subjective assessment of the quality of television pictures," int. telecommunication. Jan. 2012.

[41] ITU-T. Itu-t rec. p.910 subjective video quality assessment methods for multimedia applications," int. telecommun. union, geneva, switzerland. 2008.

[42] ITU-T. Recommendation. p.800.2. mean opinion score interpretation and reporting. 2016.

[43] Jaekyung Kim, Woojae Kim, Sewoong Ahn, Jinwoo Kim, and Sanghoon Lee. Virtual reality sickness predictor: Analysis of visual-vestibular conflict and vr contents. In *2018 tenth international conference on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE, 2018.

[44] Barbara Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26, 2004.

[45] Georgios Kougioumtzidis, Vladimir Poulkov, Zaharias D Zaharis, and Pavlos I Lazaridis. A survey on multimedia services qoe assessment and machine learning-based prediction. *IEEE Access*, 10:19507–19538, 2022.

[46] Asif Ali Laghari, Hui He, Shehnila Zardari, and Muhammad Shafiq. Systematic analysis of quality of experience (qoe) frameworks for multimedia services. *IJCSNS*, 17(5):121, 2017.

[47] Asma Ben Letaifa. Adaptive qoe monitoring architecture in sdn networks: Video streaming services case. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1383–1388. IEEE, 2017.

[48] Yu-Ting Lin, Eduardo Mucelli Rezende Oliveira, Sana Ben Jemaa, and Salah Eddine Elayoubi. Machine learning for predicting qoe of video streaming in mobile networks. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2017.

[49] Hao Liu, Ce Li, Dong Zhang, Yannan Zhou, and Shaoyi Du. Enhanced image no-reference quality assessment based on colour space distribution. *IET Image Process.*, 14(5):807–817, 2020.

[50] Manuel Lopez-Martin, Belen Carro, Jaime Lloret, Santiago Egea, and Antonio Sanchez-Esguevillas. Deep learning model for multimedia quality of experience prediction based on network flow packets. *IEEE Communications Magazine*, 56(9):110–117, 2018.

[51] Asma Lounis, Farid Alilat, and Nazim Agoulmine. Neural network model of qoe for estimation video streaming over 5g network. In *6th International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE 2018)*, pages 21–27, 2018.

[52] Octavian Machidon, Jani Asprov, Tine Fajfar, and Veljko Pejovic. Context-aware adaptation of mobile video decoding resolution. *arXiv preprint arXiv:2202.09772*, 2022.

[53] Tarun Mangla, Ellen Zegura, Mostafa Ammar, Emir Halepovic, Kyung-Wook Hwang, Rittwik Jana, and Marco Platania. Videonoc: Assessing video qoe for network operators using passive measurements. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pages 101–112, 2018.

[54] Taha Mansouri, Ali Nabavi, Ahad Zare Ravasan, and Hamid Ahangarbahan. A practical model for ensemble estimation of qos and qoe in voip services via fuzzy inference systems and fuzzy evidence theory. *Telecommunication Systems*, 61(4):861–873, 2016.

[55] Gebremariam Mesfin, Nadia Hussain, Alexandra Covaci, and Gheorghita Ghinea. Using eye tracking and heart-rate activity to examine crossmodal correspondences qoe in mulsemedia. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 15(2):1–22, 2019.

[56] Gilson Miranda, Daniel Fernandes Macedo, and Johann M Marquez-Barja. Estimating video on demand qoe from network qos through icmp probes. *IEEE Transactions on Network and Service Management*, 2021.

[57] Sebastian Möller and Alexander Raake. *Quality of experience: advanced concepts, applications and methods*. Springer, 2014.

[58] Yuzhen Niu, Yini Zhong, Wenzhong Guo, Yiqing Shi, and Peikun Chen. 2d and 3d image quality assessment: A survey of metrics and challenges. *IEEE Access*, 7:782– 801, 2018.

[59] Irena Orsolic, Dario Pevec, Mirko Suznjevic, and Lea Skorin-Kapov. A machine learning approach to classifying youtube qoe based on encrypted network traffic. *Multimedia tools and applications*, 76(21):22267–22301, 2017.

[60] Irena Orsolic, Mirko Suznjevic, and Lea Skorin-Kapov. Youtube qoe estimation from encrypted traffic: Comparison of test methodologies and machine learning based models. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2018.

[61] Katerina Pandremmenou, Muhammad Shahid, Lisimachos P Kondi, and Benny Lövström. A no-reference bitstream-based perceptual model for video quality estimation of videos affected by coding artifacts and packet losses. In *Human Vision and Electronic Imaging XX*, volume 9394, pages 486–497. SPIE, 2015.

[62] Sohee Park, Arani Bhattacharya, Zhibo Yang, Samir R Das, and Dimitris Samaras. Mosaic: Advancing user quality of experience in 360-degree video streaming with machine learning. *IEEE Transactions on Network and Service Management*, 18(1):1000–1015, 2021.

[63] Kandaraj Piamrat, Cesar Viho, Jean-Marie Bonnin, and Adlen Ksentini. Quality of experience measurements for video streaming over wireless networks. In *2009 Sixth*

International Conference on Information Technology: New Generations, pages 1184–1189. IEEE, 2009.

[64] Peter Reichl, Sebastian Egger, Raimund Schatz, and Alessandro D'Alconzo. The logarithmic nature of qoe and the role of the weber-fechner law in qoe assessment. In *2010 IEEE International Conference on Communications*, pages 1–5. IEEE, 2010.

[65] Estêvão Bissoli Saleme, Celso AS Santos, and Gheorghita Ghinea. A mulsemedia framework for delivering sensory effects to heterogeneous systems. *Multimedia Systems*, 25(4):421–447, 2019.

[66] Michael James Scott, Sharath Chandra Guntuku, Yang Huan, Weisi Lin, and Gheorghita Ghinea. Modelling human factors in perceptual multimedia quality: On the role of personality and culture. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 481–490, 2015.

[67] Michael Seufert, Pedro Casas, Nikolas Wehner, Li Gang, and Kuang Li. Streambased machine learning for real-time qoe analysis of encrypted video streaming traffic. In *2019 22nd Conference on innovation in clouds, internet and networks and workshops (ICIN)*, pages 76–81. IEEE, 2019.

[68] Michael Seufert, Nikolas Wehner, Florian Wamser, Pedro Casas, Alessandro D'Alconzo, and Phuoc Tran-Gia. Unsupervised qoe field study for mobile youtube video streaming with yomoapp. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2017.

[69] Muhammad Shahid, Joanna Panasiuk, Glenn Van Wallendael, Marcus Barkowsky, and Benny Lövström. Predicting full-reference video quality measures using hevc bitstream-based no-reference features. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 1–2. IEEE, 2015.

[70] Zhanwei Sun, Chi Harold Liu, Chatschik Bisdikian, Joel W Branch, and Bo Yang. Qoi-aware energy management in internet-of-things sensory environments. In *2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, pages 19–27. IEEE, 2012.

[71] Christian Timmerer, Anatoliy Zabrovskiy, and Ali Begen. Automated objective and subjective evaluation of http adaptive streaming systems. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 362–367. IEEE, 2018.

[72] Vladislav Vasilev, Jérémie Leguay, Stefano Paris, Lorenzo Maggi, and Mérouane Debbah. Predicting qoe factors with machine learning. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2018.

[73] Maria Torres Vega, Cristian Perra, Filip De Turck, and Antonio Liotta. A review of predictive quality of experience management in video streaming services. *IEEE Transactions on Broadcasting*, 64(2):432–445, 2018.

[74] M Vijayalakshmi and Linganagouda Kulkarni. Analysis of quality of experience (qoe) in video streaming over wifi in real time. In *Innovations in Computer Science and Engineering*, pages 79–85. Springer, 2021.

[75] Cisco VNI. Forecast and methodology. 2017.

[76] Sarah Wassermann, Michael Seufert, Pedro Casas, Li Gang, and Kuang Li. Let me decrypt your beauty: Real-time prediction of video resolution and bitrate for encrypted video streaming. In *2019 Network Traffic Measurement and Analysis Conference (TMA)*, pages 199–200. IEEE, 2019.

[77] Sarah Wassermann, Nikolas Wehner, and Pedro Casas. Machine learning models for youtube qoe and user engagement prediction in smartphones. *ACM SIGMETRICS Performance Evaluation Review*, 46(3):155– 158, 2019.

[78] Qingbo Wu, Hongliang Li, Fanman Meng, and King N Ngan. Toward a blind quality metric for temporally distorted streaming video. *IEEE Transactions on Broadcasting*, 64(2):367–378, 2018.

[79] Junfeng Xie, F Richard Yu, Tao Huang, Renchao Xie, Jiang Liu, Chenmeng Wang, and Yunjie Liu. A survey of machine learning techniques applied to software defined networking (sdn): Research issues and challenges. *IEEE Communications Surveys & Tutorials*, 21(1):393–430, 2018.

[80] Tatsuya Yamazaki. Quality of experience (qoe) studies: Present state and future prospect. *IEICE Transactions on Communications*, page 2020CQI0003, 2021.

[81] Jiachen Yang, Chunqi Ji, Bin Jiang, Wen Lu, and Qinggang Meng. No reference quality assessment of stereo video based on saliency and sparsity. *IEEE Transactions on Broadcasting*, 64(2):341–353, 2018.

[82] Mingzhe Yang, Shangguang Wang, Rodrigo N Calheiros, and Fangchun Yang. Survey on qoe assessment approach for network service. *Ieee Access*, 6:48374– 48390, 2018.

[83] Matt Yu, Haricharan Lakshman, and Bernd Girod. A framework to evaluate omnidirectional video coding schemes. In *2015 IEEE international symposium on mixed and augmented reality*, pages 31–36. IEEE, 2015.

[84] Anatoliy Zabrovskiy, Evgeny Kuzmin, Evgeny Petrov, Christian Timmerer, and Christopher Mueller. Advise: Adaptive video streaming evaluation framework for the automated testing of media players. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, pages 217–220, 2017.

[85] Saman Zadtootaghaj, Nabajeet Barman, Rakesh Rao Ramachandra Rao, Steve Göring, Maria G Martini, Alexander Raake, and Sebastian Möller. Demi: Deep video quality estimation model using perceptual video quality dimensions. In *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2020.

[86] Tiesong Zhao, Qian Liu, and Chang Wen Chen. Qoe in video transmission: A user experience-driven strategy. *IEEE Communications Surveys & Tutorials*, 19(1):285–302, 2016.

[87] YI ZHU. On predicting individual video viewing experience.

[88] Yi Zhu. *On predicting individual video viewing experience: The value of user information*. PhD thesis, Delft University of Technology, 2019.

[89] Yi Zhu, Ingrid Heynderickx, and Judith A Redi. Understanding the role of social context and user factors in video quality of experience. *Computers in Human Behavior*, 49:412–426, 2015.