Received August 5, 2019, accepted August 18, 2019, date of publication August 21, 2019, date of current version September 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936470

Why Is Multimedia Quality of Experience Assessment a Challenging Problem?

ZAHID AKHTAR[1], (Senior Member, IEEE), KAMRAN SIDDIQUE[2], (Member, IEEE), AJITA RATTANI[3], (Member, IEEE), SYAHEERAH LEBAI LUTFI[4], AND TIAGO H. FALK[5], (Senior Member, IEEE)

[1]Department of Computer Science, University of Memphis, Memphis, TN 38152, USA
[2]Department of Information and Communication Technology, Xiamen University Malaysia, Sepang 43900, Malaysia
[3]Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS 67260, USA
[4]School of Computer Science, Universiti Sains Malaysia, Penang 11800, Malaysia
[5]INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada

Corresponding authors: Zahid Akhtar (zmomin@memphis.edu) and Kamran Siddique (kamran.siddique@xmu.edu.my)

This work was supported by the Office of Research and Innovation, Xiamen University Malaysia under the XMUM Research Program Cycle 3, under Grant XMUMRF/2019-C3/IECE/0006.

ABSTRACT Quality of experience (QoE) assessment occupies a key role in various multimedia networks and applications. Recently, large efforts have been devoted to devise objective QoE metrics that correlate with perceived subjective measurements. Despite recent progress, limited success has been attained. In this paper, we provide some insights on why QoE assessment is so difficult by presenting few major issues as well as a general summary of quality/QoE formation and conception including human auditory and vision systems. Also, potential future research directions are described to discern the path forward. This is an academic and perspective article, which is hoped to complement existing studies and prompt interdisciplinary research.

INDEX TERMS Audiovisual perception, data-driven analysis, multimedia quality, objective quality metric, quality of experience, quality of services (QoS).

I. INTRODUCTION

Multimedia services (e.g., Internet Protocol TeleVision, social networks, immersive and virtual reality games, multimedia conferencing, etc.) not only have gained importance in this wireless era but also are expected to grow exponentially. Such services and applications may provide poor experience to a user due to multiple reasons, e.g., network (network overload) and terminal devices known as *quality elements* [1]. Thus, research and industry communities are devising different schemes to understand how users perceive and experience degradations, and to subsequently provide them with a better quality of experience (QoE) [64], [65].

Majority of the multimedia applications are comprised of two main elements, i.e., audio and video signals. In spite of latest technical advances, core elements of multimedia are still impaired usually by lossy signal encoding (decoding) and/or fallible transmission channels, which may lead to poor quality and user experiences [2], [3]. Accurately estimating QoE of the communicated audio-visual modalities will help service providers to further ameliorate their networks and services. In fact, digital signals' QoE assessment is becoming a paramount and demanding issue in multimedia signal processing and practices.

Majority of existing studies on QoE are based on quality assessment (QA) of audio, video or audio-visual modalities that may deteriorate under different processes such as acquisition, preprocessing, transmission, etc. There exist subjective and objective QA techniques. In subjective QA schemes, a bunch of trained/naïve end-users are asked to provide quality grades [4]–[6]. Whereas, signal fidelity measures (e.g., packet loss proportions) are utilized in objective methods to estimate quality. Though objective schemes are fast and simple, they are poor indicator of perceived QoE owing to the fact that they overlook human's multi-media perception as well as viewing circumstances [7]. It has been widely agreed upon that signal fidelity measures alone do not register well with human QoE gradings of multimedia contents [7]–[11]. The QoE decided by humans' perception and contentment is much more compounded than the statistics that may be obtained using a typical signal QA method and network management system. Since the human auditory and visual systems are the ultimate receivers and the human evaluator observes the transmitted multimedia contents, user-oriented perceptual and cognitive correlates of QoE perception are needed that is still a long way odyssey owing to various

The associate editor coordinating the review of this article and approving it for publication was Jinsong Wu.

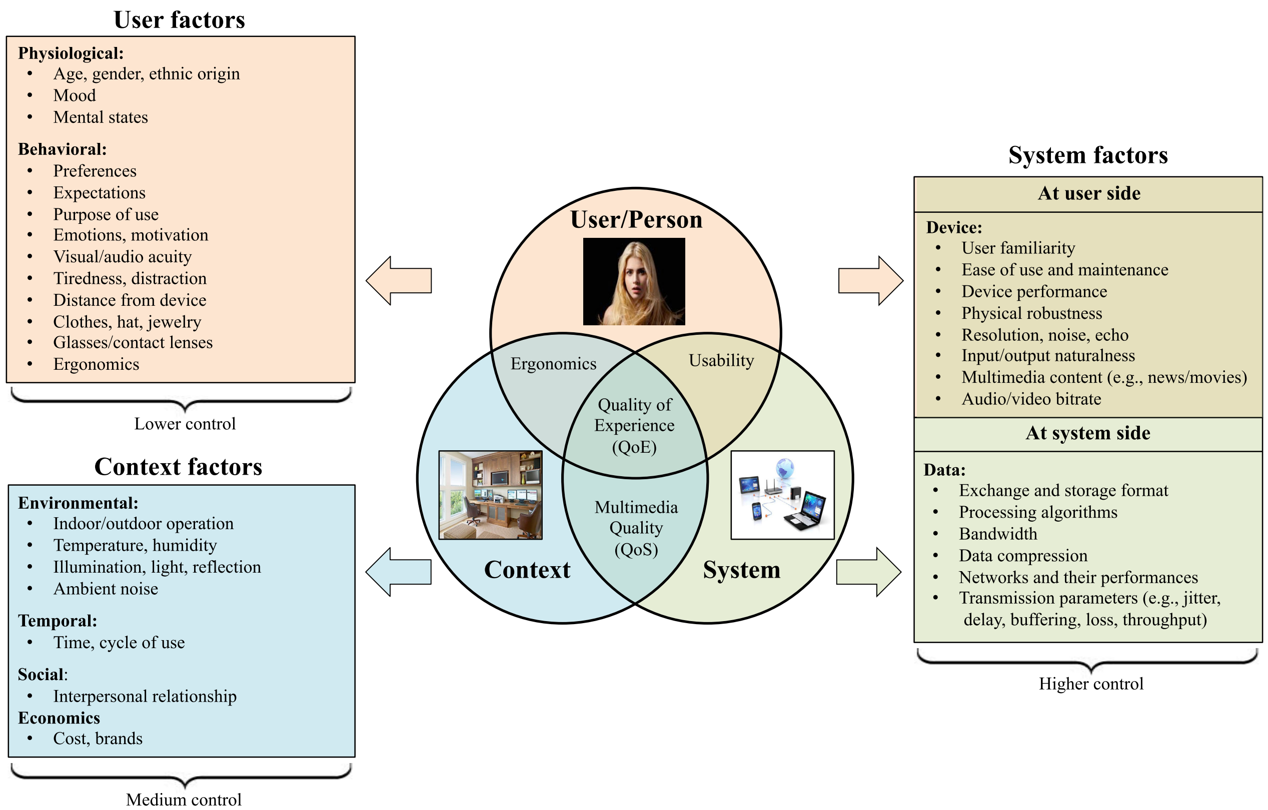


FIGURE 1. Illustration of the different factors furnishing user's QoS and QoE.

covariates such as user-, context-, and system factors as also depicted in Fig. 1. Nonetheless, current trends in user-centric experiences and burgeoning of wearables have allowed such models to be developed [12] and opening door to future innovations.

QoE is a multidisciplinary domain that needs input from human science (e.g., psychology, physiology, sociology, neuroscience and cognitive science) and information technology (e.g., digital signal processing, machine learning, computer vision and signal synthesis). There are many issues still open in the QoE field. Therefore, this article first presents an synopsis of various interpretations of quality, QoE formation and perception including human auditory and vision systems, quality of service (QoS), QoE, quality of perception (QoP), and main issues and research opportunities in this field. All in all, it is an academic and position paper that is expected to complement prior studies and prod interdisciplinary research.

II. WHAT ARE QUALITY, QUALITY OF SERVICE, QUALITY OF PERCEPTION AND QUALITY OF EXPERIENCE?

A. WHAT IS QUALITY?

Quality is a metaphysical notion and viewed as a construct of the mind that is simple to comprehend but hard to expound. According to Merriam Webster dictionary, the term quality comes from Latin word qualitas that refers to 'an inherent feature' or 'a property' or 'a degree of excellence' [13], which nowadays may be called ''character'' of an entity [12]. While few philosophers state that quality could not be elucidated, the notion of quality remains controversial in modern philosophy. In multimedia domain, quality is usually utilized with the engineering objectives, since it is a dominant gauge for assessing systems, applications or services throughout both development and operational stages [12]. All in all, quality could be contemplated as an umbrella term, as many elements play a part in constructing quality's cognizance. For example, covariates like loudness, tone, listening effort, and intelligibility are critical for audio quality. While, factors like image size, packet loss, frame rate, level of audio-visual synchronization are pivotal for visual or audiovisual quality. The concepts related to quality have been slowly evolving and continuously refined leading to several definitions, which contrast moderately with one another subject to usage's context [14]. It is imperative to revisit common expressions from this vantage point, since quality has varying overtones determined by the context where it is utilized and some meanings of those concepts have been progressively lost and deserve clarification. Therefore, in the following various interpretations of quality throughout literature are presented:

Definition 1 (Dictionary/Lexical View): ''Measure of excellence or state of being free from defects, deficiencies, and significant variations'' [15].

Definition 2 (Transcendental/Metaphysical View): This view reckons quality to be synonymous with innate excellence, which can be recognized but not defined. The transcendental view of quality is described as *"… even though Quality cannot be defined, you know what it is. … Quality is neither a part of mind, nor is it a part of matter. It is a third entity, which is independent of the two. … Quality isn't a substance. Neither is it a method. It's outside of both. … It's the goal toward which method is aimed"* [16]. But, suchlike perspective of quality has small empirical use, since it cannot be computed by professionals and can solely be discerned by virtue of experience.

Definition 3 (ISO 8402-1986 View): According to ISO (International Organization for Standardization) 8402-1986 standard view, quality is defined as *"the totality of features and characteristics of a product or service that bears its ability to satisfy stated or implied needs"* [17].

Definition 4 (Product-Based/ISO-9004 View): This view defines quality as quantifiable and measurable characteristics or attributes, and presume that an artifact, which has fine interior attributes has also fine exterior attributes. For instance, ISO-9004 describes quality as *"fitness for use, performance, safety and dependability. Sometimes this view is called quality of design"* [18]. However, disregarding a specific user's preferences and considering certain attribute's presence/absence entails elevated quality are main limitations of this strategy. Another definition of quality under this view is *"Quality refers to the amounts of the unpriced attributes contained in each unit of the priced attribute"* [19].

Definition 5 (Manufacturing View): It contemplates quality as congruence to prerequisites stipulations in which prerequisites are specified mainly in scientific, goal and quantifiable terms. Dale *et al.* stated quality as *"Strict and consistent adherence to measurable and verifiable standards to achieve uniformity of output that satisfies specific customer or user requirements"* [20]. While, Gilmore *et al.* described it as *"the degree to which a specific product conforms to a design or specification"* [21].

Definition 6 (Value-Based View): Under this view, quality is considered as *"the degree of excellence at an acceptable cost* [22]. It uses a trade-off-model between quality and cost with the aim to provide the quality as per customer's willingness to bear the cost. However, this approach has great difficulty in combining two linked yet discrete notions, i.e., measure of excellence (i.e., quality) and value (measure of worth). Though, the result may be hybrid and termed as 'affordable excellence', it lacks well-defined limits [23].

Definition 7 (Objective/Quantitative View): "Measurable and verifiable aspect of a thing or phenomenon, expressed in numbers or quantities, such as lightness or heaviness, thickness or thinness, softness or hardness" [24].

Definition 8 (Subjective/Qualitative/Perceptual/User-Based View): Here, quality is specified as fitness for objective, and considered as an independent matter and objects that best gratify preferences of user to have the biggest quality. In other words, quality is hinged on particular product/object's features like ergonomics, efficiency, dependability and that are assessed through user's perspective, thereby equating customer/user satisfaction with the quality. According to it, quality is *"Attribute, characteristic, or property of a thing or phenomenon that can be observed and interpreted, and may be approximated (quantified) but cannot be measured, such as beauty, feel, flavor, taste"* [24].

Definition 9 (ISO 9000 View): ISO 9000 is a group of quality management techniques standards, which interprets quality as *"something that can be determined by comparing a set of inherent characteristics with a set of requirements of customers and other interested parties* [25]. The quality is considered high if attributes meet prerequisites, while it's low if attributes do not satisfy all prerequisites.

Definition 10 (Experiencing View): This view defines quality to explicitly inscribe the perception involving sensory processing of exterior stimuli. In [1] author defines quality (depending on experiencing) as *"judgment of the perceived composition of an entity with respect to its desired composition"*.

The discerned composition denominates 'the entirety of an entity's features', while a feature is interpreted as an 'an entity's identifiable and mentionable attribute'. The preferred composition depicts 'the features' entirety of a person's anticipations and/or pertinent requirements and/or social demands' [14].

Definition 11 (Assumption View): The presumed quality looms up as *"quality and quality features that users, developers, manufacturers or service providers assume regarding a system, service or product that they intend to be using, or will be producing, without however grounding these assumptions on an explicit assessment of quality based on experiencing"* [12]. The underlying assumptions/expectations may be located at a distinct rank of cognitive/perceptual framework than actual emotional and sensory references.

Definition 12 (QUALINET View): According to QUALINET view, *"quality is the upshot of a person's comparison and adjudication procedure, which incorporates perception, contemplation about the conception, and the narrative of the upshot"* [26].

Counter to concepts/definitions where quality is discerned as "qualitas" (i.e., a set of intrinsic attributes), QUALINET contemplates quality with regard to gauged goodness or distinction, degree of requirement's fulfillment, and a "quality event", wherein event is a noticeable happening and ascertained in space (i.e., where it happens), character (i.e., what could be seen) and time (i.e., when it happens) [27].

From the above-enlisted definitions, one can infer several key aspects of quality: i) it is usually decided by user/consumer of a service according to context of usage, ii) it is a result of a correlative evaluation of a multifaceted entity, namely it comes via estimation of the discerned features (attributes) to expected characteristics, iii) contemporary institutionalized definitions either from standards or organizations include 'perception' as a pivotal parameter for quality judgment process, where perception is interpreted as the intentional processing of sensorial data.

Quality, essentially speaking, is the consequence of a human discernment depending on different criteria. Few of them are relied on quantifiable inherent particulars of the signal, whereas others are built on cognitive procedure thereby typically unmeasurable. Overall, definitions 7, 8 and 12 (i.e., objective view, subjective view and QUALINET view, respectively) are mostly adopted in multimedia services and research [5], [7].

With the exponential growth and consumption of multimedia applications and services, it becomes ever more paramount for service providers to deliver an optimal user experience. The very common assumption is that high performance and transmission quality leads to high acceptance (larger number of users) of the systems, services and applications. Namely, low-quality services result in low acceptance, which is not always true, e.g., the early days of YouTube that regardless of low quality precipitated in a gigantic success. It means that the correlation among quality, usability and acceptance of any multimedia system/services/application is yet poorly understood. Therefore, industry and research communities have lately moved in the direction of embracing the end-subject as the topmost significant component in multimedia quality evaluation to procure wider facets, e.g., QoE or QoP instead of only QoS. The elementary notions of QoS, QoP and QoE are overviewed in this section.

B. QUALITY OF SERVICE (QoS)

QoS is utilized as a measure of multimedia networks and applications performance. The term QoS was first defined by the ITU within the field of telephony as ''The totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service''. The Internet Engineering Task Force (IETF) defined QoS based on a network-centric view as ''A set of service requirements to be met by the network while transporting a flow''. These two definitions not only have very narrow scopes but also lack comprehensible explanation of user characteristics and context of usage, thereby leading to a gap between definitions and their common usage. The QoS standard definition is based on quantifiable networks performance factors including delivery platforms is ''a set of networking technologies and quantification instruments, which permit for the network to undertake bringing predictable outcomes [12]''. The expression QoS describes the concepts, quantification of network performance (e.g., delay, jitter) and techniques like Integrated Services.

The features like performance, responsiveness, reliable, availability, adaptivity, application and security aspects comprise the QoS. This section presents the QoS taxonomy, performance and impacting factors aspects. QoS is further divided into user, application, and resource, based on a multimedia end-to-end architecture.

User-Layer: A user-layer QoS specification enables the user to specify the QoS requirements such as frame sampling rate, cost, resolution, and security, even using a graphical user interface at an abstract level. The user can also provide QoS parameters such as resolution, multimedia content detail, video color accuracy, video/audio synchronization, audio smoothness and an overall rating.

Application-Layer: The QoS requirements specified by the user is translated and mapped to parameters at lower layer known as application-layer. This layer is hardware and platform independent which means that no assumption is made regarding network conditions and operating systems. The type of features used at the application-layer can be classified into: quantitative parameters for performance assessment for instance, behavior-specific qualitative and resolution parameters, i.e., service management in the event of network bandwidth problem. A guidelines language is utilized for coding the features obtained, and to give definite views to system architect.

Application-layer QoS specification languages can be classified into seven groups based on their paradigm. First, script-based paradigm: it is quite easy to specify requirements in high level using script languages due to their abstractness. Authors in [51] devised a method to add QoS specification into Windows NT applications without altering either the operating system or its applications. Second, parameter-based paradigm: it utilizes QoS management architecture to act on the qualitative/ quantitative parameters defined in the form of data structures. In this category, two parties in communication determine a service contract that comprises various aspects, such as flow specification, QoS adaptation, QoS commitment, QoS maintenance, reservation style, and cost. Third, process-oriented paradigm: a process (i.e., unit of execution) is employed to link QoS for synchronizing, negotiating and communicating end ports. The QuAL (Quality-of-service Assurance Language) [12] is a well-known example of process-based QoS specification language. Fourth, control-based logic approach: this category is adopted to control adaptive QoS policies and flow in adaptive systems, e.g., use of Proportional-Integral-Derivative controller to restraint fine granularity in tasks. Fifth, markup-based approach: here, markup language documents structured information containing both matter (e.g., words) and their role explanation, e.g., Extensible Markup Language (XML). Sixth, aspect-oriented approach: aspects are attributes that influence performance or components' semantics are coded using aspect-Oriented Programming (AOP) paradigm. Seventh, object-oriented approach: the object-oriented concept (as in traditional languages) helps specification refinement, i.e., specification reusability into the designs. For instance, QML (QoS Modeling Language) was developed at HP Laboratories for specification features refinement purpose.

Resource-Layer: The QoS requirements specified at the abstract layer are rendered into extra solid resource demands, i.e., physical resources narration required for the application covering their transport protocols, allocation and mechanism.

The resource-layer descriptions could be categorized into fine and coarse granularity. At fine granularity level,

TABLE 1. Synopsis of QoS characteristics for the user layer, application layer, and resource layer.

QoS Layers	Parameter	Example
User Layer (subjective criteria)	Perceptive media quality Window size Pricing model Range of price	Bad, fair, good, excellent Small, medium, big Per transmitted byte charge, flat rate Low, medium, high
Application Layer (platform- and hardware-independent)	Quantitative issues Qualitative issues Adaptation rules	Audio/image resolution, video frame rate Inter/intra-stream synchronization strategies Drop all B frames if video quality is good
Resource Layer (platform- and hardware-dependent)	Quantitative issues Qualitative issues Adaptation rules	Delay, output, delay jitter, resource demands timing, memory size Reservation technique, OS scheduling, loss recovery/detection procedures Drop x video frames if 80ms<skew<160ms

resources' detailed description is required, including specific description of qualitative and quantitative QoS demands, adaptation rules and allocation time. Coarse granularity includes meta-level guideline, for instance only specifying resource requirements without allocating time or resource instances. Table 1 summarizes QoS layers with their QoS parameters. The QoS performance can be evaluated at the quality formation process both at the system/structure and the subject-side.

The performance at the system-level could be measured in terms of precision (i.e., performance of biometrics/emotion/behavior classifiers), modality suitableness (i.e., theoretical acquaintance with properties of the trait and its applicability to the domain), elucidation precision (i.e., performance of semantic notions), dialogue management precision (i.e., interchange success ratio), contextual aptness (i.e., quantification of Grice's Cooperativity Principle), appropriateness of outcome modality (i.e., interrelations between traits) and suitability of form (i.e., the outcome supplied to the user measurable via factors such as intelligibility and comprehensibility). The interaction performance at the user side could be measured as to efforts (i.e., cognitive, physical and perceptual) demanded by the subject and its liberty of interaction. The taxonomy of QoS, as defined in [28], is illustrated in the top diagram of Fig. 2.

C. QUALITY OF PERCEPTION (QoP)

QoS and QoE describe the technical quality of the system as well as degree of user's satisfaction based on their expectation. However, these measures do not consider the fidelity and utility aspects of a user. For addressing the limitation, authors in [54] suggested the idea of Quality of Perception (QoP). It is described as ''it is an expression that comprises a subject's contentment with the grade of multimedia renderings and their capacity to probe, synthesize and comprehend the informational component of multimedia presentations''.

The multimedia quality using objective or subjective components is not adequate due to big dimensional nature of multimedia ingredients. QoP unites both subjective estimation of the contentment level along the level of multimedia dispositions (indicated by QoP-S), and objective estimation of subject's aptitude to probe, symphonize and acquire the information element of multimedia content (indicated as QoP–IA).

QoP-S comprises of QoP–LOE defined as level of enjoyment experienced by the user in the multimedia constituent, and QoP–LOQ defined as subject's decree regarding the objective grade of quality allotted to the multimedia component being encountered. Peculiarly, QoP–IA is quantifiable in percentage as a measure of level of knowledge acquired by the user from the multimedia ingredient. QoP–LOE and QoE–LOQ are procured by subjects' classical rating methods. Impact of the multimedia presentation frame rates on subject's QoP and subject's eye paths was studied in [55]. The obtained outcomes suggested that aloft frame rates do not necessarily yield elevated QoP or user knowledge acquired. However, the net subject amusement and quality conception is enhanced. Impact of demonstrating videos at changing bandwidths and frame rates on subject's QoP phrased as 'user watchability' was analyzed in [56]. Authors explicitly mentioned that video content and fidelity remarkably impacts user's QoP.

D. QUALITY OF EXPERIENCE (QoE)

QoS mainly focuses on end-to-end system performances, and does not ascertain human user contentment in terms of cause-and-effect relationship. User perception and satisfaction, however, are designed by variety other facets that may not be surely be managed by the service components performance. Quality of Experience (QoE) has lately been stated as the subject's conception of the satisfaction, usability and acceptability of the utility [48]. QoE proceed beyond traditional end-to-end QoS parameters by covering a lot of divergent facets (e.g., subject's mental state) to ameliorate the quality underwent by the subject. QoE can be stated as the perceptive QoS from view of the subjects as shown in Fig. 1. The ITU Standardization Sector (ITU-T) defined QoE as ''the overall acceptability of an application or service, as discerned subjectively by the subject'. The Qualinet White paper [26] defined QoE in context of the enjoyment and utility related to application or service considering the user's current state and personality. However, this definition is limited in the sense that it did not include system aspects but only application and service aspects. Moreover, it used the term ''user'' for the end client, which sometimes is also referred for system utility. Thus, in [12] an updated QoE is described for a global view as ''the grade of glee or nuisance of a user whose experiencing includes a service, system or application.

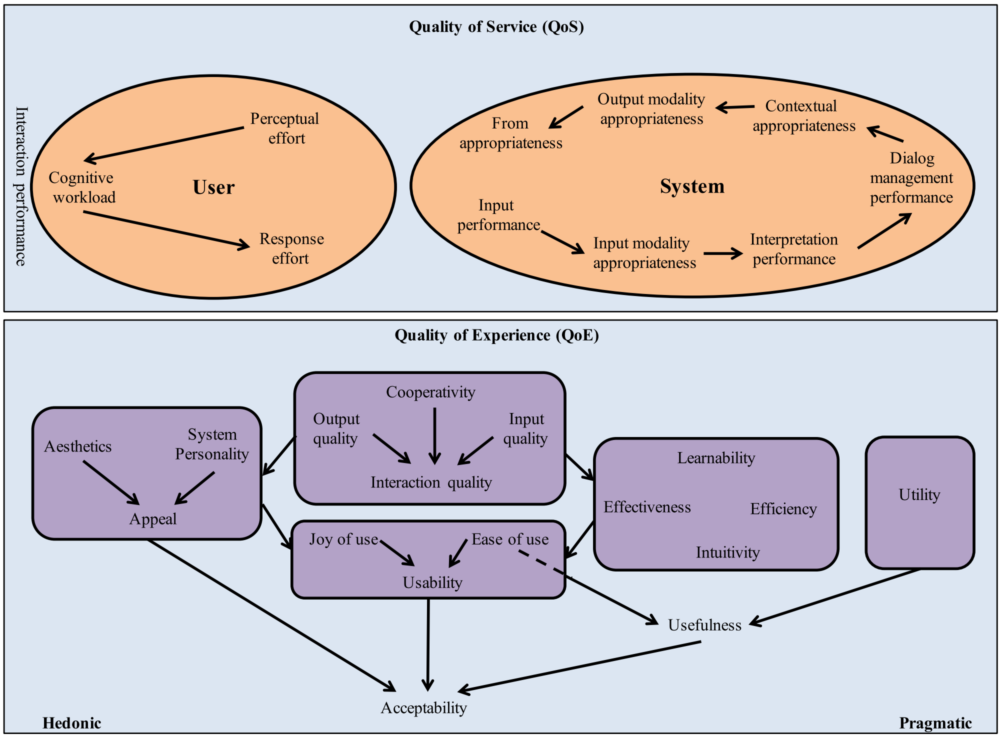


FIGURE 2. Taxonomy of QoS and QoE aspects of multimodal human-machine interaction [28].

It outcomes via the user's assessment of the accomplishment of their expectancies and requirements regarding the use and/or pleasure concerning the subject's personality, current state and context''. Whereas, the International Telecommunication Union in Recommendation ITU-T P.10 has defined QoE as ''the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state.''

The quality perception of a multimedia sample can vary with respect to subject, the content producer, or the service provider. The meaningful prediction of QoE can be realized through QoE assessment algorithms/metrics that are consistent for any type of impairments and end-users. To this end, recently introduced QoE assessment methods are utilizing information related to both signal degradations and human quality perception process information. This section details the components that may affect perceived QoE. Further, audio and visual determinants that are typically used in automatic QoE analysis are briefly studied.

The fundamental goal of a system designer and service provider is to fulfill human user's needs. The authors in [52] argued that user's QoE assessment is impacted by pragmatic and hedonic quality aspects, which should be judged by real or test subjects to attain perceived ratings. The desirable QoE aspects are interaction quality (i.e., perceived input and output qualities including cooperativity to understand respectively input comfort level, appropriateness and system's support in reaching a common aim beyond perceived quality), usability (i.e., ease and joy of use that incorporates both pragmatic and hedonic qualities, e.g., hedonic aspects like the appeal of the system), utility and usefulness (i.e., a comparison between user's functional requirements and functions offered by the system), efficiency-related aspects (i.e., effectiveness/accuracy/efficiency of the system and learnability with intuition (intuitively) for user), system's aesthetics, personality and appeal (i.e., sensory experience elicited by the system, users' perception about system characteristics, and system's ability to inherit user's interest, respectively), and acceptability (i.e., number of potential readily users as an economic measure). The taxonomy of QoE, as defined in [28], can be seen in the bottom diagram of Fig. 2.

QoE is decided by cognitive and psychological determinants such as feelings, habits, requirements and expectations. Quantified QoE can be obtained by coding system's performance with user's perception in the shape of interpretable and statistical values, which can be acquired via 'direct QoE computations rating by actual users also named subjective QoE or 'indirect QoE computations by recording subject's conduct and associating it with discerned QoE (known also as objective QoE) [116], [119]. In fact, several efforts for quantifying and assessing the QoE in multimedia networks and multimedia content consumption have been placed. For instance, Piamrat *et al.* [64] devised a QoE assessing metric for video streaming application in wireless networks that is useful in various network situations with different loss rate and distributions. In reality, various studies concerning QoE for wireless communications have been performed [131]–[133]. For example, a subjective QoE assessment for VoIP applications in a real wireless environment was carried out in [134]. While, authors in [135], [136], and [137], respectively, proposed a mapping scheme from SVC layers to DASH layers that can provide improved QoE, a QoE model for VR video transmission over wireless network through the relations between subjective QoE and latitude, and an analytical model to estimate the QoE for encrypted YouTube Live service from packet-level data collected in the interfaces of a wireless network.

Authors in [64] presented a hybrid approach called Pseudo Subjective Quality Assessment (PSQA) using random neural network (RNN) that helps to keep benefits of both objective and subjective methods. The reported results showed that PSQA not only outperforms Peak Signal to Noise Ratio (PSNR) metric but also yields outcome like subjective test. Bentaleb *et al.* [121] and Huang *et al.* [122] proposed a bandwidth broker solution for any type of content delivered to any kind of consumer device via HTTP adaptive streaming, and a pure buffer-based DASH scheme to optimize user QoE, respectively. While, Pouli *et al.* [65] developed a multimodal content retrieval scheme to provide enhanced QoE. The proposed framework uses personalization and relevance feedback techniques for retrieving and offering multimedia content tailored to specific user's preferences and/or attributes. Considering more real-world situations, Tsiropoulou *et al.* [66] treated the museum environment as a cyber-physical social system, namely a distributed non-cooperative game to maximize visitors QoE. For visitor's Recommendation Selection and Visiting Time Management (RSVTM), authors proposed a two-stage distributed scheme using game theory and reinforcement learning that learns from visitor behavior to make on-the-fly recommendation selections. Similarly, Wamser *et al.* [67] designed a model for real-time YouTube QoE estimation based on the buffer level at the YouTube application layer. The proposed model can be employed by operators to evaluate the performance of networks supplying YouTube videos. Also, works have focused on immersive QoE, e.g., Doumanoglou *et al.* [68] studied neural networks based immersive QoE prediction of real-time 3D media content streamed to VR headsets for entertainment purposes. Other studies tried to recognize the interrelation between QoE and QoS. For example, authors in [138] presented the equation 1 to capture the exponential relation between the QoE and QoS parameters such as function of loss and reordering ratio caused by jitter:

$$\frac{\partial QoE}{\partial QoS} \sim (QoE - \gamma). \tag{1}$$

Similarly, Shaikh *et al.* [139] presented linear relationship between the QoE and QoS parameters (e.g., bandwidth, throughput and delay) as:

$$\log(QoE) = a_0 + a_1 QoS_1 + a_2 QoS_2 + \ldots + a_n QoS_n. \tag{2}$$

Later, applying an exponential transformation on (Eq. 2) yielded QoE/QoS exponential correlation as:

$$QoE = e^{a_0} + e^{a_1 QoS_1 + a_2 QoS_2 + \ldots + a_n QoS_n}, \tag{3}$$

Whereas, the non-linear relationship between QoE and QoS parameters was defined by Alberti *et al.* in [140] as:

$$QoE = \sum_{i=0}^{N-1} a_i QoS_i^{k_i}, \tag{4}$$

where a_i are the constants and k_i are the exponents for N parameters.

Besides, for indirect QoE computations, utilizing physiological estimates have been lately probed in various studies [53]. The approaches related to cognitive centric group utilize neurophysiological insight to obtain perceived QoE via human torso field sensors and networks. The techniques includes magnetoencephalography (MEG), electroencephalography (EEG), near-infrared spectroscopy (NIRS), and functional magnetic resonance imaging (fMRI).

QoE can be determined both at the service provider and the client side. QoS and QoE depict factors associated with acceptableness and user's sentiment levels, a service or application, respectively. Comprehending human quality conception procedures helps to perceive impression created in the user's mind. Next subsection discusses the human perception procedure.

III. QoE FORMATION PROCESS

A vital aim in designing an audio-visual multimedia transmission, coding, decoding and display technique is to yield video and audio signals of a grade acceptable to the end-user. The genesis of QoE highly reply on human perception procedure [48]. Number of studies and theories exist that try to delineate the perception of physical events by humans via their sensory system [29]–[31]. In other words, understanding the mechanism involved in human users view, interpretation and response to audio and visual stimuli would assist to devise the design doctrines underlying video and audio encoding, decoding, displaying as well as methods for the perceived quality evaluation.

Human QoE perception is elucidated as a combination of sentient sensory experience as well as procedure consisting of higher-level and low-level sensory cognitive operation [30]. The physical stimulus, for example, a noise wave for an auditory signal, are transformed into electric signals by low-level sensorial technique in the nervous system. On the other hand, high-level cognitive processing involves the conscious processing involves elucidation and comprehension of the neural waves that build a discerned quality discernment. The quality discernment although originate via the neuronal processing of a corporeal stimulus, it also is impacted via contextual particulars, i.e., mental states, physical environment of the user (e.g., attitude, emotions, goals, mood, intentions), other modalities, and prior knowledge/experience.

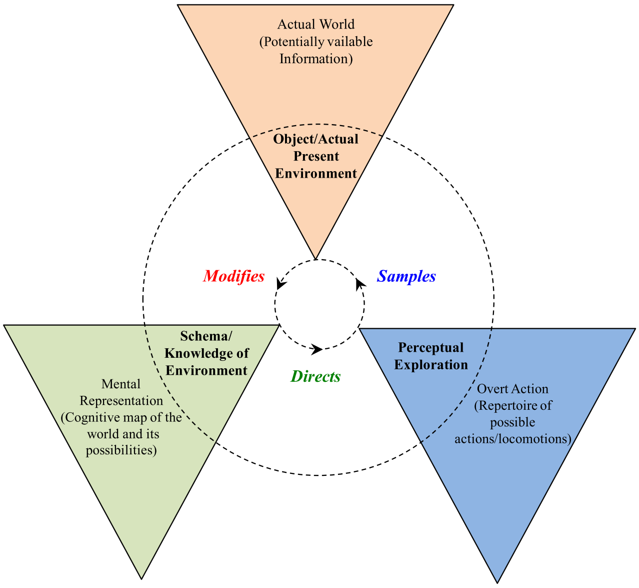


FIGURE 3. **Perceptual model cycle embedded in cognitive map.**

A model to explain the influence of knowledge on perception was proposed by Neisser [32], which is called the Perceptual Cycle. The Perceptual Cycle embedded in cognitive map is shown in Fig. 3. The model is composed of three concepts namely, the actual environment (e.g., object or stimulus), the knowledge of the actual environment (also referred as 'schema') and the perceptual exploration. The schema directs the perceptual exploration (i.e., analyzing stimuli and the environment), which involve short-term memory and the focus of attention. The perception is a continuous procedure during which the schema is continuously updated based on new information and modality dominance (relevance, resolution, reliability).

The operation of human QoE perception, vision and auditory systems is a large and complex field of study. Some of their important features are discussed below.

1) HUMAN VISUAL SYSTEM (HVS)

By far vision is the most important sense among human senses. In fact, 80-90% of all neurons in the human brain are approximated to be entailed in visual perception [33]. The human visual system can be subdivided into two parts: the eyes and the visual pathways in the brain. The eyes capture light and transform it into signals to be understood by the nervous system; while the visual pathways transmit these signals to the brain for processing. The human eye and the visual pathways cross section are shown in Figs. 4a and 4b, respectively.

From an optical point of view, the eye is analogous to a photographic camera, which is made of a system of lenses and a variable aperture to focus images on the light-sensitive retina. Vision starts when light enters the eye via the cornea, which is a transparent protective layer and main refractive surface of the eye with refractive index of 1.38 [34]. The light then passes through the pupil (i.e., the hole in the center

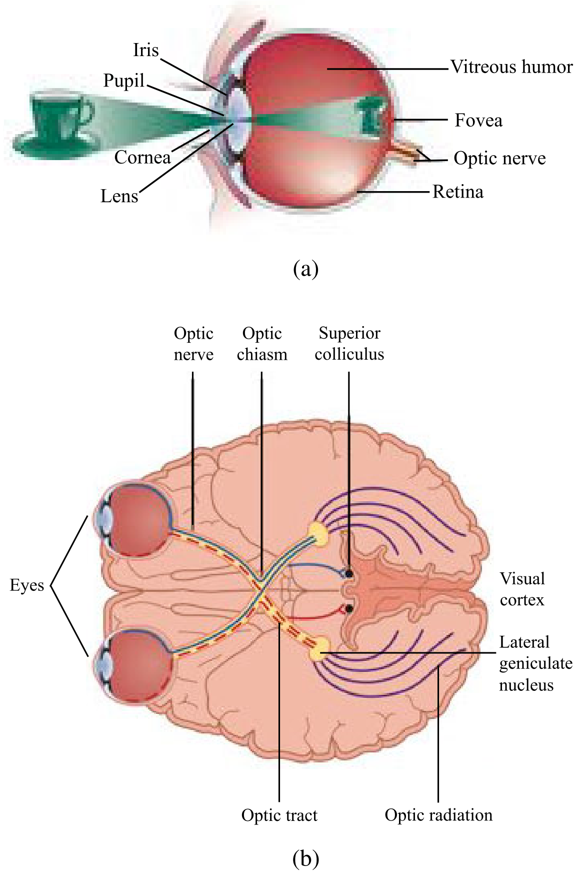


FIGURE 4. **(a) The human eye transverse section. Also, an image of a cup is focused on the retina that lines the back of the eye. (b) Visual pathways in the human brain transverse section [38].**

of the eye with diameter ranging from about 3 to 7 mm) to be focused onto the retina by lens (of refractive index 1.40). The lens changes its shape with 'accommodation' to focus the image. The pupil diameter changes based on the prevailing light levels, and the iris (i.e., the circular pigmented that surrounds the pupil and gives our eyes their characteristic color) controls aperture of the lens and thereby the amount of light entering into the eye.

The central body of the eyeball is encompassed of gelatinous vitreous humor (refractive index 1.34) that maintains the eye's shape. The retina is composed of an array of cones (photoreceptors sensitive to color at high light levels) and rods (photoreceptors sensitive to luminance at low light levels) cells that convert light energy into electrochemical signals, which are then transmitted to and interpreted by the brain. The names rod and cone are given because of their physical shape. In the human eye, there are approximately 125 million rods and 6 million cones [35], [36]. The rods are distributed throughout the retina, while cones are mostly concentrated in the center of the retina (approximately area of 0.3 mm), also known as the fovea. The cones are more sensitive and responsible for high visual acuity and color vision. The total optical power of the eye is around 60 diopters. The nerves connecting to the retina exit the eyeball via the optic nerve to the brain.

The visual information from the eyeball is delivered to various processing centers in the brain through the retinal ganglion cells' axons in the optic nerve, as also shown in Fig. 4b. A partial crossover of optic nerves projecting from each eye happens at the optic chiasm to rearrange the fibers. Each retina's fibers from the nasal halves cross to the opposite side to join the other retina's fibers from the temporal halves to form the optic tracts. The left and right visual fields are processed in the right and left hemisphere, respectively, owing to the inverted retinal images by the optics. Most of the signals from each optic tract synapse in the lateral geniculate nucleus (LGN) in the thalamus. From LGN, fibers pass by the way of the optic radiation to the visual cortex. The visual information receiving area is also known as the striate cortex due to the white stripes that are created within this area of cortex by nerve fibers that run through it [37]. Throughout the visual pathways, the neighborhood relations of the retina are preserved. Namely, the input from certain parts of the retina is processed in a specific part of the LGN and of the primary visual cortex. The human cortex comprises several areas that are involved in responding to visual stimuli (approximately covering 950 cm^2 or 27% of the cortical surface [38]) and processing different modes of vision (e.g., shape, location, color). For example, neurons in the primary visual cortex show responses to particular sizes and orientations [39]. Fig. 4b also depicts the location of the superior colliculus; it is the area that governs eye movements and other visual behaviors, which acquires about 10% of the fibers from the optic nerve [38]. All in all, the human brain processes and interprets visual information partially based on the image detected by the retina and partially using prior knowledge (e.g., shape of the object). We refer the interested reader to [40] for more details.

Visual perception is defined as the capacity of interpreting the neighboring environment by what we see. Owing to the complexity involved, numerous theories related to the relation between various visual psychological occurrences are in hypothesis phase. However, various studies suggest that radiance non-linearity, masking effects, multi-channel visual and attention, and contrast reactivity are the important building blocks of visual perception [41]–[43]. The luminance non-linearity study in [42] found that human eyes have poor and strong judgment, respectively, for the absolute brightness of the observed item and the relative differences in brightness discretion. It was also observed that there is a linear relationship between the logarithm of the perceived brightness and the objective brightness under specific range. Contrast sensitivity refers to the capability of the HVS to distinguish the differences in intensity. The HVS response depends less on the absolute luminance and more on the relation of its variations in the local surrounding luminance. According to the study on Campbell-Robson contrast sensitivity function map [43], contrast sensitive function (CSF) could be viewed as spatio-temporal band-pass filtering process. Masking effect occurs when a stimulus, which is visible by itself, may not be detected because of the presence of another. Several studies on multi-channel parallel theory [44] imply that different visual information constituents are processed via independent neural channels before entering the visual cortex. For instance, neurons in primary visual cortex are sensitive to the specific orientation stimuli. Visual attention refers to the cognitive operation involved in the selection of relevant visual information by filtering out the irrelevant one. Existing theories related to visual attention may be categorized in space where attention is aimed at discrete areas of space inside visual field of view, and object-based where attention is aimed at the article instead of its position. Through a psychology viewpoint, visual attention could be bottom-up strikingness that is affected by low-level attributes of the target or top-down protruding that is affected by user's cognitive processing.

2) HUMAN AUDITORY SYSTEM (HAS)

Anatomically, human ears consist of three components termed external (or outer), middle, and inner ears. As depicted in Fig. 5a, the external ear comprises the pinna (the visible flesh informally called ear), the auditory canal, and the tympanic membrane (the eardrum). The middle ear is made of three small connected bones called malleus (hammer), the incus (anvil), the stapes (stirrup), and the Eustachian tube. This part of the ear acts as a mechanical transformer, which converts the sound pressure into vibrations. The inner ear comprises the cochlea that transduces the mechanical vibrations to nerve impulses.

The process of audition begins when sound waves enter the ear and proceed down the auditory canal, triggering the oval-shaped tympanic membrane to vibrate. The surface area of tympanic membrane is 68 mm^2 in humans, and exquisitely sensitive to vibrations that displace the membrane by only 1/100,000,000 of a centimeter (i.e., the width of a single hydrogen atom) [45]. In other words, it is sensitive to the increase in sound pressure level before the eardrum about 10 dB higher over the frequency range of 3-3.5 kHz. In the middle ear, auditory ossicles (hammer, anvil and stirrup) transmit the tympanic vibrations into the lymph fluid of the inner ear.

The inner ear acts as an apparatus to transform vibrations to properly coded neural impulses. Particularly, the cochlea (meaning snail) is a spiral-shaped structure that produces traveling waves along the basilar membrane based on the vibrations in the lymph fluid. In order to produce maximum vibration in different parts of the membrane at different frequencies, the basilar membrane becomes broader towards the end of the cochlea. On the surface of the basilar membrane, there are approximately 15,000 hair cells, which are the peripheral extremities of the auditory nerves. There are two groups of hair cells known as inner hair cell and outer hair cells. A single inner hair cell is attached to around 20 auditory nerves thus contributing majorly in transmitting information from the cochlea to the auditory center. While, the outer hair cells are attached to the basilar membrane through about 6 auditory nerves and play vital role in high sensitivity and

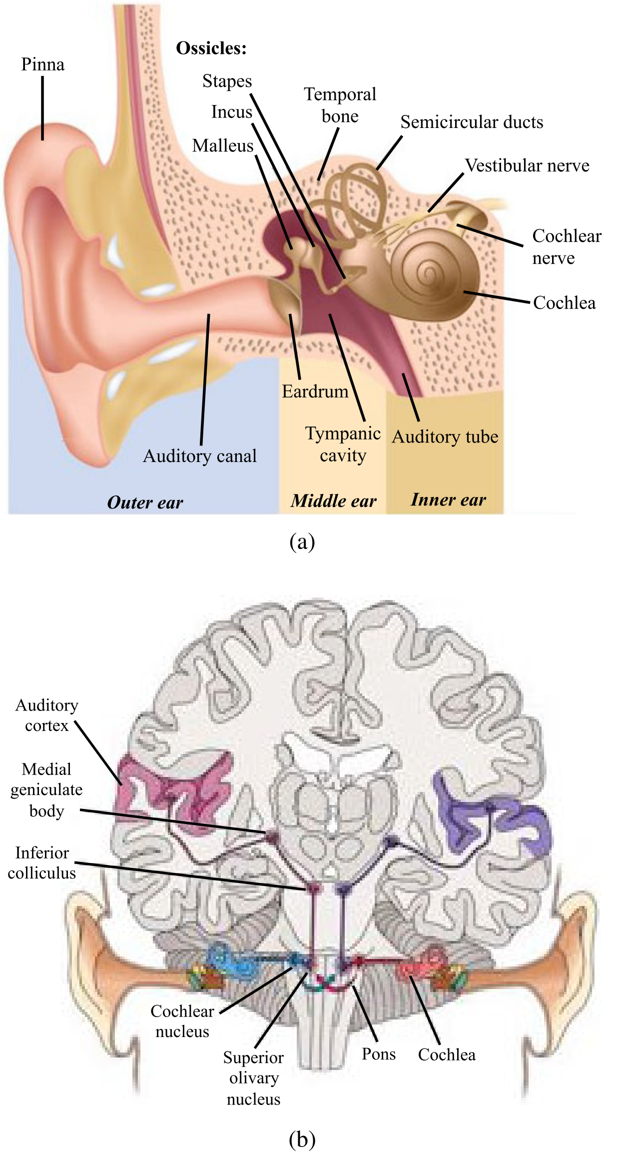


FIGURE 5. (a) Schematic illustration of the auditory anatomy. (b) Afferent pathways to the auditory cortex of the cerebral cortex [45].

acute resonance characteristics [41]. The responses in the auditory nerve fibres are very frequency selective, since the hair cells are selective for specific sound frequencies.

Fig. 5b shows the various pathways from the cochlea to the auditory nerve system and to the auditory center of the cerebral cortex. The auditory signals from the cochlea reach the cerebral cortex (situated in the temporal lobes) through the cochlear neurons, superior olivary nucleus, inferior colliculus and medial geniculate body [45]. The signals on this pathway are subject to increasingly complex processing in various parts of the brain. The nerve cells originating from the cochlea enter into the cochlear nucleus in the brain stem; at this stage decoding of the basic signal also occurs, such as intensity, duration and frequency. The superior olivary complex is another crucial relay in the brain stem, where most of the auditory fibres synapse thereby crossing the midline. The message from the superior olivary complex is then delivered to the inferior colliculus, which also plays a role in the localization of sound. Before the auditory cortex, a nucleus rely occurs in the medial geniculate body (MGB) that influences sound recognition and localization, emotional responses to sounds, auditory attention, and preparation of a motor response (e.g., vocal response). The final neuron of the MGB links the thalamus to the auditory cortex, where the message (that already decoded mostly throughout its passage in the pathway) is recognized, memorized and integrated in response.

Auditory perception is managed by two eminent components namely, binaural and masking hearing [46] apart from attention. Audio masking is a visceral episode in which user could not respond to a perceived auditory stimulus in the presence of another lower level signal. Masking can be described as a phenomenon where the threshold of audibility of sound is increased in the presence of other sounds. In some instances, lower level signals are much harder to hear which is called partial masking. Further, in some cases they become fully inaudible in the presence of other signal, which is known as temporal masking. When the sound signal and the masker (sound signal that precipitates the masking) exist simultaneously, the effect is called simultaneous masking. The masker pattern which is a psychophysical measure of the magnitude of spreading in the strength of nerve excitation, is obtained by moving frequency of signal and keeping masker fixed during search of the masking threshold.

Namely, masking pattern transpire when the masker stimulates the auditory system [47].

The sound source orientation in the area encompassing blur is practicable owing to *binaural hearing*. Experimentally, it has been demonstrated that contrasts in the timing and magnitude of sounds discerned by human ears can be utilized as cues for directional perception, as experimentally proven in many studies [41]. Specifically, the time difference and difference in intensities are main cues for direction perception under lower frequencies below 1.5 kHz. The sound localization blur is shaped by inter-aural cross correlation [47].

In summary like several operations of the nervous system, there are various video and audio perception theories existing in the literature. The two main processing schemes commonly adopted are namely, bottom-up and top-down. The top-down and bottom-up processing theories describe that the higher-level cognitive procedures and the low-level sensory data are the most major factors contributing to humans perception.

IV. WHAT ARE MAIN ISSUES AND FUTURE RESEARCH OPPORTUNITIES

To evolve the state-of-the-art in the multimedia QoE domain, few of the foremost research directions and issues are discussed in this section. Also, Table 3 is included to identify the listed directions, issues and challenges in the field of QoE together with related research works that have identified and faced them, as well as the corresponding approaches that they have used.

TABLE 2. Comparison of QoE modeling approaches.

QoE modelling approach	Human influencing factors		System influencing factors				Context influencing factors	Application domain
	Low-level processing	*High-level processing*	*Content*	*Media*	*Network*	*Device*		
Perkis *et al.* [58]	×	×	✓	×	✓	✓	×	Voice on demand, Mobile TV
Kim *et al.* [59]	×	×	×	×	✓	✓	×	IP television
Möller *et al.* [57]	✓	✓	×	×	✓	✓	✓	Multimedia
De Moor *et al.* [60]	×	✓	✓	✓	✓	✓	✓	Multimedia
Song *et al.* [61]	✓	✓	✓	✓	✓	✓	✓	Video
Reichl *et al.* [62]	✓	✓	✓	✓	✓	✓	✓	Broadband and Internet

A. INFLUENCING FACTORS

QoE is affected by diverse set of inevitable and inherent covariates of the signals and factors. They are usually grouped as human (user), technological (system) and contextual influential factors as shown in Fig. 1. The human (user) influential factors are composed of low level (e.g., gender, age, user's emotions) and high level (e.g., educational background, knowledge, previous experiences) cognitive processing. The former is easy to interpret, while latter is relatively hard to estimate. The technological (system) influencing factors are made up of content related (e.g., color, texture, 2D/3D), media related (e.g., synchronization, sampling rate), network related (e.g., delay, bandwidth) and device related (e.g., personalization, security, privacy) properties. The contextual influencing factors include temporal (e.g., usage frequency, time), physical (e.g., location), economic (e.g., cost, brand), technical and information (e.g., availability of networks), task (e.g., multitasking), and social (e.g., involvement of group) context.The issue of assessing and quantifying the end-users' QoE has been studied in different applications, including more realistic (i.e., close to real-life) situations. For example, Tsiropoulou *et al.* [69] and Lykourentzou *et al.* [70] explored QoE in social cyber-physical systems. In particular, authors in [70] presented a smart routing and recommendations model to improve the QoE of museum visitors. The proposed system takes into account the users' interests and their visiting styles together with crowd locations in the museum that impact each visitor's QoE. For individual QoE, the visitors are grouped into four distinctive categories: ant (i.e., visitors that move linearly and visit almost all exhibits), butterfly (i.e., visitors that move nonlinearly, do not follow the curator's suggestions and often change the direction of their movement), fish (i.e., visitors that move in the center of rooms, seek to see the 'larger' picture, do not approach most exhibits and do not stop very frequently) and grasshopper (i.e., visitors that show particular interests, approach certain exhibits, cross empty spaces and spend a significant amount of time in front of items of interest). Reported empirical results showed that fish visitor type benefited the most by the presented model in [70].

Wehner *at al.* [71] analyzed the mobile QoE on the long run for YouTube videos in smartphones using a publicly available app for crowsourced QoE measurements called YoMoApp. Authors determined that over time users utilizing YoMoApp obtain improved QoE if cellular network performance and streaming behavior improved. Similarly, Wassermann *et al.* [72] suggested that YouTube mobile QoE can be realized via machine learning models (e.g., random forests) with high accuracy using only network-related features. In turn, Algar *et al.* [73] proposed an automated real-time video QoE management framework for mobile users depending on the network traffic policies and user actions. Besides above wroks, Mrvelj and Matulin [74] studied user's QoE in real life environments by examining the impact of packet loss related issues in User Datagram Protocol (UDP) when 1-h multimedia content (e.g., a TV program) were streamed. The reported results revealed that users negatively perceive QoE when packet loss rate is $\geq 1\%$ and the video contains 7 or more packet loss occurrences.

Table 2 summarizes few well-known representative prior attempts of QoE modelling approaches considering different parameters. All in all, it is worth noticing in the literature that majority of QoE methods focused only on relationship between QoS and QoE without taking account of contextual and business domains, and human characteristics and roles. Very few studies contemplated all (user, system and contextual) factors. However, they are still with very limited scope, as they are efficient for either specific contextual aspects or do not function for multiple roles or domains. Therefore, QoE conceptualization requires further improvement with more concepts, taxonomy, and inter domain mapping. Since, explicit relationships between various factors (and sub factors) are not established broadly, thus the most significant challenge is developing fast and accurate QoE tools covering all the varying combination of relevant factors, also the non-technical ones. As such, there is ample room for improvement. Moreover, there exist no systematic large scale meta investigation of effects of influence elements on QoE.

B. DETERIORATION OF VISUAL AND AUDIO SIGNALS

Since audio and visual signals are main modalities in multimedia content, they play one of the biggest roles in QoE. Particularly, degradation in audio (e.g., reverberation, delay) and/or video (e.g., blurring, jerkiness) and desynchronization between them leads to degradation in perceptual QoE. Thus, QoE models are recommended to perform synchronization/alignment separately or within its feature extraction technique. Some research has been conducted to

TABLE 3. Representative research works that attempted to resolve the listed challenges and open issues in quantifying and assessing the QoE.

Challenges/Open Issues	Study	Approach/Scheme
Influencing factors	Perkis *et al.* [58]	Framework using terminal software and hardware, QoS parameters and codec information
	Reichl *et al.* [62]	Scheme based on logarithmic psychophysics using the Weber-Fechner law
Deterioration of visual and audio signals	Martinez and Farias [75]	The weighted Minkowski based model
	Hayashi *et al.* [76]	Synchronization based multimedia perceived quality model depending on audiovisual delay
	Korhonen *et al.* [115]	Method integrating subjective inspired artifacts features into an objective metric, which can be used to minimize the subjective distortion in video broadcasting over error-prone packet-switched networks
Mean opinion score (MOS)	Laghari *et al.* [77]	MOS/QoE model employing direct effect (prediction factors) and indirect effect (mediating factors)
	Gomez *et al.* [78]	Modified MOS with initial buffering time, rebuffering frequency and mean rebuffering time in Youtube
QoE taxonomy	Moller *et al.* [57]	Framework with three layers: (i) The QoS-influencing factors; (ii) the QoS interaction performance; (iii) the quality perception and judgment processes taking place inside the user
	Weiss *et al.* [79]	Taxonomy of quality aspects, quantitative metrics covering user and system performance, quality, usability, and acceptability
Human physiology and QoE	Falk *et al.* [80]	QoE System using near-infrared spectroscopy and physiological biosignal sensors
	Moldovan *et al.* [81]	Use of EEG to correlate perceived QoE of multimedia with varying properties
Standardization efforts	ITU-T [82]	Requirements for an objective perceptual multimedia QoE model via audio and video components
	Coverdale *et al.* [83]	Parametric planning (i.e., opinion model for video streaming (G.OMVS)) and packet-layer (i.e., non-intrusive parametric model for the assessment of performance of multimedia streaming (P.NAMS)) models for IPTV services
Nonintrusive (no-reference) QoE	Singh *et al.* [84]	QoE module in DVB-H networks utilising both network and video parameters without original signal
	Cherif *et al.* [85]	Method estimating end-user's perceived QoE using a trained Neural Network in H264/SVC stream
	Wu *et al.* [116]	Blind QoE metric for stalled streaming video without using its original sequence but only using global intensity and local texture features
	Nafchi *et al.* [120]	Higher orders of Minkowski distance and entropy based metric for contrast distorted samples
Wide-reaching evaluation	Gottron *et al.* [86]	Quality of Experience of Voice Communication in large-scale mobile ad hoc networks
	Ahmed *et al.* [87]	QoE analysis of a large-scale live video streaming event
Better mapping/fusion schemes	Menkovski *et al.* [88]	Machine learning fusion techniques via network and application QoS parameters
	Martin *et al.* [89]	A combination of a convolutional neural network, recurrent neural network, and Gaussian process classifier for multimedia QoE using network flow packets
Immersive 3D/virtual reality QoE	Doumanoglou *et al.* [68]	QoE of real-time 3-D media content streamed to augmented virtual reality headsets for interactive games
	Venkatraman *et al.* [90]	Multi-modal QoE prediction approach for 3D Tele-Rehabilitation system with Microsoft Kinect cameras and haptic devices
	Battisti *et al.* [117]	QoE assessment framework to understand the visual effect of asymmetric and symmetric encoding for immersive media by exploiting the concept of Preference of Experience and standardized QoE protocols
	Sawahata *et al.* [118]	Empirical study to estimate depth range required for 3D displays to show depth-compressed scenes without inducing sense of unnaturalness
Saliency-based QoE	Nauge *et al.* [91]	QoE metric using interest points extracted from different regions obtained via a hierarchical saliency map
	Luo *et al.* [92]	Spatial distortions measured by MSE and SSIM in terms of saliency weighted images, saliency maps and texture maps for video QoE
	Yang *et al.* [114]	A QoE model for stereo video that combines diverse features based on 3D saliency and sparsity
Long duration QoE	You *et al.* [7]	A weighted temporal averaging QoE method for long-term sequences
	Borowiak *et al.* [93]	A QoE method for mono- and multi-modal, long-term of audiovisual content with method of adjustment
Speech intelligibility	Taal *et al.* [94]	A objective intelligibility measure depending on time frequency (TF) weighting
	Goetze *et al.* [95]	An evaluation of several objective quality measures for predicting the quality of the dereverberated speech
Data-directed QoE	Abiru *et al.* [96]	Decentralised computing framework for mobile/IoT devices to improve QoE
	Kumar *et al.* [97]	A QoE driven rate adaptation approach for adaptive HTTP streaming considering both bandwidth savings and video quality adaptation
EEG-based QoE	Perez *et al.* [98]	A study on direct measurement of perception of image QoE from frontal electroencephalography
	Moldovan *et al.* [81]	Use of EEG to correlate perceived QoE of multimedia with varying properties
Gaming QoE	Moller *et al.* [103]	ITU-T standards for gaming QoE evaluation and management
	Sabet *et al.* [104]	A weighted PSNR objective quality method for perceptually-coded cloud gaming video
Mulsemedia QoE	Yuan *et al.* [99]	Subjective tests investigating user QoE with intensity of certain mulsemedia components
	Murray *et al.* [100]	A study on influence of users' age and gender on QoE considering various scent types and categories
	Murray *et al.* [109]	An empirical data-based model to estimate user QoE of olfaction-enhanced multimedia by considering the influence of system, user and content factors
	Jalal *et al.* [110]	A IoT paradigm-based prototype for a real smart home scenario with real consumer devices, which allows a subjective test measurement campaign to assess the QoE of the users and the feasibility of the proposed multi sensorial media TV service
QoE for hearing and vision impaired people	Fullerton *et al.* [101]	An empirical study on measuring perceived video QoE of MPEG enhancement by people with impaired vision
	Coutinho *et al.* [102]	An analysis of audio information in FPS game and its impact on deaf players QoE
QoE in Light Field Multimedia	C. Perra [127]	Evaluation study of subjective and objective QoE when viewing rendered decompressed light field images
	Viola *et al.* [128]	A framework to assess the visual quality of light field tensor displays on conventional 2D screens
	Tamboli *et al.* [129]	Univariate generalized Gaussian distribution based objective quality assessment method for key frames extracted from light field video contents
	Kara *et al.* [130]	Concept of dynamic adaptive streaming for light field video and its evaluation through QoE studies
Towards Advanced Machine Learning and Artificial Intelligent based QoE	M. Narwaria [141]	A study showing how well the ML based QoE models exploit data structure for learning
	Rego *et al.* [142]	An AI system to detect and correct errors in multimedia transmission to guarantee the QoS and QoE
5G-QoE	Martin *et al.* [143]	A network resource allocator system that enables autonomous network management aware of QoE in a 5G network scenario
	Nightingale *et al.* [144]	A 5G-QoE framework to model QoE for ultra-high-definition video flows in 5G networks
Crowdsourcing	Wu *et al.* [105]	Combination of paired comparison, probabilistic choice modeling, consistency checking, and cheat detection mechanisms
	Lin *et al.* [106]	Platform using a measuring device and crowdsourcing tools to measure the objective and subjective metrics under different scenarios
Reproducible research	Aldahdooh *et al.* [107]	Platform for video QoE using interesting subsets among coding environments and quality measures
	Aldahdooh *et al.* [108]	Largescale QoE using SoftwareX framework based on PSNR implementations

determine audio-visual synchronization acceptability threshold. However, a wide range of thresholds that are application dependent have been identified by different studies, which work better only for talking head videos. Therefore, future studies should focus on devising techniques to identify application and content independent universal threshold.

C. MEAN OPINION SCORE (MOS)

Though MOS is most extensively utilized as an indicator of perceived QoE, it is employed often without much consideration of its scope and limitations, e.g., negligence of contents and constraints foisted by the subjective tests' design while reporting the performances. MOS tuning process correlates (using a fitting function) subjective and objective MOS values, thus it is hard to estimate how well these MOS digits correspond in definite terms. It is still an open question whether the fitting function should be considered part of the objective QoE models. A further issue is the MOS outcomes against discrete studies coupled with discrete subjective tests can not be directly compared.

D. QoE TAXONOMY

There is no standard taxonomy of available objective QoE aspects (associated with human judgment and perception). Basic insufficient efforts have been put forward, e.g., [57]. Thus, it is vital to put in place correct taxonomy standards, which will help not only to facilitate quality measures but also to standardize the framework of commercial QoE systems. For better usage, the taxonomy should define the same features and feature levels for both objective QoE engines and human users.

E. HUMAN PHYSIOLOGY AND QoE

Understanding user's physiology (e.g., acuteness of vision and hearing, cognitive, emotional, or behavioral phenomena, etc.) may engender a superior comprehension of different influence factors and its relationship with QoE. Electroencephalography (EEG), functional magnetic resonance imaging (fMRI), skin temperature are examples of physiological observables. One vital issue of physiological measurements is intrusiveness, especially the ones that require attaching sensors to subjects, thereby possibly causing the user to change their natural behavior and feel some discomfort. However, recent advanced less intrusive technologies could help mitigate the above concern, e.g., estimating ECG from subject's face video. It is high time to lay the ground work in analyzing physiological measurements in relation to QoE for future multimedia services and technologies.

F. STANDARDIZATION EFFORTS

The QoE standardization ensures that certain QoE models, concepts, taxonomy and protocols are able to be used to efficiently implementing the assessment methods, products and user's feedback reporting. There exist standardization efforts that have resulted in few QoE standards (some of them are still in the making), but majority of these QoE standards do not yield predictions that are significantly more correlated with MOS as they do not cover all aspects of QoE perception.

G. NONINTRUSIVE (NO-REFERENCE) QoE

Though reduced- or full reference-based QoE assessment techniques that require the reference (original/ideal) information achieve top-level precision, they are not applicable in all applications. Therefore, blind or no-reference QoE assessment is more important but challenging problem as there is no reference sample information. The human nervous, visual and hearing systems are able to perform no-reference QoE assessment primarily owing to immense prior knowledge and superior understanding of multimedia samples. Efforts are afoot to design methods that are able to estimate nonintrusive subjective QoE with enhanced accuracy.

H. WIDE-REACHING EVALUATION

There is lack of large-scale assessment of existing QoE systems to manifest statistical gravity for published results, particularly under different impairments and application scenarios. There is an urgent need as any such comparative assessment of QoE techniques would not only help to report baseline performances without giving a false sense of progress but also be very instructive for future technical advancements.

I. BETTER MAPPING/FUSION SCHEMES

Meaningful mapping/fusion function is required to map not only objective quality measures into the predicted subjective score but also the relationship between QoS and QoE. Mapping functions can be divided into linear and non-linear. There is a requirement of devising better unsupervised learning or non-learning based mapping functions and audio-visual fusion schemes in this regard, e.g., those depend on deep learning or Bayesian belief graphical models, which is an extremely challenging task. To further amplify the accuracy of intra-application systems, audio-visual devices information and features may be comprised in dynamic mapping and fusion schemes.

J. IMMERSIVE 3D/VIRTUAL REALITY QoE

Emerging immersive technologies (e.g., 3D virtual reality (VR360) with 360 degree) are bringing many new challenges in QoE field. Conventional objective QoE techniques cannot be put in straightaway for these emerging immersive applications, since synthetic environment is assembled by dynamic 3D binocular and binaural or stereo rendering that is different that 2D techniques. Few recently devised preliminary immersive 3D QoE methods are computationally complex, have lower accuracy, and compromise user's comfort level and assessment. Dearth of public data sets including multi-label metadata (e.g., subjective MOS, lifestyle, occupation and geography) has further hindered research and development on this subject. Moreover, new technologies (e.g., holographic screens, multi sensorial media and Internet of Things (IoT)) may be utilized to create immersive environments, which are expected to soon become the next frontier of advanced broadcast services [113]. Therefore, there is a need of dedicated research not only for evaluating the level of enhancement in user QoE but also comparing functionalities and requirements in such environments in order to properly design devices, architectures, networks, and methods that will be helpful to regulate services and obtain the expected or promised user QOE.

K. SALIENCY-BASED QoE
Physiological and psychological studies have shown that humans do not pay equal attention to all exposed multimedia information, but only focus on certain parts known as saliency regions or focus of attention (FOA). While, most current QoE methods analyze all spatial and temporal information equally, therefore perhaps leading to poor match with subjective perceived-QoE. Deeper understanding about the audiovisual FOA and perceived-QoE is still a very much open issue, especially how to incorporate saliency of content into no-reference based QoE. Recently, limited analysis on eye-tracking for Saliency-based QoE has been performed. Other compelling aspect to investigate could be instantaneous perceived-QoE that may aid analyzing the relationship between the quality/QoS/QoE and the content in detail.

L. LONG DURATION QoE
Very limited attention has been devoted to QoE for long-duration sequences. Rather, exiting short-duration based techniques are generally utlized at small temporal portions and later averaged for overall QoE with identical weights for every portion, which might not be much practical where quality content and FOA fluctuate over time. There is a need to continue devising novel assessment techniques and objective metrics that yield perceived-QoE as accurate as possible. Especially, tools for continuous and multi-modal long duration QoE assessments.

M. SPEECH INTELLIGIBILITY
Speech intelligibility (i.e., how well or clearly a native speaker with healthy hearing and cognition can understand what is being said) is another dimension of QoE assessment, since the intelligibility is related to the information content that plays significant role in QoE. Most existing noise suppression techniques may improve speech quality but hardly improve speech intelligibility. Moreover, speech intelligibility is mainly investigated by automatic speech recognition research community and has received limited attention from QoE community.

N. DATA-DIRECTED QoE
Development of data-driven QoE research is yet at its early stages, thus there is yet substantial room for developments. Several aspects are interesting to be explored, e.g., dynamic metric selection such that new QoE metrics are selected as the combination of QoS and external factors change, since user expectations and network alter with time. Also, in-depth analysis and understanding of day, time and user behaviors may lead to better user QoE, e.g., some users' QoE might be extra sensitive concerning the video content and waiting for a video to start.

O. EEG-BASED QoE
Due to ubiquitous technological advances, the requirement for ingenious solutions to estimate user's QoE with multimedia utilities is increasing. One such solution is use of brain cortical measurements for QoE, namely use of EEG for user's quality perception and experiences. The main advantage of EEG based QoE methods is estimating and quantitating the effect of different components on user's QoE in a non-intrusive manner such that users need not to give input on their discerned qualities and experiences. Despite several benefits, brain measurements for QoE have not been much explored. It would also be enthralling to investigate how quality expectations play role in early level of EEG-QoE. Another direction showing promising preliminary QoE outcomes based on neurophysiological measure is NIRS.

P. GAMING QoE
With the advent of video games, the demand of producing high gaming (user-perceived) QoE is rising dramatically. Nonetheless, the user-discerned gaming QoE has not been rigorously analyzed compared to other multimedia utilities; mainly because computerized gaming is a human-machine interaction procedure, rather than just a media delivery one. Therefore, standard multimedia QoE techniques are not sufficient. The problem of gaming QoE is compounded by diverse components such as game platform, user interface software and hardware, transmission channels, and single- and multi-player games, whose exact correlations on user-perceived gaming QoE is yet to be formularized. Also, recent internet connections based mobile games QoE has seen limited research efforts.

Q. MULSEMEDIA QoE
Mulsemedia (MULtiple SEnsorial MEDIA) is a promising advancement of multimedia, which enriches traditional multimedia content with novel objects such as olfactory, haptic and thermal ones [12]. Mulsemedia is becoming integral part of several emerging multimedia technologies, especially immersive applications. Ongoing mulsemedia QoE research is focused on formidable topics such as synchronization of mulsemedia data with audio and video channels and capturing and rendering mulsemedia signals. Studies, e.g., [49], [111] and [112], have shown that presence of scents, considering crossmodal correspondences in a mulsemedia setup, and delivering sensory effects to heterogeneous systems improves user's QoE. However, most olfactory information based systems and methods are still only workable in the specialized laboratories. Likewise, haptic interactions based QoE needs further refinement and advancement since it demands a sound understanding of both kinesthetic and tactile perceptual doctrines. Also, interrelationships between special mulsemedia effects, e.g., wind, motion and illumination, on QoE have not systematically analyzed in any of the existing studies.

R. QoE FOR HEARING AND VISION IMPAIRED PEOPLE

Unfortunately, some people suffer from hearing or vision impairment caused by aging, accident or birth abnormality. User dissatisfaction with most commercially available hearing and vision aids are fairly high [63], thereby exacerbating need of novel signal processing and QoE techniques tuned to impaired users. Future QoE methods should take impaired people into consideration during development, such techniques will not only help assessing hearing/vision aids but also have a great impact in society.

S. QoE IN LIGHT FIELD MULTIMEDIA

Light field imaging has experienced a surge of popularity owing to its capacity of rendering the 3D world in a more immersive fashion. Quintessential photography captures a two-dimensional projection of the light, whereas light fields gather the luminosity of light rays along different directions [123]. The prospective for light field applications spans across numerous fields such as gaming, telepresence in video conferencing, and medical imaging. Recently, few light field capture, processing and display/rendering devices and techniques have been developed [124], [125]. Extremely large magnitudes of light field dimensional data bring new challenges in capture, compression, editing, transmission and display. While other domains of QoE have received relatively more attention, the research on QoE in light field yet remains mostly unexplored; specially QoE in light field view even though no headgears or special glasses are needed to view the visualized content in 3D. Since success of light field display and services essentially relies on the QoE they provide, there is huge arena waiting to be explored by scientific community. For instance, finding correlation between coding artifacts, characterization features, view-synthesis and refocusing, and changing viewpoint. Also, comprehensive light field QoE databases, which compromise subjective and objective scores, contents, artifacts, angular and spatial resolutions, etc., should be produced and publicly released [126]. Frameworks to evaluate the light field QoE should be proposed, which can be used both subjective and objective analyses and quantitative metrics with multilayer components as well as single and double stimulus methodologies. Such frameworks would assist not only in selecting the rendering parameters in evaluation scenario in order to remove any bias towards a specific solution but also testing streaming schemes of cost-effective real-time light field videos transmissions.

T. TOWARDS ADVANCED MACHINE LEARNING (AML) AND ARTIFICIAL INTELLIGENT (AI) BASED QoE

Quintessential QoE frameworks are frequently depend on explicit modeling of the highly non-linear conduct of human perception, which lead to overfitting or uncertain overall reliability. Whereas, AML/AI based QoE schemes attempt to emulate its perception for better accuracy and services. Despite AI and AML's recent success in various fields, objective QoE models using AML/AI (e.g., dictionary/unsupervised/deep learning) have been explored in a very limited manner. AML and AI paradigms should be studied for representation learning, including feature extraction/selection, classification, mapping functions as well as determining temporal correlations within and between different modalities, to attain enhanced generalization and interoperability capability of the frameworks. AML/AI based QoE's inherent limitations such as lack of focus on quantifying the learning ability and qualitative aspects of training and testing datasets must be studied in order to develop alternative approaches to overcome some of them.

U. 5G-QoE

Next generation networks (i.e., 5G networks) is aiming to merge heterogeneous network technologies with smart devices to create concept of ubiquitous computing with Internet of Things (IoT). 5G will have to deal with dynamic, diverse, fast, and multi-tier networks, but they also should meet the desired QoE. Some recent works focused on investigating the significance of the foremost challenges of 5G on QoE [144]. The 5G-QoE methods using different neural networks techniques should be designed, as neural networks have potential to fully learn the interrelation between resulting QoE, devices and network parameters. AI and AML algorithms can be explored for better self-configuring and self-healing, and self-optimizing mechanisms and corresponding 5G-QoE. Future works should also concentrate on establishing pros and cons of existing 3G/4G-QoE frameworks over 5G-QoE metrics, going deep-rooted topology mutation based on the present topology, and developing convenient schemes to collect 5G-QoE data on run time basis.

V. CROWDSOURCING

An ongoing challenge in QoE evaluation is lack of publicly available big datasets with larger subjective ratings. Crowdsourcing approaches (e.g., Amazon's Mechanical Turk, Facebook, Microworkers) may be a potential solution for this challenge. Crowdsourcing approaches move the evaluation from the lab territory into the Internet, thus permitting to obtain a global and diverse pool of subjects as well as including real life territories into the assessment task and reducing the turn-around time. Nonetheless, the reliability of ratings could be lower owing to the anonymity of users on the crowdsourcing platforms and loss of overall control. Crowdsourcing in QoE is a vital tool and regardless of recent research works, there still remains significant works to be done in order to it replaceing traditional subjective testing in the lab.

W. REPRODUCIBLE RESEARCH

In QoE research community, reproducible research trend should be encouraged by providing public large databases with larger subjective ratings, open source software and experimental setups. It will greatly serve to study scalability, since majority of the published results allude to comparatively small evaluation data sets that may not demonstrate the problems that may appear in actual applications.

V. CONCLUSION

Due to tremendous growth, popularity and usefulness of multimedia services and applications during the last decade, perceived-QoE at the end-user has become a vital criterion. Multimedia broadcasters and commercial service providers are increasingly embracing QoE principles owing to they may bring more end-users and revenue. Today, multimedia QoE is a thriving field with several practical applications and active research topics, including 5G/IoT-QoE, and advanced machine learning and artificial intelligence-based QoE frameworks.

Despite notable progress, more accurate assessment of QoE is proved to be a difficult endeavor, since QoE is user-centric, individual, multidimensional and multisensorial. This paper provides some insights on why QoE assessment is so strenuous by presenting different interpretations of quality, a thorough illustration of QoE perception and genesis including human auditory and vision systems. An overall synopsis of QoS, QoP and QoE within the framework of audio-visual QoE estimation is also discussed. The years ahead are full of challenges and opportunities to ameliorate perceived-QoE further and bring it to new frontiers, thus main open challenges are also highlighted with a discussion on potential future research directions. In recent years, various new QoE research possibilities using intelligent methods and significant immaturity of the available frameworks and tools have been pointed out, which demand great deal of research efforts to bring out the interesting insights, novel QoE schemes and improved users' satisfaction. Therefore, this article also seeks to motivate budding scientists, researchers and engineers to consider multimedia QoE as their field of study.

REFERENCES

[1] U. Jekosch, *Voice and Speech Quality Perception: Assessment and Evaluation* (Signals and Communication Technology). New York, NY, USA: Springer-Verlag, 2005.
[2] D. S. Hands, "A basic multimedia quality model," *IEEE Trans. Multimedia*, vol. 6, no. 6, pp. 806–816, Dec. 2004.
[3] A. R. Avila, Z. A. Momin, J. F. Santos, D. O'Shaughnessy, and T. Falk, "Feature pooling of modulation spectrum features for improved speech emotion recognition in the wild," *IEEE Trans. Affective Comput.*, to be published.
[4] R. Gupta, K. Laghari, H. Banville, and T. H. Falk, "Using affective brain-computer interfaces to characterize human influential factors for speech quality-of-experience perception modelling," *Hum.-Centric Comput. Inf. Sci.*, vol. 6, no. 1, 2016, Art. no. 5.
[5] S. Bech and N. Zacharov, *Perceptual Audio Evaluation—Theory, Method and Application*. Hoboken, NJ, USA: Wiley, 2006.
[6] Z. Akhtar and T. H. Falk, "Audio-visual multimedia quality assessment: A comprehensive survey," *IEEE Access*, vol. 5, pp. 21090–21117, 2017.
[7] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio–visual services: A survey," *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 482–501, 2010.
[8] B. Girod, "What's wrong with mean-squared error?" in *Digital Images and Human Vision*. Cambridge, MA, USA: MIT Press, 1993.
[9] D. Tian and G. AlRegib, "FQM: A fast quality measure for efficient transmission of textured 3D models," in *Proc. Int. Conf. Multimedia*, 2004, pp. 684–691.
[10] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[11] J. G. A. Barbedo and A. Lopes, "A new cognitive model for objective assessment of audio quality," *J. Audio Eng. Soc*, vol. 53, nos. 1–2, pp. 22–31, 2005.
[12] S. Möller and A. Raake, *Quality of Experience: Advanced Concepts, Applications and Methods*. Basel, Switzerland: Springer, 2014.
[13] (2011). *Merriam Webster Dictionary*. [Online]. Available: http://www.merriam-webster.com
[14] B. Belmudez, *Audiovisual Quality Assessment and Prediction for Videotelephony*. Basel, Switzerland: Springer, 2014.
[15] A. K. Rai, *Customer Relationship Management: Concepts and Cases*. New Delhi, India: PHI Learning Pvt Ltd., 2012.
[16] R. M. Pirsig, *Zen and the Art of Motorcycle Maintenance*. New York, NY, USA: Bantam Books, 1974.
[17] *Quality—Vocabulary*, Standard ISO 8402:1986, 1986. [Online]. Available: http://www.iso.org/iso/catalogue_detail.htm?csnumber=15570
[18] *Managing for the Sustained Success of an Organization—A Quality Management Approach*, Standard ISO 9004:2009, 2009. [Online]. Available: http://www.iso.org/iso/catalogue_detail?csnumber=41014
[19] K. B. Leffler, "Ambiguous changes in product quality," *Amer. Econ. Assoc.*, vol. 72, no. 5, pp. 956–967, 1982.
[20] B. G. Dale, K. D. Barber, R. T. Williams, and T. van der Wiele, "Managing quality in manufacturing versus services: A comparative analysis," *Manag. Service Qual., Int. J.*, vol. 7, no. 5, pp. 242–247, 1997.
[21] H. L. Gilmore, "Product conformance cost," *Qual. Prog.*, vol. 7, no. 15, pp. 16–19, 1974.
[22] S. T. Akinyele, "The spiritual perspective of quality: A scriptural dimension," *IFE Psychol.*, vol. 6, no. 2, pp. 62–77, 1974.
[23] W. Abramowicz, R. Hofman, W. Suryn, and D. Zyskowski, "SQuaRE based Web services quality model," in *Proc. Int. MultiConf. Eng. Comp. Scientists I*, 2008, pp. 827–835.
[24] E. Touradj, "Quality of multimedia experience: Past, present and future," in *Proc. 17th ACM Int. Conf. Multimedia*, 2009, pp. 3–4.
[25] D. Hoyle, *ISO 9000 Quality Systems Handbook*. Oxford, U.K. Butterworth-Heinemann, 2001.
[26] K. Brunnström *et al.*, "Qualinet white paper on definitions of quality of experience," in *Proc. 5th Qualinet Meeting*, 2013, pp. 1–24.
[27] H. Martens and M. Martens, *Multivariate Analysis of Quality: An Introduction*. Hoboken, NJ, USA: Wiley, 2001.
[28] S. Möller, *Quality of Telephone-Based Spoken Dialogue Systems*. Boston, MA, USA: Springer, 2010.
[29] J. Zanker, *Sensation, Perception and Action: An Evolutionary Perspective*. Basingstoke, U.K.: Palgrave, 2010.
[30] E. B. Goldstein, *Sensation and Perception*. Boston, MA, USA: Cengage Learning, 2009.
[31] R. W. Fleming, "Visual perception of materials and their properties," *Vis. Res.*, pp. 62–75, Jan. 2014.
[32] U. Neisser, *Cognition and Reality: Principles and Implications of Cognitive Psychology*. San Francisco, CA, USA: Freeman, 1976.
[33] R. A. Young, "Oh say, can you see? The physiology of vision," *Proc. SPIE*, vol. 1453, pp. 92–123, Jun. 1991.
[34] A. C. Guyton and J. E. Hall, *Textbook of Medical Physiology*, vol. 9. Philadelphia, PA, USA: Elsevier, 1991.
[35] D. A. Baylor, "Photoreceptor signals and vision. Proctor lecture," *Investigative Ophthalmol. Vis. Sci.*, vol. 28, no. 1, pp. 34–49, 1987.
[36] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human photoreceptor topography," *J. Comparative Neurol.*, vol. 292, no. 4, pp. 497–523, 1990.
[37] M. Glickstein, "The discovery of the visual cortex," *Sci. Amer.*, vol. 259, pp. 118–127, Sep. 1988.
[38] D. C. Van Essen, "Organization of visual areas in macaque and human cerebral cortex," in *The Visual Neurosciences*. Cambridge, MA, USA: MIT Press, 2003, pp. 507–521.
[39] S. W. Kuffler, J. G. Nicholls, and A. R. Martin, *From Neuron to Brain*. Sunderland, MA, USA: Sinauer Associates Inc., 1984.
[40] H. Davson and E. S. Perkins, "Human eye," in *Encyclopdia Britannica*. Chicago, IL, USA: Encyclopædia Britannica, 2010.
[41] D. A. Hall, I. S. Johnsrude, M. P. Haggard, A. R. Palmer, M. A. Akeroyd, and A. Q. Summerfield, "Spectral and temporal processing in human auditory cortex," *Cerebral Cortex*, vol. 12, no. 2, pp. 140–149, 2002.
[42] A. A. Michelson, *Studies in Optics*. Chicago, IL, USA: Univ. of Chicago Press, 1927.
[43] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol.*, vol. 197, no. 3, pp. 551–566, 1968.

[44] E. A. DeYoe and D. C. Van Essen, "Concurrent processing streams in monkey visual cortex," *Trends Neurosci.*, vol. 11, no. 5, pp. 219–226, 1988.

[45] G. Paxinos and J. K. Mai, *The Human Nervous System*. New York, NY, USA: Academic, 2004.

[46] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1999.

[47] *The E-Model: A Computational Model for Use in Transmission Planning*, document ITU-T Rec. G.107, 2005.

[48] C. W. Chen, P. Chatzimisios, T. Dagiuklas, and L. Atzori, *Multimedia Quality of Experience (QoE): Current Status and Future Requirements*. Hoboken, NJ, USA: Wiley, 2015.

[49] N. Murray, Y. Qiao, G.-M. Muntean, and B. Lee, "Multiple-scent enhanced multimedia synchronization," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 1, pp. 1–28, 2014.

[50] *Cisco Visual Networking Index: Forecast and Methodology, 2013–2018*, Cisco, San Jose, CA, USA, 2014.

[51] T. Roscoe and G. Bowen, "Script-driven packet marking for quality-of-service support in legacy applications," *Proc. SPIE*, vol. 3969, pp. 166–176, Dec. 1999.

[52] M. Hassenzahl, M. Burmester, K. Lehner, and A. Platz, "Hedonic and Ergonomic quality aspects determine a software's appeal," in *Proc. Conf. Hum. Factors Comput. Syst.*, 2000, pp. 201–208.

[53] K. U. R. Laghari, R. Gupta, J.-N. Antons, R. Schleicher, S. Möller, and T. H. Falk, "Objective characterization of human behavioural characteristics for QoE assessment: A pilot study on the use of electroencephalography features," in *Proc. IEEE Globecom Workshops*, Dec. 2013, pp. 1168–1173.

[54] G. Ghinea and J. P. Thomas, "Quality of perception: User quality of service in multimedia presentations," *IEEE Trans. Multimedia*, vol. 7, no. 4, pp. 786–789, Aug. 2005.

[55] S. R. Gulliver and G. Ghinea, "Stars in their eyes: What eye-tracking reveals about multimedia perceptual quality," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 34, no. 4, pp. 472–482, Jul. 2004.

[56] R. T. Apteker, J. A. Fisher, V. S. Kisimov, and H. Neishlos, "Video acceptability and frame rate," *IEEE Multimedia Mag.*, vol. 2, no. 3, pp. 32–40, Fall 1995.

[57] S. Möller, K.-P. Engelbrecht, C. Kuhnel, I. Wechsung, and B. Weiss, "A taxonomy of quality of service and quality of experience of multimodal human-machine interaction," in *Proc. Int. Workshop Qual. Multimedia Exper.*, 2009, pp. 7–12.

[58] A. Perkis, S. Munkeby, and O. I. Hillestad, "A model for measuring quality of experience," in *Proc. 7th Nordic Signal Process. Symp.*, 2006, pp. 198–201.

[59] H.-J. Kim, K.-H. Lee, and J. Zhang, "In-service feedback QoE framework," in *Proc. 3rd Int. Conf. Commun. Theory, Rel., Qual. Service*, 2010, pp. 135–138.

[60] K. De Moor, I. Ketyko, T. Deryckere, L. De Marez, L. Martens, G. Verleye, and W. Joseph, "Proposed framework for evaluating quality of experience in a mobile, testbed-oriented living lab setting," *Mobile Netw. Appl.*, vol. 15, no. 3, pp. 378–391, 2010.

[61] W. Song, D. Tjondronegoro, and M. Docherty, "Understanding user experience of mobile video: Framework, measurement, and optimization," in *Mobile Multimedia—User and Technology Perspectives*. Rijeka, Croatia: InTech, 2012, pp. 1–30.

[62] P. Reichl, B. Tuffin, and R. Schatz, "Logarithmic laws in service quality perception: Where microeconomics meets psychophysics and quality of experience," *Telecommun. Syst. J.*, vol. 52, no. 2, pp. 587–600, 2013.

[63] V. López-Ludeña, R. Barra-Chicote, S. Lutfi, J. M. Montero, and R. San-Segundo, "LSESpeak: A spoken language generator for Deaf people," *Expert Syst. Appl.*, vol. 40, no. 4, pp. 1283–1295, 2013.

[64] K. Piamrat, C. Viho, J.-M. Bonnin, and A. Ksentini, "Quality of experience measurements for video streaming over wireless networks," in *Proc. 6th Int. Conf. Inf. Technol., New Gener.*, 2009, pp. 1184–1189.

[65] V. Pouli, S. Kafetzoglou, E. E. Tsiropoulou, A. Dimitriou, and S. Papavassiliou, "Personalized multimedia content retrieval through relevance feedback techniques for enhanced user experience," in *Proc. 13th Int. Conf. Telecommun. (ConTEL)*, 2015, pp. 1–8.

[66] E. E. Tsiropoulou, G. Kousis, A. Thanou, I. Lykourentzou, and S. Papavassiliou, "Quality of experience in cyber-physical social systems based on reinforcement learning and game theory," *Future Internet*, vol. 10, no. 11, p. 108, 2018.

[67] F. Wamser, P. Casas, M. Seufert, C. Moldovan, P. Tran-Gia, and T. Hossfeld, "Modeling the YouTube stack: From packets to quality of experience," *Comput. Netw.*, vol. 109, no. 2, pp. 211–224, 2016.

[68] A. Doumanoglou, D. Griffin, J. Serrano, N. Zioulis, T. K. Phan, D. Jiménez, D. Zarpalas, F. Alvarez, M. Rio, and P. Daras, "Quality of experience for 3-D immersive media streaming," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 379–391, Jun. 2018.

[69] E. E. Tsiropoulou, A. Thanou, and S. Papavassiliou, "Modelling museum visitors' quality of experience," in *Proc. 11th Int. Workshop Semantic Social Media Adaptation Personalization (SMAP)*, 2016, pp. 77–82.

[70] I. Lykourentzou, X. Claude, Y. Naudet, E. Tobias, A. Antoniou, G. Lepouras, and C. Vassilakis, "Improving museum visitors' Quality of Experience through intelligent recommendations: A visiting style-based approach," in *Proc. Intell. Environ. (Workshops)*, 2013, pp. 507–518.

[71] N. Wehner, S. Wassermann, P. Casas, M. Seufert, and F. Wamser, "Beauty is in the eye of the smartphone holder a data driven analysis of YouTube mobile QoE," in *Proc. 14th Int. Conf. Netw. Service Manage. (CNSM)*, 2018, pp. 343–347.

[72] S. Wassermann, N. Wehner, and P. Casas, "Machine learning models for YouTube QoE and user engagement prediction in smartphones," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 46, no. 3, pp. 155–158, 2019.

[73] M. J. Algar, I. M. de Diego, A. Fernández-Isabel, M. A. Monjas, F. Ortega, J. M. Moguerza, and H. Jacynycz, "A quality of experience management framework for mobile users," *Wireless Commun. Mobile Comput.*, vol. 2019, Jan. 2019, Art. no. 2352941.

[74] S. Mrvelj and M. Matulin, "Impact of packet loss on the perceived quality of UDP-based multimedia streaming: A study of user quality of experience in real-life environments," *Multimedia Syst.*, vol. 24, no. 1, pp. 33–53, 2018.

[75] H. B. Martinez and M. C. Farias, "Full-reference audio-visual video quality metric," *J. Electron. Imag.*, vol. 23, no. 6, 2014, Art. no. 061108.

[76] T. Hayashi, K. Yamagishi, T. Tominaga, and A. Takahashi, "Multimedia quality integration function for videophone services," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2007, pp. 2735–2739.

[77] K. U. R. Laghari, N. Crespi, and K. Connelly, "Toward total quality of experience: A QoE model in a communication ecosystem," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 58–65, Apr. 2012.

[78] G. Gómez, L. Hortigüela, Q. Pérez, J. Lorca, R. García, and M. C. Aguayo-Torres, "YouTube QoE evaluation tool for Android wireless terminals," *EURASIP J. Wireless Commun. Netw.*, vol. 2014, Oct. 2014, Art. no. 164.

[79] S. Möller, K.-P. Engelbrecht, C. Kühnel, I. Wechsung, and B. Weiss, "Evaluation of multimodal interfaces for ambient intelligence," in *Human-Centric Interfaces for Ambient Intelligence*. New York, NY, USA: Academic, 2009, pp. 347–370.

[80] T. H. Falk, Y. Pomerantz, K. Laghari, S. Möller, and T. Chau, "Preliminary findings on image preference characterization based on neurophysiological signal analysis: Towards objective QoE modeling," in *Proc. 4th Int. Workshop Qual. Multimedia Exper.*, 2012, pp. 146–147.

[81] A.-N. Moldovan, I. Ghergulescu, S. Weibelzahl, and C. H. Muntean, "User-centered EEG-based multimedia quality assessment," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast*, Jun. 2013, pp. 1–8.

[82] *Requirements for an Objective Perceptual Multimedia Quality Model*, document ITU Rec. J.148, ITU Telecommunication Standardization Sector, 2010.

[83] P. Coverdale, S. Möller, A. Raake, and A. Takahashi, "Multimedia quality assessment standards in ITU-T SG12," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 91–97, Nov. 2011.

[84] K. D. Singh and G. Rubino, "No-reference quality of experience monitoring in DVB-H networks," in *Proc. Wireless Telecommun. Symp. (WTS)*, 2010, pp. 1–6.

[85] W. Cherif, A. Ksentini, and D. Négru, "No-reference quality of experience estimation of H264/SVC stream," in *Proc. IEEE Globecom Workshops*, 2012, pp. 1346–1351.

[86] C. Gottron, A. König, M. Hollick, S. Bergsträßer, T. Hildebrandt, and R. Steinmetz, "Quality of experience of voice communication in large-scale mobile ad hoc networks," in *Proc. 2nd IFIP Wireless Days (WD)*, 2009, pp. 1–6.

[87] A. Ahmed, Z. Shafiq, and A. Khakpour, "QoE analysis of a large-scale live video streaming event," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 44, no. 1, pp. 395–396, 2016.

[88] V. Menkovski, G. Exarchakos, and A. Liotta, "Machine learning approach for quality of experience aware networks," in *Proc. Int. Conf. Intell. Netw. Collaborative Syst.*, Dec. 2010, pp. 461–466.
[89] M. Lopez-Martin, B. Carro, J. Lloret, S. Egea, and A. Sanchez-Esguevillas, "Deep learning model for multimedia quality of experience prediction based on network flow packets," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 110–117, Sep. 2018.
[90] K. Venkatraman, S. Raghuraman, Y. Tian, B. Prabhakaran, K. Nahrstedt, and T. Annaswamy, "Quantifying and improving user quality of experience in immersive tele-rehabilitation," in *Proc. IEEE Int. Symp. Multimedia*, Dec. 2014, pp. 207–214.
[91] M. Nauge, M.-C. Larabi, and C. Fernandez, "Quality estimation based on interest points through hierarchical saliency maps," in *Proc. Eur. Workshop Vis. Inf. Process.*, 2011, pp. 186–191.
[92] Q. Luo, Y. Geng, J. Liu, and W. Li, "Saliency and texture information based full-reference quality metrics for video QoE assessment," in *Proc. IEEE Netw. Oper. Manage. Symp.*, May 2014, pp. 1–6.
[93] A. Borowiak and U. Reiter, "Quality evaluation of long duration AV content—An extended analysis using a novel assessment methodology," *Multimedia Tools Appl.*, vol. 74, no. 2, pp. 367–380, 2015.
[94] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 4214–4217.
[95] S. Goetze, A. Warzybok, I. Kodrasi, J. O. Jungmann, B. Cauchi, J. Rennies, E. A. P. Habets, A. Mertins, T. Gerkmann, S. Doclo, and B. Kollmeier, "A study on speech quality and speech intelligibility measures for quality assessment of single-channel dereverberation algorithms," in *Proc. Int. Workshop Acoustic Signal Enhancement*, 2014, pp. 233–237.
[96] K. Abiru, H. Ueno, and K. Amemiya, "Improving quality of experience for users through distributed service platform technology," *FUJITSU Sci. Tech. J*, vol. 52, pp. 56–63, Apr. 2016.
[97] Venkata Phani Kumar M and S. Mahapatra, "Quality of experience driven rate adaptation for adaptive HTTP streaming," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 602–620, Jun. 2018.
[98] J. Perez and E. Delechelle, "On the measurement of image quality perception using frontal EEG analysis," in *Proc. Int. Conf. Smart Commun. Netw. Technol.*, 2013, pp. 1–5.
[99] Z. Yuan, S. Chen, G. Ghinea, and G.-M. Muntean, "User quality of experience of mulsemedia applications," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 1, pp. 1–19, 2014.
[100] N. Murray, B. Lee, Y. Qiao, and G. Miro-Muntean, "The influence of human factors on olfaction based mulsemedia quality of experience," in *Proc. 8th Int. Conf. Qual. Multimedia Exper.*, 2016, pp. 1–6.
[101] M. Fullerton, R. L. Woods, F. A. Vera-Diaz, and E. Peli, "Measuring perceived video quality of MPEG enhancement by people with impaired vision," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 24, no. 12, pp. 174–187, 2007.
[102] F. Coutinho, R. O. Prates, and L. Chaimowicz, "An analysis of information conveyed through audio in an FPS game and its impact on deaf players experience," in *Proc. Brazilian Symp. Games Digit. Entertainment*, 2011, pp. 53–62.
[103] S. Möller, S. Schmidt, and S. Zadtootaghaj, "New ITU-T standards for gaming QoE evaluation and management," in *Proc. 10th Int. Conf. Qual. Multimedia Exper.*, 2018, pp. 1–6.
[104] S. S. Sabet, M. R. Hashemi, S. Shirmohammadi, and M. Ghanbari, "A novel objective quality assessment method for perceptually-coded cloud gaming video," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr.*, Apr. 2018, pp. 75–79.
[105] C.-C. Wu, K.-T. Chen, Y.-C. Chang, and C.-L. Lei, "Crowdsourcing multimedia QoE evaluation: A trusted framework," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1121–1137, Aug. 2013.
[106] Y. C. Lin and Z. S. Shih, "Crowdsourcing system on measuring quality of experience for Internet video streaming," in *Proc. IEEE Int. Conf. Consum. Electron.-Taiwan*, May 2016, pp. 1–2.
[107] A. Aldahdooh, E. Masala, G. Van Wallendael, and M. Barkowsky, "Reproducible research framework for objective video quality measures using a large-scale database approach," *SoftwareX*, vol. 8, pp. 64–68, Jul./Dec. 2018.
[108] A. Aldahdooh, E. Masala, G. Van Wallendael, and M. Barkowsky, "Framework for reproducible objective video quality research with case study on PSNR implementations," *Digit. Signal Process.*, vol. 77, pp. 195–206, Jun. 2018.
[109] N. Murray, G.-M. Muntean, Y. Qiao, S. Brennan, and B. Lee, "Modeling user quality of experience of olfaction-enhanced multimedia," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 539–551, Jun. 2018.
[110] L. Jalal, M. Anedda, V. Popescu, and M. Murroni, "QoE assessment for IoT-based multi sensorial media broadcasting," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 552–560, Jun. 2018.
[111] G. Mesfin, N. Hussain, A. Covaci, and G. Ghinea, "Using eye tracking and heart-rate activity to examine crossmodal correspondences QoE in mulsemedia," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 2, 2019, Art. no. 34.
[112] E. B. Saleme, C. A. S. Santos, and G. Ghinea, "A mulsemedia framework for delivering sensory effects to heterogeneous systems," *Multimedia Syst.*, vol. 25, no. 4, pp. 421–447, 2019.
[113] M. Murroni, R. Rassool, L. Song, and R. Sotelo, "Guest editorial special issue on quality of experience for advanced broadcast services," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 335–340, Jun. 2018.
[114] J. Yang, C. Ji, B. Jiang, W. Lu, and Q. Meng, "No reference quality assessment of stereo video based on saliency and sparsity," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 341–353, Jun. 2018.
[115] J. Korhonen, "Study of the subjective visibility of packet loss artifacts in decoded video sequences," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 354–366, Jun. 2018.
[116] Q. Wu, H. Li, F. Meng, and K. N. Ngan, "Toward a blind quality metric for temporally distorted streaming video," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 367–378, Jun. 2018.
[117] F. Battisti, M. Carli, P. Le Callet, and P. Paudyal, "Toward the assessment of quality of experience for asymmetric encoding in immersive media," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 392–406, Jun. 2018.
[118] Y. Sawahata and T. Morita, "Estimating depth range required for 3-D displays to show depth-compressed scenes without inducing sense of unnaturalness," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 488–497, Jun. 2018.
[119] Y. Shishikui and Y. Sawahata, "Effects of viewing ultra-high-resolution images with practical viewing distances on familiar impressions," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 498–507, Jun. 2018.
[120] H. Z. Nafchi and M. Cheriet, "Efficient no-reference quality assessment and classification model for contrast distorted images," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 518–523, Jun. 2018.
[121] A. Bentaleb, A. C. Begen, and R. Zimmermann, "QoE-aware bandwidth broker for HTTP adaptive streaming flows in an SDN-enabled HFC network," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 575–589, Jun. 2018.
[122] W. Huang, Y. Zhou, X. Xie, D. Wu, M. Chen, and E. Ngai, "Buffer state is enough: Simplifying the design of QoE-aware HTTP adaptive video streaming," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 590–601, Jun. 2018.
[123] I. Viola, M. Řeřábek, and T. Ebrahimi, "Impact of interactivity on the assessment of quality of experience for light field content," in *Proc. Int. Conf. Qual. Multimedia Exper. (QoMEX)*, 2017, pp. 1–6.
[124] A. Cserkaszky, P. A. Kara, A. Barsi, and M. G. Martini, "Expert evaluation of a novel light-field visualization format," in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, 2018, pp. 1–4.
[125] C. Perra, W. Song, and A. Liotta, "Effects of light field subsampling on the quality of experience in refocusing applications," in *Proc. Int. Conf. Qual. Multimedia Exper. (QoMEX)*, 2018, pp. 1–3.
[126] I. Viola and T. Ebrahimi, "VALID: Visual quality assessment for light field images dataset," in *Proc. 10th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, 2018, pp. 1–3.
[127] C. Perra, "Assessing the quality of experience in viewing rendered decompressed light fields," *Multimedia Tools Appl.*, vol. 77, no. 16, pp. 21771–21790, 2018.
[128] I. Viola, K. Takahashi, T. Fujii, and T. Ebrahimi, "A comprehensive framework for visual quality assessment of light field tensor displays," *Electron. Imag.*, vol. 2019, no. 1, pp. 1–6, 2019.
[129] R. R. Tamboli, P. A. Kara, A. Cserkaszky, A. Barsi, M. G. Martini, B. Appina, S. S. Channappayya, and S. Jana, "3D objective quality assessment of light field video frames," in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, 2018, pp. 1–4.
[130] P. A. Kara, A. Cserkaszky, M. G. Martini, A. Barsi, L. Bokor, and T. Balogh, "Evaluation of the concept of dynamic adaptive streaming of light field video," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 407–421, Jun. 2018.

[131] S. Baraković and L. Skorin-Kapov, "Survey and challenges of QoE management issues in wireless networks," *J. Comput. Netw. Commun.*, vol. 2013, Dec. 2013, Art. no. 165146.

[132] J. B. Ernst, S. C. Kremer, and J. J. P. C. Rodrigues, "A survey of QoS/QoE mechanisms in heterogeneous wireless networks," *Phys. Commun.*, vol. 13, pp. 61–72, Dec. 2014.

[133] G. M. Su, X. Su, Y. Bai, M. Wang, A. V. Vasilakos, and H. Wang, "QoE in video streaming over wireless networks: Perspectives and research challenges," *Wireless Netw.*, vol. 22, no. 5, pp. 1571–1593, 2016.

[134] M.-D. Cano and F. Cerdan, "Subjective QoE analysis of VoIP applications in a wireless campus environment," *Springer Telecommun. Syst.*, vol. 49, no. 1, pp. 5–15, Jan. 2012.

[135] M. Zhao, X. Gong, J. Liang, W. Wang, X. Que, and S. Cheng, "QoE-driven cross-layer optimization for wireless dynamic adaptive streaming of scalable videos over HTTP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 451–465, Mar. 2015.

[136] J. Li, R. Feng, Z. Liu, W. Sun, and Q. Li, "Modeling QoE of virtual reality video transmission over wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[137] L. R. Jiménez, M. Solera, and M. Toril, "A network-layer QoE model for YouTube live in wireless networks," *IEEE Access*, vol. 7, pp. 70237–70252, 2019.

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

[139] J. Shaikh, M. Fiedler, and D. Collange, "Quality of experience from user and network perspectives," *Ann. Telecommun.*, vol. 65, no. 1, pp. 47–57, 2010.

[140] C. Alberti, D. Renzi, C. Timmerer, C. Mueller, S. Lederer, S. Battista, and M. Mattavelli, "Automated QoE evaluation of dynamic adaptive streaming over HTTP," in *Proc. 5th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, 2013, pp. 58–63.

[141] M. Narwaria, "Toward better statistical validation of machine learning-based multimedia quality estimators," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 446–460, 2018.

[142] A. Rego, A. Canovas, J. M. Jiménez, and J. Lloret, "An intelligent system for video surveillance in IoT environments," *IEEE Access*, vol. 6, pp. 31580–31598, 2018.

[143] A. Martin, J. Egaña, J. Flórez, J. Montalbán, I. G. Olaizola, M. Quartulli, R. Viola, and M. Zorrilla, "Network resource allocation system for qoe-aware delivery of media services in 5G networks," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 561–574, Jun. 2018.

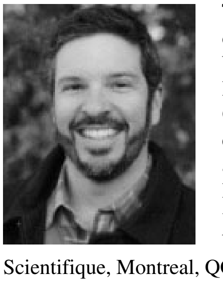
[144] J. Nightingale, P. Salva-Garcia, J. M. A. Calero, and Q. Wang, "5G-QoE: QoE modelling for ultra-HD video streaming in 5G networks," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 621–634, Jun. 2018.

ZAHID AKHTAR (SM'19) received the Ph.D. degree in electronic and computer engineering from the University of Cagliari, Italy. He was a Postdoctoral Fellow with INRS-EMT, University of Quebec, Canada, University of Udine, Italy, Bahcesehir University, Turkey, and the University of Cagliari. He is currently a Research Assistant Professor with the University of Memphis, USA. His research interests include the areas of computer vision and machine learning with applications to biometrics, affect recognition, image and video processing, audio-visual multimedia quality assessment, and cybersecurity. He is also a member of the IEEE Computer and Signal Processing Societies.

KAMRAN SIDDIQUE (M'19) received the Ph.D. degree in computer engineering from Dongguk University, South Korea. He is currently an Assistant Professor with Xiamen University Malaysia. His research interests include cybersecurity, machine learning, and big data processing.

AJITA RATTANI (M'19) received the Ph.D. degree from the University of Cagliari, Italy. She did her postdoctoral Research at Michigan State University. She was an Adjunct Graduate Faculty with the University of Missouri-Kansas City. She has been an Assistant Professor with the Department of Electrical Engineering and Computer Science, Wichita State University, since 2019. She is a Co-Editor of the Springer books: *Adaptive Biometric Systems: Recent Advances and Challenges* and *Selfie Biometrics: Advances and Challenges*. Her research interests include the fields of biometrics, machine learning, deep learning, image processing, and computer vision.

SYAHEERAH LEBAI LUTFI received the Ph.D. degree from the Universidad Politécnica de Madrid, Spain, in 2013. She is currently the Manager of Industrial Training, Student Activities at the School of Computer Sciences, Universiti Sains Malaysia. Her general research interest includes human–computer interactions, specifically focusing on affective computing (behavior analytics, personality, culture, mood, and emotion analysis and modeling).



TIAGO H. FALK (SM'14) received the B.Sc. degree from the Federal University of Pernambuco, Recife, Brazil, in 2002, and the M.Sc. and Ph.D. degrees from Queens University, Kingston, ON, Canada, in 2005 and 2008, respectively, all in electrical engineering. From 2009 to 2010, he was an NSERC Postdoctoral Fellow with the Holland-Bloorview Kids Rehabilitation Hospital, affiliated with the University of Toronto. Since 2010, he has been with the Institut National de la Recherche Scientifique, Montreal, QC, Canada, where he heads the Multimodal Signal Analysis and Enhancement Laboratory. His research interests include multimedia/biomedical signal analysis and enhancement, pattern recognition, and their interplay in the development of biologically inspired technologies.