How passive image viewers became active multimedia users

New trends and recent advances in subjective assessment of Quality of Experience

Judith A. Redi, Yi Zhu, Huib de Ridder, Ingrid Heynderickx

Abstract Subjective assessment of Quality of Experience (QoE) is key to understanding user preferences with respect to multimedia fruition. As such, it is a necessary step to multimedia delivery optimization, since QoE needs to take into account technology limitations as well as user satisfaction. The study of QoE appreciation dates back to the twentieth century, when it exploded with the advent of CRT first and LCD displays later. For a long time, this branch of research was targeted at determining user sensitivity to impairments induced in the media by suboptimal delivery. The media recipient was considered a passive observer, whose appreciation of the video material was determined primarily by the degree of annoyance due to the impairments affecting it. With the advent of mobile technology and internet-based media delivery, this impairment-centric concept of QoE has shown to be incomplete. The media recipient became an active user who creates content, interacts with the system and selects the media he/she wants to have delivered. As a result, elements such as visual semantics, user personality, preferences and intent, social and environmental context of media fruition also concur to the final experience assessment. The role played by these elements in QoE, and the cognitive/affective processes that underlie them are still to be understood, although several models of QoE apprecia-

Judith Redi
Dept. Intelligent Systems, Delft University of Technology, The Netherlands e-mail: j.a.redi@tudelft.nl

Yi Zhu
Dept. Intelligent Systems, Delft University of Technology, The Netherlands e-mail: Y.Zhu-1@tudelft.nl

Huib de Ridder
Dept. Industrial Design, Delft University of Technology, The Netherlands e-mail: H.deRidder@tudelft.nl

Ingrid Heynderickx
Dept Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, The Netherlands and Philips Research Laboratories, Eindhoven, The Netherlands e-mail: I.E.J.Heynderickx@tue.nl

tion have already been proposed. In this paper, we review the evolution of subjective QoE assessment and models from the impairment-centric approach to a more user-centric approach. We analyze relevant features and factors influencing QoE, and point out future directions for subjective QoE assessment research.

1 Introduction

Billions of digital images and videos are produced, broadcasted, shared, and enjoyed by users every day. Especially with the advent of internet-based image and video delivery, the amount of multimedia content consumed every day has dramatically increased [24], and will continue to grow in the foreseeable future. This enormous amount of information needs to be handled (i.e., captured, stored, transmitted, retrieved and delivered) in a way that meets the end-users' expectations. However, technology still shows limitations, such as limited spatial, temporal and bit rate resolution in displays, bandwidth and storage constraints introducing compression related artifacts, or error-prone transmission channels resulting in network related artifacts. As a result, multimedia material is often delivered affected by impairments which disrupt the overall appearance of the visual content. Impairments provoke a sense of dissatisfaction in the user [175, 129, 54], which, in turn, may decrease the willingness to pay for/use the multimedia application, service, or device [161].

That's why, in the last three decades, a lot of effort has been devoted to the development of technologies that can either prevent the appearance of impairments, or repair for it when needed. Following initial attempts based on the quantification of signal errors [53], it became soon clear that a better understanding of how humans experience images and videos was necessary to properly optimize media delivery. As a result, multimedia delivery optimization was researched from its early days through collaboration between engineers and vision scientists. In fact, this community can be considered a pioneer in user-centered multimedia design and engineering (for an accurate historical overview, see [21]). Within this effort, dedicated psychometric techniques were developed [3, 81, 40] and standardized [138, 73, 137, 82] to support a reliable quantification of visual quality (i.e., the perceived overall degree of excellence of the image, [40]) from a subjective point of view. With these techniques a large body of psychophysical data was collected to unveil the perceptual functions of the human visual system (HVS) that regulate the sensitivity to impairments. The outcome of these experiments served as inspiration for designing objective visual quality assessment metrics [105, 61], whose output would steer then impairment concealment (i.e. image/video restoration) and technology tuning.

It is interesting to point out that the common, underlying assumption for those studies is that having an understanding (possibly a model) of the perceptual processes that regulate impairment sensitivity suffices to predict the impairments' annoyance. In practice, being able to measure impairment sensitivity is considered to be substantially equivalent to predicting the overall quality of the viewing experience. This impairment-centric definition of visual quality (also referred to as perceptual quality in the following) has yielded remarkable results [105, 61]. Still, large room for improvement exists [114, 144]. Furthermore, new imaging and media technologies are challenging this impairment-centric notion of visual quality. Visual media are nowadays consumed in more and more immersive contexts (e.g., 3DTV, virtual and augmented reality) or in social, interactive and customizable contexts (e.g., social media, video on demand, mobile). The judge of the visual experience cannot be regarded as a mere passive observer anymore, but rather as an active user interacting with the systems on the basis of specific expectations from them. In such a scenario, impairment sensitivity cannot be expected to be the sole factor contributing to the final user satisfaction on viewing experience.

In fact, several models have been proposed during the last decade that attempted at expanding the concept of visual quality to a more encompassing idea of Quality of the (Viewing) Experience (QoE) [81, 144, 51, 128, 133, 149, 99]. In general, QoE is defined as a multidimensional quantity, depending on a number of attributes or features (i.e., quantifiable properties of the viewing experience, such as blockiness, aesthetic appeal, subject uniqueness), which are not necessarily mutually independent. Of these features, only a subset addresses traditional impairment sensitivity issues; others take into account rather cognitive and affective aspects of the experience. Attributes of the experience can be in turn influenced by external factors (i.e., factors independent of the media visualization) such as context of usage, user background, personality or task. Indeed, it has been recently shown that elements such as context of fruition [79] or user affective state [180] have an impact on visual quality appreciation, actually compensating in some cases for visual impairments. For example, football fans were shown to be highly tolerant to low frame-rates, as long as they were watching a football video [126].

Unfortunately, despite a working framework for QoE seems to be established, neither agreement has been reached on a precise taxonomy of attributes and external factors, nor much knowledge has been developed on how these quantities are interrelated. As a result, more subjective studies are needed to unveil interdependencies of QoE attributes and external factors, towards defining a precise model of how these elements concur to the final QoE judgment. In addition, integration with qualitative and quantitative user study techniques developed for other fields (e.g., human computer interaction) and including existing results from image psychology [47] are needed to fully characterize visual experiences.

In the following, we review the steps that led, throughout the last few decades, to the evolution of the concept of visual quality (typically, impairment-centric) into that of Quality of Experience. We first summarize the research done to quantify visual quality and impairment acceptability in the fields of display, signal processing and network optimization (Section 2). We then review in Section 3 the models that over the years have attempted at extending the impairment-centric conception of visual quality, finally converging into an operative definition of Quality of Experience, which takes into account also the influence of external factors on the final user satisfaction. The existing knowledge on these factors and their impact on QoE is summarized in Section 4. Finally, Section 5 outlines new trends in subjective assessment of QoE, discussing in more detail QoE of immersive imaging technologies (such as stereoscopic displays), unveiling the role of affective processes in QoE judgments, and pointing out a methodological shift from lab-based to real-world- and crowd-based subjective experiments.

2 Subjective Assessment of Visual Quality

Within the past decades, visual impairments produced by technological limitations (e.g., lossy compression, sub-optimal pixel size, unreliable network transmissions) have been for long considered the principal cause of user dissatisfaction with multimedia systems; as a result, subjective assessment studies have mainly focused on quantifying the annoyance of such impairments as a function of technology variables.

A framework that supported this research was Engeldrum's Image Quality Circle (IQC), depicted in Figure 1 [40]. Such framework aimed at providing an effective methodology for linking experienced visual quality to the setting of technological variables of a multimedia system. In the case of displays, technological variables of interest were, for example, pixel size, colour filter thickness, driving voltages, etc; when dealing with processing algorithms (e.g., compression or sharpness enhancement), such variables could be identified as the relevant parameters in these algorithms; in the field of network optimization for video streaming, bandwidth allocation was a main technological constraint. By varying the setting of technological variables, the eventual quality of the media delivery is affected. As a result, a holy grail for the multimedia community was (and still is) to infer relationships that, given a change in technological variables, accurately predict the experienced quality of the delivered media. In introducing the IQC, Engeldrum [40] argued that this problem was ill-posed: it aimed at linking a multi-dimensional description of a system (through its technological variables) to a one-dimensional overall quality preference. This relationship is not unique to begin with, and unveiling it requires an almost endless trial-and-error approach (i.e., a subjective test for every single change in a technological variable, which is costly and ineffective). Thus, rather than directly modeling the relationship between the technological variables and overall quality preferences, the IQC proposes a divide-and-conquer approach, involving three intermediate steps: (1) linking overall image quality to the (often unconsciously weighted) combination of underlying perceived attributes of the image, (2) linking each image attribute to the physical characteristics of the system output, and (3) linking the physical description of the system output to the system technological variables. By defining these three intermediate relationships, simulation and more accurate prediction of the effect that variations in technological variables have on the eventual perceived visual quality are allowed, limiting the need for subjective testing during system development.

Initially designed for display optimization, it is possible to adapt the IQC framework to the type of multimedia system under consideration (Figure 2). Although rarely applied in practice, the IQC building blocks can be easily translated from a

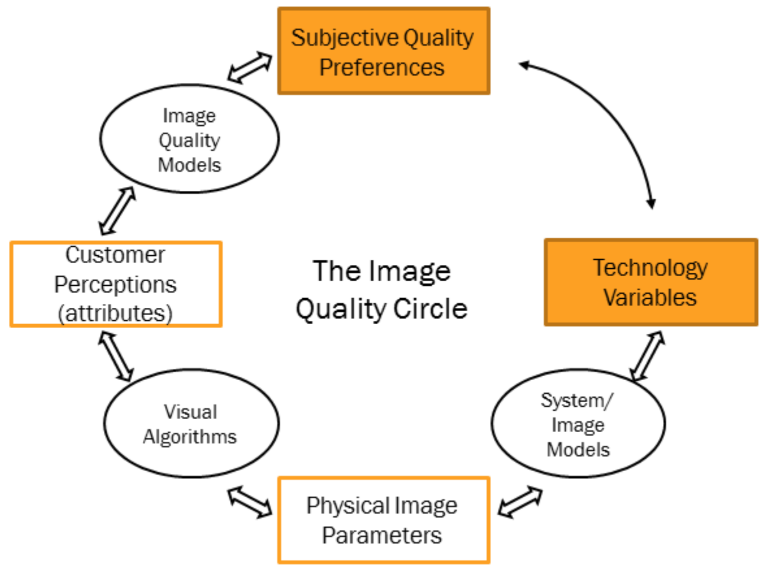

Fig. 1 Engeldrum's Image Quality Circle

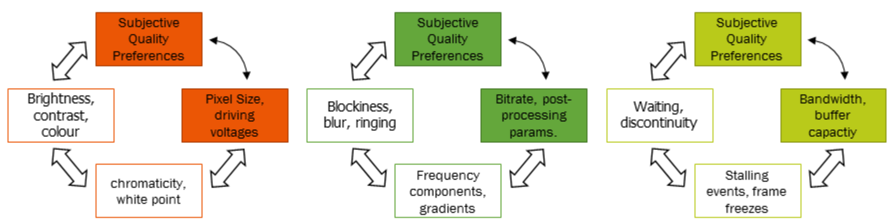

Fig. 2 Engeldrum's Image Quality Circle adapted to different problem domains: display quality assessment (a), signal processing algorithm optimization (b) and network parameter optimization (c)

display context to a signal or network context. Indeed, also technological signal variables can be related to physical characteristics of the light output of the multimedia system, resulting in partly different (e.g., compression artefacts such as blockiness and ringing) and partly similar (e.g., blur and noise) perceptual attributes, yielding an overall quality preference. Similarly, in a network context, physical media characteristics can be identified from technological Quality of Service parameters (e.g., packet-loss ratio or number of stalling events), which result in attributes (such as jerkiness or perceived waiting time for the video to load/progress) that are weighted towards an overall quality preference. In the following, we review the efforts done within the display, signal processing and networking communities to unveil relationships between technological variables, physical output characteristics, related features or attributes and the eventual user preference in terms of visual quality.

2.1 Visual Quality Preference in Displays

The impact of display related artifacts on the experience of viewers has commonly been associated with the concept of image quality, being a well-recognized concept to consumers [40], who considered it the key driver in selecting one device over the competitor's. The RaPID method first [9] and the IQC later constituted a solid framework to improve display quality. When applied to displays, a few issues needed to be addressed to make the IQC framework useful in practice: first of all, relevant image quality attributes, i.e., specific perceptual characteristics of the image, needed to be defined; then, a strategy to combine these attributes in a single overall quality measure needed to be determined.

With respect to the first issue, the RaPID method, more than the IQC model, contributed to determining unambiguous descriptions for the image perceptual attributes. These descriptions arose from a two-step procedure. First, expert (trained) viewers would discuss in team sessions which attributes characterized the image quality for a given set of images, the meaning (appearance) of these attributes and a meaningful way to quantify them. Naive viewers (a sample of standard customers) were then requested to score the image quality of the same set of stimuli. The relation with image quality was then established by means of a multivariate analysis and regression of the expert-defined attributes onto the overall quality scores. Applied to spatial resolution scaling on LCD monitors, for example, this method revealed that within this context image quality was a weighted sum of perceived blur, perceived pixellation (i.e., the fact that blocks of pixels were visible on diagonal lines), aliasing (geometrical deformations of spatially high frequency patterns), artifacts in letters (i.e., missing parts in letters/text) and increased ringing visibility [171]. An alternative approach to establish the attributes, focusing particularly on naive viewers, was described as the Interpretation Based Quality (IBQ) method [134]. This approach combined qualitative (i.e., free image sorting and interviewing) and quantitative (i.e., magnitude estimation) methodologies, and as such allowed establishing the relationship between subjective preferences and the underlying features/attributes. The method has especially added value in detecting content and context dependency in high-quality images.

The second crucial issue is the combination of judgments of perceptual attributes into a single image quality value. Historically, several attempts have been made at creating a model of quality judgment based on assessment of image degradations and impairments. Allnatt's "Law of Subjective Addition" [3] was the first major achievement in this sense. This "law" stated that impairment annoyance was inversely proportional to image quality and that distinct impairments added up when they occurred simultaneously, to then linearly map into image quality. Minkowski metrics were later proved to be a better way of combining simultaneously occurring impairments into one overall impairment score [29, 30]. This relationship was shown to hold also for extreme levels of impairment (i.e., such that the image content was almost unrecognizable) [193], as in the case of overlapping blur due to low-pass filtering using a 2D separable binomial filter [120] and noisiness due to normally-distributed spatial noise. Finally, in a refined version of the Minkowski metrics accommodating an upper bound for impairment [30] it was shown that setting the exponent to approximately 2 yielded a good description of both quality and impairment judgments. In other words, it was concluded that the accumulation of perceptually distinguishable impairments could be described by a vector-sum model. As a result, the Minkowski metric provided a valuable tool for combining perceived display quality attributes into overall image quality [121, 41].

de Ridder et al. [149] pointed out later that, although successfully combining annoyance of different artifacts into an overall quality score, Minkowski metrics assumed image quality preference to be context and application independent. To verify whether this assumption was correct, they suggested to consider image quality as an indicator of the degree to which an observer could exploit images, i.e., to regard images "... not as signals but instead as carriers of information..." [74]. This information-processing oriented definition assessed the degree of identifiability (i.e., the naturalness constraint) and discriminability (i.e., usefulness constraint) of the elements in the image. Fulfillment of both requirements would ensure high quality of the image; nevertheless, the degree to which constraints were expected to be fulfilled was highly content and task dependent [75]. In a series of experiments aimed at establishing color rendering preference in images, Janssen et al. manipulated the color characteristics of natural images, and then asked a pool of observers to judge their image quality [75, 33, 31]. Interestingly, image quality (Q) turned out to be a weighted sum of perceived naturalness (N) and colorfulness (C), or

$$Q = w \times N + (1 - w) \times C \tag{1}$$

where w is a weighting factor between zero and one. Moreover, it was observed that, when varying color contrast, participants showed a clear preference for more colorful, yet slightly unnatural images. Similar observations were made in a later study for the perceived quality of stereoscopic images where, under certain conditions, the quality of depth was found to be a weighted sum of naturalness of depth and perceived strength of depth [71, 72].

The implications of the findings above were that (1) image quality is typically judged based on a comparison of the experienced quality to an internal 'reference' image, and that (2) this internal reference is not necessarily the most realistic one (i.e., a high fidelity reproduction of reality) but is rather influenced by other factors such as memory, content and context of usage. These conclusions were later integrated in the so-called FUN model [149, 39]. In essence, this model assumes the existence of three major constraints determining image quality: Fidelity (i.e., degree of apparent match with an external reference, e.g., an original), Usefulness (i.e., degree of visibility of details) and Naturalness (i.e., degree of apparent match with an internal reference, e.g., memory colors). Overall image quality is then modeled as a weighted sum of the three constraints whereby the weighting depends on task, context, image content etc. In fact, different people, different types of images and different tasks may require different combinations of these weights, which implies that there is no single standard criterion for image quality, nor absolute perceptual preferences. It is interesting to note that this conclusion fits remarkably well with the ideas behind the so-called interface theory of human perception [88, 63], which states that perception is not about accurately reconstructing the physical world, but about constructing the properties and categories of an organism's perceptual world. Hoffman [63] argues that these perceptual structures are not intended to accurately match the physical world but, instead, are fast, intention-driven explorations of the meaningless physical world in preparation of "optically guided potential behavior", thus striving for utility and efficiency, not veridicality.

2.2 Visual Quality Preference of Processed Signals

The signal processing community evolved in many aspects rather distinct from the display community, and until recently its researchers largely focused on a signal fidelity approach. The goal of this branch of subjective studies was to understand the impact on perceived quality of a specific type of processing algorithm (e.g. compression, scaling, de-noising and so on) towards identifying the optimized setting of the algorithm's parameters to produce the best visual quality.

In 1987, Watson [179] made an important distinction between perceptually lossless and perceptually lossy image coding, thus acknowledging the relevance of understanding and modelling the impact of coding artifacts on perceived image quality. Research into image integrity [145] or artifactual quality [81] referring to the relation between coding artifacts and quality preference, became an important focus of the image and video processing community (for a thorough overview, see [21]). The approach consisted of incorporating models of low-level features of the human visual system (HVS) into image quality metrics. Subjective studies were therefore aimed at generating ground-truth data, and with that determining which HVS mechanisms were triggered by the appearance of impairments, leading to the identification and modeling of e.g. contrast [160, 60, 8] and luminance masking mechanisms [127, 106], spatial pooling strategies [178], and image structure perception [177, 48]. Temporal and movement effects on artifact visibility have also been studied [122, 164]. Furthermore, initial attempts to understand the perceptual impact of overlapping video signal impairments (i.e., co-presence of e.g., blur, blockiness and noise) were carried out by Farias et al. [45], concluding Minkowski metrics were a powerful modeling tool in this context, as already proven for displays [44].

The large body of work done on artifact visibility estimation was inspired by the idea that the HVS remained constant over time [21], i.e., despite personal preferences, our visual processing strategies have barely evolved over the course of human history, thus it should be possible to model them in a meaningful and objective way (that is, independent of individual subject differences) such that HVS-based models could accurately predict and describe image quality. Interestingly, this soon turned out to be not true, even at threshold level. In an experiment on visibility of compression artifacts [182] it was shown that non-expert observers were less sensitive to compression-related artifacts (i.e., blockiness, blur and quantization noise) than trained observers with detailed knowledge of the algorithm employed. Even more interestingly, it was shown that subjects who actually developed the algorithm were very sensitive to artifact visibility only in the images that were used within the algorithm design and test phases. Apparently, the well-informed experts knew exactly where to look for the impairments, a finding that cannot be accommodated by most of the low-level HVS-based models.

An initial attempt at studying the role of higher-level HVS features in signal impairment annoyance and related quality appreciation targeted visual attention mechanisms [42, 143]. When observing a scene, the human eye optimizes the information acquisition by focusing on specific, meaningful areas of the scene, and neglecting poorly informative areas [35]. As a consequence, it was hypothesized that signal impairments located in the visually attractive areas of an image were more likely to be noticed during the visual experience, resulting in a more negative judgment of visual quality (or higher annoyance). Evidence of this has been provided for e.g. blocking artifacts in images [2]. As a consequence, the interplay between visual attention and visual quality assessment mechanisms was thoroughly studied. Research showed that the visual quality assessment task had a significant impact on visual attention deployment [123], which was also found to be the case for image aesthetic appeal assessment [145]. These findings stressed the importance of having control, task-free eye-tracking recordings to fairly evaluate the impact of signal impairment appearance on visual attention, and the impact of visual attention on the eventual quality judgment. Several eye-tracking studies reported information in this sense, yet without a clear consensus. In the work by Vuori and others [174] the quality of the judged image was shown to have an impact on the saccades' duration. In [35] the authors showed that saliency maps of pristine images obtained from free-looking eye-tracking data were poorly correlated to the maps derived from the image quality scoring of slightly impaired versions of the same images. This correlation was shown to increase with the amount of impairment visible in the images, and to be independent on the type of signal impairment. Vu et al. [173] identified instead an effect of the type of signal impairment (i.e., blur, compression or noise) on the location of the fixations while scoring, though without quantifying it. As far as videos are concerned, Le Meur et al. [100] found that the quality evaluation task had a more limited impact in the video domain than in the image domain. Later, though, Mantel et al. [111] showed that the strength of signal impairments had an impact on the dispersion of the fixations (i.e., increasing with decreasing video quality) and was positively correlated with the duration of the fixations [111].

Despite the diversity of the abovementioned results, the study of visual attention in relation to signal impairment annoyance enabled the design of a wide range of image and video quality metrics, enhanced with either saliency or visual importance data (for a complete overview, see [42]). The added value of incorporating such information in quality metrics was clearly shown for images [107, 142], whereas it was found to be less relevant for video [42]. Furthermore, this activity produced an abundance of subjective data, most of which have been made publicly available for further research [184]. These data may be precious in further understanding the role of high-level HVS mechanisms in viewing experience appreciation.

2.3 Subjective Assessment of Network-Related Impairments.

Nowadays quality of the (broadband) broadcasted or stored video content and of the displays used for their rendering is in most circumstances of such a high level that naive consumers hardly see improvements. The latter, however, is not yet true for multimedia content distributed over (mobile) IP networks. Bandwidth limitations, along with network unreliability (i.e., the possibility of losing parts of the streamed signal/packets) can cause impairments during visualization of the image/video, including frame freezes, deformations of the spatial and temporal structure of the content, and long stalling times.

Rather than based on subjective assessment of visual quality, network parameters have been for long optimized towards keeping an acceptable Quality of Service (QoS), by taking into account parameters such as packet loss ratio, delay, jitter and available bandwidth [153], as well as video QoS parameters, such as buffering time and buffer ratio [5]. Lately, researchers have been aiming at correlating the QoS parameters to Quality of Experience (QoE) measurements (typically, identified once again with visual quality subjective ratings [138]) by using fitting functions [49, 158, 84]. In general, low QoS performance leads to low QoE [70]. For example, it has been shown that the buffer ratio (i.e., the fraction of time spent in buffering over the total session time, including playing plus buffering) consistently had a high impact on user QoE [37]. Reduced buffering times resulted in higher user satisfaction. Similar conclusions were found for other QoS parameters, such as the join time in multicast video delivery, the buffering duration, the rate of buffering events, the average bit-rate and the packet loss rate [70, 113].

In general, QoS metrics succeed in estimating QoE from a network efficiency point of view, but they do not necessarily reflect the overall viewing experience. In fact, QoS parameters fail in capturing all subjective aspects associated with the viewing experience [34, 131]. Note that typically QoS parameters are computed based on the encoded bit-stream, whereas no pixel information is analyzed; therefore, the impact of signal impairments such as blockiness and blur (see Section 2.2) is not taken into account in these approaches. In the case of packet loss, for example, it was found that the same packet loss ratio yielded more or less annoyance depending on the video encoding and video content [162]. The loss of bit-stream packets indeed can result in specific, spatiotemporal visual impairments, due to the poor concealment of the lost packet at the bit-stream decoder side. This type of impairments is more or less noticeable depending on the amount of movement in the video and can be annoying [140, 162] even more so when in combination with strong compression artifacts [80]. Furthermore, when studied in conjunction with visual attention, packet loss artifacts have been shown to be more annoying when located in visually important regions of the image [43], and to have a high potential for becoming salient, then altering the natural visual attention deployment [140]. Quite interestingly, the entity of this alteration has been shown to be negatively correlated with the perceived visual quality of the video [140].

3 From Visual Quality to Quality of (Viewing) Experience

As pointed out in Section 2, subjective studies from different communities (displays, signal processing and networking) converged eventually towards a similar conclusion: quantifying impairment sensitivity, even by means of accurate HVS models, is necessary yet not sufficient to quantify the overall quality of the viewing experience. In fact, a few models have been proposed throughout the last decade to extend the impairment-centric notion of visual quality to a broader, more representative concept of quality of the viewing experience.

Keelan [81] defined visual quality as a multidimensional quantity evolving along a number of visual attributes, comparable to Engeldrum's IQC attributes. Keelan distinguished four different families of attributes: artifactual (e.g., blockiness and blurriness), preferential (e.g., brightness and contrast), aesthetic (e.g. symmetry or harmony [46]) and personal (e.g., user emotional connection and engagement with the visual content [90, 180]). Of those, the first two were highly related to perceptual quality, whereas the latter two would contribute to the visual quality assessment by taking into account more implicit experiences of the viewer [110]. Because of this, aesthetic and personal attributes were considered "too subjective" and "unlikely to yield to an objective description" of an image, leading Keelan to the conclusion that their quantification would be "too cumbersome and expensive to use for routine image quality research" ([81], p. 6). As a result, Keelan privileged the investigation of artifactual and preferential attributes, leaving unexplained the contribution of personal and aesthetic attributes to the overall visual quality.

Ghinea and Thomas also attempted at reaching a more encompassing definition of visual quality by proposing the concept of Quality of Perception (QoP) [51]. Their reasoning started from the assumption that multimedia are primarily consumed for infotainment; therefore, viewing experience has a twofold purpose: that of transferring information to the user, and that of granting a sufficiently high level of satisfaction in terms of entertainment. To properly optimize viewing experience, then, both (1) the level of Information Assimilation (QoP-IA) and (2) the overall user satisfaction with respect to the media presentation (QoP-S) should be taken into account. QoP-IA represents the level of the user's understanding of the media content. It is typically measured as the performance (in terms of number of correct responses) on a questionnaire about the (semantic) content of the viewed media. QoP-S depends instead on two elements. The subjective Level of Quality (QoP-LoQ) measures the perceptual impact of losses on visual quality (e.g., due to the appearance of impairments), independent of the media content. The Level of Enjoyment (QoP-LoE) measures instead the overall enjoyability of the media presentation, taking into account also cognitive and affective aspects of the visual experience, such as watchability, ease of understanding and level of interest in the subject matter. Throughout multiple studies [105, 57, 58] it was found that low QoP-LoQ did not impact the level of information assimilation (i.e., despite the presence of impairment, users could still fully understand the content of the media) and had limited impact on the level of enjoyment (QoP-LoE), indicating that other elements compensated for artifact appearance in viewing experience. Unfortunately, up to date, there is little known and studied on these elements that compensate for artifact visibility in overall viewing experience: it is not known yet which are these elements and how they contribute to the eventual experience appreciation.

From a different perspective, Pereira [128] proposed a three-level model for visual experience appreciation. In Pereira's model, visual experience is first evaluated at the sensorial level, which responds to the purely physical properties of the media (i.e., comparable to the IQC physical image characteristics and to some extent to Keelan's artifactual and preferential attributes). This level of evaluation contributes therefore to the first perceptual quality impression (a concept similar to QoP-LoQ [51]). Next, the viewing experience is evaluated at the "perceptual" level. Here, the media content and the potential for creating knowledge out of it are assessed (with similarities to QoP-IA in [51]). Note that the word "perceptual" here entails also cognitive processes such as content recognition and interpretation, and is not to be confused with the classic notion of perceptual quality, which in this framework is addressed at the sensorial level. Finally, the viewing experience is assessed at the emotional level, where the way in which the media impacts the user's affective state is evaluated. Pereira suggests that viewing experience appreciation results from a linear combination of a quantification of these three aspects; however, he did not provide an empirical validation of this hypothesis. Pereira's model was further extended in [133] to assess augmented reality experiences. The extended model accounted also for implicit experiences of the user (such as cultural background), and the context and goal of the viewing experience. As a result, on top of the three levels in the model of Pereira [128], the authors of [133] suggest to take into account both usability of the multimedia system and ethnographical assessment to obtain an accurate measure of the quality of the augmented reality experience.

The FUN model of de Ridder and Endrikhovski [149], already mentioned in Section 2, can also be considered a milestone in the road that took visual quality to evolve into Quality of Experience. The model is the first to introduce the concept of finality of usage of media (and user motivation for having the viewing experience), and to suggest that viewing experience cannot be quantified without taking this concept into account. The quality of a viewing experience, indeed, should depend on the degree to which the visual information can be successfully exploited by the user towards his/her goal. This in turn is quantified in terms of the fulfillment of the Fidelity, Usefulness and Naturalness criteria already described in Section 2.2. Whereas the Fidelity criterion can be to a large extent equated to the assessment of impairment sensitivity, the Usefulness and Naturalness criteria introduce two rather new concepts in QoE evaluation. The Usefulness constraint indicates the maximum discriminability of perceived items in the image (or video); thus, the degree to which this criterion should be fulfilled is highly application and task-dependent, as for example, the fulfillment threshold for Usefulness of a consumer display is different from that of a microscope. The Naturalness constraint refers instead to the fidelity of the media to what the authors call an ”internal-reference,” or an internal representation of how the media ”should look like”. Here, previous (quality) experiences and expectations come into play. Thus, the fulfillment threshold for Naturalness is intrinsically user-dependent. The paradigm shift in this model lays in the fact that constraints are not anymore assumed to be universal, but rather application and user dependent. Hence, factors external to the viewing experience (i.e., not directly related to vision), have an impact on its appreciation, and should be studied in relation to it.

This idea has been recently picked up by the Qualinet consortium, which has proposed a rather encompassing model for Quality of multimedia Experience [99]. Note that the experience here is not limited to vision, but is multisensory, thus it is not necessarily related to imaging systems only. In fact, the Qualinet model combines elements of all models described above: to begin with, QoE is described as a multidimensional quality, that can be decomposed in a set of perceptual attributes called features. QoE features are defined as ”perceivable, recognized and namable characteristics of the individual’s experience of a [multimedia] service which contributes to its quality” and can be classified into four categories: features at the level of perception, at the level of interaction, at the level of usage and at the level of service. The features at the perceptual level entail experience characteristics that can be evaluated from immediate perception (e.g., blockiness, blurriness, brightness and contrast). The features at the interaction level account for human-technology interaction aspects of the experience (e.g., responsiveness and communication efficiency between the user and the multimedia system). Features at the level of usage assess the accessibility and the stability of the service during usage. Finally, the long term characteristics of the service beyond the single instance of usage, such as ergonomics, usability, and ease of use, are accounted for by the level of service, similarly to what was suggested in [133]. All these features are assumed not to be independent. Features appreciation is in turn mediated by a set of interrelated quantities called Influence Factors (IF). Influence Factors are defined as ”characteristics of a user, system, service, application, or context whose actual state or setting may have influence on the Quality of Experience for the user”. As such, they pre-exist the fruition of the media; nevertheless, they condition the final user satisfaction. IFs can be grouped into three categories, depending on whether they represent properties of the user, of the system (or application or service) or of the context of usage. User IFs entail characteristics of the user such as demographics, personality or emotional state, and can condition both the appreciation of technical quality (thus modulating the features of the level of perception) and that of the overall experience, also impacting on the interpretation and understanding of the media content. System IFs are those properties of the multimedia system/service that are responsible for the resulting technical quality: media encoding configuration, network parameters, display functions, etc.. Finally, context IFs encompass all situational properties of the environment in which the experience takes place. Examples of context IFs are location and space, time of the day, task and social context. For a detailed overview of known influencing factors of QoE, see Section 4.

Similarly to the FUN model [149], the Qualinet model [99] also assumes the existence of an internal 'reference' experience to which the real one is compared. All QoE features have an internal reference value that is modulated by IFs; the extent to which the features of the current experience match the reference ones builds the eventual user satisfaction. Although mapping the interactions of (reference and current) QoE features and influence factors is still beyond reach, a simplified representation of the model is attempted in Figure 3. In this figure, the model is visualized as a network of computing units that modulate the assessment of the difference between the reference and current (i.e., "quality") feature values. Influencing factors not only determine the value of the "quality" (experienced) features, but also modulate the importance that the difference between their value and that of the internal reference has on the final quality judgment (fusion module). To draw a parallel with the models reviewed so far, we can consider the level of perception similar to the artifactual attributes in [81], the QoP-LoQ in [51] and the Fidelity dimension in [149]; they all consider the impact of system IFs on perceptual features (green unit in the figure). The emotional level of [128] could result from the impact of Human and Context IFs on the level of perception and the level of service features. Finally, the Usefulness dimension in [149] could be intended as the result of context IFs on the level of perception features. Some of these interactions are further explored in Section 4, but we first want to clarify an operative definition of Quality of Experience that we will adopt throughout the rest of this chapter.

3.1 Definition of Quality of Experience

The concept of Quality of Experience (QoE) arose from the field of Telecommunication Engineering. In the past decades, the effectiveness of communication services was linked to the notion of Quality of Service (QoS), which is defined as the "totality of characteristics of a telecommunication service that bears on its ability to satisfy stated and implied needs of the user of the service" [139]. QoS is mainly operationalized in terms of system and network performance-related measures (e.g., packet loss ratio, jitter or delay). However, with the booming of online multimedia services, the notion of QoS has started showing its limitations, and was found to be poorly correlated to user satisfaction. As a result, the QoE concept emerged, and was initially defined by ITU [167] as "the overall acceptability of an application or service, as perceived subjectively by an end-user". This definition suggests that the scope of QoE has shifted from a rather narrow perspective of telecommunication systems to a broader perspective of multimedia services. Furthermore, this definition not only takes the complete end-to-end system in consideration to define QoE, but also includes the user's expectations and his/her context. Recently, the Qua-

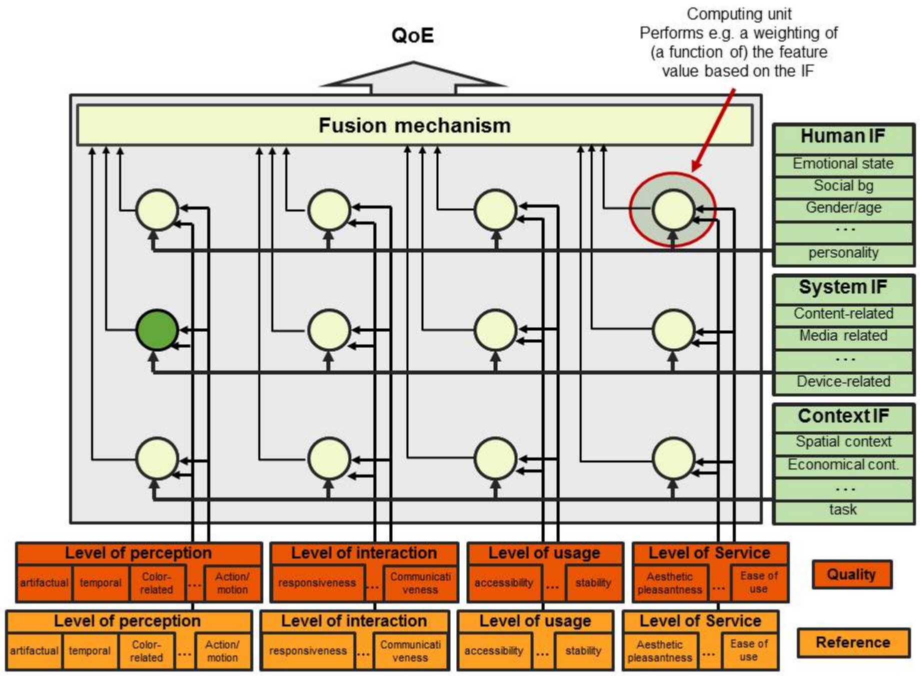

Fig. 3 Schematic representation of the Qualinet model for Quality of Experience [99, 144], fully connected into a network for the final QoE prediction. The green computing unit covers the influence of system factors (technology variables [81]) on the features of the level of perception; most research in the QoE domain has been focusing so far on this specific facet of QoE.

linet White Paper [99] proposed a more explicit definition of Quality of Experience: ”Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user’s personality and current state”. This second definition explicitly refers to the concepts of ”personality”, entailing long-term traits of the user such as feelings, thinking attitude and behavior (as per [130]), as well as user ”current state”, i.e., the punctual set of feelings, thoughts and behavior contextual to the viewing experience [99]. Note that the current state is both an influencing factor of QoE and a consequence of the experience. Although both definitions describe a similar phenomenon, the definition given by Qualinet seems to be more complete than the one of ITU-T. In the ITU-T definition, QoE is related to acceptability in terms of the ”characteristic of a service describing how readily a person will use the service”. The Qualinet definition, instead, emphasizes that human factors, such as personality and current state, may significantly influence QoE. Given the evidence of the importance of such factors in properly estimating user satisfaction and QoE (which will be documented in Section 4 of this paper), we adopt the Qualinet definition as operational definition of QoE in the remainder of this chapter. Along with this operative definition, it is worth mentioning a few other concepts that closely relate to QoE. The term engagement, for example, often refers to positive aspects of user experience. Attfield [6] gives a definition of engagement as "a quality of the user experience that emphasizes the positive aspects of interaction - in particular the fact of being captivated by a resource". In prior research, engagement was described as the experience of a user who highly focused on the video and was affectively involved with it [153]. Studies showed that engagement played a crucial role in determining user satisfaction [37]. As a result, in Section 4 we also refer to literature on engagement with multimedia content to complement the knowledge existing on factors influencing QoE. Finally, it is interesting to relate QoE to the concept of endurability. The term endurability has been used to describe the consequence of a satisfactory experience, and specifically the likelihood of remembering it and the willingness to repeat it or recommend it [124]. Read et al. [136] studied endurability in children's satisfaction with tourist attractions. They organized a group of 45 children going on a school trip to a themed tourist attraction, where they could engage in nine activities. Children scored each activity before and after performing it, and ticked "yes, maybe, no" in response to the question "Would you like to do it again?". A week after the task, children were asked to recall the separate activities that had made up the event. They were also asked to name the activity that they had liked the best. The results showed that 81% of the children recalled the activity that they had previously identified as worth repeating. This suggests that high QoE leads to endurability [124]: people remember enjoyable, useful, engaging experiences and want to repeat them.

4 Influencing Factors of Quality of Experience

As introduced in Section 3, Quality of Experience is a multifaceted quality, resulting from the interaction of multiple influencing factors. Besides factors that have been proven to influence QoE in the context of video, we also review elements of the experience found to be relevant in other research fields (e.g., psychology of gaming experience). It is important to remark that most of these factors are not independent. They may interact with each other, and as such, influence QoE in a complex way (see also Figure 3). Following the model proposed in [99], the factors are arranged into three categories (shown in 1), namely System factors, User factors and Contextual factors. Each group of factors is described in more detail in the remainder of this section.

Table 1 Factors influencing QoE discussed in this chapter

System factors (4.1)	User factors (4.2)	Contextual factors (4.3)
Devices [153, 86, 85, 87]	Interest [110, 124, 90, 126, 159, 67, 104]	Physical environment [161, 183]
Signal and network variables [175, 194, 56, 1, 103]	Personality [180, 26]	Economic conditions [14, 83, 190]
	Age/gender [185, 118, 13, 117, 69, 12]	Social motivation [124, 151, 50, 91, 101, 115, 108, 23, 25, 125, 150, 10, 16, 18, 64]
	Affect/mood [124, 180]	

4.1 System factors

System factors are to a large extent comparable to Engeldrum's technological variables, and include all those characteristics of the system (or application or service) that contribute to determine the "technically produced quality" [78] of the eventual media presentation. As such, they also determine the presence of impairments in it. In the most general formulation, system factors can address characteristics of the device on which a video is viewed (e.g., a mobile phone, PC, tablet or television), of the technological signal variables (i.e., the video format or parameters in signal processing algorithms) and of the network configuration (i.e., the so-called QoS parameters). Each of these contributions to QoE is discussed here in some detail (although not fully encompassing all the existing literature).

Devices

Nowadays, users watch videos through a diversity of devices. Studies showed that user acceptability of video quality varied with the type of device used to watch the video [86, 85, 87]. For example, See-To et al. [153] showed that user's QoE of the same video was significantly different on a desktop than on a mobile device. User expectations for mobile device performance were lower than for desktop performance, and as a consequence, a higher QoE for a video with the same amount of introduced impairments was found on the mobile device than on the desktop. In more general terms, the impact of display technology variables on image and video quality was investigated thoroughly during the last decades, at least with respect to artifactual quality [81, 40]. Pixel size and arrangement, static and dynamic contrast, white point and color gamut, motion blur and other motion artifacts, response characteristics and flicker, and finally viewing angle range are all elements known to highly impact perceived image quality [19, 93, 165]. The extension from artifactual quality to quality of experience is only limitedly addressed for device optimization. In 3D displays, stereoscopic depth has been shown to increase the appreciation for the viewing experience (see Sec.5.1) but at the same time, disparity may generate visual discomfort [94]. McCarthy et al. [112] reported that a decrease of display resolution yielded user dissatisfaction.

Signal and network variables

Functional characteristics of video streaming (e.g., frame rate, resolution and encoding) directly influence users' QoE [194], as also already discussed in Section 2.3. Gulliver et al. [56] conducted a subjective experiment to investigate the impact of different multimedia frame rates on the user's (impairment-centric) visual quality. Participants were asked to view video clips at different frame rates, and to answer for each clip some questions evaluating whether the participants understood the video content, and to give per clip an overall quality score and a score for the level of enjoyment. The results showed that the assimilation of video information was not significantly affected by frame rate, but the user's perceived visual quality and enjoyment were. In other words, higher frame rates improve overall user enjoyment and quality perception. Similar aspects have been investigated within the context of Scalable Video Coding (SVC). The SVC specification of the H.264 coding scheme [1] adapts the video stream along the temporal, spatial and Signal-to-Noise Ratio (SNR) dimensions to obtain an optimal trade-off between frame-rate, spatial resolution and (spatial) impairment visibility. For a given spatial resolution, the optimal trade-off between temporal and SNR quantization is known to depend, among other factors, on motion [175]. For fast motion videos, a decrease in SNR is preferred over a loss in smoothness resulting from low frame-rates. For static videos, the opposite happens. The trade-off between spatial resolution and frame-rate has instead been shown to depend on the video bitrate, with a preference for large spatial resolution at low bitrates ($<$800 kbps) [103].

4.2 User factors

A user factor is defined as "any variant and invariant characteristic of a human user" that influences the viewing experience, such as demographic, personality, or interest related characteristics [99]. User factors determine for a large part the user "current state" mentioned in the QoE definition reported in Section 3. These factors were largely overlooked for a long time, because they were judged too difficult to quantify in both a subjective and objective way [81]. Nonetheless, lately researchers have started investigating them more systematically, also thanks to the large amount of personal data made available by the users themselves in Social Media. We review in the following some of the main findings with respect to the influence of user factors to QoE.

Interest

In psychology literature, interest has been considered as an emotion. Silvia [159] suggested that interest comes from two appraisals: novelty and coping potential. Novelty is the tendency to seek elements that are new, or unusual in one's environment, and evoke in the user a sense of curiosity. Huang et al. [67] showed that incorporating novel elements into a website attracted curious users and brought out enjoyable experiences. Coping potential is the ability to understand unfamiliar, complex objects, and as such is strongly user-dependent [159]. In the field of aesthetic appreciation of art, it has been shown for example that abstract, unfamiliar works of art were poorly appreciated by the average user [104]; nevertheless, the stronger the background art knowledge of the user, the higher the aesthetic appreciation would get [110]. O'Brien [124] indicated that QoE was often triggered when something resonated with a user's interest. Kortum & Sullivan [90] employed a total of 100 participants and 180 movie clips encoded at nine compression levels from 550kbps up to DVD quality. After viewing the clips, participants were asked to rate the (impairment-related) visual quality and desirability of the movie content. The results showed a general increase in quality rating as the desire for content increased, at a given bitrate. Thus, personal interest in the video significantly influenced user judgments [90]. Palhais et al. [126] used videos of sport events, encoded in four different bitrate/resolution combinations. Participants chose three sports that they were more interested in and three sports that they liked less. Then they watched all videos and rated the (impairment-related) visual quality at the end of each video. The results demonstrated that the interest level had a strong influence on the subjective assessment of the visual quality: users tended to value a video with the same bitrate as higher in QoE when they were more interested in the content of the video.

Personality

Personality is "the particular combination of emotional, attitudinal, and behavioral response patterns of an individual". One of the effective ways to determine personality is the five-factor model (FFM) [26] or "Big Five", consisting of the dimensions openness (i.e., degree of intellectual curiosity and creativity), conscientiousness (i.e., tendency to show self-discipline), extraversion (i.e., the level of orientation towards other people), agreeableness (i.e., the tendency to be compassionate and cooperative) and neuroticism (i.e., the tendency to experience unpleasant emotions easily). Wechsung et al. [180] conducted an experiment asking 33 participants to perform a series of tasks (such as play, pause and stop) in front of an IP-TV. The results showed that the personality of participants influenced their performance. For example, neuroticism was negatively correlated with performance, while agreeableness enhanced it. In contrast, the results also indicated no correlation of personality with impairment annoyance.

Age/Gender

Evidence exists that age influences QoE. Wolters et al. [185] found older adults to be more critical than younger users, which may suggest that elderly people have higher requirements for QoE. However, Naumann et al. [118] observed the opposite: they found that older users tended to rate the (impairment-centric) visual quality more positively than younger users. As far as gender is concerned, little has been done to investigate its effect on QoE appreciation. Males and females are known to react differently to emotional pictures [13] and to have different perception of olfactory and visual media synchronization [117]. As a consequence, it is reasonable to expect that optimal QoE settings may depend on gender too. An initial evidence in this sense can be found in [69]: within the context of 3D audio telephony and teleconferencing services, it was found that males and females have different preferences, in terms of experienced QoE, with respect to the size of the room where the teleconference is held. Closer to the field of image quality Campanella Bracken [12] showed that women viewing a video clip on either an HDTV or a (at that time) more standard resolution NTSC TV reported more perceived realism than men, which may imply that women evaluate at least part of the television content as more real than men.

Affect/Mood

During the interaction with online (video) services users may experience positive or negative affective states (or moods). The interaction itself may induce such a state, but it is also possible that people are already in a particular affective state, such that it may impact the way they experience the interaction with the video service. Positive affective states relate to enjoyment, satisfaction, and fun. For example, a lack of fun can act as a barrier to shop online or enjoyment during a webcast can draw the user in [124]. Negative affective states, such as frustration, anxiety, and boredom, may lead to low QoE. For example, participants that feel frustration towards a technology report lower QoE ratings [180].

4.3 Contextual factors

Contextual factors describe all aspects of the environment within which the user consumes the media, e.g., physical location, economical aspects or social context. The following are considered prominent contextual factors for QoE evaluation.

Physical environment

Many aspects of the physical environment may affect QoE; these aspects may range from characteristics of the seating position (e.g., viewing distance and viewing height) to disturbances that occur in the environment a viewer is in. Viewing distance is a balancing act between two aspects: a shorter viewing distance increases the field of view, and makes the viewer more involved with the content, but may make impairment better visible as well [183]. Staelens et al. [161] investigated subjective quality under viewing conditions, in which television is typically watched.

The results showed that the interruption of phone calls and SMS alerts could prevent a person from getting engaged into a video, which would result in low QoE.

Economic aspects

The economic aspects relate to key concepts of marketing, such as the product and brand strategy, the pricing strategy, the positioning of the product in the market, and the market segmentation and identification of target groups [14]. These aspects are closely related to the notion of Quality of Customer Experience, introduced by Kilkki [83][123]. He indicated that the economic aspects of a product or service, such as price and brand, can have a high impact on QoE, also due to customer loyalty (think about, e.g., Apple). According to [190], there is a positive correlation between the willingness to pay for a multimedia product/service and the (impairment-centric) visual quality of the video offered to the user. The study clearly showed that users were inclined to pay less if they were offered a video with a lower visual quality. When users felt they were overpaying for their service with regard to the quality they experienced, they reacted in different ways, all eventually leading to a decrease in revenues for the operator of those services.

Social context

Social context refers to the fact that a user is affected by the interaction with a group of other people [151], being family, friends or even strangers. In the past, many studies have reported the social impact of traditional TV watching [50, 91, 101]. Co-located co-viewing is a rather common way of consuming the more traditional media, such as TV programs [115], having great potential as a social activity and conversational topic [108]. Co-viewing when enjoying each other's company can increase user's overall satisfaction [115]. Recently, a concept of social TV - as implemented in ”Amigo TV” - has emerged: it provides multiple viewers with a joint TV watching experience by adding communication features via audio conferencing, graphic symbols, and avatars [23, 25]. User studies of social TV have confirmed the high acceptance of such technology, because it allows users to communicate with friends even when they are not physically co-located [125, 150]. Far less is known about the impact on QoE of a newer concept of social context, namely the one arising from recommendations and opinions (e.g., Facebook ”likes”) of friends and/or strangers. Watching online videos is not so often done by multiple people sitting together, nonetheless, a sort of social influence (or pressure) may manifest itself through the opinions of peers or friends gathered via e.g., social networks. Having input from peers or friends is indeed already very common on shopping websites or in gaming communities. For example, O'Brien [124] noted that in one of his studies an interviewed person mentioned reading book reviews from ”certain reviewers that I know I can trust that have similar taste to me” [124]. In other words, this interviewed person created his own social context to support him in deciding which books to read. In the gaming industry, social interaction is explicitly designed in the game. Lively virtual societies are built around multiplayer online games (e.g., World of Warcraft), and these games are highly successful [10, 16]. Several studies even claimed that digital games also can increase social interaction in gamers' real life (e.g., they talk to friends about the game strategy) [18]. Either playing games together or watching others play a game can bring enjoyment to gamers [64]. Although very little has been studied so far, all these studies point towards the impression that social context may strongly impact Quality of Experience.

5 Beyond visual quality: new trends in subjective QoE Assessment

Despite the body of work on influencing factors of QoE described in Section 4, it is clear that unveiling a reliable model of user QoE preference is still beyond reach. The profound transformation that media consumption underwent in the last decade opens countless questions and applications in which influencing factors and features of the viewing experience still have to be determined. In facing this major challenge, subjective assessments represent the core instrument to learn more about the interplay between perceptual, cognitive and affective mechanisms that underlie the appreciation of viewing experience. Nevertheless, subjective QoE assessment methodologies need now to be integrated with knowledge developed from traditionally very different fields (e.g., human computer interaction, affective computing, behavioral psychology, but also media production, computer graphics, and lighting design) to overcome the traditional impairment sensitivity paradigm. We identify three major directions in which the subjective QoE assessment community should seek for the paradigm shift: the technological one, the psychological one, and the methodological one.

5.1 Beyond the traditional screen technology: QoE of immersive viewing experience.

Imaging technologies are evolving quickly. In recent years, new display technologies have been developed that provide a more immersive viewing experience by enhancing specific experience features. High Dynamic Range (HDR) displays, for example, magnify perceived contrast by means of different backlight technologies and the usage of a larger number of bits to represent luminance information [4, 154]. Similarly, stereoscopic and autostereoscopic displays (3D) enhance perceived depth [137, 119], and upcoming 4k and 8k devices display images at a ultra-high resolution [17]. These features come with an undoubted added value for the viewing experience. Nevertheless, to ensure the full enjoyment of the enhanced experience, immersive technologies need optimization both at a display and at a signal level.

In the case of HDR imaging for example, problems such as optimal design of the backlight dimming algorithm [89] as well as of tone mapping operators that can display a high dynamic range image on a regular display [191] are still under investigation. Furthermore, to drive the optimization of such immersive technologies, it is essential to (1) properly understand the impact of an enhanced dimension on the eventual QoE and (2) assess whether such attribute enhancement modifies the impact of other attributes on QoE. In the following, we attempt at exemplifying why these two points are crucial for QoE optimization, by looking at that technology among the aforementioned ones that was most thoroughly studied in the last decade, i.e., stereoscopic displays.

Although introduced halfway the twentieth century, 3D displays became accessible to the general public within the last decade. Initially based on anaglyph projectors in movie theatres, stereoscopic display technology underwent major optimization efforts to finally enter consumers' living rooms. Initially, optimization was again driven by the concept of "image quality", of which a large body of knowledge was already available from 2D displays. Soon enough, it became clear that this concept was insufficient to properly quantify user satisfaction with respect to the overall experience provided by 3D displays. When asked to assess image/video quality of a stereoscopic display, users limited their judgment to the annoyance of the impairments introduced by the technology, not valuing the experience created by the stereoscopic depth [155, 95, 92, 163]. Seuntins [155] showed, for example, how image quality scores decreased due to visible impairments as a consequence of compression similarly in 2D and 3D displays. A similar effect was found for blur, even for different camera-base distances [95]. Kuijsters [92] also showed that increasing stereoscopic depth in images of various levels of colourfulness didn't affect or slightly reduced the perceived image quality (as shown in Figure 4).

At no point in the abovementioned results users judged instead the (added) value of the increased depth on the experience. Hence, to allow manufacturers to optimize, and where needed balance, the full experience of stereoscopic content, a higher level concept was needed. The concepts of naturalness, already introduced in the FUN model described in Section 3 [149], as well as that of "viewing experience" [157] were investigated to cover the user's perceptual and cognitive experience of stereoscopic displays. Based on the series of experiments described above, a higher level evaluation criterion EC was defined and modelled as a combination of image quality IQ and perceived depth D, i.e.,:

$$EC = \alpha \times IQ + \beta \times D \tag{2}$$

When using naturalness as EC, it was found to incorporate depth information for about 25% (while 75% of the judgment still depended on image quality), assessment of viewing experience consisted instead for about 82% out of image quality and for about 18% out of stereoscopic depth [95].

The importance of naturalness as an evaluation criterion for stereoscopic displays is in line with results from applying the IBQ method to stereoscopic content assessments. Hkkinen [59] showed that for many viewers stereoscopy changed the

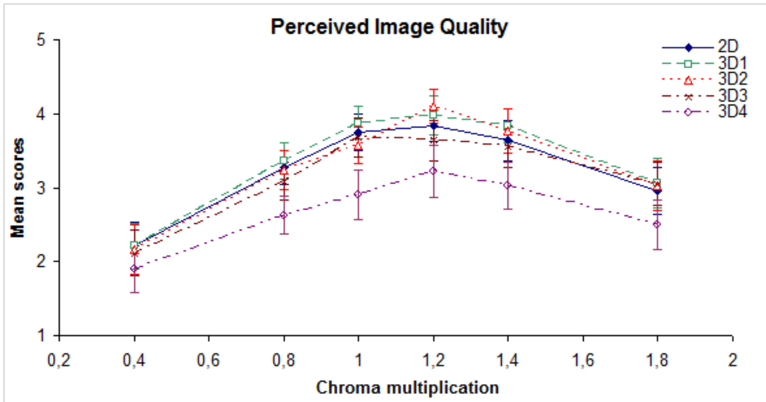

Fig. 4 Illustration of the effect of impairments (chroma multiplication) and stereoscopic depth on image quality scores for chroma affected images [92]. The different lines indicate mean scores for 2D content (solid line), and 3D content with a disparity of 1 mm (3D1), 2 mm (3D2), 3 mm (3D3) and 4 mm (3D4), respectively.

life-likeness of the content, although in some cases stereoscopy also introduced artificiality or ”unrealness”, depending on the specific content. Partly based on all these findings, the ITU [137] defined a new recommendation for the subjective evaluation of 3D-TV systems, prescribing that viewers should score three factors separately, i.e., picture quality, perceived depth and visual comfort.

Immersive viewing technologies are these days evolving even beyond the display. Virtual and Augmented Reality technologies, for example, are currently used for a multitude of applications, ranging from Mental Health Computing [168] to Google Glasses [7], and an understanding of QoE with respect to those highly immersive contexts has to be achieved. Immersive experiences are also evolving by incorporating other types of technologies, traditionally uncorrelated to multimedia delivery. Starting with the Philips Ambilight TV for example, LED lights were incorporated in the TV display to increase the field of view and to give therefore viewers a more cinematic experience, also reducing the eye fatigue. As for stereoscopic displays, image quality was found to relate only to visual impairment, neglecting the added value of the light effect. The term viewing experience, on the contrary, was proven to cover both image quality and the added value of Ambilight, also in combination with increased depth (i.e., when mounted on stereoscopic displays) [156].

It is reasonable to expect that viewing experience will even further extend beyond the display in the future. Solid State Lighting (SSL) technology, for example, is already being used to embed low-resolution ”displays” in our whole environment. Because of their fast-switching and spectrally tuneable characteristics, LEDs can be spatially distributed and embedded in (semi)-transparent materials: this allows designers to present information on walls, floors and/or ceilings around us. Although still used mainly towards functional purposes, we can foresee for the near future that such technology will provide enhanced entertainment experiences, based on the creation of (affective) atmospheres through the combination of visual content, sound and lighting. Ideas to use atmospheric light in combination with video or games (via scripting) or with music (e.g. by associating colors to terms from the lyrics) are indeed emerging. Nevertheless, assessment of the full experience of these systems will require - most probably - a new higher level concept, overarching aspects of image quality, sound quality and light experience [172].

5.2 Beyond perception: the role of aesthetics and emotion in QoE appreciation

A second important evolution in subjective QoE assessment is the inclusion of affective evaluations within QoE measurements. As mentioned in Sections 3 and 4, the affective state of the user (i.e., his/her mood or specific emotional state) may impact the way a viewing experience is appreciated [180]. In turn, the potential for the viewing experience to impact on the affective state of the user (e.g., increase the arousal of the emotion as well as improve its valence) should be considered in QoE assessment paradigms. Efforts in this direction are currently growing (e.g. [148, 135, 181]), also reaching out to the affective computing community. Nevertheless, major challenges are ahead. First, appropriate methodologies to measure the affective impact of media in relation to QoE have yet to be determined. Both self-reporting instruments such as the Self-Assessment Manikin [96] or the Affect Button [15], and more ”objective” tools such as physiological measurements (e.g., EEG or skin conductance) are being evaluated at the moment. Existing results are however scattered and more structured efforts are needed to identify a pool of affective measurements that can complement existing standards [138, 73, 137, 82] for subjective QoE measurement. A second important challenge lays in the need to decouple, within the QoE judgment, the effect of affective states pre-existing the visual experience from that of the emotional state induced by the viewing experience itself. The ability of the media to induce emotion, creating an empathic experience with the video content, may be positively taken into account in user QoE judgments. A valuable tool to this end would be to use stimuli of which the emotional impact on the user can be controlled (e.g., IAPS emotional slides, or standardized excerpts from movies with the potential to induce specific mood states [97, 55]). They could be used to induce specific moods prior to the visual experience, to investigate the impact of pre-existing mood on QoE judgments; also, they could constitute test stimuli for the viewing experience, to allow an understanding of how induced emotional states alter QoE. Nevertheless, mood induction practices have to be carefully designed in order to carry out experiments that are still ethically acceptable.

At the same time, it is interesting to research which properties of the image have the potential to impact the affective state of the user, along with information on the changes in arousal and valence of this affective shift. Color, for example, is well known for having an impact on people’s mood, both from a psychological and a physiological point of view [169, 192]. Similarly, contrast or content arrangement of an image may generate changes in the mood state. Understanding and quantifying the relationship between physical properties of the image, their perception and their impact on the user affective state is therefore a key challenge for upcoming QoE research.

Some work in this sense has been carried out within the scope of understanding the aesthetic appeal of media. Aesthetic appreciation is generally recognized to be related to both perceptual and affective mechanisms, and it has been for long studied independently from the concept of Quality of the Visual Experience. Mastered for a long time by artists and then also addressed by psychologists [11], lately it has started to attract the attention of the media engineering community. Predicting the aesthetic appeal of images has become interesting especially towards improving information retrieval, computer graphics and automatic management of image collections [77]. As a consequence, studies have been carried out first to identify image and user attributes impacting on aesthetic appeal and then to model them. Perceptual image features (as per [51], see Section 3) such as color saturation, brightness and amount of details (i.e., texture and visual crowding) were found to contribute to the final aesthetic quality judgment [27, 109, 76]. In particular, the deployment of visual attention has been shown to be related to image clutter [20], in turn negatively correlated with image aesthetic appeal scores [146]. The same study showed that visual importance [176] is to some extent predictive of compliance of the image to photographical compositional rules, which in turn has a beneficial effect on aesthetic appeal. Impairment generated by specific media configuration (System Influence Factors, see Section 4.1), such as blockiness [145] have been shown to negatively affect aesthetic appeal. Similarly, user Influence Factors such as experience, cognitive bias, and personal opinions and memories [132] have been found to strongly condition the appreciation of the aesthetic experience. Correlation between aesthetic ratings and familiarity with the image subject has been reported in [110], and content recognizability (i.e., the level of abstraction of the content) has been shown to have an influence on aesthetic appeal in works of art [98].

Interestingly, very little work has been carried out in trying to link the aesthetic appeal of an image to the overall Quality of the visual Experience. Nevertheless, initial evidence exists that the aesthetic appeal of an image does influence not only QoE, but also the judgment in terms of annoyance of impairments presented in the image itself [144]. In this study, Redi asked a pool of participants to judge ”integrity” (namely, the traditional, impairment-related concept of visual quality) of a set of images, including (i) a group of pristine images and (ii) a group of images derived by those in group (i) by applying JPEG compression to them. As a result, the images in this second group presented the same content as those in the first one, but affected by visible compression artifacts. The pristine images of group (i) had already been evaluated in terms of aesthetic appeal in a separate study [145]. Redi correlated then the integrity scores of both groups of images with the aesthetic appeal scores of the pristine ones. It was found that the integrity judgments of the pristine images (group i) were influenced by the level of aesthetic appeal of the image, and that the two quantities were negatively correlated (see Figure 5.a). Conversely, when impairments were present (group ii, figure **??**.b) integrity judg-

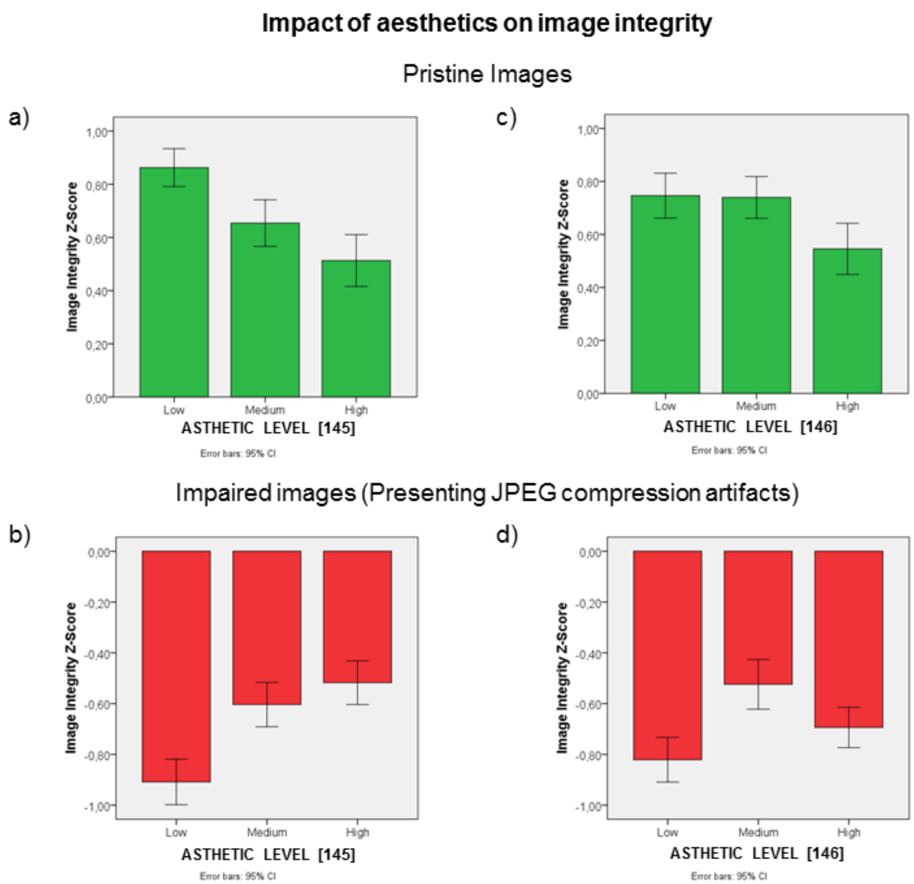

Fig. 5 Relationship between aesthetic appeal and image integrity [144, 145, 146]: dependency of integrity scores of pristine images on the aesthetic appeal level as assigned in experiment [145] (a) and in experiment [146] (c); dependency of integrity scores of impaired (JPEG-compressed) images on the aesthetic appeal level as assigned in experiment [145] (b) and in experiment [146] (d).

ments increased as the aesthetic appeal increased. It should be mentioned that, in instructing the participants to score aesthetic appeal, the concept of integrity was also explicitly mentioned, and distinguished from aesthetic appeal. This may have primed participants in (unconsciously) taking into account integrity in their aesthetic appeal judgments, partially explaining the results found in [144]. To check this, we repeated the analysis performed in [144], but by using a different set of aesthetic appeal scores, obtained from study [146]. There, participants were again asked to score aesthetic appeal, but now without reference to image integrity. The consistency in aesthetic appeal scores between experiments [145] and [146] turned out to be 0.68.

Although still acceptable in terms of predictive power of one set of scores for the other, this number is far from correlations typically found across experiments for e.g. impairment annoyance scores (typically, ~ 0.9). There are several possible reasons for this discrepancy: (1) the highly personal component of the aesthetic appeal judgment (since different participants were used in both experiments), (2) the difference in experimental protocol or (3) range effects (since only a subset of the images of the first experiment were used in the second) [149]. Despite these possible deviations, also the results of the second experiment showed that the aesthetic appeal level had an impact on integrity [144] for both the pristine and impaired images. Pristine images that were highly aesthetically appealing were scored significantly lower in integrity than the others (Figure 5.c), whereas impaired images with a low aesthetic appeal were scored significantly lower than the others (figure 5.d), confirming the trend identified in Figure 5.b. This consistency points out how aesthetic appeal plays a role in QoE and in tolerance to impairment; as a result, including aesthetic appeal information in QoE metrics may help in improving their accuracy and QoE optimization thereafter.

5.3 Beyond lab-based studies: methodological shifts for reliable QoE quantification

Research on Quality of Experience has relied for long (and now more than ever) on determining user preferences with respect to the sensitivity to visual impairment through subjective studies. The main goal of subjective testing is to sort stimuli (i.e., media) according to their perceived properties or attributes [40] on a given scale.

Multiple psychometric methodologies have been developed with this purpose, and adapted for the measurement of QoE in standardized conditions [138, 73, 137, 82], and choosing the most appropriate one for a test is far from trivial. In discriminating among methodologies, Engeldrum [40] suggests to take into account aspects such as the confusion level within the set of stimuli and the effort required to the participant to complete his/her task. The confusion level is determined by how closely the test stimuli are spaced in quality. The narrower are the quality gaps among them, the higher is the probability of inducing confusion (disagreement, possibly inversions) in across-participants judgments. Methods able to accurately measure the quality of stimuli with high confusion (e.g., paired comparison) are typically unable to measure large quality gaps. The effort required to participants to complete their task depends on the number of judgments needed per stimulus, and hence it is related to the number of stimuli involved. Methods requiring a high number of judgments per stimulus are not suitable for experiments involving large datasets, as they could be prone to fatigue and learning errors. Other desirable properties of the methods, depending on the goal of the experiments, may be the minimization of inter-participant variability [62] or the robustness to range effects [32] (e.g. in case results of multiple, separate experiments need to be merged into a single set of data [141]).

Psychometric Methods for QoE Measurement

The Paired Comparison (PC) method [28, 166] is a classic psychometric technique that allows measuring distances among stimuli in terms of Just Noticeable Differences (JNDs) [40]. The experimental procedure consists of asking subjects to compare each stimulus with all other stimuli in the set. As a result, even small differences between the stimuli can be detected. On the other hand, the judgment effort grows as the square of the number of stimuli, hence this number must be limited. Moreover, in analyzing the results, complications may arise due to the ”zero and one problem” [116], or inconsistencies in the selected model [52]. Lately, the QoE has shown growing interest around PC, [102, 186, 187], and methods have been developed for establishing confidence intervals to the quality scores provided by the PC tests [186]. The Double Stimulus Impairment Scaling (DSIS) methodology [138] is also often chosen for the assessment of visual impairments. DSIS judgments are expressed on an interval scale (typically, a five-point categorical scale, ACR), as a (conscious) comparison of each impaired stimulus with its undistorted version. Being a double stimulus method (i.e., reference and test stimuli are both shown during the judgment), the DSIS requires a moderate effort per judgment, but still allows the assessment of large datasets. A possible drawback of the method may be the categorical scale used for the assessment: the boundaries among categories (e.g. ”good” and ”fair”) are blurred an depend on the participant; this may result in low inter-participant agreement [81, 40]. The ACR scale is however to date the most used method for scaling stimuli, also in a Single Stimulus setting (i.e., without an explicit reference to be presented to the participant) [141]. Both DSIS and Single Stimulus scaling can be performed also with numerical scales, both discrete or continuous [138, 68]. In all these cases, The results of the tests are reported in terms of average score per stimulus (Mean Opinion Scores), expressed in the scale used for the experiment. These scores reflect human preference, though do not have a precise psychophysical meaning. Indeed, the obtained scores may vary with the definition of the scale [40], as well as with the quality range spanned by the stimuli (range effects, [32]). This suggests that comparing results of different experiments may be problematic, possibly inducing inconsistencies when merging these data in a single, larger dataset.

Among classic scaling methodologies, The Quality Ruler (QR) method deserves a mention, as an middle-ground alternative between the direct scaling methodologies (DSIS, Single Stimulus) and PC. The QR method was first described by Keelan in [81], and subsequently adopted as an international ISO standard for psychometric experiments for image quality estimation [82]. The core idea of the QR method is to provide the participant with a set of reference images, anchored along a calibrated quality scale, to compare a test image with. The task of the participant is to find the reference image closest in quality to the test image by visual matching. Reference images (1) depict a single scene and vary in only one perceptual attribute (i.e., blur, blockiness, color saturation); (2) are closely spaced in quality, but altogether span a wide range of quality. They are presented in a way that easily allows detection of the quality difference between them, and their close spacing in quality should allow the participant to score with higher confidence, decreasing the risk of inversions and range effects. In practice, participants perform several comparisons reference-test stimuli to complete a single assessment, until they find the reference stimulus that matches the quality of the test one. The advantage of this procedure is that, as long as the referencer stimuli are kept the same, subjective scores obtained from a quality ruler experiment always refer to the ruler scale, and not to the quality range spanned by the test stimuli. This minimizes range effects. Furthermore, it has been shown that the visual matching procedure reduces inter-participant variability [141]. Unfortunately, this method has been successfully implemented for images [82, 141], but it is of hard applicability for video QoE assessment.

Subjective testing outside the Lab

To obtain reliable results, psychometric experiments have usually been performed in highly controlled, standardized environments [138, 73, 137, 82], typically within laboratory facilities. This allowed to control for lighting and viewing position, minimizing the effect of environmental contextual factors (see Section 4.3) and making visibility conditions homogeneous across participants.

The evolution of multimedia technology calls now for a shift in the traditional lab-based study paradigm. Since the advent of mobile technology (smartphones and tablets) and internet-based video delivery, the visual experience is now consumed in very different environments, and should be studied within realistic usage conditions to be properly optimized [161, 152]. Furthermore, with the acceptance of the more encompassing definition of QoE presented in Section **??**, where personal differences are taken into account and need to be understood, more attention should be given to the demographic composition of the pool of participants used in the subjective studies. This pool should be indeed as representative as possible of existing differences in terms of user Influencing Factors (Section 4.1). Recruiting such a diverse pool of participants may prove difficult, and the eventual amount of individuals to be involved in the study may explode, thus making a lab-based experiment unfeasible in terms of time consumption and cost. In this scenario, interest in using Crowdsourcing [66] for subjective tests of QoE has grown significantly [65]. Crowdsourcing was originally conceived to outsource small and repetitive tasks (so-called microtasks) to a multitude of people (so-called microworkers) who, online and for a small compensation, could perform these tasks in a time- and cost-effective way. Platforms such as Microworkers, Amazon Mechanical Turk and CrowdFlower were created to facilitate the recruitment of microworkers, and soon enough it became clear that Crowdsourcing had an enormous potential for the performance of Human Intelligence Tasks (e.g., image labeling) in Multimedia research [36]. As a result, it became of interest for subjective QoE assessment as well. Compared to traditional subjective QoE assessment in a controlled lab environment, which is time-consuming and high-cost, Crowdsourcing tasks can be accomplished within a few minutes and do not require a long-term employment of the participants. More importantly, Crowdsourcing gives the opportunity to collect data from populations with very diverse demographics, enabling therefore the investigation of User Influencing Factors in QoE judgments [65, 170].

Nevertheless, Crowdsourcing still has some challenges ahead. Besides the reliability issues related to payment scheme, worker selection and ease of sloppiness in carrying out the experimental task, which are extensively discussed in [65], there are a few other points that should be made. First, Crowdsourcing tasks should be fairly short (up to 10 minutes) to avoid boredom and unreliable behavior. Traditional QoE tests typically involve tens or hundreds of stimuli, requiring participants to score for much longer timespans (typically between 30 minutes and one hour). Thus, to collect QoE scores for a large set of stimuli, experimenters usually have to decompose the scoring task in a set of smaller tasks (i.e., campaigns), each one including a sub-set of the stimuli. Although it is common practice to merge all QoE scores from different campaigns as if they were expressed on the same QoE scale, this may be a dangerous practice. Range as well as environmental effects [149] may occur, making the merging meaningless. A possible solution to this is to select a number of stimuli to be scored in all campaigns [147]. If carefully chosen (e.g. some with excellent quality, others with very low quality) these stimuli can function as anchors for the scoring scale, limiting range effects; furthermore, they can be used for realignment purposes.

It should also be considered that, by dividing the traditional design over many people, we intrinsically generate mixed-subjects designs. So, it may be the case that the measurements themselves loose in accuracy, because one cannot fully exploit within-subjects variance. An interesting solution to that has been recently proposed in [187, 188], which involves randomized paired comparison [38] to accommodate incomplete and imbalanced data. Paired Comparison has been found to be an effective methodology for measuring QoE via crowdsourcing, due to the simplicity of the task [102] and the availability of tools for the analysis of incomplete preference matrices. Furthermore, it has been shown to be a suitable methodology for embedding worker reliability checks [102, 22, 189]. This property is especially desirable for crowdsourcing, given that the trustworthiness of workers is often doubtful and that, due to the lack of supervision, workers responses may be inaccurate or erroneous (see also [65]). On the other hand, for scaling large sets of stimuli (in the order of hundreds), the applicability of paired comparison is still limited, as the number of pairs to be judged may be intractable also for such a far-outreaching methodology. Finally, although this may be partially compensated by the larger demographic spread, it should be considered that at the present stage Crowdsourcing attracts only a subset of the population, leaving out, e.g., elderly people (who, for the time being, cannot master the technology).

6 Conclusion

Although subjective assessment of Quality of Experience has been investigated for over 50 years, this field is in continuous expansion. In this chapter we documented the evolution of the impairment-sensitivity centric understanding of QoE to a more encompassing one, which takes into account both attributes of the experience as well as external influencing factors that modulate the user appreciation for a specific delivered media. In a world where media fruition is tightly related to social networks and social media, as well as to immersive but also mobile viewing systems, the user cannot be considered as a simple, passive observer anymore. Users are active agents which interact with the system, e.g., selecting the content and/or the modality with which they desire the media to be delivered. As a result, and as we documented in this chapter, elements such as visual semantics, user personality, preferences and intent, social and environmental context of media fruition also concur to the final experience assessment. While (theoretical) models of QoE appreciation exist that take these elements into account, in practice little is known, still, on how attributes of the experience combine into a final QoE judgment and how external factors influence this combination. Similarly, light still needs to be shed on the cognitive and affective processes that underlie viewing experience appreciation. Subjective quality assessment is therefore now more than ever a core element in pushing forward the research on QoE optimization. Empirical studies are needed to unveil and thoroughly understand the mechanisms underlying the interplay between attributes and influencing factors of viewing experiences, and existing as well as new methodologies will have to be deployed towards that goal. Methods designed for studying affect, human-machine interaction, as well as human cognitive processing will have to be adopted by the QoE community and integrated with existing subjective quality measurement tools to properly quantify viewing experiences in their context of usage. Moreover, these methods will have to be adapted to be deployed in large scale experiments that reach out to big amounts of users world-wide, by means of web-based technologies such as crowdsourcing. A different type of adaptation will also be necessary to perform subjective studies in less controllable but more realistic contexts of usage. Data analysis techniques will also have to be upgraded to allow the merging of heterogeneous data derived from lab-, real world- and web-based studies. Content metadata as well as user/context information retrievable in internet (e.g. social media profiles, or textual comments to media material) will also provide useful information to better understanding of user preferences. Finally, appropriate modeling tools will have to be deployed to construct, based on the abovementioned information, an accurate, encompassing model of Quality of Viewing Experience appreciation that could steer a better delivery of the richness of digital multimedia content that is available nowadays

References

1. Nicola Adami, Alberto Signoroni, and Riccardo Leonardi. State-of-the-art and trends in scalable video compression with wavelet-based approaches. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(9):1238–1255, 2007.
2. Hani Alers, Judith Redi, Hantao Liu, and Ingrid Heynderickx. Studying the effect of optimizing image quality in salient regions at the expense of background content. *Journal of*

Electronic Imaging, 22(4):043012–043012, 2013.
3. John Allnatt. *Transmitted-picture assessment*. Wiley Chichester,, UK, 1983.
4. Munisamy Anandan. Progress of led backlights for lcds. *Journal of the Society for Information Display*, 16(2):287–310, 2008.
5. Javed Asghar, Francois Le Faucheur, and Ian Hood. Preserving video quality in iptv networks. *Broadcasting, IEEE Transactions on*, 55(2):386–395, 2009.
6. Simon Attfield, Gabriella Kazai, Mounia Lalmas, and Benjamin Piwowarski. Towards a science of user engagement (position paper). In *WSDM Workshop on User Modelling for Web Applications*, 2011.
7. Roberto Baldwin. Google glasses face serious hurdles, augmented-reality experts say. *Wired Magazine*, 2012.
8. Peter GJ Barten. *Contrast sensitivity of the human eye and its effects on image quality*, volume 72. SPIE press, 1999.
9. Soren Bech, Roelof Hamberg, Marco Nijenhuis, Kees Teunissen, Henny Looren de Jong, Paul Houben, and Sakti K Pramanik. Rapid perceptual image description (rapid) method. In *Electronic Imaging: Science & Technology*, pages 317–328. International Society for Optics and Photonics, 1996.
10. Steve Benford, Chris Greenhalgh, Tom Rodden, and James Pycock. Collaborative virtual environments. *Communications of the ACM*, 44(7):79–85, 2001.
11. Daniel E Berlyne. Aesthetics and psychobiology. 1971.
12. Cheryl Campanella Bracken. Presence and image quality: The case of high-definition television. *Media Psychology*, 7(2):191–205, 2005.
13. Margaret M Bradley, Maurizio Codispoti, Dean Sabatinelli, and Peter J Lang. Emotion and motivation ii: sex differences in picture processing. *Emotion*, 1(3):300, 2001.
14. Frances Brassington and Stephen Pettitt. *Principles of marketing*. FT Prentice Hall, 2005.
15. Joost Broekens and Willem-Paul Brinkman. Affectbutton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies*, 71(6):641–667, 2013.
16. Barry Brown and Marek Bell. Play and sociability in there: Some lessons from online games for collaborative virtual environments. In *Avatars at Work and Play*, pages 227–245. Springer, 2006.
17. ITUR BT2020. Parameter values for ultra-high definition television systems for production and international program exchange, 2012.
18. Diane Carr, Gareth Schott, Andrew Burn, and David Buckingham. Doing game studies: A multi-method approach to the study of textuality, interactivity and narrative space. *Media International Australia, Incorporating Culture & Policy*, (110):19, 2004.
19. Joseph A Castellano. *Handbook of display technology*. Elsevier, 1992.
20. Cathleen D Cerosaletti, Alexander C Loui, and Andrew C Gallagher. Investigating two features of aesthetic perception in consumer photographic images: clutter and center. In *IS&T/SPIE Electronic Imaging*, pages 786507–786507. International Society for Optics and Photonics, 2011.
21. Damon M Chandler. Seven challenges in image quality assessment: past, present, and future research. *ISRN Signal Processing*, 2013, 2013.
22. Kuan-Ta Chen, Chen-Chi Wu, Yu-Chun Chang, and Chin-Laung Lei. A crowdsourceable qoe evaluation framework for multimedia content. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 491–500. ACM, 2009.
23. Konstantinos Chorianopoulos and George Lekakos. Introduction to social tv: Enhancing the shared experience with interactive tv. *Intl. Journal of Human–Computer Interaction*, 24(2):113–120, 2008.
24. I Cisco. Cisco visual networking index: Forecast and methodology, 2011–2016. *CISCO White paper*, pages 2011–2016, 2012.
25. T Coppens, Frie Vanparijs, and K Handekyn. Amigotv: A social tv experience through triple-play convergence. *Alcatel Technology white paper*, 2005.
26. Paul T Costa and Robert R McCrae. The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment*, 2:179–198, 2008.

27. Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Studying aesthetics in photographic images using a computational approach. In *Computer Vision–ECCV 2006*, pages 288–301. Springer, 2006.
28. Herbert Aron David. *The method of paired comparisons*, volume 12. DTIC Document, 1963.
29. Huib de Ridder. Subjective evaluation of scale-space image coding. In *Electronic Imaging'91, San Jose, CA*, pages 31–42. International Society for Optics and Photonics, 1991.
30. Huib de Ridder. Minkowski-metrics as a combination rule for digital-image-coding impairments. In *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pages 16–26. International Society for Optics and Photonics, 1992.
31. Huib de Ridder. Naturalness and image quality: saturation and lightness variation in color images of natural scenes. *Journal of imaging science and technology*, 40(6):487–493, 1996.
32. Huib de Ridder. Cognitive issues in image quality measurement. *Journal of Electronic Imaging*, 10(1):47–55, 2001.
33. Huib de Ridder, Frans JJ Blommaert, and Elena A Fedorovskaya. Naturalness and image quality: chroma and hue variation in color images of natural scenes. In *IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology*, pages 51–61. International Society for Optics and Photonics, 1995.
34. Carlo Demichelis and Philip Chimento. Ip packet delay variation metric for ip performance metrics (ippm). 2002.
35. Robert Desimone and John Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222, 1995.
36. Anhai Doan, Raghu Ramakrishnan, and Alon Y Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4):86–96, 2011.
37. Florin Dobrian, Vyas Sekar, Asad Awan, Ion Stoica, Dilip Joseph, Aditya Ganjam, Jibin Zhan, and Hui Zhang. Understanding the impact of video quality on user engagement. *ACM SIGCOMM Computer Communication Review*, 41(4):362–373, 2011.
38. Alexander Eichhorn, Pengpeng Ni, and Ragnhild Eg. Randomised pair comparison: an economic and robust method for audiovisual quality assessment. In *Proceedings of the 20th international workshop on Network and operating systems support for digital audio and video*, pages 63–68. ACM, 2010.
39. S.N. Endrikhovskij. Image quality and colour categorization. *Colour image sci-ence: exploiting digital media*, pages 363–382, 2002.
40. Peter G Engeldrum. *Psychometric scaling: a toolkit for imaging systems development*. Imcotek Press, 2000.
41. Peter G Engeldrum. A theory of image quality: The image quality circle. *Journal of imaging science and technology*, 48(5):447–457, 2004.
42. Ulrich Engelke, Hagen Kaprykowsky, H Zepernick, and Patrick Ndjiki-Nya. Visual attention in quality assessment. *Signal Processing Magazine, IEEE*, 28(6):50–59, 2011.
43. Ulrich Engelke, Romuald Pepion, Patrick Le Callet, and Hans-Jürgen Zepernick. Linking distortion perception and visual saliency in h. 264/avc coded video containing packet loss. In *Visual Communications and Image Processing 2010*, pages 774406–774406. International Society for Optics and Photonics, 2010.
44. Mylene CQ Farias and Sanjit K Mitra. No-reference video quality metric based on artifact measurements. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, pages III–141. IEEE, 2005.
45. Mylène CQ Farias, Sanjit K Mitra, and John M Foley. Perceptual contributions of blocky, blurry and noisy artifacts to overall annoyance. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, volume 1, pages I–529. IEEE, 2003.
46. Elena Fedorovskaya, Carman Neustaedter, and Wei Hao. Image harmony for consumer images. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 121–124. IEEE, 2008.
47. Elena A Fedorovskaya and Huib De Ridder. Subjective matters: from image quality to image psychology. In *IS&T/SPIE Electronic Imaging*, pages 86510O–86510O. International Society for Optics and Photonics, 2013.

48. Rony Ferzli and Lina J Karam. A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb). *Image Processing, IEEE Transactions on*, 18(4):717–728, 2009.
49. Markus Fiedler, Tobias Hossfeld, and Phuoc Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *Network, IEEE*, 24(2):36–41, 2010.
50. David Gauntlett and Annette Hill. *TV living: Television, culture and everyday life*. Routledge, 2002.
51. George Ghinea and JT Thomas. Quality of perception: user quality of service in multimedia presentations. *IEEE Transactions on Multimedia*, 7(4):786–789, 2005.
52. WILFRED A Gibson. A least-squares solution for case iv of the law of comparative judgment. *Psychometrika*, 18(1):15–21, 1953.
53. Bernd Girod. What's wrong with mean-squared error? In *Digital images and human vision*, pages 207–220. MIT press, 1993.
54. Lutz Goldmann, Francesca De Simone, Frederic Dufaux, Touradj Ebrahimi, Rudolf Tanner, and Mauro Lattuada. Impact of video transcoding artifacts on the subjective quality. In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, pages 52–57. IEEE, 2010.
55. James J Gross and Robert W Levenson. Emotion elicitation using films. *Cognition & Emotion*, 9(1):87–108, 1995.
56. Stephen R Gulliver and George Ghinea. Stars in their eyes: What eye-tracking reveals about multimedia perceptual quality. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 34(4):472–482, 2004.
57. Stephen R Gulliver and Gheorghita Ghinea. Defining user perception of distributed multimedia quality. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 2(4):241–257, 2006.
58. Stephen R Gulliver and Gheorghita Ghinea. The perceptual and attentive impact of delay and jitter in multimedia delivery. *Broadcasting, IEEE Transactions on*, 53(2):449–458, 2007.
59. Jukka Häkkinen, Takashi Kawai, Jari Takatalo, Tuomas Leisti, Jenni Radun, Anni Hirsaho, and Göte Nyman. Measuring stereoscopic image quality experience with interpretation based quality methodology. In *Electronic Imaging 2008*, pages 68081B–68081B. International Society for Optics and Photonics, 2008.
60. Andrew M Haun and Eli Peli. Is image quality a function of contrast perception? In *IS&T/SPIE Electronic Imaging*, pages 86510C–86510C. International Society for Optics and Photonics, 2013.
61. Sheila S Hemami and Amy R Reibman. No-reference image and video quality estimation: Applications and human-motivated design. *Signal processing: Image communication*, 25(7):469–481, 2010.
62. T Hobfeld, Raimund Schatz, and Sebastian Egger. Sos: The mos is not enough! In *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, pages 131–136. IEEE, 2011.
63. Donald Hoffman. The interface theory of perception: Natural selection drives true perception to swift extinction. *Object categorization: Computer and human vision perspectives*, pages 148–165, 2009.
64. Robyn M Holmes and Anthony D Pellegrini. Childrens social behavior during video game play. *Handbook of Computer Game Studies. The MIT Press, Cambridge, Massachusetts*, 2005.
65. T Hoßfeld, C Keimel, M Hirth, B Gardlo, J Habigt, K Diepold, and P Tran-Gia. Crowdtesting: a novel methodology for subjective user studies and qoe evaluation. *University of Würzburg, Tech. Rep*, 486, 2013.
66. Jeff Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.
67. Ming-Hui Huang. Designing website attributes to induce experiential encounters. *computers in Human Behavior*, 19(4):425–442, 2003.
68. Quan Huynh-Thu, M-N Garcia, Filippo Speranza, Philip Corriveau, and Alexander Raake. Study of rating scales for subjective quality assessment of high-definition video. *Broadcasting, IEEE Transactions on*, 57(1):1–14, 2011.

69. Mansoor Hyder, Noel Crespi, Michael Haun, Christian Hoene, et al. Are qoe requirements for multimedia services different for men and women? analysis of gender differences in forming qoe in virtual acoustic environments. In *Emerging Trends and Applications in Information Communication Technologies*, pages 200–209. Springer, 2012.
70. Selim Ickin, Katarzyna Wac, Markus Fiedler, Lucjan Janowski, Jin-Hyuk Hong, and Anind K Dey. Factors influencing quality of experience of commonly used mobile applications. *Communications Magazine, IEEE*, 50(4):48–56, 2012.
71. Wijnand A IJsselsteijn, Huib de Ridder, and Roelof Hamberg. Perceptual factors in stereoscopic displays: the effect of stereoscopic filming parameters on perceived quality and reported eyestrain. In *Photonics West'98 Electronic Imaging*, pages 282–291. International Society for Optics and Photonics, 1998.
72. Wijnand A IJsselsteijn, Huib de Ridder, and Joyce Vliegen. Subjective evaluation of stereoscopic images: effects of camera parameters and display duration. *Circuits and Systems for Video Technology, IEEE Transactions on*, 10(2):225–233, 2000.
73. P ITU-T RECOMMENDATION. Subjective video quality assessment methods for multimedia applications. 1999.
74. Ruud Janssen. *Computational image quality*, volume 101. SPIE press, 2001.
75. TJWM Janssen and FJJ Blommaert. Image quality semantics. *Journal of imaging science and Technology*, 41(5):555–560, 1997.
76. Wei Jiang, Alexander C Loui, and Cathleen Daniels Cerosaletti. Automatic aesthetic value assessment in photographic images. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 920–925. IEEE, 2010.
77. Dhiraj Joshi, Ritendra Datta, Elena Fedorovskaya, Quang-Tuan Luong, James Z Wang, Jia Li, and Jiebo Luo. Aesthetics and emotions in images. *Signal Processing Magazine, IEEE*, 28(5):94–115, 2011.
78. Satu Jumisko-Pyykkö. *User-centered quality of experience and its evaluation methods for mobile television*. PhD thesis, Doctoral thesis, Tampere University of Technology, Tampere, 2011.
79. Satu Jumisko-Pyykkö and Miska M Hannuksela. Does context matter in quality evaluation of mobile television? In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, pages 63–72. ACM, 2008.
80. Sandeep Kanumuri, Pamela C Cosman, Amy R Reibman, and Vinay A Vaishampayan. Modeling packet-loss visibility in mpeg-2 video. *Multimedia, IEEE Transactions on*, 8(2):341–355, 2006.
81. Brian Keelan. *Handbook of image quality: characterization and prediction*. CRC Press, 2002.
82. Brian W Keelan and Hitoshi Urabe. Iso 20462: A psychophysical image quality measurement standard. In *Electronic Imaging 2004*, pages 181–189. International Society for Optics and Photonics, 2003.
83. Kalevi Kilkki. Quality of experience in communications ecosystem. *J. UCS*, 14(5):615–624, 2008.
84. Hyun-Jong Kim, Dong Hyeon Lee, Jong Min Lee, Kyoung-Hee Lee, Won Lyu, and Seong-Gon Choi. The qoe evaluation method through the qos-qoe correlation model. In *Networked Computing and Advanced Information Management, 2008. NCM'08. Fourth International Conference on*, volume 2, pages 719–725. IEEE, 2008.
85. Hendrik Knoche and John McCarthy. Mobile users" needs and expectations of future multimedia services. Technical report, 2004.
86. Hendrik Knoche and M Angela Sasse. Getting the big picture on small screens: Quality of experience in mobile tv. Technical report, Information Science Reference, 2008.
87. Hendrik Knoche and M Angela Sasse. The big picture on small screens delivering acceptable video quality in mobile tv. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 5(3):20, 2009.
88. Jan Koenderink. Vision as a user interface. In *IS&T/SPIE Electronic Imaging*, pages 786504–786504. International Society for Optics and Photonics, 2011.

89. Jari Korhonen, Claire Mantel, Nino Burini, and Soren Forchhammer. Searching for the preferred backlight intensity in liquid crystal displays with local backlight dimming. In *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, pages 118–123. IEEE, 2013.
90. Philip Kortum and Marc Sullivan. The effect of content desirability on subjective video quality ratings. *Human factors: the journal of the human factors and ergonomics society*, 52(1):105–118, 2010.
91. Robert Kubey and Mihalyi Csikszentmihalyi. *Television and the quality of life: How viewing shapes everyday experience*. Routledge, 2013.
92. Andre Kuijsters, Wijnand A Ijsselsteijn, Marc TM Lambooij, and Ingrid EJ Heynderickx. Influence of chroma variations on naturalness and image quality of stereoscopic images. In *IS&T/SPIE Electronic Imaging*, pages 72401E–72401E. International Society for Optics and Photonics, 2009.
93. Wen-Hung Kuo, Po-Hung Lin, and Sheue-Ling Hwang. A framework of perceptual quality assessment on lcd-tv. *Displays*, 28(1):35–43, 2007.
94. Marc Lambooij, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53(3):30201–1, 2009.
95. Marc Lambooij, Wijnand IJsselsteijn, Don G Bouwhuis, and Ingrid Heynderickx. Evaluation of stereoscopic images: beyond 2d quality. *Broadcasting, IEEE Transactions on*, 57(2):432–444, 2011.
96. Peter J Lang. Behavioral treatment and bio-behavioral assessment: Computer applications. 1980.
97. Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. NIMH, Center for the Study of Emotion & Attention, 2005.
98. Julie Lassalle, Laetitia Gros, Thierry Morineau, and Gilles Coppin. Impact of the content on subjective evaluation of audiovisual quality: What dimensions influence our perception? In *Broadband Multimedia Systems and Broadcasting (BMSB), 2012 IEEE International Symposium on*, pages 1–6. IEEE, 2012.
99. Patrick Le Callet, Sebastian Möller, Andrew Perkis, et al. Qualinet white paper on definitions of quality of experience. *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, 2012.
100. Olivier Le Meur, Alexandre Ninassi, Patrick Le Callet, and Dominique Barba. Do video coding impairments disturb the visual attention deployment? *Signal Processing: Image Communication*, 25(8):597–609, 2010.
101. Barbara Lee and Robert S Lee. How and why people watch tv: Implications for the future of interactive television. *Journal of advertising research*, 35(6):9–18, 1995.
102. Jong-Seok Lee, Francesca De Simone, and Touradj Ebrahimi. Subjective quality evaluation via paired comparison: application to scalable video coding. *Multimedia, IEEE Transactions on*, 13(5):882–893, 2011.
103. Jong-Seok Lee, Francesca De Simone, Naeem Ramzan, Zhijie Zhao, Engin Kurutepe, Thomas Sikora, Jörn Ostermann, Ebroul Izquierdo, and Touradj Ebrahimi. Subjective evaluation of scalable video coding for content distribution. In *Proceedings of the international conference on Multimedia*, pages 65–72. ACM, 2010.
104. Congcong Li and Tsuhan Chen. Aesthetic visual quality assessment of paintings. *Selected Topics in Signal Processing, IEEE Journal of*, 3(2):236–252, 2009.
105. Weisi Lin and C-C Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312, 2011.
106. Hantao Liu and Ingrid Heynderickx. A perceptually relevant no-reference blockiness metric based on local image characteristics. *EURASIP Journal on Advances in Signal Processing*, 2009:2, 2009.
107. Hantao Liu and Ingrid Heynderickx. Visual attention in objective image quality assessment: based on eye-tracking data. *Circuits and Systems for Video Technology, IEEE Transactions on*, 21(7):971–982, 2011.

108. James Lull et al. *Inside family viewing: ethnographic research on television's audiences.* Routledge, 1990.
109. Yiwen Luo and Xiaoou Tang. Photo and video quality evaluation: Focusing on the subject. In *Computer Vision–ECCV 2008*, pages 386–399. Springer, 2008.
110. Wendy Ann Mansilla, Andrew Perkis, and Touradj Ebrahimi. Implicit experiences as a determinant of perceptual quality and aesthetic appreciation. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 153–162. ACM, 2011.
111. Claire Mantel, Nathalie Guyader, Patricia Ladret, Gelu Ionescu, and Thomas Kunlin. Characterizing eye movements during temporal and global quality assessment of h. 264 compressed video sequences. In *IS&T/SPIE Electronic Imaging*, pages 82910Y–82910Y. International Society for Optics and Photonics, 2012.
112. John D McCarthy, M Angela Sasse, and Dimitrios Miras. Sharp or smooth?: comparing the effects of quantization vs. frame rate for streamed video. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 535–542. ACM, 2004.
113. Ricky KP Mok, Edmond WW Chan, and Rocky KC Chang. Measuring the quality of experience of http video streaming. In *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*, pages 485–492. IEEE, 2011.
114. Anush Krishna Moorthy and Alan Conrad Bovik. Visual quality assessment algorithms: what does the future hold? *Multimedia Tools and Applications*, 51(2):675–696, 2011.
115. Margaret Morrison and Dean M Krugman. A look at mass and computer mediated technologies: Understanding the roles of television and computers in the home. *Journal of Broadcasting & Electronic Media*, 45(1):135–161, 2001.
116. JH Morrissey. New method for the assignment of psychometric scale values from incomplete paired comparisons. *JOSA*, 45(5):373–378, 1955.
117. Niall Murray, Yuansong Qiao, Brian Lee, Gabriel-Miro Muntean, and AK Karunakar. Age and gender influence on perceived olfactory & visual media synchronization. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6. IEEE, 2013.
118. Anja B Naumann, Ina Wechsung, and Jörn Hurtienne. Multimodal interaction: A suitable strategy for including older users? *Interacting with Computers*, 22(6):465–474, 2010.
119. A Neil. Autostereoscopic 3d displays. *Computer*, 8:32–36, 2005.
120. MRM Nijenhuis. *Sampling, interpolation, images: a perceptual view*. PhD thesis, Eindhoven University of technology, Eindhoven, 1993.
121. MRM Nijenhuis and FJJ Blommaert. Perceptual error measure for sampled and interpolated images. *Journal of Imaging Science and Technology*, 41(3):249–258, 1997.
122. Alexandre Ninassi, Olivier Le Meur, Patrick Le Callet, and Dominique Barba. Considering temporal variations of spatial visual distortions in video quality assessment. *Selected Topics in Signal Processing, IEEE Journal of*, 3(2):253–265, 2009.
123. Alexandre Ninassi, Olivier Le Meur, Patrick Le Callet, Dominique Barba, Arnaud Tirel, et al. Task impact on the visual attention in subjective image quality assessment. In *Proceedings of European Signal Processing Conference*, 2006.
124. Heather L O'Brien and Elaine G Toms. What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6):938–955, 2008.
125. Lora Oehlberg, Nicolas Ducheneaut, James D Thornton, Robert J Moore, and Eric Nickell. Social tv: Designing for distributed, sociable television viewing. In *Proc. EuroITV*, volume 2006, pages 25–26, 2006.
126. Joana Palhais, Rui S Cruz, and Mário S Nunes. Quality of experience assessment in internet tv. In *Mobile Networks and Management*, pages 261–274. Springer, 2012.
127. Thrasyvoulos N Pappas, Robert J Safranek, and Junqing Chen. Perceptual criteria for image quality evaluation. *Handbook of image and video processing*, pages 669–684, 2000.
128. Fernando Pereira. Sensations, perceptions and emotions towards quality of experience evaluation for consumer electronics video adaptations. In *Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005.
129. Pablo Pérez, Jesús Macías, Jaime J Ruiz, and Narciso García. Effect of packet loss in video quality of experience. *Bell Labs Technical Journal*, 16(1):91–104, 2011.

130. Lawrence A Pervin and Oliver P John. *Handbook of personality: Theory and research*. Elsevier, 1999.
131. Kandaraj Piamrat, Cesar Viho, J Bonnin, and Adlen Ksentini. Quality of experience measurements for video streaming over wireless networks. In *Information Technology: New Generations, 2009. ITNG'09. Sixth International Conference on*, pages 1184–1189. IEEE, 2009.
132. Scott Plous. *The psychology of judgment and decision making.* Mcgraw-Hill Book Company, 1993.
133. Jordi Puig, Andrew Perkis, Frank Lindseth, and Touradj Ebrahimi. Towards an efficient methodology for evaluation of quality of experience in augmented reality. In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pages 188–193. IEEE, 2012.
134. Jenni Radun, Tuomas Leisti, Jukka Häkkinen, Harri Ojanen, Jean-Luc Olives, Tero Vuori, and Göte Nyman. Content and quality: Interpretation-based estimation of image quality. *ACM Transactions on Applied Perception (TAP)*, 4(4):2, 2008.
135. Benjamin Rainer, Markus Waltl, Eva Cheng, Muawiyath Shujau, Christian Timmerer, Stephen Davis, Ian Burnett, Christian Ritz, and Hermann Hellwagner. Investigating the impact of sensory effects on the quality of experience and emotional response in web videos. In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pages 278–283. IEEE, 2012.
136. JC Read, SJ MacFarlane, and Chris Casey. Endurability, engagement and expectations: Measuring children's fun. In *Interaction design and children*, volume 2, pages 1–23. Shaker Publishing Eindhoven, 2002.
137. ITU-R BT Recommendation. 2021, subjective methods for the assessment stereoscopic 3dtv systems. *International Telecommunication Union, Geneva, Switzerland*, 2012.
138. ITURBT Recommendation. 500-11, methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union, Geneva, Switzerland*, 4:2, 2002.
139. ITUT Recommendation. E. 800: Terms and definitions related to quality of service and network performance including dependability. *ITU-T 2008*, 2008.
140. Judith Redi, Ingrid Heynderickx, Bruno Macchiavello, and Mylene Farias. On the impact of packet-loss impairments on visual attention mechanisms. In *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, pages 1107–1110. IEEE, 2013.
141. Judith Redi, Hantao Liu, Hani Alers, Rodolfo Zunino, and Ingrid Heynderickx. Comparing subjective image quality measurement methods for the creation of public databases. In *IS&T/SPIE Electronic Imaging*, pages 752903–752903. International Society for Optics and Photonics, 2010.
142. Judith Redi, Hantao Liu, Paolo Gastaldo, Rodolfo Zunino, and Ingrid Heynderickx. How to apply spatial saliency into objective metrics for jpeg compressed images? In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 961–964. IEEE, 2009.
143. Judith Redi, Hantao Liu, Rodolfo Zunino, and Ingrid Heynderickx. Interactions of visual attention and quality perception. In *IS&T/SPIE Electronic Imaging*, pages 78650S–78650S. International Society for Optics and Photonics, 2011.
144. Judith A Redi. Visual quality beyond artifact visibility. In *IS&T/SPIE Electronic Imaging*, pages 86510N–86510N. International Society for Optics and Photonics, 2013.
145. Judith A Redi and Ingrid Heynderickx. Image integrity and aesthetics: towards a more encompassing definition of visual quality. In *IS&T/SPIE Electronic Imaging*, pages 829115–829115. International Society for Optics and Photonics, 2012.
146. Judith A Redi and Isabel Povoa. The role of visual attention in the aesthetic appeal of consumer images: A preliminary study. In *Visual Communications and Image Processing (VCIP), 2013*, pages 1–6. IEEE, 2013.
147. Judith Alice Redi, Tobias Hoßfeld, Pavel Korshunov, Filippo Mazza, Isabel Povoa, and Christian Keimel. Crowdsourcing-based multimedia subjective evaluations: a case study on image recognizability and aesthetic appeal. In *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia*, pages 29–34. ACM, 2013.

148. Ulrich Reiter and Katrien De Moor. Content categorization based on implicit and explicit user feedback: combining self-reports with eeg emotional state analysis. In *Quality of multimedia experience (QoMEX), 2012 fourth international workshop on*, pages 266–271. IEEE, 2012.
149. Huib Ridder and Serguei Endrikhovski. 33.1: Invited paper: image quality is fun: reflections on fidelity, usefulness and naturalness. In *SID Symposium Digest of Technical Papers*, volume 33, pages 986–989. Wiley Online Library, 2002.
150. Raimund Schatz, Siegfried Wagner, Sebastian Egger, and Norbert Jordan. Mobile tv becomes social-integrating content with communications. In *Information Technology Interfaces, 2007. ITI 2007. 29th International Conference on*, pages 263–270. IEEE, 2007.
151. Jose A Scheinkman. Social interactions. *The New Palgrave Dictionary of Economics*, 2, 2008.
152. Dimitri Schuurman, Katrien De Moor, Lieven De Marez, and Tom Evens. A living lab research approach for mobile tv. *Telematics and Informatics*, 28(4):271–282, 2011.
153. Eric WK See-To, Savvas Papagiannidis, and Vincent Cho. User experience on mobile video appreciation: How to engross users and to enhance their enjoyment in watching mobile video clips. *Technological Forecasting and Social Change*, 79(8):1484–1494, 2012.
154. Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matthew Trentacoste, Abhijeet Ghosh, and Andrejs Vorozcovs. High dynamic range display systems. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 760–768. ACM, 2004.
155. Pieter Seuntiens, Lydia Meesters, and Wijnand Ijsselsteijn. Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric jpeg coding and camera separation. *ACM Transactions on Applied Perception (TAP)*, 3(2):95–109, 2006.
156. Pieter Seuntiens, Ingrid Vogels, and Arnold van Keersop. Visual experience of 3d-tv with pixelated ambilight. *Proceedings of PRESENCE*, 2007, 2007.
157. Pieter J Seuntiëns, Ingrid E Heynderickx, Wijnand A IJsselsteijn, Paul MJ van den Avoort, Jelle Berentsen, Iwan J Dalm, Marc T Lambooij, and Willem Oosting. Viewing experience and naturalness of 3d images. In *Optics East 2005*, pages 601605–601605. International Society for Optics and Photonics, 2005.
158. Mario Siller and John Woods. Improving quality of experience for multimedia services by qos arbitration on a qoe framework. In *in Proc. of the 13th Packed Video Workshop 2003*. Citeseer, 2003.
159. Paul J Silvia. Interestłthe curious emotion. *Current Directions in Psychological Science*, 17(1):57–60, 2008.
160. Joshua A Solomon, Andrew B Watson, and Albert Ahumada. Visibility of dct basis functions: Effects of contrast masking. In *Data Compression Conference, 1994. DCC'94. Proceedings*, pages 361–370. IEEE, 1994.
161. Nicolas Staelens, Stefaan Moens, Wendy Van den Broeck, Ilse Marien, Brecht Vermeulen, Peter Lambert, Rik Van de Walle, and Piet Demeester. Assessing quality of experience of iptv and video on demand services in real-life environments. *Broadcasting, IEEE Transactions on*, 56(4):458–466, 2010.
162. Nicolas Staelens, Glenn Van Wallendael, Karel Crombecq, Nick Vercammen, Jan De Cock, Brecht Vermeulen, Rik Van de Walle, Tom Dhaene, and Piet Demeester. No-reference bitstream-based visual quality impairment detection for high definition h. 264/avc encoded video sequences. *Broadcasting, IEEE Transactions on*, 58(2):187–199, 2012.
163. Wa James Tam, Lew B Stelmach, and Philip J Corriveau. Psychovisual aspects of viewing stereoscopic video sequences. In *Photonics West'98 Electronic Imaging*, pages 226–235. International Society for Optics and Photonics, 1998.
164. KT Tan, Mohammed Ghanbari, and Donald E Pearson. An objective measurement tool for mpeg video quality. *Signal Processing*, 70(3):279–294, 1998.
165. Cornelis Teunissen. Flat panel display characterization: a perceptual approach. 2009.
166. Louis L Thurstone. A law of comparative judgment. *Psychological review*, 34(4):273, 1927.
167. International Telecommunication Union. Itu-t rec. 109: Definition of quality of experience (qoe). *Liaison Statement, Ref.: TD 109rev2 (PLEN/12)*, 2007.

168. Vanessa Vakili, Willem-Paul Brinkman, and Mark A Neerincx. Lessons learned from the development of technological support for ptsd prevention: a review. *Stud Health Technol Inform*, 181:22–6, 2012.
169. Patricia Valdez and Albert Mehrabian. Effects of color on emotions. *Journal of Experimental Psychology: General*, 123(4):394, 1994.
170. Martin Varela, Toni Maki, Lea Skorin-Kapov, and Tobias Hossfeld. Towards an understanding of visual appeal in website design. In *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, pages 70–75. IEEE, 2013.
171. Elena Vicario, Ingrid Heynderickx, Giulio Ferretti, and Paola Carrai. 17.1: Design of a tool to benchmark scaling algorithms on lcd monitors. In *SID Symposium Digest of Technical Papers*, volume 33, pages 704–707. Wiley Online Library, 2002.
172. Ingrid MLC Vogels. How to make life more colorful: from image quality to atmosphere experience. In *Color and Imaging Conference*, volume 2009, pages 123–128. Society for Imaging Science and Technology, 2009.
173. ECL Vu and DM Chandler. Visual fixation patterns when judging image quality: Effects of distortion type, amount, and subject experience. In *Image Analysis and Interpretation, 2008. SSIAI 2008. IEEE Southwest Symposium on*, pages 73–76. IEEE, 2008.
174. Tero Vuori, Maria Olkkonen, Monika Pölönen, Ari Siren, and Jukka Häkkinen. Can eye movements be quantitatively applied to image quality studies? In *Proceedings of the third Nordic conference on Human-computer interaction*, pages 335–338. ACM, 2004.
175. Demin Wang, Filippo Speranza, Andre Vincent, Taali Martin, and Phil Blanchfield. Toward optimal rate control: a study of the impact of spatial resolution, frame rate, and quantization on subjective video quality and bit rate. In *Visual Communications and Image Processing 2003*, pages 198–209. International Society for Optics and Photonics, 2003.
176. Junle Wang, Damon M Chandler, and Patrick Le Callet. Quantifying the relationship between visual salience and visual importance. In *IS&T/SPIE Electronic Imaging*, pages 75270K–75270K. International Society for Optics and Photonics, 2010.
177. Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.
178. Zhou Wang and Xinli Shang. Spatial pooling strategies for perceptual image quality assessment. In *Image Processing, 2006 IEEE International Conference on*, pages 2945–2948. IEEE, 2006.
179. Andrew B Watson. Efficiency of a model human image code. *JOSA A*, 4(12):2401–2417, 1987.
180. Ina Wechsung, Matthias Schulz, Klaus-Peter Engelbrecht, Julia Niemann, and Sebastian Möller. All users are (not) equal-the influence of user characteristics on perceived quality, modality choice and performance. In *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*, pages 175–186. Springer, 2011.
181. Amaya Becvar Weddle and Hua Yu. How does audio-haptic enhancement influence emotional response to mobile media? In *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, pages 158–163. IEEE, 2013.
182. J.H.D.M. Westerink. Influences of subject expertise in quality assessment of digitally coded images. In *SID International Symposium Digest of Technical Papers*, volume 20, pages 124–127. SID, 1989.
183. Joyce HDM Westerink and Jacques AJ Roufs. Subjective image quality as a function of viewing distance, resolution, and picture size. *SMPTE journal*, 98(2):113–119, 1989.
184. Stefan Winkler and Subramanian Ramanathan. Overview of eye tracking datasets. In *QoMEX*, pages 212–217, 2013.
185. K Maria Wolters, Klaus-Peter Engelbrecht, Florian Gödde, Sebastian Möller, Anja Naumann, and Robert Schleicher. Making it easier for older people to talk to smart homes: the effect of early help prompts. *Universal Access in the Information Society*, 9(4):311–325, 2010.
186. Chen-Chi Wu, Kuan-Ta Chen, Yu-Chun Chang, and Chin-Laung Lei. Crowdsourcing multimedia qoe evaluation: A trusted framework. *IEEE transactions on multimedia*, 15(5):1121–1137, 2013.

187. Qianqian Xu, Qingming Huang, Tingting Jiang, Bowei Yan, Weisi Lin, and Yuan Yao. Hodgerank on random graphs for subjective video quality assessment. *Multimedia, IEEE Transactions on*, 14(3):844–857, 2012.
188. Qianqian Xu, Qingming Huang, and Yuan Yao. Online crowdsourcing subjective image quality assessment. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 359–368. ACM, 2012.
189. Qianqian Xu, Jiechao Xiong, Qingming Huang, and Yuan Yao. Robust evaluation for quality of experience in crowdsourcing. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 43–52. ACM, 2013.
190. Kyoko Yamori and Yoshiaki Tanaka. Relation between willingness to pay and guaranteed minimum bandwidth in multiple-priority service. In *Communications, 2004 and the 5th International Symposium on Multi-Dimensional Mobile Communications Proceedings. The 2004 Joint Conference of the 10th Asia-Pacific Conference on*, volume 1, pages 113–117. IEEE, 2004.
191. Akiko Yoshida, Volker Blanz, Karol Myszkowski, and Hans-Peter Seidel. Perceptual evaluation of tone mapping operators with real-world scenes. In *Electronic Imaging 2005*, pages 192–203. International Society for Optics and Photonics, 2005.
192. Ai Yoto, Tetsuo Katsuura, Koichi Iwanaga, and Yoshihiro Shimomura. Effects of object color stimuli on human brain activities in perception and attention referred to eeg alpha band response. *Journal of Physiological Anthropology*, 26(3):373–379, 2007.
193. H van Zee and DW Kaandorp. Kwaliteit en degradatie. ipo rapport no. 853. Technical report, Institute for Perception Research, Eindhoven, 1992.
194. Thomas Zinner, Oliver Hohlfeld, Osama Abboud, and Tobias Hoßfeld. Impact of frame rate and resolution on objective qoe metrics. In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, pages 29–34. IEEE, 2010.