QUALITY OF SERVICE AND QUALITY OF EXPERIENCE IN BROADCASTING

1. INTRODUCTION

Consumers select a specific product or service based on the quality of their experience with it. Every user evaluates the products and services according to their own perception. The expression Quality of Experience (QoE) was recently incorporated by the International Telecommunications Union (ITU), which defined it as The degree of delight or annoyance of the user of an application or service (1). Being a preliminary definition, the ITU also states that Recognizing on-going research on this topic, this is a working definition which is expected to evolve for some time. This definition was taken from (2), where it is also stated that the QoE results from the fulfillment of the user's 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.

In the telecommunications field it is more common to use the concept of Quality of Service (QoS), also defined by the ITU as The totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service (3). The characteristics to which the concept of QoS refers can be directly related to the technical aspects of the communications network (e.g., bit rate, delays, and error rate) or to the way of providing the service (e.g., service provisioning time and response time for claims handling). Although these aspects influence the final perception of the users (i.e., QoE), the definitions and measures of QoS are considered from the perspective of the service provider and not from the perspective of the end user.

The QoE concept can be applied to any field and is something essentially subjective, since it is associated with what the user feels regarding the use of some application, some product, or some service. The sensation we experience (i.e., the quality of our experience) will depend on a large number of factors: the type and characteristics of the application or service, the context of use, our own expectations, cultural background, socioeconomic factors, psychological profiles, and emotional status at the moment, among others.

The engineering challenge is to be able to evaluate and measure QoE, even though it is a clearly subjective characteristic. For engineering purposes, the QoE has to be represented by quantitative values. This can be through a numerical value (i.e., a single value that represents the QoE), a multidimensional representation (i.e., several numerical values, each representing a "dimension" or aspect of the QoE), and/or using verbal descriptors.

The best way to measure QoE is to ask people directly about their specific experience, on a previously defined scale. Historically, a scale with five possible responses has been used, ranging from Excellent (associated with a score of 5) to Bad (with a score of 1), as described in Table 1.

Beyond the individual experiences, it is interesting to evaluate what most people think using a statistical representation. The average of the opinions of several people is known by the acronym MOS (Mean Opinion Score), also defined by ITU (3). This historically scale is widely used for QoE evaluations, but other scales and ranges can also be used. Nevertheless, it may be questionable if the QoE of the new multimedia services can be qualified with a single and unique dimension, and some other proposals are arising (4).

In broadcasting services there are many different QoS parameters that can be measured. For example, Report ITU-R BT.2389 (5) describes the guidelines for measurements of different QoS parameters for digital terrestrial television in broadcasting systems, which are measurement of signal-to-noise ratio (SNR), carrier-to-noise ratio (CNR), modulation error rate (MER), phase jitter (PJ), bit error rate (BER), packet error rate (PER), and frame error rate (FER), among others. All of these parameters can be established and controlled by the broadcasters. Nevertheless, they are meaningless for the end users. The users can be comfortable or dissatisfied, according to other very different parameters that affect the QoE (e.g., zapping time, picture, and video quality). Since QoE depends on a very important number of factors, it is common to try to separate them and evaluate how each of them affects the quality of the overall experience. In broadcasting, the audio, image, and video quality are of paramount relevance. In Multimedia Services, the delay and stalling effects are also important. Even though other factors affect the QoE in Broadcasting and Multimedia Services, this article focuses on the audio and video quality aspects.

2. AUDIO QUALITY

Audio is of primary importance in broadcasting. Audio is a signal that is intended to carry sound information. It is expected to be converted to sound in an appropriate transducer and listened by subjects.

Voice, music, or ambience may be affected by coding and transmission constraints imposed by different standards in use.

There are many features of the signal that can affect the quality. In this section, we review the parameters of the audio signal that have special incidence on audio quality and the artifacts that can be introduced when it is coded or transmitted.

2.1. Audio Signal Parameters Relevant for Quality

A large number of parameters define the quality of a given audio signal. Recommendation ITU-R BS.644-1 (6) lists the parameters that are considered to be the most important in the analogue environment:

- Nominal bandwidth
- Amplitude/frequency response
- Group-delay variation
- Nonlinear distortion
- Error in reconstituted frequency
- Error in amplitude/amplitude response
- Level of stability

J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*. Copyright © 2021 John Wiley & Sons, Inc.
DOI: 10.1002/047134608X.W8415

Table 1. Typical Five-Point Scale for Quality Evaluation

Score	Description
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

- Noise (and single-tone interference)
- Disturbing modulation by power supply
- Stereo: level difference between A and B channels
- Stereo: phase difference between A and B channels
- Stereo: crosstalk between A and B channels

For decades, broadcast engineers considered these parameters as the primary indicators for audio quality. Precise measurement methods were used to measure each one.

Digitalization brought other characteristics of the processes imposed by engineers to the audio signal that define its quality. Digitalization processes, coding, and transmission losses are the main sources of degradations. The following list of parameters is added to the previous one:

- Quantization resolution
- Sampling frequency
- Tandem capability (due to codec cascading)
- Coding delay
- Error resilience
- Recovery time

Recommendation ITU-R BS.1548-7 (7) specifies the requirements that are relevant to the use of audio source coding systems in sound broadcasting, including television. It distinguishes the case of Contribution and Distribution, and Emission applications. All the categories should comply with certain considerations of audio quality.

The following three categories of audio quality listed in Table 2 are assumed at ITU-R BS.1548-7 (7) for broadcasting applications.

The quantization resolution shall be 20 bits for contribution, 18 for distribution, and 16 for emission.

For Categories (1) and (2), the sampling frequency shall be 48 kHz, and the bandwidth of the main audio channels 20–20 000 Hz and of the Low-Frequency Effects (LFE) channel 15–120 Hz. On the other hand, for intermediate quality emission, frequency shall be 32 or 48 kHz, but for very low bit rates, lower sampling frequencies are admitted (8, 11.025, 12, 16, 22.05, and 24 kHz), and the bandwidth depends on the selection.

Cascading is permitted, and it shall be possible to apply three codecs in cascade when in distribution applications and five codecs in cascade when in contribution.

Coding delay for all channels in a program must be identical and should be as low as possible. It should be noted that the computations implied in coding impose minimum delays. So, the coding performance required by the particular application (i.e., amount of bit rate reduction) may impose larger delays. Considering coding delay in the case of television sound (7), the delay of audio must be matched with the delay of video. It is desirable that the audio coder produces encoded audio frames (access units) that correspond exactly to the time period of the matching video frame.

A mechanism must be provided in the audio bit stream to allow the decoder to identify residual channel errors and to adopt proper concealment methods.

Table 2. Audio Categories

Category	Audio quality	Application	Bandwidth	Sampling frequency
(1)	Very high quality, with sufficient quality margin to allow cascade (concatenation) and postprocessing. Should be subjectively indistinguishable from the source for most types of audio program material.	Contribution, distribution, production, and postproduction	20–20 000 Hz (main channels) 15–120 Hz (low-frequency effects channels)	48 kHz
(2)	Subjectively transparent quality, sufficient for the highest quality broadcasting. Should be subjectively similar to the original signal for most types of audio program material.	High-quality (“CD quality”) emission	20–20 000 Hz (main channels) 15–120 Hz (low-frequency effects channels)	48 kHz
(3)	Equivalent to or better than good FM service quality, or equivalent to or better than good AM service quality. Using the MUSHRA method described in Recommendation ITU-R BS.1534, the mean score corresponding to “excellent” or “good” grade may be required.	Intermediate quality emission	The bandwidth depends on the sampling frequency	32 or 48 kHz For very low bit rates, lower sampling frequencies are admitted (8, 11.025, 12, 16, 22.05, and 24 kHz)

Source: Modified from ITU, “ITU-R BS.1548-7 - User requirements for audio coding systems for digital broadcasting.” 2019.

Table 3. Minimum Recommended Bitrates for Different Codecs in Use in Contribution and Distribution Applications

Codec	Minimum recommended bitrate per channel (kbits s^{-1})
MPEG-1 Layer II	180
MPEG-4 AAC	144
MPEG-H 3D	144
AC-4	128

For Contribution and Distribution applications, Recommendation ITU-R BS.1196 (8) recommends minimum bitrates for different codecs to achieve minimum audio quality, as summarized in Table 3.

For Emission applications, ITU-R BS.1548-7 (7) lists various audio codecs, declaring that all the included codecs fulfill the minimum requirements of audio quality as well as the minimum bitrate for each codec. Tables 4 and 5 show these minimum bitrates depending on the application.

2.2. Audio Artifacts

When coding and transmission is not subjectively transparent, there appears certain "artifacts" in the audio signal. Those artifacts make the audio signal to be perceived as different from the original one, thus affecting the audio quality. Feiten et al. (9) include a listing of the audio artifacts that are listed in Table 6.

3. IMAGE AND VIDEO QUALITY

Digital video and multimedia content, distributed through communications networks or transmitted by broadcasting, suffer various types of distortions or degradations during the process of acquisition, compression, processing, transmission, and reproduction. For example, the techniques commonly used in digital video coding introduce loss of information to reduce the bandwidth required for its transmission, which generates distortions. On the other hand, the packet networks on which the video is transported, for example, the Internet or the wired and wireless digital channels, may introduce additional distortions, due to delays, errors, and packet losses, among other factors.

3.1. Video Signal Parameters Relevant for Quality

In old analog video times, there were many parameters of the video signal that affected video quality such as:

- Bandwidth
- Dynamic range
- Chroma to luminance delay
- Autointerference of the delayed signal (ghost)
- Noise

However, the use of digitalization in production and transmission made them obsolete. Now, the parameters that affect video quality are related to the acquisition, compression, and transmission of the digitized video

Table 4. Minimum Recommended Bitrates for Different Codecs in Use in High-Quality Emission Applications

AAC LC profile	AAC LC with MPEG surround	AC-3/E-AC-3	MPEG-2 Layer II	AC-4	MPEG-H LC profile	DTS-UHD
Fulfilled at 144 kbit s^{-1} per 2 channels	Fulfilled at 384 kbit s^{-1} per 5 channels	Fulfilled at 192 kbit s^{-1} per 2 channels	Fulfilled at 256 kbit s^{-1} per 2 channels	Fulfilled at 96 kbit s^{-1} per 2 channels, at 192 kbit s^{-1} per 5 channels, and 288 kbit s^{-1} per 11.1 channels (system J)	Fulfilled at 768 kbit s^{-1} per 22.2 channels (system H)	Fulfilled 128, 192, 288 kbit s^{-1} per 2, 5, and 11 channels, respectively

Table 5. Minimum Recommended Bitrates for Different Codecs in Use in Intermediate Quality Emission Applications

HE-AAC	HE-AAC with MPEG surround	HE-AAC v2	Extended HE AAC	AC-4	MPEG-H LC profile	DTS-UHD
Fulfilled (excellent) at 48 kbit s^{-1} per 2 channels; fulfilled (good) at 32 kbit s^{-1} per 2 channels; fulfilled (good) at 24 kbit s^{-1} per 1 channel	Fulfilled (good) at 64 kbit s^{-1} per 5 channels	Fulfilled (good) at 24 kbit s^{-1} per 2 channels	Fulfilled (good) at 16 kbit s^{-1} per 2 channels; fulfilled (good) at 12 kbit s^{-1} per 1 channel	Fulfilled (excellent) at 48 kbit s^{-1} per 2 channels; fulfilled (excellent) at 128 kbit s^{-1} per 5.1 channels; fulfilled (excellent) at 256 kbit s^{-1} per 11.1 channels	Fulfilled (excellent) at 48 kbit s^{-1} per 2 channels; fulfilled (excellent) at 128 kbit s^{-1} per 5.1 channels (system B)	Fulfilled (excellent) at 64, 144, and 192 kbit s^{-1} per 2, 5, and 11 channels, respectively

Table 6. Audio Artifacts

Processing	Artifacts	Observations
Coding	Quantization noise	This noise is perceived as roughness when above a threshold
	Binaural unmasking distortions	When either the phase or level differences of the signal at the two ears are not the same as those of the masker
	Aliasing artifacts	—
	Timbre distortion (birdies)	Due to single bands being switched off
	Muffled audio (band limitation)	Due to high frequency loss
	Preechoes	Appears when an impulsive sound appears at the middle of a long frame. The decoded quantization noise spreads over the entire frame
	Rasping	Due to the use of spectral band replication
	Metallic sound	
	Tone trembling	
	Sparkling	
	Bubbling	
	Change of stereo impression	Original level and delay differences between the transmission channels are not exactly reconstructed
Transmission	Interruptions frame repetition	Due to packet loss. Concealment methods are not standardized. Hence, for a given audio frame loss, the different concealment implementations may result in different levels of perceived audio quality
	Stalling	While buffering, playing is interrupted, originating a perceived quality issue
	Asynchrony	Loss of audio packets may result in a shorter audio stream and may cause asynchrony in the case of an audiovisual signal

Source: Adapted from B. Feiten, M.-N. Garcia, P. Svensson, and A. Raake, 2014, "Audio Transmission," in Quality of Experience: Advanced Concepts, Applications and Methods, S. Möller and A. Raake, Eds. Cham: Springer International Publishing, pp. 229–245, 2014.

signal. Table 7 summarizes the parameters affecting video quality and their corresponding stages.

In the case of digital video broadcasting services delivered over broadband Internet protocol networks (10), some quality parameters at the IP layer can affect the image quality and the smooth playout of video content:

- Packet loss ratio
- Latency
- Jitter

The design of an IP multimedia delivery service should contemplate the definition of these parameters:

- Maximum packet rate per stream
- Maximum number of sustainable streams
- Maximum bandwidth per stream (or packet rate for a given packet size)
- Transport protocol to be used
- Frame size (transport layer)
- Packet size
- Allowed interpacket gap profile
- Maximum burst size

3.2. Video Artifacts

The artifacts that appear in the video signal due to digitalization and compression are various. As early as in 1992, Rec ITU-R BT.813 (11) enumerated the following picture quality degradation factors:

- Image blur
- Edge busyness
- False contouring
- Granular noise
- "Dirty window" effect
- Movement blur
- Jerkiness (interruptions of fluid motions)

Besides, some others can be added:

- Blockiness
- Staircase effect
- Mosaic pattern effect

When the video is digitally transmitted in frames or packets, some of them can be lost or delayed, causing other artifacts and impairments, such as:

- Slicing (slices that freeze or show garbage for a period of time due to packet loss)
- Freezing or stalling

The following paragraphs briefly describe the meaning of the most popular video artifacts that affect the QoE in digital images and video.

The blockiness effect is the most notorious of the perceived degradations in digital video. This effect has its origin in the fact that the image is divided into small blocks to perform the digital encoding process. This encoding uses a mathematical transformation, generally based on the Discrete Cosine Transformation (DCT). The block effect is presented as discontinuities at the edges of adjacent blocks when reconstructing the image. Within the same frame, the coarser the quantification is, the more visible the blockiness effect is. The quantification threshold from which the block effect is perceived depends on the type of image, the Spatial Information (SI) and the movement content, and the Temporal Information (TI). Generally, the effect is less perceived in images with high movements or in places with very high or very little brightness.

Table 7. Parameters Affecting Video Quality

Processing	Parameters
Acquisition	Number of pixels per frame (resolution) Sampled signals (mainly R,G,B or Y,CR,CB) Sampling frequency (for each signal) Sampling structure Bits per sample Bitrate of the output (uncompressed signal) Number of frames per second
Coding	The codec The implementation of the codec (codecs have many options and tools which a manufacturer has to configure and can include or not in a particular implementation) The bitrate of the output The Group of Pictures (GOP) structure The coded signals (mainly R,G,B or Y,CR,CB)
Transmission	Transmission bandwidth Packet losses that may be caused, depending on the service and medium, by: • low signal to noise • impulsive noise • cochannel Interference • adjacent-channel interference • low signal • congestion • delayed packets Protection against noise that involves (between others): • coding scheme • type of modulation • error correction techniques

Blurriness is basically loss of image details. While this may be due to images taken out of focus, it can also be an effect introduced by the digitalization process, occurring when the high orders of the digitalized coefficients are suppressed, which are the ones that provide the fine details within their blocks. For example, many popular codecs use the DCT for the digitalization process, and in order to reduce the bitrate of the coded video, some coefficients related to the fine details may be discarded. This degradation can also contribute to other effects, such as blockiness and mosaic.

Edge busyness is the distortion concentrated at the edges of objects. It is often caused by the use of coarse quantization levels during the encoding process within a block containing both edges delimiting smooth areas and pixels with a significantly different average level. The result is to produce distortion concentrated at the edges of objects, characterized by temporally varying sharpness or spatially varying noise.

The effect of quantizing the luminance values of the pixels leads to false contours being generated in the areas of gradual transitions in the places of transition from one quantized value to another. The effect is seen as abrupt changes of luminance in places where there should be a gradual transition. It can be especially visible on large monitors or televisions.

When the image contains diagonal borders or lines with respect to the vertical or horizontal axis, the “staircase effect” can be presented. Since the base images of the digitalization process are formed using horizontal and vertical patterns, they are not best suited to diagonal edges. If compression techniques that eliminate high spatial detail information are used, a staircase effect can be observed in the diagonals. When the sections adjacent to the edge have high contrast, the effect is especially noticeable and takes the name of “ringing.”

The “mosaic” effect occurs when the edges of all or a large part of the blocks of an image seem to not coincide. This effect is closely related to the blockiness effect.

Besides the video artifacts that are produced during the encoding process, the transmission over error prone networks can cause packet losses. These losses of information can produce slicing (i.e., slices that freeze or show garbage for a period of time) or even freezing of all the images for some period of time.

4. BROADCASTING AND MULTIMEDIA QUALITY

When a user watches TV programs or movies, audio and video together form a unitary experience. Beyond the considerations made for audio alone and for video alone, in each corresponding section, a user perceives a sense of unity that has its own QoE when they are combined.

There are some parameters that are identified as affecting the QoE of the whole experience (12).

- Correlation between picture and sound images:
  - Correlation of source positions derived from visual and audible cues (including azimuth, elevation, and depth).
  - Correlation of spatial impressions between sound and picture.
  - Time relationship between audio and video.
  - Effect of visual image on basic audio quality.
- Harmony of spatial impressions of picture and sound.
- Assessment of listening and viewing arrangements.

The relative delay between video and audio, if large enough, is especially notable when performers are talking, as lip movements do not coincide with the speech.

The permissible time difference between sound and picture is defined in Recommendation ITU-R BT.1359 (13). This recommendation establishes the values for detectability and acceptability thresholds for asynchrony that result, respectively, in 125 and 200 ms when video leads the audio, and 45 and 100 ms when audio leads.

Synchronization of digital audio signals is a necessary function for the exchange of signals between equipment. The objective of synchronization is primarily to time align sample clocks within digital audio signal sources and align them with video frames/fields. Recommendation ITU-R BS.2032 (14) provides methods for synchronizing interconnected digital audio equipment and to address synchronization of the audio sample clocks to video reference signals.

5. SUBJECTIVE EVALUATIONS

The most reliable way to measure the quality of an audio, image, or video is through subjective evaluation, carried out by a group of people who judge about their perception when exposed to the content. The MOS is the metric generally accepted as a measure of quality. This subjective evaluation must be carried out under controlled conditions, so that all the people who participate in the evaluation procedure listen to the audio or see the image or video under the same conditions and are instructed to carry out their evaluations in the same way. This leads to the need to standardize the way in which these evaluations are carried out, through international recommendations, which then allow comparing the results of different evaluations carried out in different places and by different people. For example, Recommendation ITU-R BT.500 (15) has been specially developed for video quality measurements in TV applications; Recommendation ITU-R BT.710-4 (16) for video in High-Definition (HD) TV; and Recommendation ITU-T P.913 (9) for the subjective assessment of video, audio, and audiovisual quality of television in any environment. For stereoscopic TV, also known as 3DTV, Recommendation ITU-R BT.2021 (17) describes the way to make subjective tests. Recommendation ITU-R BS.1284-2 (18) establishes the general guidelines for subjective testing of audio quality.

The subjective evaluations are performed according to the corresponding Recommendation. A set of reference material is selected. These references consist of high-quality short audio, video, or multimedia clips, according to the type of test. In most cases, 10–12 seconds long is enough for each reference (an exception to this rule is when long-term continuous quality tests are performed). The content of each reference is selected according to the desired aspect to be evaluated in the test. These references are known as the Sources (SRC). For each SRC, a set of degraded materials are prepared. Each of the degraded sequences represents some type of Hypothetical Reference Circuit (HRC), a fixed combination of an encoder, operating at a given bit rate, with some kind of network condition and with a specific decoder. The HRC and SRC are generically called “stimulus” and are presented to the evaluators to obtain their judgment. In the Recommendations, different methods for presenting the stimulus and different evaluation scales are presented. The most simple and commonly used methods for subjective evaluation can be classified as Single Stimulus (SS) and Stimulus Comparison (SC) and are described in the following subsection.

5.1. Single Stimulus (SS)

In the SS methods, the subject is presented with one stimulus and rates that stimulus in isolation. The stimuli may or may not include the reference. If the reference is included, it is presented as an independent stimulus that qualifies as any other. Samples are presented randomly for each observer. In Figure 1 a timeline of an SS test is presented. The subjects are informed for a few seconds that a new video will be presented. The video identifier (i.e., a consecutive number) is shown, and then the video is presented (i.e., a 10-seconds clip). Immediately after the end of the video clip, a gray screen instructs the subjects to rate the video quality. After the subjects make their rating, the cycle starts again, and for a few seconds the announcement of the next video is presented.

In a variant of these methods called Single Stimulus with Multiple Repetition (SSMR), the stimuli are presented in the same way as in the SS methods, but the same stimulus is repeated more than once. In this way, the dependence of the order of presentation of the stimuli is minimized.

One of the most commonly used methods is the Absolute Category Rating (ACR). It is one of the most widespread SS methods due to its simplicity. The method consists in presenting the samples consecutively, one after the other, and in asking the subject to evaluate them independently, using a discrete five-level scale, similar to the one presented in Table 1. For the evaluation, the subjects can use predefined templates in paper, or specific software applications, as those presented in (19, 20).

A variant of ACR, called Absolute Category Rating with Hidden Reference (ACR-HR), includes a reference between the samples without telling the evaluators where it is located.

The Single Stimulus Continuous Quality Evaluation (SSCQE) method (15) replicates the typical conditions in which the stimulus is presented continuously and without reference. It allows to assess long sequences (i.e., more than 5 minutes), which is more similar to real services. For its execution it is necessary to use a specially designed electronic device that allows the evaluators to perform qualifications while the stimulus is reproduced on a continuous scale. The device records the user’s continuous ratings on a graph that presents the “waveform” of the ratings as a function of time. For audio, this method

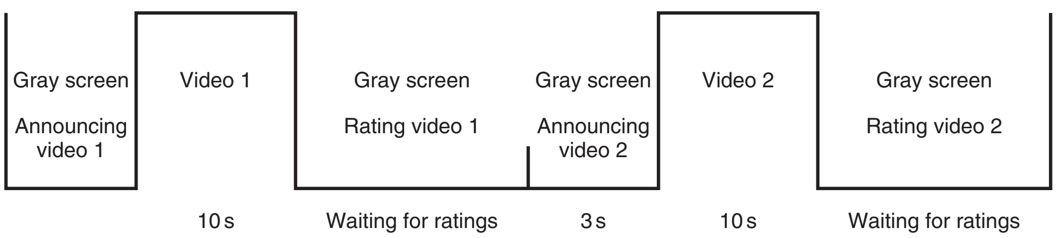

Figure 1. Timeline for Single Stimulus subjective evaluation of video.

is often called Continuous Evaluation of Time-Varying Speech Quality (CETVSQ) (21).

5.2. Stimulus Comparison (SC)

In the SC methods, the evaluators must compare two stimuli that are presented to them. They can be presented simultaneously or one after the other.

The Degradation Category Rating (DCR) method presents stimuli in pairs. The first stimulus presented is always the reference, while the second is the degraded. DCR uses a Double Stimulus Impairment Scale (DSIS), typically with five levels, according to Table 8.

In the Comparison Category Rating (CCR), also known as Double Stimulus Comparison Scale (DSCS), the stimuli are presented in pairs and in random order, and the evaluators rate the second sample with respect to the first one. It can be used to compare a sample with the reference or two degraded samples. The evaluators should rate the degradation of the second stimulus in relation to the first using a seven-level scale as described in Table 9.

The Double Stimulus Continuous Quality Scale (DSCQS) method presents videos in pairs (reference/sample) without indicating what the reference is. The reference/sample pairs are repeated once or several times (depending on the duration of the sequences), and it is requested to evaluate both videos on a continuous scale with anchors. In this method, the difference between the ratings is evaluated.

The Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) method aims to be more similar to the real user experience in which the perception of quality occurs continuously, so it is suitable for long sequences. The procedure consists of showing both sequences simultaneously, where the first sequence is the reference and the other one is the sequence under study. Evaluators must qualify continuously using a special hardware device such as that used in the SSCQE method.

Other methods that are not described here can also be used, such as MUltiple Stimulus with Hidden Reference and Anchor (MUSHRA), Subjective Assessment of Multimedia VIdeo Quality (SAMVIQ), and ABX Double Blind Comparator System.

Table 8. DCR Values

Score	Description
5	Imperceptible
4	Perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Table 9. CCR Values

Score	Description
−3	Much worse
−2	Worse
−1	Slightly worse
0	The same
1	Slightly better
2	Better
3	Much better

6. OBJECTIVE EVALUATIONS

Subjective measurement methods, performed according to the international recommendations in controlled environments, are the most reliable way to measure the user's QoE. But, on the other hand, they are difficult to implement and cannot be used for real-time applications. It takes a long time to prepare the conditions under which the test is carried out, to involve a group of "nonexpert" people who have the will and the time to carry out the evaluations, and an important pre- and postprocessing time. In the past years, different evaluations and standardized efforts have been made, and are currently ongoing, in order to obtain *objective* models and algorithms to predict the perceived audio, video, and multimedia quality in different scenarios. Objective evaluations are based on the information of the multimedia content, the network, the encoding process, and/or other measurable values.

The Video Quality Experts Group (VQEG) (22) and the Qualinet Organization (23) have been working in the design and comparison of different objective models applied to different scenarios. Based on these and other works, different standards of objective models for QoE prediction have been approved.

In a general way, the quality models that predict the users QoE can be classified into Full Reference (FR), Reduced Reference (RR), and No Reference (NR) models (24). In the first class, FR models, the original and the degraded samples are directly compared by the algorithm. In the RR models, some reduced information about the original audio or video is needed, and is used along with the degraded one, in order to estimate the perceived multimedia quality. NR models are based only on the degraded media in order to make an estimation of the perceived quality.

The first standardized video quality objective model was approved in 2004 in the Recommendations ITU-T J.144 (25) and ITU-R BT.1683 (26). These recommendations are suitable for Standard Definition (SD) TV in digital environments without network errors and were developed in the "Full Reference TV" (FRTV) project performed by VQEG between 2000 and 2003. Other remarkable standardizations are ITU-T J.246 (27) (perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference) and ITU-T J.247 (objective perceptual multimedia video quality measurement in the presence of an FR) (27), both result of the Multimedia Phase I work of VQEG between 2005 and 2008. ITU-T J.249 (perceptual video quality measurement techniques for digital cable television in the presence of an RR) (28) was the result of the "Reduced Reference and No Reference TV" (RRNR-TV) VQEG group in 2009. Similarly, ITU-T J.341 and ITU-T J.342 were the standardization of the work performed by the "High-Definition TV" (HDTV) VQEG project ended in 2011. Finally, Recommendation

ITU-T J.343 (Hybrid perceptual bitstream models for objective video quality measurements) (29) was the result of the "Hybrid Perceptual/Bitstream" work of VQEG in 2014.

Although a great effort has been made in the field of objective models for QoE evaluation, there is still a long way to go. Currently, there is no "general" model that can be applied to any scenario. Also, NR models are still not good enough to be standardized.

7. TRENDS

With the convergence of second-screen adoption and the abundance of real-time news consumption via social channels, the broadcast landscape underwent a major transformation in the past years and will continue to do so in the near future. One thing that has become clear: Viewers have begun to demand highly customized experiences that meet their individual needs. According to a recent report conducted by Nielsen (30), 84% of smartphone and tablet owners say they use their devices as second screens while watching television at the same time. Others are growing wary of traditional broadcast television packages altogether and choose other avenues, such as Netflix and YouTube, to watch TV.

Home entertainment systems continue to evolve in size and sophistication, delivering new levels of experience and adventure to consumers. Home entertainment has expanded to become a true indoor electronic playground for children and adults alike, with systems incorporating large-screen displays, gaming consoles, audio equipment, and docking stations. Often, a single remote control gives users complete command of these environments. Delivering innovative capabilities induce consumers to upgrade and expand their systems. It has been this insatiable consumer appetite for a better life that has driven technologists, content delivery companies, and service providers to place greater emphasis on the technologies and standards required to deliver, transfer, and store entertainment and multimedia content into and throughout the home, all with high QoS. Global service providers have begun to offer advanced whole-home video delivery, on the customer premise, enabling consumers to enjoy the emergence of new services offered by IPTV, 3DTV, SU-U-HDTV, HbbTV, advancements in cloud services, and over-the-top (OTT) content providers. Consumers can get significantly higher network speeds these days from Internet providers. OTT providers (such as Netflix, Amazon, and others) are offering movies and TV shows for either download or direct streaming over the Internet, which are the type of shows that consumers want to watch on a big-screen HD TV. In addition, television makers and retailers are betting big on Web-connected televisions (such as Apple TV and Google TV).

In short, the evolving wants and needs of the viewer is the future of broadcast television. Consumers are in control. In near future, this will become even more evident, with more people demanding customized television experiences through user-generated content and the option of microbundled packages. To keep up, broadcasters must stay updated with the latest innovations to adequately engage their customers.

For years, technology companies have been predicting linkage of TVs, computers, digital video recorders, game consoles, and other electronic gadgets together in a home network, allowing consumers to share content among devices. Now consumers have finally begun to enjoy such networked entertainment as multiroom digital video recorder services offered by payTV providers such as cable operators, Satellite operators, and Telcos around the world. New services such as video conferencing and telepresence, which were once seen only in science fiction, are also finally getting into the consumer's home. New technologies now enable every conceivable area of control for residential environments: control of utilities such as lighting; heating, ventilating, and air conditioning (HVAC); automated window treatments; pool and spa controls; control of security (such as garage and access controls); control of home appliances locally or remotely from a smartphone; and entertainment systems (such as home theaters and digital video recorders).

Nowadays, the concept of networked system can be leveraged by cyber-physical system (CPS) and Internet of Things (IoT). A CPS is a system of collaborating computational elements controlling physical entities. Today, a precursor generation of CPSs can be found in areas as diverse as aerospace, automotive, chemical processes, civil infrastructure, energy, healthcare, manufacturing, transportation, entertainment, and consumer appliances. This generation is often referred to as embedded systems. In embedded systems, the emphasis tends to be more on the computational elements and less on an intense link between the computational and physical elements. Unlike more traditional embedded systems, a full-fledged CPS is typically designed as a network of interacting elements with physical input and output instead of as standalone devices (31). The notion is closely tied to concepts of robotics and sensor networks with intelligence mechanisms proper of computational intelligence leading the pathway. Ongoing advances in science and engineering will improve the link between computational and physical elements by means of intelligent mechanisms, dramatically increasing the adaptability, autonomy, efficiency, functionality, reliability, safety, and usability of CPSs (32). This will broaden the potential of CPSs in several dimensions, including intervention (e.g., collision avoidance); precision (e.g., robotic surgery and nanolevel manufacturing); operation in dangerous or inaccessible environments (e.g., search and rescue, firefighting, and deep-sea exploration); coordination (e.g., air traffic control and war fighting); efficiency (e.g., zero-net energy buildings); augmentation of human capabilities (e.g., healthcare monitoring and delivery); and enhancing human experiences (33).

Over recent past years there has been a growing interest in the ability of embedded devices, sensors, and actuators to communicate and create a ubiquitous cyber-physical world. The growth of the notion of the IoT and the rapid development of technologies such as short-range mobile communication and improved energy efficiency are expected to create a pervasive connection of "things"

(34). Drastic growth in the number of smart devices and sensors (i.e., CPS) connected to the IoT has the potential to change how consumers interact with all networked technology, including their media and entertainment platforms. This represents an opportunity for the entertainment industry to assimilate the growing volume of customer insight that will be constantly generated by IoT technologies throughout the market in order to drive more responsive and interactive offerings. The proliferation of connected devices and IoT technologies to support them could redefine the level of interaction between entertainment providers and their customers. It can enhance our lives by optimizing our entertainment experiences in an intuitive and automated manner. In short, IoT and the cloud-based resources that connect devices to powerful real-time analytical engines are rapidly evolving to truly surround each of us to create an immersive environment that will augment a rapidly growing array of our experiences.

Within this framework, there is a need of metrics able to evaluate the level of enhancement of these experiences. As stated earlier in this article, QoE can be seen as an evolution from the QoS, both defined by the ITU-T in P.10/G.100 (1). QoS is defined as the totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service, whereas QoE is defined as the overall acceptability of an application or service, as perceived subjectively by the end-user. Although this definition was largely used (but not necessarily agreed), one could easily understand that acceptability is only one aspect of quality, as one may accept a service – depending on the context – but not necessarily be happy or satisfied. Therefore, the COST Action IC1003 – QUALINET (23) goes a step beyond and defines QoE as the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state. The QUALINET white paper even goes further and defines influence factors as any characteristic of a user, system, service, application, or context whose actual state or setting may have influence on the Quality of Experience for the user which are grouped into human, system, and context influence factors. Additionally, features of QoE are provided depending on the level of direct perception, interaction, the usage situation, and service. A QoE feature is thus defined as a perceivable, recognized and nameable characteristic of the individual's experience of a service which contributes to its quality. Moreover, 5G is gaining more and more attention as a possible vehicle for these new features, whereas artificial intelligence (AI) through machine learning (ML) is starting to be adopted to predict user/network behavior. New technologies such as multisensorial media, light field, virtual reality/augmented reality (VR/AR), holographic screens, and the proliferation of connected devices through the IoT paradigm could create an immersive environment that will enrich a rapidly growing array of customer experiences and become the next frontier of advanced broadcast services. In Murroni et al. (35) a comprehensive overview of advances in research and in the state-of-the art technologies of fundamental areas that are critical to the QoE for advanced broadcast services is provided.

8. CONCLUSIONS

The concept of QoS is directly related to the technical aspects of a communications network (e.g., bit rate, delays, and error rate) or the way the operator provides the service (e.g., service provisioning time and response time for claims handling). On the other hand, the concept of QoE is directly related to the user's feelings and results from his or her satisfaction, which in turn is related to the fulfillment of the user's expectations with respect to the perception of the application or service, in the light of his or her personality and current state. When applying these concepts to broadcasting and multimedia services, the video and audio quality are of paramount importance. Current broadcasting and multimedia services relay on digitalized media and transmission systems that natively introduce different types of degradations, resulting in impairments to the video and audio quality. Controlling the QoS (e.g., bitrates and error rates), some aspects of the digital chain (from encoding, transmission, and reception) can be controlled. But the way in which the remainder degradations and impairments affect the user's QoE is not trivial. This article has described the main concepts of QoS and QoE, specifically applied to audio and video in the Broadcasting and Multimedia field.

RELATED ARTICLES

Methods for Image Quality Assessment; Image and Video Coding; Multimedia Audio; Multimedia Video

BIBLIOGRAPHY

1. ITU. ITU-T P.10/G.100 – Vocabulary and Effects of Transmission Parameters on Customer Opinion of Transmission Quality, 2017.
2. K. Brunnström and S. A. Beker. Qualinet White Paper on Definitions of Quality of Experience, in *Presented at the Fifth Qualinet Meeting*, 2013. Available at https://hal.archives-ouvertes.fr/hal-00977812.
3. ITU. ITU-T E.800 – Definitions of Terms Related to Quality of Service, 2008.
4. R. C. Streijl, S. Winkler, and D. S. Hands. *Multimed. Syst.* **2014**, *22*, pp 213–227.
5. ITU. ITU-R BT.2389-0 – Guidelines on Measurements for Digital Terrestrial Television Broadcasting Systems, 2016.
6. ITU. ITU-R BS.644-1 – Audio Quality Parameters for the Performance of a High-Quality Sound-Programme Transmission Chain, 2002.
7. ITU. ITU-R BS.1548-7 – User Requirements for Audio Coding Systems for Digital Broadcasting, 2019.
8. ITU. ITU-R BS.1196 – Audio Coding for Digital Broadcasting, 2019.
9. B. Feiten, M.-N. Garcia, P. Svensson, and A. Raake. In *Quality of Experience: Advanced Concepts, Applications and Methods*;

Möller, S.; Raake, A., Eds.; Springer International Publishing: Cham, 2014; pp 229–245.
10. ITU. ITU-R BT.1720 – Quality of Service Ranking and Measurement Methods for Digital Video Broadcasting Services Delivered Over Broadband Internet Protocol Networks, 2005.
11. ITU. ITU-R BT.813 – Methods for Objective Picture Quality Assessment in Relation to Impairments from Digital Coding of Television Signals, 1992.
12. ITU. ITU-R BS.2126-0 – Methods for the Subjective Assessment of Sound Systems with Accompanying Picture, 2019.
13. ITU. ITU-R BT.1359 – Relative Timing of Sound and Vision for Broadcasting, 1998.
14. ITU. ITU-R BS.2032 – Synchronization of Digital Audio Sample Clock to Video References, 2013.
15. ITU. ITU-R BT.500-14 – Methodology for the Subjective Assessment of the Quality of Television Pictures, 2019.
16. ITU. ITU-R BT.710-4 – Subjective Assessment Methods for Image Quality in High-Definition Television, 1998.
17. ITU. ITU-R BT.2021 – Subjective Methods for the Assessment of Stereoscopic 3DTV Systems, 2015.
18. ITU. ITU-R BS.1284-2 – General Methods for the Subjective Assessment of Sound Quality, 2019.
19. R. Sotelo, J. Joskowicz, and A. Gonzalez. 4K Video Quality Assessment: An Application to Support Subjective Tests, in *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Valencia, Spain, 2018.
20. J. Joskowicz, R. Sotelo, M. Juayek, D. Durán, and J. P. Garella. Automation of Subjective Video Quality Measurements, in *LANC'14 – Latin America Networking Conference*, Montevideo, Uruguay, 2014.
21. ITU. ITU-T P.880 Continuous Evaluation of Time Varying Speech Quality, 2004.
22. Video Quality Experts Group (VQEG). Available at https://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx (accessed December 15, 2019).
23. QUALINET – COST IC 1003. Qualinet. European Network on Quality of Experience in Multimedia Systems and Services. Available at http://www.qualinet.eu/ (accessed December 15, 2019).
24. Y. Chen, K. Wu, and Q. Zhang. *IEEE Commun. Surv. Tutor.* **2015**, *17*(2), pp 1126–1165. doi: 10.1109/COMST.2014.2363139.
25. ITU. ITU-T J.144 – Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference, 2004.
26. ITU. ITU-R BT.1683 – Objective Perceptual Video Quality Measurement Techniques for Standard Definition Digital Broadcast Television in the Presence of a Full Reference, 2004.
27. ITU. ITU-T J.247 – Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference, 2008.
28. ITU. ITU-T J.249 – Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Reduced Reference, 2010.
29. ITU. ITU-T J.343 – Hybrid Perceptual Bitstream Models for Objective Video Quality Measurements, 2018.
30. Nielsen Group. The Nielsen Total Audience Report Q2 2018. Available at https://www.nielsen.com/us/en/insights/report/2018/q2-2018-total-audience-report/ (accessed April 10, 2020).
31. E. Lee. Cyber Physical Systems: Design Challenges. University of California, Berkeley. Technical Report No. UCB/EECS-2008-8, 2008.
32. C. Alippi. *Intelligence for Embedded Systems*. Springer Verlag, 2014.
33. The National Science Foundation. 'Cyber-Physical Systems'. Program Announcements & Information, 2008.
34. L. Atzori, A. Iera, and G. Morabito. *Comp. Net.* **2010**, *54*(15), pp 2787–2805.
35. M. Murroni, R. Rassool, L. Song, and R. Sotelo. *IEEE Trans. Broadcast.* **2018**, *64*(2), pp 335–340.

Maurizio Murroni
DIEE-University of Cagliari, Cagliari, Italy

Jose Joskowicz
Universidad de la Republica, Montevideo, Uruguay

Rafael Sotelo
Universidad de Montevideo, Montevideo, Uruguay