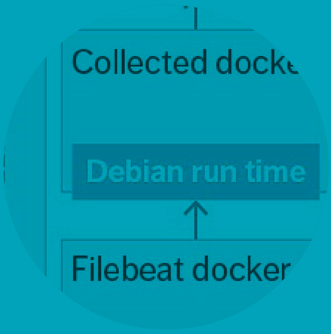
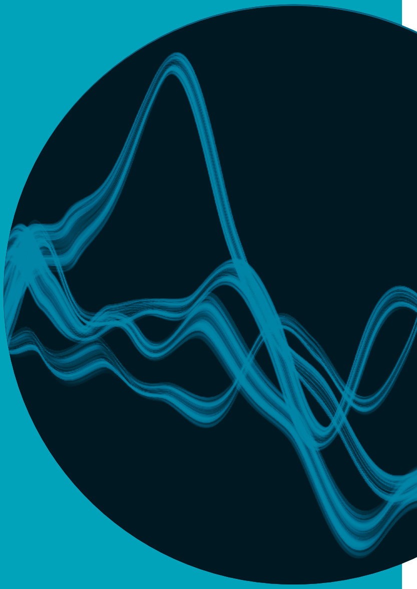
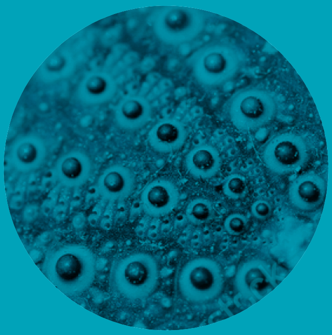
CHARTING THE FUTURE OF INNOVATION | #6·2017

# ERICSSON TECHNOLOGY Review





## LEVERAGING STANDARDS FOR VIDEO QoE



ERICSSON

# Video QoE

## LEVERAGING STANDARDS TO MEET RISING USER EXPECTATIONS

'How happy are our users with their video experience?' has become a vital question for mobile network operators and media service providers alike. New standards for QoE testing have the potential to help them ensure that they are able to meet user expectations for the service that will account for three-quarters of mobile network traffic in five years' time.

GUNNAR HEIKKILÄ,
JÖRGEN GUSTAFSSON

**In 2016, Ericsson ConsumerLab found that the average person watches 90 minutes more TV and video every day than they did in 2012 [1]. While traditionally broadcast linear TV is still popular with many viewers, internet-based TV and video delivery and video on demand (VOD) services are all growing rapidly. In fact, 20 percent of video consumption occurred on handheld devices in 2016 [1].**

■ As users become more accustomed to video streaming services, their quality expectations rise, presenting a big challenge for media service providers (MSPs) and mobile network operators (MNOs). While delivering high-quality video over a fixed connection can be difficult, doing so wirelessly is a much more demanding undertaking. Yet by 2022, 75 percent of all mobile data traffic is expected to come from video, according to the 2016 Ericsson Mobility Report [2].

At the same time, TV screens are getting larger, which requires higher video resolution. High definition (HD) is now the new baseline, and ultra high definition (UHD) is coming to both fixed and mobile devices, demanding higher bandwidth. To provide a consistently high QoE – particularly in wireless cases with significant fluctuations in available bandwidth – both MSPs and MNOs must have a clear understanding of which impairments their users are experiencing and be able to accurately assess their perception of video quality.

### Adaptive HTTP-based streaming

Adaptive HTTP-based video streaming variants such as DASH and HLS are the dominant video delivery method used today. Their streaming servers provide several versions of the same video, each separately encoded to offer varying levels of quality. The streaming client in the user's phone, tablet or PC dynamically switches between the different versions during playout depending on how much network bitrate is available.

If the available network bitrate is high, the client will download the best-quality version. If the bitrate suddenly drops, the client will switch to a lower-quality version until conditions improve. The purpose of this is to avoid stalling (when the client stops playout intermittently to fill its video buffer) – which is known to annoy users.

While the ability to switch intermittently between versions of a video significantly decreases the risk of stalling, the quality variations that this leads to can also be annoying for the user. *Figure 1* provides an example of a worst-case scenario with both quality variations and a stalling event, which would result in an overall low-quality experience for the user.

### Subjective quality

ITU-T P.910 [3] is the recognized standard for performing subjective video quality tests. The tests are carried out in a lab equipped with mobile phones, tablets, PCs or TVs, where a number of videos are shown to a group of individuals. Each individual then grades each video according to their subjective perception of its quality, selecting one of the following scores: 5 (excellent), 4 (very good), 3 (good), 2 (fair) or 1 (poor). Finally, the average score for each video is calculated. This number is known as the mean opinion score (MOS).

> “THE TESTS ARE DESIGNED TO MAKE ANALYSIS AS ACCURATE AND STRAIGHTFORWARD AS POSSIBLE”

The video sequences are typically produced in a lab environment so that all types of impairments can be included. High-quality sports, nature and news videos are used as a starting point. Impairments (codec settings, rate and resolution changes, initial buffering, stalling and so on) are then emulated by varying the bitrate over a certain range, for example, or placing stalling events of different lengths at various points in the videos.

The tests are designed to make analysis as accurate and straightforward as possible. Devising subjective tests is a time consuming and expensive process, though, and lab tests can't assess exactly what an MNO's real users are experiencing. The best way to overcome these challenges is by using objective quality algorithms.

**Terms and abbreviations**

**APN** – Access Point Name | **AVC** – advanced video coding | **DASH** – Dynamic Adaptive Streaming over HTTP | **DM** – device management | **eNB** – eNodeB (LTE base station) | **ETSI** – European Telecommunications Standards Institute | **HD** – high definition | **HEVC** – High Efficiency Video Coding | **HLS** – HTTP Live Streaming | **HTTP** – Hypertext Transfer Protocol | **ITU-T** – International Telecommunication Union Telecommunication Standardization Sector | **MNO** – mobile network operator | **MOS** – mean opinion score | **MPD** – media presentation description | **MSP** – media service provider | **OMA** – Open Mobile Alliance | **Pa** – short-term audio predictor | **Pq** – long-term quality predictor | **Pv** – short-term video predictor | **QoE** – quality of experience | **RRC** – Radio Resource Control | **SMS** – short message service | **UHD** – ultra high definition | **VOD** – video on demand | **VP9** – An open-source video format and codec

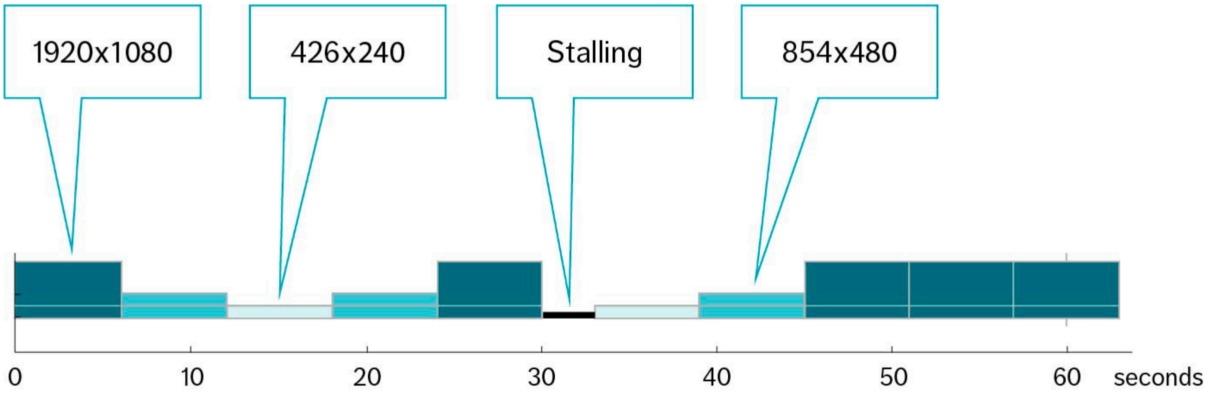


*Figure 1* A 60-second video with quality variations (resolution) and a three-second stall in the middle

“IDEALLY, DIFFERENT LEVELS OF COMPLEXITY SHOULD BE USABLE BY A SINGLE MODEL”

### Objective quality

As the wording suggests, objective quality algorithms (also known as objective models) are designed to mimic the behavior and perception of humans. The goal is to produce the same scores as the MOS values that would result from running a subjective test on the same videos.

Many different types of objective models can be adopted, depending on the intended usage and the kind of input data employed. Those using the most limited set of input parameters base the objective quality estimation on encoding rates, video resolution, frame rates, codecs and stalling information, as these factors provide the minimum amount of information about the video playout that is required to estimate a quality score. More complex models might use the complete encoded video bitstream, or even the full received video signal, to further increase the estimation accuracy.

The objective models described above are no-reference models, where input is taken only from the receiving end of the media distribution chain. Full-reference models can also be adopted, where the video originally transmitted is compared with the one that is received. Another variant is the reduced-reference model, where the original video is not needed for reference, but certain information about it is made available to the model.

Traditionally, objective models are used to evaluate quality based on relatively short video sequences: approximately 10 seconds long. However, with adaptive video streaming, where quality can vary significantly during a given session, the model must also assess how this long-term variation affects user perception. To do this, model evaluation of much longer video sequences (up to several minutes) is required.

Ideally, different levels of complexity should be usable by a single model – that is, from only a few input parameters up to the full bitstream, depending on deployment. This is the scope that applies with the new no-reference ITU-T P.1203 standard.

### Standardization of ITU-T P.1203

The P.1203 standard [4] was developed as part of an ITU-T competition between participating proponents (Ericsson and six other companies), where each one sent in its proposed quality models. The models that performed best were then used as a baseline to create the final standard. An internal model architecture was also defined to facilitate the creation of a model that would be as flexible as possible.

#### Architecture

The standard includes modules for estimating short-term audio and video quality, and an integration module estimating the final session quality due to adaptation and stalling, as shown in *Figure 2*.

The short-term video and audio predictor (Pv and Pa) modules continuously estimate the short-term audio and video quality scores for one-second pieces of content. This means that for a 60-second video, there will be 60 audio scores and 60 video scores. These Pv and Pa modules are specific to each type of codec.

The Pv and Pa modules operate in up to four different modes, depending on how detailed the input is from the parameter extraction. For the least complex mode, the main inputs are related to resolution, bitrate and frame rate, while the most complex mode performs advanced analysis of the video payload.

The short-term scores from those modules are fed into the long-term quality predictor (Pq) module, together with any stalling information, and the final session quality score for the total video session is then estimated. The Pq module also produces a number of diagnostic outputs, so that the underlying causes of the score can be analyzed. The Pq module is not mode- or codec-dependent and is therefore common for all cases.

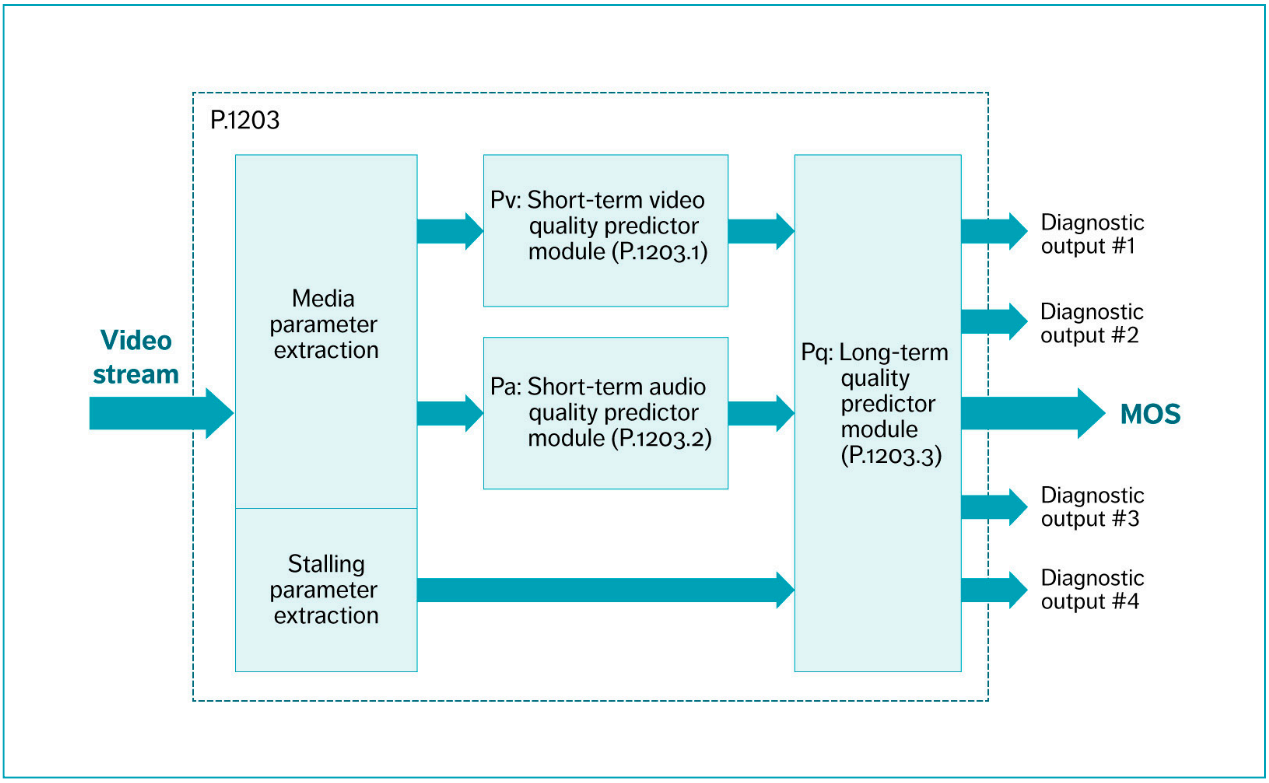

P.1203
Video stream
Media parameter extraction
Stalling parameter extraction
Pv: Short-term video quality predictor module (P.1203.1)
Pa: Short-term audio quality predictor module (P.1203.2)
Pq: Long-term quality predictor module (P.1203.3)
Diagnostic output #1
Diagnostic output #2
MOS
Diagnostic output #3
Diagnostic output #4

*Figure 2* ITU-T P.1203 architecture

### Training

In the development and standardization of P.1203, a large number of subjective test databases were created, each containing videos that were graded by at least 24 test individuals. An important goal set in the development of P.1203 was to handle the long-term perceived effects of stalling events and quality adaptations.

Thus, in contrast to traditional subjective tests in which a few 10-second videos are typically repeated, the videos used for P.1203 were all unique and between one and five minutes long. It was important to avoid repetition and continuously present new test videos that the viewers found interesting enough to pay attention to throughout the playout. The tests were done on mobile devices as well as on computer monitors and TV screens, to cover all the different use cases.

### Validation

Since the development of the P.1203 standardization was run as a competition, the performance of the different proponent models required validation, and the validation could not be carried out on any existing databases. Instead, after the submission of all of the proponents' candidate models to ITU-T, all the proponents worked together to design a new set of databases, with each step distributed to different proponents so that none had complete control over any individual database.

The new set of validation databases were then used to evaluate the models, and the top-performing ones were selected to form the P.1203 standard. In cases where the respective performance of several proponent modules was so close that a statistical test could not tell them apart, the modules were merged to form a single standard without alternative implementations.

### Final model

While the Pv and Pa modules were developed using traditional analytical methods and implemented as a series of mathematical functions, the Pq module is more advanced. This module is divided into two separate estimation algorithms: one using a traditional functional approach and the other based on machine-learning concepts.

The functional variant models human perception, as influenced by the effect of quality oscillations, deep quality dips, repeated quality or stalling artifacts and the ability to memorize. All of these effects are described by mathematical functions (as they are for the Pv and Pa modules), which are then combined to estimate the user's total perception of quality.

Machine learning is a method that solves a problem with the support of self-learning computer algorithms. It is well suited for problems where the relationship between the input and the output is complex, as in the Pq module. During the design and training phase, the algorithms automatically identify how various characteristics of the input data (Pv/Pa scores and stalling parameters) are reflected in changes to the output data (the test panel MOS values). The algorithm then automatically builds a black-box algorithm, which implements the final machine-trained solution and estimates the user score.

The final Pq MOS estimate is a weighted average of the output from the traditional functional algorithm and the machine-learning-based one. One of the advantages of using two different Pq algorithms is that they have statistically independent estimation errors, and when the two scores are averaged, the actual error becomes smaller.

### Future standardization work

The video module currently supports H.264/AVC video codecs up to HD (1920x1080). A new work item has been started in ITU-T, which will result in a recommendation that also supports H.265/HEVC and VP9 video codecs up to UHD resolution (3840x2160). This work item is running in a similar fashion to P.1203, with a competition giving participating companies the opportunity to submit their own proposed models.

“THE USEFULNESS OF HAVING A STANDARDIZED CLIENT FEEDBACK MECHANISM HAS BEEN RECOGNIZED BY 3GPP”

### Implementing a quality model

Successful implementation of a quality model is dependent on access to the input data required by the model itself. The most demanding models, such as full-reference variants, are usually implemented close to the video streaming client, inside the device, so that the complete received video can be compared with the one sent. This is typically done for manual testing scenarios, where a special test phone is used in which a quality model has been implemented.

This method is not feasible for passive measurements, where all or a large part of live video traffic is monitored, so model input data needs to be collected in another way. For example, an MNO that wants to gain a better understanding of its overall perceived video streaming quality would need to collect data from all streaming sessions. One way of doing this would be to intercept the traffic at certain network nodes and use the traffic content and pattern to try to infer which service is being used and the quality level being delivered. This can be difficult, however, owing to the fact that many services are now encrypted, which significantly limits access to the data required to do a detailed quality estimation.

One way to overcome this challenge is with the help of the streaming client, which has full knowledge of what is happening during the video session. A feedback link from the streaming client can be used to report selected metrics to the network (or the original streaming server) where the quality can be estimated. This technique is already used internally for many streaming services. For example, when a user clicks on a video link on the internet, the client typically continuously measures different metrics and sends them to the server. Unfortunately for MNOs, though, these feedback channels are usually encrypted and available only to the MSP. Even if an MSP were to make the metrics available to the MNO, they would still be proprietary, and it would be difficult for the MNO to compare them due to the fact that the content and level of detail typically differ between MSPs.

### 3GPP QoE reporting

The usefulness of having a standardized client feedback mechanism has been recognized by 3GPP, and its technical specification TS 26.247 describes how this can be implemented for a DASH streaming client [5]. The 3GPP concept is called QoE reporting, as the metrics collected and reported are related specifically to the quality of the session. Sensitive or integrity-related data such as the user’s position and the content viewed cannot be reported.

The basic concept is that the streaming client can receive a QoE configuration that specifies the metrics to be collected, how often collection will occur, when reporting will be done and which entity to report to. There are three ways to send a QoE configuration to the client to facilitate different deployment cases:

#### 1. Media presentation description

In this case, the client downloads a media presentation description (MPD) when streaming starts. The MPD specifies how the media is structured and how the client can access and download the media chunks. The MPD can also contain a QoE configuration that makes it possible to get feedback from the client. Since the MPD is usually controlled by the content or service provider, the QoE reports from the client are typically configured to go back to their servers and are not always visible to the MNO.

#### 2. Open Mobile Alliance Device Management

Open Mobile Alliance Device Management (OMA-DM) has defined methods for how an MNO can configure certain aspects of connected devices such as APNs, SMS servers and so on. These methods also include an optional QoE

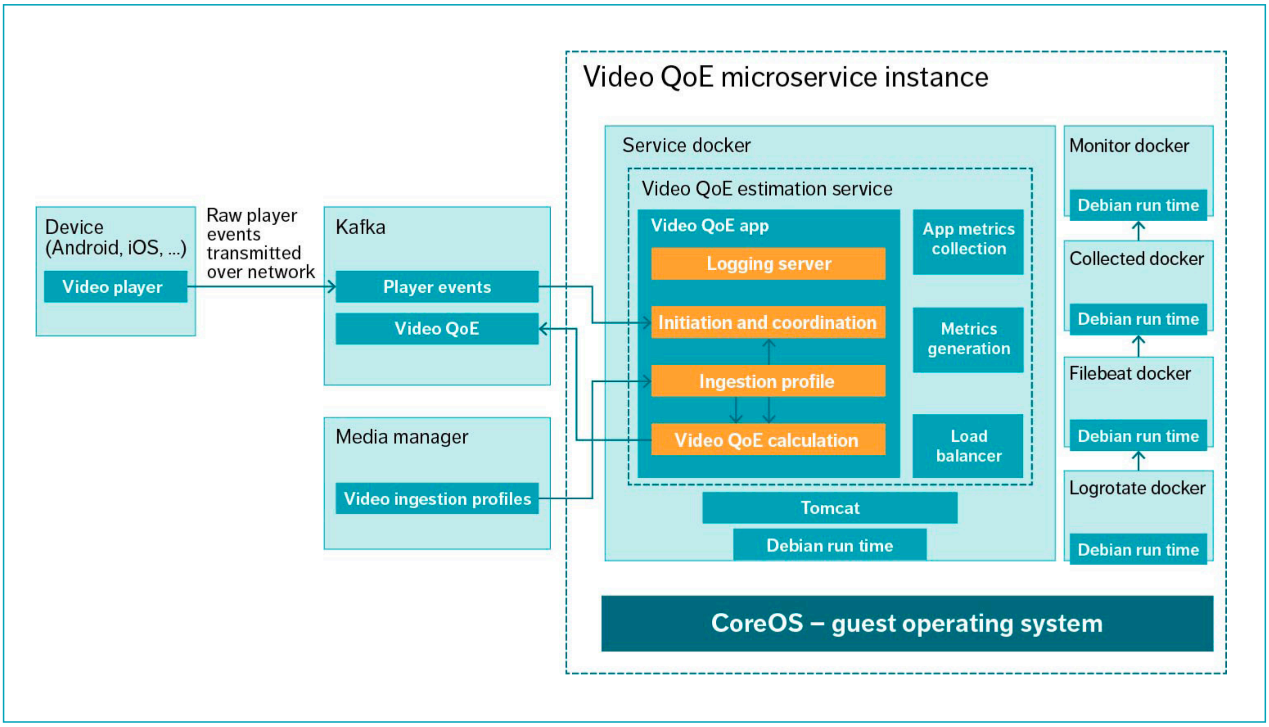


*Figure 3* Example of a video QoE microservice

configuration that activates QoE reporting from the client. However, not all MNOs deploy OMA-DM in their networks.

### 3. Radio Resource Control

The Radio Resource Control (RRC) protocol [6] is used between the eNB and the mobile device to control the communication in the RAN. The possibility of including a QoE configuration was added in 3GPP rel-14, giving MNOs the ability to use RRC to activate QoE reporting. As a result, the QoE configuration can be handled like many other types of RAN-related configurations and measurements.

The QoE metrics reported by the client in each of these three methods are well aligned with the input requirements in P.1203. This means that, in principle, any of them would enable an MNO to gain a good overview of the video streaming quality experienced by its users. Before this can happen, though, the new standards will need to be deployed both in the network and in the clients.

## Deployment of video QoE estimation

The delivery of over-the-top video today most commonly uses a cloud-based microservices platform with inbuilt video-quality estimation features. The P.1203 algorithm for estimating video quality is most suitable for implementation as a microservice that estimates video quality (in terms of MOS) for all video streams and for all individual video streaming sessions. If the estimated video quality distribution shows that there is a quality issue, a root cause analysis can be carried out and the necessary measures can be taken to improve quality.

> THE INPUT PARAMETERS TO THE P.1203 ALGORITHM ARE TAKEN DIRECTLY FROM THE VIDEO CLIENT

When running a media service, and providing the video client as part of the service, the input parameters to the P.1203 algorithm are taken directly from the video client that has all the details about the playout of the stream and reported over the network to an analytics backend. *Figure 3* outlines an example of such an implementation, in which player events are reported to a stream processing system (Kafka) where the video QoE is calculated in a video QoE microservice. The output from the video QoE calculation is then posted back into the stream processing system to be used for monitoring, visualization, root cause analysis and other purposes.

For an MNO, the standard architecture is to install a probe inside the network – in the core network, for example. The probe monitors video traffic using shallow or deep packet inspection and gathers information that can be used as input to a QoE estimation algorithm (P.1203). If a video service is not encrypted, the relevant metrics and events from the video streams can often be measured or estimated. The task becomes more challenging if the video streams are encrypted, but quality estimations of these video streams can still be done to some extent by using a combination of probes and standardized and proprietary algorithms and models. However, the ability to report quality-related metrics direct from the video clients enables much more accurate estimation of quality.

## Conclusion

MNOs and MSPs stand to gain a great deal from developing a better understanding of how users experience video quality, and the ITU-T P.1203 and 3GPP TS 26.247 standards provides the framework that is necessary to help them do so. Implementation of these standards by MNOs, MSPs and device manufacturers will enable the efficient and accurate estimation of video QoE required to meet continuously rising user expectations. The standard's smart handling of the effects of stalling events and quality adaptations makes it well suited to overcome the challenges presented by VOD and by adaptive video streaming in particular.

## THE AUTHORS

### Gunnar Heikkilä

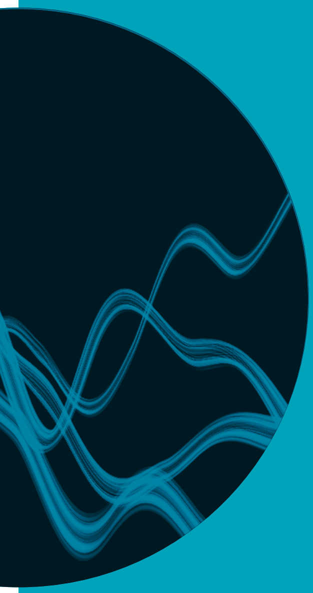
◆ is a senior specialist in machine intelligence at Ericsson Research. Since 1996, he has focused on user experience quality assessment and measurements, including standardization in 3GPP, ETSI and ITU-T. He joined Ericsson in 1987 and has previously worked with control system software for military defense radar systems, and with software design for synchronous digital hierarchy optical fiber transmission systems. Heikkilä holds an M.Sc. in computer science from Luleå University of Technology, Sweden.

### Jörgen Gustafsson

◆ is a research manager at Ericsson Research, heading a research team in the areas of machine learning and QoE. The research is applied to a number of areas, such as media, operations support systems/business support systems, the Internet of Things and more. He joined Ericsson in 1993. He is co-rapporteur of Question 14 in ITU-T Study Group 12, where leading and global standards on parametric models and tools for multimedia quality assessment are being developed, including the latest standards on quality assessment of adaptive streaming. He holds an M.Sc. in computer science from Linköping University, Sweden.

## References:

1. **Ericsson ConsumerLab report, TV and Media 2016, available at:** *https://www.ericsson.com/networked-society/trends-and-insights/consumerlab/consumer-insights/reports/tv-and-media-2016*
2. **Ericsson Mobility Report, November 2016, available at:** *https://www.ericsson.com/en/mobility-report*
3. **ITU-T P.910, April 2008, Subjective video quality assessment methods for multimedia applications, available at:** *http://www.itu.int/itu-t/recommendations/rec.aspx?rec=9317*
4. **ITU-T P.1203, November 2016, Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport, available at:** *http://www.itu.int/itu-t/recommendations/rec.aspx?rec=13158*
5. **3GPP TS 26.247, January 2015, Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH), available at:** *http://www.3gpp.org/DynaReport/26247.htm*
6. **3GPP TS 25.331, January 2016, Radio Resource Control (RRC); Protocol specification, available at:** *http://www.3gpp.org/DynaReport/25331.htm*



ISSN 0014-0171
284 23-3300 | Uen

SE-164 83 Stockholm, Sweden
Phone: +46 10 719 0000