

User Measures of Quality of Experience: Why Being Objective and Quantitative Is Important

Peter Brooks, Teolys, France
Bjørn Hestnes, Telenor Research & Innovation, Norway

Abstract

This article develops the concept of quality of user experience by focusing on requirements for its measurement and communication for use in industry. Whereas more typical work involves either modeling media quality based on technical measures or user subjective assessment of media quality, it is argued that QoE measurement should include objective measures of the process and outcomes of usage. Also, subjective measures should be collected as quantitative data rather than qualitative data as is currently common, because quantitative data enables combined analyses of objective and subjective measures for deriving global QoE ratings. Global ratings are important for optimizing the communicability of results to target audiences. Examples of measures and scales are provided. A structured approach to defining and measuring QoE in relation to QoS is explained and illustrated.

Currently, network service evaluation according to the International Telecommunication Union (ITU) Recommendations is distinguished by objective modeling of measurable technical system performance and subjective testing with people.

The general advantages and limitations of objective modeling and subjective testing are summarized below prior to introducing a complementary approach based on combining objective measures of user performance with quantitative subjective measures. This approach is developed in the framework of the European Telecommunications Standards Institute (ETSI) standardization [1]. Particular attention is given to the industry stakeholders for quality of (user) experience (QoE) data and their requirements for easy-to-use guidance on QoE in order to make good decisions about the users (and potential users) of products and services. It is argued that a quantitative global QoE rating from multiple measures

of user performance enables deriving alternative QoE scales for presenting alternative, more meaningful conclusions about QoE to different stakeholders. Also, a structured approach to defining and measuring QoE in relation to quality of service (QoS) is explained and illustrated as a basis for deriving QoE guidelines for network operators, equipment manufacturers, and service providers.

Advantages and Disadvantages of Current Approaches

It is useful to distinguish and consider the validity of three existing approaches for measuring network service quality from a user perspective:

- Testing user-perceived QoS
- Surveying subjective QoE
- Modeling media quality

The first two approaches collect subjective data from users. The third approach is based on objective technical measurements.

Testing User-Perceived QoS

Measures of user-perceived QoS are subjective because they obtain the opinion of test participants when asked to rate the quality of a medium or detect a change in quality. This approach is often also referred to as subjective quality evaluation. Although there are various methods standardized [2, 3], they share in common a focus on user perception, a self-report rating measurement technique, and a mean opinion score (MOS) as the main measurement.

There are some important advantages of the user-perceived

The work on quality of user experience presented in this article is performed at ETSI. ETSI produces globally applicable standards for information and communications technologies (ICT), including fixed, mobile, radio, converged, broadcast, and Internet technologies, and is officially recognized by the European Commission as a European Standards Organization. ETSI is a not-for-profit organization whose 720+ ETSI member companies and organizations benefit from direct participation and are drawn from 60 countries worldwide. ETSI Specialist Task Forces (STFs) are teams of highly skilled experts working together over a predefined period to draft an ETSI standard under the technical guidance of an ETSI Technical Body and with the support of the ETSI Secretariat. The task of the STFs is to accelerate the standardization process in areas of strategic importance and in response to urgent market needs.

QoS approach. By using a direct measure of user behavior (user perception), it is very user-centered. Because the user perception data is collected in laboratory experiments with standardized methods, there is a high level of control, and this body of work can benefit from the replication, exchange, and comparison of data. As user perception is recorded for known levels of QoS, such as transmission delay or packet loss, a direct comparison may be made between systems of different technical performance. Furthermore, MOS scores may be analyzed using statistical techniques to develop predictive models of media quality.

There are, however, two critical disadvantages of the approach:

- First, the end-user focus is entirely in the perceptual domain. This means that the measures concern human ability to detect or recognize different media. Typically, work in this area does not measure user behaviors and interaction that result from usage after perception. Also, important QoS changes may be small enough to be unconsciously perceived but still affect behavior; for example, even when people do not notice 170 ms audio-video asynchrony in a video, their ratings of the speaker are negatively affected [4]. Indeed, the added value of new network services is largely expected to result from the creation of interactions that are life-like and human-like, in which case many of the factors involved in natural human interaction with other people and the physical world represent automatic behaviors of which we are not conscious. For example, we are not usually aware of our human reliance on the normal frequency and duration of head nods and speakers' lip-shape information when communicating: artificial modification of the frequency and duration of head nods can disrupt abilities to communicate; and without lip-shape information, listeners cannot tolerate as much noise interference because humans automatically blend lip-shape information with heard speech in order to understand what is said. Therefore, user performance and experience may not be the same as user perception.

- The second disadvantage of the user-perceived QoS approach is related to the first: because many of the user behaviors enabled by network services are associated with unconscious psychological factors, there should not be a reliance on the test participants' self-reporting of opinion. Just as in other areas of psychological research, accepted good practice requires identifying and measuring objective behaviors, with subjective measures of user opinion used mainly as a supplement.

Surveying Subjective QoE

Whereas QoS concerns measures of technical characteristics that affect service performance [5], QoE should concern measures from users. The concept of subjective QoE extends beyond user-perceived media quality to include measures such as usability and user satisfaction [6, 7]. However, this work views QoE as entirely subjective. For example, current definitions of QoE are:

- "...the term to describe user perceptions of the performance of a service" [8, p. xiv]. "QoE is how a user perceives the usability of a service when in use — how satisfied he or she is with a service..." [8, p. 3]
- "The overall acceptability of an application or service, as perceived subjectively by the end user" [9, p. 5].

With this conceptual focus on subjective user perception, the methodological focus has been to survey user opinion via questionnaires and rating scales. However, as presented earlier, the study of user behavior should not rely on user opinion. Typically, the only attention to objective measures concerns the technology and not the users, such as measures that can be

automatically collected by QoS monitoring tools (e.g., call blocking and terminal streaming buffer status) [8]. In these cases QoE is inferred from indirect technical measures, and user tests are necessary to identify and validate the relationship between the technical parameters and actual user behavior.

Modeling Media Quality

Computational models of media quality are objective because they model measurable technical parameters [10]. They share in common the aim to assess the combined effects of variations in several technical parameters, inclusion of technical parameters that are believed to affect quality, and some scalar rating of quality as the main output.

There are some important advantages of the computational model approach. It is user-centered because it selects and models parameters on their basis to influence quality from a user perspective. Indeed, modeling user satisfaction is often an objective. The models are standardized, as with user-perceived QoS rating procedures, so this body of work can benefit from confidence in the comparison of results. Once a valid and reliable model has been developed and learned, it may be quicker and cheaper to apply than performing tests with people.

Although current standardized models have been validated against user test results, this validation should be done when any new parameter is added or aspect of the model is updated to reflect changes in the technology or usage context. Therefore, the advantages of speed and economy to apply compared to tests with people are reduced somewhat because of the requirement to continually validate against user test data. Also, any modeling validated only against user perception data (as is currently the case) will be limited for the same reasons presented earlier.

Objective Measures of QoE

The concepts of user experience and quality of experience were originally promoted by human-computer interaction researchers to emphasize concern with the *outcomes* of people's experience with — or through — technology, for example [11]. The key point is that measures of user experience or QoE should concern user performance based on actual usage.

There will be different user performance measures for different network services because of the different reasons and situations for use. To give some examples, a distinction shall be made between services offered for people to interact with content (e.g., IPTV and video on demand) and services offered for person-to-person interaction (e.g., VoIP and videoconferencing). The user performance measures listed in Table 1 are objective because they are not dependent on the test participants' opinions. Rather, they are direct measures of either the process or outcome of user behavior. For each measure an example is given for both a person-content interaction situation and a person-to-person interaction situation.

The objective data illustrated in Table 1 might be collected in different ways for laboratory and field tests, and for different services and usage situations. Taking *accuracy of task* as an example, for Mobile TV this could be the correct identification of a football player, whereas for use of a machine-based avatar by hearing-impaired people who lip-read, this could be the correct understanding of spoken information.

Subjective Measures of QoE

The main point so far has been that objective measures of QoE can and should be collected from user tests and that these measures enable us to extend beyond user perception

Measure	Examples	
	Person-content interaction	Person-to-person interaction
Accuracy of user task completion	Correct recognition of visual information	Extent to which a shared problem is correctly solved
Gain or loss to user	Game score	Purchase price negotiated by a salesperson
Time taken for goal completion	Game duration	Communication task time
Number of user inputs	Mouse clicks	Words spoken
Time between user inputs	Time between mouse clicks	Pauses in conversation
Number of turns taken to communicate	Conversational turn-taking with an avatar	Conversational turn-taking with another person
Number of interruptions	Conversational interruptions with an avatar	Conversational interruptions with another person
Number of simultaneous actions	User input coinciding with machine output	Simultaneous speech during conversation

Table 1. *Example objective measures of user performance.*

Measure	Examples	
	Person-content interaction	Person-to-person interaction
Effectiveness	Rating of how accurately visual information was received	Rating of the extent to which the communication goal was achieved
Efficiency	Rating of the effort required to watch mobile TV	Rating of the effort required to communicate by video
Satisfaction	Rating of overall satisfaction with the content	Rating of overall satisfaction with the communication
Enjoyment	Rating of engagement with a website	Rating of fun with a real-time game
Social presence	Rating of whether an avatar seems real	Rating of impression of personal contact with other persons
Impression of communication partner	Rating of the intelligence of an avatar	Rating of the trustworthiness of other persons

Table 2. *Example subjective measures of user performance.*

to user experience. However, understanding user opinion remains important, and a combination of objective and subjective variables should better reflect the complexity of QoE. For example, although many user communication behaviors are unconscious and should be measured objectively (e.g., turn-taking during interaction), other behaviors are conscious (e.g., visual effort) or subjectively experienced (e.g., satisfaction), and may be measured by directly asking the user. As with the objective measures shown in Table 1, subjective variables can also be identified for measuring the process and outcomes of usage. Table 2 lists some subjective measures. As with the objective measures, examples are provided for both person-content interaction and person-to-person interaction, and these measures would be implemented in different ways for different applications, services, and usage situations.

Table 2 also illustrates that measuring subjective user experience of new and future network services extends beyond traditional measures of user perception, user satisfaction, and usability to incorporate, for example, feelings of social presence, and the formation of impressions about our machine and human communication partners.

Deriving Quantitative and Global QoE Measures

It is possible to identify specific users of QoE data who are concerned with improving the quality of network services. Here we call these people stakeholders, both to avoid possible confusion with the service users and because it emphasizes that these people have an interest in understanding user experience in order to succeed with their work. Stakeholders have diverse roles (Table 3). They include persons from more technical to more financial work areas, and from more design-oriented to management functions. In order to communicate QoE data in an effective way to diverse and busy people, meaningful summary statements of QoE are required that may differ from the original measures used in a user test [12]. Therefore, although a range of objective and subjective measurements should be used to measure QoE (e.g., there are 13 variables if we combine the examples in Tables 1 and 2), there is a requirement for techniques that combine these multiple variables into more global QoE measures.

Statistical techniques can be used to derive a single global

Technically oriented	Customer oriented	Management
Product strategist Strategic service developer System integrator Project leader Service level agreement (SLA) negotiators Technical researcher Interaction designer Development engineer Audio/video codec engineer	Marketing Service portfolio specialist Salesperson Sales support Customer support Project leader Human factors researcher Service host	Chief enterprise officer Chief technical officer Chief marketing officer Chief of strategy Chief operation officer Chief research officer Chief development officer Sales director

Table 3. Stakeholders for QoE data have diverse roles and backgrounds.

QoE level from multiple QoE variables. For the arithmetic involved, user subjective responses should be measured in such a way that the ratio between any two intervals can be calculated. Although there are exceptions, ordinal scales are typically employed in work on user-perceived QoS and subjective QoE, and therefore do not allow ratio calculations. Ordinal scales provide ordered scale positions against which test participants provide their opinion. Such a scale therefore has rank-order relations, but conclusions about the distance between the scale positions cannot be made. For example, the MOS quality scale (Fig. 1a) is an ordinal qualitative scale and not a quantitative scale. Even if the five scale positions of *Bad*, *Poor*, *Fair*, *Good*, and *Excellent* are numbered from 1 to 5 [2] or from 0 to 100 [3], the numbers may not accurately reflect the distance between the scale positions, so it may be invalid to calculate statistics such as the mean, standard deviation, and correlation coefficient because these statistics require interval or ratio data. It would also be invalid to use useful techniques such as factor analysis for data grouping and multiple regression for predictive modeling.

A subjective rating scale with interval or ratio properties can be achieved by using a label-free scale or a label at each end of the scale (Fig. 1b). The actual choice of scale may depend on many aspects of a particular user test, and also deliberate attempts to reduce possible errors and biases from user test participants, for example [13].

The objective measures of user performance listed in Table 1 provide quantitative data. With subjective variables also measured quantitatively on an interval or ratio scale, it is possible to statistically summarize, compare, and combine the total set of QoE measures. Examples are:

- Calculating an average score per measure for different services
- Computing a global score as a weighted linear combination of the original variables
- Testing for a significant difference in QoE between QoS levels

Measuring both subjective and objective user experience variables in a quantitative way also ensures that a large range of techniques may be applied afterward for deriving different global QoE ratings. This is important for enabling the selection of optimal ways of communicating with particular stakeholders. For example, it may be appropriate to report to certain stakeholders the average QoE for two different services (calculation of means), while other stakeholders may wish to know the amount to which the QoE of a service varies across particular types of users (calculation of standard deviations). Case studies with stakeholders [14] indicate the value of being able to summarize QoE qualitatively and with alternative numerical scales that may be easier to understand and apply. Examples are:

- QoE measured on a scale ranging from -5 to $+5$
- Summary statements such as:
–Task effectiveness with service X is better than with service Y.

- QoE is *good* but not *very good*.
- A service is *unusable*.

Even if a global QoE rating is expressed more qualitatively in order for it to be more meaningfully communicated, it can be derived from a statistical analysis of quantitative data that benefits from more objectivity. Also, if the original data is quantitative, many options for recategorizing and labeling are kept available. Figure 2 illustrates that alternative scales offering different types of meaning for stakeholders can be derived from an originally quantitative QoE rating. There are many possibilities. The valid derivation of categorical scales from a quantitative QoE scale is a topic for further research.

Expressing QoE in Terms of QoS

Although user experience embodies psychological measures, it is important that it is expressed in relation to the networks and equipment that influence user behavior and result in a certain level of QoE. Therefore, QoE data should succeed where possible in combining both user experience and technical measures; for example, to provide an expression of the user experience when using a particular service with known levels of

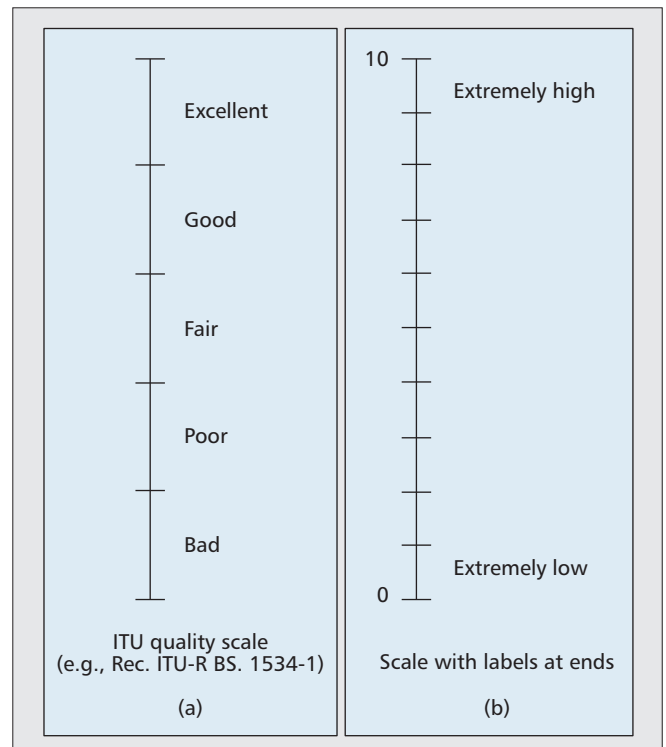


Figure 1. Example scales for measuring subjective response: a) qualitative scale; b) quantitative scale.

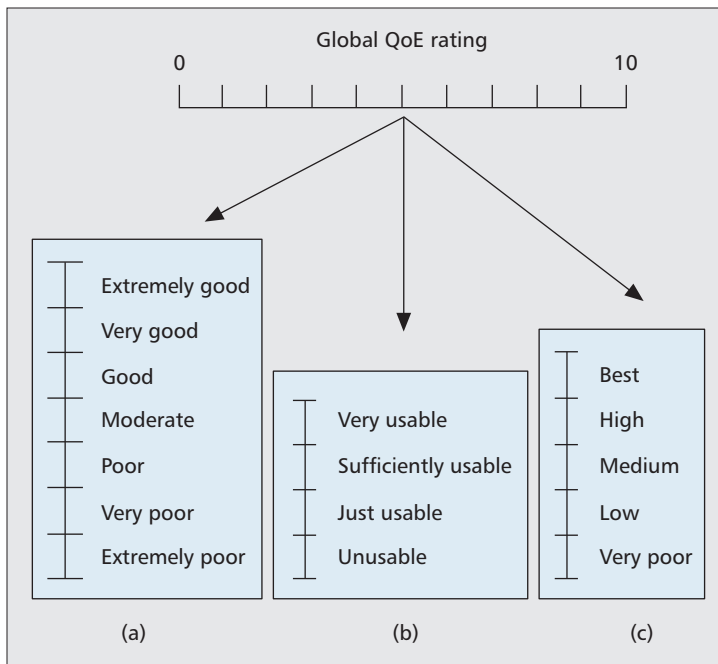


Figure 2. Example alternative qualitative scales that may be derived from data collected on a quantitative scale.

QoS. The amount of variables concerned therefore becomes very large, and some structure is important. A structured approach to QoE can be achieved by applying the clause:

```
IF <communication situation>;
  USING <service prescription>;
  WITH <technical parameters>;
  THEN <user experience>.
```

These attributes can have many parameters; some of the main ones are listed with examples in Table 4. As can be seen, the attribute <technical parameters> includes QoS parameters, and the attribute <user experience> includes whichever objective and subjective measures from Tables 1 and 2 were appropriate for a particular user test.

If we apply this clause when designing and reporting a user test, it helps keep explicit the attributes of QoE and break QoE down into its key set of parameters. Furthermore, it is possible to use this structure to build a database of user test results from which QoE guidelines for stakeholders may be derived. Such guidelines for real-time communication services have recently been made available by ETSI [1].

Bringing This Together in a Definition of QoE

As indicated earlier, although there are existing definitions of QoE, they are restricted to collecting subjective measures from users. Therefore, here is a definition of QoE that takes into account the arguments of the previous sections.

QoE is a measure of user performance based on objective and subjective psychological measures of using a service or product.

NOTE 1: It takes into account technical parameters (e.g., QoS) and usage context variables (e.g., communication task), and measures the process and outcomes of usage (e.g., user effectiveness, efficiency, satisfaction, and enjoyment).

NOTE 2: The appropriate psychological measures will be dependent on the communication context. Objective psychological measures do not rely on the opinion of the user (e.g., task completion time measured in seconds, task accuracy measured

in number of errors). Subjective psychological measures are based on the opinion of the user (e.g., perceived quality of medium, satisfaction with a service).

EXAMPLE: A network operator may conclude that a service with a certain level of QoS used for a particular communication situation offers users excellent QoE, while with a different level of QoS it provides poor QoE.

Here the notion of user performance is central to QoE because QoE is based on actual usage, and a networked service has the potential to help or hinder the process and outcomes of human behavior that will determine user satisfaction.

Summary and Conclusions

The current article aims to develop the concept of QoE and its measurement. Particular attention is given to developing measures of QoE that:

- Have higher validity because they include objective measures of user performance
- Are communicable to the intended target audiences of QoE data within the network services industry

Subjective measures should be collected as quantitative data that are consistent with numerical objective data because this enables statistical descriptions and

combinations of the multiple subjective and objective variables. It should be possible to derive global QoE measures that enable industry professionals to compare services, products, and QoS levels. A structured approach to QoE can be based on four essential attributes (the communication situation, service prescription, technical parameters, and user experience) that combine user and technical aspects.

From the current literature there appears to be a widespread misunderstanding that objective measures can only be collected from technology and that data from users is necessarily subjective. However, as in other areas of psychological research, it is possible to devise a test methodology with objective measures of user behavior.

When subjective data is collected from users, there currently appears to be a tendency to rely on qualitative data using ordinal rating scales. By using methods for collecting quantitative data (e.g., ratio rating scales), we can increase the range and validity of the statistical techniques that can be applied and, very important, more usefully combine subjective and objective data for global QoE measures.

One reason improving QoE for network services is a challenge is because there are a large number of human and technical variables that should be understood. Just in terms of the human element, there are complex interactions of conscious and unconscious psychological factors. To measure this complexity there are roles for many methods, and the traditional user perception approaches involving subjective assessment and technical parameter modeling are considered complementary to the QoE approach outlined here. For example, the *user perception* and *user experience* approaches may be aligned with two complementary phases of users' self-evaluation of a network service:

- Potential users can be expected to assess quality when deciding to buy or first use a service.
- Actual users can be expected to continually assess quality and possibly decide to stop using a service or change to a competitor.

Rating the acceptability of media quality may be part of the process of first choosing to use a service or purchase a product (e.g., trying a videoconferencing service, selecting between different televisions) and is consistent with the user perception approaches. In comparison, the QoE approach described here

Attribute	Parameter	Examples
Communication situation	User task	Give instruction, negotiate an outcome
	User group	Business people, elderly people
	User environment	Conference room, in a parked car
Service prescription	Service type	Video call, audio call, video on demand, IPTV
	Terminal type	Laptop computer, mobile handset
Technical parameters	Bit rate	1 Mb/s, 64 kb/s
	Media protocol	H.264, MPEG2, AAC
	Network protocol	TCP-IP, UDP-IP, RTSP
	Delay	50 ms, 500 ms, 1 s
	Audio-video asynchrony	0, 50 ms, -100 ms
	Jitter	50 ms, 100 ms, 1 s
	Packet loss	0.5%, 1%, 5%
	Video frame rate	7 frames/s, 25 frames/s, 30 frames/s
	Video resolution	CIF, 1920 × 1080, XGA
User experience	Task effectiveness	Task accuracy, value of negotiated agreement
	Task efficiency	Task time, number of speech interruptions
	User satisfaction	Acceptability of the service, satisfaction with communication
	User enjoyment	Level of engagement, level of fun

Table 4. *Attributes of QoE and their main parameters.*

should help determine the user aspects of network services that could be improved for enhancing the longer-term user experience, thereby reducing product rejections and churn.

Acknowledgment

This work is co-financed by the EC/EFTA in response to the European Commission's ICT Standardization Work Programme. The author's would like to thank the ETSI members and other companies and individuals who have contributed to the work.

References

- [1] ETSI STF 354, "Guidelines and Tutorials for Improving the User Experience of Real-time Communication Services," http://portal.etsi.org/STFs/STF_HomePages/STF354/STF354.asp
- [2] ITU-T Rec. P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," Geneva, Switzerland, 1996.
- [3] ITU-R BS.1534-1, "Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems," Geneva, Switzerland, 2003.
- [4] B. Reeves and C. Nass, "The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places," Cambridge Univ. Press, 1996.
- [5] ITU-T Rec. E.800, "Definitions of Terms Related to Quality of Service," Geneva, Switzerland, 2008.
- [6] Nokia, "Quality of Experience (QoE) of Mobile Services: Can It Be Measured and Improved?," White Paper, Finland, 2004.
- [7] R. Jain, "Quality of Experience," *IEEE Multimedia*, vol. 11, no. 1, 2004, pp. 96–95.
- [8] D. Soldani, M. Li, and R. Cuny, *QoS and QoE Management in UMTS Cellular Systems*, Wiley, 2006.

- [9] ITU-T Rec. P.10/G.100, "Amendment 2: New Definitions for Inclusion in Recommendation ITU-T P.10/G.100," Geneva, Switzerland, 2008.
- [10] ITU-T Rec. P.862, "Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs," Geneva, Switzerland, 2001.
- [11] L. Alben, "Quality of Experience," *ACM Interactions*, vol. 3, 1996, pp. 11–15.
- [12] P. Brooks *et al.*, "Communicating Quality of Experience Data for the Development of Multimedia Services," *Proc. 20th Int'l. Symp. Human Factors Telecommun.*, Sophia Antipolis, France, Mar. 21–23, 2006.
- [13] S. Zielinski, F. Rumsey, and S. Bech, "On Some Biases Encountered in Modern Audio Quality Listening Tests — A Review," *J. Audio Eng. Soc.*, vol. 56, no. 6, 2008, pp. 427–51.
- [14] ETSI TR 102 643, "Human Factors; Quality of Experience (QoE) Requirements for Real-time Communication Services," Sophia Antipolis, France, 2009.

Biographies

PETER BROOKS (Peter.Brooks@teolys.com) is a Chartered Psychologist and Associate Fellow of the British Psychological Society. Since his Ph.D. in applied psychology he has over 20 years of experience that spans the main research environments for human factors in telecommunications: in industry (ITT, Great Britain, and Alcatel, Germany), in an independent research institute (SINTEF, Norway), and as a university lecturer (Cranfield, Great Britain). Since 2003 he has focused on independent research and consultancy based in France.

BJØRN HESTNES is a senior research scientist at Telenor Research & Innovation in Norway. He has over 25 years of experience with Telenor and has been a project leader for more than 20 user-centered projects. He has special interests in video and mobile communications, with work that spans electronic classrooms, remote inspection, videotelephony for blind people, and heterogeneous and 3G video calls. He has a Master's degree in system development from the University of Oslo.