Report from Dagstuhl Seminar 15022

Quality of Experience: From Assessment to Application

Edited by

Katrien De Moor[1], Markus Fiedler[2], Peter Reichl[3], and Martín Varela[4]

1 **NTNU – Trondheim, NO,** katrien.demoor@item.ntnu.no
2 **Blekinge Institute of Technology – Karlskrona, SE,** markus.fiedler@bth.se
3 **Universität Wien – Wien, AT,** peter.reichl@univie.ac.at
4 **VTT Technical Research Centre of Finland – Oulu, FI,** martin.varela@vtt.fi

Abstract

This report provides an overview of the program, discussions and outcomes of Dagstuhl Seminar 15022 "Quality of Experience: From Assessment to Application", which took place from 4–7 January 2015 at Schloss Dagstuhl – Leibniz Center for Informatics. The seminar and the challenges that were addressed have their roots in the earlier Dagstuhl Seminars 09192 "From Quality of Service to Quality of Experience" and 12181 "Quality of Experience: From User Perception to Instrumental Metrics". The main goal of the seminar was to strengthen and go beyond the current understanding on Quality of Experience (QoE) and its assessment, in order to start tackling the logical yet highly challenging next steps: to move from assessment to application and to translate insights on QoE and knowledge from this research field into forms of economic and/or societal value. This report contains the personal statements and main challenges brought forward by the participants, who were on the fly clustered into six main discussion topics. We here report on the discussions and outcomes from the group work, organized around these bottom-up generated topics: "QoE theory and modeling", "QoE methodologies", "User factors and QoE", "QoE management", "Monetization of QoE" and "QoE in new domains".

Seminar January 4–7, 2015 – http://www.dagstuhl.de/15022
1998 ACM Subject Classification H.5.1 Multimedia Information Systems, H.5.2 User Interfaces
Keywords and phrases Quality of Experience, Usability, User experience, Content, Network monitoring, Quality measurement, Service pricing, Network management, Application management
Digital Object Identifier 10.4230/DagRep.5.1.57

1 Executive Summary

Katrien De Moor
Markus Fiedler
Peter Reichl
Martín Varela

Within the past few years, Quality of Experience (QoE) has gone through an explosive growth and established itself as an independent, multidisciplinary field of research, both in academic and industrial communities. Significant advances have been made with respect to the conceptual understanding of QoE as well as in terms of methodology and instrumentation, and the earlier Dagstuhl seminars 09192 "From Quality of Service to Quality of Experience"

Quality of Experience: From Assessment to Application, *Dagstuhl Reports*, Vol. 5, Issue 1, pp. 57–95
Editors: Katrien De Moor, Markus Fiedler, Peter Reichl, and Martín Varela
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

and 12181 "QoE: From User Perception to Instrumental Metrics" have played a catalyzing role in this process: for example, by putting key challenges on the agenda, by stimulating (collaborative) activities that address them and by contributing to the establishment of a multi-disciplinary community around the topic, involving a range of actors with sometimes very different, yet complementary perspectives, priorities and motivations in relation to QoE. The main goal of this seminar was to strengthen and go beyond the current understanding on Quality of Experience (QoE) and its assessment, in order to address a logical yet highly challenging next step, namely to move from assessment to application and to translate insights in QoE and knowledge from this research field into forms of economic and/or societal "value". The main underlying motivation is that – even though the conceptual grounds and methodological implications of QoE are a very interesting and worthy research topic as such – they also represent milestones on the road to reach another ultimate goal: translating the theoretical and empirical understanding of QoE, its assessment and measures, into "value". This value can be rather explicit and concrete (e.g., increased revenue, or reduction of number of customer complaints), but it can also be intangible and more latent (e.g., customer loyalty, strengthened relation between a customer and a provider, enabling user empowerment, contributing to well-being).

The seminar brought together 27 participants to work towards this challenging goal. They were representing 13 different countries and 17 different institutions, resulting in a variety of different backgrounds and specific expertise domains. The seminar took place over 2.5 days and was organized in such a way that time for group discussion and interaction was maximized, while the time for individual presentations was kept to a minimum. At the beginning of the seminar, every participant was invited to write down three challenges fitting within this overall scope of the seminar. Thereupon, a concise presentation round was organized. Every participant was asked to make a short statement (5 minutes/1 slide) related to her or his challenges. These personal statements are included in the form of short abstracts in this report.

The main challenges and questions put forward by the participants were clustered on the fly into six broader topics, around which the seminar group work was organized, namely: "Theory and modeling", "QoE methodologies", "User factors and QoE", "QoE management", "Monetization of QoE" and "QoE in new domains". The group work was divided into two parts, with three topics being discussed in parallel in both parts of the group work. The initial assignment of participants to the six groups was deliberately organized randomly instead of thematically. The intention was to mix up participants with different backgrounds and interests as much as possible in order to stimulate open discussions and flow of thoughts. Participants had the possibility to switch to another group by exchanging with another participant in case they had a strong preference for another group. Every participant was involved in two discussion groups.

In between part 1 and 2 of the group work, a plenary reporting session was organized. During this plenary session, each group briefly presented the main points discussed and potential joint activities. During the final plenary reporting and closing session, the main points and outcomes from the second part of the group work were presented. Extensive summaries of the discussions and main outcomes for each of the six working groups are presented in Section 4 of this report. Due to the time constraints, there was unfortunately not enough time for deep follow-up discussions in the plenary sessions. The seminar as such was also very briefly evaluated in the final plenary gathering. One important factor which would have further improved the participants' QoE and which was mentioned several times, is more time for "digestion" and "reflection" between the sessions (which was indeed limited,

given the duration of the seminar). Overall however, and supported by the participants' feedback during and after the seminar, we can look back on a successful and productive seminar during which plans for several future and follow-up activities were made.

2 Table of Contents

3 Overview of Talks

3.1 Individual experience? Adaptive QoE for monetization

Jan-Niklas Antons (TU Berlin – Berlin, DE)

QoE has been a rapidly growing topic in the last few years. One big future topic of research is how to show the value (in terms of monetization) to for instance industry or customers. Customer behaviour – in terms of media usage – is important as one key performance indicator when it comes to monetization. The known models of QoE for quality estimation and quality driven technology development are mainly focused on larger customer groups, but these average values must not always represent the majority of customer perception or behaviour. Therefore adaptive QoE estimation strategies could be a solution. As additional estimates and moderator variables, user groups could be identified and the influence of the individual user's state could be utilized.

3.2 QoE from active network management

Alemnew Asrese (Aalto University – Espoo, FI)

The existing frameworks Quality of Experience models for Internet applications such as web applications are mainly based on subjective assessment. This subjective assessment needs much resources and time. On the other hand, today's Internet applications are becoming more complex, there are many in number and type, and they have different requirements for QoE. One of the challenges is how to model and predict the perceived QoE from the low level (active) network measurement. Thus we need to have a framework that could correlate the subjective QoE with the one which is predicted from active network measurement results.

3.3 QoE and the re-assessment of the assessment

Katrien De Moor (NTNU – Trondheim, NO)

The new definition of QoE [1, 2] represents an important step forward for the field: QoE is equaled to a dynamic affective state (a *degree* of *delight* or *annoyance*), and its transient and relative character is fully underlined. Explicit reference is made to the extent to which a person's expectations and needs are fulfilled (with respect to the utility and/or enjoyment) through the experiencing of an application, service or system, relative to the user's context, personality and current state (e.g., mood). It is furthermore acknowledged that a range of complex and strongly interrelated factors may influence QoE and a classification of potential influence factors has been presented. However, this more holistic and humanistic theoretical

perspective on QoE has implications for the way in which QoE is and should be evaluated, measured and predicted, and can be linked to new challenges:

1. **Put the new definition of QoE into practice.** The traditional, standardized measures used in quality assessment and subjective testing are insufficient and need to be extended with robust and validated alternative measures of QoE (beyond Mean Opinion Scores, MOS) that allow to grasp QoE in terms of human affect (including delight and annoyance) and that enable to capture also hedonic (and not only utilitarian) features of QoE. Only in a limited fraction of the literature, this challenge is addressed and as a result, there is a barrier to advancing the state of the art regarding fundamental relations that are highly of interest in research on QoE, for instance between (indicators of) technical performance, perceived technical quality, delight and annoyance, enjoyment, user engagement. The lack of such shared standard measures furthermore limits the comparability of studies.
2. **Increase the ecological validity of QoE research.** Most QoE experiments take place in a controlled and artificial laboratory setting, meaning that the human subject and the experience as are completely taken out of their natural, real world environment and that ground truth data are collected in research settings characterized by a very low ecological validity. There is therefore a need to study and understand QoE also in real world contexts and over longer periods of time in order to better current understanding the relevance, impact and weight of individual or combined influence factors.
3. **Investigation of QoE in less traditional application domains targeting higher level goals (e.g., continuous care, learning, . . .) instead of monetary gains.** One important question is what should/could be the outcome of striving for high QoE. From a business perspective, QoE is linked to a potential for increased revenues, enhanced customer loyalty and satisfaction, and considered as a strategy for reducing the risk of churn. However, striving for QoE could also be relevant in non-profit contexts and sectors with goals that go beyond purely economic drivers, such as enabling positive, valuable experiences as a goal in itself, provision of continuous care (e.g., in a hospital setting), learning, . . . Investigating QoE in such less traditional application domains requires an adaptation of the tools and measures that are currently used.

References

1 Patrick Le Callet, Sebastian Möller, and Andrew Perkis. Qualinet white paper on definitions of quality of experience, version 1.1, june 3, 2012. *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, 2012.
2 Alexander Raake and Sebastian Egger. Quality and quality of experience. In Sebastian Möller and Alexander Raake, editors, *Quality of Experience*, T-Labs Series in Telecommunication Services, pages 11–33. Springer International Publishing, 2014.

3.4 Challenges

Philip Eardley (British Telecom R&D – Ipswich, GB)

Within BT we measure the performance of our broadband network by using active measurements from probes at the home gateway in the homes of a few thousand volunteers. This is proving very useful to our BT Retail operational division, and I see several challenges to extending its capabilities. Firstly, how to help the human expert who studies the measurement

results, for example by automatically identifying likely issues in the network that should be investigated more closely. Secondly, how to measure aspects that the current tests struggle with, such as the performance of home networks and of end-to-end services like gaming; I suspect this needs passive or hybrid tests. Thirdly, how to ensure the security and privacy of such a monitoring capability, especially as it becomes more widespread and pervasive.

3.5 The human, the technology and the business: bridging QoE, user experience, technology experience and customer experience

Sebastian Egger (AIT – Wien, AT)

Since the advent of QoE (and the first Dagstuhl seminars on that topic) several things have changed. QoE is nowadays a well accepted research strand in telecommunications and is properly addressed by a number of research institutions. Despite this valuable achievements, a large share of nowadays technological and telecommunication systems is still not fully considering user requirements with respect to quality as well as utility. Therefore, my main aim for a future research agenda on the topic of QoE is to subdue technology from a human perspective. This aims to continually assess existing as well as future applications for their subordination with respect to aforementioned utility and quality requirements. As existing QoE models and assessment methodologies only tackle a rather limited (and telecommunication centric) subset of applications, the first challenge is to extend the types of applications that are currently investigated in the QoE research context to get a broader range of human technology interaction considered. The second major challenge will be to bridge the current gap between QoE, user experience (UX), technology experience (TX) and customer experience (CX). This will include identification of appropriate (experience) time spans for each of these four experience categories as well as the understanding of interrelations and differences between these concepts. Addressing these two challenges will lead to a more holistic picture of experience of human subjects with technological systems as well as the contribution of different dimensions (QoE, CX, TX) to the overall customer experience of a service as a whole.

3.6 Teletraffic-inspired QoE models: the basis for more effective SLA

Markus Fiedler (Blekinge Institute of Technology – Karlskrona, SE)

Main reference M. Fiedler, J. Shaikh, V. J. D. Elepe, "Exponential On-Off Traffic Models for Quality of Experience and Quality of Service Assessment," PIK-Praxis der Informationsverarbeitung und Kommunikation, 37(4):297–304, 2014.
URL http://dx.doi.org/10.1515/pik-2014-0031

From a teletraffic perspective, Quality of Experience (QoE) modeling has many yet unexploited facets. Some traditional QoE modeling efforts and models suffer from loose couplings between cause (e.g. bit error rates in a wireless channel) and consequence (e.g. delayed delivery due to intermediate repair actions instead of pure image quality reduction). Teletraffic models that capture and keep track of impacts on QoE all through the communication system

and stacks help to create stronger links between observable Quality of Service (QoS) issues and QoE estimations. In particular, the model parameters tell stories about sensitivities and thresholds regarding observable QoS parameters.

Such teletraffic-inspired QoE models address a first challenge on how to pave the way from QoE assessment to application, namely to devise effective QoE-based Service Level Agreements (SLA). Further challenges are to take QoE concepts to new domains, such as the Internet of Things (IoT), Machine-to-Machine (M2M) and Business-to-Business (B2B) communications, and to express (good or lousy) QoE in monetary values.

3.7 Joint QoE assessment for software validation

Farnaz Fotrousi (Blekinge Institute of Technology – Karlskrona, SE)

QoE in the context of telecommunication networks is affected by the quality of the software applications as a proxy for the user experience of those networks.That makes a joint collaboration between telecommunication and software engineering areas. Validations of software features that integrate network features could be linked to the user acceptance or churn. Correlation between quality (software and network) and its impact on the users (QoE) contributes for conceiving, evolving and maintaining software systems. For this purpose, one of the challenges is about mapping the definition of quality in software engineering and telecommunication domains where a different categorization for quality definition has been established. The second challenge would be capturing a joint measurement where some quality attributes are dominated in software engineering and/or telecommunication. Defining the quality threshold for an acceptable software would be another challenge depends on the software functionality and relevant network elements.

3.8 Challenges for quality-driven content delivery

Pantelis Frangoudis (INRIA Rennes – Bretagne Atlantique, FR)

The availability of QoE models for various application domains makes it possible to utilize them for quality-driven service delivery. These models may require information from multiple layers, which should be provided by different stakeholders (user, content provider, ISP) in a timely manner, to drive application behaviour adaptation. We focus on an over-the-top (OTT) content delivery environment, where data are disseminated over the infrastructure of ISPs, which is outside the content provider's (CP) control. In this environment, it is critical to address issues of ISP-CP cooperation, since network awareness is important for QoE-driven delivery, while at the same time decisions by the OTT provider can impact the operation of the ISP. Significant challenges thus emerge. From an operational perspective, the level of ISP-CP cooperation defines the granularity of network-level information (e.g., congestion in network segments in the user-data center path) that the ISP is willing to share with the CP, but also affects the process of placing and controlling probes in the ISP network, so that accurate QoE estimation can be achieved. At the business level, an open issue is whether it is feasible and if there are incentives for ISPs and OTT providers to offer QoE-driven SLAs.

3.9 Towards smart determination of appropriate quality

Samuel Fricker (Blekinge Institute of Technology – Karlskrona, SE)

Joint work of Fricker, Samuel; Farnaz, Fotrousi; Fiedler, Markus
Main reference F. Fotrousi, S. Fricker, M. Fiedler, "Quality Requirements Elicitation based on Inquiry of Quality-Impact Relationships," in Proc. of the 22nd IEEE Int'l Requirements Engineering Conference (RE'14), pp. 303–312, IEEE, 2014.
URL http://dx.doi.org/10.1109/RE.2014.6912272

The telecommunication and software engineering disciplines approach each other increasingly to integrate knowledge and concepts across disciplines. This partnership creates opportunities to address hard problems in surprisingly new ways. An important challenge is the determination of appropriate quality for software products and services. Too little quality leads to churn where users look for alternatives. Too much quality leads to poor return of investment because the development and operation of systems becomes unnecessarily expensive. The software engineering body of knowledge did not contain appropriate methodology to address this problem. It was just recently that a first method was proposed.

The QoE community would approach this challenge by studying the relationships between software quality and the impact generated by such quality experimentally. Unfortunately, such experiments are costly and hard to implement for software products and services with a non-trivial set of features. Too many experiments would need to be performed and the requirements on the expertise of the average software practitioner would be excessive.

We are looking for ways to alleviate the problem by sensible selection of QoE experiments to be performed and smart timing of requests for user feedback. Beyond doing the obvious, the automated execution of QoE tests, we aim at answering the following questions: (1) How can we generalize results from QoE tests across users, features, and contexts? (2) When is QoE data valuable enough to be collected?

The answers to these two questions will allow us to determine good-enough quality in a more lightweight manner than we do it today. While automation will ease the adoption of the techniques by practitioners, the reduced need for experimentation will reduce cost and annoyance that are generated with them. As a result, a much broader range of software products can benefit from QoE methodology and delight rather than disturb.

3.10 QoE for education services at primary and middle school

Juan Pablo González Rivero (Plan Ceibal – Montevideo, UY)

The success of one to one educational models relies heavily on the learning services and communications infrastructure that support them. The appropriation of new educational tools supported by technology requires users to be motivated and attracted to use these new services. When these models are carried at a national level, as in the case of Plan Ceibal in Uruguay, user heterogeneity, vast differences in experience with technology, cultural, socioeconomic and geographic variety make it particularly difficult to understand the different assessments about the quality of services experienced in the educational process. Another particular aspect of the case is that the dynamic is not the usual in which users pay for the service they want, in this case there is no payment and services are imposed. There is a

need for having quality indicators from the user perspective with the objective of acting proactively on services to ensure success in the appropriation of tools and in the improvement of learning. To cope with this, the techniques and tools of QoE are a promising way. This in turn brings big challenges:

1. **To develop methodologies and models of QoE for educational services.** The main services provided in education can be separated into two classes, videoconferencing based services and web adaptive learning platforms. Currently there are no models oriented to educational applications. The differences between both types of services and the particularities of each one enlightens the need for development of methodologies and specific models to monitor the QoE. In turn, the models to allow proper management should enable the root analysis and require adequate incorporation of the factors associated with the users. Regarding methodologies, classic laboratory tests would be inadequate. There is a need to develop ecological and indirect methods of assessing the QoE taking into account the particular characteristics of children and adolescents.
2. **Development of new QoE estimation tools.** For QoE-based services management, there is a need to develop and operate tools to estimate QoE in each of the relevant services. In the case of Plan Ceibal it is possible to develop tools in a distributed manner, deployed on network equipment, user terminals and centralized platforms. The challenges would be in the definition of the architecture and management of tools.
3. **Measure the relationship between QoE and its impact on learning.** QoE measurement as an input for deployment and optimization services is a key but not the ultimate goal. We sense that there is an intimate relationship between QoE services and its impact on learning, (ie. The relationship between the quality of audio in an English videoconferencing class and intonation learned by students). In order to understand the educational value related to the QoE the link between these two concepts needs to be taking into account.

3.11 QoE in digital ecosystems

Poul Einar Heegaard (NTNU – Trondheim, NO)

Most services are now provided over what is sometime referred to as a *digital ecosystem*. This is a heterogeneous system with multiple stakeholders/actors with one or more roles, over multiple domains, including content and network providers, and the end users. In digital ecosystems the various technical systems can be tightly coupled, while the control and management in the different domains exchange the management information only through what is given in a Service Level Agreements (SLA).

In this context it is important to understand what can be observed and controlled to enhance the QoE and the underlying QoS. This can be studied both from the user and provider perspective. The user perspective typically focuses on how the users perceive the service, but also more insight in what the user her/himself can do to change and improve her/his QoE is important. The (content/network) providers are more and more concerned with what they can do with respect to technical presentation and delivery, and on customer support and relations, to enhance the QoE from their users' perspective. For instance, the providers are interested in a "loyalty curve" over time and to understand how and why the

loyalty changes, and how the loyalty develops over time. Of particular interest is to get more insight in what the *long term effects* on QoE are, including use of both subjective test and objective measurement in real life environments.

A large number of QoE factors have been identified and the effects of these been (objectively and subjectively) studied and measured. The technical factors and impairments are mostly related to the performance of the services, but it is important also to increase the understanding of the effect on QoE of other technical, non-performance, such as *trustworthiness* factors, which includes dependability/reliability, (information) security, and safety.

The relative weight between the various QoE factors needs to be studied more. One approach is to use a modeling framework where a weighted convolution of the empirical distribution of the opinion scores of different factors, which gives an overall *distribution of the opinion scores*. The sensitivity of changes in the weights, and potentially in their cross-correlation can then be studied analytically. In general, it is very useful for the providers to know more about the the uncertainty of the opinion scores, measured for instance as standard deviation, quantiles in the distribution, or the probability of observations below an acceptance threshold.

It is a great challenge to define a *functional relation between QoE and QoS*. From studies of QoS, it is know that this is really hard to do even for QoS factors at different technical (protocol) levels, and across different domains. What is the best approach to define measurements and management strategies from the providers' perspective to enhance QoE? Furthermore, how can we operationalize the new QoE definition? For example, if we assume that the technical factors influence the annoyance level only, and that the annoyance is low and the delight will dominate. How can we then determine what level of annoyance will make the QoS-related factors, measured by technical attributes and MOS-score, to become a significant, dominating effect on the QoE (and not "just one of many factors")?

3.12 QoE++: From ego- to eco-system?

Tobias Hoßfeld (Universität Duisburg-Essen – Essen, DE)

QoE research has advanced significantly in recent years with a focus on the QoE ego-system. This means that QoE has been mainly addressed within a single session on a short-time scale for a single user of one concrete application. Thereby, different facets have been addressed by the research community, like subjective user studies to identify QoE influence factors for particular applications like video streaming, QoE models to capture the effects of those influence factors on concrete applications, QoE monitoring approaches at the end user site but also within the network to assess QoE during service consumption and to provide means for QoE management for improved QoE. Recent research has focused for example on HTTP adaptive video streaming which monitors QoE at the application layer at the end user site and adapts the video to the current network conditions. As major QoE influence factors stalling events, initial delays, and video quality switches have been identified and basic QoE models have been derived which build the basis for quality adaptation mechanisms. This enables service providers to improve resource utilization and QoE by incorporating information from different layers in order to deliver and adapt a video in its best possible quality.

However, in order to progress in the area of QoE, new research directions have to be taken. There is a need for QoE++. The application of QoE in practice needs to consider the entire QoE eco-system and the stakeholders along the service delivery chain to the end user. In comparison to the traditional QoE ego-system thinking, the QoE eco-system addresses among others the following research topics: in-session vs. global system perspective, short- vs. long-time scales when considering QoE, single vs. multi-user QoE, single vs. concurrent usage of applications and services, user vs. business perspective by addressing all key stakeholder goals. From the user's perspective, current QoE models mainly quantify the influence of various parameters on the perceptual quality. However from a service provider's perspective, it may be more relevant how the user is behaving, as a consequence of the experienced QoE, but also as a consequence of other context factors like pricing, privacy, etc. Thus, QoE++ requires (*a*) to extend current QoE models by the different perspectives of the QoE eco-system including the service provider perspective, (*b*) to incorporate user behavior as part of the model, and (*c*) to identify and include relevant internal and external context factors including physical, cultural, social, economic context.

QoE++ faces the following three major challenges. (1) Can we utilize QoE for network & service management? Or is it more appropriate to consider user engagement or user behavior? Which context factors are relevant or are such context-factors even more important for network & service management, e.g. in order to foresee and react on flash crowds? (2) Can we transform QoE into business models, SLAs, etc.? Or is it possible to 'trade' QoE? For example, offering WiFi sharing at home, a user may get improved service delivery and QoE by its ISP. (3) Do we understand fundamental models and natural relationships of QoE++? How can we extend existing QoE models to take into account the service provider's perspective? How can we include user behavior in the models? What is the relationship between QoE and user behavior?

Following QoE++ will shift from ego- to eco-systems and give answers to those questions.

3.13 Quality of Experience – objectives and challenges

Kalevi Kilkki (Aalto University – Espoo, FI)

QoE has become a popular topic during the last years, at least, in certain circles of communication networks and services. One of the main reasons for introducing QoE as a "scientific concept" was the limited usage of the concept of QoS. In practice, QoS was used to describe some technical attributes of communication services like packet loss ratio, delay and the available bit rate.

QoE is aimed to extend, compared to QoS, the analysis towards human and to economic aspects. This objective poses immediately a hard measurement challenge, because human beings are strange objects to measure. Our behavior is complex, often seemingly irrational, and always highly context-dependent. Thus an artificial (laboratory) arrangement is just one very specific context in which humans behave in certain specific way. Then if human behavior is observed in real, everyday situations, there can be huge variations which makes it extremely hard to make clear, statistically significant conclusions.

And finally, we are still lacking a general framework to assess in a systematic way what we prefer in our lives, not just what we prefer in a specific situation, for instance, when

watching to a video clip through Internet. Thus, the QoE framework shall locate in the area of human life instead of the area of technology (where QoS naturally to locate). Any QoE measurement result shall truly describe human experience.

3.14 QoE from a telecom operator's perspective

Eirini Liotou (National Kapodistrian University of Athens – Athens, GR)

In telecommunication networks, despite the catholic presence of Quality of Service (QoS) mechanisms, QoE has always been an "afterthought". This means, that although QoS provisioning has been an integral part of these networks, QoE has never been an original design intention. However, a QoS-based, system-centric view of the network is no longer sufficient, and it needs to be complemented or replaced with more user-centric approaches. This shift is currently an emerging, open challenge.

The acquiring of QoE awareness and the management of a network in a QoE-centric way may be beneficial not only to the end-users and telecom providers, but also to any other stakeholders involved in the service provisioning chain, such as service, content and cloud providers, or even customer care and support agents. Focusing on the telecom operator's perspective, we may identify three potential opportunities (and incentives, thereof) that derive from acquiring QoE and controlling a network in a QoE-centric way: (*a*) to increase the loyalty curve of the customers and, equivalently, to decrease customer churn, (*b*) to drive business operations and Customer Experience Management solutions, and (*c*) to cut costs by identifying and exploiting the non-linear relationships between QoS parameters and the perceived QoE.

Towards this direction, various research questions need to be addressed: (1) How can QoE be measured, monitored and controlled in telecommunication networks? Such a QoE management framework is essential before any operator-specific business decisions are made. (2) What kind of business opportunities are created for the operator and other stakeholders, such as OTT service providers, assuming that QoE can be managed? New QoE-based business models need to be designed, carefully considering Network Neutrality issues. (3) What is the new (more active) role of the end-user in such a QoE-aware/QoE-centric network (e.g., users may provide feedback about their preferences, priorities and experience)? Moreover, how can the end-user be convinced to "buy" QoE? Potential strategies of the network operator may include to "personalize" each end-user and provide QoE accordingly, or to build more aggregated user-profiles, facing the fact that the "average user" does not exist.

Among others, these questions should be considered as fundamental, before starting to design the QoE-centric networks of the future.

3.15 Motivations, consumer behaviour and QoE

Toni Mäki (VTT Technical Research Centre of Finland – Oulu, FI)

Quality of Experience is already well understood regarding the multimedia and telephone services. This understanding is established on good understanding of the human sensory

system, related mental processes and also well defined activities and so called motivational objects people are pursuing. Relation of these concepts is important when trying to understand the QoE of a given service. E.g., for multimedia services it is, while not easy, possible to see how motive (e.g. enjoyment) and mental processes handling sensory inputs are closely related. When moving towards non-multimedia services this coupling becomes looser. Gaining better understanding on this is a one challenge in modelling contemporary services. We are working towards a methodology that allows analysing these relations and how much of the QoE of a service can be accounted for technical quality in systematic manner (based on activity theory). My other interest and more widely a great challenge is quality-related consumer behaviour. How people experience services and how that translates into deliberate and also unintentional behaviour and decisions is a complex process.

3.16 What to do with Quality of Experience? A proposal for new research directions

Sebastian Möller (TU Berlin – Berlin, DE)

Whereas considerable progress has been made in defining, measuring and predicting Quality of Experience (QoE) for a number of commercial services, QoE still comes in the form of a self-directed object of measurement: QoE is measured in order to improve QoE. However, little attention has been drawn to what else can be done in case that QoE gets below or above a certain threshold. One could think of a mobile telephone service which switches to a different communication channel (such as SMS) when a call is lost during a train ride, and which automatically re-installs the call when the network connection is back for a sufficiently long period of time. Or, a video service which automatically starts recording and time-delayed playback in case that the IP-based television line is overloaded. Identifying potential use cases where QoE monitoring leads to new, proactive types of services first requires user studies in the field to identify hidden needs. Then, the technical feasibility needs to be analyzed, and finally the implemented services need to be evaluated as to whether they really provide additional value. In order to evaluate such pro-active services, it will be important to move beyond the classical borders of QoE measurement, by investigating utility and acceptance as a trade-off between user need fulfillment and QoE provision.

3.17 From Service Level Agreements (SLA) to Experience Level Agreements (ELA)

Peter Reichl (Universität Wien – Wien, AT)

Based on several years of advances especially concerning applications around media and Web, QoE research has now reached a state of maturity which allows (and at the same time demands) turning these insights into operational reality. In order to achieve this transition in a smooth way, it is proposed to learn from the rather successful concept of Service

Level Agreements (SLA) which today form a well-established broad base for a common understanding of quality provision on a networking level, i.e. with respect to Quality of Service (QoS) and the related characteristic parameters. Similarly, as far as Quality of Experience is concerned, it is now necessary to design and specify corresponding "Experience Level Agreements" (ELA) taking account of the specific user-centric perspective on service quality. This step still poses several key challenges, including the appropriate parametrization of quality features and user context factors as well as the applicability towards a broad spectrum of new fields approaching the horizons of QoE research, ranging from the Internet of Things up to arts and culture.

3.18 Acceptability towards a better User Experience

Miguel Ríos Quintero (TU Berlin – Berlin, DE)

In order to improve next generation audiovisual services, telecommunication companies traditionally focus on improving parameters based on Quality of Service (QoS). Unfortunately, QoS mostly focuses on performance and reliability parameters of the service. So far, this approach has proved to be insufficient to predict customer satisfaction.

The main concern is that every user has a different perception, based on several influencing factors. Factors such as emotional state and intrinsic user characteristics are often not considered or vaguely explored in traditional evaluation methods. Here, it is necessary to expand the view from a performance-driven approach towards actual Quality of Experience (QoE), addressing what a technical quality means to a subject in terms of accepting that quality for actual usage. In turn, an increased QoE may be beneficial in terms of how long a user spends with the service per usage, whether he actually purchases the service at all, and in terms of whether he stays with the service or drops it for the benefit of a concurrent service.

3.19 From momentary "within-session" QoE to a more global and ecologically valid picture

Werner Robitza (TU Berlin, DE)

In the domain of Quality of Experience we have reached a level where we are well equipped to give valid predictions of perceived momentary audiovisual quality or delight/annoyance, that is answering the question of "What would the quality rating be if we asked an average person now?" However, this only works for small time frames, from several seconds up to a few minutes, and only considers the typical user, since in traditional tests, all ratings are combined into the Mean Opinion Score (MOS). However, the remembered quality and experience for a user is not necessarily just the average (or a simple function) of the hypothetical momentary quality in terms of a Mean Opinion Score.

When we deal with the challenge to predict experienced quality for a longer time period, a specific region, or a group of people, how is the value of QoE meaningfully integrated over

time? From a similar perspective, we could also ask: How does it manifest for different kinds of users? This is a challenge to be solved in the near future.

To this aim, we also need ecologically valid test settings and models that go beyond short term stimuli of just a few seconds, content that was purposely created to have no influence on quality ratings by being boring, and "within-session" experiments, where users have no way to interact with a service. By putting the human into the role of a passive viewer, when in reality they would actively experience a service, we are invariably getting test results that may not be representative. Standardization efforts in this direction are crucial to enable a global understanding of what these concepts entail. For example, a collection of general guidelines for more ecologically valid tests would have huge benefits for the future evaluation of existing services and creation of new experiences altogether. In the long term, we expect QoE models to be based on such tests.

But what else could be improved when performing tests with users? Humans behave differently in the presence of good or bad QoE. They may adjust their behavior so as to improve their positive affect towards a service, or they may show signs of aggression or confusion. The inclusion of behavioral patterns in QoE studies will help improving services and delivering better QoE for the specific user in the long run.

3.20 YouSlow – why and where is YouTube slow?

Henning Schulzrinne (Columbia University – New York, US)

Joint work of Nam; Hyunwoo; Kim, Kyung-Hwa; Calin, Doru; Schulzrinne, Henning
Main reference H. Nam, K.-H. Kim, D. Calin, H. Schulzrinne, "YouSlow: A Performance Analysis Tool for Adaptive Bitrate Video Streaming," in Proc. of the 2014 ACM SIGCOMM Conf. (SIGCOMM'14), pp. 111-112, ACM, 2014.
URL http://dx.doi.org/10.1145/2619239.2631433

YouTube is one of the most popular media delivery platforms, and many users equate the QoE of YouTube with their Internet experience overall. The "buffering" circle of YouTube and similar applications has become iconic. However, it is often very difficult or impossible for users to determine why their experience is bad or highly variable – is it a local Wi-Fi problem? Maybe a microwave oven or baby monitor? Access network issues? NATs, proxies and firewalls? Network interconnection disputes? We built YouSlow, a follow-on effort to our DYSWIS (Do You See What I See), to map the performance of YouTube on a global scale, and see clear patterns in user-visible QoE artifacts, i.e., re-buffering.

3.21 Potential for integrated/cooperative QoE management schemes

Lea Skorin-Kapov (University of Zagreb – Zagreb, HR)

We are witnessing many different players involved in end-to-end service delivery, ranging from various cloud providers, content providers, network operators, etc. We are further witnessing what has been referred to as a paradigm shift towards an Internet of Services, envisioning everything on the Internet as a service. Such a transition will potentially lead

to new services being realized as service chains combining and integrating the functionality of (potentially many) other services offered by third parties (e.g., infrastructure providers, software providers, platform providers, etc.).

QoE models dictate the parameters to be monitored and measured, with the ultimate goal being effective QoE optimization strategies. Hence, the question is how can new QoE models be exploited (in a practical sense) in the context of QoE management schemes? The majority of QoE-based management approaches to-date may be primarily related to either network management (based on monitoring and control on access and core network level) or application management (adaptation of quality and performance on end-user and application level). For example, OTT services (e.g., YouTube, Netflix) delivered by third party service/content providers commonly implement QoE control schemes on the application level. As a result, different QoE optimization and control loops are involved (e.g. dynamic application adaptation, network management), with the question being how do they interact? A further question is to what extent can cooperative management schemes and underlying business models involving multiple players achieve more efficient management of network/system resources, while enhancing customer QoE? What are the stakeholder incentives? What (novel?) solutions are needed for coordination and information exchange among actors involved in the service delivery chain in order to provide channels for effective QoE control/improvement? Importantly, the aforementioned questions should also be considered in the context of regulatory policy and the ongoing network neutrality debate.

3.22 QoE assessment in new generation of multimedia services: the need for adaptations

Samira Tavakoli (Universidad Politécnica de Madrid – Madrid, ES)

Nowadays QoE is a valuable and well-known concept in the multimedia services. Because of the trade-off between the user's QoE and the cost for providing the high quality services, content providers find themselves in need of understanding the perceived quality of provided service to the user. Numerous researches in this regard have been conducted with an ultimate goal of optimizing the user's viewing experience. Considering the subjective assessment as one of the common approaches to evaluate the QoE, several test methodologies have also been provided as international recommendations. However, by immersion of new generation of multimedia services, the suitability (even validity) of these methods to accurately illustrate the perceptual quality of these services, is questionable. As addressed by previous studies, it is quite likely that the relative impact of impairment type changes with the setting of subjective experiment.

As an example of new multimedia services, the HTTP Adaptive Streaming (HAS) can be mentioned, which has gained widespread popularity as a cost-efficient way to distribute pre-encoded video content. With adaptive streaming, it is probable that the quality switching takes over periods up to several minutes, providing a novel type of impairment which is time-varying quality sequence.

Most common evaluation methodologies, like Absolute Category Rating (ACR) recommend the use of short video sequences to be evaluated by the test subjects. However, in the case of some techniques such as HAS, using short test sequences cannot be appropriate. It is

not clear whether the perceptual quality of individual stimuli (cf. only adaptation event) would be the same as evaluating the stimuli when occurring in a longer sequence. Other approaches such as Single Stimulus Continuous Quality Evaluation (SSCQE) could also not be suitable. This is because the effect of recency and hysteresis of the human behavioral responses while continuously evaluating the quality could lead to an unreliable evaluation through this methodology. On the other hand, in traditional testing methodologies, the quality of the video in audiovisual services is often evaluated separated and not in the presence of audio. Nevertheless, the requirement of jointly evaluating the audio and video within a subjective test has been addressed by some previous studies.

From another side, in regard to subjective experiment planning, having multiple technical variables in such services (in the case of adaptive streaming these parameters could be dimension, amplitude, frequency and period of adaptation), makes the full-matrix design not feasible. This implies that the experimental results are limited to few number of content types or test conditions, so that making a generic conclusion becomes impractical. Another question that can be raised is whether lab/crowdsourcing studies give a reliable knowledge about "actual" application of these techniques (e.g. evaluating the perceptual quality in live streaming event). To summarize, novelties of new multimedia services such as adaptive streaming, highlight the need to investigate new assessment approaches compatible to the nature of application under study to reliably picture the real user's perceived quality.

3.23 Decompiling QoE

Christos Tsiaras (Universität Zürich – Zürich, CH)

Joint work of Tsiaras, Christos; Stiller, Burkhard
Main reference C. Tsiaras, B. Stiller, "A Deterministic QoE Formalization of User Satisfaction Demands (DQX)," in Proc. of the 39th IEEE Conf. on Local Computer Networks (LCN'14), pp. 227–235, IEEE, 2014.
URL http://dx.doi.org/10.1109/LCN.2014.6925776

Measuring the impact of technical variables, such as latency, bandwidth, or resources priority-access, on Quality of Experience (QoE) of various services demands an extensive feedback from end-users, when those variables change. Estimating QoE in a given scenario becomes harder, when non-technical variables, such as price, need to be considered in addition to technical ones. In any case, detailed feedback that correlates all variables affecting QoE is needed by end-users for each service separately. A deterministic mathematical model (DQX) [1] encapsulating user demands, service characteristics, and variable specifications is proposed to formalize the QoE calculation, considering one or multiple and diverse variables. The output of QoE functions presented in DQX can be normalized such that results will be compatible with the five-point scale Mean Opinion Score (MOS), proposed by the ITU-T.

References

1 Christos Tsiaras, Burkhard Stiller, *A Deterministic QoE Formalization of User Satisfaction Demands (DQX)*. The 39th IEEE Conference on Local Computer Networks (LCN), Edmonton, Canada, 8–11 September, 2014.

3.24 Application of QoE to create business value in telecom

Astrid Undheim (Telenor Research and Development – Trondheim, NO)

Network quality is seen as one of the most important drivers for customer satisfaction in Telenor, and the user perceived network quality is measured using the Net Promoter Score (NPS) framework. The results are valuable for insight into capacity problems for specific base stations etc, however the NPS measurements are based on limited sms-feedback. There is clearly a need to introduce real-time, per-application QoE measurements based on available data from the network, handset and application.

At the same time, QoE measurements should be operationalized, and without explicit user feedback. One way to achieve this is to look into new QoE measures such as: how often a service is used, how long sessions, etc. These are implicit QoE data that can more easily be mapped to the status of the network, and that even allows for longitudinal studies.

Machine-learning methods and big data technologies are attractive to go one step further in operationalizing QoE, through building hypotheses on the influencing factors for QoE, and testing these in real operation. The resulting model will give insight into which factors are most important for customers in given situation, as well as help predicting the QoE in similar situations. These models and methods will provide actionable insight for a large set of customers, with increasing accuracy as the model learns from new customers, situations and service usage.

3.25 $elling QoE

Martín Varela (VTT Technical Research Centre of Finland – Oulu, FI)

QoE has been a trending research topic for a few years now, and it has become a buzzword in some business domains as well. However, most (almost all, really) of the work done on the field has been either for modeling/estimation/measurement or as a tool to guide technical improvements (e.g. QoE-driven management). I believe that the next big advance in making QoE operational will come from showing to the different stakeholders that QoE is worth ($$$) caring about. My question is now, how do we go about it?

3.26 QoE in next-generation networks and services

Min Xie (Telenor Research – Fornebu, NO)

As a research scientist at Telenor, an operator with a strategy prioritizing Customers, we aim to provide services that are perceived by users as satisfactory services. To do that, we need to understand how users experience and then operate to optimize user experience.

The first challenge is how to measure and analyze QoS and QoE in the mobile networks. Mobile operators can gather data from both the network side and the user side. We need measurements from the users perspective to reflect their perception of the delivered services. However, the user end device has limited functionality and resource whereas the number of variables needed to be measured could be huge. It is therefore our task to decide which metrics to measure and where to measure them, i.e., to distribute the measurement tasks properly between operators and users (also service providers). The collected measurements need to be analyzed efficiently, e.g., via machine learning, to identify the role and significance of different variables for different services. Identifying the key QoS metrics and their thresholds would help operators to adapt the network operations more effectively to serve users.

The analytical results serve as the base to solve the second challenge, predicting QoE without or with limited user feedback. In a mobile network, it is not realistic to perform subjective QoE assessment towards individual users to get the real-time QoE indicator. Instead, it is more practical to create a QoS-QoE map and estimate the QoE based on available QoS measurements. The challenge lies in the limitation of available measurements and the influence of contextual factors such as emotion, environment, and social status. An accurate QoE prediction would help operators to allocate resource more efficiently and react to potential QoE degradation proactively.

With a better understanding of user experience, we then aim to define delivered service quality from the perspective of Experience Level Agreement (ELA), as comparison to SLA. The ELA will describe the service quality in terms of "experience", which could be better understood by the users. On the other hand, it will keep the quantitative specifications of the service quality as references for operators. This ELA concept is expected to bridge the gap between operators and users so that they can communicate in a same language on the service quality.

3.27 On QoE monetization and forming a global customer satisfaction metric

Patrick Zwickl (Universität Wien – Wien, AT)

Main reference P. Reichl, P. Maillé, P. Zwickl, A. Sackl, "A fixed-point model for QoE-based charging," in Proc. of the 2013 ACM SIGCOMM Workshop on Future Human-centric Multimedia Networking (FhMN'13), pp. 33–38, ACM, 2013.

URL http://dx.doi.org/10.1145/2491172.2491176

The transition from theory to practical applicability of QoE is exacerbated by the following challenges: (1) monetization and utilization/exploitation strategies for QoE information, (2) formation of a global QoE picture (bridging the bounds of individual test ranges, model bounds, and current biases), and (3) concepts easing the handling of the QoE marketisation complexity.

The monetization or other kind of utilization of QoE information is notoriously difficult, as it involves the fixed-point problem between willingness-to-pay and service perception [1], but also immanently requires the integration of other factors, e.g., socio-economic classification of users and the service usage preference (in the current context or in general). This is further hampered by the non-linearity of QoE metrics, which is contrasted by the requirement for linear ISP and user utility presentations whenever feeding QoE results into economic optimization models.

A related major obstacle is the absence of a global QoE view, which overcomes isolated QoE results, which are bound to single test case, test scenario, methodology, test limitations or biases, and parameter settings. Today's QoE results are hardly comparable across service types and cannot be regarded to be a generically comparable measure, e.g., comparing the perception across two different kind of services. However, the mapping from QoE to any kind of utility representation will require an understanding of the service usage preferences across service types and quality levels by individual users for various kind of quality conditions, forming a key block for the overall valuation of services, i.e., utility. This will likewise also affect the ISP utility perspective in terms of prices that can be charged. For instance, when comparing fictive test results around input QoS test range interval $[A, B']$ where $B' > B$ ("$>$" means better) with a separate testing of $[B, B']$, then we can expect to receive inconsistent results. This is even more problematic for across service type comparisons, which mind the service usage preference of the user.

Complexity is at least a three-dimensional problem in operation, which involves technology, economics, but also users:

- "Simplicity" is an important factor from a technological point of view – reducing the complexity wherever possible, increasing the scalability whenever doable. For this reason, it is important to analyse the tradeoffs between profiting from QoE-aware service differentiation and rising complexities in parts of the system (e.g., client, access or core network) when doing so. In particular, strategic decisions like shifting complexity from the network to the clients (e.g., adaptive streaming) may strongly affect the technical complexity.
- On the business axis, we can differentiate in long term and short term effects: In the short run the generation of new revenue streams exceeding the costs of additional resources is in focus. In the long-run, the loyalty of customers may be influenced by adequate QoE levels. Both short and long term perspective have to be addressed in concert by appropriate QoE marketisation frameworks in order to obtain a sustainable QoE market configuration.
- On the user side, a high "usability" when offering QoE upgrades is crucial but difficult to obtain, as network quality is an experience good that can hardly be communicated in advance. In practice, experience simulators or past experiences of similar users may provide helpful indications on adequate quality and price combinations prior to the actual purchase and experience thereafter.

In general, a separate willingness-to-pay testing for network quality is more realistic than meaningful approximations, however, efforts should be dedicated to the better interrelation of individual test results and their projection to untested cases.

References

1 Peter Reichl, Patrick Maillé, Patrick Zwickl, and Andreas Sackl. A fixed-point model for QoE-based charging. In *Proceedings of the 2013 ACM SIGCOMM Workshop on Future Human-centric Multimedia Networking (FhMN)*, pages 33–38, 2013.

4 Working Groups

4.1 QoE theory and modeling

Sebastian Egger (AIT – Wien, AT)
Peter Reichl (Universität Wien – Wien, AT)

While the modeling of QoE has made significant advances over the last couple of years, currently existing models still lack an integration of user behavior aspects and user context factors along with the consideration of appropriate temporal scales. Therefore, during the discussions of this group, a comprehensive QoE and user behavior model has been developed, providing a framework which allows joining a multitude of existing modeling approaches under the perspectives of service provider benefit, user well-being and technical system performance. In addition, the role of a broad range of corresponding influence factors has been discussed, with a specific emphasis on user and context issues, and a series of related use cases have been identified which are suitable to validate the proposed model. During and after the discussions in this group, joint efforts resulted in a conceptual paper [1].

References

1 Peter Reichl, Sebastian Egger, Sebastian Möller, Kalevi Kilkki, Markus Fiedler, Tobias Hoßfeld, Christos Tsiaras and Alemnew Sheferaw Asrese. Towards a comprehensive framework for QoE and user behavior modelling. In *Proceedings of the 7th International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015.

4.2 QoE methodologies

Katrien De Moor (NTNU – Trondheim, NO)
Werner Robitza (TU Berlin – Berlin, DE)

The goal for this working group was to discuss challenges, problems and potential ways forward in the context of QoE methods and methodological approaches. The challenges that were identified during the initial presentation sessions were further clustered into 3 main and partly overlapping discussion topics: (1) the need to increase the ecological validity of QoE research inside and outside of the lab, (2) specific operationalization issues (measuring the "real" QoE), and (3) other issues regarding the "when" and how to measure QoE? For each of these topics, main issues and potential ways to address them in a better way were discussed.

4.2.1 Increasing ecological validity inside and outside the lab

Ecological validity refers to the extent to which a study design resembles and reflects the real-world situation and thus, the extent to which empirical findings can be generalized to real-world settings. In traditional (standardized) QoE studies, the ecological validity is very low and problematic in several ways:

- The results of such lab studies may be strongly biased (leading to "wrong" results and prediction models) because the stimuli and conditions are artificial and not necessarily representing the real-world situation.
- Test participants are very focused on quality degradations or specific properties of the system, as they are usually instructed and primed by the researcher to do so (e.g., in the briefing and instructions, in questionnaires). This is usually not the case in a real-life situation.
- It is very difficult to keep people engaged in a test which takes place in such an unnatural controlled lab situation. Yet, users' engagement is crucial in the context of QoE, so this poses a challenge which needs to be addressed.
- The current standards and recommendations are not adapted to ensure a higher ecological validity of research findings, yet they play a crucial role as they essentially define the terms of common understanding, methodologies and measures, which will be used by different stakeholders (from research, industry, ...) and for different purposes.

Without abandoning these current standards completely, there are certain aspects that could be changed relatively easily in order to work towards the goal of creating a more ecologically valid setting inside the lab. This is a relatively short-term goal which could be reached by e.g.,

- Introducing more immersive test paradigms: not repeating any content, because repetitions are unnatural, thus leading to boredom and increased attention to minuscule details or the test procedure itself. Immersive testing has already been shown to be practical and leading to less frustrated test participants [1].
- Having clear strategies to try to keep the attention high. For instance, give the participants a task that helps to create sufficient attention throughout the whole test (e.g., incentive per watched video, remembering details or replying to questions about the content afterwards and when correct answers are given, there is an additional reward, ...).
- Including longer duration and thus more representative test stimuli and using meaningful and real content (e.g., in audiovisual tests).
- Trying to make the evaluation process as unobtrusive as possible. This could be done by e.g., also capturing implicit feedback from users (e.g., through behavioral measures) and matching it with the captured explicit feedback (e.g., user ratings). A mismatch between both feedback sources could help to identify issues that threaten the validity (e.g., the observers were not paying attention, experimenter bias, ...).

The future vision (longer term) goes even further and includes:

- Development of new recommendations (or amendments to existing ones, such as ITU-T P.913) on how to investigate QoE in different real life environments (e.g., home, mobile, ...), with guidelines on appropriate methods to use, the combination of qualitative and quantitative approaches in an optimal way, tools that can be used for capturing behavior and monitoring technical parameters, etc.
- Evaluating QoE over longer time periods (cumulative or longitudinal QoE) instead of focusing on one particular moment in time (as in traditional test settings).
- Additional issues that have to be considered in this respect relate to e.g., privacy concerns and motivation (e.g., how to motivate people to participate in longer duration studies).

4.2.2 Operationalization issues

When it comes to the application of QoE, several issues were brought up:

- In the new, broadly-supported definition, QoE is defined as a degree of delight or annoyance, which is relative to a range of aspects (e.g., the user's expectations, current state, personality, ...) but the dominant measure (perceived overall quality) does not allow to capture delight or annoyance.
- The decision on which measures are most appropriate to include is depending on various issues, e.g., which modality/modalities? Which type of use case (e.g., mainly enjoyment vs. mainly utility-oriented?), Which target group (e.g., children, elderly people, ...).
- Depending on the modality there may be perceptual differences (e.g., age- or gender-based) but to gain deeper insights, we need to systematically gather and report more information about the participants and their characteristics: currently this is very limited.

What is needed?

- Complementary measures that operationalize the new QoE definition as well as guidelines on (1) how to use these measures in practice, and (2) on how to select the most relevant measures. Examples are e.g., behavioral measures (during experience, but also behavioral intentions after use), but also complementary self-report measures, such as Self Assessment Manikin (Valence, Arousal, Dominance), validated scales measuring abstract constructs such as user engagement, immersion, flow, expectations, mood, ...
- Guidelines on how to instruct and train test subjects (e.g., "don't overthink", which is really important when including e.g., measures of emotion, engagement, ...).
- In order to be able to really advance the understanding of how both stable and more dynamic human factors may influence or be related to QoE, we need to gather more information about the test subjects, so that we can describe the test panel better and look into different groups or segments (e.g., depending on affinity with technology, adopter profile, attitudes, personality traits, ...). There is also a strong need for guidelines on what to include, when and how, and how to use and report on this information later on. This topic was further discussed in the group discussion on user characteristics.
- Better guidelines on the number of participants that are needed: this strongly depends on the setup of the study and the goals (in some cases, a limited number of test subjects is enough, but in other cases it is not).
- Inclusion of qualitative questions in order to gain better insights into the "why" dimensions (and e.g., to check at the end of a test whether the "right" questions were included).

4.2.3 When and how should QoE be evaluated?

When should we measure QoE? This question comes up in several situations, both in a lab context and in real-life environments. When should a network operator do active or rather passive monitoring? When conducting a test, should we collect implicit or rather explicit feedback (or a combination of both)?
Specific questions were brought up:

- Should we measure QoE only when it is a complete "unknown"? Should we measure it only when a problem has occurred or rather pro-actively? Or rather in situations when a certain QoE-based decision has to be made (e.g. reduce resources while still guaranteeing an acceptance level of 80%)? How about in the context of next-generation services where tolerance levels are unknown?
- Is there an optimum point of asking during a subjective experiment or evaluation study?
- How to motivate and engage users to participate, without biasing them or unintendedly increasing certain acceptance thresholds?

- What are the most optimal incentives? By giving too many incentives, the risk is that people may be trying to simply "please" the experimenter. This was a problem e.g., in a study conducted in Uruguay (Plan Ceibal): because they received a free computer, the children involved in the study always rated the quality good despite intrinsically showing different behavior in the presence of bad network conditions.
- When including simple binary questions (e.g., "Is this quality level acceptable?"), there is a risk that internal quality threshold information gets lost with each further question.

We discussed several solutions and developed suggestions on how and when to measure:

- In general, it may be useful to consider first of all whether asking users for direct feedback could be avoided (e.g., by using other approaches, such as A/B testing). In some cases it may also be left up to the user, so that they can give feedback only when they want to (e.g., Skype).
- In the best case, the experience which is under investigation (e.g., game experience) should not be interrupted. In case explicit feedback is gathered, rather place it between two tasks.
- Another rule of thumb is to always try to lower the burden for the participant as much as possible, so for example in selection of measures to include: be selective and only include the constructs that you are most interested in (based on previous research, theory, ...).
- It may be useful (e.g., in a video quality test in the lab) to record rating and viewing behavior during a test and show it again to the test subject afterwards, while encouraging the subject to reflect on why he or she gave a certain score (did you see a change, how would you describe it, ...). This may provide valuable additional insights.
- Both in real-life and lab settings, it would be valuable to more explicitly inquire after personal priorities of users and customers of a service, as these priorities may differ a lot from user to user. However, gaining such richer information does not necessarily mean that only individual differences need to be considered. There may be patterns that can be studied for small groups of users and deeper insights in individual priorities could help to identify and to understand various user profiles/user groups, and to potentially derive different QoE models per profile or, at least, adjust the QoE provisioning procedure accordingly.
- When asking for explicit feedback related to QoE (e.g., in a real life setting), this should in an ideal case also have added value for the user or customer. However, this is not always easy to put into practice.
- It is also important to give users acknowledgments for their feedback, so that they know and believe that their feedback will really be considered and is useful.
- To increase the motivation and avoid biases, one strategy is to provide an intrinsic value instead of an extrinsic incentive, e.g., status (e.g., just mere participation), the value for the user is a functionality of the service itself (e.g., a free movie when evaluating a video on demand service).
- Involving users as early as possible may lower QoE issues with end product or service, and thus put a kind of "QoE by design" strategy into practice.

4.2.4 Conclusions

During the constructive and fruitful discussion during this group session, we identified a set of open issues and challenges related to the current methodological approaches and ways in which they could be adapted in order to increase the ecological validity of QoE research, put the new definition of QoE into practice and hereby enable new theory building

and better understanding of fundamental underlying relationships between QoE and its assumed influence factors. As follow-up work, bringing together practical guidelines and recommendations for more ecologically valid QoE research, and developing a decision tree for setting up experiments and selecting appropriate methods and measures, would be highly valuable.

References

1 Margaret H. Pinson, Andrew Catellier, and Marc Sullivan. A new method for immersive audiovisual subjective testing. In *Proceedings of the 8th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Jan. 2014.

4.3 User factors and QoE

Katrien De Moor (NTNU – Trondheim, NO)
Toni Mäki (VTT Technical Research Centre of Finland – Oulu, FI)
Werner Robitza (TU Berlin, Berlin, DE)

There are various reasons for considering user factors within QoE research. Regarding the economic and business values the following motivations were found important during our discussion:

- How to ensure customers are staying?
- How is the "threshold" for churning changing over time and impacted by rival services or adaptation over time?
- Incorporating user factors is important for marketing strategies, e.g., in customer segmentation.

From theoretical point of view following reasons to record and consider user factors were identified:

- User factors may help explain differences in ratings.
- It may be possible to create adaptive QoE models based on user factors (i.e., models targeted for different user groups).

Generally, internal reference models and internal criteria of human beings are influenced by previous experiences and differ among users. For example, the same rating behavior may be displayed but internal decision criteria may be different.

Practically, however it is still largely unresolved, how user factors can be integrated in the cycle of developing and conducting experiments, creating models and predictions, applying them in practice, and refining them for new services or new user groups.

4.3.1 Experiment lifecycle

The Figure 1 illustrates a typical flow of an experiment. The emerged ideas related to each phase are shortly discussed below (implicitly showing the focal points of discussion as well).

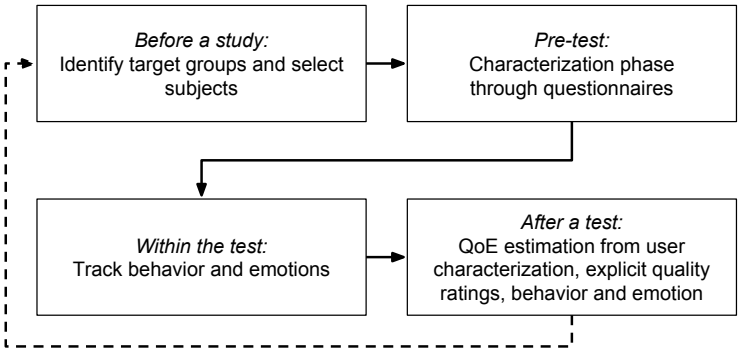

Figure 1 Experiment Lifecycle.

a) Before a study

The purpose of the study to be typically defines some ideal target groups from which the subjects can be selected. For example, in case of business oriented study, it is important to cover the real market segments. Here, we refer to marketing literature where a lot of existing information can be exploited.

b) Pre-test

Typically some demographic data is collected in QoE tests, including features like age, gender and expertise. Still, they are rarely even considered and typically not collected in standard fashion. There are numerous metrics or factors that could be considered:

- (Socio-)demographic factors such as age, gender and profession.
- Physiological factors such as long-term personality traits or short-term moods.
- Expectations which relate to various aspects such as motivations of users (gratification), attitudes towards technology or other people, and experience and relation to the service (e.g., frequent vs. casual gamer).
- Willingness to pay and spending behavior (how much and when does the user spend typically?)

At the moment, there is too little data, often lacking structure, to be able to determine which factors are influencing quality ratings significantly. Typical experiments use a small number of subjects, and both test stimuli and duration are short. All this makes it difficult to identify relevant factors. However, we think that the community should continue to encourage experimenters to continue to collect and report at least the typical factors, *in a systematic manner*, in order to gather at least descriptive and qualitative data.

In the long run, though, it would be highly desirable to identify and validate the important factors with quantitative results from tests with higher number of subjects and long-term duration. There are validated questionnaires in our field and related domains, which could serve as starting point. It may be necessary to define a new set of questionnaires or complement the existing ones. With understanding gained from such a study, we could come up with reduced and standard pre-test questionnaire.

It is also possible that the clustering of users based on user factors is sensitive to the service. Can we expect that they are the same for different services or not? Currently, such knowledge does not exist. One may be able to group similar media types, such as audio and video.

c) Within a test

Users have an internal reference of quality which they compare the currently experienced quality against. This internal reference varies per person, but may be clustered into groups. In addition to internal reference, subjects differ, e.g., in terms of sensitivity to stimuli and many physiological aspects.

Moving towards exploiting QoE, and going beyond perceived quality, user behavior was also discussed. Monitoring behavior during the tests may give insight of how people react to different quality levels.

Validity and reliability of rating. Each person assesses quality according the their internal reference and judgment process. So even if users agree on the rating, how they get there and because of what reasons, may differ. It is important to also note that these internal references change over time. So, how can we know do ratings present what we think they represent? How to create the picture of an internal reference of the observer?

There are several existing measures, for example: self-reports such as confidence ratings, recording rating behavior such as rating times, and using passive objective methods such as EEG measures, to identify internal cognitive thresholds, i.e., did the user notice anything?

Behavior. As with user factors, there are multitude of behaviors and metrics. To investigate the effect of QoE on behavior we need to include behavioral measures, but which ones? How to classify different types of behavior? Could the measures from related domains, such as UX, be exploited to complement QoE studies? The following issues were discussed:

- Two main types of behavior were identified: Adaptive behavior that may help in maintaining higher QoE (e.g. increasing pauses in conversation when the delay is high), or expressive behavior that a user may display, but that does not help to improve the QoE (e.g., aggressive clicking).
- Including behavioral measures (and observing user behavior) is possible to some extent in the lab, but may be more interesting in more realistic settings. Longer and realistic settings could enable us to track "default" behavior for certain users or user groups and then detect anomalies.
- Including behavioral patterns as part of user factors (i.e., user groups), would enable to create panels of users where behavioral patterns are better known. Such a capability would be highly valuable, e.g., for campaigns studying economic viability of services or features.
- Currently, most test paradigms enforce behavior in the sense of asking people for a quality rating in defined scenario, where the subject is told to imagine himself or herself performing a certain task. More ecologically valid tests would aim at giving realistic tasks and observing more "natural" user behavior.
- Also, the tests typically force people to focus on one service, and do not allow them to use other services or applications at the same time. How could we integrate the possibility of multi-tasking into a lab test? For example, allow them to do whatever they want at the side while doing a test and observing their behavior?

d) After a test

Typically, we would like to complement the existing QoE predictions with the collected user characteristics, behaviors and emotions to provide better accuracy. Understanding the relations between user factors, technical factors and behavior and emotions allows also creating prediction models related to behavioral aspects, such as willingness to pay.

4.3.2 Research questions and open issues

- Which user factors are actually the important ones?
- Can we use existing validated questionnaires to capture these factors or do we need our own set of validated questionnaires?
- What is the minimal set of factors that everyone should collect and report?
- What types of behavior are there and how they should be measured?
- How can we track behavior across multiple services and longer time intervals?

4.4 QoE management

Henning Schulzrinne (Columbia University – New York, US)
Martín Varela (VTT Technical Research Centre of Finland – Oulu, FI)

4.4.1 Architecture

We believe that many networked applications can benefit from following a QoE-enhanced architecture, shown below. In that architecture, the application, shown as "application", is observed by an application-specific observer function. The observer may, for example, gather information about video stalls, start-up latency or user interactions for a video playback applications, or audio gaps and speech behavior for a conferencing application. For web-based applications, many of these observations can be made on the web server, but JavaScript-heavy applications may also need local instrumentation. The applications, using operating system interfaces to be defined, also provide information about network characteristics (carrier, packet loss) to the application provider. The application is augmented by an advisor function that detects when the application is not performing well, provides an indication to the user, preferably indicating the underlying cause at an appropriate level of technical detail, and offer recommendation to the user on how to improve the experience. In some cases, the application can adjust the behavior automatically. The predictor function tries to anticipate problems, particularly for applications with long-lived sessions, such as video playback and interactive communication applications or multi-player games. For example, it might warn the user that the network conditions are unlikely to support a smooth application experience. The application provider filters the incoming application and low-level observations, possibly with the assistance of machine learning. The filtered information then may be visualized, but more importantly, feeds the advisor and predictor functions. Individual user observations may be combined by appropriate crowd-sourcing, e.g., correlating metrics within the same household, the same ISP or same computing platform.

4.4.2 QoE repair models

Traditionally, consumers notice degradation in QoE, contact their application or network provider by phone, and the help desk walks the customer through a series of manual steps, including short-term actions ("reboot") and longer-term decisions ("upgrade your service"). The help desk staff is essentially blind and limited to indirect observations. More recently, the help desk functionality has been enhanced by the ability to see some customer-specific metrics or install remote login tools. The long-term goal should be to avoid the need to call customer support – problems are detected, diagnosed and repaired automatically.

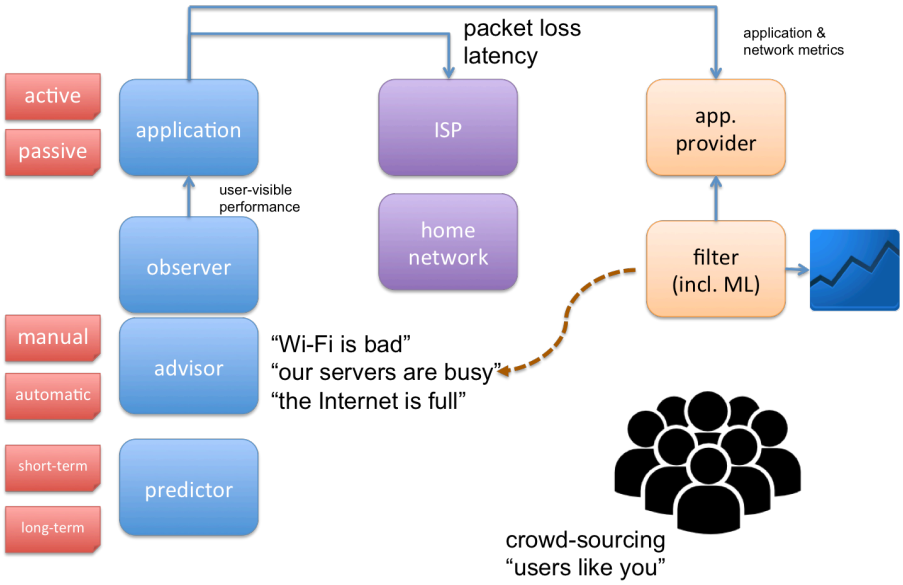

Figure 2 High-level architecture for QoE management.

4.4.3 Privacy concerns

Effective feedback loops may expose additional user data, such as user behavior, to the ISP or application provider. Thus, such systems must be designed to minimize information leakage to third parties, and ISPs and application providers must communicate clearly what information is gathered, how long it is retained in individualized or aggregate form and whether it is also used for purposes other than diagnostics.

4.4.4 Research questions and open issues

- **Intersection with other research disciplines.** Understood within a closed-loop context, the approach has strong overlap with other traditional and emerging research communities, such as classical network management, Software Defined Networking (SDN), smart data pricing and cross-layer design.
- **IWhat are the interfaces?** All key system components, including the operating system, home network components, the ISP, application and browser need suitable APIs, whether local (JavaScript) or HTTP-based. It is not yet clear what information should be exposed, to whom or at what time scale or granularity, or what functions need to be controllable.
- **What are the actuators?** Many of the components listed only allow users to set properties through configuration, command line or human-focused web interfaces, not APIs. For example, it would be useful if ISPs had interfaces that allowed authorized applications to increase the data cap or temporarily increase the connection speed. Home network gateways may need the ability to differentiate various users, e.g., provide priority to work-related applications.
- **What are the control loop properties?** Control loops needs to fail safe, so that applications continue to functions reasonably well even if some of the external functions cannot be observed or controlled. Control functions need to avoid adding to congested network and contribute to signaling storms. Feedback may be directed to more than one recipient in the control loop (e.g. ISP and application provider).

- **How can we support root cause analysis?** Today, determining the cause of intermittent QoE problems is often tedious and time consuming. What kind of information can components gather to simplify post-incident root cause analysis? Should there be an emulation capability so that users or programmers can easily replicate possible root causes, such as packet loss or intermittent connectivity? Like CCTV, better logging may not prevent the QoE crime, but help in apprehending the QoE perpetrator.
- **How can we anticipate and prevent problems?** Systems should be able to predict and anticipate flash crowds, for example, on the timescale of at least minutes, sufficient to marshal additional computing and network resources.
- **We need a new design-focused approach for QoE.** Research is needed to identify the appropriate overall system architecture that allows components to communicate in order to improve QoE. Standards bodies then need to define appropriate component-specific standards and data formats. Finally, both networking and software engineers need to be trained to towards a measurement-focused, whole-system approach of system design.

4.5 Monetization of QoE

Martín Varela (VTT Technical Research Centre of Finland – Oulu, FI)

4.5.1 Introduction

The goal of this working group was to discuss issues related to improving business (for any type of online service providers) by exploiting our knowledge of QoE. Given the group composition – with many participants from a networking background – the discussion had a distinct networking flavor, but OTT and cloud services services were also considered.

4.5.2 Main issues found...

a) In QoE-based business models

There are several key aspects to understand if we are to use QoE as a business-improving tool. The first one of those is how to convey what quality *feels like* to the end users. A lack of reference points for quality in most services (e.g., as "toll quality" was for the PSTN) presents a major roadblock for selling QoE to end users.

Secondly, from the operator's perspective, it is important to understand and measure the users' willingness to pay for the service, possibly at different quality levels. We should note here that the different quality levels might imply not just a different average MOS score, but rather differences in service dependability (consistency) or trustworthiness, for example. Metrics for these factors are currently under-developed topics in our community, and their impact may differ depending on the class of user, such as personal vs. professional use.

Thirdly, in competitive markets, the ability to segment users is limited, in theory, by the incremental cost of providing the improved quality (if a provider in a competitive market were to charge more, a competitor would underbid for the same quality step). Often, the cost of additional bandwidth is low once a network has been built, and costs for customer support and administrative overhead do not depend on usage. The ability to differentiate customer segments is strongest where customers have limited choices and where the need for quality strongly correlates with the ability to pay. Sometimes, providers also restrict

who can access low-cost, low-quality offerings, such as the low-bandwidth Internet access restricted to low-income families in the US. As another example for networks, prepay cellular packages may limit roaming, thus offering reduced coverage at a lower price point. Examples of market segmentation of applications by QoE include Netflix's UHD option, or Spotify's subscriber tiers. Some research [2] also suggests that quality-based market segmentation is possible.

On the technical side, there is a lack of standardized, well-engineered and robust measurement and monitoring tools for QoS and QoE, which are needed to successfully exploit QoE in a business context. The architecture proposed in the QoE Management working group could be a good starting point for addressing this gap.

b) Combining QoE and SLAs

QoE knowledge could be exploited in several places to facilitate SLAs:

- Facing the user (e.g., sell a consistently-guaranteed quality level);
- As an internal SLA by the service provider (e.g., use it to drive resource management so as to provide a certain quality level to the users);
- As a means of supporting SLAs in multi-provider scenarios (e.g., SaaS vendors who require guarantees from upstream providers).

Given the first issue discussed above on how to communicate the meaning of QoE to the users, the first option seems the hardest to work out currently. The second and third options could use QoE models as tools in determining which SLAs or resource management mechanisms would best suit the service provider's goals e.g., [1].

Another limiting factor related to SLAs is that in many contexts, they simply don't exist or are very poor. For example, user-facing network plans rarely have any service level guarantees (e.g., "*Up to* 100 Mbps", which in practice is never realized). Even in some commercial contexts, such as cloud hosting, SLAs can be almost non-existent (e.g., Amazon's AWS). Offering SLAs in these commercial contexts might also come with liability implications that require a careful risk analysis.

4.5.3 Discussion

a) Economic differentiators

ISPs mainly use speed as a low-level differentiator, while cellular providers compete on coverage and network technology (e.g., 4G). For enterprise users, reliability (and speed of recovery from failure), or sharing (virtual circuits vs. best effort) seem most promising.

For cloud services, there are more possibilities to differentiate quality offerings for different customer tiers, but they are – as can be expected – very service-dependent. For OTT video, for example, different quality levels and content availability could be a way to tier subscriptions.

b) Information disclosure

A common recurring topic during the discussion was that of properly communicating with the customers on quality and performance issues. The main issue related to this is about communicating what a certain level of QoE *feels like* to the users, if QoE-based differentiation is to be done at all.

The requirements for information disclosure are also different when dealing with end users, or with "edge providers" (e.g., SaaS providers), both in terms of metrics, and of language.

c) Public policy

When thinking about QoE-based differentiation, the issue of network neutrality appears very quickly in many cases. There are several examples where delivering a service with proper quality might require deals between the service and network providers involved, which could lead to neutrality issues (e.g., Netflix in the USA).

In some countries (e.g., Chile, Finland), telecom regulators have started imposing certain performance levels and/or availability of clearly defined metrics with which ISPs need to comply. This concept could be expanded upon to include QoE indicators for a group of services representative of what most users use, for example. Other regulators, such as Ofcom and the FCC, have promulgated voluntary or mandatory transparency requirements for ISPs, sometimes based on measurements ("Measuring Broadband America").

d) Non-ICT domains

An interesting parallel was drawn between the OTT services and ISPs relation and that of power generation and distribution in the Nordic countries. It would be interesting to study these in more detail, as the problems faced are remarkably similar in several aspects, and have been successfully solved in the case of power grids.

Other non-ICT domains, such as retail, might also provide interesting tools to apply to online services.

4.5.4 Research questions and open issues

- **In order to enable QoE-aware business models and SLAs, how can we convey the "meaning" of quality to the end users?** Three different levels of information disclosure were identified:

  End-users: A "QoE Emulator" would allow users to experience the different quality levels and make a decision based on them. This would allow to demonstrate how different services would be experienced at peak times. Another option is a crowd-sourced approach, where ratings of similar users under similar conditions might be used as guidelines.

  Professionals and regulators: For professionals, a MOS-like quantitative estimate may be sufficient, but may not reflect variability in time and across the service territory.

  Engineers: Multiple lower-level metrics, and if possible causal links between them and QoE, allow root cause analysis.

- **What to measure, and how to measure it?** From the ISP's point of view, most issues tend to occur in the "middle mile" which is shared by multiple users. For users, issues are often found in their home networks, which cannot be easily monitored by the providers. Measurements would then be needed in several points, including the end user's premises, the last mile, and the links between the POP and the actual services. The types and number of services to be measured in order to obtain representative results is also an open question. Once this has been solved, there remains the issue of communicating this to end-users in an understandable manner. Some ideas to deal with this could be a "food label"-like approach[1] that would put some technical aspects in terms understandable by users (e.g., 5-star ratings). An objective third party could collect the measurements

[1] Inspired by the nutritional information labels on food products.

and provide ratings for network and service providers available in a certain area, so that consumers would be able to gauge the differences (in terms of quality) between the different providers and their different plans.

- **How to identify different customer segments?** In this context, willingness-to-pay studies might be a good starting point, showing how users are willing to spend money on services at different qualities. Parallels drawn from other domains might also provide interesting insight, such as the case of Nordic countries' power grids mentioned above, for instance.

References

1 Pantelis A. Frangoudis, Aggeliki Sgora, Martín Varela, and Gerardo Rubino. Quality-driven optimal SLA selection for enterprise cloud communications. In *Proceedings of the IEEE ICC 2014 – Workshop on QoE-centric Network and Application Management (QoENAM)*, June 2014.

2 Andreas Sackl, Patrick Zwickl, and Peter Reichl. The trouble with choice: an empirical study to investigate the influence of charging strategies on content selection on QoE. In *Proceedings of the 9th International Conference on Network and Service Management (CNSM)*, Oct. 2013. DOI: 10.1109/CNSM.2013.6727850

4.6 QoE in new domains

Markus Fiedler (Blekinge Institute of Technology – Karlskrona, SE)

4.6.1 Starting points

The collection of topics and questions that were raised from the participants revealed interest in new domains for QoE. Initially, the areas of eHealth, education, and smart grids were identified, and complemented by issues regarding software, installation, billing, and security.

4.6.2 Leaving the QoE comfort zone of multimedia

The classical domain for QoE is multimedia experience with particular focus on audio and video, which is nicely reflected in the most recent, Dagstuhl-born and hedonic-centered definition of QoE as "degree of delight and annoyance" [1, 2, 3]. On the other hand, vendors, providers and operators are interested in QoE as a means of assessing and controlling the risk of user churn. Indeed, the risk of user annoyance and consequent churn nowadays exists in many ICT service domains. When users face suboptimal service performance, they may abandon the service, which makes providers, operators and in the end even vendors face unfortunate economical consequences. In particular, the introduction of new types of services has to be successful, or put the other way round, should not be "overshadowed" by QoE issues. Thus, from the viewpoints of vendors, providers and operators, QoE (and in particular its connection to churn, Service Level Agreements and business aspects) has become a focus area for both existing and emerging services. This poses the particular challenge to the QoE community of how the QoE definition, focusing on *well-being* (delight and annoyance), can be applied to any emerging kind of service, far beyond audio and video. However, purely focusing on well-being and QoE is not sufficient. It is evident that even aspects of *usability*, in particular regarding installation and usage, as well as *utility* need to be addressed in order to pave the way towards acceptance and success of new services.

4.6.3 Emerging areas

Areas showing an increased interest in QoE and related issues are amongst others:

- *Internet of Things* (IoT), with the challenge of implementing Machine-to-Machine (M2M) communication that satisfies the (human) end user and enables new Business-to-Business (B2B) relationships;
- *E-Health*, with the challenge of designing and implementing applications and services that are accepted by all stakeholders under rather complex usage and legal conditions;
- *E-Learning*, with the challenge of having services at hand that do not impede, but strongly support the learning process;
- *Smart data*, where a key value is found in the right information in the right place at the right time;
- *Serious games*, where the "gamification" of use cases (such as memory training) needs to be done in an appealing and convincing way;
- *Arts and culture*, where well-working and -perceived services open up for new channels to reach existing and new audiences.

All those areas have in common that well-being, usability and utility are key ingredients for successful service delivery.

4.6.4 Demands, net benefit, and utility

A review of demands in the different areas has shown that *time-based quality and consistency issues* dominate. One of the delegates pointed out that if one cannot provide the service in a consistent manner, one will face issues with experience and utility. Indeed, the approach in [4] introduces the *net benefit* as principal determinant of user behavior and – through user satisfaction – customer behavior. Net benefit is the result of *gross benefit* – impacted by fulfillment of needs and service quality – and *sacrifice* – impacted by service quality (again!), value of time and usage price. The ambivalent role of service quality as both enabler (if good) and disabler (if bad) can also be seen in utility values, which can be positive (gains) or negative (losses). Obviously, time plays a key role as sacrifice to be paid by the user (e.g. in the form of boredom or lost opportunities). Thus, waiting time issues translate into negative utility issues, which is counterproductive to the demand of positive utility for the successful implementation of new services, and these waiting times have to be modeled and evaluated. Related concepts from the area of Software Engineering, combining time, error, effort and usage, can contribute to this task.

4.6.5 Research questions and open issues

Given the rapid development of new fields with the need of QoE, usability and utility considerations, the following research questions emerged:

- Once a new field appears, how to develop a quality concept for that?
- How should QoE in those new domains be assessed?

There is an obvious need to deepen these issues, potentially in the form of another QoE-related Dagstuhl Seminar.

References

1 Dagstuhl Seminar 09192. http://www.dagstuhl.de/09192
2 COST Action IC 1003 QualiNet. http://www.qualinet.eu
3 Dagstuhl Seminar 12181. http://www.dagstuhl.de/12181
4 Kalevi Kilkki. *An Introduction to Communications Ecosystems*. 2012. http://kilkki.net

5 Concluding remarks

Katrien De Moor (NTNU – Trondheim, NO)

The main objectives of this third Dagstuhl seminar on Quality of Experience were to strengthen and go beyond the current understanding on QoE, and to push the transition from the assessment of QoE to the practical application of QoE knowledge and mechanisms. Keeping these objectives in mind and referring back to the detailed summaries from the different working groups above, we can conclude that significant steps forward were made. The fruitful discussions amongst the participants resulted in the identification of both points of crystallization concerning our current understanding of QoE and open questions and challenges for future joint endeavors.

The joint work on the fundamentals of QoE ("QoE theory and modeling") resulted in a comprehensive QoE and user behavior model, providing a high-level integrative framework which was to date still lacking. The discussions on "QoE methodologies" were centered around a number of caveats and challenges related to the traditional methodological approaches in QoE research and more specifically, on concrete proposals on how these could be better addressed in order to increase the ecological validity of QoE research. Similarly, in the group work on "User factors", and based on the limitations of current approaches, the discussion resulted in concrete suggestions on how to better take user factors into account (in terms of measuring and capturing user factors, incorporating these factors into QoE analyzes and exploiting the related insights in the application of QoE principles and insights) and in the identification of open questions that can steer future joint work on this matter.

Complementing and going beyond the research perspective, the fourth working group identified challenges related to "QoE Management". As a concrete and important step forward, a high-level and QoE-enhanced architecture was developed. In addition, a set of related research questions and requirements, including the need for a new design-focused approach for QoE, were discussed and put on the agenda. The business perspective and challenge of turning QoE-related knowledge into economic value, were the main focus of the discussions in the group on the "Monetization of QoE". Barriers to the development and implementation of QoE-aware business models and SLAs, as well as key issues that need to be tackled in order to overcome these barriers (such as e.g., the identification of different customer segments), were identified and discussed. Finally, the working group on "QoE in new domains" explored the relevance and application of QoE for fields that go beyond the "safe and familiar" multimedia domain. The concrete domains that were considered in the discussion include Internet of Things, E-Health, E-Learning, Smart data, Serious games and Arts and Culture. Each of these new domains brings along distinct challenges for research on QoE, which need to be further investigated and deepened in the future. By putting

these domains explicitly on the future QoE-agenda, we believe that the seminar may play a triggering and accelerating role in this respect.

During the wrap-up session on the last day, which featured the high-level summaries from the different working groups, the seminar in itself was also briefly evaluated. More specifically, every participant was asked to make a short statement about the seminar and its organization. From this short feedback session, it became clear that the group work, and the fact that the time for individual presentations was limited, was in general very much appreciated by the participants. However, it was also commonly repeated that there was perhaps not enough time for "digestion", reflection and further discussion due to the tight schedule and limited time (2.5 days). Similarly, it was argued that there should also be more time to report back to the whole group after the break-out discussions, and more time to take up some of the discussed issues further in plenum.

Altogether, we can look back at a successful and inspiring seminar, which triggered lively discussions and cross-disciplinary exchange, and which resulted in plans for joint follow-up activities in several of the discussion groups. 2.5 months after the seminar, a number of concrete outcomes can already be included in the report. For example, one paper has already been accepted for presentation at the IEEE ICC 2015-workshop on "Quality of Experience-based Management for Future Internet Applications and Services" [4], and three other joint papers that either directly originate from the discussions during the seminar, or have been inspired by them, have been accepted for presentation at the 7th International Workshop on Quality of Multimedia Experience (QoMEX 2015)[1, 2, 3]. In addition, many open issues, which will require deeper discussions (e.g., at potential future Dagstuhl seminars and at other venues), and which may lead to future collaboration between the participants, have been put on the agenda. In line with the previous Dagstuhl seminars on QoE, we are therefore confident that the Dagstuhl QoE community has been strengthened by the seminar and that several of the ideas discussed during this seminar will find their way into the literature, or become visible in another way in the future.

References

1 Jan-Niklas Antons, Sebastian Arndt, Katrien De Moor and Steffen Zander. Impact of perceived quality and other influencing factors on emotional video experience. In *Proceedings of the 7th International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015.

2 Tobias Hoßfeld, Poul E. Heegaard and Martín Varela. QoE beyond the MOS: Added value using quantiles and distributions. In *Proceedings of the 7th International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015.

3 Peter Reichl, Sebastian Egger, Sebastian Möller, Kalevi Kilkki, Markus Fiedler, Tobias Hoßfeld, Christos Tsiaras and Alemnew Sheferaw Asrese. Towards a comprehensive framework for QoE and user behavior modelling. In *Proceedings of the 7th International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015.

4 Martín Varela, Patrick Zwickl, Peter Reichl, Min Xie, and Henning Schulzrinne. Experience Level Agreements (ELA): the challenges of selling QoE to the user. In *Proceedings of the IEEE ICC 2015 – Workshop on Quality of Experience-based Management for Future Internet Applications and Services (QoE-FI)*, June 2015.

Participants

- Jan-Niklas Antons
TU Berlin – Berlin, DE
- Alemnew Asrese
Aalto University – Espoo, FI
- Katrien De Moor
NTNU – Trondheim, NO
- Philip Eardley
British Telecom R&D – Ipswich, GB
- Sebastian Egger
AIT – Wien, AT
- Markus Fiedler
Blekinge Institute of Technology – Karlskrona, SE
- Farnaz Fotrousi
Blekinge Institute of Technology – Karlskrona, SE
- Pantelis Frangoudis
INRIA Rennes – Bretagne Atlantique, FR
- Samuel Fricker
Blekinge Institute of Technology – Karlskrona, SE
- Juan Pablo González Rivero
Plan Ceibal – Montevideo, UY
- Poul Einar Heegaard
NTNU – Trondheim, NO
- Tobias Hoßfeld
Universität Duisburg-Essen – Essen, DE
- Kalevi Kilkki
Aalto University – Espoo, FI
- Eirini Liotou
National Kapodistrian University of Athens – Athens, GR
- Toni Mäki
VTT Technical Research Centre of Finland – Oulu, FI
- Sebastian Möller
TU Berlin – Berlin, DE
- Peter Reichl
Universität Wien – Wien, AT
- Miguel Ríos Quintero
TU Berlin – Berlin, DE
- Werner Robitza
TU Berlin – Berlin, DE
- Henning Schulzrinne
Columbia Univ. – New York, US
- Lea Skorin-Kapov
Univ. of Zagreb – Zagreb, HR
- Samira Tavakoli
Universidad Politécnica de Madrid – Madrid, ES
- Christos Tsiaras
Universität Zürich – Zürich, CH
- Astrid Undheim
Telenor Research and Development – Trondheim, NO
- Martín Varela
VTT Technical Research Centre of Finland – Oulu, FI
- Min Xie
Telenor Research – Fornebu, NO
- Patrick Zwickl
Universität Wien – Wien, AT