Deriving QoE in Systems

From Fundamental Relationships to a QoE-based Service-level Quality Index

Tobias Hoßfeld · Poul E. Heegaard · Lea Skorin-Kapov · Martín Varela

Abstract With Quality of Experience (QoE) research having made significant advances over the years, service and network providers aim at user-centric evaluation of the services provided in their system. The question arises how to derive QoE in systems. In the context of subjective user studies conducted to derive relationships between influence factors and QoE, user diversity leads to varying distributions of user rating scores for different test conditions. Such models are commonly exploited by providers to derive various QoE metrics in their system, such as expected QoE, or the percentage of users rating above a certain threshold. The question then becomes how to combine a) user rating distributions obtained from subjective studies, and b) system parameter distributions, so as to obtain the actual observed QoE distribution in the system? Moreover, how can various QoE metrics of interest in the system be derived? We prove fundamental relationships for the derivation of QoE in systems, thus providing an important link between the QoE community and the systems community. In our numerical examples, we focus mainly on QoE metrics. We furthermore provide a more generalized view on quantifying the quality of systems by defining a QoE-based Service-level Quality Index (SQI). This index exploits the fact that quality can be seen as a proxy measure for utility. Following the assumption that not all user sessions should be weighted equally, we aim to provide a generic framework that can be utilized to quantify the overall utility of a service delivered by a system.

Tobias. Hoßfeld
University of Würzburg, Germany, E-mail: tobias.hossfeld@uni-wuerzburg.de

Poul E. Heegard
NTNU, Trondheim, Norway, E-mail: poul.heegaard@ntnu.no

Lea Skorin-Kapov
University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia, E-mail: lea.skorin-kapov@fer.hr

Martín Varela
callstats.io, Helsinki, Finland, E-mail: martin@varela.fi

1 Introduction

One of the main research challenges faced by the the QoE community is deriving QoE models for various applications and services, whereby ratings collected from subjective user studies are used to model the relationship between tested influence factors and QoE. With it being well known that different users perceive both quality and value differently [1], *user diversity* will inherently impact the distribution of rating scores for a given test condition [2], [3]. However, the majority of user studies to-date still report only on MOS (Mean Opinion Score) values and confidence intervals, and utilize these values to derive QoE models. When focusing on technical Quality of Service (QoS) influence factors, this leads to the common reporting of so-called *QoS-to-MOS* mapping functions.

Previous work has argued that from a service/network provider perspective, there is a likely interest in additional metrics beyond MOS values, thus providing deeper insight into rating distributions and how various conditions are perceived by the user population [4, 5, 3] (as opposed to how conditions are perceived by an "average user"). As an example, the *GoB* metric gives the probability that for a given condition, the user rating will be "good or better" [6] (e.g., on a 5 pt. Absolute Category Rating, ACR, scale, this corresponds to a rating of 4 or 5). In addition to a *QoS-to-MOS* mapping function, the results of a user study could thus be used to derive and report also a *QoS-to-GoB* mapping function. Such a mapping function could subsequently be used by a service or network provider in the context of QoE management when aiming to maximize the percentage of "happy" users in the system [7]. To generalize, subjective studies are used to derive QoS-to-QoE mapping functions, where QoE in this context can refer to any QoE metric of interest (e.g., MOS, GoB).

1.1 Fundamental Relationships for Deriving QoE in Systems

Moving from the domain of *user studies* to the *systems domain*, we consider service/network providers interested in deriving various QoE metrics in their system, given (a) the system performance, and (b) QoE models available from user studies. To put it in a mathematical context, we observe certain system parameters, which is described by a random variable (RV). Assuming, for illustration purposes, a Web-based service, system performance may be quantified by the response time X experienced by the end user. As a result, we have a response time distribution in the system, meaning various users will experience different response times. On the other hand, going back to the results of subjective studies, we know that the user ratings for a certain test condition

(response time) also follow a distribution. Hence, due to user diversity, the experienced QoE for a certain response time $X = t$ is again a distribution $Q|_t$. The question arises as to what is the observed QoE distribution Q in the system, when X is a random variable of the system's performance and $Q|_t$ is a random variable of the user's QoE for $X = t$? Moreover, how can various QoE metrics in the system be derived, such as expected QoE and expected GoB?

To this end, we highlight the following key contributions of the paper:

✓ We prove a **fundamental relationship** showing that the **expected QoE in the system is equal to the expected MOS in the system**, despite the fact that the actual QoE distribution in the system is not (necessarily) equal to the MOS distribution in the system. We note that the MOS distribution in the system is obtained by mapping response times of the system to MOS values as per a given QoS-to-MOS mapping function.

✓ We show that to derive additional QoE metrics in the system it is **necessary to use corresponding mapping functions** derived from user rating distributions in subjective studies. In particular, to derive the expected GoB metric in the system, a QoS-to-GoB mapping function is needed. If only a QoS-to-MOS mapping function is available, it is not possible to derive the expected GoB in the system.

✓ Going beyond our previous work [8], in which we derived fundamental relationships assuming that QoE depends on a single QoS parameter only, we now show that **these relationships can be extended to an arbitrary number of parameters**. Novel use cases are considered to demonstrate how to use those fundamental relationships. Firstly, the dimensioning of a web server based on a target GoB ratio is discussed. Secondly, we consider HTTP video streaming QoE as an example for a multidimensional QoE relationship.

To stress the implications of these contributions, we again highlight the link between the QoE community, systems community, and end users: if researchers conducting subjective user studies provide different QoS-to-QoE mapping functions for QoE metrics of interest, **this is enough to derive corresponding QoE metrics from a system's perspective**. This holds for any system parameter distribution, as long as the corresponding values are captured in the reported QoE models.

The term 'service' has evolved over the years from being simple transport of data to a model where access to an application may be delivered as a service. When deriving QoE-related metrics in a system, we consider the system as offering a *single service* (e.g., web browsing, video streaming, etc.). Multiple users using the service each experience their own QoE. Adopting a more generalized approach (portrayed in Figure 1), we may consider a *system* as potentially offering multiple *services*. During a certain time period, multiple *sessions* corresponding to a given service may be active in the system. A session involves one or more *users*. For example, in a network (*system*) we could consider HTTP streaming (a *service*). Over the course of one hour, multiple users are each watching their own video streams for a given duration (*session*).
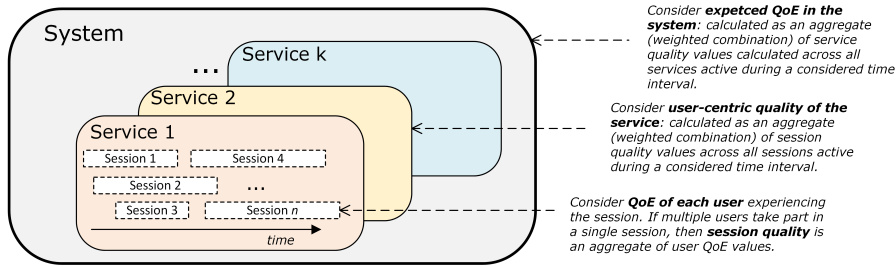
Fig. 1 The relationship between systems, services, and sessions, as considered in this paper.

For a **given service**, we distinguish between the following:

- the QoE of an individual user experiencing a session, and
- a measure indicating the overall service quality over a target time period (covering multiple sessions).

It is important to note that QoE is inherently linked to an end user experience. Thus, when measuring aggregate quality over multiple users, we are no longer referring to QoE *as such*, but rather to a QoE-based quality aggregate.

1.2 Not All User Sessions Are Created Equal

In our aim to calculate expected QoE in a system and prove fundamental relationships, in Sections 2 and 3 we will consider a system as offering a single service, where all individual QoE values (inherently linked to individual users) are *equally weighted*.

In a realistic setting, however, where a system operator is aiming to utilize estimated QoE values of users in the system for purposes such as dimensioning, monitoring, or benchmarking, the *relative importance* of user sessions may need to be considered. We therefore go a step further and provide a **generic framework** for quantifying the utility of a given service from a QoE-based perspective, discussed in Sec 4. We address cases when not all user sessions (and consequently corresponding user QoE values) should be weighted equally.

Continuing with the previously given example of HTTP video streaming, we consider the service delivered to multiple users in a network over a one-hour period. We utilize QoE models that map system performance to QoE for each individual user viewing session. One option is to calculate the QoE for each user session, and then consider the average value to be the expected QoE in the system *for that service* (as reported in Sections 2 and 3). However, in a realistic case, some users may watch short 1—3 minute video clips, while others may watch videos lasting 30 min — 1 hour. Such sessions widely differ both in terms of duration, as well as in terms of consumed system resources. Consequently, we argue that when calculating an aggregate quality index for the given time frame, it may be relevant for the service provider to consider

some sessions to be "more important" than others, i.e., they do not contribute to the same extent to a measure of the overall service quality.

We therefore define a QoE-based ***Service Quality Index (SQI)*** as: *a measure indicating the overall utility of a service delivered by a system and derived as a weighted combination of quality values estimated per user session.* Individual QoE values are weighted according to factors that are deemed relevant by a service provider, and may be related to session characteristics (e.g., session duration), resource consumption (costs), number of users involved in a session, etc. We note here that we are working off the assumption that QoE is indeed a good proxy for the users' utility. While we think that this is indeed a reasonable assumption for most services, it should be verified for individual services when necessary. For simplicity's sake, we also assume a linear correlation between QoE and utility. This might be simplistic, as shown by Kilkki in [9], however we note that a non-linear relationship between QoE and utility does not affect the core of our proposal, it simply implies a slightly more complex weighting function. The fundamental relationship extended to SQI is also formulated in that way that *the SQI value in a system is equal to the expected utility values.*

1.3 Definition of Terms

Several different QoE-related terms and concepts are used throughout the paper. For the sake of clarity, we first define relevant terms according to their usage in this paper. Please note that proper mathematical definitions are provided in the later sections.

The term **Quality of Experience (QoE)** refers to a complex multidimensional construct comprised of various perceptual features contributing to the quality of an individual's experience. A commonly cited definition of QoE is provided in the Qualinet Whitepaper [10] stating that "*Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state.*" Moreover, in the context of communication services, it is stated that QoE is influenced by various underlying influence factors related to the service, content, device, application, and context of use.

The quantification of QoE for an individual user is obtained through **individual QoE user ratings**. While we are aware that an individual's rating may in fact be influenced by a wide range of underlying system, context, and human factors, we note that network and service providers commonly rely on simplified QoE models, developed to estimate QoE based on measurable QoS parameters in the system. Thereby, we consider a user as quantifying their QoE for a given condition (in our case expressed in terms of QoS parameters) on an underlying rating scale, e.g., a 5-point Absolute Category Rating scale (1: poor, 2: bad, 3: fair, 4: good, 5: excellent). This quantification is necessary for the analysis of systems and for the derivation of fundamental relationships.

For the sake of readability, we often simply say 'QoE of a user' instead of using the correct notion 'QoE rating of a user on the used rating scale'. We further note that the term 'QoE distribution' refers to the 'distribution of QoE user ratings on the used rating scale'. When using the term **QoE metrics**, we refer to an aggregation of individual QoE user ratings across multiple users for a particular condition. Examples for common QoE metrics are the Mean Opinion Score (MOS), which refers to the average rating over all users for a given test condition; the Good-or-Better (GoB) ratio indicating the ratio of users rating good or better on the given rating scale; and the Poor-or-Worse (PoW) ratio indicating the ratio of users rating poor or worse. A more detailed overview of QoE metrics and their definitions is provided in [3]. As stated previously, the majority of conducted user studies only report such aggregated metrics. Most commonly, studies report MOS values per test condition, thereby lacking insights into the distribution of QoE ratings for the condition. The reporting of such aggregate metrics hides the underlying diversity of actual user scores. Consequently, the reporting of only MOS values makes it impossible to infer the GoB ratio from a MOS value. With the MOS alone, we cannot say anything about the GoB. For example, a MOS of 3 is reached (a) if 100% of the users are rating 3 or (b) if 50% of the users are rating 1 and 50% are rating 5. Although resulting in the same MOS, a system provider may for example be more interested in the PoW ratio.

In a system, users will experience different conditions. When referring to the **expected QoE in the system**, we are in fact referring to the expected value of the QoE user ratings in the system. Hence, we are using the mathematical definition of *expected value*, which should in this context not be confused with the notion of *the users' expectations* (in relation to the perceiving subject's frame of reference) as used in the Qualinet definition and addressed in a number of studies [11]. Figure 2 visualizes the different terms and provides the relationship between individual users in sessions and the expected QoE in the system.

Finally, in this article we introduce the **Service Quality Index (SQI)** in a system, which we define as being based on session **utility functions** (further details provided in Sec. 4). Utility functions have generally been used to specify the relation between relative user satisfaction and consumption of a certain resource [12]. With the concept having been adopted from economics, utility functions offer a way to formalize the correlation between network performance and user perceived quality (QoE), by defining a formal mathematical vehicle for expressing a user's degree of satisfaction with respect to corresponding multi-criteria service performance [13, 14]. In this paper, when considering the *individual user perspective*, we consider a QoE mapping function as relating user perceived quality to system conditions, i.e., QoS parameters (please note that when assuming a broader view of QoE, this relationship can of course depend on additional context, user factors, etc.). In the context of customer retention and avoiding churn, it is well known that QoE is one of the main drivers for service and network providers when deploying their services [15, 16, 17]. Further, considering the *provider's perspective*, ensuring a certain

level of QoE for a given user session entails a certain cost, results in a certain revenue, etc. Thus, as previously mentioned, not all user sessions are necessarily considered equal in terms of their relevance and *value* to the provider. Consequently, focusing on the provider perspective, we refer to the *utility of a given user session* as the estimated QoE of the session, but scaled using an assigned weight factor, so as to reflect additional factors deemed relevant by the provider (e.g., cost in terms of resource consumption, profit). For example, a session with higher costs for the provider could result in a decreased session utility (obtained by scaling the QoE estimate with a lower weight factor). For a provider aiming to quantify the *overall quality* (or utility) of a service delivered in their system, we propose to integrate individual session utility values into a metric that we refer to as SQI.

The remainder of the paper is structured as follows. Section 2 provides the one-dimensional fundamental relationship between the QoE in the system and the subjective user studies for arbitrary QoE metrics. The multi-dimensional case is considered in Section 3. Section 4 then provides a generic framework for quantifying SQI from a user oriented perspective, building on the assumption that user QoE values in the system may need to be weighted differently, depending on the target use case. A final discussion and conclusions are given in Section 5.

2 Fundamental Relationship: QoE in the System for a Single Parameter

This section revisits fundamental relationships to quantify QoE in the system for a single service. Thereby, the users in the system consume the service in a similar way. In this section, we assume that the QoE depends on a single system parameter only. As a consequence, the QoE of the users, Q, may be derived by applying a one-dimensional mapping function to the single system parameter, X. As a concrete example, we consider web QoE. The users are consuming certain web pages from a server, whereby the response time of the server is the relevant QoE parameter. We use the web QoE example for illustrating how to use the QoE in the system for dimensioning a web server in Section 2.2. Before that, we summarize the fundamental one-dimensional relationships for QoE in the system, as originally contributed in [8]. The limitation to a single parameter is no longer required and will be extended to an arbitrary number of parameters in Section 3.

2.1 Fundamental One-Dimensional QoE Relationship

Figure 2 provides an overall picture on deriving QoE in a system. In a system, its users will experience different performance measures, such as response times, throughput, etc. For the sake of readability, in the following we will use
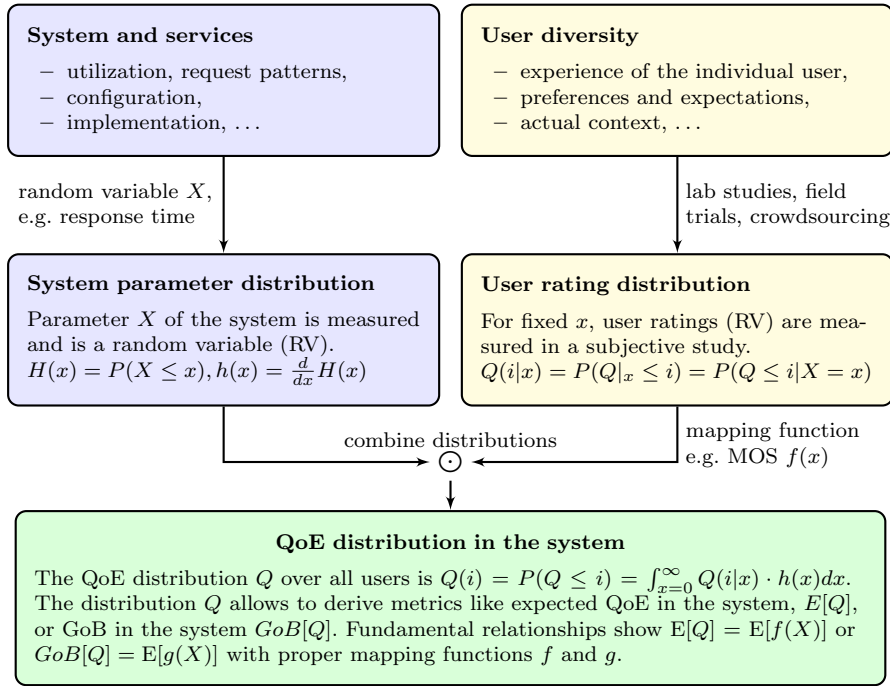
Fig. 2 Overview of system parameter distributions as observed in a real system, and user rating distributions in a subjective study. They are combined into a QoE distribution as observed in the system.

response times for web QoE as an example of a system parameter X. The system's performance depends on both its configuration and its implementation. However, since the system utilization varies as the offered load (requests) varies, the users will experience different response times, which can be represented by a continuous random variable X. The cumulative distribution function (CDF), $H(x)$, and the probability density function (PDF), $h(x)$, of the response time is

$$H(x) = P(X \leq x), \ h(x) = \frac{d}{dx}H(x) \,. \tag{1}$$

Two different users experiencing the same system condition (e.g., response time) x may rate the situation differently due to user diversity. The rating scale may be either discrete, like the typical 5-point Absolute Category Rating (ACR) scale, or continuous. Thus, we obtain a (discrete or continuous) QoE user rating distribution that depends on x. This is represented by a random variable $Q|_x$ for the QoE user ratings, given that the system parameter is $X = x$, with the CDF $Q(i|x)$ and PDF $q(i|x)$, as follows

$$Q(i|x) = P(Q|_x \leq i) = P(Q \leq i|X = x), \ q(i|x) = \frac{d}{dx}q(i|x) \,. \tag{2}$$

In the case of a discrete rating scale, $q(i|x)$ is the probability mass function (PMF) indicating the probability $P(Q = i|X = x)$ that the user rating is $Q = i$ for the system parameter $X = x$.

Now let us do the following **Gedankenexperiment**. All users in the system are rating the QoE on the same 5-point rating scale after a session. This could be implemented by the service provider with a proper interface in the provided application, although this may be very annoying in practice. Then, the service provider observes a random variable Q of all QoE user ratings as well as a random variable X of the corresponding QoS measurements. Hence, the provider also obtains the conditional QoE $Q|x$ which is a random variable reflecting the user ratings in the system for a particular QoS x. Then, we obtain the MOS mapping function $f(x)$ as the average over all user ratings for x, which is the expected value of the random variable $Q|x$. Hence, $f(x) = \mathrm{E}[Q|x]$. In a similar way, the GoB mapping function $g(x) = P(Q|_x \geq 4)$ is obtained for the underlying 5-point rating scale – and analogously other QoE metrics of interest like PoW. The fundamental relationships show how the service provider can combine the QoS measurements and QoE mapping functions and which statements can be derived on the expected QoE in the system or the ratio of users in the system with a QoE user rating which is good or better. To be more precise, the expected QoE in the system means the expected value of the random variable Q of the QoE users rating, i.e. $\mathrm{E}[Q]$. Those fundamental relationships are derived in the following.

Q is the unconditional random variable for the QoE user ratings over all the system performance conditions, with the CDF $Q(i)$ and PDF $q(i)$ as follows.

$$Q(i) = P(Q \leq i) = \int_{x=0}^{\infty} Q(i|x)h(x)dx \tag{3}$$

$$q(i) = P(Q = i) = \int_{x=0}^{\infty} q(i|x)h(x)dx \tag{4}$$

In practice, the service provider does not obtain the QoE user rating distribution Q. Instead, the provider may collect QoS measurements, i.e. the random variable X. The probabilities $q(i|x)$ and $Q(i|x)$ may be estimated from user ratings obtained by means of subjective studies, e.g., in the laboratory, via crowdsourcing, or by field trials, as long as the system condition x is observed. In user studies, the QoE user rating distribution typically has been obtained under certain (controlled or observed) conditions, which do not reflect the current system parameter distribution $H(x)$. The $H(x)$ might change due to reconfiguration or reimplementation of the system and its service, or due to changes in the offered load or system utilization.

A service provider is interested in the QoE distribution $Q(i)$ of all users in the system. The stochastic components of the QoE distribution are 1) *system parameter distribution X* (i.e. response time in our example) and 2) *user rating diversity yielding user rating distribution $Q(i|x)$*. To be more precise, the service provider is interested in QoE metrics like the expected QoE in the system over all users or the ratio of users in the system who rated the experience

QoE good or better. In practice, the service provider would not ask all users as we do in the Gedankenexperiment. But the service provider may be able to measure QoS and utilizes existing QoE models from literature, which provides e.g. a mapping $f(x)$ from QoS to MOS, a mapping $g(x)$ from QoS to GoB, or a mapping $q(i|x)$ from QoS to conditional user rating distributions.

> **Fundamental question:** what kind of information is required from subjective studies, such that the system provider may derive the metrics of interest from the distribution of Q, also when the system parameter distribution $H(x)$ changes?

While answering this question from a mathematical point of view, we need to be aware of the inherent limitations of such indirect QoE measurements. The service provider relies on the provided QoS mapping functions in order to estimate the QoE user rating distribution or QoE metrics of interest. However, in practice it may be sufficient to get a rough estimate of the true QoE of the users in the system to get a QoE oriented perspective rather than a QoS-driven focus.

2.1.1 Expected QoE vs. Expected MOS

We consider here the case of a discrete rating scale like a 5-point ACR scale for the sake of simplicity. In particular, we use a discrete rating scale with items $0, \ldots, n$ where 0 indicates the lowest QoE and n indicates the highest QoE of the scale. Please note that those fundamental relationships can also be derived for continuous rating scales analogously.

In Eq. (2) the distribution of the QoE user ratings i under a specific system parameter x is given. The expected user rating, given x, is the MOS value.

$$E[Q|_x] = \sum_{i=0}^{n} i q(i|x) \tag{5}$$

Let $f(x) = E[Q|_x] = \sum_{i=0}^{n} i q(i|x)$ be the mapping function between the condition x and the MOS rating (the mean opinion score for a given value of x), which may be derived from subjective user studies. In practice, subjective studies will typically cover only a few instants of the response time only due to cost reasons. However, the mapping function $f(x)$ is continuous, since x is continuous.

$$f : x \to \mathbb{R} \quad x \in \mathbb{R}_{\geq 0} \tag{6}$$

Then, a continuous mapping function f like the exponential function suggested by the IQX hypothesis [15] needs to be fitted to the collected MOS values. Please note that no assumptions are required for this mapping function $f(x)$, besides continuity.

The MOS mapping function allows us to derive the following fundamental one-dimensional relationship between the expected QoE in the system and the system parameter X. The expected QoE in the system means the expected

value of the random value Q reflecting the QoE user ratings in the system as observed in the Gedankenexperiment. For the sake of readability, we just use the term 'expected QoE in the system'.

Expected QoE in the system. The expected QoE in the system is equal to the expected MOS, $\mathrm{E}[Q] = \mathrm{E}[f(X)] = \mathrm{E}[M]$

---

where the random variable M of MOS ratings is the transformation from the random variable, X using the MOS mapping function, $M = f(X)$. This equality follows from

$$\mathrm{E}[Q] = \sum_{i=0}^{n} iq(i) = \int_{x=0}^{\infty} (\sum_{i=0}^{n} iq(i|x))h(x)dx$$

$$= \int_{x=0}^{\infty} f(x)h(x)dx = \mathrm{E}[f(X)] = \mathrm{E}[M] \ . \tag{7}$$

The $\mathrm{E}[Q]$ is the expected value of Q over the distribution of Q, while $\mathrm{E}[M]$ is the expected value of M over the distribution of X. This equation can be read from both sides. (a) $\mathrm{E}[Q] = \mathrm{E}[f(X)]$. This implies that the QoE user rating Q in the system is obtained as illustrated in the Gedankenexperiment. Then, we may derive the MOS mapping $f(x)$ from the user ratings in the system for any condition x and the equality holds. (b) $\mathrm{E}[f(X)] = \mathrm{E}[Q]$. If we only have the QoS to MOS mapping function, then we may apply the mapping to the QoS values X. Assuming that this MOS mapping function provides the average user rating of the users in the system for any condition x, then this equality holds. Hence, the service provider relies on the MOS mapping function to properly include the relevant parameters and context for the users in its system. This is typically a simplification in practice, but gives the service provider the possibility to evaluate the system in a more QoE-centric way.

QoE and MOS distributions inequality. Although the expected QoE in the system is equal to the expected MOS, this *does not imply* that the QoE and MOS distributions are necessarily equal, i.e., $\mathrm{E}[Q] = \mathrm{E}[M]$ does not imply $Q \stackrel{d}{=} M$.

This can be seen from the simple fact that for a discrete rating scale, Q is a discrete random variable, while M is a continuous random variable. Please note that we do not need any assumptions on the user rating distribution $Q|_x$, response time distribution X or the MOS mapping function $f(x)$ for this fundamental relationship $\mathrm{E}[Q] = \mathrm{E}[f(X)]$.

In practice, it is tempting to measure the expected response time $\mathrm{E}[X]$ and then to apply the MOS mapping function f to get the expected MOS (i.e. the expected user rating). However, the relation between the system parameter (e.g.m response time) and MOS is in general a non-linear function, which implies that

$$\mathrm{E}[Q] = \mathrm{E}[f(X)] \neq f(\mathrm{E}[X]) \ . \tag{8}$$

In general, $E[f(X)] \neq f(E[X])$, except when f is a linear transformation. E.g., if $f(x) = x^2$ then $f(E[X]) = (E[X])^2$ is not the same as $E[f(X)] = E[X^2]$. Only for a deterministic distribution with a constant value, the variance is $Var[X] = E[X^2] - E[X]^2 = 0$ and $E[X^2] = E[X]^2$.

2.1.2 GoB Ratio

The probability that the QoE Q in the system is rated good or better is denoted as $GoB[Q] = P(Q \geq \eta) = \sum_{i=\eta}^{n} q(i)$ where η is set accordingly to reflect good or better. Commonly, a value of $\eta = \frac{3}{4}n$ is chosen on a rating scale with minimum value of 0 and maximum value of n [3]. For example, on a 5-point ACR scale, 4 indicates a value of good and $GoB[Q] = P(Q \geq 4)$.

> **GoB in the system.** The GoB ratio in the system is equal to the expected GoB, $GoB[Q] = E[g(X)]$.

For deriving the QoE metric GoB for the QoE distribution Q over all users in the system and stochastic response times, it is necessary to provide a continuous GoB mapping function $g(x) = GoB[Q|_x]$ for the response time x.

$$GoB[Q] = P(Q \geq \eta) = \sum_{i=\eta}^{n} \int_{x=0}^{\infty} q(i|x)h(x)dx = \int_{x=0}^{\infty} g(x)h(x)dx = E[g(X)]$$

(9)

In general, it is not possible to derive the GoB from the MOS distribution M, although approximations may exist under certain assumptions and conditions, see [18]. The fundamental GoB relationship means that the system parameter X may be measured and can then be mapped to $g(X)$ to derive $GoB[Q]$.

2.1.3 Higher Moments and User Distribution

In a similar way, we may derive those fundamental results for variances or distributions, see [8]. To derive the complete distribution of the random variable Q, it is necessary to have the distribution $q(i|x) = P(Q = i|X = x)$.

$$q(i) = P(Q = i) = \int_{x=0}^{\infty} P(Q = i|X = x)h(x)dx$$
$$= \int_{x=0}^{\infty} P(Q|_x = i)h(x)dx , \quad \forall i = 0, \ldots, n \qquad (10)$$

Please note that we included an example for deriving the QoE user rating distribution Q from QoS measurements in [8]. In that example, we consider a single web server offering a certain service. Previous subjective studies have shown that for a given waiting time x, the conditional distribution $Q|x$ can be approximated with a binomial distribution. The fundamental relationship

allows to obtain then the distribution Q with the equation above. Based on this example, we show in Section 2.2 how to derive the GoB mapping function $g(x)$ to analyze the GoB ratio in the system.

For the k'th order moments of the r.v. Q:

$$E[Q^k] = \int_{x=0}^{\infty} \sum_{i=1}^{n} i^k q(i|x) h(x) dx = \int_{x=0}^{\infty} \sum_{i=1}^{n} f_k(x) h(x) dx = E[f_k(X)] \quad (11)$$

where $f_k(x) = \sum_{i=1}^{n} i^k q(i|x) = E[(Q|_x)^k]$

2.2 Example: Web QoE Dimensioning

In the following example, we use the GoB in the system to dimension a web server. The goal of the dimensioning is to determine the required service rate of the server such that a certain target GoB ratio is ensured for a given system request rate. We consider a single web server offering users a certain service like access to a web site. We chose this simple use case, since there exist web QoE models that may be utilized for this example, and the system itself can be be modeled as a queueing system, for which analytical results are well known. The response time of the system, i.e. the web page load time, is the single parameter X influencing the QoE in this example.

The system is modeled as an M/M/1-FCFS queueing system. User requests arrive according to a Poisson process with rate λ. The server has a single processing unit which serves request in a first-come-first-serve (FCFS) manner with service rate μ. If the server is occupied, arriving requests need to wait until they are served. An unlimited waiting room for the incoming requests is assumed. With the request interarrival times and the service times following an exponential distribution, with intensities λ and μ, respectively, this is a classical M/M/1-FCFS waiting queue with the well known response time distribution $X \sim Expo(\mu - \lambda)$, and the probability density function (PDF):

$$h(x) = (\mu - \lambda)e^{-(\mu-\lambda)x} . \quad (12)$$

under the assumption $\mu > \lambda$ (where the system is said to be stable).

In [3], it is shown that the opinion scores (for a given system parameter x), $Q|_x$ for a web QoE study can be very well approximated with a binomial distribution for various system response times x. Thereby, the user ratings are shifted to a scale from 0 to 4 (instead of the typical 1 to 5 scale). In that subjective study, a set of different page load times x was rated by 72 subjects. Thus, for any response time x, we may approximate the distribution of $Q|_x$ with a Binominal distribution, $Q|_x \sim Bino(n, p_x)$ with MOS $E[Q|_x] = np_x$ and $n = 4$ on the discrete rating scale $(0; 4)$. due to the used rating scale. The parameter p_x follows as $p_x = E[Q|_x]/n$. The MOS for a certain response time follows the IQX hypothesis, as shown in [19], which finally leads to $Q|_x \sim Bino(n, e^{-\beta x})$, with probability mass function (PMF)

$$q(i|x) = P(Q|_x = i) = \binom{n}{i} e^{-i\beta x}(1 - e^{-\beta x})^{n-i} \quad \text{for } i = 0, 1, 2, 3, 4 \quad (13)$$

with the sensitivity parameter $\beta = 0.25$ of the IQX hypothesis for this distri-
bution, see [8]. The MOS mapping function is $f(x) = E[Q|_x] = \sum_{i=0}^{4} iq(i|x)$
which is $f(x) = ne^{-\beta x}$.

Considering the scale from $0, \ldots, 4$, and with $\eta = 3$, the GoB mapping
function from Eq.(9) is

$$g(x) = \sum_{i=\eta}^{n} q(i|x) = q(3|x) + q(4|x)$$
$$= e^{-4\beta x} + 4e^{-3\beta x}(1 - e^{-\beta x}) = 4e^{-3\beta x} - 3e^{-4\beta x}. \tag{14}$$

The GoB of all users in the M/M/1 system example can be derived by com-
puting $E[g(X)] = GoB[Q]$. Combining the response time distribution $h(x)$,
Eq.(12), with the GoB mapping function $g(x)$, Eq.(14), then we can in our
example derive $g(X)$ and hence $E[g(X)]$ both analytically and numerically:

$$G_\beta(\lambda, \mu) = E[g(X)] = \int_{x=0}^{\infty} g(x)h(x)dx$$
$$= (\mu - \lambda)\left(\frac{4}{3\beta - \lambda + \mu} - \frac{3}{4\beta - \lambda + \mu}\right) \tag{15}$$

where λ is the arrival rate, μ the service rate, and $\beta = 0.25$ the fitted IQX
sensitivity parameter.

Now, we want to dimension the web server system in such a way that the
GoB ratio is larger than a target value, e.g. $G_\beta(\lambda, \mu) \geq G^*$ where $G^* = 90\%$
in this example. This requirement can either be expressed by the service rate
μ as a function of the arrival rate λ and G^*, or by λ as a function of μ and
G^*. In this example we use the required service rate $\mu(G^*)$ as a function of
the arrival rate and the target GoB ratio G^*. Figure 3 provides the numerical
results and allows the service provider to determine the required service rate
depending on the actual arrival rate in the system. Hence, the service provider
needs to estimate the arrival rate to finally decide on the required service rate,
according to the following equation for a target GoB ratio of $0 < G^* < 1$:

$$\mu(G^*) \geq \lambda + \frac{7\beta}{2}\left(\sqrt{\frac{1 - G^*/49}{1 - G^*}} - 1\right) > \lambda \tag{16}$$

Please note that arrival rate λ and the sensitivity parameter β have the unit
$1/s$, while the target GoB ratio G^* is dimensionless.

3 Extension of the Fundamental Relations to Multiple Parameters

The web QoE example in the previous section considered the response time as
single parameter influencing the QoE. Thus, X is a one-dimensional random
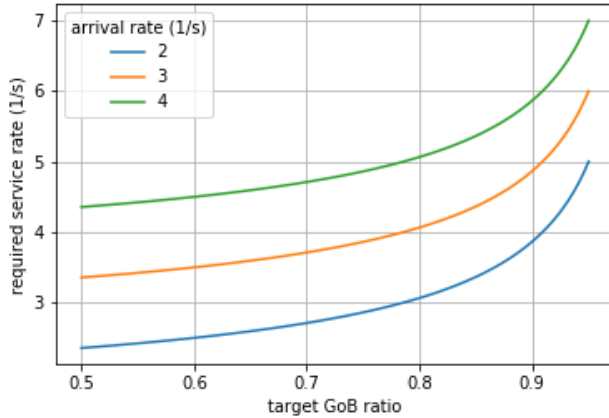
Fig. 3 Dimensioning of the required service rate for the web QoE example (Sec.2.2) such that the GoB is above a certain target GoB ratio G^* for a given arrival rate (Eq.(16)). In practice, the provider needs to fix G^* and estimate the arrival rate, e.g. overestimating.

variable which needs to be measured in the system. Then, the one-dimensional QoE mapping function can be applied to derive QoE metrics in the system.

For other applications and services, several parameters influencing QoE must be considered. As a result, several random variables need to be observed in the system. Then, a multidimensional QoE mapping function may be applied to quantify the QoE in the system. As an example, we consider non-adaptive HTTP video streaming in Section 3.3. The QoE is mainly determined by the number of stalls and the total stall duration [20] which are not independent of each other. In practice, those two parameters need to be measured in the system to derive the joint probability density function for both parameters. This use case demonstrates how to conduct a QoE evaluation of a running service in a system. In Section 3.1, we derive fundamental relationships for the QoE in the system for multiple parameter. The underlying multidimensional QoE model may have certain characteristics, e.g. an additive or multiplicative QoE model, which will allow to simplify the fundamental relationships (Section 3.2). The video streaming example in Section 3.3 will also show how to utilize the relationships in real-world system when measuring the system performance.

We want to highlight that we still consider in this section that all sessions are comparable. This allows to quantify the QoE for the offered service in the system across all users.

3.1 Fundamental Multidimensional QoE Relationships

We follow the same line of thinking as sketched in Figure 2. The system performance distribution is described by the random variables $X_1, X_2, \dots$ which may

be dependent on each other. Hence, we have a multivariate random variable which is a list of random variables with the corresponding joint probability density function $h(x_1, x_2, ...)$.

$$\mathbf{X} = (X_1, X_2, \dots) \tag{17}$$

In subjective studies, the conditional probabilities $Q|\mathbf{x}$ are obtained for the test condition $\mathbf{x} = (x_1, x_2, \dots)$. From the subjective studies, relevant mapping functions are derived, e.g., a MOS mapping function $f : \mathbf{x} \to [1; 5]$. In practice, it is a difficult problem to sample the parameter space $\mathbf{x}$ properly to derive appropriate models such as the MOS mapping function. A large set of parameters (reflected in the dimensionality of the parameter space of $\mathbf{x}$) will require too high costs for conducting subjective experiments which would cover a representative combination of system parameters in $\mathbf{x}$. If one or more of the system parameters are continuous and not discrete, this will increase this challenge even further. This sampling problem is interesting and investigated in recent literature. As a starting point, the interested reader may review for example [21, 22, 23, 24] developing active learning algorithms for multidimensional QoE models.

3.1.1 Expected QoE

For the sake of simplicity, we consider in the following only two parameters, e.g., X_1 (number of stalls) and X_2 (stall duration), to derive the fundamental multidimensional QoE relationships. The same thoughts can be generalized to an arbitrary number of possibly dependent parameters. We further assume a discrete rating scale yielding a discrete random variable of user ratings, i.e., $Q \in \{0, \dots, n\}$. Again, the derivation can be generalized to continuous rating scales analogously.

The expected QoE in the system requires the joint two-dimensional PDF $h(x_1, x_2)$ of the two parameters $\mathbf{X} = (X_1, X_2)$ as well as the conditional QoE distribution $Q|_x$ with the corresponding probability mass function $q(i|\mathbf{x}) = q(i|x_1, x_2) = P(Q|_\mathbf{x} = i) = P(Q = i|X_1 = x_1, X_2 = x_2) = (Q = i|\mathbf{x})$. Then, we obtain the following fundamental relationship for the expected QoE $\mathrm{E}[Q]$ in the system.

$$\begin{aligned}
\mathrm{E}[Q] &= \sum_{i=0}^{n} i \cdot q(i) = \sum_i i \int_{x_1} \int_{x_2} q(i|x_1, x_2) \cdot h(x_1, x_2) dx_1 dx_2 \\
&= \int_{x_1} \int_{x_2} h(x_1, x_2) \sum_i i q(i|x_1, x_2) dx_1 dx_2 \\
&= \int_{x_1} \int_{x_2} h(x_1, x_2) \mathrm{E}[Q|x_1, x_2] \, dx_1 dx_2 \\
&= \int_{x_1} \int_{x_2} h(x_1, x_2) f(x_1, x_2) dx_1 dx_2 = \int_{\mathbf{x}} h(\mathbf{x}) f\mathbf{x}) d\mathbf{x} \\
&= \mathrm{E}[f(X_1, X_2)] = \mathrm{E}[f(\mathbf{X})]
\end{aligned} \tag{18}$$

Similar to the one-dimensional relationship, a MOS mapping function f is required which maps the system performance $\mathbf{X}$ to a distribution of MOS values $M = f(\mathbf{X})$. Note that the relation given in Eq.(18) generally holds for k-dimensional parameter sets (in Section 2, we consider $k = 1$-dimensional, while in this section we consider $k = 2$-dimensional sets). The k-dimensional MOS mapping function is $f(\mathbf{x}) = \mathrm{E}[Q|_x]$ and the expected QoE in the system is then the expected MOS.

> **Expected QoE in a system with k-dimensional parameter set X.**
> The expected QoE is equal to the expected MOS, $\mathrm{E}[Q] = \mathrm{E}[f(\mathbf{X})]$.

3.1.2 GoB Ratio

For other QoE metrics, similar derivations as in the one-dimensional case lead to the fundamental relationships. We show the example of the GoB ratio in the following.

$$GoB[Q] = P(Q \geq \alpha) = \sum_{i=\eta}^{n} P(Q = i) \tag{19}$$

$$= \sum_{i=\eta}^{n} \int_{x_1} \int_{x_2} q(i|x_1, x_2) h(x_1, x_2) dx_1 dx_2$$

$$= \int_{x_1} \int_{x_2} h(x_1, x_2) \sum_{i=\eta}^{n} q(i|x_1, x_2) dx_1 dx_2$$

$$= \int_{x_1} \int_{x_2} h(x_1, x_2) g(x_1, x_2) dx_1 dx_2 = \int_{\mathbf{x}} h(\mathbf{x}) g(\mathbf{x}) d\mathbf{x}$$

$$= \mathrm{E}[g(\mathbf{X})] \tag{20}$$

Hence, we need the multidimensional GoB mapping function g and the multivariate PDF h for calculating the GoB over all users in the system. It is $g(x_1, x_2) = GoB[Q|_x]$.

> **GoB in the system with k-dimensional parameter set X.** The GoB
> in the system is equal to the expected GoB $g(\mathbf{X})$, i.e. system performance
> mapped to GoB, $GoB[Q] = \mathrm{E}[g(\mathbf{X})]$.

3.1.3 QoE Distribution

For the derivation of the complete distribution of the random variable Q, it is necessary to have the distribution $q(i|x_1, x_2)$.

$$q(i) = P(Q = i) = \int_{x_1} \int_{x_2} q(i|x_1, x_2) h(x_1, x_2) dx_1 dx_2 \quad \forall i = 0, \ldots, n \tag{21}$$

3.2 Multiplicative and Additive QoE Models

The fundamental relationships indicate that the joint PDF h is required to compute the metrics for the QoE in the system. Now, let us consider that the two random variables X_1 and X_2 are independent of each other. Then, the joint probability density function $h(x_1, x_2)$ is the product of the marginal PDFs $h_1(x_1)$ and $h_2(x_2)$.

$$h(x_1, x_2) = h_1(x_1) \cdot h_2(x_2) \tag{22}$$

Another simplification considers the underlying QoE model. The literature often provides additive and multiplicative QoE models, see [25] for detailed discussions. To this end, we focus on the expected QoE in the system and corresponding MOS mapping functions in the following for additive and multiplicative MOS models.

3.2.1 Additive MOS Model and Expected QoE in the System

The literature suggests additive QoE models for different services like mobile web browsing [26] or speech quality. The E-model [27] is a commonly used parametric planning model for predicting expected speech quality. The underlying principle for handling multiple different types of impairments came from the OPINE model proposed by NTT, assuming quality degradation factors are summed on a psychological scale [28]. For video quality estimates, the amendment of ITU-T Rec. P.1201 [29] suggests the use of an additive model whereby degradations resulting from stalling and initial delay are subtracted from a maximum MOS value.

Let us consider an additive QoE model which maps the parameters x_1 and x_2 to corresponding MOS values. In general, we may have m different mapping functions f_j^*, $(j = 1, \cdots, m)$ which are defined on a k-dimensional parameter space $\mathbf{x}$, where m and k might be different. For instance, in the combined additive model we weights the two terms with corresponding factors a_1 and a_2. There are no assumptions on the functions f_1 and f_2, which may be e.g. non-linear functions. We arrive at the following (non-linear) two-dimensional MOS mapping function f.

$$f(\mathbf{x}) = a_1 f_1^*(\mathbf{x}) + a_2 f_2^*(\mathbf{x}) \tag{23}$$

In the following we consider two mapping functions, f_1 and f_2, and two parameters x_1 and x_2, and let $f_i^*(\mathbf{x}) = f_i(x_i)$, $i = 1, 2$, i.e., one function f_i per parameter x_i. Hence,

$$f(x_1, x_2) = a_1 f_1(x_1) + a_2 f_2(x_2) \tag{24}$$

The expected QoE in the system can be derived based on the fundamental relation $E[Q] = E[f(X)]$. For such additive QoE models, we obtain then

$$
\begin{aligned}
E[Q] &= \int_{x_1} \int_{x_2} (a_1 f_1(x_1) + a_2 f_2(x_2)) \cdot h(x_1, x_2) dx_1 dx_2 \\
&= \int_{x_1} \int_{x_2} a_1 f_1(x_1) h(x_1, x_2) dx_1 dx_2 + \int_{x_1} \int_{x_2} a_2 f_2(x_2) h(x_1, x_2) dx_1 dx_2 \\
&= a_1 E[f_1^*(\mathbf{X})] + a_2 E[f_2^*(\mathbf{X})]
\end{aligned}
\tag{25}
$$

On a k-dimensional parameter set $\mathbf{X}$, the $E[Q] = \sum_{d=1}^{k} a_d f_d^*(\mathbf{X})$.

We may use the marginal distributions to further simplify the summands. Let us consider the marginal distribution $h_1(x_1) = \int_{x_2} h(x_1, x_2) dx_2$. The conditional density is defined as $h(x_2|x_1) = h(x_1, x_2)/h_1(x_1)$. In other words, the joint density of both random variables is the product between the conditional density and the marginal density.[1]

$$
\begin{aligned}
E[f_1^*(\mathbf{X})] &= \int_{x_1} \int_{x_2} f_1(x_1) h(x_1, x_2) dx_1 dx_2 \\
&= \int_{x_1} \int_{x_2} f_1(x_1) h(x_2|x_1) h_1(x_1) dx_1 dx_2 \\
&= \int_{x_1} f_1(x_1) h_1(x_1) \int_{x_2} h(x_2|x_1) dx_2 dx_1 \\
&= \int_{x_1} f_1(x_1) h_1(x_1) dx_1 = E[f_1(X_1)]
\end{aligned}
\tag{26}
$$

Thus, the expected QoE in the system can be described with the marginal density functions.

$$
E[Q] = a_1 E[f_1(X_1)] + a_2 E[f_2(X_2)]
\tag{27}
$$

On a k-dimensional parameter set $\mathbf{X}$, the $E[Q] = \sum_{d=1}^{k} a_d f_d(X_d)$.

In practice, the usage of the marginal distributions may be required, when the complete information is not available. The complete information could be provided as a tuple of measurement values for each session or as a multi-dimensional histogram. Due to the large amount of measurement data, in practice the marginal distributions may be only captured. Dash boards for example may only provide the aggregated histogram of measurement values for the individual parameters.

3.2.2 Multiplicative QoE Model

For multiplicative QoE models, we obtain similar results as in the additive case, except the simplified formulation with marginal distributions is not possible as we will see later. Multiplicative models have been demonstrated for audiovisual

---

[1] This relation is well known regarding conditional probabilities.

quality whereby the multiplicative term between audio and video qualities is generally sufficient to estimate audiovisual quality [30]. This was confirmed in a survey comparing integration models [31].

The video streaming example in Section 3.3 relies also on such a multiplicative MOS model. In particular, the multidimensional IQX model [21] relies on a multiplicative model. The parameters x_1 and x_2 are contributing to the MOS according to the single dimensional IQX hypothesis. Thus, $f_1(x_1) = a_1 e^{-\beta_1 x_1}$ and $f_2(x_2) = a_2 e^{-\beta_2 x_2}$. In addition, the interaction between the two parameters is considered $f_{12}(x_1, x_2) = a_{12} e^{-\beta_{12} x_1 x_2}$. Then, the MIQX model suggests $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \cdot f_{12}(x_1, x_2) = a e^{-(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2)} = a e^{-\boldsymbol{\beta}\mathbf{x}}$. Hence, the MIQX is an extension of the IQX to a vector of parameters $\mathbf{x}$ and the sensitivity parameters $\boldsymbol{\beta}$ with $\boldsymbol{\beta}\mathbf{x}$ as a linear model with interaction. The result of the MIQX is a multiplicative model.

We consider the multiplicative model for the MOS and two potentially dependent parameters X_1, X_2 in the following. Here, the MOS mapping function takes into account an additive offset b and a discrete rating scale.

$$f(x_1, x_2) = a \cdot f_1(x_1) \cdot f_2(x_2) + b \tag{28}$$

The expected MOS is then the system can be derived with the fundamental relationship and the joint PDF $h(x_1, x_2)$.

$$
\begin{aligned}
\mathrm{E}[Q] &= \sum_{i=0}^{n} i q(i) = \sum_{i=0}^{n} i \int_{x_1} \int_{x_2} q(i|x_1, x_2) h(x_1, x_2) dx_1 dx_2 \\
&= \int_{x_1} \int_{x_2} h(x_1, x_2) \sum_{i=0}^{n} i q(i|x_1, x_2) dx_1 dx_2 \\
&= \int_{x_1} \int_{x_2} h(x_1, x_2) f(x_1, x_2) dx_1 dx_2 \\
&= a \int_{x_1} \int_{x_2} f_1(x_1) f_2(x_2) h(x_1, x_2) dx_1 dx_2 + b \\
&= a \int_{x_1} f_1(x_1) h_1(x_1) \int_{x_2} f_2(x_2) h(x_2|x_1) dx_2 dx_1 + b \tag{29}
\end{aligned}
$$

However, this equation cannot be simplified anymore as in the additive case, since the latter integral leads to a function ϕ_2 of x_1. Thus, we have the function $\phi_2(x_1) = \int_{x_2} f_2(x_2) h(x_2|x_1) dx_2$ based on f_2 and therefore

$$\mathrm{E}[Q] = a \int_{x_1} f_1(x_1) h_1(x_1) \phi(x_1) dx_1 + b \,. \tag{30}$$

In the case that the parameters X_1 and X_2 are independent, Eq.(29) can be simplified, since $h(x_1, x_2) = h_1(x_1) h_2(x_2)$ or $h(x_2|x_1) = h_2(x_2)$. Then, $\phi_2(x_1) = \mathrm{E}[X_2]$.

$$
\begin{aligned}
\mathrm{E}[Q] &= a \int_{x_1} f_1(x_1) h_1(x_1) \int_{x_2} f_2(x_2) h_2(x_2) dx_2 dx_1 + b \\
&= a \cdot \mathrm{E}[f_1(X_1)] \cdot \mathrm{E}[f_2(X_2)] + b \tag{31}
\end{aligned}
$$

3.3 Example: HTTP Video Streaming QoE

This example shows how to quantify the expected QoE for an HTTP video streaming service in a system for two-dimensional QoS measures. This example also depicts how a service provider may monitor the QoS in the system and then apply the two-dimensional fundamental relationship in order to estimate the expected QoE in the system.

The service provider measures the stalling pattern on application layer, which is the total stall duration t and the number n of stalls, i.e., the system parameter set is $\mathbf{x} = (x_1 = n, x_2 = t)$. Previous work has already shown that stalling is a major QoE influence factor [20]. In particular, for (non-adaptive) HTTP streaming[2], a two-dimensional QoE mapping function $f(\mathbf{x}) = f(n, t)$ is provided which maps the number of stalls and the total stall duration to MOS values [20].

$$f(n,t) = 3.5 \cdot e^{-0.15t - 0.19n} + 1.5 \tag{32}$$

which implies that $f : \mathbf{x} \rightarrow [1.5; 5]$

For obtaining the stalling patterns, we conduct a simulation and we compare the measurement results for the expected QoE in the system with theoretical results from the known simulation parameters. The observed number of stalls N in the system is a random variable which is assumed to follow a geometric distribution: $N \sim Geom(p)$ with $P(N = k) = (1 - p)^k \cdot p$ for $k = 0, 1, 2, \ldots$. The total stall duration follows an Erlang distribution which is composed of N exponential phases of average length L. Thereby, L indicates the average stall duration for each of the N stall events. Thus, $T \sim Erlang(N, L)$. We select the parameters of the stalling simulation as follows: $E[N] = 2.5, E[L] = 2$ s. This results into $E[T] = E[N] \cdot E[L] = 5$ s.

For each simulated video j, we sample the set $\mathbf{x}_j = \{n_j, t_j\}$, where n_j is the number of stalls for video j, and t_j is the total stalling duration for the same video. We simulated $M = 100,000$ replications, i.e., M video sessions. We estimated the number of stalls of $\bar{N} = 2.47$ over all videos, the total stall duration $\bar{T} = 4.91$ s. The probability of no stalling, $P(N = 0) = 0.2866$. The total stall duration and the number of stalls are correlated and the coefficient of correlation is $corr(N, T) = 0.88$.

We are interested in the average QoE of the video service (which is the expected MOS) in the system, with the random variables N and T which are mapped to the random variable $Q = f(N, T)$ by Eq. (32). Thus, we need to compute $E[Q] = E[f(N, T)]$ which requires the joint PDF $h(\mathbf{X}) = h(n, t)$. In practice, however, we simply use the measurement values $\mathbf{x} = \{n_j, t_j\}$ and obtain q_j from $q_j = f(\mathbf{x}_i) = f(n_j, t_j)$, which is the corresponding MOS value for the j-th video. Please remember: With the MOS mapping function, we can only derive the expected system QoE (but not system GoB etc.).

---

[2] There are more advanced HTTP streaming models available which also consider adaptive streaming like [32]. For the sake of simplicity and for illustrating the application of the fundamental multidimensional relationships, the simpler non-adaptive HTTP video streaming case is considered.

The average QoE of the video service obtained from the measurements is therefore

$$\bar{Q} = \frac{1}{M} \sum_{j=1}^{M} q_j \, . \tag{33}$$

For the QoE of the system, it is also recommended to compute the confidence intervals based on the measurements. In this example, the average QoE is $E[Q] = 3.380$ with the 95% confidence interval $(3.299, 3.460)$. As already discussed previously, it is not possible to compute the average QoE by mapping the average number of stalls and the average total duration to MOS, unless the mapping is a linear function. In our example, we estimated $\bar{Q} = 3.38$, which is not equal to $f(\bar{N}, \bar{T}) = 2.55$.

The theoretical expected QoE in the system can be derived for our scenario with the joint PDF $h(n, t)$ for the MOS mapping function with parameters $\alpha, \beta_t, \beta_n, \gamma$. It is $f(n, t) = \alpha e^{-\beta_t \cdot t - \beta_n \cdot n} + \gamma$. The joint PDF $h(n, t)$ can be expressed with the conditional density $a_n(t)$ which is the total stall duration under the condition of n stalls. Thereby, the number of stalls follows the geometric distribution with parameter p and $a_n(t)$ follows the Erlang distribution with n phases of average length $L = 1/\lambda = 2\,\mathrm{s}$. Figure 4 shows the results for a parameter study on the average number of stalls in the system according to the following equation.

$$E[Q] = \sum_{n=0}^{\infty} \int_{t=0}^{\infty} f(n,t)h(n,t)dt \quad = \frac{e^{\beta_n}(\beta_t + \lambda)(\alpha p + \gamma) + \gamma\lambda(p-1)}{e^{\beta_n}(\beta_t + \lambda) + \lambda(p-1)} \tag{34}$$

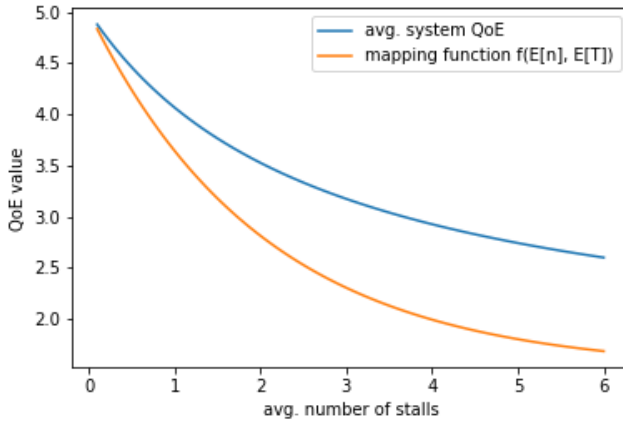**Fig. 4** Parameter study for the HTTP video streaming example. Again, $f(\mathrm{E}[N], \mathrm{E}[T])$ cannot be used to derive the average system QoE, but $\mathrm{E}[f(N, T)]$ due to the non-linear QoS-MOS mapping function.

In practice, for monitoring the QoE in the system, it is required that the corresponding mapping function is used (e.g. MOS mapping function f as in the example above or GoB mapping function g). Then the input parameters of the mapping function need to be measured in the system. Then, the QoE in the system can be derived according to the fundamental relationships. Hence, the average of the measurement values mapped to QoE needs to be calculated, as in Eq.(33).

4 A QoE-based Approach to Service-Level Quality

In the calculations and discussions given in Sections 2 and 3, the QoE in the system is obtained using estimated per-user QoE values, e.g., by applying MOS or GoB mapping functions to a given set of system parameters. Therefore, the QoE of the individual users are all taken into account in the same way and weighted equally. However, a service provider (or system operator) may aim at utilizing the estimated QoE values of the users for various purposes (e.g., network/server dimensioning, quality monitoring, system benchmarking). In this context, a user interacts with the service within a certain session. Those sessions may be very different in terms of session duration, resource demands, or emerging costs for the proivder. For some services, such as multiparty video calls, several users may take part in the same session, but each user will have their own individual experience of that session.

It may be relevant to consider that certain sessions are more "important" than others for the provider, e.g., due to resource demands or costs for a session, or due to the number of affected users in a session. Thus, such sessions may be more *indicative* of an aggregate service quality measure, which is then in turn used by an operator as a proxy measure when performing tasks such as network/server dimensioning, quality monitoring, or system benchmarking. In this section, we first provide a generic definition of a **Service Quality Index** (SQI), and then provide two example use cases to demonstrate scenarios in which a service provider can utilize such an index.

4.1 Defining a QoE-based Service Index (SQI)

In Section 1.2, we defined the SQI as *a measure indicating the overall utility of a service delivered by a system and derived as a weighted combination of quality values estimated per user session.* Rather than considering only a QoE mapping function (which estimates QoE per session), we extend this notion to a *utility function*, which weights individual session QoE values in a manner that is deemed relevant by a service provider (e.g., based on session duration, resource consumption, etc.). For a session i, we define the utility function as:

$$u(\mathbf{x}_s) = m^*(\mathbf{x}_s) \cdot w^*(\mathbf{x}_s) \tag{35}$$

where $m^*(\mathbf{x}_s)$ is QoE mapping in session s on a subset of $\mathbf{x}$, and $w^*(\mathbf{x}_s) = w_s$ is the weight of session s. In a generic sense, the weight factor is assigned to

indicate the relative importance of a given session quality value. Hence, we may use normalized weights for the utility function, e.g., when having S sessions, $w_s^* = w_s / \sum_{s=1}^{S} w_s$. Please note that x_s may also include other aspects than QoS parameters like the duration of sessions or the resource consumption for that session. Then the QoE mapping function may only consider a subset of those parameters. Similarly, the weighting function may also consider a subset of the parameters.

We then define the SQI as the expected session utility values which requires the joint PDF $h(\mathbf{x})$ of the parameters $\mathbf{X}$ (RV) which are the input values for the QoE mapping function m^* and the weighting function w^*. Then, we have

$$SQI = \int_{\mathbf{x}} u(\mathbf{x})h(\mathbf{x})d\mathbf{x} = \int_{\mathbf{x}} m^*(\mathbf{x})w^*(\mathbf{x})h(\mathbf{x})d\mathbf{x}\,. \qquad (36)$$

> **Service Quality Index (SQI).** The SQI value in a system is equal to the expected utility values, $SQI = \mathrm{E}[u(\mathbf{X})]$.

If all sessions have the same weight, $w(\mathbf{x}) = 1$, and we use the MOS mapping function f, then $SQI = \mathrm{E}[f(\mathbf{X})] = \mathrm{E}[Q]$ and SQI represents the expected QoE in the system. If we use the GoB mapping function g instead, then $SQI = \mathrm{E}[g(\mathbf{X})] = GoB[Q]$.

In the following sections, we illustrate two example use cases with different weights illustrating how the SQI may be utilized in a meaningful way by service providers.

4.2 Example Use Case: Utilizing SQI for Web QoE Dimensioning

As an example, we once again consider a web service provider aiming to dimension a web server (as explained previously in Section 2.2). For each loaded page, we have a waiting time t_w and a service time t_s to process the user request at the web server. The page load time is $t = t_w + t_s$ and the MOS is from Eq (31)

$$f(t_w, t_s) = a f_1(t_w) f_2(t_s) = 4e^{-\beta(t_w + t_s)}$$

where $a = 4$, $f_1^*(\mathbf{x}) = f_1(t_w) = e^{-\beta t_w}$ and $f_2^*(\mathbf{x}) = f_2(t_s) = e^{-\beta t_s}$. Rather than relying simply on calculating expected QoE in the system, the provider may assign weights to individual MOS values that are proportional to the service times. One example for this time-dependent consideration is the processing costs. Le us consider ad impressions during service consumption. A service provider may have contracts with ad companies which pay per impression time. While a user is served, ads are displayed and the provider may get revenue from the ads company [33].

Thus, we consider the weight of a user request to correspond to $w^*(t_w, t_s) = w(t_s) = t_s$, i.e., only the time it takes to process the user request, i.e. to download actual contents. The joint PDF is $h(t_w, t_s)$. In the FIFO system, the waiting time and the service time are independent (different than in a

processor sharing system). Hence, we simply multiply the PDFs of the waiting time and the service time $h(t_w, t_s) = h_w(t_w) \cdot h_s(t_s)$. The PDF of the service time is $h_s(t_s) = \mu e^{-\mu \cdot t_s}$. The PDF of the waiting time is $h_w(t_w) = \rho (\mu - \lambda) e^{-(\mu - \lambda)t_w}$. Putting the different parts together, we can derive the SQI based on the random variables T_w and T_s reflecting the waiting time and the service time.

$$\begin{aligned} SQI = \mathrm{E}[u(T_w, T_s)] &= \int_{t_w, t_s} w^*(t_w, t_s) \cdot f(t_w, t_s) \cdot h(t_w, t_s) dt_w dt_s \\ &= \frac{4\lambda (\mu - \lambda)}{(\beta + \mu)^2 (\beta - \lambda + \mu)} \end{aligned} \tag{37}$$

For dimensioning the service rate μ (how fast user requests can be processed), the service provider first needs to estimate the service demand rate λ, i.e. the number of user requests per time. Figure 5 shows the SQI as well as the expected QoE in the system in relation to the server rate μ for $\lambda = 3$. The system is stable, i.e., not in overload, when $\mu > \lambda$. As expected, the higher the server rate is, the higher the expected QoE in the system is. However, taking into account the session weights, we have a different picture. Low server rates μ close to $\lambda = 1$ lead to highly loaded systems with bad QoE and low SQI values. However, high server rates lead to high QoE values, but shorter service times. In that case, the provider may lose revenue from the ads company. As a result, the SQI value is also low. In Figure 5, the SQI curve indicates the **optimal operating point wrt. the service rate**, which is considered best for the operator according to the SQI.
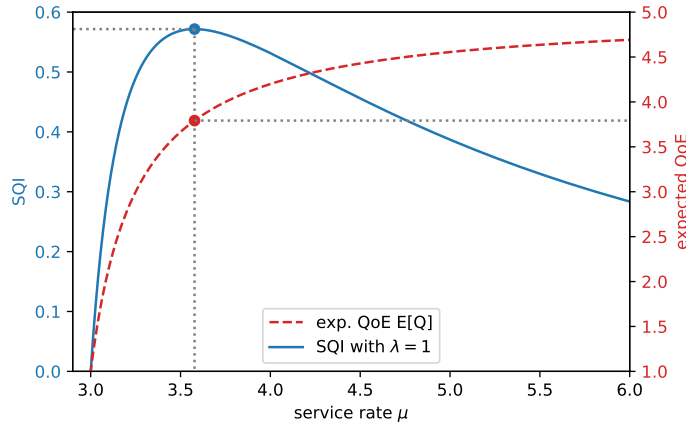

Fig. 5 Utilizing SQI for the web QoE example in Section 2.1 for $\lambda = 3$. We use the service time as weighting function $w(t_w, t_s) = t_s$ and use the MOS mapping function $f(t_w, t_s)$. The optimal server rate is $\mu = 3.58$ leading to $SQI = 0.57$ and $\mathrm{E}[Q] = 3.80$. The QoE values are shifted to a 5-point scale from $[1; 5]$.

4.3 Example Use Case: Utilizing SQI for Benchmarking a System Offering
Multiparty Calls

For complex applications, such as WebRTC conferences, where many users
can participate in a session, using several media modalities at once, we face
some additional complexities in defining the "quality of a session", since unlike
the case of video streaming or web browsing, each session itself (a "call")
potentially involves: a) many users, and b) several QoE models (e.g., for voice
and video) that need to be combined. As is the case with e.g., video streaming,
calls can have widely varying durations, which in and of themselves also affect
the QoE of the participating users (who in addition can join or drop at different
times during the call).

In such a context, if from the service level we want to see the "call" as
our session unit, there needs to be some quality aggregation done already
at the session level (since anyway we are considering the session quality as
a function of the QoE of the participating users). In this case, we can shift
the inter-session variation into the session quality estimation itself (assuming
that different weights should be assigned to different participating user QoE
values), and then we can simply average session qualities over the service as
a whole. If, on the other hand, we wanted to consider the session granularity
at the user level, as opposed to the call level, we would need to include some
aspects of the call itself (e.g., duration of the user's participation, number of
users in the call) in the weighting function.

Let us assume that a service provider would like to apply the SQI frame-
work for the purpose of benchmarking their service. We use the following nota-
tion: S: total number of sessions; n_s: number of users in session s; $n = \sum_{s=1}^{S} n_s$:
total number of users across all sessions; $\mathbf{x}_{s,i}$: a set of objectively measured
parameters related to user i in session s; $u(\mathbf{x}_{s,i})$: the utility function for user i
in session s; $m(\mathbf{x}_{s,i})$: the mapped QoE value for user i in session s; and $w(\mathbf{x}_{s,i})$:
the weight for user i in session s. We then calculate the SQI as a weighted
average of all utility values summarized across all users and sessions:

$$SQI = \frac{1}{n} \sum_{s=1}^{S} \sum_{i=1}^{n_s} u(\mathbf{x}_{s,i}) = \frac{1}{n} \sum_{s=1}^{S} \sum_{i=1}^{n_s} m(\mathbf{x}_{s,i}) \cdot w(\mathbf{x}_{s,i}) \qquad (38)$$

The weight assigned to a given user QoE value can be derived based on the
relevance of a user *within a particular session* (as compared to other session
users, for example, if there is a so-called *dominant speaker*), as well as based
on the relevance *of the session itself* (as compared to other sessions), or some
combination thereof. By now considering each user QoE value individually, we
can denote the SQI value as the following:

$$SQI = \frac{1}{n} \sum_{i=1}^{n} u(\mathbf{x}_i) = \frac{1}{n} \sum_{i=1}^{n} m(\mathbf{x}_i) \cdot w(\mathbf{x}_i) \qquad (39)$$

where $\mathbf{X}_i$ is the set of objectively measured parameters for user i, $i = 1, ..., n$.
It thus once again follows that $SQI = \mathrm{E}[u(\mathbf{X})]$.

We note that if all weights are equal, then $w(\mathbf{x_j}) = 1$ for all users in the system. In such a case, if we use the MOS mapping function, then it follows that $SQI = \mathrm{E}[f(\mathbf{X})] = \mathrm{E}[Q]$. Analogously, if we use the GoB mapping function, it follows that $SQI = \mathrm{E}[g(\mathbf{X})] = GoB[Q]$.

5 Discussion and Conclusions

Service and network providers rely on QoE models (often in the form of QoS-to-MOS mapping functions) for estimating and / or predicting user perceived service quality in their systems. A common approach is to use the distribution of MOS scores in the system (as obtained from a QoS-to-MOS mapping function) to draw conclusions with respect to the QoE distribution (or other QoE metrics) of users in the system. These metrics are then further used to drive QoE optimization and management decisions [34, 35, 36]. Similarly, [37] analyzes MOS distributions, but states that "[...] the ultimate goal is to predict the distribution of user ratings". This will "[...] give operators and service providers a holistic view of service quality." Especially in 5G, a user-centric design is foreseen, requiring to consider system QoE [38].

In this paper, we draw the attention of the systems community to the fact that the actual QoE distribution in a system is not (necessarily) equal to the MOS distribution in the system. The current systems literature however, indicates that there is clearly lack of a common understanding as to *what are the implications of using MOS distributions rather than actual QoE distributions.* For example, it is not possible to derive the ratio of users experiencing good or better (GoB) quality in the system by utilizing the MOS mapping function to obtain the MOS distribution. Instead, a QoS-to-GoB mapping is required. We provide important insights to raise awareness and foster further research in this area; targeting also the QoE community, and once again highlight the need for reporting QoE metrics and mapping functions beyond just those relying on MOS (e.g., GoB).

The contribution of this paper are proven fundamental multi-dimensional QoE relationships providing an important link between the QoE community and the systems community. If researchers conducting subjective user studies provide different QoS-to-QoE mapping functions for QoE metrics of interest (e.g., MOS or GoB), this is enough to derive corresponding QoE metrics from a system's perspective. This holds for any QoS distribution in the system, as long as the corresponding QoS values are captured in the QoE models.

In addition, we propose a framework for network/service providers which provides guidelines for taking into account the *relative importance* of user sessions in a system. We define a QoE-based Service Quality Index (SQI) based on individual user QoE values, which are weighted according to factors that are deemed relevant by a service provider. Such factors may be related to session characteristics (e.g., session duration), resource consumption (e.g., costs), number of users involved in a session, etc. The index considers *utility*

as an extension of quality, and is consistent with the definition of QoE metrics in the system when all QoE values are weighted equally.

> ✓ Expected QoE in the system: $E[Q] = E[f(X)]$
> ✓ GoB QoE ratio in the system: $GoB[Q] = E[g(X)]$
> ✓ Service quality index: $SQI = E[u(X)]$

Acknowledgments

Conflict of Interests

On behalf of all authors, the corresponding author, Tobias Hoßfeld (University of Würzburg, Germany), states that there is no conflict of interest.

References

1. V. A. Zeithaml, "Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence," *The Journal of marketing*, vol. 52, no. 3, Jul. 1988.
2. L. Janowski and Z. Papir, "Modeling subjective tests of quality of experience with a generalized linear model," in *Quality of Multimedia Experience (QoMEx 2009)*. IEEE, 2009.
3. T. Hoßfeld, P. E. Heegaard, M. Varela, and S. Möller, "QoE beyond the MOS: an in-depth look at QoE via better metrics and their relation to MOS," *Quality and User Experience*, vol. 1, no. 1, 2016.
4. J. Xu, L. Xing, A. Perkis, and Y. Jiang, "On the properties of mean opinion scores for quality of experience management," in *2011 IEEE International Symposium on Multimedia*, 2011, pp. 500–505.
5. R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives," *Multimedia Systems*, vol. 22, no. 2, pp. 213–227, 2016.
6. S. Möller, "Models for predicting speech communication quality and service-related models," in *Assessment and Prediction of Speech Quality in Telecommunications*. Springer, 2000, pp. 89–114.
7. T. Hoßfeld, P. E. Heegaard, L. Skorin-Kapov, and M. Varela, "No Silver Bullet: QoE Metrics, QoE Fairness, and User Diversity in the Context of QoE Management," in *Quality of Multimedia Experience (QoMEX 2017)*. IEEE, 2017.
8. ——, "Fundamental relationships for deriving qoe in systems," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–6.
9. K. Kilkki, *An Introduction to Communications Ecosystems*. CreateSpace Independent Publishing Platform, 2012.
10. P. Le Callet, S. Möller, A. Perkis *et al.*, "Qualinet white paper on definitions of quality of experience," *European network on quality of experience in multimedia systems and services (COST Action IC 1003)*, vol. 3, no. 2012, 2012.
11. A. Sackl, R. Schatz, and A. Raake, "More than i ever wanted or just good enough? user expectations and subjective quality perception in the context of networked multimedia services," *Quality and User Experience*, vol. 2, no. 1, p. 3, 2017.

12. S. Shenker, "Fundamental design issues for the future internet," *IEEE Journal on selected areas in communications*, vol. 13, no. 7, pp. 1176–1188, 1995.
13. P. Reichl, B. Tuffin, and R. Schatz, "Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience," *Telecommunication Systems*, vol. 52, no. 2, pp. 587–600, 2013.
14. L. Skorin-Kapov, K. Ivesic, G. Aristomenopoulos, and S. Papavassiliou, "Approaches for utility-based qoe-driven optimization of network resource allocation for multimedia services," in *Data traffic monitoring and analysis*. Springer, 2013, pp. 337–358.
15. M. Fiedler, T. Hoßfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, 2010.
16. A. Ahmad, A. Floris, and L. Atzori, "Qoe-centric service delivery: A collaborative approach among otts and isps," *Computer Networks*, vol. 110, 2016.
17. M. Varela, P. Zwickl, P. Reichl, M. Xie, and H. Schulzrinne, "From Service Level Agreements (SLA) to Experience Level Agreements (ELA): The Challenges of Selling QoE to the User," in *2015 IEEE International Conference on Communication Workshop (ICCW)*. IEEE, 2015.
18. T. Hoßfeld, M. Fiedler, and J. Gustafsson, "Betas: Deriving quantiles from mos-qos relations of iqx models for qoe management," in *Integrated Network and Service Management (IM), 2017 IFIP/IEEE Symposium on*. IEEE, 2017, pp. 1011–1016.
19. T. Hoßfeld, F. Metzger, and D. Rossi, "Speed index: Relating the industrial standard for user perceived web performance to web QoE," in *Quality of Multimedia Experience (QoMEX 2018)*, May 2018, pp. 1–6.
20. T. Hoßfeld, R. Schatz, E. Biersack, and L. Plissonneau, "Internet video delivery in youtube: From traffic measurements to quality of experience," in *Data Traffic Monitoring and Analysis*. Springer, 2013.
21. H.-S. Chang, C.-F. Hsu, T. Hoßfeld, and K.-T. Chen, "Active learning for crowdsourced qoe modeling," *IEEE Transactions on Multimedia*, vol. 20, no. 12, pp. 3337–3352, 2018.
22. P. Ye and D. Doermann, "Active sampling for subjective image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
23. B. Osting, J. Xiong, Q. Xu, and Y. Yao, "Analysis of crowdsourced sampling strategies for hodgerank with sparse random graphs," *Applied and Computational Harmonic Analysis*, vol. 41, no. 2, 2016.
24. M. J. Khokhar, N. A. Saber, T. Spetebroot, and C. Barakat, "On active sampling of controlled experiments for qoe modeling," in *Proceedings of the Workshop on QoE-based Analysis and Management of Data Communication Networks*. ACM, 2017, pp. 31–36.
25. T. Hoßfeld, L. Skorin-Kapov, P. E. Heegaard, M. Varela, and K.-T. Chen, "On additive and multiplicative qos-qoe models for multiple qos parameters," in *5th ISCA/DEGA Workshop on Perceptual Quality of Systems, PQS 2016*, 2016.
26. S. Baraković and L. Skorin-Kapov, "Multidimensional modelling of quality of experience for mobile web browsing," *Computers in human behavior*, vol. 50, pp. 314–332, 2015.
27. ITU-T Recommendation G. 107, "The e-model, a computational model for use in transmission planning," 2015.
28. N. Osaka and K. Kakehi, "Objective model for evaluating telephone transmission performance," *Review of the Electrical Communication Laboratories*, vol. 34, no. 4, pp. 437–444, 1986.
29. ITU-T Recommendation P.1202.2, "Parametric non-intrusive bitstream assessment of video media streaming quality - higher resolution application area," 2013.
30. B. Belmudez and S. Möller, "Audiovisual quality integration for interactive communications," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013, no. 1, p. 24, 2013.
31. M. H. Pinson, W. Ingram, and A. Webster, "Audiovisual quality components," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 60–67, 2011.
32. W. Robitza, S. Göring, A. Raake, D. Lindegren, G. Heikkilä, J. Gustafsson, P. List, B. Feiten, U. Wüstenhagen, M.-N. Garcia *et al.*, "Http adaptive streaming qoe estimation with itu-t rec. p. 1203: Open databases and software," in *Proceedings of the 9th ACM Multimedia Systems Conference*. ACM, 2018, pp. 466–471.

33. D. L. Hoffman and T. P. Novak, "Advertising pricing models for the world wide web," *Internet publishing and beyond: The economics of digital information and intellectual property*, vol. 5, p. 2, 2000.

34. R. Imran, M. Odeh, N. Zorba, and C. Verikoukis, "Quality of experience for spatial cognitive systems within multiple antenna scenarios," *IEEE Transactions on Wireless Communications*, vol. 12, no. 8, pp. 4153–4161, 2013.

35. P. Tang, P. Wang, N. Wang, and V. N. Ngoc, "QoE-based resource allocation algorithm for multi-applications in downlink LTE systems," in *Int. Conf. on Computer, Communications and Information Technology (CCIT 2014)*, 2014, pp. 1011–1016.

36. A. W. Yusuf-Asaju, Z. M. Dahalin, and A. Ta'a, "Framework for modelling mobile network quality of experience through big data analytics approach," *Journal of Information and Communication Technology (JICT)*, 2018.

37. M. Varela and T. Mäki, "Embracing uncertainty: A probabilistic view of HTTP video quality," in *First IFIP/IEEE International Workshop on Quality of Experience Management (QoE-Management 2017)*, Lisbon, Portugal, May 2017.

38. E. Liotou, H. Elshaer, R. Schatz, R. Irmer, M. Dohler, N. Passas, and L. Merakos, "Shaping QoE in the 5G ecosystem," in *Quality of Multimedia Experience (QoMEX 2015)*. IEEE, 2015, pp. 1–6.