Economics of Quality of Experience

Peter Reichl[1], Bruno Tuffin[2], and Patrick Maillé[3]

[1] FTW Telecommunications Research Center Vienna, Austria
[2] INRIA Rennes Bretagne-Atlantique, Campus de Beaulieu, France
[3] Télécom Bretagne, Cesson-Sévigné Cedex, France
`reichl@ftw.at`, `bruno.tuffin@inria.fr`,
`patrick.maille@telecom-bretagne.eu`

Abstract. While the recent strong increase of interest in Quality of Experience both in industry and academia has managed to place the end user again into the center of service quality evaluation, corresponding economic implications have not received similar attention within the research community yet. Therefore, in this section we point out some of the key links between the quality as delivered by the network, as perceived by the user, and as valued by the market. The example of logarithmic utility functions allows demonstrating how this broad interdisciplinary approach is able to provide significant contributions in describing and analyzing future telecommunication ecosystems in a holistic way.

Keywords: telecommunication ecosystem, utility function, Weber-Fechner Law, WQL hypothesis, charging for QoE.

1 From Quality of Service to Quality of Experience

While the notion of "Quality of Experience" (QoE) has managed to become one of the key buzzwords in the networking community over the last few years, it may be surprising to learn that the concept itself is far from being new. Indeed, already back in 1994, ITU-T Rec. E.800 [5] has defined service quality as "*the collective effect of service performance which determines the degree of satisfaction of a user of the service*". However, research on "Quality of Service" (QoS), as it has been called since, started to neglect this explicit user-centric focus very soon and instead has put a clear emphasis on quantitative (and easily measurable) network parameters like, e.g., packet loss rate, throughput, delay, jitter and/or bandwidth.

Of course, this reduction of research scope is strongly based on the implicit assumption that improving the quality of packet delivery with respect to one or more of these parameters will automatically lead to some sort of increase in user satisfaction. While this assumption in general is not unreasonable, we have to note on the other side that reality sometimes may be much more complex. Let us, for instance, consider the user-perceived quality of file download in a mobile broadband scenario, where downlink bandwidth can be safely assumed to play a key role as QoS parameter. Depending on the complexity of the task, it has been experimentally shown that users

A.M. Hadjiantonis and B. Stiller (Eds.): Telecommunication Economics, LNCS 7216, pp. 158–166, 2012.

demonstrate two different types of behaviour: as long as they are performing very simple download tasks (e.g. download of an mp3 or zip file), it seems that they are simply evaluating their waiting time until task completion (i.e. file download completed), leading to a logarithmic law as expected from what we know already from psychology [1].

However, if it comes to web browsing, this indirect proportional relationship between downlink bandwidth and user waiting time (i.e. doubling the bandwidth reduces waiting time by 50%) is no longer valid, basically for two reasons [1]: Firstly, due to complex interactions of the HTTP and TCP protocols with network performance, the network-level page load time for web pages does not directly depend on the available bandwidth. Moreover, rendering and displaying the web page on the local machine leads to additional non-linearities also for the resulting application-level page load time. Secondly, there is also a noticeable difference between perceived subjective page load time and the page load time on the application level, caused by the simple fact that while browsing through web pages, users regularly perceive the load process of a web page as already finished while in reality content is still being retrieved (e.g. because the browser window may be too small to display the web page in its entirety, or because the user, due to progressive rendering of the browser, does not anticipate that there might be additional content still under way). Indeed, experimental results indicate that the technical page load time typically differs from the perceived page load time by a surprisingly high factor between 1.5 and 3 [1].

Therefore, recently the strict user-centred emphasis of the original QoS concept has been reinforced, underlined by the new terminology of QoE whose fundamental definition again is due to ITU-T and reads as "*overall acceptability of a service or application, as perceived subjectively by the end-user*" [6]. Note, however, that the research community is still far from agreeing on this definition – among the various attempts to improve it, maybe the proposal elaborated in 2009 during a related Dagstuhl seminar is most noteworthy, describing QoE as "*degree of delight of the user of a service, influenced by content, network, device, application, user expectations and goals, and context of use*" [2].

2 Microeconomic Service Valuation

Indeed, we can assume an even broader interdisciplinary perspective while discussing appropriate concepts for determining the value of a communication service, leading us directly into microeconomic utility theory, which aims at describing the preferences of user i with the help of a so-called "utility function" $u_i(x)$. To this end, we formally define $u_i(x)$ to be a mapping of the consumption set X, i.e. the set of all resources user i could possibly consume, to real numbers such that $u_i(x) \leq u_i(y)$ implies that the user prefers y over x, with $x, y \in X$.

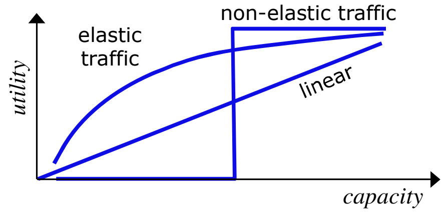

Fig. 1. Typical examples of utility functions

Figure 1 displays some typical shapes of utility functions for the case of network capacity: while linear utility assumes a direct proportionality between capacity and user satisfaction, the non-elastic case refers to applications (e.g. certain codecs) which require a certain minimum capacity but do not gain from additional capacity. Probably the most typical case is referred to as "elastic traffic", where initially some small capacity has quite a positive effect, whereas the impact of additional capacity is decreasing along with the level of already offered capacity (formally speaking, this mapping is assumed to be monotonically increasing, continuous - or at least right-continuous - and concave).

Amongst the many examples, weighted logarithmic utility functions of the form $u_i(x_i) = w_i \log x_i$ have been traditionally playing a key role in illustrating this concept. Note that in this case, if we assume a total number of users *N*, i.e., $i \in \{1,...,N\}$, the overall social welfare, defined as sum of user utilities, equals

$$U(x_1,...,x_N) = \sum_i w_i \log x_i \tag{5.1.1}$$

As has been shown by Kelly [5], the flow allocation x_i^* which maximizes (5.1.1) under capacity constraints of the form $\sum x_i \le C_j$, where the sum is taken over all flows *i* with routes *r*(*i*) using link *j* with its associated capacity C_j, fulfils the criterion of *weighted proportional fairness*, i.e. there is no other allocation x with positive sum of proportional rate changes:

$$\sum_{i:j\in r(i)} w_i \frac{x_i - x_i^*}{x_i^*} = \sum_{i:j\in r(i)} w_i \frac{\Delta x_i^*}{x_i^*} \le 0 \tag{5.1.2}$$

Hence, as a primary conclusion we note that optimizing the overall social welfare under the assumption of logarithmic utilities leads to a proportionally fair allocation of the available network bandwidth.

3 WQL Hypothesis and Weber-Fechner Law

While so far, the assumption of a logarithmic utility function has been made mainly for reasons of mathematical tractability, we will now discuss whether and to which extent this can be linked to fundamental laws governing the perception of service quality in real life. Indeed, there is a significant number of examples in the related literature which report on such logarithmic relationships between a certain networking parameter (used as trigger/stimulus) and the resulting Quality of Experience, usually expressed in terms of Mean Opinion Score (MOS) values, including scenarios like Voice over IP quality depending on varying bitrates as evaluated by Rubino et al.'s Pseudo-Subjective Quality Assessment (PSQA) tool [14], web browsing under IP latency, using an experimental design where IP latency can be assumed to be equivalent to user waiting time [3], download of files (different sizes) or pictures in mobile broadband scenarios [1], and connection setup times for accessing a 3G mobile network [1].

It turns out that especially for simple tasks like "download this file", "connect to this mobile network", "start a Google search on the term 'xyz'", "go to the next picture in the gallery", etc., we observe a fundamental logarithmic law for QoE evaluation throughout. This has been recently termed "WQL hypothesis", stating that "*the relationship between **W**aiting time and its **Q**oE evaluation on a linear Absolute Category Rating (ACR) scale is **L**ogarithmic*". As argued in [1], there is sufficient empirical evidence suggesting that the WQL hypothesis cannot be rejected.

Should we be surprised by this result of logarithmic laws governing our perception of service quality at least to some extent? The answer to this question leads us even much deeper into the interdisciplinary arena, i.e. to the field of psychophysics as the science of quantifying the general behaviour of the human sensory system. The establishment of psychophysics dates back well into the middle of the 19th century when the German physiologists Ernst Heinrich Weber and Gustav Fechner first described what soon should become a very fundamental contribution to psychology of perception, i.e. the so-called Weber-Fechner Law. Their theory is based on the key concept of "just noticeable differences" which are assumed to be at the core of the human sensory system. According to this principle, sensory differences can be observed only if the corresponding trigger (physical stimulus) is changed by at least a certain proportion of its current value. Formally speaking, the differential perception dP is assumed to be directly proportional to the relative change dS/S of a physical stimulus of magnitude S. Assuming k as constant of proportion, straightforward integration then yields

$$dP = k \cdot \frac{dS}{S} \quad \Rightarrow \quad P = k \cdot \ln \frac{S}{S_0} \tag{5.1.3}$$

where P describes the magnitude of perception as a function of the stimulus size S and a stimulus threshold (constant of integration) S_0.

The Weber-Fechner Law (WFL) can be applied to a surprisingly broad range of scenarios, ranging from vision (logarithmic stellar magnitudes), hearing (logarithmic dB scale), tasting, smelling, touching etc. However, its validity is not only restricted to actual human senses, but includes as well numerical cognition [8] and time perception [15].

4 The Fixed Point Problem of QoE Charging

Summarizing briefly what we have discussed so far, we have identified fundamental logarithmic dependencies between certain network parameters (QoS parameters) serving as stimulus and the user-centric perceptional evaluation of the resulting service quality for a broad variety of scenarios and evaluation methods (ranging from standards over learning tools to actual user trials). Moreover, microeconomic theory suggests that this kind of logarithmic law has further very interesting properties, most notably the fact that distributing network resources in a (proportionally) fair manner leads to the maximization of overall user benefit (social welfare). Bringing both lines of argumentation together, we have thus been able to show that indeed proportional fairness of resource allocation is equivalent to maximizing overall Quality of Experience – a relationship which to the best of our knowledge has not yet been formulated expressively so far.

While we started this section with a formal introduction of the utility function $u_i(x)$ as a user-specific preference function, there are different approaches if it comes to quantifying $u_i(x)$, e.g. by estimating the reselling value of a resource, or the willingness-to-pay of user *i*. Sticking to this latter approach, $u_i(x_i) = w_i \log x_i$ can hence be interpreted as the maximal charge/tariff user *i* would be willing to pay for the delivered QoS (e.g. bandwidth) x_i.

In this way, we have a basic mechanism to charge for QoS – determine the actually delivered QoS parameter, apply a suitable tariff function and calculate the resulting price. In this simple model, there is a primary (network level) feedback cycle as the charged price triggers the overall demand which itself is the key factor for the degree of congestion in the network and thus influences the provided Quality of Service (see Fig. 1 left).

This process is fundamentally different for the case of QoE charging (see Fig. 2 right): on the one hand, here we still pay for quality (now: Quality of Experience), i.e. we have to provide an estimation of the delivered QoE and calculate a charge from that. What makes life complicated, however, is the fact that the QoE estimation may heavily depend on the expected price itself (as well as on the QoS delivered by the network and also other factors). In this sense, the charge is serving both as input and as output of the QoE evaluation, thus constituting a secondary (user level) feedback cycle: high prices let the QoE expectations grow, hence the actual QoE evaluation will deliver relatively low results which by itself cannot justify the initial high prices. Similarly, low prices do not constitute major expectations, and as a consequence, the user is positively surprised by the experienced quality and would even be willing to pay more for it than what is actually charged.

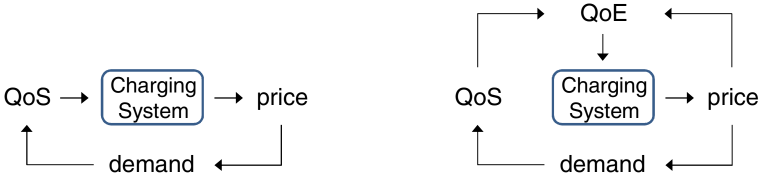

Fig. 2. Charging for QoS versus Charging for QoE

It is pretty obvious that this user-level feedback leads us to a very general fixed point problem – determining the precise charge which causes the right amount of expectations such that the subsequent QoE evaluation leads to a result for which the user is willing to pay exactly the amount she is asked for from the very beginning. While the determination of willingness to pay for certain perceived quality has experienced a comparably broad treatment in the related work, this is not the case for the other side of the medal, i.e. the influence of user predisposition caused by knowledge about the tariff structure onto her evaluation of perceived quality. To the best of our knowledge, the only experimental evidence so far is due to a series of user trials conducted in 2000/2001 as part of the European FP5 project M3I (Market-Managed Multiservice Internet).

Here, the idea was to offer users short video clips delivered with different frame rates (ranging from 1/sec to 25/sec) and ask them to indicate by a slider their willingness to pay. In addition, before starting the trial each user has been informed that she is member of one out of three categories (gold/silver/bronze) with VIP treatment at high charges for the gold users, preferential treatment at medium charges for the silver users and ordinary treatment at low charges for bronze users. Note that, actually, during the experiment no user differentiation whatsoever has occurred. For further details on the trial setup we refer to [3], which also provides the source for Figure 3.

From these results, we observe that – independently of the level of quality delivered – user expectations have indeed some influence on the quality evaluation. As far as acceptability is concerned (Fig. 3 top), gold users, who have been suggested to experience preferential ("VIP") treatment, show significantly higher expectations compared to silver/bronze users, while, at the same time, they are also willing to pay significantly more than their non-VIP colleagues (Fig. 3 bottom). Hence we may conclude that the user predisposition has indeed significant influence on her QoE evaluation, thus confirming the need for a detailed analysis of the mentioned secondary feedback cycle.

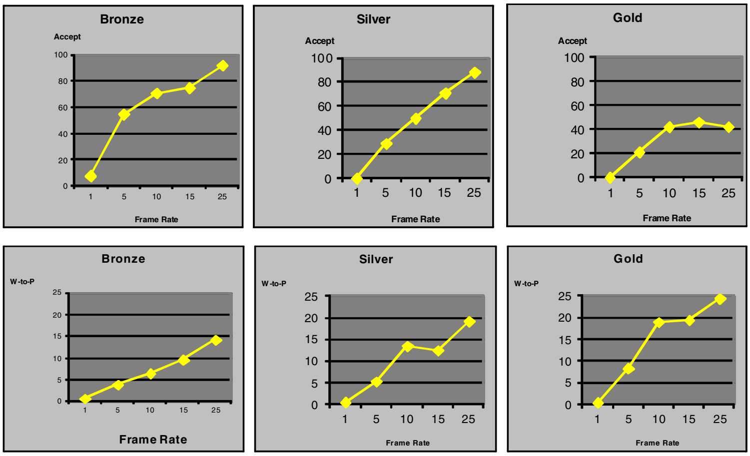

Fig. 3. Acceptability (in percent, first row) and Willingness-to-Pay (in pence per min, second row) for Different User Categories (source: [3])

5 Mechanisms for QoE-Based Charging

Depending on the location where the fixed point problem introduced in the previous section is supposed to be solved, we may distinguish two main approaches for QoE-based charging mechanisms: either, the transformation from QoS measurements to QoE estimations is performed by the system using an appropriate tool, while the result serves as input into a conventional Charging and Accounting System (CAS), or we leave this difficult task entirely to the end user who is put in charge to decide in real-time on his willingness-to-pay for the currently experienced service quality.

Whereas the first approach basically requires extending the CAS by some automatic QoE evaluation mechanism, e.g. based on learning algorithms (PSQA [14] and similar tools could provide invaluable help in this respect), we consider the second option to be by far more interesting. For instance in [10] we have proposed a feedback-based approach which is reducing the input required from the end user to a single bit of information indicating that both (a) the current QoE is unsatisfactory and (b) the end user is willing to pay for better quality. In terms of technical implementation, this solution only requires the installation of a "hot button" to be pressed by the end customer if both (a) and (b) are fulfilled, and as long as the additional quality (at the additional price) is in equilibrium with the user's expectations and needs. For further details about the prototypical realization of this concept in the framework of an IMS test-bed we refer to [10].

6 Toward the Future Telecommunications Ecosystem

This section has been devoted to exploring research questions at the precise intersection of microeconomics, psychology of perception and networking technology, demonstrating how such an interdisciplinary approach is able to make us aware of and to open us to a plethora of links and bridges between these disciplines which we initially had hardly hoped for. At the same time, we consider this as a very typical feature for the way telecommunications research will have to be conducted in the future in order to be able to contribute significantly to the sustainable success of this industry as a whole. From this point of view, it is no longer sufficient to restrict our research to more or less pure communications engineering, but on the contrary it is essential to continuously integrate the economic and user perspective as well.

This holistic approach requires at the same time a paradigm change which can be best described by the transition from communication systems to *communication ecosystems* (see [8] for a pivotal contribution to the establishment of this new overarching framework notion). Remember that the concept of an "ecosystem" as we know it from biology has turned out very useful in describing a community of organisms together with their environment, viewed as a system of interaction and interdependent relationships between the inhabitants as well as with the environment. It is most remarkable that the typical structure of a biological ecosystem, viewed as a pyramid composed of several layers (inorganic matter, basic source of food, primary consumers, secondary consumers and tertiary consumers = carnivores), which describe the most general interaction as of “eating or being eaten”, strikingly resembles the classical layered structure of communication systems (physical layer, link layer, network and transport layer up to application layer) as we know it for instance from the ISO/OSI model. However, we now realize that the ISO/OSI approach only describes the environmental (= network technology) part of our communication ecosystem, whereas the holistic integration of its inhabitants, be it end costumers, business units or network, service, application, content providers, will require a significant extension way beyond layer 7. As a first consequence, this novel approach does no longer focus technology evolution as such, but rather on investigating how we can use it and what we can do with it.

Of course, putting such a holistic and interdisciplinary approach into practice is a different (and equally interesting) story. To this end, already almost a decade ago we have proposed to describe this interrelation between Network efficiency (economics), User acceptance (usability/user-perceived quality) and Technological feasibility (technology) as a kind of dynamic triangle of forces which we used to call “NUT Trilemma” [11]. However, it is essential to acknowledge that, while we believe that technology as such can no longer serve as the ultimate goal of our work, starting from solid technological grounds is still considered the indispensable first step which allows for the subsequent integration of economic and user aspects into a holistic framework.

Acknowledgments. Part of this work has been funded by the Austrian government and the City of Vienna within the COMET funding program. Additional support from COST Action IS0605, SISCOM/UEB (in the framework of the International SISCOM Research Chair 2010/11), INRIA Rennes and LAAS-CNRS Toulouse is gratefully acknowledged.

References

1. Egger, S., Reichl, P., Hoßfeld, T., Schatz, R.: Time is Bandwidth? Narrowing the Gap between Subjective Time Perception and Quality of Experience. In: Proc. IEEE International Conference on Communications (ICC 2012) – Communication QoS, Reliability and Modeling Symposium, Ottawa, Canada (June 2012)
2. Fiedler, M., Kilkki, K., Reichl, P.: From quality of service to quality of experience. Dagstuhl Seminar Proceedings 09192, `http://drops.dagstuhl.de/opus/volltexte/2009/2235/pdf/09192`
3. Hands, D., Gale, C., De Bruine, A.: User Experiments and Trials. FP5 project M3I, Deliverable 15.2 (November 2001)
4. Ibarrola, E., Liberal, F., Taboada, I., Ortega, R.: Web QoE evaluation in multi-agent networks: validation of ITU-T G.1030. In: Proc. ICAS 2009, Valencia, Spain (April 2009)
5. ITU-T Rec. E.800: Terms and definitions related to quality of service and network performance including dependability (1994)
6. ITU-T Rec. G.100/P.10: Vocabulary for performance and quality of service, amendment 2: new definitions for inclusion in Recommendation P.10/G.100 (2008)

7. Kelly, F., Maulloo, A., Tan, D.: Rate Control in Communication Networks: Shadow Prices, Proportional Fairness and Stability. Journal of the Operations Research Society 49, 237–252 (1998)
8. Kilkki, K.: Quality of Experience in Communications Ecosystems. Journal of Universal Computer Science 14(5), 615–624 (2008)
9. Longo, M.R., Lourenco, S.F.: Spatial attention and the mental number line: evidence for characteristic biases and compression. Neuropsychologia 45, 1400–1406 (2007)
10. Reichl, P., Fabini, J., Kurtansky, P., Stiller, B.: A Stimulus-Response Mechanism for Charging Enhanced Quality-of-User Experience in Next Generation All-IP Networks. In: Proc. XIII Conferencia Latino-Ibero-Americana de Investigación de Operaciones (CLAIO 2006), Montevideo, Uruguay (November 2006)
11. Reichl, P., Hausheer, D., Stiller, B.: The Cumulus Pricing Model as an Adaptive Framework for Feasible, Efficient and User-friendly Tariffing of Internet Services. J. Computer Networks (2003)
12. Reichl, P., Egger, S., Schatz, R., D'Alconzo, A.: The Logarithmic Nature of QoE and the Role of the Weber-Fechner Law in QoE Assessment. In: Proceedings of the 2010 IEEE International Conference on Communications, pp. 1–5 (May 2010)
13. Reichl, P., Tuffin, B., Schatz, R.: Logarithmic Laws in Service Quality Perception: Where Microeconomics Meets Psychophysics and Quality of Experience. Telecommunication Systems Journal 55(1), June 18 (2011), doi:10.1007/s11235-011-9503-7 (to appear, January 2014)
14. Rubino, G.: Quantifying the Quality of Audio and Video Transmissions over the Internet: the PSQA Approach. In: Barria, J. (ed.) Design and Operations of Communication Networks: A Review of Wired and Wireless Modelling and Management Challenges. Imperial College Press (2005)
15. Takahashi, T.: Time-estimation error following Weber-Fechner law explain subadditive time-discounting. Medical Hypotheses 67(6), 1372–1374 (2006)
16. Weber, E.H.: De Pulsu, Resorptione, Auditu Et Tactu. Annotationes Anatomicae Et Physiologicae. Koehler, Leipzig (1834)