



A Theoretical Approach to the Formation of Quality of Experience and User Behavior in Multimedia Services

Werner Robitza¹, Sabine Schönfellner², Alexander Raake³

¹Telekom Innovation Laboratories, Deutsche Telekom AG

²Institute of German Studies, Karl-Franzens-Universität Graz

³Audiovisual Technology Group, Technische Universität Ilmenau

werner.robitza@telekom.de, sabine.schoenfellner@uni-graz.at, alexander.raake@tu-ilmenau.de

Abstract

This paper presents a conceptual model that relates the quality formation process—as established in literature—to human behavior in multimedia consumption. It gives theoretical definitions of behavioral aspects to allow a common understanding of the terms in the User Experience and Quality of Experience communities. The work creates a basis for the predictive modeling of behavior, identifies challenges for its practical assessment, and poses research questions that will need to be answered.

Index Terms: Quality of Experience, user behavior, multimedia, psychology

1. Introduction and Motivation

In today's world, service providers, content producers and distributors strive to increase the time users spend with their service, for example watching videos or having video conversations with friends and family. We call this *user engagement*. Engagement is typically described from a long-term perspective, but in principle, it is the result of short-term behavior that can be observed in users. Here, we deal with questions such as: How long are users' calls? How many minutes of a certain video do they watch? Even more importantly, service providers would like to know when and why people *abandon* services: for example, when will users stop tolerating long video loading times? Large-scale studies (e.g., as described in [1–4]) have already analyzed user behavior, but they are inadequate at determining underlying QoE factors.

In order to be able to predict this form of engagement, we first need to understand how the current level of Quality of Experience (QoE) results in short-term behavioral (inter)actions. However, the simple assessment of a Mean Opinion Score (MOS) will not be enough to be able to infer specific user actions. In fact, the typical quality assessment models we use today—being developed on MOS ratings only—cannot predict behavior at all. This also highlights another problem: the user test methods that are normally applied do not even allow users to behave as they do during real life usage, causing a strong demand for new experimental paradigms, such as studied in [5, 6].

Theoretical concepts on how quality forms in the human mind have already been described in previous works from Raake *et al.* [7–9] and in the Qualinet White Paper [10], based

This work was funded from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 643072, Network QoE-Net.

on [11]. Yet, there is still a need for relating the perceptual quality formation process to human behavior—that is, short- and long-term behavior. Reichl *et al.* [12] describe a first high-level framework that considers QoE and behavior. The work takes a technical perspective: QoE affects user state, which in turn influences user behavior. The authors give examples of scenarios where the application of behavior models can be useful. However, the paper does not detail the specific interplay between quality and behavior and does not address different behavioral dimensions. Considering the current state of the art, we believe that it is necessary to formulate the underlying hypotheses for behavioral research.

In this paper, we first explain the theoretical background on behavior in Section 2, then present a conceptual model of user behavior and QoE in Section 3. It serves as a basis for further investigations on the subject of human behavior in QoE and highlights which factors need to be studied in more detail, as explained in Section 4.

2. Psychological Background on Behavior

Although many definitions of “behavior” exist in biology (and scientists do not agree on one [13]), we can generally describe it as any response of a human to a stimulus. However, in order to facilitate understanding of different forms of behavior, it is useful to classify behavioral acts from different perspectives.

We primarily distinguish invisible (covert) from visible (overt) behavior; we consider the latter to be of most interest to our research. Also, behavior can be interactive or non-interactive. For interactive behavior, we discern between human-to-human interaction and human-to-computer interaction. Note that in the domain of QoE, human-to-human behavior is typically mediated by human-to-computer interaction, for example in a phone conversation. Finally, the dimension of time needs to be mentioned: when we speak about behavior, in this paper we mean short-term individual actions of a few seconds duration at most. Behavioral *patterns* [14] group individual behavioral actions together.

Different types of behavior can be associated with different levels of consciousness [14, 15]. Note that in psychology, there are diverging views on how human consciousness is structured, resulting in different models of information processing and behavior generation (see for example [16]).

- *Unconscious behavior* happens without thought. Generally, two classes of unconscious behavior exist: First, there is pre-cognitive *reflexive/intuitive behavior*, which cannot be directly influenced. We typically only indirectly perceive it through inference and can reflect upon it later. One example

of such behavior would be automatically (i.e., unconsciously) closing the eyes in response to a flash of light. Second, we find *subconscious behavior*, which is induced by societal norms rather than pre-wired processes [14], e.g., using headphones so as not to disturb others.

- *Conscious behavior* always requires a *thought action*. However, most daily behaviors are *post-conscious*; they have become habits (“learned acts”) and at some point do not require explicit thought anymore. In situations that deviate from the usual script, these behaviors demand conscious attention and can still be reflected upon. This may be the case when unexpected quality problems with media services occur. Here, we can link behavior to the quality formation process. For example, using a video chat demands conscious attention until a habit is formed; the usage becomes post-conscious until a certain event occurs that deviates from the habitual, for example when the voice quality drops noticeably (see *Quality Awareness* in Section 3.2) and the conversation partners have to consciously decide whether or not to quit and re-establish the call.

3. The Quality and Behavior Formation Process

In Figure 1, we present our proposed model for quality and behavior formation, based on [8] and related models. The model depicts a person in the state of experiencing a multimedia service.

3.1. Perception and Reflexive/Intuitive Behavior

In the *perception* process, perceptual events are generated through sensory processing, which acts as a selective feature extractor for our surrounding world: it selects only relevant stimuli for further processing. Here, stimuli are characterized on the basis of those perceptual events, which are compared to internal references. Perception comprises all the five senses, but in the classic telecommunication quality research domain, sound and vision have been of primary interest.

The output of the perception process are the perceptual events themselves as well as the perceived character of those events. During perception, unconscious and reflexive behavioral events are formed, such as eye movements towards regions of interest or the already mentioned example of turning the head towards a sound source. In the classic model [8], these are called “exploratory actions”. They typically directly influence the perception of signals, for example by changing visual angle.

3.2. Quality Formation

The *quality formation* process assumes the user being in a state of experiencing the perceptual events generated before [8]. In this case, we build upon an extended version of the process, which includes a component of quality attention focus [7]. It determines the specific quality features a quality judgment will be based on.

In this contribution, we introduce a new state of *immersion*. It describes a feeling of being “sucked” into the media. Immersion is also related to the concept of *flow* [17]. When users are immersed, they may be less aware of their surroundings. Typically, this term is used for Virtual Reality (VR) environments, where it relates to the “objective level of sensory fidelity” of the VR system [18, 19]. However, to some extent, users can feel immersed into “traditional” media, too.

Immersion is influenced by several factors: content-related and technical factors of the media itself, as well as external factors such as the context of use (e.g., the user’s task and location). Those factors are implicitly contained in the *assumptions* block. Distraction also plays a role here. In our model, the immersion state is mainly determined by the process of experiencing. For example, users get more and more immersed into a movie they are watching. When the user achieves awareness of the current quality, a process is started in which he or she reflects on the events and attributes the quality features (*reflection and attribution*). This process could be triggered by the character of the perceptual events deviating from the user’s assumptions, for example when the video stream suddenly becomes blurry. Also, an intervention from the outside may cause quality awareness (e.g., in a subjective quality assessment task). The level of immersion also influences the quality awareness process: when users are deeply immersed, they may not be easily distracted by quality problems.

3.3. (Post-)conscious and Affective Behavior Formation

When the above-mentioned reflection and attribution process is triggered, it feeds into a *behavioral matching* process. Here, the person evaluates which possible action in their behavioral repertoire will likely yield which result. The repertoire comprises all actions they are capable of, based on their memory. In fact, we can distinguish different types of long-term memory, depending on whether we can consciously recall it or not: 1) *Semantic memory* contains our knowledge about the world in general; *episodic memory* is what we colloquially call “memories”: it stores specific instances in our life that we can recall. Both types of memory are available to us consciously; they are called “explicit” [20]. 2) *Procedural memory* on the other hand is a form of implicit memory [21]. It contains basic (bodily) actions such as walking, eating, or using a mouse to click a button. They can be conscious if required, but we usually do not (have to) think about them.

According to Brewin [15], behavioral matching is a “look-ahead” process. It requires cognition—in contrast to pre-wired behaviors—and considers all previous experience of the user. The matching process may result in behavioral events, if it is determined that they help the user in reaching a goal. In normal usage, we would expect goals such as watching a video or starting a phone call, but in the context of multimedia QoE, we hypothesize that this goal typically is to increase the user’s quality of experience or immersion; more specific goals however depend on the concrete application and the user’s task. Whether users become consciously aware of this matching process or not depends on the context in which it is carried out.

Generally speaking, the more severe the quality problems or the lower the experience level of the user in a given setting, the more likely it is that their actions require more conscious deliberation. Imagine a user deciding to pause a video stream to let it buffer in order to prevent further stalling events. We assume that the user at this point has through their deliberation decided that it was worth sacrificing waiting time for a more enjoyable streaming experience. In more complex cases such as video telephony, we also have to consider corrective actions that adapt the behavior to the current QoE, such as changing the pace of speaking [22].

The output of this matching process (i.e., the behavioral events themselves) may of course affect the stimulus—or its perception—and ultimately yield a higher level of QoE (after perception and quality formation). This process could happen

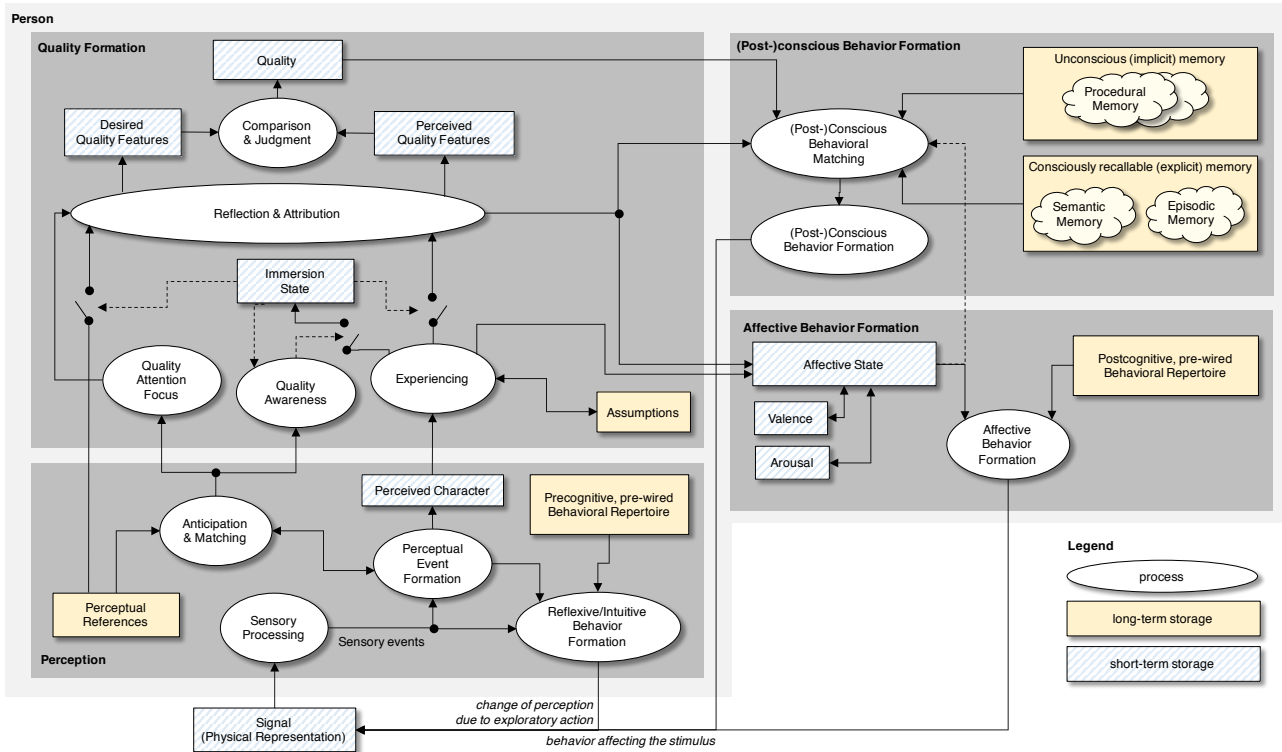


Figure 1: Conceptual model of quality and user behavior formation.

almost immediately (e.g., when increasing the volume of music to hear a certain part) or develop over time (e.g., increasing immersion into a movie by playing it in fullscreen). If the behavioral events do not help in attaining the goal of a better experience, the user may learn from such an event and in the future decide to behave differently. Here we have to consider certain aspects of learning, which are not yet part of the proposed model.

In addition to the (post-)conscious behavioral component, there is an *affective* behavioral component, in which an unconscious process determines the pre-wired response to a certain affect (felt emotion). This response—mostly in the form of covert behavior—strongly depends on the valence and arousal potential of the stimulus and the user’s affective state. In our conceptual model, one example for such a response is the development of anger or frustration, which could result in an increased heart rate, or even hitting the mouse on the table. This process is generally unconscious, but it can be perceived by inference through its results. For example, a video conferencing user may notice that he or she is becoming angry due to a strong audio delay or echo. The current affective state also is considered in the (post-)conscious anticipation process.

4. Challenges for Future Research

Theory, assessment and prediction of human behavior are not new concepts per se: we find related research in the fields of psychology (e.g., cognitive and clinical psychology, learning, behavior change in psychotherapy) and User Experience (UX, e.g., test methodologies, Human–Computer Interaction). However, once we link behavior to QoE, new challenges emerge.

In order to assess behavior and experienced quality, the currently often-used (and standardized) subjective test methodologies need to be adapted. First and foremost, passive viewing-

only or listening-only tests (e.g., described in ITU-T recommendations P.910, P.800 and others) do not allow users to interact with systems that would be interactive in real life, since only the audiovisual stimulus itself is assessed passively. However, once we allow interaction, we can identify several influencing and confounding factors that have to be addressed. They are not necessarily orthogonal in how they impact results; we expect a significant interplay between some of those factors.

4.1. Ecological Validity

Current subjective viewing or listening quality assessment methods often rely on the experiment leader explaining users to “imagine using a service”, by constructing a context of use that simulates real life (which cannot be replicated in the lab). Other test methods require subjects to rate the quality of recorded interactions instead of taking part in an interaction themselves. All those tests are often performed in dedicated, neutral test rooms. This so-called decontextualization of the test situation [23] creates a lack of *ecological validity*, which means that the gathered results cannot necessarily be applied in real life, since they are just valid in the lab context they have been acquired in.

Apart from the use of more natural test environments (which is more common in the field of UX), a more ecologically valid quality assessment test would obviously have to allow users to interact, too. This calls for a shift from dedicated test platforms—which only emulate a given (type of) service—to the use of real services or entirely emulated instead. Such a paradigm change would also mean that instead of viewing or listening to stimuli, users would be engaging in *sessions* (e.g., they would have to browse through content, click to start a video playback, end a session). It is therefore not only the stimulus itself that will be subject to rating, but larger parts of

an application interaction.

Increasing ecological validity often goes hand in hand with the introduction of more confounding factors in the test, or making it too specific to a certain system under study. A great challenge will therefore be to define ecologically valid test paradigms for behavioral research.

4.2. Observer Effects and Demand Characteristics

The specific context of a test situation influences how test participants behave. The notion of *demand characteristics* is well summarized by Orne [24]: “[The] subject is not a passive responder to stimuli and experimental conditions. Instead, he is an active participant in a special form of socially defined interaction which we call ‘taking part in an experiment.’” This effect relates to how humans interpret the hypotheses of the study they are taking part in and how their responses reflect upon themselves. For example, users may be trying to please experiment leaders by giving favorable answers to questionnaires. Or, users may be reluctant to interact with systems in ways that they were not explicitly allowed to. Those experimental biases may be present in passive audiovisual quality tests as well; however, in interactive and behavioral test procedures they are exacerbated by another fact: we are even more interested in assessing participant reactions. Consequently, subjects may be even more influenced by demand effects in their behavior.

The confounding factor introduced by telling users that they are monitored may be limited in its effect by hiding the real experiment purpose, for example through a mock task (that is, not telling the users about monitoring behavior and giving them another task instead). This is a so-called “deceptive” study [25]. However, our QoE studies related to behavior in web video services show that this paradigm may not suffice—users seem to stay more passive in a laboratory setting than in real life, when they are forced to use services and equipment they are not familiar with [6]. It should be noted that ethical concerns may prevent researchers from fully hiding research hypotheses from their subjects [25], therefore, the revelation of the real purpose after the test to the subjects should be mandatory.

4.3. Designing Tasks and Integrating Quality Ratings

Subjects can take part in passive quality assessment tests with their only task being to provide a quality rating. In conversational speech quality tests, the use of specific tasks is already much more common than for the assessment of audiovisual quality services. The tasks (e.g., having to order plane tickets via telephone) are intended to drive the conversation itself naturally, which is then rated for its quality. However, this task only facilitates human-to-human behavior. Once we allow human-to-computer interactivity, we may have to give users another, more complex task. Those tasks would typically be aligned with use cases of the application under study. For example, participants would be asked to freely select test videos from a web portal instead of being presented a random order automatically. Giving another (more realistic) task in addition to rating quality may yield a higher motivation for users, therefore increasing ecological validity. At the same time, a new question arises: how do we compare ratings for tasks executed differently, that is, with different actions and/or outcomes?

It is not trivial to introduce a quality rating task in addition to a task that drives the session naturally. When users know they are expected to rate a system’s quality, they may focus on different system aspects, usually being much more critical in their ratings (see the *quality attention focus* in our model). This

may of course be intentional—it is useful for the development of instrumental quality models to be more sensitive. However, when the goal is a *natural* interaction and high ecological validity, the inclusion of quality ratings may change the participant’s behavior. Vice-versa, another research question relates to how certain behavior affects quality perception and thus the delivered quality ratings, for example whether subjects rate a session better if they were able to actively mitigate problems through their behavior. At this point however, we can ask ourselves whether a quality rating is even necessary, or whether the display of certain behavior (e.g., quitting a session) is also a useful indicator (as shown, e.g., in [1]).

4.4. Measurement of Immersion and Quality Awareness

In Section 3.2 we hypothesized that the current state of immersion impacts the user’s awareness of quality. When a user is deeply immersed, they may not be aware of quality issues that could otherwise be easily perceived. For example, if the person had been asked to provide a rating of the blurriness of a video stimulus, they would likely be less immersed into the presented content, due to their given task.

To our knowledge, there is no commonly accepted and validated method to assess the level of immersion from a QoE perspective. While it may seem straightforward to ask users to provide verbal or written feedback on how immersed they are feeling, the mere task providing a rating may make users already feel less immersed. Hence, a non-intrusive way to quantify immersion would be preferred. This may call for physiological measurements, for example with eye tracking, electroencephalography (EEG), heart rate, or galvanic skin response (GSR), although the use of those in a test could also impact the level of immersion.

4.5. User Factors and Individuality

Summarizing the points raised in the previous sections, it becomes apparent that user factors will have a strong impact on the expected demand effects of test situations on the one hand, and the expected behavioral responses on the other hand. Current quality assessment methodologies only capture basic demographic information, which is understandable when the models trained on such tests shall only predict the MOS of an “average user”. Here, the inclusion of a high number of subjects is important to even out differences in rating.

However, as we progress towards a more holistic view on QoE, we will likely see more individualized predictions from models, which in turn also require standardized measures of user factors through appropriate questionnaires. The challenging questions therefore are: Which factors should be more explicitly considered in tests? How can they be obtained in real life for a successful application of new types of models?

5. Conclusion

In this paper, we presented a theoretical model that extends the quality formation process with a model of human behavior. As a further contribution, we described a number of challenges which will have to be addressed for a successful assessment and prediction of behavior and QoE. They are based on the development of new, interactive test methodologies and rooted in experimental biases that are likely to be introduced by adapting known and standardized test procedures. We listed specific research hypotheses that we hope will be addressed in the coming years of QoE research.

6. References

- [1] F. Dobrian, A. Awan, D. Joseph, and A. Ganjam, "Understanding the impact of video quality on user engagement," *Communications of the ACM*, vol. 56, no. 3, pp. 91–99, 2013.
- [2] J. Shaikh, M. Fiedler, P. Arlos, and D. Collange, "Modeling and analysis of web usage and experience based on link-level measurements," in *Teletraffic Congress (ITC . . . , 2012*.
- [3] K.-T. Chen, C.-Y. Huang, P. Huang, and C.-L. Lei, "Quantifying Skype user satisfaction," *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 4, p. 399, 2006.
- [4] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 2001–2014, 2013.
- [5] R. K. P. Mok, E. W. W. Chan, X. Luo, and R. K. C. Chang, "Inferring the QoE of HTTP video streaming from user-viewing activities," in *Proceedings of the first ACM SIGCOMM workshop on Measurements up the stack - W-MUST '11*, 2011, p. 31.
- [6] W. Robitza and A. Raake, "(Re-)Actions Speak Louder Than Words? A Novel Test Method for Tracking User Behavior in Web Video Services," in *Eighth International Workshop on Quality of Multimedia Experience (QoMEX)*, Lisbon, 2016.
- [7] J. Skowronek and A. Raake, "Conceptual Model of Multiparty Conferencing and Telemeeting Quality," in *QoMEX*, 2015.
- [8] A. Raake and S. Egger, "Quality and Quality of Experience," in *Quality of Experience*. Springer, 2014, pp. 11–34.
- [9] A. Raake, *Speech quality of VoIP: assessment and prediction*. John Wiley & Sons, 2007.
- [10] P. Le Callet, S. Möller, and A. Perkis, "Qualinet White Paper on Definitions of Quality of Experience," European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Tech. Rep. March, 2013.
- [11] U. Jekosch, *Voice and speech quality perception: assessment and evaluation*. Springer Science & Business Media, 2006.
- [12] P. Reichl, S. Egger, S. Möller, K. Kilkki, M. Fiedler, T. Hossfeld, C. Tsirias, and A. Asrese, "Towards a Comprehensive Framework for QoE and User Behavior Modelling," in *QoMEX*, 2015.
- [13] D. A. Levitis, W. Z. Lidicker, and G. Freund, "Behavioural biologists do not agree on what constitutes behaviour," *Animal Behaviour*, vol. 78, no. 1, pp. 103–110, 2009.
- [14] J. Heimlich and N. Ardoin, "Understanding behavior to understand behavior change: a literature review," *Environmental Education Research*, vol. 14, no. 3, pp. 215–237, 2008.
- [15] C. R. Brewin, "Cognitive Change Processes in Psychotherapy," vol. 96, no. 3, pp. 379–394, 1989.
- [16] M. Velmans, "Is Human Information Processing Conscious," *Behavioral and Brain Sciences*, vol. 14, pp. 651–726, 1991.
- [17] M. Csikszentmihalyi, *Creativity: Flow and the Psychology of Discovery and Invention*. Harper Perennial, 1996.
- [18] M. Slater, "A note on presence terminology," *Presence connect*, vol. 3, no. 3, pp. 1–5, 2003.
- [19] D. A. Bowman and R. P. McMahan, "Virtual Reality: How much Immersion is enough?" *Computer*, vol. 40, no. 7, pp. 36–43, 2007.
- [20] E. Tulving, "Episodic and Semantic Memory," *Organization of Memory*. London: Academic, vol. 381, no. 4, 1972.
- [21] H. L. Roediger, "Implicit memory. Retention without remembering," *The American Psychologist*, vol. 45, no. 9, pp. 1043–1056, 1990.
- [22] K. Schoenenberg, "The Quality of Mediated-Conversations under Transmission Delay," Ph.D. dissertation, TU Berlin, 2015. [Online]. Available: <http://dx.doi.org/10.14279/depositonce-4990>
- [23] M. Peeters, C. Megens, C. Hummels, A. Brombacher, and W. Ijsselstein, "Experiential Design Landscapes: Design Research in the Wild," in *Nordic Design Research Conference 2013*, 2013.
- [24] M. T. Orne, "Demand characteristics and the concept of quasi-controls," *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books*, p. 110, 2009.
- [25] American Psychological Association, "2010 Amendments to the 2002 "Ethical principles of psychologists and code of conduct"." *The American Psychologist*, vol. 65, no. 5, p. 493, 2010.