Towards a New ITU-T Recommendation for Subjective Methods Evaluating Gaming QoE

Sebastian Möller[1], Jan-Niklas Antons[1], Justus Beyer[1], Sebastian Egger[2], Elena Núñez Castellar[3,4],
Lea Skorin-Kapov[5], Mirko Sužnjević[5]

[1] Quality and Usability Lab, Telekom Innovation Labs, TU Berlin, Germany
[2] AIT Austrian Institute of Technology, Vienna, Austria
[3] Department of Communication Sciences, iMinds-MICT-Ghent University, Belgium
[4] Department of Data Analysis, Ghent University, Belgium
[5] Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

Abstract— **This paper reports on activities in Study Group 12 of the International Telecommunication Union (ITU-T SG12) to define a new Recommendation on subjective evaluation methods for gaming Quality of Experience (QoE). It first resumes the structure and content of the current draft which has been proposed to ITU-T SG12 in September 2014 and then critically discusses potential gaming content and evaluation methods for inclusion into the upcoming Recommendation. The aim is to start a discussion amongst experts on potential evaluation methods and their limitations, before finalizing a Recommendation. Such a recommendation might in the end be applied by non-expert users, hence wrong decisions in the evaluation design could negatively affect gaming QoE throughout the evaluation.**

Keywords—Computer game, gaming QoE, subjective evaluation method, standard, ITU-T

I. INTRODUCTION

In addition to voice, video and web browsing, increasingly popular applications running on top of IP-based networks are online computer games. However, in contrast to the mentioned applications, little information exists on a) factors influencing the Quality of Experience (QoE) of online computer games, b) on subjective methods for assessing gaming QoE, or c) on instrumental quality prediction models. For this reason, ITU-T SG12 has decided to start work on three work items, one on the factors affecting QoE in gaming applications (G.QoE-gaming), one on the definition of subjective methods for evaluating the quality experiences during gaming activities (P.GAME) defined for ITU-T SG12 Question Q.7/12, and the third one on an opinion model for gaming applications (G.OMG), defined in ITU-T SG12 Question Q.13/12. The main focus of this paper will however be on P.GAME.

For this work item, a structure has been drafted and put for discussion in September 2014 [1]. This structure follows the general guidelines of ITU-T Recommendations in terms of the scope, definitions, abbreviations and conventions, and sets out for a description of gaming QoE aspects, the experimental set-up to be used in the subjective evaluation, and questionnaires for quantifying user perception throughout the evaluation. The draft finishes with some general considerations on the usefulness of performance measurements as well as physiological response measurements which may complement the evaluation. It is amended by exemplary questionnaires and bibliographic references.

Whereas the draft has been produced by an ITU-T member with an interest in evaluation methods in gaming QoE, it does not yet reflect the experience which the entire scientific community has gained by applying evaluation methods in practice. The authors think that such a draft should be discussed in the scientific community at large, as the resulting recommended methods may be used by anyone interested – but not necessarily experienced – in gaming QoE evaluation.

It is the aim of the present paper to foster the discussion on the draft, and to critically review potential candidate content and methods for inclusion into the upcoming Recommendation. As the draft is mainly available to ITU-T members, we start with a brief overview of the current status in Section 2. Section 3 then discusses issues regarding the structure of evaluation experiments, the test environment and set-up, the test participants, and questionnaires to quantify different QoE aspects. Section 4 concludes with a proposal for next steps to be started in order to substantiate and improve the methods to be recommended in the future.

II. CURRENT DRAFT

The current draft of the Recommendation considers gaming QoE as a multidimensional construct which includes several features. These features have been partially derived from the Game Experience Questionnaire (GEQ, see below), and have been put into a relationship in the taxonomy proposed in [2]. The current draft considers the features aesthetics (the sensory experience the system elicits, and the extent to which this experience fits individual goals and spirit, influenced e.g. by the visual appeal), interaction quality (playability of the game, in terms of the degree to which all functional and structural elements of a game enable a positive experience to the gamer), playing quality (game usability, i.e. the degree to which a player is able to learn, control, and understand a game), flow [3] (focused motivation leading to a feeling of spontaneous joy under conditions of balance between challenge and skills), immersion, and positive and negative affect. These features may lead to a positive game experience, i.e. the degree of delight or annoyance of the player after the gaming experience.

This definition is congruent to the one defined in the Qualinet White Paper [4] for QoE, and applied to gaming services.

It should be noted that the aforementioned features are not orthogonal, but partially overlapping and interrelated. For example, the responsiveness of the game towards user commands may be considered as a part of the interaction quality, but it may have a severe impact on the flow experience, especially in case that the game rules are violated because of the late reaction of the game towards user commands. On the other hand, it is also related to the quality of the graphical output, in case that it is mainly affecting the rendering of the graphical output, and not the game rules. Flow, in turn, may be strongly influenced by the immersion of the game. In fact, this is related to the definition of flow in positive psychology, where it is considered as "the mental state of operation in which a person performing an activity is fully immersed in a feeling of energized focus, full involvement, and enjoyment in the process of the activity. In essence, flow is characterized by complete absorption in what one does." [3].

In order to evaluate the mentioned quality features, the draft recommends the use of laboratory tests under con-trolled experimental conditions. The test environment should follow the requirements from ITU-T Rec.s P.910 [5] and P.911 [6] for the subjective evaluation of video and audio-visual quality of multimedia services, in terms of lighting and acoustic conditions. No recommendations are yet given with respect to the test system set-up. In turn, the draft focuses on the impact of the test participants and the game type. Whereas for the test participants characteristics like the skill, game preference, average playing time, and perceptual abilities are mentioned as important criteria, the characteristics of the game are less well structured. Furthermore, the draft describes that – apart from general video characteristics such as spatial and temporal complexity – the game genre, structure and rules can determine the impact degradations may have on gaming QoE. It is recommended that representative game scenes are selected which are adapted in complexity to the players' skills.

Furthermore, the draft contains recommended questionnaires to quantify selected quality aspects. These include methods for integral quality assessment (mainly following the Absolute Category Rating (ACR) methods described in ITU-T Rec. P.911 [6]), custom-built scales for diagnostic degradation assessment (such as the annoyance caused by network delay), scales for measuring flow (such as the flow short scale), the Self-Assessment Manikin [7] for assessing the affective dimensions valence, arousal and dominance, as well as the Game Experience Questionnaire (GEQ, [8]). This latter questionnaire is the most comprehensive one and makes use of 42 items to assess the features immersion, tension, competence, flow, negative affect, positive affect, and challenge; cf. the discussion on the GEQ in Section 4.

Finally, the draft mentions the possibility to assess performance and physiological responses of gamers. Since it has been found that the game outcome affects player experience, performance metrics such as the obtained points, number of kills, goals, units built/destroyed etc. may be related to QoE. In addition, parameters such as game session length or frequency may be employed as performance metrics. Physiological responses are a promising way to assess certain quality features without the interruption of the game process caused by questionnaires or interviews. They are, however, not yet well understood, and standardized experimental set-ups for their measurement in the gaming process are still lacking.

III. POINTS OF DISCUSSION

Whereas the present structure and topical selection of the draft is considered as a proper starting point, we see a number of points which are worth a more in-depth discussion. These points are raised in the following paragraphs in an enumerative fashion within each subsection. Conclusions and a priorization of these points follow in Section 4.

A. Evaluation procedure

As in any other evaluation paradigm, its definition should be guided by the purpose of the evaluation. For example, if the purpose of the evaluation is to determine the impact of different network conditions on the user's willingness to continue playing, this should put the playing activity into the focus of the evaluation, and the interruptions resulting from user feedback (e.g. using rating scales) should be kept to an absolute minimum. In turn, if the purpose is to identify and understand perceptual features underlying the gaming activity, then more detailed questionnaires are necessary.

Gaming is an interactive activity, thus evaluation experiments should reflect the interactive nature of gaming in order to provide valid results. In addition to the general interactive nature of games, the majority of players are using games for leisure, and therefore an interactive scenario is necessary to ensure the ecological validity. Usually, test participants are asked to play limited scenes of selected games (see below), and to provide quality ratings after each scene. Scenes are chosen to keep the level of difficulty, interactivity, length of the experience, and other factors which might potentially influence interaction behavior constant and comparable between participants. They may be interlinked to avoid boredom of the user during the experiment. Common scenes vary in length between one and several minutes. However, adding the duration of the rating procedure to the scene length, this may result in only a dozen of scenes (and corresponding gaming conditions, such as network or device settings) which can be evaluated within one test session. Thus, interactive tests limit the number of system influence factors to be tested.

A current proposal from an ITU member to overcome this problem are, viewing-only tests, listening-only tests, or listening-and-viewing tests similar to the ones defined for rating audio-visual material in a passive setting. Stimulus material in such tests is commonly limited to around 10 s length, so a significantly higher number of test conditions can be tested in one session. As an example, [9] proposes to use the SAMVIQ ("Subjective Assessment Methodology for Video Quality") method recommended in ITU-R Rec. BT.1788 [10] for rating video quality degradations on video material extracted from online games. In the SAMVIQ test paradigm, participants compare a game scene which has been processed

with different video transmission conditions on a continuous rating scale, by directly switching between the different processing conditions. Whereas such a method will be able to resolve small differences in video quality, other quality features can obviously not be quantified. In addition, it is still unknown how the judgments obtained in such a passive setting compare to the ones obtained in an interactive setting. In order to gain such knowledge, it is indispensable to carry out both active and passive tests in parallel on material extracted from the same games. As long as this knowledge is missing, it may be advisable to carry out both active and passive tests in combination, as it has also been proposed in [9].

The aforementioned methods are either based on a small selection of scenes, or break down the game in even smaller pieces (10 seconds). Thus, some of the main gaming QoE features such as flow, competence and immersion can probably not be well established in these methods. Another approach would be to let players play the game for a longer time period (approximately 10 minutes). Game characteristics (e.g. game success) could be used to identify scenes which could be used to let the gamers rate the corresponding material either in retrospect (video play back), or the rating is triggered during the game after a certain time period which corresponds to a scene.

B. Evaluation environment

Classical multimedia lab tests are commonly carried out in somewhat "neutral" environments, such as sound-shielded rooms with daylight imitation. Whereas this environment creates controlled conditions for each participant, it is obviously not representative for a real-life gaming situation. In particular in the case of mobile gaming on portable devices (smartphone, tablets) such an environment may generate misleading results with respect to the impact of device and display size on sub-aspects of QoE. Experiments described in [11] showed a significant impact of display size on several QoE dimensions, whereas the impact of the usage environment – neutral lab room versus simulated metro environment – showed no significant influence. It was concluded that the pure physical simulation, consisting of back-ground noise and space restrictions for the participants, might have missed a social impact of co-travellers in a real metro. Experiments should clarify the differences in perception of games in real-life and laboratory use and determine how realistic simulated usage environments need to be in order to generate ecologically valid results.

A relatively new paradigm for gaming QoE evaluation is to conduct crowdsourcing experiments, through online platforms like Amazon Mechanical Turk, Microworkers or Crowdee. Such a paradigm offers the potential for a large number of geographically distributed test subjects to participate in subjective user studies in their realistic environment. Consequently, a diversity of environmental conditions is represented which may be realistic for real-life gaming activities as well, and which cannot be tested to such an extent in the lab. In turn, also the network conditions are not fully controllable, and users may operate the games from different devices which equally affect gaming QoE. First experiences

[26] illustrate the potential of such studies for increasing the ecological validity of the evaluation.

C. Characteristics of test participants

The taxonomy of [2] lists 4 user factors which potentially influence gaming QoE: Experience, playing style, intrinsic motivation, as well as other static or dynamic user factors such as age, gender, native language, current emotional status, boredom, distraction, or curiosity. Experience can be divided into general game experience (i.e., how much time the test participant spends weekly playing digital games), experience related to game type (i.e., how much experience does the test participant have related to the type of the game under test) and experience playing the exact game under test. For example, a player who predominantly plays games of one genre may not be experienced and skilled at playing games from another genre, and should hence be considered a player with low experience for that particular genre. Additionally, players that have experience playing a particular game are in the best position to rate the game performance, as they know exactly how the virtual world should behave.

The impact of experience on QoE scores has been experimentally confirmed in the case of Massively Multiplayer Online Role-Playing Games (MMORPGs) [12]. Whereas the playing experience can be expected to significantly vary with age, playing style might be more related to a player's personality, which could be measured e.g. with the "Big 5" inventory, using standard screening questionnaires such as NEO-FFI [13] or alike. Player experience is commonly classified with regard to the number of weekly hours spent with playing activities, and a typical threshold to differentiate between casual players and experts seems to be around 10h per week. Gender dynamics has been shown to affect playing experience of women in [14]; it can be expected that this factor is moderated by playing style or personality.

Playing motivation is commonly steered in laboratory tests with the help of scenarios, by giving precise playing tasks to the participants. In addition, Self-Regulation Questionnaires for assessing motivation based on the Self Determination Theory might be helpful for this purpose, such as [15]. Also, frameworks for motivation assessment have been developed for certain game types such as MMORPG's [16].

In general it seems to be not completely clear yet what user characteristics are valid measures to distinguish be-tween groups that have varying gaming QoE due to certain user traits or states.

So far, the emotional state has been considered as a user state which can result from and influence the gaming experience. However, the emotional state is also quite important as an input variable, i.e. as a human influence factor. In the natural gaming situation, a game will be initiated when a certain emotional state is present. Therefore, a valid result can only be obtained by considering the incoming emotional state of player (i.e., emotional state at the time players arrive for the test session). It can be evaluated on the one hand by using subjective methods, e.g., with the already mentioned Self-Assessment Manikin, or on the other hand by using

physiological indices such as the alpha asynchrony index which is a neural indicator of liking/disliking [17].

Workload could be assessed by questionnaires or physiological methods. As subjective measure an adapted version of the NASA-TLX can be used, but also other measures, such as performance indices can be employed. Boredom and/or fatigue could be assessed using EEG-derived physiological measures, such as activities on the alpha or theta bands [18]. Experiments clarifying the applicability of such measures in a gaming context are still missing.

D. Selection of games

Players usually have preferences for specific games or game types, and it can be expected that such a preference will be reflected in gaming quality assessments (similar to video content in video quality assessments). However, it is yet unclear how such preferences can be measured, and how games can be selected or constructed which reflect different, controlled degrees of preference. User studies on these aspects might help to clarify, and to develop potential screening questionnaires which might be used in order to select test participants and/or corresponding games. These questionnaires might be combined with the questionnaires for gaming experience, as players most likely will prefer the games they actually play.

In contrast to the semantics of the game content, the relationships between games and technical system characteristics have been an object of numerous studies. Taking the example of a cloud game where video content is streamed over the network, the temporal and/or spatial complexity of the video material has an impact on the video quality achievable with different video codecs. Indices such as the Spatial Information Index or the Temporal Information Index described in [5] may be adequate metrics to quantify these impacts. Further, the interactivity of the game content was shown to influence game experience, e.g. in conjunction with network bandwidth or delay [19]. Hence, a classifier of the degree of interactivity for each utilized game is necessary and should be reported along the results of the QoE evaluation. In turn, it was also shown that the game type may have an impact on network traffic, and in this way also on gaming QoE [20]. Such indirect relationships have to be investigated further to come up with a meaningful categorization of games.

E. Standardized game platform

A key issue in conducting gaming QoE studies is the design of the game platform itself. The main tasks attributed to digital game platforms include the following: gathering the player's input, calculating the virtual world state (based on the defined logic of the virtual world), rendering the virtual scene, and displaying the resulting real-time video to the user. Depending on the distribution of these tasks on the entities of the gaming platform, we differentiate between:

- Local games – all the tasks are performed on the local computer or game console.

- Online games – the calculation of the virtual world state is done on a remote server while other tasks

(input, rendering of the virtual scene, and displaying) are performed locally. The player's inputs and information regarding the next virtual world state are transferred to and from a remote server, respectively.

- Cloud games – the calculation and the rendering of the virtual scene are done remotely, while gathering the player's input and displaying are done locally. The player's inputs are transferred from the client to the remote server. Video or 3D graphical information is streamed from the remote server to the client. In this case, the server also performs encoding of the video stream, while the decoding is done locally.

For each of the above-listed scenarios, a standardized game platform should be defined when conducting QoE tests, due to significant differences in processing and network requirements.

Another important aspect to address is the differentiation of user devices, which can be grouped into: handheld consoles, mobile phones, tablets, consoles, and personal computers. As some elements of the technical platform (such as computational and graphical processing capacity of involved devices) undergo rapid technological development, a standardization of particular devices is not deemed worthwhile. Instead, technical parameters have to be selected and standardized which describe user-perceivable aspects of the end-to-end gaming platform including the end user devices. Such parameters which may be of interest in QoE studies include (some parameters may apply only for some types of task distributions):

- Input characteristics and devices (touchscreens, controllers, keyboard and mouse)

- (Minimum) frame rate (usually an average number of frames per second)

- Video jerkiness (whether and at which frequency and duration a stall in the video display has occurred)

- Network bandwidth

- Network delay and jitter (i.e. network delay variation)

- Network packet loss

- Server processing delay

- Game type/genre

- Game perspective/ camera (first person linear perspective, third person linear perspective, and third person isometric perspective)

- Spatial and Temporal Perceptual Information of the game's visual output (SI, TI) [5]

Of further interest is the effect of dynamically changing values for these parameters.

F. Questionnaires and Interviews

Game experience is usually enquired by means of questionnaires, which are filled out at the end of each experimental condition. The Game Experience Questionnaire (GEQ), which was developed, implemented and validated during the "Fun of Gaming" (FUGA) project, has been so far a popular tool to assess game experience. Complete information about the GEQ can be found in [21], However, some concerns have recently been raised regarding the reliability of the GEQ core module [22] and about the inadequate length of the questionnaire (42 items) when using within-subjects experimental designs. As a result, shorter questionnaires aimed to asses flow experience are gaining popularity such as the scale reported in Engeser and Rheinberg [23] or the 12 items flow scale reported by Sherry and colleagues [24]. The scale of [23] collapses characteristics of flow into two dimensions labeled absorption (e.g., involvement, distorted sense of time, optimal challenge, absent-mindedness) and smooth and automatic running (e.g., concentration and focus, control, clarity) while in the 12 items flow scale reported by Sherry and colleagues, four questions are designed to measure boredom resulting from game play, 4 questions to address frustration resulting from game play, and 4 to assess the subjective experience of flow (see [25]).

Questionnaires with appropriate scales are not the only way to evaluate QoE. In addition to such quantitative measurement instruments, qualitative interviews may help to find out aspects which affect QoE most. In [10], it is proposed to use guided interviews which address four aspects related to the visual quality: visual fatigue, fluidity, visual discomfort, and gameplay (i.e. playability of the game). These aspects are solicited in structured interviews and then rated by the evaluator on dedicated rating scales. Other aspects might be addressed through questions asked by the evaluator.

IV. NEXT STEPS AND FUTURE WORK

The discussion above should have made clear that we consider the current status of the Recommendation draft as insufficient for an approval. In particular, we foresee work to be necessary to enhance the following points of the draft:

A. Evaluation procedure

The Recommendation draft should include both passive and interactive procedures. Whereas interactive procedures are deemed indispensable to ensure ecological validity, passive procedures should complement these, especially for assessing audio-visual streaming of the game content. The characteristics of the scenes needs to be described for both procedures, and anchor conditions should be included in the evaluation which allow to relate results from both procedures to each other.

B. Evaluation environment

We foresee most gaming QoE evaluations to be carried out in the lab, and as long as no significant difference to field test results can be shown, we consider this as an acceptable paradigm. However, information should be included into the Recommendation on factors which potentially have an impact on the results, as long as no sound evaluation of those factors is available.

C. Characteristics of test participants

As there are indications of a relatively large number of characteristics to be influential on gaming QoE, we propose to compile a screening questionnaire which quantifies these characteristics as far as possible. Additional questionnaires could quantify the player's emotional state (before and after the gaming activity), and might be complemented by physiological measurements in the future.

D. Selection of games

Although the game type or specific implementation details have been shown to moderate the effects of e.g. network impairments, no established classification of games exists so far. This complicates the long-term repeatability of tests, because player expectations towards technological sophistication and aesthetical style change, and because particular games quickly become outdated.

E. Standardized game platform

Technical platform specifications for the mentioned task distributions and different classes of user devices have to be developed. Such specifications should be understandable to non-expert users who need to perform subjective evaluations of gaming systems. To reach this aim, the essential technical and procedural parameters need to be defined first, and these parameters should remain valid despite the rapid technological change in the game domain.

F. Questionnaires and interviews

Many of the presented questionnaires are quite redundant, or at least they seem to capture overlapping aspects of gaming QoE. For example, there is a flow dimension in the GEQ, the flow short scale, and also possible physiological measures that are intended to capture similar variations. For a finalized Recommendation, it seems to be one important step to identify the redundant or overlapping methods, and to include a suggestion of what minimal method mixture is recommended. Such a method mixture should also aim at ensuring that all important data is captured, while the time needed for the experiments should be kept as short as possible.

G. Final considerations

Despite these open issues, we think that the aim of an open Recommendation for gaming QoE evaluation is a worthwhile task. The common understanding reached by setting up such a Recommendation will increase the comparability of evaluation results across different laboratory environments, and will help to collect a sufficient amount of subjective QoE measurements which could be used as a basis for instrumental models for predicting Gaming QoE, e.g. from parametric descriptions of the gaming set-up. We hope that the methods included in the final Recommendation text will provide sufficient guidance for researchers and practitioners alike to make evaluation results comparable across evaluation sites, and to finally improve gaming QoE for the end user.

The valued reader interested in commenting or contributing to the recommendation can contact the responsible editor under sebastian.moeller@telekom.de.

REFERENCES

[1] ITU-T Contr. COM 12-203, "Proposal for a Draft New Recommendation on a Subjective Evaluation Methods for Gaming Quality (P.GAME)", Source: Deutsche Telekom AG, ITU-T SG12 Meeting, 2-11 Sept. 2014, Geneva.

[2] Möller, S., Schmidt, S., Beyer, J., "An Overview of Concepts and Evaluation Methods for Computer Gaming QoE", in: *Fifth Int. Workshop on Quality of Multimedia Experience 2013 (QoMEX 2013)*, Klagenfurt, 3-5 July, 236-241, 2013.

[3] M. Csikszentmihalyi, "Flow: The Psychology of Optimal Experience", Harper & Row, 1990.

[4] Qualinet White Paper on Definitions of Quality of Experience, European Network on Quality of Experience in Multimedia Networks and Services (COST Action IC 1003), Lausanne, version 1.2, March 2013.

[5] ITU-T Rec. P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications", International Telecommunication Union, Geneva, 2008.

[6] ITU-T Rec. P.911, "Subjective Audiovisual Quality Assessment Methods for Multimedia Applications. International Telecommunication Union, Geneva, 1998.

[7] M.M. Bradley, P.J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential", *J. Behavior Therapy and Exp. Psychiatry 25*(1), 49-59, 1994.

[8] W. Ijsselsteijn, Y. De Kort, K. Poels, F. Bellotti, A. Jurgelionis, "Characterising and Measuring User Experiences in Digital Games", in: *Proc. Int. Conf. on Advances in Computer Entertainment Technology*.

[9] ITU-T Contr. COM 12-166, "QoE and Perceptive Quality of Video Game in Passive Mode", Source: Orange SA, ITU-T SG12 Meeting, 2-11 Sept. 2014, Geneva.

[10] ITU-R Rec. BT.1788, "Methodology for the Subjective Assessment of Video Quality in Multimedia Applications", International Telecommunication Union, Geneva, 2007.

[11] J. Beyer, V. Miruchna, S. Möller, "Assessing the Impact of Game Type, Display Size, and Usage Context on Mobile Gaming QoE", in: *Proc. 6th Int. Workshop on Quality of Multimedia Experience (QoMEx 2014)*, Sept. 18-20, Singapore, 2014.

[12] M. Sužnjević, L. Skorin-Kapov, M. Matijašević, "The Impact of User, System, and Context factors on Gaming QoE a Case Study Involving MMORPGs", in: The 12th Annual Workshop on Network and Systems Support for Games, 2013.

[13] P.T. Costa, R.R. McCrae, "Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual", Psychological Assessment Resources, 1992.

[14] L. Vermeulen, E. Núñez Castellar, J. Van Looy, "Challenging the other: Exploring the role of opponent gender in digital game competition for female players", Cyberpsychology, Behavior, and Social Networking, 17(5), 303-309, 2014.

[15] R.M. Ryan, J.P. Connell, "Perceived locus of causality and internalization: Examining reasons for acting in two domains", *Journal of Personality and Social Psychology 57*, 749-761, 1989.

[16] N. Yee, "Motivations for Play in Online Games", CyberPsychology & Behavior 9.6: 772-775, 2006.

[17] S. Arndt, J.-N. Antons, R. Gupta, K. Laghari, R. Schleicher, S. Möller, T. H. Falk, "The Effects of Text-to-Speech System Quality on Emotional States and Frontal Alpha Band Power," in: Proc. of International IEEE/EMBS Conference on Neural Engineering (NER), pp. 489–492, 2013.

[18] J.-N. Antons, R. Schleicher, S. Arndt, S. Möller, and G. Curio, "Too Tired for Calling? A Physiological Measure of Fatigue Caused by Bandwidth Limitations," in Proceeding of International Workshop on Quality of Multimedia Experience (QoMEX), pp. 63–67, 2012.

[19] S. Möller, D. Pommer, J. Beyer, J. Rake-Revelant, "Factors Influencing Gaming QoE: Lessons Learned from the Evaluation of Cloud Gaming Services", in: *Proc. 4th Int. Workshop on Perceptual Quality of Systems (PQS 2013)*, Wien, 2-4 Sept, 173-177, 2013.

[20] M. Suznjevic, J. Bayer, L. Skorin-Kapov, S. Möller, N. Sorsa, "Towards Understanding the Relationship Between Game Type and Network Traffic for Cloud Gaming", in: *Proc. 2014 IEEE Int. Conf. on Multimedia and Expo Workshops (ICMEW)*, Jul. 14-18, Chengdu, 2014.

[21] W. IJsselsteijn, K. Poels, Y.A.W. de Kort, "The Game Experience Questionnaire: Development of A Self-report Measure to Assess Player Experiences of Digital Games", TU Eindhoven, Eindhoven, 2008.

[22] F. De Grove, J. Van Looy, C. Courtois, "Towards a serious game experience model: Validation, extension and adaptation of the GEQ for use in an educational context", Playability and Player Experience 10: 47-61, 2010.

[23] S. Engeser, F. Rheinberg, "Flow, Performance and Moderators of Challenge-skill Balance", Motivation and Emotion 32, 158-172, 2008.

[24] J.L. Sherry, S. Rosaen, N. Bowman, S. Huh, "Cognitive skill predicts video game ability", in: Proc. Annual Meeting of the International Communication Association, Dresden, 2006.

[25] R. Weber, R. Huskey, "Attentional capacity and Flow experiences: Examining the attentional component of synchronization theory", in: Proc. Annual conference of the International Communication Association, London, 2013.

[26] K.-T. Chen, C.-C. Wu, Y.-C. Chang, C.-L. Lei, "A crowdsourceable QoE evaluation framework for multimedia content", in: Proc. 17th ACM Int. Conf. on Multimedia (MM'09), ACM, New York NY, pp. 491-500.