Factors influencing video Quality of Experience: measurements and theoretical model

KAMIL KONIUCH, AGH University of Science and Technology, Institute of Information and Communication Technologies, Poland

Over the last decade, user subjective perception attracted interest in video quality studies. Researchers infer about the Quality of Experience (QoE) based on participants' statements, behaviors, and psychophysiological reactions. This is because an objective evaluation of QoE is impossible due to its subjective nature. Thus, clear operationalization of variables in QoE studies is crucial. For that purpose theoretical background is necessary. Current descriptive models of QoE consist of many strongly correlated variables and omit the role of essential factors such as content. In my Ph.D., I investigate factors influencing QoE which are important from the perspective of the user. To be able to conclude about those multiple, complex variables I am working on a new video QoE model inspirited by Structural Causal Models. This model help to generate hypotheses and operationalize variables. It also could be used to design statistical analyses for complex datasets.

Additional Key Words and Phrases: Quality of Experience, QoE influence factors, ecological validity, Causal Models, Modeling

1 INTRODUCTION

The growing popularity of video services generates a high demand for transmission bandwidth. One of the responses from providers is the use of a user perception study to distinguish the efficacy of different compression methods. For example, user assessments helped build an objective full-reference video quality metric called VMAF which allows for automatic prediction of subjective ratings [15]. These types of metrics are necessary for the evaluation of new video compressing methods. Furthermore, quality-aware service optimization may lead to a more efficient energy consumption of video-on-demand services [14]. The sustainability of video services and their impact on climate is a growing concern in studies [8], [3], [6].

Nevertheless, measuring quality based on human perception brings some challenges. For instance, VMAF is based on standardized laboratory experiments with short, soundless clips as video stimuli [5]. A series of single subjective judgments are collected on a 5-point Absolute Category Rating scale. The same content is presented multiple times but on different levels of distortion. The goal of this methodology is to limit the number of factors that could influence ratings. This approach is focused on providing reliable, repetitive results.

In consequence, the context of the study is far different from how people usually consume video. In other words, the ecological validity of mentioned studies is limited. In effect, appreciation of video quality in a natural environment might be different from the predictions made based on a strict laboratory experiment. By adding more factors that are important from a user perspective we could predict everyday satisfaction from video quality better.

Moreover, human assessments might be moderated by numerous factors even in a controlled laboratory environment. Theoretical models play an important role in the classification of those factors by aggregating knowledge from previous studies. As I will present more broadly in the next section the influence of essential factors is systematically overseen in those models.

2 LITERATURE REVIEW

QoE, defined as *"the degree of delight or annoyance of the user of an application or service"* [4], is a latent variable. Therefore, we can only conclude about the QoE based on measurable variables such as self-report questionnaires, subjective scales, psychophysiological measurements, or the behavior of participants. This type of measurement is always an approximation of latent variables burdened with error. In a study of latent variables, the theoretical

IMX '22 - Doctoral Consortium, 2022, Portugal

2022 ACM International Conference on Interactive Media Experiences. Copyright held by the owner/author(s).

background is crucial. Theoretical models allow for comparable operationalization between studies and are a source of hypothesis. A good theory is also falsifiable and measurable. To be able to achieve this requirements, simplicity is necessary[10].

The main source of the theoretical background for QoE was described in[4], where authors define it as: *"..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."*. This definition takes into account more human-related variables such as delight or enjoyment, expectations, personality, and current state than technical namely utility. As with most psychological variables these are latent too. Additionally, QoE is influenced by numerous other factors, including system-, context-, and user-related factors [19] which could correlate with each other. In consequence, models build based on this theory are very complex e.g.[21].

The complexity of models has several consequences. Firstly, it generates a great need for building a strict control laboratory environment to provide reliable results. Every procedure that would be less strict should be unreplicable due to the vast amount of uncontrolled variables influencing the measurement. Secondly, measuring factors influencing QoE (IFs) is very hard with quantitative methods due to their strong correlation. In effect, every experimental setup that is more observatory could be uninterpretable. On top of that data obtained from real networks should be too noisy to be useful to conclude about QoE. However, research practice shows that it is possible to draw useful conclusions from for example crowdsource QoE experiments [20]. Moreover, there is a new effort in linking QoS metrics with user behavior to better predict QoE [13]. In this case, the authors underline the strong need for a novel approach with better IFs classification. Such a theoretical gap between user behavior and quality was also stated before in[21]. Yet, both of mentioned approaches are still limited to rearranging current classification or they even add more complexity. In [21] authors added (post-)conscious and affective behavior formation to the perception of the quality model. Both of these "submodels" consist of multilevel correlations and latent variables. The model described in [13] adds new service-related influencing factors divided into 5 subcategories.

Even though, most descriptive QoE models omits the role of content in generating delight or irritation or in quality formation process[4], [17], [16], [22], [18]. Moreover, models describing factors influencing QoE (IFs) oversees the content or focuses on its' technical aspects only [19][7], [18].

However, a survey of predictive QoE models [1] shows that on 13 reviewed content-related models, only in one case content was not influential. Moreover, in the questionnaire study described in section Own research we ask users about factors influencing their QoE. Among the most influential factors, interest in content was stated as most important for VOD and live stream. These results picture the important role of the content in generating delight or annoyance, which is overseen by mentioned complex models.

On the other hand, much simpler models can be found in QoE-related literature. For instance, in [2] authors present QoE browsing models as part of the literature review. This model has a less complex structure and provides evidence for every connection. Similarly, much simpler models including the influence of content could be found in video user experience research [9]. Similarly, in my work, I am building an uncomplicated model, including the most important factors from the perspective of the users. Additionally, I plan to use participant behavior as the outcome of QoE. Relations between every variable are unidirectional and have causal meaning. Moreover, every relation, or lack of it, is strongly augmented. Two main requirements that I have is falsifiability and measurability of the model. They are necessary to evaluate the model with real data in the future.

The main source of inspiration for this content-based QoE model was Structural Causal Models (SCM) presented by Pearl[12]. This method provides an opportunity to use qualitative causal structure to explain phenomenons constructed with latent variables. This structure can be represented in form of a graph consisting of nodes and arrows. Arrow drawn from A to B represents the assumption that A *may be* a direct cause of B. In example 1 delight or annoyance might be *caused* by Quality. Additionally, arrow representation does not imply that process

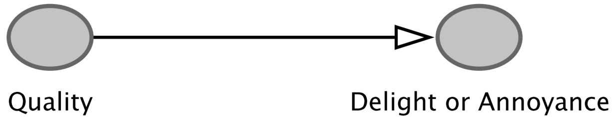

Fig. 1. Example of causal relation presented on graph.

is simple. Relations between A and B can be complex, multi-staged, and non-linear. What we assume is the direction of the relation. Moreover, by drawing an arrow from A to B we do not assume that A is the only thing that *cause*s B. In fact, this approach assumes that every variable could be influenced by unspecified factors not represented in a graph. This allows to include measurement error and the influence of unknown factors in the model.

On the other hand, the absence of an arrow represents a stronger assumption, namely, the lack of relation between variables represented by nodes. In fig. 1 amount of delight or irritation does not change the quality. Unless not directly. Every lack of arrow between variables has to be well-argued. That is why diagrams are especially useful in verifying hidden assumptions of the models. From this perspective, current QoE models are full of assumptions that are impossible to validate. Structural Causal Models are not the only statistical procedure based on graphs. Generally, methods using path analysis are widely use for model testing and are very good described [11].

In summary, current models of QoE are complex and full of hidden assumptions. Their goal is rather descriptive, so it is impossible to verify them with data. Moreover, role of content is systematically omit and their ability to explain user behavior is limited. In ecologically valid studies both content and behavior are crucial to understand users delight or annoyance. That is why I am working on new video QoE model which could be use not only to describe QoE but is also measurable and falsifiable. This will help to challenge it against the data from experiments in future.

3 OWN RESEARCH

My Ph.D. is part of the project TOWARDS BETTER UNDERSTANDING OF FACTORS INFLUENCING THE QOE BY MORE ECOLOGICALLY-VALID EVALUATION STANDARDS (TUFIQoE) funded by the Norwegian Financial Mechanism 2014-2021 under project 2019/34/H/ST6/00599. As a member of an international, cross-disciplinary team I collaborate with experts in communication, psychologists, IT, and telecommunication specialists. Our goal is to understand better which factors play a role when people use video services. I am participating in over 10 experiments but 4 of them are most crucial for my dissertation.

3.1 Research questions

In the course of work, I formulated three main research questions.

- What are the most important factors influencing QoE from perspective of user?
- What type of behavior could be use to predict annoyance of the user?
- How accurately objective metrics can predict subjective QoE assessments of a ecologically valid content?

To be able to answer those questions, I will propose a model combining above mentioned factors.

3.2 Questionnaire

To choose which factors we should look into in the laboratory, we started with a questionnaire study. This broad screening method was created especially for this purpose. 140 participants rated 29 various parameters that influence their past experience with video services. We asked them how strongly these factors influenced their satisfaction or irritation. Among the most influential scales, participants reported numerous content-related factors: interest in content, appreciation of the content, the emotion evoked, or importance of the content. A detailed paper describing this research is in the review process. In the future, we plan to use our questionnaire on a massive scale.

3.3 Experiment Your YouTube our Lab

Based on these conclusions from the questionnaire, we built a laboratory setup that allows users to watch the content of their choice on the YouTube service. During watching, a special browser extension changes the bit rate. In addition, we track "nerd statistics" provided by YouTube. What is our main innovation, we measure participants' interest in the content, quality of the video, and most interestingly participants' behavior. We track behavioral data such as comments reading, rewinding, change of content, etc. With this data, we hope we will be able to answer questions about behaviors indicating the annoyance of the user. At this stage of work, we formulated our first conclusions after the pre-test in our laboratory. We also work on a system of classification of the participants' behavior.

3.4 Experiment Watching with friends

The next step will be conducting a similar experiment but in pairs and with Netflix instead of YouTube. We know from our questionnaire that Netflix is most popular for watching with friends purpose, so this would be more natural for testers. Participants will come to our laboratory in groups of two. Before the experiment, we will ask them to choose TV series that they have not seen yet but like too. In the first stage, they will be watching one episode in separate rooms. Secondly, participants will see the second episode together. During watching, we will manipulate the bit rate and display the ACR scale to collect participants' opinions about quality. Apart from the operationalization of the influence of other people on quality ratings, we hope to understand better how people access quality. That is why our participants will have to negotiate on ratings at loud in the second part of the experiment. At this stage, we are working on an extension that will allow us to manipulate bit rate, capture subjective ratings of quality, and most interestingly download the VMAF provided by Netflix. This approach will help us investigate the performance of this objective metric in a real usage scenario with interesting content.

3.5 New video QoE Model

As I mentioned in the literature review section it is impossible to use current QoE models to analyze the data. Especially, in more ecologically valid research setups. That is why my main objective for the Ph.D. is to propose and validate the new video QoE model. We have a generalized model that we were able to specify for every type of experiment in the TUFIQoE project. The main components are content, quality of service, quality, level of delight or annoyance, and behavior. For every experiment, we specify the context and add adequate variables to the model. At this stage, we are preparing an opinion paper with a detailed description of the model. In the future, we would like to challenge it against data from the experiments mentioned above. This is possible with path analysis or Causal Structural Models. We also hope that our work will be an inspiration for future studies and a source of a new hypothesis.

4 CONCLUSION AND FUTURE WORK

The questionnaire study helped us investigate how important for users are factors related to the content of the video. Similarly, the first analysis of participants' behavior during Your YouTube Our Lab experiment, shows potential patterns of behavior depending on the type of the content. It is too soon to draw strong conclusions tho. Nevertheless, the number of interactions with the interface during this experiment and the variability of the behavior was very high. This fact pushed our next experiments in the direction of less interactive platforms with longer content. This shift will enable better investigation of objective quality matrices performance with natural, interesting content.

We working on a model combining main factors and behavior as the result of QoE in one path diagram. We hope that our model will help to answer questions like how important is quality or why people stop watching? Is it due to content or because the video quality is insufficient? We are convinced that answers to those questions will help provide more sustainable solutions for video streaming in the future. Moreover, content recommendation algorithms could be even better trained if they could "recognize" the loss of engagement caused by video quality drops. Last but not least, we hope that our model will be a source of inspiration for new hypotheses and paradigms in the QoE study. Building a new framework might structure discussions and education of video QoE in a creative way. This new structure could possibly facilitate the dialogue between specialists from different domains relative to video QoE.

REFERENCES

[1] Mohammed Alreshoodi and John Woods. 2013. Survey on QoE\QoS correlation models for multimedia services. *arXiv preprint arXiv:1306.0221* (2013).

[2] Sabina Baraković and Lea Skorin-Kapov. 2017. Survey of research on Quality of Experience modelling for web browsing. *Quality and User Experience* 2, 1 (2017), 1–31.

[3] Altanshagai Batmunkh. 2022. Carbon Footprint of The Most Popular Social Media Platforms. *Sustainability* 14, 4 (2022), 2195.

[4] Kjell Brunnström, Sergio Ariel Beker, Katrien De Moor, Ann Dooms, Sebastian Egger, Marie-Neige Garcia, Tobias Hossfeld, Satu Jumisko-Pyykkö, Christian Keimel, Mohamed-Chaker Larabi, et al. 2013. Qualinet white paper on definitions of quality of experience. (2013).

[5] RECOMMENDATION ITU-R BT. 2002. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union* (2002).

[6] Jigna Chandaria, Jeff Hunter, and Adrian Williams. 2011. The carbon footprint of watching television, comparing digital terrestrial television with video-on-demand. In *Proceedings of the 2011 IEEE International Symposium on Sustainable Systems and Technology*. IEEE, 1–6.

[7] Sebastian Egger, Peter Reichl, and Katrin Schoenenberg. 2014. Quality of experience and interactivity. In *Quality of experience*. Springer, 149–161.

[8] Oche Ejembi and Saleem N Bhatti. 2015. The energy cost of your Netflix habit. In *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*. 225–226.

[9] WK Eric. 2012. User experience on mobile video appreciation: How to engross users and to enhance their enjoyment in watching mobile video clips. *Technological Forecasting and Social Change, In Press, Corrected Proof, Available online April* 18 (2012), 2012.

[10] Xavier Gabaix and David Laibson. 2008. The seven properties of good models. *The foundations of positive and normative economics: A handbook* (2008), 292–319.

[11] G David Garson. 2013. *Path analysis*. Statistical Associates Publishing Asheboro, NC.

[12] Madelyn Glymour, Judea Pearl, and Nicholas P Jewell. 2016. *Causal inference in statistics: A primer*. John Wiley & Sons.

[13] Fatima Laiche, Asma Ben Letaifa, and Taoufik Aguili. 2021. QoE-aware traffic monitoring based on user behavior in video streaming services. *Concurrency and Computation: Practice and Experience* (2021), e6678.

[14] Dayoung Lee, Jungwoo Lee, and Minseok Song. 2019. Video quality adaptation for limiting transcoding energy consumption in video servers. *IEEE Access* 7 (2019), 126253–126264.

[15] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara. 2016. Toward a practical perceptual video quality metric. *The Netflix Tech Blog* 6, 2 (2016).

[16] Sebastian Möller, Marcel Wältermann, and Marie-Neige Garcia. 2014. *Features of Quality of Experience*. Springer International Publishing, Cham, 73–84. https://doi.org/10.1007/978-3-319-02681-7_5

[17] Alexander Raake and Sebastian Egger. 2014. *Quality and Quality of Experience*. Springer International Publishing, Cham, 11–33. https://doi.org/10.1007/978-3-319-02681-7_2

[18] Peter Reichl, Sebastian Egger, Sebastian Möller, Kalevi Kilkki, Markus Fiedler, Tobias Hoßfeld, Christos Tsiaras, and Alemnew Asrese. 2015. Towards a comprehensive framework for QoE and user behavior modelling. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.

[19] Ulrich Reiter, Kjell Brunnström, Katrien De Moor, Mohamed-Chaker Larabi, Manuela Pereira, Antonio Pinheiro, Junyong You, and Andrej Zgank. 2014. Factors influencing quality of experience. In *Quality of experience*. Springer, 55–72.

[20] Werner Robitza, Alexander M Dethof, Steve Göring, Alexander Raake, André Beyer, and Tim Polzehl. 2020. Are you still watching? streaming video quality and engagement assessment in the crowd. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.

[21] Werner Robitza, Sabine Schönfellner, and Alexander Raake. 2016. A theoretical approach to the formation of quality of experience and user behavior in multimedia services. In *5th ISCA/DEGA Workshop on Perceptual Quality of Systems*. 39–43.

[22] Marwin Schmitt, Dick CA Bulterman, and Pablo S Cesar. 2018. The contrast effect: QoE of mixed video-qualities at the same time. *Quality and User Experience* 3, 1 (2018), 1–17.