On predicting individual video viewing experience
The value of user information

Zhu, Yi

DOI
10.4233/uuid:8c1c8eae-43ca-405b-b075-6f0a7a8d2ab9

Publication date
2019

Document Version
Final published version

Citation (APA)
Zhu, Y. (2019). *On predicting individual video viewing experience: The value of user information*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:8c1c8eae-43ca-405b-b075-6f0a7a8d2ab9

On Predicting Individual Video Viewing Experience:

The Value of User Information

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on Tuesday, 2 July 2019
at 12:30 hours

By

YI ZHU

Master of Engineering in Information and Communication Engineering, Southeast University, China
Born in Wuxi, China

This dissertation has been approved by the

promotor: Prof. dr. A. Hanjalic
promotor: Prof. dr. I. E. J. Heynderickx

copromotor: Dr. J. Redi

Composition of the doctoral committee:

Rector Magnificus	Chairman
Prof. dr. A. Hanjalic,	Delft University of Technology, promotor
Prof. dr. I. E. J. Heynderickx,	Eindhoven University of Technology, promotor
Dr. J. Redi,	Delft University of Technology, copromotor
Independent members:	
Prof. dr. D. C. A. Bulterman,	Vrije Universiteit Amsterdam
Prof. dr. M. A. Larson,	Radboud University, Nijmegen
Prof. dr. P. LeCallet,	University of Nantes, France
Prof. dr. H. de Ridder,	Delft University of Technology

The research reported in this thesis is supported by Chinese Scholarship Council (No. 201206090028).

ISBN/EAN: 978-94-6384-052-1

Cover was designed by Qian Liu.
An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

To my family

SUMMARY

Experience prediction is one key component in today's multimedia delivery. Knowing user's viewing experience allows online video service providers (e.g., Netflix, YouTube) to create value for their customers by providing personalized content and service. However, individual experience prediction is a challenging problem since viewing experience (defined as Quality of Experience in this thesis) is a multifaceted quantity and it is rather personal and subjective. The existing methods for quantifying Quality of Experience (QoE) target at estimating how the video quality is perceived by users, neglecting the hedonic part of experience (the degree of enjoyment of a user watching a video). Quite naturally, these methods consider only factors related to video perceptual quality (purely from video), which is insufficient to properly assess viewing experience. The research reported in this thesis attempts for the first time at shifting the paradigm for perceptual quality modeling, towards measuring and predicting the level of enjoyable viewing experience a user has with a video. In particular, it focuses on exploiting the potential value of user factors (information from users) and investigate their influences on QoE prediction.

The goal of this thesis is to develop a feasible method for predicting the individual viewing experiences in terms of perceptual quality and enjoyment by taking multiple influencing factors into account. Here, the influencing factors are taken from both video (e.g., related to perceptual quality) and user (user factors, e.g., interest. personality). We take three major steps to accomplish this goal. We first deploy a subjective experiment to understand the relationship between perceptual quality and enjoyment, and how their influencing factors form the final viewing experience. With a set of identified influencing factors, we then propose a new QoE prediction model which processes both user and video information to predict individual experience (i.e., either perceptual quality or enjoyment). We show that combining information from video and user enables better prediction performance as compared to only considering information from video related to perceptual quality. Our third step tackles the problem of reliable data collection for the individual QoE research. We developed an open-sourced Facebook application, named YouQ, as an experimental platform for automatic user information collection from social media while performing an online QoE subjective experiment. We show that YouQ can produce reliable results as compared to a controlled laboratory experiment, both in terms of QoE and of quantification of user factors and traits. As a result, a complete, feasible method for individual QoE prediction is presented in this thesis.

Based on the findings presented in this thesis, we reflect on the contribution and make recommendations for future research directions, which we think are substantial and promising for individual QoE prediction.

Contents

1

Introduction

1.This chapter consists of material adapted from J. Redi, Y. Zhu, H. de Ridder, I. Heynderickx. *How Passive Image Viewers Became Active Multimedia Users*, in C. Deng, L. Ma, W. Lin, K.N. Ngan: Visual Signal Quality Assessment pp 31-72, Springer, 2015.
2.While we realize that video in general consists of audiovisual content, we focus in this thesis on the visual data stream of a video only.

1

1.1. MOTIVATION

ACCORDING to Cisco's forecasts, video delivery will account for 82% of the overall internet consumer traffic by 2021 [1]. This enormous amount of data needs to be handled (i.e., captured, stored, transmitted, retrieved and delivered) in a way that meets the end-users' expectations. However, technology still shows limitations, such as limited spatial, temporal and bit rate resolution in displays, bandwidth and storage constraints, or error-prone transmission channels. As a result, video material is often delivered affected by impairments of different nature (from blocking artifacts due to compression over jerkiness to audiovisual errors due to packet loss) which disrupt the overall appearance of the video content. Impairments provoke a sense of dissatisfaction in the user [2, 3, 4], which, in turn, may decrease the willingness to pay for/use an online application, service, or device [5, 6].

As a consequence of the above, effort has been devoted to the development of technologies that can either prevent the appearance or reduce the visibility of impairments. Following the initial attempts based on the quantification of signal errors [7], it soon became clear that a better understanding of how users experience the consumption of video signals was necessary to properly optimize the video delivery and that this understanding could only be achieved through a collaboration between engineers and vision scientists [8]. Through this collaboration, dedicated psychometric techniques were developed [9, 10] and standardized [11, 12, 13, 14] to support a reliable quantification of visual quality from a subjective point of view (i.e., the perceived overall degree of excellence of a visual content item [11]). With these techniques, a large body of psychophysical data was collected to unveil the perceptual functions of the human visual system (HVS) that regulate the sensitivity to impairments [15]. The outcome of these experiments served as inspiration for designing objective/automatic visual quality assessment models [16, 17, 18], the output of which would steer the technology tuning towards impairment concealment (e.g., video restoration).

The common, underlying assumption for the visual quality assessment models is that having an understanding of the perceptual processes that regulate impairment sensitivity suffices to predict the impairments' annoyance and in this way also the overall quality of the video viewing experience [16]. This impairment-centric definition of visual quality (also referred to as perceptual quality throughout this thesis) has been considered effective for a long time [19, 20], but is now being challenged. With the development of new media technologies, the quality standards of media consumers have changed. Video viewing experience is tightly related to the usage of social media, as well as to mobile, interactive and immersive viewing systems. For example, users nowadays would watch video on one screen, while communicating with friends on social media (typically Facebook and/or Twitter) at the same time [21, 22]. New technology and new modalities of media consumption led to a change in the expectations regarding media services: perceptual quality is not any more seen as the sole aspect of the user's video viewing experience. Instead, video viewing experience is increasingly seen as a multidimensional concept, having other aspects as well, and also being influenced by various factors, such as visual semantics [23, 24], user personality [25, 26], preferences [27, 28, 29, 30], social [30, 31] and environmental context [5]. In this thesis, we attempt for the first time at shifting the paradigm for objective perceptual quality modeling, towards mea-

suring and predicting individual viewing experience with a given video by considering such multiple aspects and influencing factors.

1.2. THESIS SCOPE

1.2.1. ASPECTS OF VIDEO VIEWING EXPERIENCE

In this thesis, we approach the modeling and assessment of the individual video viewing experience under the formal and more general framework of ***"Quality of Experience"***. The concept of Quality of Experience (QoE) arose from the need to assess the quality of online media services from the point of view of user satisfaction [32]. QoE was initially defined by ITU [33] as "*the overall acceptability of an application or service, as perceived subjectively by an end-user*", and later on, in the Qualinet White Paper [34], 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*".

In the context of this thesis, the Qualinet definition points to the need to look at "*the degree of delight or annoyance*" of a user watching a video. This emphasizes the need to look at the entire video viewing experience, which reaches beyond the perceived video quality alone. Evidence has indeed shown that focusing purely on perceptual quality fails to describe user's overall video viewing experience [8, 34, 35]. For example, a user may have high viewing experience because of the enjoyable video content and despite low perceived perceptual quality. This reflects another aspect of QoE - ***enjoyment*** – that reflects the hedonic part of QoE, taking into account cognitive and affective aspects of the video viewing experience [36].

In view of the above, we consider two aspects of the individual video viewing experience in this thesis, ***perceptual quality*** and ***enjoyment***. The perceptual quality, as mentioned earlier, reflects the annoyance generated in the user by the presence of perceivable impairments in the media (e.g., blockiness, or disruptions due to packet loss in transmission). This aspect has been extensively researched in the past, in relation to e.g. video codec properties and/or network conditions [16]. Enjoyment, however, has not been sufficiently investigated, and neither its relation to perceptual quality. In particular, little is known about how the two aspects balance each other in determining the final QoE.

1.2.2. QOE: THE INFLUENCING FACTORS

In addition to different QoE aspects discussed above, there are also different factors that influence these aspects, their mutual relations and the QoE judgment in general. According to the Qualinet White Paper [34], these factors can be grouped into three categories: *system factors*, *user factors* and *contextual factors*. System factors include all characteristics of a system (or application or service) that contribute to the "*technically produced quality*" [37] of a media presentation. As such, system factors influence the presence of impairments in the video. In the most general formulation, system factors can address characteristics of the device on which a video is viewed (e.g., a mobile phone, PC, tablet or television [38]), of the technological signal variables (i.e., the video format or parameters in signal processing algorithms [4, 16]) and of the network configuration

(e.g., buffering time [32, 39]). Contextual factors describe all aspects of the environment within which the user consumes the media [34], e.g., physical location, economical aspects or social context. Finally, a user factor is defined as "*any variant and invariant characteristic of a human user*" that influences the viewing experience, such as demographic, personality, or interest related characteristics [34]. User factors determine for a large part the user "*personality and current state*" mentioned in the QoE definition [34].

System factors were investigated thoughtfully in the past since they were used to reflect and determine the perceptual quality of a video [33, 34]. However, similar to the conclusion that perceptual quality alone is not sufficient any more to reflect QoE, we also need to expand the scope of influencing factors in order to secure optimal QoE assessment in the modern technological and consumer landscape. The user factors as well as contextual factors were largely overlooked for a long time. In this thesis, we focus on ***user factors*** and investigate their influence on QoE.

There are many user factors that may potentially influence QoE. Taking the user's interest as an example, sports fans have been shown to tolerate incredibly low perceptual quality of a video when watching a match of a club they support [29]. Personality is also found to influence individual video viewing experience [26, 33]. For example, neurotic people are more likely to be affected by the pressure of being tested and therefore switch the TV channel or change the volume of the TV much more quickly than agreeable people in an experiment [25]. Gender [40], age [41], and affective state [30, 42] have also been shown to influence visual experiences. Despite these remarkable findings stemming from research fields such as Media Psychology and Human Computer Interaction, we are still far away from having a full picture of which user factors have a main impact on video experiences, and the extent to which they interact with system and context factors. This lack of completeness poses a major obstacle to the design of individual QoE assessment models.

1.2.3. METHODS TO ASSESS INDIVIDUAL VIDEO VIEWING EXPERIENCE

Like in the general case of QoE, individual video viewing experience can be assessed either using objective or subjective methods. Objective methods are based on computational models that link the physical properties of the media system directly to QoE. For example, poor network conditions (e.g., long buffering time) are assumed to lead to low QoE [43, 44]. Thus, efforts have been devoted to assessing QoE based on one (or several) network parameter(s), e.g., buffering time, delay time [32, 39]. Similarly, other methods focus on video signal quality assessment, operate at a decoded bitstream level to predict the appearance of impairments (e.g. due to the video codec used) and estimate how annoying these impairments are for the end user [16, 17, 18].

Subjective methods involve asking users to self-report their QoE with a given video. To date, they remain the most reliable way to quantify QoE. Subjective methods often quantify QoE in terms of Mean Opinion Scores (MOS) [11], indicating the average QoE perception across (a sample of) users. Since the MOS quantifies QoE in a commonly understandable way [45, 46], it is widely accepted for benchmarking the existing objective methods [45], which in fact, are in vast majority targeted at predicting MOS. However, the fact that a MOS is just an average makes it unsuitable for assessing individual video viewing experience.

In addition, the QoE research community has developed standardized methodologies and experimental settings to evaluate QoE subjectively, which are typically tied to a controlled lab environment [11, 12, 13, 14]. However, with the advent of mobile technology (smartphones and tablets) and internet-based video delivery, video is nowadays consumed in very different environments and the video viewing experience should be studied within realistic usage conditions in order to be properly optimized [5, 47]. Furthermore, a lab experiment usually fails to include a sufficient number and diversity of users. In order to compensate for these deficiencies, we are building in this thesis on the increasing effort in the QoE field to evaluate QoE via internet-based tools (e.g., via social media [48, 49] or crowdsourcing [50, 51]). Studies have shown that an internet-based experiment can, if carefully designed, lead to results being as reliable as those obtained in a controlled lab environment [24, 49, 50]. At the same time, it allows to recruit users from a larger, more diverse group and to let them interact with a video in their real-life conditions, leading to more representative results [52].

1.3. THESIS OBJECTIVE AND CONTRIBUTION

IN view of the discussion in the previous section, we define the main challenge addressed in this thesis as to ***develop a feasible method for predicting the quality of individual video viewing experiences that considers multiple QoE aspects and multiple factors influencing these aspects***. Specifically, we focus on combining the aspects of perceptual quality and enjoyment and consider a number of different user factors together with system and contextual factors. We pursue this challenge in three majors steps, each contributing to answering a specific (set of) research questions.

The first step stems from the need, outlined in Section 2.2, of furthering our understanding of how different factors, and specifically user factors, influence different aspects of QoE. Here we ask:

1.a. Which user factors influence enjoyment and perceptual quality?

1.b. How do user factors interact with each other and with context and system factors in forming the final QoE impression?

We investigate these questions in **Chapter 2**, which reports the outcomes of a large empirical study investigating the impact of social context and user factors on different aspects of the (social) video viewing experience. The contribution of this study is twofold. First, we show that enjoyment and perceptual quality indeed can be seen as different aspects of QoE. While users were able to clearly distinguish various levels of video quality used in our study, these levels did not necessarily affect the user's enjoyment. Second, user's enjoyment is shown to be significantly influenced by the level of interest of the user in the video, the social context as well as the video content (in this case, the video genre).

With the encouraging results from our empirical study, we move forward to answering the crucial research question of this thesis:

2. Can we design an objective quality model that, by processing user, system, and context information, is able to predict individual QoE?

In **Chapter 3** we propose a new QoE model that takes different QoE aspects and influencing factors into account. Our model takes as input not only the information from the video (i.e., information related to visual impairments and video content), but also the information from the user (i.e., personality, gender, interest and cultural background of the specific user who watched the video). The model predicts not only the perceptual quality of the video, but also the level of enjoyment that a user experiences with the video. Most importantly, our model targets at individual viewing experience, i.e., the experience that a specific user has of a specific video. We validate our model based on the data we collected from our empirical study as well as another public dataset [26]. The results show that combining information from video and user enables better QoE prediction performance as compared to only considering information from video related to visual impairments, both when targeting perceptual quality and enjoyment.

Finally, we tackle the problem of reliable data collection in empirical studies aimed at investigating individual QoE. Information on user factors (e.g. demographics or personality) is usually collected via self-report questionnaires in lab-based experiments. Therefore, the data size typically is not big enough for validating an individual QoE prediction model. Since there are not many public datasets available that provide individual QoE scores (most of them only provide MOS as we explained in section 2.3) and the corresponding user factors, this lack of data introduces a bottleneck for researchers to validate our proposed model. In addition, to achieve an automatic QoE prediction by considering user factors, another big challenge lies in collecting user factors in an unobtrusive fashion. Here, internet-based data collection can provide a suitable alternative. Social media, in particular, can provide a platform where to collect unobtrusively (and of course given the user consent) accurate information on user interest and preferences [49, 53]. This leads us to formulate our last two research questions:

3.a. Can we use social media as a platform to perform online experiments aimed at collecting reliable subjective assessments of QoE?

3.b. Can social media-based experimentation support the reliable and unobtrusive collection of user factor data?

In **Chapter 4**, we report on how we implemented an open-sourced Facebook application, named YouQ, as an experimental platform for studying individual video viewing experience and collecting user information from social media while performing an online QoE subjective experiment. We show that YouQ can produce reliable results as compared to a controlled laboratory experiment, both in terms of QoE and of quantification of user factors and traits. It is important to remark here that YouQ was developed following strong ethical principles and with great care with respect to privacy aspects. Before being made publicly available, YouQ has passed the Facebook review process [1]. All users who access it are obliged to read a short introduction (regarding our experimental purpose and our privacy policy) and to give their consent to let YouQ retrieve their personal information for research purposes only. Users are free to drop out and stop sharing their information with YouQ at any time. We also made sure that YouQ can be used even when someone refused to share his/her personal information. All information that YouQ has

[1] Please find more information at: https://developers.facebook.com/docs/facebook-login/review

collected is anonymized and can in no way be traced back to a specific user.

The thesis concludes with a reflection on its contributions in **Chapter 5** and with suggestions for future research.

1.4. How to read the thesis

The technical part of this thesis, represented by chapters 2, 3 and 4, consists of original publications that have been adopted in their original form. The publication source of each chapter is indicated in a footnote on the first page of that chapter. As a consequence of working with original publications, the terminology and notation may vary across chapters. For the same reason, some sections from different chapters, typically the introduction and related work sections, may be similar in terms of argumentation and the material they cover.

1.5. List of publication related to the thesis

The following papers have been published by the author of the thesis while pursuing a Ph.D. degree in the Multimedia Computing Group at the Delft University of Technology. This publications directly serving as chapters of the thesis, or contributing to thesis chapters, are indicated accordingly in each chapter.

Book Chapters

1. Redi J A, Zhu Y, de Ridder H, et al. How passive image viewers became active multimedia users [M]//Visual Signal Quality Assessment. Springer International Publishing, 2015: 31-72.

Journals

1. Zhu Y, Heynderickx I, Redi J A. Understanding the role of social context and user factors in video Quality of Experience [J]//Computers in Human Behavior, 2015, 49: 412-4.

2. Zhu Y, Guntuku S, Lin W, Ghinea G, Redi J. Measuring Personal QoE using Facebook: A Survey of the State-of-Art and Proposal for Future Direction [J] //The ACM Transactions on Multimedia Computing, Communications, and Applications 2017

Conference papers

1. Zhu Y, Heynderickx I, Redi J A. Alone or together: measuring users' viewing experience in different social contexts [C] //IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2014: 90140W-90140W-11.

2. Zhu Y, Heynderickx I, Hanjalic A, et al. Towards a comprehensive model for predicting the quality of individual visual experience [C] //SPIE/IS&T Electronic Imaging. International Society for Optics and Photonics, 2015: 93940A-93940A-12.

3. Kaptein R, Zhu Y, Koot G, et al. Retrieving Relevant and Interesting Tweets During Live Television Broadcasts [C] //International Conference on Web Engineering. Springer International Publishing, 2015: 175-185.

4. Zhu Y, Hanjalic A, Redi J A. QoE Prediction for Enriched Assessment of Individual Video Viewing Experience [C]//Proceedings of the 2016 ACM on Multimedia Conference. ACM, 2016: 801-810.

1.6. REFERENCE

[1] Cisco Visual Networking Index: Forecast and Methodology, 2016–2021. available at: https://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf .2017.

[2] Pérez, P., et al. Effect of packet loss in video quality of experience. Bell Labs Technical Journal, 16, 1 , 91-104, 2011.

[3] Goldmann, L., et al. Impact of video transcoding artifacts on the subjective quality. IEEE Second International Workshop on Quality of Multimedia Experience (QoMEX), 2010.

[4] Pozueco, Laura, et al. Subjective video quality evaluation of different content types under different impairments. New Review of Hypermedia and Multimedia 23.1: 1-28, 2017.

[5] Staelens, N., et al. Assessing quality of experience of IPTV and video on demand services in real-life environments. IEEE Transactions on broadcasting, 56, 4, 458-466, 2010.

[6] Reichl, Peter, et al. What you pay is what you get?. Quality and User Experience 3.1:5, 2018.

[7] Girod, B. What's wrong with mean-squared error? Digital images and human vision, 207-220, 1993.

[8] Chandler, D. M. Seven challenges in image quality assessment: past, present, and future research. ISRN Signal Processing, 2013.

[9] Keelan, B. Handbook of image quality: characterization and prediction. CRC Press, 2002.

[10] Engeldrum, P. G. Psychometric scaling: a toolkit for imaging systems development. Imcotek, 2000.

[11] ITU-R. BT.500-13, Methodology for the subjective assessment of the quality of television pictures, 2012

[12] ITU-T, P.910. Subjective video quality assessment methods for multimedia applications, 2008.

[13] ITU-T, P.913. Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment, 2016.

[14] ITU-R. BT.2021-1. Subjective methods for the assessment of stereoscopic 3DTV systems, 2015.

[15] Moon, Seong-Eun, and Jong-Seok Lee. Implicit Analysis of Perceptual Multimedia Experience Based on Physiological Response: A Review. IEEE Transactions on Multimedia 19.2: 340-353, 2017.

[16] Lin, W. and Kuo, C.-C. J. Perceptual visual quality metrics: A survey. Journal of Visual Communication and Image Representation, 22, 4, 297-312, 2011.

[17] Gu, Ke, et al. Learning a no-reference quality assessment model of enhanced images

with big data. IEEE transactions on neural networks and learning systems 29.4: 1301-1313, 2018.

[18] Liu, Min, et al. Perceptual reduced-reference visual quality assessment for contrast alteration. IEEE Transactions on Broadcasting 63.1: 71-81, 2017.

[19] Moorthy, A. K. and Bovik, A. C. Visual quality assessment algorithms: what does the future hold? Multimedia Tools and Applications, 51, 2, 675-696, 2011.

[20] Shahid, Muhammad, et al. No-reference image and video quality assessment: a classification and review of recent approaches. EURASIP Journal on Image and Video Processing 2014.1: 40, 2014.

[21] Oviedo, Vanessa, et al. Effects of media multi-tasking with Facebook on the enjoyment and encoding of TV episodes. Computers in Human Behavior 51: 407-417, 2015.

[22] Neate, Timothy, et al. Cross-device media: a review of second screening and multi-device television. Personal and Ubiquitous Computing 21.2: 391-405, 2017.

[23] You, Junyong, et al. Perceptual-based quality assessment for audio–visual services: A survey. Signal Processing: Image Communication 25.7: 482-501, 2010.

[24] Siahaan, Ernestasia, et al. Semantic-aware blind image quality assessment. Signal Processing: Image Communication 60: 237-252, 2018.

[25] Wechsung, I., et al. All users are (not) equal-the influence of user characteristics on perceived quality, modality choice and performance. In Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop (pp. 175-186). Springer New York, 2011.

[26] Scott, Michael James, et al. Do personality and culture influence perceived video quality and enjoyment?. IEEE Transactions on Multimedia 18.9: 1796-1807, 2016.

[27] O'Brien, H. L., and Toms, E. G. What is user engagement? A conceptual framework for defining user engagement with technology. Journal of the American Society for Information Science and Technology, 59(6), 938-955, 2008.

[28]Kortum, P., and Sullivan, M. The effect of content desirability on subjective video quality ratings. Human factors: the journal of the human factors and ergonomics society, 52(1), 105-118, 2010.

[29] Palhais, J., Cruz, R. S., and Nunes, M. S. Quality of Experience Assessment in Internet TV. In Mobile Networks and Management (pp. 261-274). Springer Berlin Heidelberg, 2012

[30] Schmitt, Marwin Robert, et al. Towards individual QoE for multi-party video conferencing. IEEE Transactions on Multimedia, 2017.

[31] Chorianopoulos, K., and Lekakos, G. Introduction to social TV: Enhancing the shared experience with interactive TV. Intl. Journal of Human–Computer Interaction, 24(2), p. 113-120, 2008.

[32] Chen, Yanjiao, et al. From QoS to QoE: A tutorial on video quality assessment. IEEE Communications Surveys & Tutorials 17.2: 1126-1165, 2015.

[33] SG12, I. Definition of quality of experience. TD 109rev2 (PLEN/12), Geneva, Switzerland, 16-25, 2007

[34] Brunnström, K., et al. Qualinet white paper on definitions of quality of experience, 2013.

[35] Gulliver, S. R. and Ghinea, G. Defining user perception of distributed multimedia quality. ACM Transactions on Multimedia Computing, Communications, and Applications, 2, 4, 241-257, 2006.

[36] Tamborini, R., et al. Defining media enjoyment as the satisfaction of intrinsic needs. Journal of communication, 60, 4, 758-777, 2010.

[37] Jumisko-Pyykkö, S. User-centered quality of experience and its evaluation methods for mobile television. Tampere University of Technology, 12, 2012.

[38] See-To, E. W et al. 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, 79(8), 1484-1494, 2012.

[39] Duanmu, Zhengfang, et al. A quality-of-experience index for streaming video. IEEE Journal of Selected Topics in Signal Processing 11.1: 154-166, 2017.

[40] Noor, Tawfeeg S., et al. The impact of gender on the Quality of Experience for video services. Automation and Computing (ICAC), 2016 22nd International Conference on. IEEE, 2016.

[41] Naumann, Anja B., et al. Multimodal interaction: A suitable strategy for including older users? Interacting with Computers 22.6: 465-474, 2010.

[42] Engelke, Ulrich, et al. Psychophysiology-based QoE assessment: a survey. IEEE Journal of Selected Topics in Signal Processing 11.1: 6-21, 2017.

[43] Dobrian, F., et al. Understanding the impact of video quality on user engagement. ACM SIGCOMM Computer Communication Review. Vol. 41. No. 4, 2011.

[44] Ickin, S., et al. Factors influencing quality of experience of commonly used mobile applications. IEEE Communications Magazine, 50, 4, 2012.

[45] Streijl, Robert C., et al. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. Multimedia Systems 22.2: 213-227, 2016.

[46] Hoβfeld, Tobias, et al. SOS: The MOS is not enough!. Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on. IEEE, 2011.

[47] Schuurman, D., et al. T. A Living Lab research approach for mobile TV. Telematics and Informatics, 28, 4, 271-282, 2011.

[48] Kosinski, M., et al. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. American Psychologist, 70, 6, 543, 2015.

[49] Gardlo, Bruno, et al. Microworkers vs. facebook: The impact of crowdsourcing platform choice on experimental results. Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on. IEEE, 2012.

[50] Egger-Lampl, Sebastian, et al. Crowdsourcing Quality of Experience Experiments. Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments. Springer, Cham,154-190, 2017.

[51] Varela, M., et al. Towards an understanding of visual appeal in website design. IEEE Fifth International Workshop on Quality of Multimedia Experience (QoMEX), 2013.

[52] Hossfeld, T., et al. Best practices for QoE crowdtesting: QoE assessment with crowdsourcing. IEEE Transactions on Multimedia, 16, 2, 541-558, 2014.

[53] Fu, Pei-Wen, et al. What makes users share content on facebook? Compatibility among psychological incentive, social capital focus, and content type. Computers in Human Behavior 67: 23-32, 2017.

2

Understanding the Role of Social Context and User Factors in Video Quality of Experience

This chapter is based on Y. Zhu, I. Heynderickx, J. A. Redi, *Understanding the role of social context and user factors in video Quality of Experience*. Computers in Human Behavior, 2015, 49: 412-4.

Abstract

Quality of Experience is a concept to reflect the level of satisfaction of a user with a multimedia content, service or system. So far, the objective (i.e., computational) approaches to measure QoE have been mostly based on the analysis of the media technical properties. However, recent studies have shown that this approach cannot sufficiently estimate user satisfaction, and that QoE depends on multiple factors, besides the media technical properties. This chapter aims to identify the role of social contextual and user factors (such as interest and demographics) in determining quality of viewing experience. We also investigate the relationships between social context, user factors and some media technical properties, the effect of which on image quality is already known (i.e., bitrate level and video genre). Our results show that the presence of co-viewers increases the user's level of enjoyment and enhances the endurability of the experience, and so does interest in the video content. Furthermore, although participants can clearly distinguish the various levels of video quality used in our study, these do not affect any of the other aspects of QoE. Finally, we report an impact of both gender and cultural background on QoE. Our results provide a first step towards building an accurate model of user QoE appreciation, to be deployed in future multimedia systems to optimize the user experience

2.1. Introduction

Online video services show a continuous growth. By 2010, over 71% of internet users had watched videos online, and this number grew from 33% in 2006 [40]. These figures are forecasted to further grow in the coming years [9, 40]. With a constantly increasing volume of streamed video data, maintaining a satisfactory video service to users at all times is challenging for internet and multimedia providers. Due to different technological limitations (e.g., bandwidth and storage constraints, network malfunctioning), visible artifacts (e.g., blockiness or blur due to compression, freezes or jerkiness due to transmission errors) can be introduced to any stage of the video delivery cycle [46, 59]. This, in turn, can severely degrade the user's satisfaction, and evidence shows that users intend to pay less if a service cannot meet their expectations [43, 65]. As a consequence, online video providers are eager to find ways to measure and predict user's satisfaction with videos in order to optimize their video delivery chains.

Quality of Experience (QoE) is a concept commonly used to describe user's overall satisfaction [34], reflecting the degree of delight or annoyance of a user with a (multimedia) system, service or application. In the past decades, user's satisfaction with videos has been estimated mainly from a technical perspective, i.e., based on either the information gathered from the network and service conditions or from image and video analysis [55]. From a network management perspective, the concept Quality of Service (QoS) has often been equated to QoE. Here, network parameters, such as packet loss or delay [2], as well as video QoS parameters, e.g., the so-called join time at the start of playing the video or the buffering time during the video [15], were monitored; their compliance to given standards was considered enough to guarantee sufficiently high QoE. The signal processing community has instead relied more often on the analysis of information extracted from the decoded image/video signal to estimate the visibility of artifacts in it [24, 36]. Artifact visibility was considered to be inversely related to perceptual quality,

and therefore to user satisfaction [7]. In both cases, user satisfaction was mainly associated to technical properties of the multimedia signal, service or system.

Lately, research has shown that this approach has limitations, and that other elements concur to guarantee user satisfaction when watching video [34, 66]. For example, recent studies claimed that QoE should also be considered from a user perspective [13]: evidence has been provided that user's interest [32] and personality [60] influence QoE too. Such findings reveal the complexity of QoE: it is a combination of many influencing factors, not limited to QoS parameters nor artifact visibility.

Influencing factors on QoE are often grouped into three categories, i.e., system, user and contextual factors [34]. System factors concern the technical aspects of a multimedia system (e.g., network parameters, media genre, media configuration). User factors refer to individual characteristics of the user who is experiencing the video (e.g., demographics, personal interest or personality). Contextual factors refer to the characteristics of the environment within which the video experience is consumed (e.g., physical features of the environment, economical factors related to the video fruition, presence or absence of co-viewers). As mentioned earlier, most research in the field has focused on system factors, leaving the contribution of user and contextual factors largely unexplored. However, the rise of online video fruition has created a shift from a passive viewing experience to a more active, personalized and shared experience, changing the traditional television market considerably [58]. Compared to traditional TV users who just watch scheduled programs, internet users are free to choose the content they want, at any point in time and space they want, through a variety of devices (e.g., tablets, smartphone or computers). Thus, it is expected that personal characteristics as well as context of fruition will play an important role in such viewing experiences. Moreover, the rise of social media has led to a new type of social viewing experience, where preferences for video content are clearly reported on social media platforms (through comments and ratings), and are visible to the rest of the (vast) online community. The social context in which the video is experienced is therefore expected to play a key role in the eventual user satisfaction.

As the optimization of online video watching requires a more in-depth understanding of the impact of user and contextual factors on QoE, we here want to contribute to the generation of this knowledge by considering the impact of social context in particular. Interestingly, very little is known about how social context (1) relates to QoE and (2) combines with system and user factors to determine the final user satisfaction with the viewing experience. We specifically focus on what we define as "direct" social context, that is, the presence or absence of co-viewers in the physical proximity of the user. We report the outcomes of an empirical study looking into the role played by direct social context in determining QoE when given system factors (i.e., video genre and bitrate) are in place. Furthermore, we analyze the interactions of direct social context with user influencing factors such as demographics, interest in the video genre and immersive tendency. We measure six different aspects of the viewing experience, namely perceptual quality, enjoyment, endurability, satisfaction, involvement and information assimilation. The outcomes should support building an accurate objective model for QoE on the longer term.

This chapter continues by presenting the related work in section 2.2, which we re-

viewed to define the hypotheses for the empirical study as described in section 2.3. We then outline our experimental methodology in section 2.4, followed by the analysis of the results in section 2.5. We discuss our findings in section 2.6, leading to the most important conclusions in section 2.7.

2.2. Related work

In the past decades, the effectiveness of multimedia services has been linked to the notion of Quality of Service (QoS), defined as the "totality of characteristics of a telecommunication service that bears on its ability to satisfy stated and implied needs of the user of the service" [50]. QoS is mainly operationalized in terms of system and network performance-related measures (e.g., packet loss ratio, jitter or delay). This approach has started showing its limitations, and was found to be poorly correlated to user satisfaction [6]. As a result, the Quality of Experience concept has emerged, being defined as "the overall acceptability of an application or service, as perceived subjectively by the end-user" by ITU-T [56]. Compared to QoS, the notion of QoE has taken a user-centric perspective, now keeping user perception into consideration. Remarkable work has been done in estimating QoE from a perceptual point of view [24, 36].

Recently, the Qualinet White Paper [34] has proposed an even more compelling definition of Quality of Experience:

"Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state".

Although both the ITU-T and the [34] definitions describe a similar concept, the latter seems more complete than the one of ITU-T, as it emphasizes how user-related factors, e.g., personality and current state, may have an impact on QoE. Given the evidence of the importance of such factors in properly estimating QoE (which will be explained in detail in section 2.2.1), we use the Qualinet definition as operational definition of QoE throughout this chapter.

2.2.1. Factors influencing QoE

Quality of Experience is a multifaceted quality, resulting from the interaction of multiple influencing factors, which are reviewed here in more detail, although not in an exhaustive way. As shown in Table 2.1, these factors can be arranged into three categories, namely system factors, user factors and contextual factors [34].

System factors refer to the system, application and media "properties and characteristics that determine the technically produced quality of an application or service" [28]. Within video delivery services, system factors can influence QoE by altering the perceptual quality of the video [55]. For example, a given type of compression (e.g. H.264/AVC), aiming at obtaining a given bitrate for the video, possibly generates compression artifacts (e.g. blockiness, blur and ringing), which, if visible, result in annoyance for the user, lowering his/her satisfaction. Similarly, network QoS parameters [15], and the media configuration [22] are known to have an impact on QoE. For example, it has been shown that the buffer ratio (i.e., the fraction of time spent in buffering over the total ses-

sion time, including playing plus buffering) is inversely related to QoE [15], and similar conclusions were reached for other QoS parameters, such as the join time in multicast video delivery, the buffering duration, the rate of buffering events, the average bitrate and the packet loss rate [27, 39]. Besides the signal/network factors, user's QoE with video also may be influenced by the nature of the video content itself [3]. Different genres (e.g., sports, comedy, etc.) show very different viewing patterns which may result in different perceptual quality. Given a certain bitrate, for example, genres characterized by high pace movement (e.g., sports or action film) usually have lower perceptual quality than genres which contain little movement [25]. Moreover, it has been shown that user's active emotions (e.g., worry, fun) were significantly higher when watching action videos compared to other genres, e.g., documentary, sports [48, 54].

Table 2.1: Factors influencing QoE discussed in this section

System factors	User factors	Contextual factors
Signal and networkparameters [1-3]	Interest [8, 9]	Physical environment [19]
Genre [5]	Personality [10]	Economic aspects [20]
	Demographics [14, 15]	Social situation [21-23]

User factors refer to individual characteristics of the user that may influence the viewing experience. Some studies indicate that QoE is triggered when something resonates with a user's interest [44] and that personal interest in video content significantly influences user's QoE judgment [32]. Moreover, it is shown that users tend to value a video with the same bitrate as higher in QoE when they are more interested in the content of the video [47]. Personality is shown to influence at least the user performance part of QoE. Neurotic people are less able to switch the TV channel or change the volume of the TV on their first attempt compared to agreeable people and/or people with technical competence or enthusiasm [60]. Demographic characteristics of the user (e.g., age, gender and cultural background) are also expected to influence QoE. At least for age, there is evidence in literature: older adults are found to be more critical than younger users, which suggests that elderly people may have higher requirements for QoE [64]. However, another study demonstrates the opposite trend: older users tend to rate video quality more positively than younger users, although the performance is worse [43]. Similar scattered results exist for gender [4, 26, 42], and no systematic investigation has been carried out, to the best of our knowledge, to clarify the role of demographic characteristics in QoE appreciation.

Contextual factors are related to the environment in which the user consumes the media. The physical environment certainly influences QoE through a number of elements, and should be characterized accordingly. The seating position (e.g., viewing distance and viewing height), lighting conditions as well as disturbances that occur in the environment a viewer is in (e.g., incoming phone calls or SMS message alerts) may influence user experience. Viewing distance for example is known to affect the overall perceptual quality: a shorter viewing distance increases the field of view, and makes the viewer more involved with the content, but may make artifacts better visible as well [61]. Economic aspects related to the experience fruition also contribute to generate expectations with respect to its quality; when unfulfilled, they may have critical consequences

on the willingness of the user to repeat the experience. [29] showed that when users feel they are overpaying for the service in regard to the quality of the experience, they react in different ways, which all eventually lead to a decrease in revenues for the operator of those customers.

In this chapter we mainly focus on one particular contextual factor, namely the social context within which a video is seen. It is well known that a user is affected by the interaction with a group of other people [52], being them family, friends or even strangers. In fact, it has been shown that co-located co-viewing (which is a rather common way for consuming TV programs [41]) may increase user's overall satisfaction with the program [45]. The social element of the viewing experience stretches even further, with users recording their favorite programs and sharing them with families and friends [33], or with people using their viewing experience as a conversation topic [35]. Recently, a concept of "social TV" has emerged: it provides multiple viewers with a joint TV-watching experience by adding communication features [8]. User studies of social TV have confirmed the high acceptance of such technology, because it allows users to communicate with friends even when they are not physically co-located [18]. All these results point towards a growing importance of quantifying the relevance of the social context in QoE. Yet, limited research reports this relevance, and does not discuss its relationship with the other user and system factors listed above.

2.2.2. Existing approaches to measure QoE

QoE has been historically measured in two ways: objectively and subjectively. Objective QoE assessment entails the estimation of QoE from the analysis of a set of system/signal parameters, in a way that is completely automated and does not involve human judgment directly, e.g., [17, 24, 30, 36]. These measurement techniques, also often referred to as quality metrics, are certainly preferred for online multimedia delivery optimization, and have proven to be effective at relating QoE to system factors. On the other hand, they have shown limitations in taking into account user and contextual factors to QoE [51, 57]. To design objective metrics that can properly reflect the influence of contextual and user factors on QoE judgments, these relationships first have to be characterized from an empirical point of view. For this type of investigation, subjective measurements are more appropriate.

Subjective QoE assessment is based on asking users to self-report their (perceptual) satisfaction with respect to a set of multimedia contents. To date, this approach is still considered to be the most reliable way to quantify QoE. Subjective ratings are often collected via psychometric experiments [16] and aim at measuring the satisfaction of an average user with respect to a given stimulus (e.g. video). As a result, subjective QoE is often expressed in terms of Mean Opinion Scores (MOS), quantifying the average rating according to a specific aspect of QoE. The image and video quality community, for example, has made use of standardized methodologies and experimental settings to quantify the annoyance of visible artifacts and/or the perceptual quality of a video [49].

In fact, many studies are based on the analysis of MOS of perceptual quality (PQ) to understand the relationship between QoE and influencing factors. Some studies suggest that the measurement of QoE should be complemented by a measurement of the level of enjoyment of the experience, which reflects how much happiness or fun a user gets

from the videos [21, 23]. Gulliver and Ghinea [23] also proposed to take into consideration user's level of Information Assimilation (IA). IA, reflecting the level of comprehension of video content, is used to measure a user's ability to understand and assimilate information from videos; as the authors argue that media are consumed for infotainment purposes, both the entertainment and the informative capabilities of the viewing experience should be evaluated in QoE assessments. More aspects that may characterize QoE have been suggested in literature, but have not been investigated directly in relation to multimedia optimization. Recently, the term "endurability" has been used to describe the consequence of satisfaction; it refers to the extent to which users remembered the experience as enjoyable, were willing to repeat it and/or advise it to others [44]. We may say that endurability measures the outcome of high QoE: favorably impressed users will be more willing to repeat and share the experience, as well as use the multimedia system/service. As such, we consider endurability as a valuable business-oriented measurement of QoE. Another concept related to QoE is involvement, which occurs when a user is psychologically immersed in a video. Involvement has been shown to be positively correlated to the experience likeability [44]. Thus, involvement can be considered as a supplementary measure of QoE. Recently, also affect-related measures (e.g., emotion or mood) have received increasing attention in QoE research [1, 11]. It has been shown that a positive mood (e.g., joy) relates to the experience of enjoyment and satisfaction, whereas a negative mood (e.g., frustration) relates to poor experiences and eventually leads to disengagement with the service [44].

The measures of QoE , as discussed above, have been administered in different forms, i.e., by means of self-report [12, 54], via interviews [14] or with physiological measurements, such as facial expression, galvanic skin response and EEG [1, 11]. It should be pointed out here, that besides for PQ, no standardized methods exist for capturing all the listed aspects of QoE. Hence, to broaden the concept of QoE from a measure of perceptual satisfaction to that represented in the Qualinet QoE definition we adopted, complementary aspects of QoE should be measured, which we propose to do by including attributes as perceptual quality, enjoyment, satisfaction, endurability, involvement and information assimilation.

2.3. RESEARCH QUESTIONS AND HYPOTHESES

BASED on the literature overview given in section 2.2, we formulate three research questions:

1. What is the effect of direct social context on QoE?
2. How is the impact of system factors on QoE affected by the direct social context?
3. How is the impact of user factors on QoE affected by the direct social context?

To answer these research questions, QoE is measured along the six attributes, mentioned above: perceptual quality, enjoyment, satisfaction, endurability, involvement and information assimilation. The system factors considered are video genre and compression bitrate, and the user factors studied are immersive tendency, user interest and demographics. The direct social context is defined here as the presence/absence of physically co-located co-viewers.

Since it has been shown that users enjoy each other's company and that co-viewing can increase their level of enjoyment while watching TV [45], we formulate our first hy-

pothesis as:

H1. The presence of co-viewers increases the user's QoE.

It is generally known that video bitrate affects perceptual quality [27, 39], and so, also QoE; the lower the bitrate level, the lower PQ, and thus QoE. It is, however, not known to what extent the lower QoE may be balanced out by the presence of co-viewers. But, since we hypothesize that co-viewing increases QoE, we also hypothesize that:

H2a. The presence of co-viewers increases the user's tolerance to artifacts present in low bitrate videos.

In addition, we may expect that the effect that co-viewing has on QoE depends on the preference of users to watch a particular video genre alone or in company. Hence, we hypothesize that:

H2b. The increase in QoE by co-viewers is bigger for video genres that are preferred to be watched in group than for video genres that are preferred to be watched alone.

Related to the third research question, literature suggests a direct impact of user factors, such as user interest, immersive tendency and demographics, on QoE. For example, previous studies indicated that the higher level of interest of a user with a video, the higher he/she rates experience satisfaction [47], which is expected to be part of QoE. Hence, we hope to confirm the hypothesis:

H3a. User interest positively correlates with user's QoE.

The immersive tendency of a user quantifies how easily he/she gets involved in common activities, and so was often used to measure involvement in virtual reality studies [62]. Similarly, it is expected that a user who has high immersive tendency becomes more involved when watching videos. In addition, evidence shows that a high level of involvement leads to high satisfaction [44], and so, high QoE. Hence, we hypothesize:

H3b. The higher the immersive tendency of a user, the higher the involvement with the video, and thus the higher the QoE.

Related to demographic factors earlier studies showed that males and females react differently to emotional pictures [5] and have different perception of olfactory and visual media synchronization [42]. Some impact of age on QoE has been demonstrated, though not all reported results in literature were consistent [43, 64]. Finally, users with a different cultural background usually have a different understanding of experience, and thus may perform differently toward a same task [38]. As a consequence, it is reasonable to expect that optimal QoE settings may depend on these demographic factors. Thus, we hypothesize:

H3c. Gender, age and cultural background have an impact on QoE.

Also in relation to the user factors under consideration in our study, it is not known to what extent their impact on QoE is affected by the direct social context of watching the video alone or in group. We may though expect that group processes are more important than personal interest or immersive tendency when judging QoE of watching TV with others. Consequently, we expect QoE to be more affected by the user factors under evaluation when watching the videos alone than in group, leading to the hypothesis:

H3d. The positive impact of personal interest and immersive tendency on QoE is more pronounced when watching the videos alone than in group.

To evaluate the above mentioned hypotheses, we designed an empirical study, controlled for the social context and for the system factors video genre and bit rate. We

Table 2.2: Overview of the experimental setup. V indicates the Video Clip tested; S indicates a group of participants that watched the video in the single viewer's condition; G indicates a group of participants that watched the video with groups of 3 people. The effect of social context and video bitrate level are investigated between subjects, whereas the effect of genre is investigated within subjects.

	Genre1		Genre2		Genre3	
	V1	V2	V3	V4	V5	V6
High Bitrate	S1	S2	S1	S2	S1	S2
	G1	G2	G1	G2	G1	G2
Low Bitrate	S2	S1	S2	S1	S2	S1
	G2	G1	G2	G1	G2	G1

measured QoE along six dimensions, and characterized the participants along the user factors interest, immersive tendency and demographics.

2.4. EXPERIMENTAL SET-UP

To test our hypotheses, we created two real-life viewing situations with varying direct social context. In the first situation, single users (hereafter indicated with S, shown in Figure 2.1.a) watched the videos alone (i.e., absence of direct social context). In the second one, a group of three friends (hereafter indicated with G, shown in Figure 2.1.b) watched the videos together. Participants who were involved in one social situation (e.g., single) were not presented with the other situation (e.g., group). As a result, we investigated social context as a between-subjects variable.

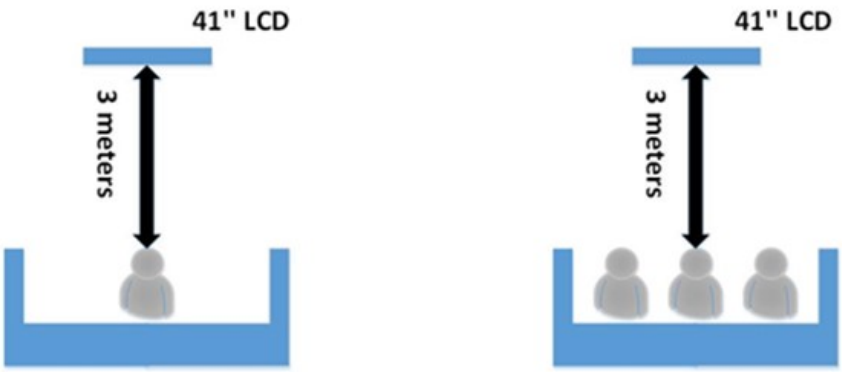

Figure 2.1: The two different social contexts investigated in the experiment. In viewing situation (a), a single participant watched videos on a 41" screen from a couch 3 meters away. In viewing situation (b), three participants, friends, watched the same video together in the same environmental conditions

Six videos distributed over three genres (i.e., comedy, sports and education) were used in our study. All videos were encoded at two quality levels (i.e., high and low). Participants within each social context were further divided into two sub-groups (S1 and S2, or G1 and G2, as shown in Table 2.2). Within each sub-group, participants watched the video content only once, at a quality level that was either high or low. As a result, the

effect of bitrate level was investigated as a between-subjects variable, while video genre was investigated as a within-subjects variable.

2.4.1. PARTICIPANTS

Sixty participants (i.e., 27 females and 33 males) from the Delft University of Technology (TUD) were recruited for this experiment. The participants' age ranged between 18 and 41 years (mean age = 26.5). Over half of the participants (60%) were of Asian origin, whereas the rest was from Western countries (i.e., European or American). It should be noted that only 9 participants (i.e., 15%) were English native speaker, but since an English proficiency certification is needed to be a student at TUD, we were confident that all participants had a sufficient English level to understand the video content as well as the questions posed for the measurements. Fifty-two participants (i.e., 88%) were frequent online video users (i.e., watching online video at least several times a week). YouTube and social websites (e.g., Facebook, twitter etc.) were the platforms most commonly used for consuming online video.

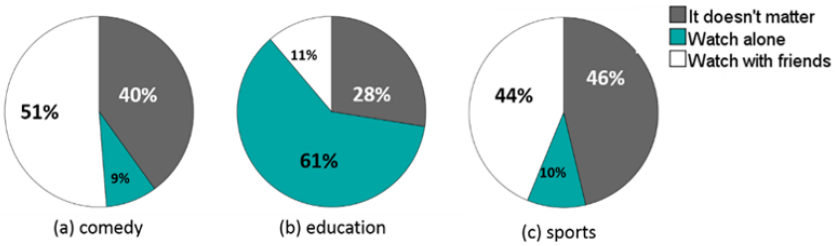

Figure 2.2: Results of the online survey on preferred social context for watching (a) comedy, (b) education and (c) sports videos. The white area in the pie plots indicates preference for watching alone, the greenish area indicates preference for watching with friends, and the grey area indicates that it doesn't matter.

2.4.2. STIMULI

Three different genres of video were used in this study. We first conducted a pilot survey to select these genres. We listed fifteen of the most common genres for online video (as indicated e.g. in YouTube), and for each genre, we asked participants to choose whether they preferred watching it alone or with friends; if they were not sure, they could also choose "it doesn't matter". We received 80 responses from PhD students of TUD. A clear preference (as in gathering over 50% of the choices) was found for two genres, as shown in Figure 2.2: 51% of our participants indicated a preference for watching comedy videos with friends, and 61% of the participants indicated to prefer watching education videos alone. The 'sport' genre, characterized by not obtaining a clear consensus for the preferred viewing situation (see Figure 2.2.c) was also used in our study.

We selected 2 different clips from each of above three genres (screenshots are given in Figure 2.3): the Jimmy Kimmel Show (JKS) and Saturday Night Live (SNL) for comedy, The Birds of Paradise (BoP) and a TED talk (TED) for education, and Soccer and

Basketball for sports[1].

All videos lasted at least 5 minutes and originated from YouTube. They were encoded with H.264/AVC, which is the most commonly used codec for online videos (Schwarz 2007). All the original videos had a temporal resolution of 30 frames per second (fps) and a spatial resolution of 1280*720 pixels. Videos were further encoded in H.264/AVC at two different bitrates: high (2000 kbps) and low (600 kbps). The reason to choose only two bitrate levels is that the relationship between video bitrate and QoE (or rather perceptual quality) has been largely investigated in the past, and it was not our interest to retrieve it or further characterize it; rather, we wanted to focus on the changes in QoE due to user and contextual factors, given a certain bitrate level. It should be noted that the original bitrate of the BoP and TED videos was less than 2000 kbps. So for these two videos, we used the original video bitrate as the high quality value. The audio of all clips was encoded in the AAC format (ISO/IEC, 2005) with a bit rate of 112 kbps to avoid any effect of the sound on QoE.

Finally, three 10s-long video samples were used for training the participants. The media configuration of these samples was the same as for the test videos (30 fps, 1280*720, H.264/AVC). The samples were also encoded at two bitrate levels (i.e., high and low). These samples were used to let participants get acquainted with the range of video quality used in the experiment.

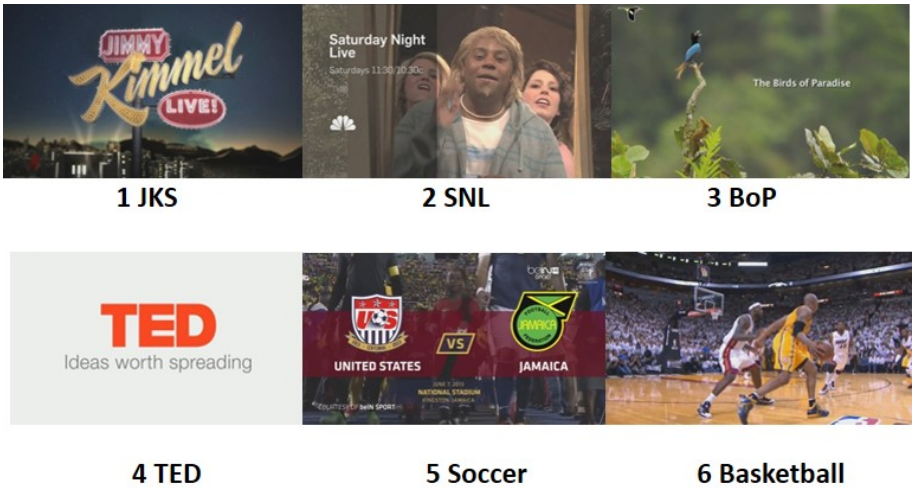

Figure 2.3: Screenshots of the six video clips tested: The Jimmy Kimmel Show (JKS) and Saturday Night Live (SNL) for comedy, The Birds of Paradise (BoP) and a TED talk (TED) for education, and Soccer and Basketball for sports.

[1] JKS, available at: http://www.youtube.com/watch?v=qc9fh-GcjMY&hd=1
SNL, available at: http://www.youtube.com/watch?v=eweXwtMIj5I&hd=1
BoP, available at: http://www.youtube.com/watch?v=YTR21os8gTA&hd=1
TED, available at: http://www.youtube.com/watch?v=H14bBuluwB8&hd=1
Soccer, available at: http://www.youtube.com/watch?v=xFVtb4G_pic&hd=1
Basketball, available at: http://www.youtube.com/watch?v=5OOqQ8YwLk4&hd=1

2.4.3. APPARATUS

All videos were presented on a 41" LCD display (LG, model No. 42LM3400). Participants were seated on a couch in front of the display. The viewing distance was 6 times the height of the screen (i.e., approximately 3 meters) in order to satisfy the preferred viewing distance [49]. The rest of the environmental settings followed the ITU-R BT.500 Recommendations and was kept the same for all participants. All of the display parameters, such as brightness and contrast, were set to their default value for the experiment. The audio volume was kept constant for all participants. Participants used a tablet (Samsung Galaxy Tab 2) to fill in all the questionnaires.

Table 2.3: An overview of the three questionnaires used in our study

Questionnaire	Aspect	Number of Questions	Measure scale	D* /I
UF questionnaire	Demographics	9	–	I
	Interest on genre	3	7-point Likert	I
	Immersive Tendency	18 [4]	7-point Likert	I
QoE questionnaire	Enjoyment	4 [6]	7-point Likert	D
	Endurablility	4 [8]	7-point Likert	D
	Satisfaction	4 [6]	7-point Likert	D
	Involvement	4 [12]	7-point Likert	D
	Perceptual Quality	2 [13]	7-point Likert	D
IA Questionnaire	Information Assimilation	24 [18]	True-False	D
Questionnaire	Assimilation			

*D stands for Dependent variable, I stands for Independent variable.

2.4.4. MEASUREMENTS

Although physiological measurements are gaining interest in QoE research, we nonetheless decided to fully rely on questionnaires, since we wanted the QoE measurements to be as unobtrusive and ineffective in changing the user experience as possible. As shown in Table 2.3, three questionnaires were used. The User Factor (UF) questionnaire included nine questions about demographics (i.e., name, age, gender, cultural background, primary language, educational background, frequency of use of online video, favorite platform and genre); three questions about personal interest on the genres we used in our study, and eighteen questions about immersive tendency, mainly adapted from the Immersive Tendency Questionnaire (ITQ), [62]. Both user interest and immersive tendency were measured on a 7-point scale. Eventually, the UF questionnaire consisted of 30 questions.

The QoE questionnaire measured user's viewing experience in terms of satisfaction, involvement, enjoyment, endurability, and perceptual quality (PQ). Information assimilation was measured with a separate questionnaire. Satisfaction and enjoyment were measured through subsets of the questionnaire proposed by See-To [54]. Involvement was measured through an adapted version of the Igroup Presence Questionnaire [53]. Items from O'Brien's questionnaire [44] were used to measure endurability. Each of these

four aspects was quantified through 4 items to be rated on a 7-point Likert scale. Perceptual quality was instead measured through 2 questions (i.e., one for the annoyance of artifacts, and another one for the overall video quality) to be rated on a 5-point scale, according to the ITU-R BT.500 [49] specification. Eventually, the questionnaire consisted of 18 items.

The information assimilation (IA) questionnaire was made of 24 yes-no questions (i.e., 4 per video), and was used to evaluate the participant's level of information assimilation with each video. This questionnaire was generated based on a pilot experiment. We made 10 content questions for each video used in our study and asked two participants to answer these questions after watching all videos. In principle, all questions could be answered by carefully watching the video, but in practice they weren't. We included in the final IA questionnaire only the questions that two participants answered differently or both answered wrong to ensure some discriminative power on these questions across the different viewing conditions.

2.4.5. PROCEDURE

Firstly, participants were welcomed and asked to sign an informed consent form. They were then seated on a couch in front of the LCD display and asked to fill out the UF questionnaire by using a tablet. Before the start of the actual experiment, an introduction was given to all the participants. In this introduction, six 10-second video samples (spanning a broad range of visual quality) were shown and a questionnaire sample was provided to let participants get acquainted with the range of artifact visibility in the videos and with the scoring scales of the questionnaires. After that, participants watched the six videos in a random order. Following the presentation of each video clip, participants were asked to fill out the QoE questionnaire. The next video was not played until all participants in a session completed the questionnaire. The items in this questionnaire were randomized for each participant. After the complete viewing session, participants were asked to fill out the IA questionnaire. We deliberately chose this timing, since we tried to avoid that participants would pay an unnatural amount of attention to the content of the video, once they discovered that we would ask detailed questions about the content. This, indeed, would have influenced their experience with following videos. It's important to note that participants were asked to fill out all the questionnaires by themselves through a tablet and, those in the group viewing situation, were not allowed to interact with their friends during the phase of answering questions. Interaction was instead welcomed during the phase of watching the videos.

2.5. RESULTS

2.5.1. DATA PREPARATION

Before discussing our results in more detail, we performed a number of bias checks on the distribution of our participants over the two social contexts, i.e., participation in the single vs. group viewing situation. Note that for some variables such as interest, immersive tendency and some demographic data, values of one participant contributing to the group viewing situation were missing. Thus, where applicable, the results of only 59 instead of 60 participants are reported. In addition, since we weren't interested in possible

differences in QoE between the two videos of a particular genre, all analyses described below concern the averaged results over the two videos per genre.

UF QUESTIONNAIRE

The UF questionnaire outputted (a.o.) immersive tendency scores for all participants. For each of them, the immersive tendency value was obtained by summing up the scores of the 18 items in the questionnaire [62]. All participants scored within the normal range, with a mean immersive tendency score of 81.5, in line with the value of 78.7 found in [37]. We inspected the presence of a bias between the two groups of participants (i.e., single vs. group social context) through an independent-samples Mann-Whitney U-test. No significant difference was found between the two social situations (U=13770, p=0.49), suggesting that the distribution of immersive tendency at the start of the experiment was similar in the two groups of participants.

From the UF questionnaire, we also obtained user interest scores on 3 genres (i.e., comedy, education and sports). The mean interest of comedy (mean = 5.2) and education (mean = 4.8) was higher than that of sports (mean = 3.2). We ran an independent-samples Mann-Whitney U-test on the interest scores between the two viewing situations (i.e., single vs. group). No significant difference was found (U=15153, p=0.593), suggesting that the distribution of user interest was similar for the two groups of participants.

QOE QUESTIONNAIRE

The QoE questionnaire consisted of five sub-questionnaires (see Table 2.3), each addressing a different aspect of QoE: satisfaction, involvement, enjoyment, endurability and perceptual video quality. We first tested the internal consistency between items in each sub-questionnaire, using Cronbach's alpha (α) [10]. Usually values of α above 0.8 represent high reliability, i.e., the different items in each aspect measure the same underlying psychological construct. As shown in Table 2.4, for four of the five sub-questionnaires, α was higher than 0.82, indicating high internal consistency. The α value of perceptual quality (PQ), on the other hand, was lower than 0.8, indicating that the two items included in the sub-questionnaire investigated a slightly different concept.

Table 2.4: Internal consistency measured by Cronbach's alpha among items in each aspect of the questionnaire

Aspects	Enjoyment	Involvement	Satisfaction	Endurability	PQ
Chronbach's α	0.917	0.82	0.961	0.939	0.773

As a result:

• For satisfaction, involvement, enjoyment and endurability, we summed the scores given by a participant to the different items in each sub-questionnaire to generate Aspect Scores (AS, one per aspect, video and participant). AS ranged between 4 and 28.

• For PQ, we instead decided to analyze the Opinion Scores (OS) separately for the two aspects. As a results, PQ was characterized by two OS per video and participant.

To investigate mutual relationships between the six different aspects of QoE (i.e., 4xAS + 2xOS), we calculated Spearman rank order correlations between them. As shown

Table 2.5: The Spearman rank order correlations between all aspects of the QoE questionnaire. Here OS1 represents the opinion score of artefact annoyance in the video and OS2 represents the opinion score of overall perceptual quality.

QoE measures	OS1	OS2	Enjoyment	Satisfaction	Endurability
OS2	0.617**				
Enjoyment	0.222**	0.280**			
Satisfaction	0.305**	0.381**	0.889**		
Endurability	0.253**	0.344**	0.868**	0.873**	
Involvement	0.179**	0.240**	0.615**	0.643**	0.582**

in Table 2.5, all aspects were significantly correlated with each other ($p<0.001$). Enjoyment, satisfaction and endurability showed a strong correlation ($r>0.85$), and were all only moderately correlated with involvement ($0.58<r<0.64$). The two PQ items (OS1 and OS2 in Table 2.5) were also moderately correlated with each other ($r = 0.62$), but poorly correlated with the rest of the QoE aspects. In other words, enjoyment, satisfaction and endurability seem to measure a very similar aspect of QoE. They are related to involvement, although not to the full extent. Picture video quality has some influence on the other QoE measures, albeit to a limited extent. Despite these results, we still analyzed all aspects of QoE separately in more detail in all following analyses.

INFORMATION ASSIMILATION QUESTIONNAIRE

The IA questionnaire outputted information assimilation scores (IAS). This IAS was calculated as the percentage of correct answers a participant gave for a video. We first did a k-independent samples Kruskal-Wallis test among all IAS scores, setting genre as independent variable. As shown in Figure 2.4, genre had a significant impact on IAS (χ^2=22.093, $p<0.001$) suggesting that participants' ability of getting right information varied among the different genres. In general, IAS of comedy videos was significantly higher than that of education (U=5153.5, $p<0.001$) and sports (U=5093, $p<0.001$) videos.

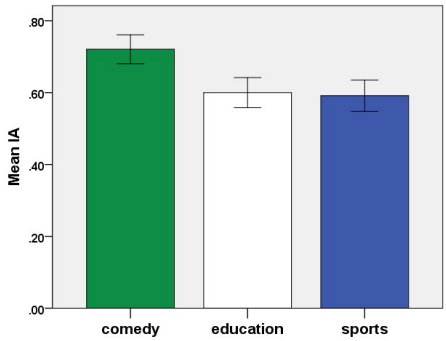

Figure 2.4: The information assimilation for the three different genres. Here the Y-axis represents the mean percentage of correct answers for all participants. Comedy videos were on average better understood than education and sports videos.

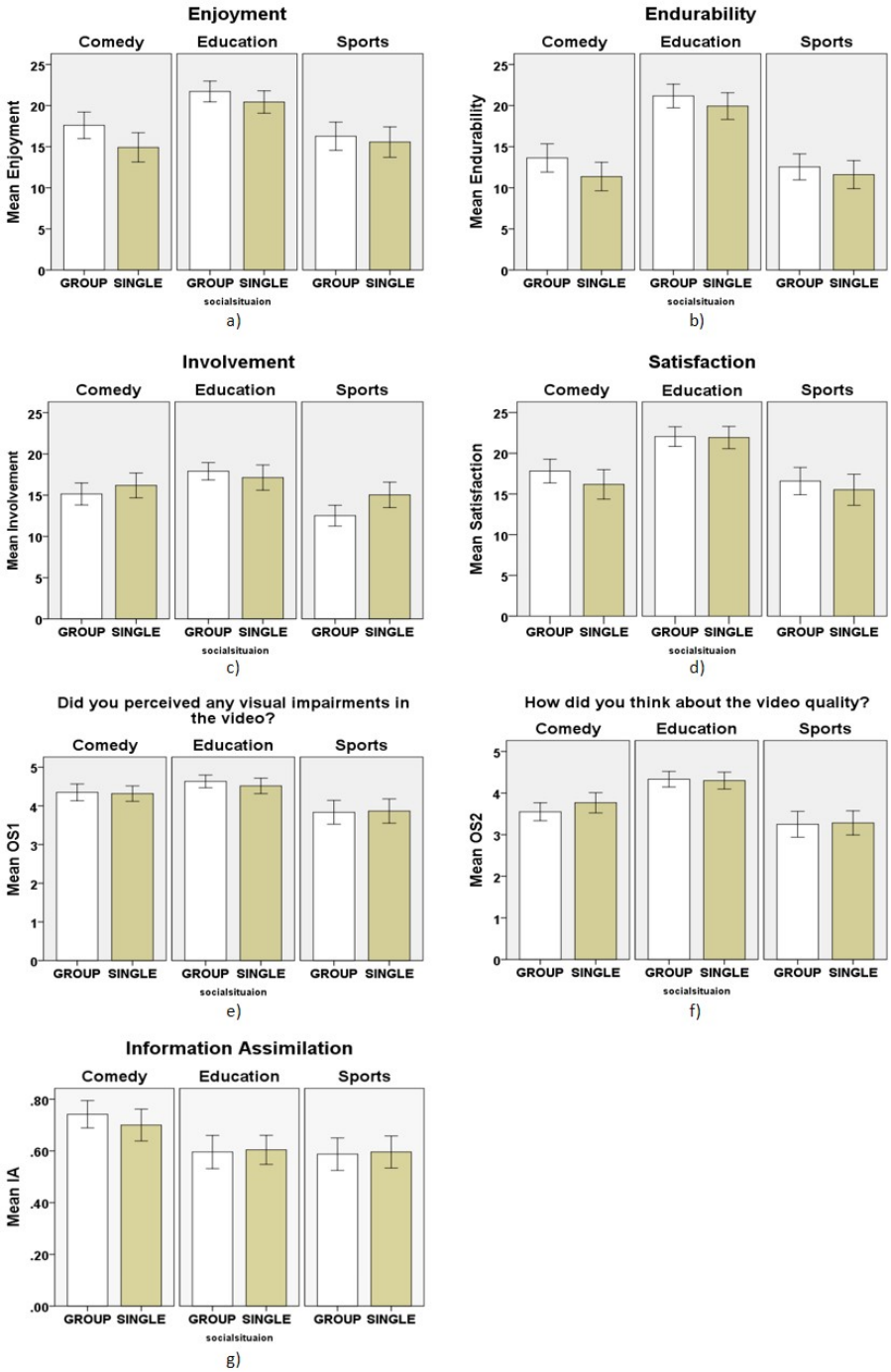

Figure 2.5: The mean AS, OS and IAS for the two social situations and three video genres: (a) the mean AS of enjoyment, (b) the mean AS of endurability, (c) the mean AS of involvement, (d) the mean AS of satisfaction, (e, f) the mean Opinion Scores of the two PQ items, and (g) the mean IA. The white bars give the score for group viewing, while the yellow bars give the score for single viewing. The scores for the three genres (i.e., comedy, education and sports) are shown in three separated columns

2.5.2. THE IMPACT OF SOCIAL CONTEXT ON QOE AND ITS INTERACTION WITH SYSTEM FACTORS

The impact of social context on QoE is visualized in Figure 2.5 for the 4 AS, the 2 OS and the IAS in separate graphs. Each graph directly compares the scores for group viewing to the scores for viewing alone for each of the three genres of video separately. As some of the dependent variables were not normally distributed, we investigated whether the observed differences were significant using the non-parametric factorial analyses on aligned rank data [63]. Here bitrate level and social situation (i.e., single or group) were set as between-subjects factors, whereas video genre (i.e., comedy, education or sport) was investigated as within-subjects factor. The QoE measurements (i.e., the aggregate AS of endurability, enjoyment, involvement and satisfaction, as well as the two OS of PQ and the IAS) were the dependent variables. All the 2-way interactions among the three factors (i.e., social situation, bitrate level and video genre) were included.

The results show that the social situation significantly influenced user's enjoyment ($F(1,116) = 4.228$, $p = 0.042$) and endurability ($F(1,116) = 4.231$, $p=0.042$). As shown in Figures 2.5.a and 2.5.b, participants that watched the videos in group rated enjoyment and endurability of the video experience higher than those participants who watched the videos by themselves. There was no significant interaction between social situation and genre for enjoyment ($F(2, 232)=0.821$, $p=0.441$) and endurability ($F(2,232)=0.577$, $p=0.562$), implying that the increase in both AS when watching the videos in group was independent on the genre of the videos. No significant effect of social situation on involvement was found, however we did find a significant interaction between social situation and genre ($F(2, 232) = 3.141$, $p = 0.045$), suggesting that user's involvement in different social situations varied with video genre. To better understand this interaction effect, the data was split up per video genre, and the Mann-Whitney U-tests were performed with only social situation as the independent factor. The results, presented in Figure 2.5c, show that people who watched the sports videos with friends were significantly less involved in the video than those who watched these videos alone ($U = 1383.5$, $p = 0.028$). We didn't find a significant difference between the two social situations for the education and comedy videos on involvement ($U = 1665.5$, $p = 0.479$ and $U = 1622.5$, $p = 0.350$ respectively). We also didn't find a significant effect of social situation on satisfaction ($F(1,116) = 1.277$, $p = 0.261$) and information assimilation ($F(1, 116)= 0.054$, $p=0.817$), while both QoE aspects also did not show a significant interaction between social situation and video genre.

No significant effect of social situation on PQ was found (for OS1: $F(1,116)=0.182$, $p=0.67$ and for OS2: $F(1,116)=2.312$, $p=0.131$, shown in Figures 2.5.e and 2.5.f), indicating that the sensitivity of participants to artifacts in the video was not affected by the presence of co-viewers. Conversely, a significant effect of bitrate level was found on both PQ items (for OS1: $F(1,116)=51.225$, $p<0.001$ and for OS2: $F(1,116)=101.855$, $p<0.001$). As shown in Figure 2.6, participants clearly recognized the 600 kbps videos as having lower quality and more visible artifacts than the 2000 kps videos with the same genre, independent on the presence or absence of co-viewers.

So, our hypothesis H1 is confirmed at least for the aspects of enjoyment and endurability of QoE. We didn't find an effect of presence of co-viewers on involvement, satisfaction, PQ and information assimilation. We have to reject hypothesis H2a, since the

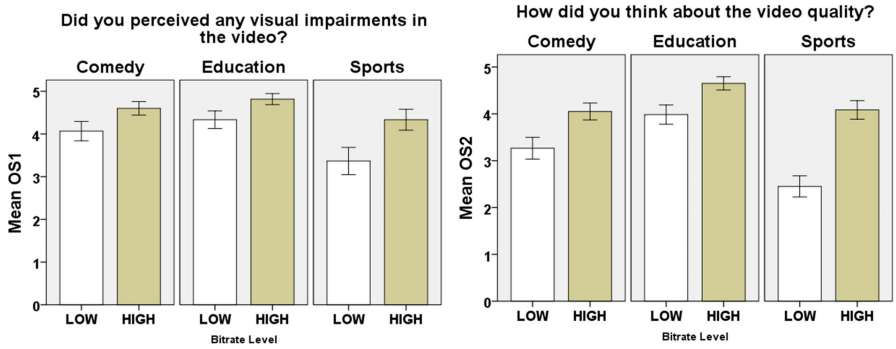

Figure 2.6: The mean OS for the two bitrate levels (600kpbs and 2000kbps) split up for the three video genres. The white bars indicate the low bitrate level, while the yellow bars indicate the high bitrate level. The mean OS per video genre (i.e., comedy, education and sport) are shown in three separated columns.

presence of co-viewers didn't affect the tolerance to artifacts, i.e., the PQ aspect of QoE. Finally, genre only affected the impact of social context on involvement, but not in the hypothesized way. We didn't find an effect of group viewing for the genre comedy, being clearly preferred to be viewed in group, nor for the genre education, being clearly preferred to be viewed alone. The only significant effect found was that involvement increased when the participants viewed the sports video alone. Hence, we also have to reject hypothesis H2b.

Surprisingly, PQ was the only QoE aspect affected by the bitrate level: we didn't find significant effects of bitrate level on enjoyment ($F(1,116) = 0.425$, $p = 0.516$), satisfaction ($F(1,116) = 0.031$, $p = 0.86$), endurability ($F(1,116) = 0.262$, $p = 0.610$), involvement ($F(1,116) = 0.412$, $p = 0.522$) and information assimilation ($F(1,116)=0.002$, $p=0.963$). Thus, we confirmed the general knowledge that low bitrate affects picture quality, but not the assumed consequence that bitrate then also affects QoE.

2.5.3. The impact of user factors on QoE and their interaction with social context

User Interest

To test whether user interest positively correlated with user's QoE, Spearman correlations between the user interest scores and the scores of the seven QoE aspects were calculated per video genre over all participants. As shown in Table 2.6 for the Overall scores, satisfaction and endurablity were consistently and positively correlated with user interest with r-values around 0.3. The r-values were somewhat higher for the video genre comedy than for the other two video genres. Enjoyment and involvement were also positively correlated with user's interest in two out of three genres. Participants interested in comedy and sports enjoyed it more, while participants interested in sports and education felt more involved. No significant correlation was found between user's interest and the two PQ items in any video genre, suggesting that artifact visibility did not change due to user's interest. Similarly, no significant correlation was found between user's in-

Table 2.6: Spearman rank order correlations between user interest and all QoE aspects for the 3 video genres separately. 'Overall' represents the correlations calculated over all participants, whereas 'Single' and 'Group' represent the correlations only over the participants that viewed the videos alone or in group, respectively

			Enjoyment	Satisfaction	Endurability	Involvement
Comedy	Overall	r	0.447**	0.400**	0.361**	0.080
		p	0.000	0.002	0.005	0.547
	Single	r	0.578**	0.452*	0.442*	0.228
		p	0.001	0.012	0.014	0.225
	Group	r	0.286	0.362	0.288	-0.028
		p	0.132	0.054	0.129	0.885
			Enjoyment	Satisfaction	Endurability	Involvement
Sports	Overall	r	0.379**	0.289*	0.319*	0.306*
		p	0.003	0.027	0.014	0.018
	Single	r	0.498**	0.369*	0.363*	0.435*
		p	0.005	0.045	0.049	0.016
	Group	r	0.189	0.192	0.190	0.216
		p	0.326	0.319	0.323	0.261
			Enjoyment	Satisfaction	Endurability	Involvement
Education	Overall	r	0.255	0.262*	0.303*	0.271*
		p	0.052	0.045	0.020	0.038
	Single	r	0.283	0.211	0.322	0.217
		p	0.130	0.263	0.083	0.249
	Group	r	0.229	0.319	0.279	0.317
		p	0.232	0.092	0.142	0.093

			OS1	OS2	IAS
Comedy	Overall	r	0.096	-0.127	-0.104
		p	0.470	0.338	0.264
	Single	r	0.086	-0.083	-0.124
		p	0.650	0.664	0.345
	Group	r	0.084	-0.150	-0.065
		p	0.666	0.436	0.63
			OS1	OS2	IAS
Sports	Overall	r	-0.189	-0.009	0.025
		p	0.151	0.946	0.789
	Single	r	-0.357	-0.270	0.064
		p	0.053	0.150	0.628
	Group	r	-0.050	0.324	-0.006
		p	0.798	0.086	0.963
			OS1	OS2	IAS
Education	Overall	r	-0.127	-0.028	0.059
		p	0.338	0.835	0.527
	Single	r	–0.145	-0.032	0.056
		p	0.445	0.866	0.669
	Group	r	-0.153	0.000	0.066
		p	0.429	1	0.624

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

terest and IA scores, suggesting that the user's ability of assimilating information from the videos did not depend on his/her interest in the video content. In summary, since some QoE aspects were clearly positively related to user's interest, we can confirm hypothesis H3a.

As we were also interested in understanding to what extent the social context might affect the impact of user's interest on QoE, we split up our data per social situation (i.e., single vs. group viewing) and recalculated the Spearman correlations. As shown in Table 2.6, user's enjoyment, satisfaction and endurability were positively correlated with user's interest for comedy and sports videos, when viewed alone. However, no significant correlations were found between the user's interest and any of the QoE aspects when the videos were viewed in group. This indicates that whereas interest may in general affect QoE of users, the presence of co-viewers may suppress this effect, and other factors such as the pleasure of having company during the experience may weigh more. This finding confirms hypothesis H3d, at least for the user factor of user interest and for the comedy and sports video genre.

Immersive Tendency

To test whether individual immersive tendency affects the user's QoE, we calculated the Spearman correlations between the participant's immersive tendency and the seven QoE aspects per video genre over all participants. No significant correlation was found between any QoE aspect and immersive tendency for any video genre. Hence, this rejects hypothesis H3b.

To evaluate hypothesis H3d, we also calculated the Spearman correlations between the personal immersive tendency and the QoE aspects for the participants viewing the videos alone or in group separately. For the comedy and sports videos, immersive tendency was positively correlated with enjoyment ($r=0.367$, $p=0.046$ and $r=0.387$, $p=0.035$, respectively) and endurability ($r=0.362$, $p=0.049$ and $r=0.475$, $p=0.008$, respectively) in the single viewing situation. But in the group viewing situation, no significant correlation was found between immersive tendency and any QoE aspect. Hence, these findings are in line with the findings on user interest; immersive tendency has a positive relation with QoE when sports and comedy videos are watched alone, but not when they are watched in group. In the latter case, group processes may overrule the impact of individual's characteristics on QoE. As such, we also confirm hypothesis H3d for immersive tendency, but again only for the sports and comedy videos.

Impact of demographics on QoE

To evaluate the impact of demographic information on QoE, we specifically focus on a possible effect of gender and cultural background. Since our sample of participants was relatively young, it didn't allow us to make a fair analysis on the effect of age on QoE.

The effect of gender on QoE was investigated using Mann-Whitney U-tests with gender as the independent variable and all QoE aspects (i.e., 4xAS, 2xOS and IAS) as dependent variables. We found a significant effect of gender on involvement ($U=13926.5$, $p=0.031$). In particular, males were easier involved in the videos than females, and as shown in Figure 2.7 this trend was independent on the video genre. No significant influence was found on any other QoE aspect.

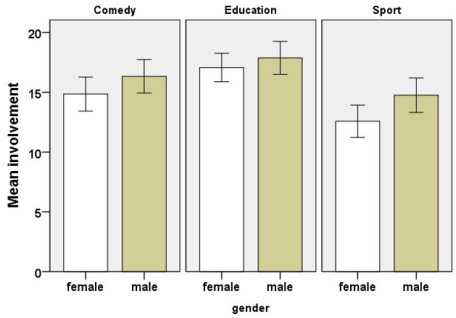

Figure 2.7: The effect of gender on involvement for the three video genres separately. Here the white bars represent the involvement scores for the female participants, while the colored bars represent the scores for the male participants.

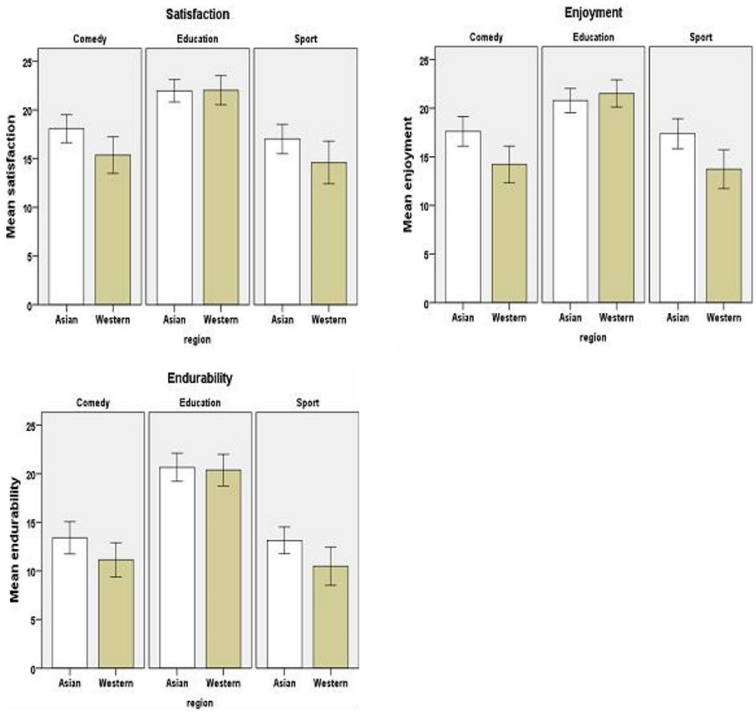

Figure 2.8: The mean score on satisfaction (a), enjoyment (b) and endurability (c) for the two cultural backgrounds and the three video genres. The white bars represent the Asian participants, while the yellow bars represent the Western participants.

The second demographic factor that we examined was the cultural background of the participants. As mentioned earlier, there were 36 Asian participants and 24 Western participants in our study. We split our data into these two groups and ran Mann-Whitney U-tests with all QoE aspects as dependent variables. We found a significant difference between the two groups on their QoE ratings in terms of satisfaction (U=13584.5, p=0.042), enjoyment (U=12979, p=0.008) and endurability (U=13357.5, p=0.023). As shown in Figure 2.8, Asian participants tended to rate their experience (i.e., enjoyment, satisfaction and endurability) higher than Western participants. This trend seems to be more pronounced for the comedy and sports videos than for the education videos. We didn't find a significant difference between the two cultural backgrounds for involvement, the two PQ items, or information assimilation.

We may therefore conclude that some demographic characteristics of the users have an influence on some of the QoE aspects we considered, and therefore our hypothesis H3c is partially supported.

2.6. Discussion

Quality of Experience is a very complex concept and its proper quantification still has several challenges ahead. Based on existing literature, we proposed to measure various aspects of Quality of Experience, including perceptual quality (along two separate dimensions of (1) artifact visibility and (2) overall quality), enjoyment, satisfaction, endurability, involvement and information assimilation. Measurement scales for perceptual quality are well established; conversely, no standardized or universally accepted scales exist for the other aspects. As a result, we based our measurements on existing questionnaires, eventually building a composite questionnaire for QoE. Such questionnaire is a first step towards the definition of a more encompassing tool to measure subjective Quality of Experience; nevertheless, it requires further validation. One of the consequences of this lack of validation is that we don't have insight on the relevance to QoE of each of the aspects we measured. To circumvent this issue, we decided to consider that an independent variable of our study significantly affects QoE when it showed a consistent significant effect on multiple QoE aspects. In practice, this implies that we considered perceptual quality, enjoyment, endurability and satisfaction as equally important aspects of QoE, but obviously more research towards a validated questionnaire is needed.

Multiple factors that potentially influence QoE have been identified in literature [34], but not yet consistently evaluated. In this study we investigated the role of the presence/absence of co-viewers in combination with several system and user characteristics. Our findings showed that the presence of co-viewers increased the participants' level of enjoyment and made them more willing to repeat the experience. Thus, social context has some impact on QoE. We also found an interaction with video genre on involvement: participants who watched the sports videos in presence of their friends were less involved with the videos than those who watched videos alone. This finding might be explained by the fact that sport videos usually have less of a storytelling component compared to education and comedy videos; hence, people may be more willing to engage in the social interaction, since the risk of missing important information in the video is less.

Bitrate level of the video was investigated as a system factor with a possible interaction with social context. We found, as expected, that bitrate level impacted the perceptual quality of the video (the lower the bitrate, the lower the quality). On the other hand, it did not influence any other aspect of QoE, implying that user's satisfaction (or enjoyment, involvement, endurability) of a video could be kept at the same level even if video quality decreased a bit. Actually, the scores on picture visual quality were also marginally correlated with the other QoE aspects as satisfaction, enjoyment, endurability and involvement. This finding provides a relevant insight for multimedia delivery optimization: for guaranteeing enjoyment, and more importantly, willingness to repeat the experience (endurability), factors other than bitrate and consequent artifact visibility play a more prominent role. Among those, the presence of co-viewers and the level of interest in the content of the video should be considered. Further research though is needed to reliably quantify the importance of the different user, system and contextual factors to QoE.

Overall we found multiple QoE aspects that were different between the various types of video genre used in our study. To some extent these findings may be a direct result of differences in experience between different video genres. But these differences may also be related to user and/or contextual factors. For example, user interest and preference for watching a video from a certain genre alone or with friends may have an impact on how a given type of video is experienced. Interest in the video content was measured per genre and was positively correlated with user's satisfaction and endurability. This confirmed the finding of [47]. Yet, interest did not correlate with perceptual video quality, contradicting previous results which showed a link between content desirability and PQ [32]. It is interesting to note in this respect that personal interest in a given genre - and also immersive tendency - correlated with QoE aspects within the single viewing situation, but not in presence of co-viewers. The latter was especially true for the video genres 'comedy' and 'sports', which were previously indicated by participants as video genres preferably watched in a group. These findings imply that personal characteristics such as interest in a video genre and immersive tendency are less important for the viewing experience when watching videos with co-viewers, with respect to other factors such as social interaction.

We investigated the effect of gender and cultural background on QoE. The results showed female participants to be more involved in the viewing experience than male participants. This was not originated by a difference in immersive tendency between males and females ($U=389$, $p=0.513$), which we could have expected given the findings of previous research [37]. In fact, immersive tendency was not found to be correlated to involvement in this study. Related to cultural background, we found that Asian participants rated their QoE higher than Western participants. This might be partially due to different rating habits across cultures. It has been shown that rating behavior might be "area-specific", and that country of origin significantly influenced the performance of a same task [19]. Although these results are in line with our preliminary findings, further research is needed to verify whether other elements, related to the viewing experience rather than to the judgment method, concur in making the differences in rating between cultures significant.

In a final attempt to summarize our results, considering the role played by each factor

in determining QoE and their interaction, we performed an automatic linear modeling including all factors investigated in this study using SPSS 20. The enjoyment was set as target variable while social situation, immersive tendency, user interest, gender, cultural background, bitrate level and video genre were set as inputs. Results showed that video genre, interest, social situation and cultural background were considered as important predictors of enjoyment. But, the resulting linear regression model provided a rather low prediction accuracy ($F(4, 353)=21.727$, $r= 0.447$, $R2=0.199$, $p<0.001$), suggesting that simple linear regression is not sufficient to properly estimate QoE. How to fuse the influencing factors still needs further investigation and possibly involves the use of non-linear modeling tools with higher modeling capabilities (as described, e.g. in Gastaldo et al. [20]).

Although we found some interesting results in this first investigation quantifying QoE aspects including the role of social context, system factors and user factors, our study also has some limitations that warrant additional research. In this first study, we investigated the impact of social context only for a "direct" social context, existing of co-viewers being friends. We intend to extend this investigation of QoE to more social situations in the future, including watching videos with strangers or watching videos while co-viewers are not physically co-located. The latter is a form of social context that currently grows tremendously as a consequence of new platforms for online video viewing and sharing. Obviously, we limited ourselves in this first study to three video genres only; these genres were selected such that we covered the gamut from strong preference for being watched alone to strong preference for being watched in group. We found some differences in QoE aspects related to video genre and in the interaction with user factors, such as interest and immersive tendency, but we need QoE scores for more videos per genre and for more video genres to get to systematic conclusions with respect to content and video genre.

2.7. CONCLUSIONS

IN this chapter, we investigated a set of influencing factors on user's QoE with videos. Our results showed that co-viewing videos with friends increased the user's level of enjoyment and enhanced the endurability of the experience, indicating that social context should be further investigated in relation to QoE and considered also in automated measurements. The presence of co-viewers did not change participant's ability to detect visual artifacts, yet the presence of visible artifacts did not affect the enjoyment and endurability of the viewing experience (as well as any of the other aspects of QoE examined). We may conclude therefore that a pure analysis of the (perceptual effects of the) bitrate is insufficient to properly characterize the entire quality of the viewing experience. User interest also showed significant correlations with the QoE aspects of enjoyment, satisfaction and endurablity. This effect however was suppressed by the presence of co-viewers, further corroborating our hypothesis that social context plays a major role in determining QoE. Finally, cultural background was shown to impact QoE ratings, whereas gender was shown to affect only involvement. These findings are a first, but concrete step towards defining an encompassing model of QoE appreciation, that goes beyond the sole analysis of system factors, and takes interaction effects of user, contextual and system factors into account as well. Such a model would be extremely appeal-

ing for optimizing online multimedia delivery, when limited bandwidth is available in the whole signal chain. Towards such a model, however, additional research questions, such as the relative importance of various QoE aspects, the impact of video genre, and the relevance of additional user and system factors, still need to be investigated.

2.8. REFERENCE

[1] Antons, J.-N., Arndt, S., Schleicher, R. and Möller, S. 2014. Brain Activity Correlates of Quality of Experience. Quality of Experience, Springer, 109-119.

[2] Asghar, J., Le Faucheur, F. and Hood, I. 2009. Preserving video quality in IPTV networks. Broadcasting, IEEE Transactions on, 55, 2, 386-395.

[3] Balachandran, A., Sekar, V., Akella, A., Seshan, S., Stoica, I. and Zhang, H. 2012. A quest for an internet video quality-of-experience metric. Proceedings of the 11th ACM workshop on hot topics in networks, ACM, 2012.

[4] Bracken, C. C. 2005. Presence and image quality: The case of high-definition television. Media Psychology, 7, 2, 191-205.

[5] Bradley, M. M., Codispoti, M., Sabatinelli, D. and Lang, P. J. 2001. Emotion and motivation II: sex differences in picture processing. Emotion, 1, 3, 300.

[6] Brooks, P. and Hestnes, B. 2010. User measures of quality of experience: why being objective and quantitative is important. Network, IEEE, 24, 2, 8-13.

[7] Chikkerur, S., Sundaram, V., Reisslein, M. and Karam, L. J. 2011. Objective video quality assessment methods: A classification, review, and performance comparison. Broadcasting, IEEE Transactions on, 57, 2, 165-182.

[8] Chorianopoulos, K. and Lekakos, G. 2008. Introduction to social TV: Enhancing the shared experience with interactive TV. Intl. Journal of Human–Computer Interaction, 24, 2, 113-120.

[9] Cisco, I. 2012. Cisco visual networking index: Forecast and methodology, 2011–2016. CISCO White paper, 2011-2016.

[10] Cortina, J. M. 1993. What is coefficient alpha? An examination of theory and applications. Journal of applied psychology, 78, 1, 98.

[11] De Moor, K., Mazza, F., Hupont, I., Quintero, M. R., Mäki, T. and Varela, M. 2014. Chamber QoE: a multi-instrumental approach to explore affective aspects in relation to quality of experience. Human Vision and Electronic Imaging XIX. Vol. 9014, International Society for Optics and Photonics.

[12] De Moor, K., Quintero, M. R., Strohmeier, D. and Raake, A. 2013. Evaluating QoE by means of traditional and alternative subjective measures: an exploratory'living room lab'study on IPTV. Vienna, Austria.

[13] De Pessemier, T., De Moor, K., Joseph, W., De Marez, L. and Martens, L. 2013. Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching. Broadcasting, IEEE Transactions on, 59, 1, 47-61.

[14] Desmet, P. 2005. Measuring emotion: Development and application of an instrument to measure emotional responses to products. Funology,Springer, 111-123.

[15] Dobrian, F., Sekar, V., Awan, A., Stoica, I., Joseph, D., Ganjam, A., Zhan, J. and Zhang, H. 2011. Understanding the impact of video quality on user engagement. ACM SIGCOMM Computer Communication Review, 41, 4, 362-373.

[16] Engeldrum, P. G. 2000. Psychometric scaling: a toolkit for imaging systems development. Imcotek Press.

[17] Fiedler, M., Hossfeld, T. and Tran-Gia, P. 2010. A generic quantitative relationship between quality of experience and quality of service. Network, IEEE, 24, 2, 36-41.

[18] Fröhlich, P., Baillie, L. and Schatz, R. 2006. Exploring the Joint iTV Experience. FTW Technical Report, FTW-TR-2006-005

[19] Gardlo, B., Ries, M., Hossfeld, T. and Schatz, R. 2012. Microworkers vs. facebook: The impact of crowdsourcing platform choice on experimental results. Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on. IEEE, 35-36.

[20] Gastaldo, P., Zunino, R. and Redi, J. 2013. Supporting visual quality assessment with machine learning. EURASIP Journal on Image and Video Processing, 2013, 1, 1-15.

[21] Ghinea, G. and Thomas, J. 2005. Quality of perception: user quality of service in multimedia presentations. IEEE Transactions on Multimedia, 7, 4, 786-789.

[22] Gulliver, S. R. and Ghinea, G. 2006. Defining user perception of distributed multimedia quality. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2, 4, 241-257.

[23] Gulliver, S. R. and Ghinea, G. 2004. Stars in their eyes: What eye-tracking reveals about multimedia perceptual quality. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 34, 4, 472-482.

[24] Hemami, S. S. and Reibman, A. R. 2010. No-reference image and video quality estimation: Applications and human-motivated design. Signal processing: Image communication, 25, 7, 469-481.

[25] Huynh-Thu, Q. and Ghanbari, M. 2008. Temporal aspect of perceived quality in mobile video broadcasting. Broadcasting, IEEE Transactions on, 54, 3, 641-651.

[26] Hyder, M., Crespi, N., Haun, M. and Hoene, C. 2012. Are QoE requirements for multimedia services different for men and women? Analysis of gender differences in forming QoE in virtual acoustic environments. International Multi Topic Conference, Springer, 200-209.

[27] Ickin, S., Wac, K., Fiedler, M., Janowski, L., Hong, J.-H. and Dey, A. K. 2012. Factors influencing quality of experience of commonly used mobile applications. Communications Magazine, IEEE, 50, 4, 48-56. [28] Jumisko-Pyykkö, S. 2011. User-centered quality of experience and its evaluation methods for mobile television. Doctoral thesis, Tampere University of Technology, Tampere.

[29] Kilkki, K. 2008. Quality of Experience in Communications Ecosystem. J. UCS, 14, 5, 615-624.

[30] Kim, H.-J., Lee, D. H., Lee, J. M., Lee, K.-H., Lyu, W. and Choi, S.-G. 2008. The QoE evaluation method through the QoS-QoE correlation model. Networked Computing and Advanced Information Management, 2008. NCM'08. Fourth International Conference on. Vol. 2, IEEE,719-725.

[31] Klaue, J., Rathke, B. and Wolisz, A. 2003. Evalvid–A framework for video transmission and quality evaluation. International Conference on Modelling Techniques and Tools for Computer Performance Evaluation.Springer, 255-272.

[32] Kortum, P. and Sullivan, M. 2010. The effect of content desirability on subjective

video quality ratings. Human factors: the journal of the human factors and ergonomics society, 52, 1, 105-118. [33] Kubey, R. and Csikszentmihalyi, M. 2013. Television and the quality of life: How viewing shapes everyday experience. Routledge.

[34] Le Callet, P., Möller, S. and Perkis, A. 2012. Qualinet white paper on definitions of quality of experience. European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003).

[35] Lee, B. and Lee, R. S. 1995. How and why people watch TV: Implications for the future of interactive television. Journal of advertising research, 35, 6, 9-18.

[36] Lin, W. and Jay Kuo, C.-C. 2011. Perceptual visual quality metrics: A survey. Journal of Visual Communication and Image Representation, 22, 4, 297-312.

[37] Ling, Y., Nefs, H. T., Brinkman, W.-P., Qu, C. and Heynderickx, I. 2013. The relationship between individual characteristics and experienced presence. Computers in Human Behavior, 29, 4, 1519-1530. [38] Marcus, A. 2006. Cross-cultural user-experience design. International Conference on Theory and Application of Diagrams. Springer, 16-24.

[39] Mok, R. K., Chan, E. W. and Chang, R. K. 2011. Measuring the quality of experience of HTTP video streaming. Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on. IEEE, 485-492.

[40] Moore, K. 2011. 71% of online adults now use video-sharing sites. Pew Internet and American Life Project.

[41] Morrison, M. and Krugman, D. M. 2001. A look at mass and computer mediated technologies: Understanding the roles of television and computers in the home. Journal of Broadcasting & Electronic Media, 45, 1, 135-161.

[42] Murray, N., Qiao, Y., Lee, B., Muntean, G.-M. and Karunakar, A. 2013. Age and gender influence on perceived olfactory & visual media synchronization. Multimedia and Expo (ICME), 2013 IEEE International Conference on. IEEE, 1-6.

[43] Naumann, A. B., Wechsung, I. and Hurtienne, J. 2010. Multimodal interaction: A suitable strategy for including older users? Interacting with Computers, 22, 6, 465-474.

[44] O'Brien, H. L. and Toms, E. G. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. Journal of the American Society for Information Science and Technology, 59, 6, 938-955.

[45] Oehlberg, L., Ducheneaut, N., Thornton, J. D., Moore, R. J. and Nickell, E. 2006. Social TV: Designing for distributed, sociable television viewing.Intl. Journal of Human–Computer Interaction 24.2: 136-154.

[46] Pérez, P., Macías, J., Ruiz, J. J. and García, N. 2011. Effect of packet loss in video quality of experience. Bell Labs Technical Journal, 16, 1, 91-104. [47] Palhais, J., Cruz, R. S. and Nunes, M. S. 2012. Quality of Experience Assessment in Internet TV. International Conference on Mobile Networks and Management. Springer, 261-274.

[48] Rainer, B., Waltl, M., Cheng, E., Shujau, M., Timmerer, C., Davis, S., Burnett, I., Ritz, C. and Hellwagner, H. 2012. Investigating the impact of sensory effects on the quality of experience and emotional response in web videos. Quality of Multimedia Experience

(QoMEX), 2012 Fourth International Workshop on.IEEE, 278-283.

[49] Recommendation, I. 2002. 500-11, Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, Geneva, Switzerland, 4, 2.

[50] Recommendation, I. 1994. E. 800: Terms and definitions related to quality of service and network performance including dependability. ITU-T August 1994.

[51] Redi, J. A. 2013. Visual quality beyond artifact visibility. In Human Vision and Electronic Imaging XVIII (Vol. 8651, p. 86510N). International Society for Optics and Photonics.

[52] Scheinkman, J. A. 2008. Social interactions. The New Palgrave Dictionary of Economics, 2.

[53] Schubert, T., Friedmann, F. and Regenbrecht, H. 2001. The experience of presence: Factor analytic insights. Presence, 10, 3, 266-281.

[54] See-To, E. W., Papagiannidis, S. and Cho, V. 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, 79, 8, 1484-1494.

[55] Serral-Gracià, R., Cerqueira, E., Curado, M., Yannuzzi, M., Monteiro, E. and Masip-Bruin, X. 2010. An overview of quality of experience measurement challenges for video applications in IP networks. International Conference on Wired/Wireless Internet Communications, Springer, 252-263.

[56] SG12, I. 2007. Definition of Quality of Experience. TD 109rev2 (PLEN/12), Geneva, Switzerland, 16-25.

[57] Staelens, N., Moens, S., Van den Broeck, W., Marien, I., Vermeulen, B., Lambert, P., Van de Walle, R. and Demeester, P. 2010. Assessing quality of experience of IPTV and video on demand services in real-life environments. Broadcasting, IEEE Transactions on, 56, 4, 458-466.

[58] Tercek, R. 2011. Trends in Multichannel TV and Online Video in the United States.Ontario Media Development Corporation.

[59] Wang, D., Speranza, F., Vincent, A., Martin, T. and Blanchfield, P. 2003. Toward optimal rate control: a study of the impact of spatial resolution, frame rate, and quantization on subjective video quality and bit rate. Visual Communications and Image Processing 2003. Vol. 5150. International Society for Optics and Photonics, 198-210.

[60] Wechsung, I., Schulz, M., Engelbrecht, K.-P., Niemann, J. and Möller, S. 2011. All users are (not) equal-the influence of user characteristics on perceived quality, modality choice and performance. Proceedings of the Paralinguistic information and its integration in spoken dialogue systems workshop, Springer, 175-186.

[61] Westerink, J. H. and Roufs, J. A. 1989. Subjective image quality as a function of viewing distance, resolution, and picture size. SMPTE journal, 98, 2, 113-119.

[62] Witmer, B. G. and Singer, M. J. 1998. Measuring presence in virtual environments: A presence questionnaire. Presence: Teleoperators and virtual environments, 7, 3, 225-240.

[63] Wobbrock, J. O., Findlater, L., Gergle, D. and Higgins, J. J. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. Proceedings of the SIGCHI conference on human factors in computing systems, ACM, 143-146.

[64] Wolters, K. M., Engelbrecht, K.-P., Gödde, F., Möller, S., Naumann, A. and Schleicher, R. 2010. Making it easier for older people to talk to smart homes: the effect of early help prompts. Universal Access in the Information Society, 9, 4, 311-325.

[65] Yamori, K. and Tanaka, Y. 2004. Relation between willingness to pay and guaranteed minimum bandwidth in multiple-priority service. Communications, 2004 and the 5th International Symposium on Multi-Dimensional Mobile Communications Proceedings. The 2004 Joint Conference of the 10th Asia-Pacific Conference on. IEEE, 113-117.

[66] Zhu, Y., Heynderickx, I. and Redi, J. A. 2014. Alone or together: measuring users' viewing experience in different social contexts.Human Vision and Electronic Imaging XIX. Vol. 9014. International Society for Optics and Photonics.

3

QoE Prediction for Enriched Assessment of Individual Video Viewing Experience

This chapter is based on Y. Zhu, A. Hanjalic , J. A. Redi, *QoE Prediction for Enriched Assessment of Individual Video Viewing Experience*, Proceedings of the 2016 ACM on Multimedia Conference. ACM, 2016: 801-810.

Abstract

Most automatic Quality of Experience (QoE) assessment models have so far aimed at predicting the QoE of a video as experienced by an average user, and solely based on perceptual characteristics of the video being viewed. The importance of other characteristics, such as those related to the video content being watched, or those related to an individual user have been largely neglected. This is suboptimal in view of the fact that video viewing experience is individual and multifaceted, considering the perceptual quality (related to coding or network-induced artifacts), but also other – more hedonic - aspects, like enjoyment. In this chapter, we propose an expanded model which aims to assess QoE of a given video, not only in terms of perceptual quality but also of enjoyment, as experienced by a specific user. To do so, we feed the model not only with information extracted from the video (related to both perceptual quality and content), but also with individual user characteristics, such as interest, personality and gender. We assess our expanded QoE model based on two publicly available QoE datasets, namely i_QoE and CP-QAE-I. The results show that combining various types of characteristics enables better QoE prediction performance as compared to only considering perceptual characteristics of the video, both when targeting perceptual quality and enjoyment.

3.1. Introduction

With the increasing volume of online video consumption, users' expectations in terms of Quality of Experience (QoE) are growing rapidly. According to its most widespread definition [18], "*Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state*". As such, QoE has been established as the main indicator of the user satisfaction with the video viewing experience. Insufficient QoE will be less and less accepted by users, leading them to quit the experience or even change the delivering service/application altogether [23]. Intelligent mechanisms to assess, in real time, the quality of the viewing experience of a user, and to enhance it when possible are, therefore, critical for the adoption of future video delivery services.

The definition given above indicates that QoE is an abstract concept, which is difficult to quantify. Because of this, in the context of video consumption, QoE has been mostly identified with the more easily quantifiable concept of perceptual (visual) quality (PQ) [3]. PQ refers to the perceptual impact of the presence of network-related impairments in video, such as buffering events or transmission errors, and/or visible artifacts resulting from e.g. video coding (e.g., blockiness or blur). Consequently, automatic QoE assessment efforts have focused on estimating the annoyance that artifacts and impairments cause to users [5]. This process has relied mostly on video and/or network condition analysis [3], extracting perceptual characteristics of the video to serve as input to the QoE assessment model. The latter has generally targeted the prediction of one PQ value per video (or service or application), in a user–agnostic way. That is, PQ predictions are the same for all users, mimicking the perception of an "average user", typically measured subjectively through Mean Opinion Scores (MOS) [32].

On the other hand, the QoE definition suggests, just as the recent evidence provided

in [20], that QoE is a multifaceted concept, of which PQ is only one aspect. This points to the need to complement PQ by measuring other aspects of QoE, such as the level of enjoyment of the experience [11, 40]. Users tend to be willing to repeat and share enjoyable experiences, thus offering enjoyable experiences can enhance user's loyalty to services and applications [22]. Enjoyment reflects the hedonic part of QoE, being a pleasurable response to media use [33]. As such, it can be considered complementary to PQ (with which it has been shown to be only poorly correlated [40]) in properly characterizing QoE.

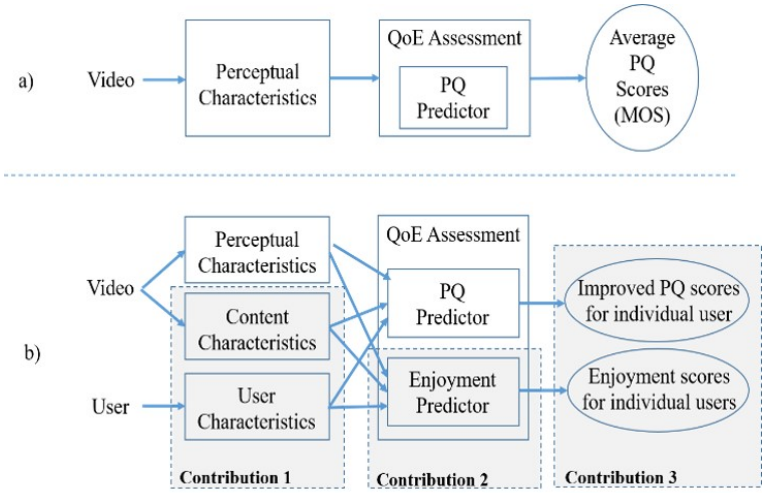

Figure 3.1: The traditional approach to assessing video QoE, b) The proposed expanded QoE assessment model targeting individual scores representing two selected QoE aspects, PQ and enjoyment.

In order to predict enjoyment, the perceptual characteristics mentioned above may not be informative enough. Better results could be achieved by also taking into account content characteristics [41], providing clues about the video content being watched, and user characteristics, such as personal interest [40], personality [29], or cultural background [21]. Especially the inclusion of user characteristics would enable us to consider individual differences when assessing QoE, resulting in more fine-grained predictions than those expressed in MOS or similar "one-fits-all" measures. In addition, since user and content characteristics [7] have been shown to have influence on viewing experience in general [26, 40], feeding QoE assessment models with this type of information could result in better predictions of various QoE aspects, thus not only enjoyment, but also PQ. In view of the rationale given above, we define the contribution of this chapter as specified in the following items:

Contribution 1: We expand the traditional QoE assessment paradigm (Figure 3.1a) to take as input not only the perceptual characteristics, but also the user characteristics and content characteristics. Specifically regarding the content characteristics, we focus on the extraction of information on the affective charge of a video [12, 34]. Regarding user characteristics, we consider personality, gender, interest and cultural background of the specific user for which the QoE needs to be assessed.

Contribution 2: We also expand the traditional QoE assessment paradigm in terms of the output: we predict not only the PQ but also the level of enjoyment that a user experiences with a video. The predictions of both PQ and enjoyment are based on the expanded set of characteristics, as explained in the previous item.

Contribution 3: In contrast to the current practice of targeting the QoE prediction for an average user via MOS, we produce individual PQ and enjoyment predictions reflecting the opinion that a specific user has of a specific video.

The contributions with respect to the traditional video QoE assessment paradigm (Figure 3.1.a) are illustrated in Figure 3.1.b. We note that we are not interested in delivering a new, fully functional QoE model at this stage, but rather in answering the following research questions:

RQ1: Does the expanded set of characteristics help improve the prediction of PQ and enjoyment for individual users?

RQ2: If so, which of these characteristics are most informative for which QoE aspect?

In the following section, we review the literature on perceptual quality and enjoyment prediction in the domain of video consumption, as well as knowledge on individual differences in QoE assessment. We then present our proposed expanded QoE prediction paradigm in Section 3.3 and the experimental setup for its validation in Section 3.4. The results are presented in Section 3.5, while the discussion in Section 3.6 summarizes the main insights and sets targets for future research.

3.2. Related work

3.2.1. Perceptual video quality assessment

Methods for the automatic PQ assessment are also known as video quality metrics and rely mainly on the analysis of the decoded bitstream (i.e., accessing pixel values), the encoded bitstream (i.e., at a packet level) or both. Metrics that analyze the decoded bitstream use information on the (audio-) visual signal to compute the PQ. Hence, they do not need any information about the video delivery system under testing and are fully unobstrusive [3, 5]. Beside the so-called “data metrics” (e.g. MSE and PSNR), which are known to correlate poorly with user perceptions, “picture metrics” [37] have been developed based on the analysis of video content and distortion types. Within this family, metrics either predict PQ by directly modeling mechanisms of the Human Vision System (HVS) deemed relevant to video quality (e.g., color perception or contrast sensitivity, as in the case of VQM [24]) or by inferring artifact visibility from statistical properties and/or visual features of the video (e.g., blockiness or blur, as in the case of e.g. MOVIE [30]). Albeit accurate in their predictions, these metrics usually rely on the availability of a lossless source video as reference. As a result, they are not applicable to optimize the quality of multimedia delivery in real time [5]. Recently, many no-reference metrics [13, 28] have also been proposed, which work only on the degraded video and do not need a reference, yet with room for improving performance.

In practice, up until now, the most reliable way to measure PQ is to perform subjective experiments. According to [32], subjective PQ can be quantified in terms of Mean Opinion Scores (MOS), i.e. the average of the (numerical) judgments expressed by users relative to their perceptual satisfaction with respect to a video [5, 32]. The MOS mea-

sure is popular since it can express the perceptual quality of videos in a commonly understandable way [32]. Therefore, most objective QoE metrics, as listed above, are designed to predict MOS, and most QoE (video) datasets report, along with the tested sequences, only the corresponding MOS, rather than the individual PQ scores [17], leaving little space for investigating individual differences in PQ perception. As indicated by our Contribution 3, in this chapter we abandon the MOS-based PQ assessment paradigm and explore possibilities for individual PQ assessment.

3.2.2. VIDEO ENJOYMENT ASSESSMENT

Video enjoyment has been tied closely to the amount of positive emotional pleasure (i.e., high arousal, positive valence) that a video presents [41]. Lately, however, evidence has been brought that video enjoyment doesn't only relate to videos with positive valence. Users may also enjoy videos portraying events with negative valence, such as horror or dramatic scenes [6]. A broad range of research has been carried out in order to measure/estimate affective content automatically [4]. For instance, several studies investigated the influence that different types of segments or shots, the use of color and audio characteristics (extensive review in [4, 39]) have in determining the video affective content. Others modeled affective content by combining multiple audio-visual characteristics extracted from the video [12].

Since the analysis of affective video content relies on features that are extracted from video data, it enables one to obtain insights about the level of enjoyment only in general, for an average user, similarly as in the case of PQ. Here, again, the challenge remains to learn about the individual enjoyment while watching a video, which can be seen as an outcome of the dynamic interaction between the affective content of the video, user characteristics (e.g., personality, personal interest), and user's current mood [7]. We pursue this challenge in this chapter.

3.2.3. USER CHARACTERISTICS INFLUENCING QOE

According to [18], a number of user characteristics (human factors) influence QoE, including demographics (e.g., age, gender), interest, personality and cultural background. With respect to demographics, males are reported to get more easily immersed in the content of a video than females [40], and older users were shown to have higher requirements for QoE as compared to younger users [26]. Interest in video content also counts, as users have been shown to be more tolerant to visual impairments when they are interested in the content of the video [26]. Regarding personality, extraversion (a personality trait describing enthusiastic and talkative individuals [2]) has been found to have a positive impact on users' video enjoyment [36]. Furthermore, cultural background was shown to capture individual differences in rating the QoE for a given video content [40]. Finally, user's mood is another influencing factor of QoE, and specifically when related to the user's intent to seek pleasant experiences [7]. For example, if a user is tense and eager to relax, he/she may enjoy a comedy show more than an intense action movie, even if he/she would normally prefer watching such movie.

To the best of the authors' knowledge, hardly any attempt has been made at incorporating user characteristics into automatic QoE assessment, mostly due to the complexity of retrieving personal user information in an unobtrusive fashion. We witness, how-

ever, rapidly emerging channels for collecting information about users via the services they subscribe to, e.g., movie preferences on Netflix (http://netflix.com), and large advances in user modeling (e.g. natural language processing to infer user personality from textual contributions to social media [10]). Therefore, the time when user information will be easily attainable in real time for automatic QoE assessment is not far, and it is worthwhile starting looking into how to incorporate this type of information in QoE assessment models. Making a substantial step in this direction is one of the objectives and contributions of this chapter.

3.3. The proposed QoE model

As illustrated in Figure 3.1a, the prediction of QoE in a video viewing scenario has been approached so far according to the following three principles:

1. QoE was simplified to Perceptual Quality (PQ) only,
2. PQ was assessed using a limited set of perceptual characteristics related to the visibility of artifacts and other video impairments,
3. QoE was expressed in terms of MOS, neglecting individual differences among users.

QoE is, however, more than PQ alone, and the set of factors potentially influencing QoE is much broader than the perceptual characteristics considered so far. Especially the user characteristics, if taken into account, could not only help improve the prediction of various QoE aspects, but could also bring differentiation in QoE predictions across individual users.

In this chapter, we propose a new QoE prediction model in which we overcome the three limitations mentioned above, targeting QoE prediction for individual users and broadening the prediction scope by producing the scores not only for PQ, but also for enjoyment. The model consists of the modules depicted in Figure 3.1b. In the following subsections, we elaborate on the realization of each of the modules, starting with the model input, i.e., the experience characteristics. Each type of experience characteristics considered in this chapter, thus either the perceptual, content or user characteristics, is represented by a set of numerical indicators that feed the prediction module. We note that the set of characteristics deployed in this chapter is not intended to be exhaustive. Instead, as indicated by the research questions posed in Section 1, we aim at discovering whether combining characteristics of different types brings improvement in predicting different QoE aspects, and which types of characteristics are more informative in this respect. Based on these insights, a more elaborate analysis of characteristics per type can be performed in future work.

3.3.1. Perceptual characteristics

A great number of perceptual characteristics has been proposed so far, mainly targeting the prediction of the PQ of videos [3, 5]. In our study, we decided to use characteristics from a well-known no-reference model working at the decoded-bitstream level [19, 28]. We chose to go for off-the-shelf indicators rather than designing new ones because the focus of this work is to unveil the role of different types of influencing factors in predicting QoE, rather than to improve the performance of PQ models according to the tradi-

tional approach (Figure 3.1a). Also note that we have not considered temporal impairments (e.g., re-buffering or packet-loss) at this stage, although this is of great interest for future study.

We consider 44 perceptual characteristics (hereafter referred to as *PC*s). Thirty-six of those are computed based on the Natural Scene Statistics (NSS) model [19] at each frame m ($m = 1, \ldots M$) and are denoted as $PC_k(m)$, where $k = 1,2,\ldots 36$. The NSS features are parameters describing the shape of the distribution of (transformed) pixel values in distorted images. Depending on the level of distortion in an image, the distribution changes, and so do the parameters describing it, revealing the perceptual impact of image distortions. They are computed as follows. First, each frame m is partitioned into squared patches of $n \times n$ pixels. The sharpness of each patch is computed as specified in [19], and only those patches whose sharpness value is higher than a certain threshold (0.75 in this study) are selected. Within the selected patches at frame m, intensity values are transformed by applying mean removal and divisive normalization. The transformed values are then used to fit a Generalized Gaussian Distribution (GGD), which is known to represent the distribution of such values in unimpaired images:

$$f_m(x;\, \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} \exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right) \tag{3.1}$$

where the gamma function is defined as $\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t}\, dt, \quad a > 0$ and x are the transformed patch intensity values. To evaluate the impact of distortions at each frame, the products of adjacent transformed values (in the horizontal, vertical, and diagonal orientations) are used to fit an Asymmetric Generalized Gaussian Distribution, described by the parameters $\gamma,\ \beta_l, \beta_r$:

$$f(x;\, \gamma,\ \beta_l, \beta_r) = \begin{cases} \frac{\gamma}{(\beta_l+\beta_r)\Gamma(\frac{1}{\gamma})} \exp\left(-\left(\frac{-x}{\beta_l}\right)^{\gamma}\right) & \forall x < 0 \\ \frac{\gamma}{(\beta_l+\beta_r)\Gamma(\frac{1}{\gamma})} \exp\left(-\left(\frac{-x}{\beta_r}\right)^{\gamma}\right) & \forall x \geq 0 \end{cases} \tag{3.2}$$

Finally, the mean of the distribution is computed as well:

$$\eta = (\beta_r - \beta_l) \frac{\Gamma(\frac{2}{\gamma})}{\Gamma(\frac{1}{\gamma})} \tag{3.3}$$

The parameters $\alpha,\ \beta, \gamma,\ \beta_l, \beta_r,\ \eta$ that shape these distributions are known to capture the differences between lossless and distorted images, as well as the severity of these distortions [19]. In order to capture multi-scale behavior, the set of 18 parameters ($\gamma,\ \beta_l, \beta_r,\ \eta$ being calculated for 4 orientations, plus $\alpha,\ \beta$) is computed for two patch sizes (i.e., 96x96 and 48x48 in this case). Thus, eventually, 36 NSS *PCs* are computed for each frame. We then average their values across the M video frames to obtain the values PC_k, $k = 1,2,\ldots 36$, that feed our model. The temporal variation of the frame mean DC coefficients is the PC_{37}. This *PC* represents sudden local changes in a video, which may arise from various temporal distortions. Six statistical DCT *PCs* (represented as PC_{38} to PC_{43}) are also computed from each frame difference using the spatio-temporal model described in [28]. These *PCs* describe the distribution of DCT coefficients across the

frame differences of the video, and have been shown to be able to reflect the perceptual impact brought about by spatio-temporal artifacts in video. Finally, the PC_{44} relates to motion estimation, and is computed based on the coherence of motion vectors in strength and direction (the predefined window size is 10). PC_{44} denotes variations in local motion due to temporal distortions. Details regarding its implementation can be found in [28].

3.3.2. Content characteristics

As described in Section 3.2.2, an insight in the enjoyment elicited in a user while watching a video could be obtained from the affective aspects of the video content being watched [41]. The underlying assumption is, however, based more on psychological studies than on empirical measurements. It is therefore unclear (a) to which extent enjoyment is indeed related to affective video content and (b) which aspects of the video related to eliciting affective reactions are most informative for drawing conclusions about the enjoyment. In this chapter we aim at providing some answers to these questions.

Due to the complexity of the problem, we focus only on one affective dimension, *arousal* (i.e., from excited to calm), and build on a set of proven arousal-related audiovisual features, namely *motion activity*, *sound energy*, *hue ratio* and *shot change rate* [12, 34]. Hereafter we refer to these features as content characteristics (*CC*). Again, we rely on existing work, as the improvement of the affective video content representation is not the goal of this chapter.

The *motion activity* [12] is computed as:

$$MA = \frac{1}{M}\sum_{m=0}^{M-1}\frac{100}{N_m|\vec{v}_{MAX}(m)|}\left(\sum_{i=1}^{N_m}|\vec{v}_i(m)|\right) \tag{3.4}$$

In (3.4), the overall magnitude of all (N_m) motion vectors $\vec{v}_i(m)$ between two adjacent frames m and m+1 is normalized by the length of the longest motion vector $\vec{v}_{MAX}(m)$ at frame m. The obtained values are then averaged across all M frames to yield the motion activity (*MA*) value.

The *sound energy* [12] is computed from the audio track of the videos. The sound energy E_m at frame m is defined as the sum of the power spectrum of the audio samples corresponding to frame m. The overall *sound energy* E is then computed as the mean of E_m across all frames.

The *hue ratio* [31] is computed as:

$$HR = \frac{\sum_{m=1}^{M}\frac{G_m}{H_m}}{M} \tag{3.5}$$

where H_m is the total number of pixels in frame m while G_m is the number of green pixels in frame m.

The *shot change rate* [12] represents the level of dynamics in the video content and is defined as

$$S = \frac{\sum_{m=1}^{M}100e\hat{}((1-(m_n-m_p))/\delta)}{M} \tag{3.6}$$

where m_n and m_p are the frame indexes of the two closest shot boundaries to the left and right of frame m. $\delta$is a constant value, set as recommended in [12].

3.3.3. USER CHARACTERISTICS

We make use of the information provided directly (through self-report) from the users that judged their video experience. Specifically, we rely on the user information collected in [40] and [29] capturing the *interest, immersive tendency, personality, cultural background* and demographics, as described in more detail below.

Personal *interest* is defined as the level of prior interest in the (genre of) the video that the user is about to experience and judge. Similarly, we include values that quantify the user's *Immersive Tendency* (*IT*), which reflects how easily a user gets involved in a particular task [38], in this case specifically in the content of the video. It is argued that a high level of involvement may result in high satisfaction [22].

The *personality* characteristics included in this study represent the "the big five" personality traits i.e., *openness, conscientiousness, extraversion, agreeableness*, and *neuroticism* [2]. *Cultural background* is also represented through a set of six characteristics, namely *power distance, individualism, uncertainty, masculinity, pragmatism*, and *indulgence*. Finally, we include some user demographics information as well, i.e., *gender* and *origin*. The *origin* was defined based on the user's nationality.

The proposed model, depicted in Figure 3.1b, targets the prediction of both PQ and enjoyment based on the same, broadened set of characteristics. The decision to predict PQ and enjoyment independently is due to the fact that, although these two aspects of QoE are not necessarily uncorrelated (they have been found to be poorly although significantly positively correlated [40]), the nature of their relationship remains mostly vague. Hence, in this exploratory study we target their prediction separately, and leave the investigation of their interdependencies to future work. The prediction module per QoE aspect can be implemented in many ways, from a linear combination of the input characteristics to more complex non-linear models. In addition, a feature selection step before the prediction process may be needed depending on the selected model [8]. In this chapter, we choose for a simple implementation using a linear classifier, and a more complex Support Vector Machine, as explained and justified in more detail in Section 3.4.2

3.4. EXPERIMENAL SETUP

IN this section we describe the experimental setup through which we implemented and evaluated the proposed QoE prediction model. The section covers dataset description (Section 3.4.1), predictor implementation (Section 3.4.2), and the evaluation procedure (Section 3.4.3).

3.4.1. DATASET DESCRIPTION

To the best of our knowledge, only two public datasets, namely i_QoE[1] and CP-QAE-I[2], meet our requirements, i.e., include user characteristics and individual QoE (PQ and

[1] available at http://ii.tudelft.nl/iqlab/iQOE.html
[2] Available at: 1drv.ms/1M1bnwU

enjoyment) ratings. These two datasets were derived from two independent user studies conducted in [40] and [29], respectively.

i_QoE

As shown in table 3.1, the i_QoE dataset uses 6 high resolution (i.e., 1280*720) videos as sources, covering three genres, i.e., sports, comedy and education. All six videos last for about 5 minutes, and are further encoded with H.264/AVC at two bitrate levels, i.e., 600kbps and 2000kbps. The two resulting versions of each video present clear differences in PQ. Fifty-nine participants evaluated the videos, split into two disjoint groups: 30 participants viewed the test sequences themselves, the remaining 29 viewed the sequences in the company of two friends (so, in groups of three; interaction among them was allowed). After a short training session in which users became acquainted to the type of artifacts they would see during the experiment, each participant viewed only one version (either 600 kbps or 2000 kbps, in random and counterbalanced order) of all six videos. For each video, participants scored their level of enjoyment through 4 questions (each to be answered on a 7-point Likert scale) [40]. They were also asked to score their perceptual quality on a 5-point ACR scale, according to [16]. In total, 354 ratings (59x6 videos) on these two QoE aspects were collected.

Table 3.1: Internal consistency measured by Cronbach's alpha among items in each aspect of the questionnaire

Video material		
	i_QoE [40]	CP-QAE-I [29]
Num. source sequences	6	12
video format	h.264/AVC	h.264/AVC
Bitrate (kbps)	600, 2000	384, 768
Resolution	1280*720	1280*720, 854*480
Framerate (fps)	30	5, 15,25
User Characteristics		
	i_QoE [40]	CP-QAE-I [29]
Participants	59	114
Gender	27 F, 32 M*	33 F, 81 M
Interest	Yes	No
Nationality	Yes	Yes
Personality	Yes	Yes
Immersive tendency	Yes	No
Cultural Background	No	Yes

*F stands for Female, M stands for Male.

Before starting the experiment, participants were asked to fill in a questionnaire investigating personal information, such as the level of *interest* they had (a priori) in the video genres they were about to see, their *immersive tendency*, *nationality*, *gender* as well

as their *personality*. *Interest* in the video genre was quantified on a 7-point Likert scale (with 7 being the highest possible level of interest in the video content). The *Immersive Tendency* was quantified via the questionnaire [38], returning an IT score on a scale ranging from 18 to 126 (the higher the score, the higher the immersive tendency). *Personality* traits were quantified via the 10-items TIPI questionnaire (also known as BFI-10), where each item is assessed on a 7-point Likert scale, and each trait is measured by a pair of opposite items [9]. For example, the trait "*Agreeableness*" was quantified by adding up the self-assessment of the user on the positive item *Sympathetic & Warm* and the inverse of the self-assessment on the negative item *Critical & Quarrelsome*.

CP-QAE-I

The CP-QAE-I dataset uses 12 short videos (i.e., around 2 minutes long) selected to cover different affective categories (i.e., sadness, anger and disgust). 12 versions of each video are included in the dataset, resulting from a combination of three system factors: bitrate, resolution and frame rate (see table 3.1 for the specific settings). 114 participants from three universities were involved in this study. They were first asked to report personal information, being *age*, *gender*, *cultural background* and *personality*. *Cultural background* was measured via the VSM-2013 questionnaire [14] on 7-point Likert scale. *Personality* was quantified through the BFI-10 [9].

Participants were encouraged (but not forced) to evaluate one of the 12 versions of each short video (i.e., they evaluated only one of the 12 combinations of system factors for specific video content). Participants were asked to report their level of enjoyment and PQ immediately after watching each video. Both QoE aspects were rated on a five-point scale. In total, 84 participants managed to finish all 12 videos. The minimum number of clips that one participant evaluated was 3. Eventually, 1232 individual ratings were recorded for both enjoyment and PQ.

COMMON AND EXCLUSIVE CHARACTERISTICS

For both datasets, *PC*s and *CC*s (48 in total) were extracted from all test videos in an identical way. With respect to *UC*, the information in the two datasets is only partially overlapping. Both datasets report information on the user's *gender* and *personality*, based on the big five model (one score per trait, normalized between 0 and 1). Those six characteristics are used as *UC*s for both datasets, leading to a total of 54 common characteristics for the two datasets.

With respect to the exclusive characteristics, as shown in Table 3.1, the i_QoE dataset presents the individual ratings on *interest* (one value) as well as on *immersive tendency* (one value). CP-QAE-I reports information on *cultural background* (one value per each of six traits) which is not given in i_QoE. Those *UC*s are considered as exclusive characteristics for each dataset in our study, with values being normalized between 0 and 1. In addition, as shown in Table 3.1, both datasets have information about user's *nationality*. However, because the majority of participants in i_QoE was from either the Netherlands or China, the nationality information for i_QoE is reported as a binary value (i.e., either Westerner or Asian [40]). The CP-QAE-I, on the other hand, assumes five categorical values for nationality, being British, Chinese, Singaporean, Indian and the rest of the world. Due to the mismatch in encoding of the nationality variable for the two datasets, the latter is considered as an exclusive indicator for the two datasets. Finally, both datasets

measure *personality* by using the same questionnaire [9]. As mentioned earlier, each personality trait (five in total) was measured by adding up the self-reported scores on two opposite items, a positive and a negative one. i_QoE reports the values of each item as well as the aggregated trait scores. CP-QAE-I dataset reports only the latter. We include the scores of the 10 items (2x5 traits) as exclusive *UC* for i_QoE. Hence, there are 7 exclusive characteristics for CP-QAE-I and 13 for i_QoE.

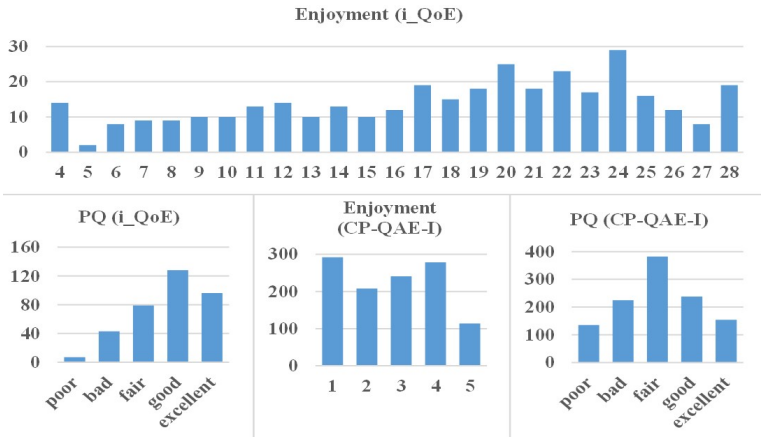

Figure 3.2: Histograms showing the distribution of individual ratings for the two datasets considered in this study. Here, the X-axis represents points on the rating scale, whereas the Y-axis represents the number of instances for each score

3.4.2. Prediction module implementation

Both datasets report discrete (ordinal) ratings of enjoyment and PQ. The rating distributions are shown in Figure 3.2. For i_QoE, enjoyment ratings range between 4 and 28, while for CP-QAE-I, they range between 1 and 5. For both datasets, perceptual quality ratings are expressed on an ACR scale, ranging between poor and excellent.

According to [16], a score of 'fair' on an ACR scale (middle point in the 5-point scale) or less indicates "Unacceptable Quality" (UQ), whereas a score of 'good' or 'excellent' (4 and 5) indicates "Acceptable Quality" (AQ). Although this distinction was originally conceived for acceptability of perceptual quality [16], we extend it to the enjoyment ratings as well. For CP-QAE-I, videos rated between 1 and 3 on the enjoyment Likert scale are considered as "Not Enjoyed" (NE) by the user, whereas scores of 4 and 5 indicate that the video was "Enjoyed" (E) by the user. In line with these settings, the threshold for identifying enjoyed video experiences for i_QoE is set at 17, i.e., experiences scored 17 or less are considered not enjoyed, whereas experiences with score above 17 are considered as enjoyed.

As shown in Table 3.2, the i_QoE has 154 and 200 instances in which a user found the video to be not enjoyable and enjoyable, respectively. In 130 of the 354 instances the user deemed the perceptual quality of the video unacceptable. The CP-QAE-I dataset included 771 instances of not enjoyed experiences, and 461 of enjoyed experiences. Similarly, in 788 cases videos were deemed by a user to be of unacceptable quality and in

Table 3.2: Overview of class distribution for the two datasets. In the parentheses we indicate the percentage of instances of the majority class across the dataset

	Enjoyment	PQ
i_QoE	154NE, 200E(56.4%)	130UQ, 224AQ (63.3%)
CP-QAE-I	771NE, 461E(62.6%)	788UQ, 444AQ (64%)

444 cases the quality was considered acceptable. Each dataset contains a percentage of instances (i.e., a specific video evaluated by a specific user) for which the user enjoyed the video, and a complementary percentage of instances for which the user did not (the same goes for perceptual quality).

The prediction module is implemented by using classification algorithms, targeting the prediction of acceptable versus unacceptable quality (AQ vs UQ) for PQ and enjoyed vs not enjoyed (E or NE) for enjoyment. Due to the exploratory nature of this study, it is still early for deploying complex machine learning algorithms to learn the relationships between *PCs*, *CCs* and *UC*s and QoE aspects to optimize the prediction. This is justified only after we have learned more about which characteristics influence what QoE aspect and to which extent. We therefore use a modeling tool that is easily interpretable (i.e, with a limited number of parameters to fit), and specifically, a Linear Discriminant Classifier (LDC). LDC can reduce the dimensionality of the input while preserving as much of the class discriminatory information as possible, and has been used to predict QoE in [1]. In addition to the LDC, we use a Support Vector Machine (SVM) to verify the added value of non-linear modeling in the prediction [15]. The advantage of SVM is that, when using a linear kernel, every input indicator gets a weigh, which indicates its importance in the classification process, hence allowing intelligibility of the trained model.

3.4.3. EVALUATION PROCEDURE

The LDC and SVM are trained independently on the two datasets for each QoE aspect. That is, each classifier is trained to predict either PQ or enjoyment. An R-fold cross-validation is performed in order to estimate the predictor performance in a robust way, especially considering the relatively small size of the datasets. Thus, data is split in R folds depending on the size of dataset. Then, R runs are performed where samples from one fold are used as test data whereas samples from the remaining folds are used as training data. Each run returns a misclassification rate (*MisRate*) on the test data. The final accuracy of the model is then defined as:

$$Accuracy = 1 - \sum_{r=1}^{R} MisRate_r / R \tag{3.7}$$

The prior probability of correctly predicting QoE (or more specifically, of correctly predicting whether a user described by *UC* would experience a video described by *CC* and *PC* as of being enjoyable, or having acceptable PQ), is defined as the percentage of the accumulation of the majority class in the total number of the ratings. For example, based on the numbers listed in Table 3.2, the prior probability for a user to find a video enjoyable in i_QoE is calculated as:

$$Enjoyment_{i\_QoE} = \frac{200E}{154UE + 200E} \times 100 = 56.4\% \tag{3.8}$$

The prior probabilities per QoE aspect and dataset are indicated in Table 3.2. Since these percentages are unbalanced, we use the prior probability to the majority class as baseline, rather than even chance (50% accuracy). In addition, we compute the Matthews Correlation Coefficient (MCC) to indicate the performance of our model [25]. The MCC is a value between -1 and 1 (i.e., -1 means total inverse prediction, 0 means no better than prior probability, 1 means perfect prediction) and is considered as a reliable measure of assessing the quality of binary classification [25].

3.5. Results

Our evaluation is carried out through two different experiments. The first experiment analyzes each dataset separately (i.e., either i_QoE or CP-QAE-I) towards:

1) evaluating the performance of the proposed model when all the available characteristics are used (i.e., all common characteristics and the exclusive ones) and

2) understanding which (types of) characteristics are most informative for the prediction of PQ and enjoyment.

In the second experiment, the two datasets are merged. The aim of this second experiment is to check the generalization potential of our approach, that is, what performance can be achieved across datasets.

3.5.1. Experiment 1: Model performance

The first experiment consists of three parts. Part I evaluates the overall performance when using all characteristics, Part II evaluates the performance for a specific type of characteristics (*PC*, *CC*, *UC*) and Part III investigates the key influencing characteristics in predicting PQ and/or enjoyment.

Part I: Overall performance

First, the LDC and the SVM were trained separately for the prediction of PQ and enjoyment, based on all available characteristics for each dataset. A 10-fold cross validation was performed for each model and dataset. For the SVM-based model, a Radial Basis Function (RBF) kernel was chosen.

The resulting accuracy for i_QoE and CP-QAE-I is reported, under the column labeled "All", in Tables 3.3 and 3.4 respectively, whereas the MCC values and the confusion matrices are reported in Table 3.5. Note that each fold in the cross-validation returned a partial confusion matrix, and that the final confusion matrices were composed of the resulting ten partial matrices. In general, SVM gives better overall accuracy than LDC, as also confirmed by the MCC values. With regard to i_QoE, better accuracy is achieved in PQ prediction (around 17% above the baseline) than in enjoyment prediction (around 13% above the baseline). For CP-QAE-I, on the other hand, the models perform better in predicting enjoyment (around 10% above the baseline) than for PQ (around 5% above the baseline). In the latter case, MCC indicates a relatively weak prediction power of the

model, highlighting how PQ prediction may be more difficult when multiple system factors impair a video, as it is the case for CP-QAE-I (whereas only bitrate is manipulated in i-QoE).

Table 3.3: i_QoE: The performance of LDC and SVM on Enjoyment and Perceptual Quality, based on all characteristics, the three characteristic categories and selected indicators

	Enjoyment (Baseline 56.4%)				
Predictor	ALL	CC	PC	UC	Selected
LDC	67.8%	63%	59.6%	66.1%	69.5%
SVM	69.5%	63%	60.2%	65.3%	69.5%

	Perceptual Quality (Baseline 63.3%)				
Predictor	ALL	CC	PC	UC	Selected
LDC	78.2%	64.9%	80.2%	60.1%	78.5%
SVM	80.2%	64.9%	80.2%	63.3%	80.2%

Table 3.4: CP-QAE-I:The performance of LDC and SVM on Enjoyment and Perceptual Quality, based on all characteristics, the three characteristic categories and selected indicators

	Enjoyment (Baseline 62.6%)				
Predictor	ALL	CC	PC	UC	Selected
LDC	71.3%	70.5%	65.4%	61.9%	63.8%
SVM	73.1%	70.4%	68.7%	65.8%	68.5%

	Perceptual Quality (Baseline 64%)				
Predictor	ALL	CC	PC	UC	Selected
LDC	65.5%	64.4%	66.1%	64%	64.8%
SVM	68.2%	65.4%	67.2%	69.6%	66.2%

PART II: PERFORMANCE PER CHARACTERISTIC TYPE

To investigate whether a specific type of characteristics is informative for the prediction of either enjoyment or PQ, we trained the classifiers by feeding them only one type of characteristics at the time (i.e., *UC*, *CC* or *PC*). The rest of the setup was kept the same as in Part I.

The results are reported in Tables 3.3 and 3.4 under the column labeled *UC*, *CC* and *PC*. For i_QoE, *UC*s perform best in predicting enjoyment (achieving an accuracy around 66%). In contrast, *PC*s perform best in predicting PQ, achieving an accuracy of 80.2% regardless of the classifier used. For enjoyment prediction in CP-QAE-I, only using *CC*s achieved better performance than only using *UC*s or *PC*s, leading to an accuracy of 70.5%. For PQ prediction in CP-QAE-I, only considering *PC*s gave slightly better accuracy compared to only using any of the other two types of characteristics. Nevertheless, in general we can observe that using a single type of characteristics is suboptimal with respect to using all three types of information together.

Table 3.5: The confusion matrices and MCC of LDC and SVM for Enjoyment and Perceptual Quality prediction based on all characteristics. The dataset the results refer to is indicated in parenthesis: i_QoE (iQ) or CP-QAE-I (CP)

	Enjoyment							
	LDC (iQ) MCC:0.35		SVM (iQ) MCC:0.39		LDC (CP) MCC:0.37		SVM (CP) MCC:0.42	
Pred./ Truth	NE	E	NE	E	NE	E	NE	E
NE	103	51	110	44	633	138	626	145
E	63	137	64	136	216	245	186	275

	Perceptual Quality							
	LDC (iQ) MCC:0.52		SVM (iQ) MCC:0.57		LDC (CP) MCC:0.19		SVM (CP) MCC:0.26	
Pred./ Truth	NE	E	NE	E	NE	E	NE	E
NE	84	46	89	41	672	116	691	97
E	31	193	29	195	309	135	295	149

Key influencing characteristics

In this section, we check to which extent specific subsets of characteristics, possibly mixtures from different types, are suitable for individual QoE prediction. To this end, we performed feature selection to identify the optimal set of key characteristics. Here, data from each dataset was randomly split into two equally sized sets. We used the first set for feature selection, and the second for model training and testing, based on the selected characteristics.

We used Sequential Forward Feature Selection (SFS) for the LDC. Characteristics were selected starting from an empty pool, to which they were sequentially added until there was no improvement in reducing the number of misclassified observations. For SVM, we exploited its intrinsic capability to identify key characteristics in the prediction when using a linear kernel. All input characteristics were assigned a relevance weight [15], based on which only the top 25 were considered to be key characteristics.

In order to evaluate the prediction accuracy on the second half of the dataset, the data was first randomized and then the QoE models were trained (and tested) by using only the key characteristics selected for each classifier. A 5-fold cross-validation was performed due to the smaller data size as compared to Part I. The rest of the setup was kept identical.

Table 3.6: Key influencing characteristics for LDC on i_QoE. Here, the number between the barckets indicates the order of feature selction process, e.g., 1 indicates the first characteristic that been selected.

QoE aspect	CC	PC	UC
Enjoyment	-	PC_{16}(2)	*Interest*(1)
		PC_{12}(3)	*Conscientiousness*(4)
Perceptual Quality	-	PC_{28}(1)	*Gender*(3)
		PC_{41}(2)	

Table 3.7: Key influencing characteristics for LDC on CP-QAE-I. Here, the number between the barckets indicates the order of feature selction process, e.g., 1 indicates the first characteristic that been selected.

QoE aspect	CC	PC	UC
Enjoyment	*Motion Activity*(1)	PC_{41}(2) PC_{34}(4)	*Uncertainty*(3)
Perceptual Quality	-	PC_1(1) PC_{22}(6) PC_{41}(4) PC_5(5)	*Indulgence*(2) *Agreeableness*(3) *Neuroticism* (7)

To verify whether the feature selection influenced accuracy, the procedure above was performed 100 times, and the same was repeated using all characteristics as input. Subsequently, we performed an independent sample t-test (over the 100 repetitions, so with df = 99) per QoE aspect and model in order to compare prediction accuracy between using the key or all characteristics. None of these four tests resulted in a p-value < 0.05, illustrating no significant loss in prediction accuracy by limiting the set of characteristics to the most important ones. Tables 3.6 and 3.7 report the key characteristics for the prediction of enjoyment and PQ, respectively, selected for the LDC. The corresponding results for SVM are presented in Figure 3.3.

With regard to enjoyment prediction for i_QoE, *Interest* was selected as the most relevant characteristic for both SVM and LDC, supporting the finding of previous studies [40]. Thus, collecting information on user's personal preferences on video content and genres (e.g., possibly via social media by tracking user's watching history, like the most watched genres) may be a key requirement when designing systems able to predict user enjoyment. With respect to *personality*, *conscientiousness* and its sub-characteristic *dependable & self-disciplined* were found to be the key influencing indicators by both LDC and SVM in enjoyment prediction of i_QoE. This resonates with findings in psychological literature that *conscientious* individuals are more likely to have enjoyable experience [35]. Two perceptual characteristics (i.e., PC_{12} and PC_{16}) were selected by LDC as well. In the case of SVM, as shown in Figure 3.3.a, the top 5 key influencing characteristics of SVM were *UC*s and one *CC* (i.e., *Hue Ratio*).

For enjoyment prediction in CP-QAE-I, *Motion Activity* was found to be the most relevant characteristic by LDC. Together with two *PC*s (PC_{34} and PC_{41}), one characteristic related to *cultural background* (i.e., *uncertainty*) was also considered to be among the key ones for LDC. As shown in Figure 3.3.c, two content characteristic (i.e., *Shot cuts* and *Hue ratio*) were among the top 25 key influencing characteristics of enjoyment prediction selected by SVM. The dominance of *CC* with respect to *UC* in enjoyment prediction for CP-QAE-I (which was already suggested by the experiment reported in Section 3.5.1 Part II) may be due to the fact that videos in this dataset were purposely selected to vary in terms of their affective charge. This may explain more variance in the data than individual user characteristics.

With regard to PQ prediction in i_QoE, two *PCs* were selected by LDC, PC_{28} and PC_{41}. These two characteristics describe both spatial (i.e., PC_{28}) and temporal (i.e., PC_{41}) aspects of distortions in the videos. Moreover, *gender* was also identified as a key influencing characteristic in PQ prediction by LDC. For SVM, instead, as shown in Figure 3.3.b,

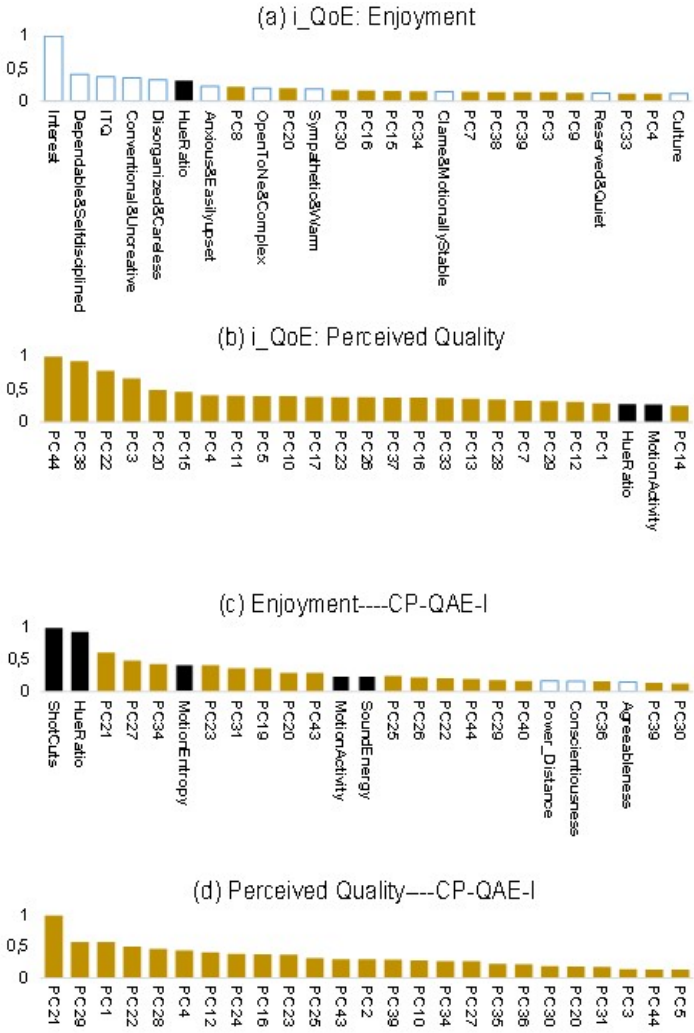

Figure 3.3: Relevance (y-axis) of the key influencing indicators per QoE aspect and dataset, by using SVM. Yellow bars indicate perceptual characteristics, black bars indicate content characteristics, and white bars concern user characteristics

two *CCs* (i.e., *Hue ratio* and *Motion Activity*) were found as the key influencing characteristics. The rest of the key influencing characteristics relate to *PCs*. Finally, for PQ prediction of CP-QAE-I, SVM selected only *PCs* as top 25 key influencing characteristics. One characteristic related to *cultural background* (i.e., *indulgence*) and two *personality* traits (i.e, *agreeableness* and *neuroticism*) were considered relevant by LDC.

Finally, the resulting accuracy for i_QoE and CP-QAE-I based on the selected characteristics is reported in Table 3.3 and 3.4, under the column labeled "Selected". No significant difference was found in the performance accuracy on the test fold of both i_QoE and CP-QAE-I, suggesting that the performance of using the key influencing characteristics was comparable to the one when using all characteristics, despite the reduced amount of input information and with the advantage of working with substantially reduced complexity of the QoE assessment process.

3.5.2. EXPERIMENT 2: GENERALIZATION

In the second experiment, only the common characteristics available for both datasets were used as input for training the predictors. In order to make the characteristics' values of the two datasets compatible with each other, the set of joint values of each characteristic from both datasets was normalized in [0,1].

First, we attempted a cross-dataset evaluation, using the data of one dataset (e.g., i_QoE) for training the models, and those of the other dataset (e.g., CP-QAE-I) for testing. In these experiments, we only used SVM (with RBF kernel), as it was giving the best performance on both datasets and for both enjoyment and PQ. The resulting accuracy, MCC values and confusion matrix are shown in Table 3.8. In general, the accuracy was found to be only around (or even lower than) the baseline, with MCC values close to 0. This might be due to the fact that each dataset uses different test videos with different media configuration (e.g., different resolution, bitrate) and content, resulting in different ranges of *PC* and *CC*. For example, the high bitrate of test videos in i_QoE (i.e., 2000kbps) is much higher than that of (test) videos in CP-QAE-I (i.e., 768kbps). In this way, videos (and users) from one dataset seem to sample an area of the video (and user) space different from that covered by the videos (and users) in the other dataset. As a result, a model trained on one dataset may be unable to extrapolate and predict PQ and enjoyment for the data in the second dataset.

In order to compensate for the differences between the datasets, we decided to merge the two datasets into a single one, possibly achieving a better coverage of the video and user space in the training phase. Enjoyment and PQ again were set as two separate targets. A 10-fold cross validation was performed by using again SVMs as predictors. Table 3.9 presents the accuracy of the SVM trained on the merged dataset as well as the corresponding MCC values and confusion matrices. The baseline for enjoyment and Perceptual quality prediction in this experiment was 58.3% and 57.8%, respectively. As the table shows, the best accuracy that SVM achieved was around 10% above the baseline. The MCC values were higher as compared to only considering one dataset as training set, but still lower than the performance achieved for both enjoyment and PQ prediction in Experiment 1 by using all characteristics (except for PQ prediction of CP-QAE-I). This lower performance might be due to the fact that the models here were trained on a smaller number of common indicators (especially *UC* characteristics), missing essential

Table 3.8: The confusion matrices, Accuracy (Acc) and MCC values of SVM for cross-dataset validation: results on the left columns refer to experiments using i_QoE as training and CP-QAE-I as test data; the rightmost columns refers to experiments using CP-QAE-I as training data and i_QoE as test data

	Enjoyment (i_QoE) Acc:61.12%, MCC:-0.04		Enjoyment (CP-QAE-I) Acc:55.37%, MCC:0.03	
Pred./Truth	NE	E	NE	E
NE	743	28	31	123
E	451	10	35	165

	PQ (i_QoE) Acc:61.28%, MCC:0.12		PQ (CP-QAE-I) Acc:59.89%, MCC:0.07	
Pred./Truth	UQ	AQ	UQ	AQ
UQ	600	188	35	95
AQ	289	155	47	177

Table 3.9: The confusion matrices, Accuracy (Acc) and MCC values of SVM for validation on merged datasets

	Enjoyment Acc:67.91%, MCC:0.33				Perceptual Quality Acc:68.91%, MCC:0.35	
Pred./Truth	NE	E		Pred./Truth	UQ	AQ
NE	731	194		UQ	805	113
E	315	346		AQ	380	288

characteristics (such as, e.g., *interest*) and possibly under-fitting the data.

3.6. Discussion and Conclusion

In this section we highlight a number of main conclusions that can be drawn from the results reported in the previous section and related to the research questions posed in Section 3.3.1.

Regarding RQ1, the results show that for accurate prediction of different aspects of individual QoE, combining the information describing different types of characteristics (perceptual, content and user) is more effective than using only one type of characteristics. As a result, we not only reached a promising performance in predicting enjoyment using multi-type characteristics, but we also managed to reach an improvement in PQ prediction compared to the traditional approaches where only PCs are deployed, e.g. in the case of CP-QAE-I.

Enjoyment was shown to be influenced by all three types of characteristics used, indicating that perceptual and affective characteristics of the video content as well as the characteristics of the user watching are all relevant in this respect. More in depth, and also touching upon RQ2, *Interest* and several *personality* traits were selected as key characteristics for the prediction of enjoyment. Additionally, a set of *PC* was selected, suggesting that *PC* also matters in influencing more hedonic aspects of QoE (i.e., enjoyment). However, since no consistent set of *PC*s has been identified across different

datasets, we note that more studies are needed to identify an optimal set of *PCs* for enjoyment prediction in a general case.

With regard to Perceptual quality (PQ) prediction, *PCs*, as expected, are dominant. Our feature selection returned the ones describing both spatial and temporal characteristics of distortions in the videos. In addition, it is interesting to point out that *gender* was identified as a key influencing characteristic in predicting PQ of i_QoE, suggesting that gender differences should be further investigated when it comes to PQ prediction in a general case. Up till now, *gender*, as one core user characteristic, is hardly investigated in the context of QoE. Most existing QoE datasets do not report the gender information of the users, and the ones that have such information, are usually imbalanced, neglecting gender differences in QoE as noted in [27].

In general, the performance of our individual PQ and enjoyment predictors is satisfactory, and maximized when *UC*, *CC* and *PC*, are used. However, room for improvement exists, and some limitations of our setup should be taken into account in future studies. First, our model only considered a limited number of *UC*s, and more could be included in future models. For example, the more dynamic (varying) *UC*s, like skills or affective state, may potentially benefit QoE prediction. Therefore, collecting more, and more diverse *UC* information is crucial for creating a future individual QoE dataset. Expanding the set of *UC*s may be also beneficial for extending the model to predict other QoE aspects, such as endurability or immersiveness, which may be influenced by other user individual traits as well as video characteristics. Secondly, our model should be trained on a larger range of videos (with different content and system configurations) covering various ranges of *PC* and *CC*. Finally, our results show that SVM in general has better performance as compared to LDC. This result may imply that the QoE prediction can be further improved if we implement the prediction module with non-linear models (e.g., random forests or neural networks).

3.7. REFERENCE

[1] Agboma, F. and Liotta, A. 2012. Quality of experience management in mobile content delivery systems. Telecommunication Systems, 49, 1, 85-98.

[2] Atkinson, R. L. 1996. Hilgard's introduction to psychology. Harcourt Brace College Publishers Philadelphia, PA.

[3] Barkowsky, M., Sedano, I., Brunnström, K., Leszczuk, M. and Staelens, N. 2014. Hybrid video quality prediction: reviewing video quality measurement for widening application scope. Multimedia Tools and Applications, 74, 2, 323-343.

[4] Calvo, R. and D'Mello, S. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. Affective Computing, IEEE Transactions on, 1, 1, 18-37.

[5] Chikkerur, S., Sundaram, V., Reisslein, M. and Karam, L. J. 2011. Objective video quality assessment methods: A classification, review, and performance comparison. Broadcasting, IEEE Transactions on, 57, 2, 165-182.

[6] De Wied, M., Zillmann, D. and Ordman, V. 1995. The role of empathic distress in the enjoyment of cinematic tragedy. Poetics, 23, 1, 91-106.

[7] Eliashberg, J. and Sawhney, M. S. 1994. Modeling goes to Hollywood: Predicting individual differences in movie enjoyment. Management Science, 40, 9, 1151-1173.

[8] Gastaldo, P., Zunino, R. and Redi, J. 2013. Supporting visual quality assessment with machine learning. EURASIP Journal on Image and Video Processing, 2013, 1, 1-15.

[9] Gosling, S. D., Rentfrow, P. J. and Swann, W. B. 2003. A very brief measure of the Big-Five personality domains. Journal of Research in personality, 37, 6, 504-528.

[10] Gou, L., Zhou, M. X. and Yang, H. 2014. Knowme and shareme: Understanding automatically discovered personality traits from social media and user sharing preferences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Toronto, 2014). ACM.

[11] Gulliver, S. R. and Ghinea, G. 2006. Defining user perception of distributed multimedia quality. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2, 4, 241-257.

[12] Hanjalic, A. and Xu, L.-Q. 2005. Affective video content representation and modeling. Multimedia, IEEE Transactions on, 7, 1, 143-154.

[13] Hemami, S. S. and Reibman, A. R. 2010. No-reference image and video quality estimation: Applications and human-motivated design. Signal processing: Image communication, 25, 7, 469-481.

[14] Hofstede, G., Hofstede, G. J., Minkov, M. and Vinken, H. 2008. Values survey module 2008 manual. Maastricht: Institute for Research on Intercultural Cooperation.

[15] Hoßfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K. and Tran-Gia, P. 2014. Best practices for QoE crowdtesting: QoE assessment with crowdsourcing. Multimedia, IEEE Transactions on, 16, 2, 541-558.

[16] ITU-R. 2002. BT.500-11, Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, Geneva, Switzerland, 4, 2.

[17] Le Callet, P. and Autrusseau, F. 2005. Subjective quality assessment IRCCyN/IVC database.

[18] Le Callet, P., Möller, S. and Perkis, A. 2012. Qualinet white paper on definitions of quality of experience. European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003).

[19] Mittal, A., Soundararajan, R. and Bovik, A. C. 2013. Making a "completely blind" image quality analyzer. Signal Processing Letters, IEEE, 20, 3, 209-212.

[20] Möller, S. and Raake, A. 2014. Quality of Experience. Springer.

[21] Nisbett, R. E. and Miyamoto, Y. 2005. The influence of culture: holistic versus analytic perception. Trends in cognitive sciences, 9, 10, 467-473.

[22] O'Brien, H. L. and Toms, E. G. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. Journal of the American Society for Information Science and Technology, 59, 6, 938-955.

[23] Pérez, P., Macías, J., Ruiz, J. J. and García, N. 2011. Effect of packet loss in video quality of experience. Bell Labs Technical Journal, 16, 1, 91-104.

[24] Pinson, M. H. and Wolf, S. 2004. A new standardized method for objectively measuring video quality. Broadcasting, IEEE Transactions on, 50, 3, 312-322.

[25] Powers, D. M. 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.

[26] Redi, J. A., Zhu, Y., de Ridder, H. and Heynderickx, I. 2015. How Passive Image Viewers Became Active Multimedia Users. Springer.

[27] Redi, J. A. a. l. C., P. 2015. Revisiting QoE for future media. www.qualinet.eu/images/stories/Qualinet-Newslet-Vol%2004.pdf

[28] Saad, M. A., Bovik, A. C. and Charrier, C. 2014. Blind prediction of natural video quality. Image Processing, IEEE Transactions on, 23, 3, 1352-1365.

[29] Scott, M. J., Guntuku, S. C., Huan, Y., Lin, W. and Ghinea, G. 2015. Modelling human factors in perceptual multimedia quality: On the role of personality and culture. ACM.

[30] Seshadrinathan, K. and Bovik, A. C. 2009. Motion-based perceptual quality assessment of video. In Proceedings of the IS&T/SPIE Electronic Imaging (2009). International Society for Optics and Photonics.

[31] Smits, E. and Hanjalic, A. 2010. A system concept for socially enriched access to soccer video collections. IEEE MultiMedia, 17, 4, 0026-0035.

[32] Streijl, R. C., Winkler, S. and Hands, D. S. 2014. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. Multimedia Systems, 1-15.

[33] Tamborini, R., Bowman, N. D., Eden, A., Grizzard, M. and Organ, A. 2010. Defining media enjoyment as the satisfaction of intrinsic needs. Journal of communication, 60, 4, 758-777.

[34] Wang, H. L. and Cheong, L.-F. 2006. Affective understanding in film. Circuits and Systems for Video Technology, IEEE Transactions on, 16, 6, 689-704.

[35] Wang, W. 2010. How Personality Affects Continuance Intention: An Empirical Investigation of Instant Messaging. In Proceedings of the PACIS (Taipei, 2010).

[36] Wang, Y.-S., Lin, H.-H. and Liao, Y.-W. 2010. Investigating the individual difference antecedents of perceived enjoyment in the acceptance of blogging. World Academy of Science, Engineering and Technology, 67, 1, 1014-1023.

[37] Winkler, S. and Mohandas, P. 2008. The evolution of video quality measurement: from PSNR to hybrid metrics. Broadcasting, IEEE Transactions on, 54, 3, 660-668.

[38] Witmer, B. G. and Singer, M. J. 1998. Measuring presence in virtual environments: A presence questionnaire. Presence: Teleoperators and virtual environments, 7, 3, 225-240.

[39] Zeng, Z., Pantic, M., Roisman, G. and Huang, T. S. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31, 1, 39-58.

[40] Zhu, Y., Heynderickx, I. and Redi, J. A. 2015. Understanding the role of social context and user factors in video Quality of Experience. Computers in Human Behavior, 49, 412-426.

[41] Zillmann, D. and Bryant, J. 2002. Entertainment as media effect. McQuail's reader in mass communication theory, 406.

4

Measuring Individual Video QoE Using Facebook: A Novel Experimental Platform for Future Directions

This chapter is based on Y. Zhu, S. Guntuku, W. Lin, G. Ghinea, J.A. Redi, *Measuring Individual Video QoE: A Survey, and Proposal for Future Directions Using Social Media* . ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 2018, 14.2s:30.

ABSTRACT

The next generation of multimedia services have to be optimized in a personalized way, taking user factors into account for the evaluation of individual experience. Previous works has investigated the influence of user factors mostly in a controlled laboratory environment which often includes a limited number of users and fails to reflect a real-life environment. Social media, especially Facebook, provide an interesting alternative for Internet-based subjective evaluation. In this chapter, we develop (and open-source) a Facebook application, named YouQ, as an experimental platform for studying individual experience for videos. Our results show that subjective experiments based on YouQ can produce results as reliable as in a controlled laboratory experiment. Additionally, YouQ has the ability to collect user information automatically from Facebook, which can be used for modeling individual QoE.

4.1. INTRODUCTION

INTERNET-BASED multimedia fruition has exploded in recent years. According to Cisco's forecasts, video delivery will account for 80% of the overall Internet consumer traffic by 2019 [1]. Newer and more immersive formats are becoming more common; high-frame-rate, panoramic, HDR video is on the verge of becoming regular (live) streamed content. Content delivery systems now face the challenge of meeting users' expectations in terms of quality while facing the current infrastructural limitations. User tolerance for impairments and buffering events in video playback is decreasing exponentially, and influences users' choice in adopting one service or content provider over the other [2]. On the other hand, resource scarcity (e.g., bandwidth) makes flawless delivery more challenging, making it necessary for a new generation of smart resource optimizing and delivery mechanisms. To do that, it is essential to devise accurate tools to measure the extent to which the user deems the multimedia experience to be of a high quality.

The concept of Quality of Multimedia Experience (QoE) has been shaped to address this necessity. According to the Qualinet White Paper [3], "*Quality of Experience is the degree of delight or annoyance of the user of an application or service. It results from the fulfilment 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*". As such, QoE is determined by the characteristics of the multimedia system that delivers the experience, as well as the context in which the experience is delivered. However, multimedia experiences are also, and more importantly, individual and their appreciation and endurability differ from user to user [4, 5].

Traditionally, multimedia delivery has been optimized for the 'average user', and QoE measured through so-called Mean Opinion Scores [6]. Multimedia quality metrics, aimed at measuring QoE based on the analysis of either the encoded or decoded media stream [7], have been designed to predict the media quality as perceived by such an average user. Individual preferences have mostly been neglected, due to the inherent difficulty in dealing with quantifying and measuring individual characteristics.

However, in reality, there is no such thing as an 'average user'. We see this in (computational) advertising, for example: the same product is advertised differently on different markets according to their demographics; products are branded differently to take into

account local sensibilities or preferences. In the domain of (multimedia) recommendation [8, 81], individual preferences and diversity [9] are taken into account. Personalized solutions for education, conservation, UX, (e-) health, to name but a few domains, are becoming the rule, rather than the exception [10, 82].

Similarly, perceived QoE varies from user to user. Recently, several studies have examined the influence of user factors on individual QoE: users with different age, gender, personal interests, and personality experience the same multimedia differently [11, 12, 74, 75, 76, 77]. Research has also shown that by incorporating user information into quality metrics, reliable estimations of individual QoE can be obtained [4], and user-based (over and above content-based) delivery optimization is within reach.

QoE, by definition, is supposed to be subjective and individual. However, we use the term "individual QoE" since the majority of the literature on QoE has not treated it as such. Whilst the literature conveniently averages QoE (just like with the average family having 2.4 children), in reality averaging ignores the fact that QoE is unique to each individual and to each individual's experience and that it could depend (or not, as the context may be) on an individual's cognitive style, personality, learning style, affective state, psychophysical acuity, as well as on any of a number of environmental factors such as lighting, ambient noise, access device, not to mention task being undertaken. The challenge is that the set of individual factors upon which an individual's QoE depends is not fixed; rather this (sub)set varies from one context to another, and it is this what justifies even more emphatically the individuality and uniqueness of a user's experience – hence the term "individual QoE."

Despite encouraging initial results, research on individual QoE is still in its infancy. The number of and the extent to which user factors may influence QoE is enormous, and a collaborative, community-wide effort is needed to test several user factors systematically. In fact, in order to get reliable results, researchers usually focus on one (or few) specific user factors (e.g., interest [11], personality [11, 12, 78]), following standard protocols and methods (i.e., ITU-R BT500 [6]) in a controlled laboratory environment. However, there are several other factors beyond the laboratory environment that may potentially influence multimedia experiences. Additionally, when it comes to individual QoE evaluation, the studies in the laboratory often include limited number of users and fail to reflect real-life consumption scenarios [13].

Lately the scientific community has devoted increasing attention to the usage of web and Internet platforms to perform user testing [14–17]. Internet-based subjective experimentation refers to all types of experiments conducted through the Internet (e.g., e-mail, websites, crowdsourcing platforms, or social media) [15]. It allows reaching users all over the world (with consequent diversity in demographics and cultural backgrounds) and in ecologically valid consumption conditions. For example, millions of users answered a personality questionnaire, which was traditionally used in the laboratory [15].

The QoE community has followed this trend, focusing especially on crowdsourcing [18] as a method to collect subjective QoE evaluations in a fast and cost-effective way. By posting short QoE evaluation tasks on crowdsourcing platforms (e.g., Mechanical Turk, Microworkers), researchers can reach out to a vast number of users across the world, who complete evaluations in a resource effective manner (both in terms of time and money). A number of studies have shown that crowdsourced subjective testing can produce QoE

assessments that are as reliable as those that would be obtained in a controlled laboratory setting [13, 19, 20]. However, they need to be short to let the user stay focused [14]. As a result, there is a limit to the number of factors which can be explored in each experiment.

Social media, especially Facebook, could serve as an interesting platform for Internet-based subjective experimentation. With over 1.65 billion monthly active users [17], Facebook has become a major platform for people to share their experiences and preferences (e.g., by expressing likes/dislikes or providing status updates) [21]. Facebook can be seen as a huge real-life, continuously growing database of human behavior. Evidence has shown that data collected from Facebook (e.g., status updates) reflects the user's actual ideas and feelings outside the social network [22, 23], and individual traits such as, e.g., personality and health, can be inferred through automatic analysis of the text of comments and posts [24, 76, 79, 80]. Different from other social media platforms, Facebook allows researchers to access user information by developing a Facebook application [17]. Through careful design, by embedding QoE subjective tests into Facebook applications, it is possible to reach potentially millions of users and collect a substantial dataset for research [17], where personal preferences and traits can also be taken into account and studied in relation to individual QoE appreciation. However, questions arise again with respect to the reliability of the data collected through such an application, since users of the application are often from an uncontrolled environment.

In this chapter, we also investigate the potential for Facebook as a reliable platform for studying individual QoE evaluations. Specifically, we are interested in answering the following question: *Can Internet-based subjective experimentation with a Facebook application produce results for individual Quality of Experience evaluation for videos, as reliable as a controlled laboratory experiment?*

To answer this question, we developed YouQ[1], an open-sourced Facebook app designed to perform subjective experiments aimed at measuring individual QoE while watching videos. To test its effectiveness, we performed an experiment investigating the impact of personality and demographics on the perceived QoE of videos with different genres and arousal levels. We performed this experiment both in the wild (i.e., openly on Facebook), and in a controlled lab setting to measure the difference that looser control and higher ecological validity bring to controlled QoE evaluations. In addition, we compared data from our online experiment with two previous independent studies [11, 12], to verify whether results from our Facebook-based experiment are in line with the literature, and whether it can provide additional insights to the growing field of research addressing user factors in QoE. Further, we also provide a meta-analysis of the state of the art literature studying the influence of user factors on multimedia experience.

4.2. RELATED WORK

4.2.1. FACTORS INFLUENCING MULTIMEDIA QUALITY OF EXPERIENCE

In the 1990s, user satisfaction with multimedia systems and services was mainly linked to the concept of Quality of Service (QoS), defined as the *"totality of characteristics of a telecommunication service that bears on its ability to satisfy stated and implied needs*

[1] YouQ: ii.tudelft.nl/qoe/FbApp/

of the user of the service" [25]. In fact, QoS focuses mainly on quantifying system and network performance, based on metrics such as frame-rate, resolution, packet loss ratio, lag, and the like. These metrics, which purely describe the system functioning, were soon shown to correlate poorly with actual user satisfaction [26]. In view of this discrepancy, a more user-centric concept of Quality of Experience (QoE) started to emerge and moved the focus towards quantifying the quality of media as perceived by the user. Research in perceptual quality has seen a substantial amount of work in the last two decades [7], incorporating the influence of core user factors such as perception and attention in QoE optimization. On the other hand, most of these efforts were still devoted to maximize QoE for an "average user", developing a single solution for any user adopting the system or service, irrespective of that users' background.

Lately, the view of what QoE is and how it should be measured has evolved dramatically. The Qualinet White Paper on definitions of Quality of Experience has painted a new picture with respect to what influences QoE, and as such should be taken into account in multimedia delivery optimization [3]. In this new view, the QoE delivered by a system is not only influenced by the properties of the system itself (system factors) but also by the context in which the experience is consumed (contextual factors), and by the current state of the user, i.e., by the individual characteristics of the user.

SYSTEM FACTORS

System factors determine *"the technically produced quality of an application or service"* [3]. They can influence QoE in video delivery services by altering the perceptual quality of the content. For example, visible artifacts such as blockiness, blur, and ringing might be generated by a certain type of compression (e.g., H.264/AVC). These can result in user dissatisfaction. Similarly, QoE can be impacted by network QoS parameters [27], and the media configuration [28]. For example, buffer ratio, i.e., the fraction of time spent in buffering over the total session time, including playing plus buffering, is inversely related to QoE [27]. Similar observations were made for other QoS parameters such as buffering events rate, buffering duration and average bitrate [29]. Besides these, the nature of video content itself can influence users' QoE [30]. For example, different genres (e.g., comedy, action, etc.) have different viewing patterns, resulting in different perceptual quality. For instance, for a given bitrate, genres which contain little movement, e.g., comedy sequences, usually have higher perceptual quality than genres which contain high-pace movement, e.g., action sequences [31].

CONTEXTUAL FACTORS

Contextual factors describe the situation in which the media is consumed by the user. They entail characteristics of the physical environment, the presence (in *situ* or remote) of other users experiencing the same media, or economic conditions regulating the service fruition. Physical surroundings that may influence users' experience are the seating position (i.e., its viewing distance and height), lighting conditions [32] and disturbances which may occur such as incoming phone calls or SMS message alerts [33]. Co-viewing videos with friends may make the entire experience more enjoyable, and therefore of higher quality [11]. Finally, depending on the amount of money paid for accessing the content, users may have different expectations in terms of quality, with lower prices corresponding to higher tolerance to impairments [34].

4.2.2. PRELIMINARY STUDIES ON USER FACTORS AND INDIVIDUAL DIFFERENCES IN QOE

Differential Psychology has been studying individual differences in user factors and their impact on human behavior and performance for over a century. Despite this large body of knowledge, individual differences and user factors have mostly been neglected in multimedia delivery optimization, with the exception of a few studies. Their origin is scattered, coming from the fields of Media Psychology and Human-Computer Interaction, and only lately from the Multimedia community. Although the literature in this area is far too thin to build a complete taxonomy of the effect of individual differences and user factors on Quality of Experience, we can identify four macro-categories within which user factors can be clustered: demographic, physiological, psychological, and socio-cultural factors.

PHYSIOLOGICAL FACTORS

Because of the configuration of the human sensory system, physiological characteristics of the individual users play a crucial role in QoE. As such, they have been thoroughly investigated, and models of visual perception and attention have been incorporated in objective video quality assessment metrics. Nevertheless, these factors have been mostly studied and modeled referring, once again, to the "average user". Fewer studies have looked into how individual physiological characteristics influence QoE. Elements such as visual acuity and losses in contrast sensitivity due to aging have been shown to influence visibility (and annoyance) of visual impairments; yet, they have hardly been included in QoE models [35]. Color blindness also alters a users' perception and, when it comes to immersive, 3D visual representations, stereo-blindness can be a core factor in determining the appreciation of new imaging technologies [36]. Similarly, human hearing characteristics can influence QoE with auditory (or audiovisual) media; in fact, international recommendations for subjective testing conditions provide a set of indications to minimize individual differences in QoE ratings due to physiological diversity, rather than acknowledging them [37].

SOCIO-CULTURAL FACTORS

Both the educational and the socio-cultural background of the user play an important roles in experiencing media. Cultural background has been shown to influence visual perception and attention, for example, between Westerners and Asians in [38]: Americans were reported to have more analytical visual perception (inclined to pay attention to details), while Asians were seen to have a more holistic visual perception (likely to be more sensitive to context). The influence of culture on optical illusion, color perception, visual attention, and brain functioning was documented in [39]. The correlation between culture and cognition was also studied in [40], by analyzing the variation of word associations given by Japanese and American participants. Previous experience with multimedia technology and devices is also critical in determining one's QoE. Users with expertise in photography were shown to be more consistent than naïve users in judging the aesthetic appeal of images; users less acquainted with mobile media consumption were found to be more critical and to have different expectations, in terms of QoE, with respect to other users who were more technologically savvy [5, 13].

DEMOGRAPHICS

Demography (e.g., age, gender or nationality) may influence QoE. For instance, older adults are found to be more critical than younger users suggesting that elderly people usually have higher requirements for QoE [41]. On the contrary, another study showed that younger users tend to rate video quality lower than older users, although the performance is better [42]. Similarly scattered results exist for biological sex [5, 13], and no uniform conclusion has been carried out, to the best of our knowledge, to clarify the role of demographic characteristics in QoE appreciation.

PSYCHOLOGICAL FACTORS

The affective state of the user plays most likely the biggest role in determining one's satisfaction with the delivered experience. The user's mood has been shown to influence quality preferences, and multimedia experiences have been shown to influence mood in turn [13]. Among emotions, the most impactful one for QoE has been shown to be interest [43]. A number of studies indicate that QoE is triggered when something resonates with a user's interest [44] and that personal interest in video content significantly influences a user's QoE judgment [45]. The more a user is interested in the video content, the higher QoE s/he attributed to the same bitrate [46]: football fans have been shown to be able to tolerate incredibly low frame-rates, when watching a match they like. Finally, user performance is proved to be influenced by personality. For example, enthusiastic people are more likely to switch the TV channel or change the volume of the TV on their first attempt compared neurotic people [41]. Researchers found that including user expectations, users' monetary budgets, and quality pricing in modeling perceptual quality evaluation leads to increased accuracy [47–49].

Further, there is additional work that studied the variability of QoE based on other factors such as the device used for streaming the content [50]. While we tried to present most of the factors in this section, this is not an exhaustive list.

4.2.3. MEASURING QOE: FROM THE LABORATORY TO THE REAL-WORLD

Subjective QoE evaluation is based on self-report ratings of one's (perceptual) satisfaction with respect to multimedia contents. Subjective ratings are often collected via psychometric experiments in the lab [51] and aim at measuring the satisfaction of an average user regarding a given video. To ensure reliability and repeatability of the experimental results, the community has developed standardized methodologies and recommendations on experimental settings for subjective testing [6]. These recommendations often include indications on the setup of the physical surroundings, assuming experiments are performed in a controlled laboratory environment for increased control over the execution of the experimental task.

Although ensuring high reliability and repeatability, lab-based experimentation also presents several inconveniences: (1) being highly time-consuming, the number of conditions (in our case, combinations of user and system factors) is necessarily limited, which typically leads to investigating at most one or two user factors at a time; (2) the demography of the users is often not representative of the general population, as most laboratories are embedded in technical faculties of universities, and most participants are students; and (3) the lab environment often fails to simulate the real-life environ-

ment where multimedia is consumed [13], thereby hampering the ecological validity of the results.

A growing number of researchers are now attempting to overcome the above inconveniences by conducting Internet-based subjective experiments (e.g., via websites, social media, and/or crowdsourcing platforms) [20]. Internet-based experimentations enable new possibilities for individual QoE evaluation, by allowing researchers to recruit users from a larger, more diverse group and users to perform the experiment in their real-life conditions, leading to more representative results [13]. In fact, psychologists have used the Internet to study human behavior for a decade [15]. The most straightforward application of online experimenting is implementing traditional lab experiment online [16], i.e., using Internet technology (e.g., E-mail, website) to deliver surveys or questionnaires [15]. Very successful examples are the outofservice.com website and the Project Implicit website[2], which both host standard personality questionnaires [52], and have reached millions of users [15].

Within QoE evaluation research, it is not straightforward to move from laboratory experiments to online experiments by simply generating a web version of the existing lab test [14], because of conceptual, technical, and motivational differences [53]. Conceptual differences are due to the fact that online experiments usually should be much shorter than a laboratory experiment, lasting at most 10 minutes to prevent the participant from losing focus and interest (where lab-based experiments can last more than 1 hour) [13]. The online experiment usually needs to be split into multiple smaller tasks, orchestrated to cover the entire condition set and avoid context effects [12]. Moreover, the online experiment is less supervised as compared to the lab experiment because it is not possible to give direct and instant feedback between users and supervisors during the experiment [13]. This implies that Internet-based experiments need to be carefully designed in order to ensure that each user understands the task without any difficulties (by providing a proper training session) [13]. The technical differences related to the web-based nature of the online experiment is that it is performed in a more real-life environment using a personal device in contrast to the standardized environment and device in the lab-based evaluation [12]. This implies that advanced experimental devices (e.g., eye tracking) are not available in the online environment. In addition, the fact that the stimuli (e.g., test videos) has to be delivered over possibly slow connections to the user has to be considered. Finally, motivational differences have to be taken into account. Whereas in the lab experiment participants are motivated by genuine interest in the research and/or by social pressure, in Internet-based experiments extrinsic motivation is to be generated via either financial incentives (i.e., from commercial online experiment platforms, e.g., Mechanical Turk, or Microworkers) or amusement (i.e., from social media platform, e.g., Facebook) [13].

Although challenging, these differences can be overcome through careful design and by embedding reliability checks within the experiments. Previous works [17, 54] summarized the recommendations and best practices for online QoE assessment of multimedia, providing extensive evidence that by paying attention to the differences listed above in the experimental design, results as reliable as those collected in the lab can be obtained in online settings [14].

[2] http://projectimplicit.net/.

4.2.4. Facebook as a research tool for online experimenting

A number of studies have shown that social media profiles, such as Facebook, convey fairly accurate impressions of the profile owners [21, 55, 79–81]. For example, Facebook profiles are shown to reflect personality of the profile owners in real-life [23], e.g., there is a positive correlation between a user's number of Facebook friends and a user's extraversion [20]. In addition, users typically use Facebook to maintain offline relationships with their friends [21] by sharing photos/videos or posting comments [56]. Videos spread much faster on social media (e.g., Facebook, Twitter) than on a video/image sharing network (e.g., YouTube) [57]. In short, Facebook provides a compelling source of measurable user information [58].

A growing number of researchers, especially in psychology and social science, have begun to collect user information via Facebook for research purposes [17]. User information from Facebook includes self-reported information (e.g., age, gender, current workplace, etc.), digital traces of behavior (e.g., likes, posts, photo uploaded, etc.) and information contributed by others (e.g., photo tagged or comments on a user's wall) [17]. Many studies have tried to estimate individual characteristics (e.g., personality, life satisfaction) based on user information collected from Facebook. For instance, user's personality traits are found to be significantly correlated with semantic features of status updates [59], Facebook profile [23], number of friends [60], number of photos uploaded [61], profile photos [83] and so on. Facebook 'likes' can be used to draw predictive profiles of the Facebook user's race, sexual orientation, religion, IQ, and the like [62]. Users' personal interest can be estimated by analyzing the self-statement in their profiles as well as their previous activities e.g., videos watched or liked [63, 64]. Further, there has been work studying the difference between recruiting users from Facebook, paid Microworkers crowdsourcing platforms, and traditional controlled lab environments [65].

Of course, due to privacy concerns, users have the right to limit access to parts of their Facebook information [17, 21]. Thus, a well-designed Facebook application is needed to engage users and make them willing to share their Facebook information with researchers. MyPersonality, developed by [66], is a successful Facebook application to collect user information through an online personality survey. The application attracted more than 10 million users, and about 30% of its users were willing to share their Facebook information with the authors. The above examples thus demonstrate the potential of Facebook applications as a research tool to run subjective experiments by simultaneously collecting users' information.

4.3. YouQ: THE STRUCTURE AND DESIGN

We developed a Facebook application (which obtained approval from Facebook), YouQ, for the collection of individual video QoE data. YouQ has multiple advantages:

(1) it can reach users with diverse demographics in real-life viewing environments;

(2) it can collect user information (e.g., age, gender, video watch history, uploaded pictures) automatically from Facebook.

In addition, the app allows checking for user reliability at every stage of the experiment. The app is open-sourced for research. Researchers are free to edit the set of stimuli and questionnaires based as need be.

4.3.1. OVERALL DESCRIPTION

YouQ organizes the entire process of online subjective testing in three main phases: (1) *Start Phase*, (2) *Test Phase*, and (3) *Score Phase* as shown in Figure 4.1. At the beginning of the Start phase, the user reaches the app page, and can log in from his/her Facebook account to access the app. New users (i.e., users who have not used the app before) are shown a short introduction and are asked consent to let YouQ retrieve the information s/he has shared on Facebook. Returning users can read this information again anytime by simply clicking the "About" button on the home page shown in Figure 4.2. A questionnaire is then presented to collect personal information: in our case, demographics, personal interest on different video genres, and Big-Five personality. By doing so, we intended to check whether such self-reported information matches the information automatically collected from Facebook. In parallel all the requested user Facebook information is retrieved and stored in a database (see Section 4.3.2 for a detailed explanation).

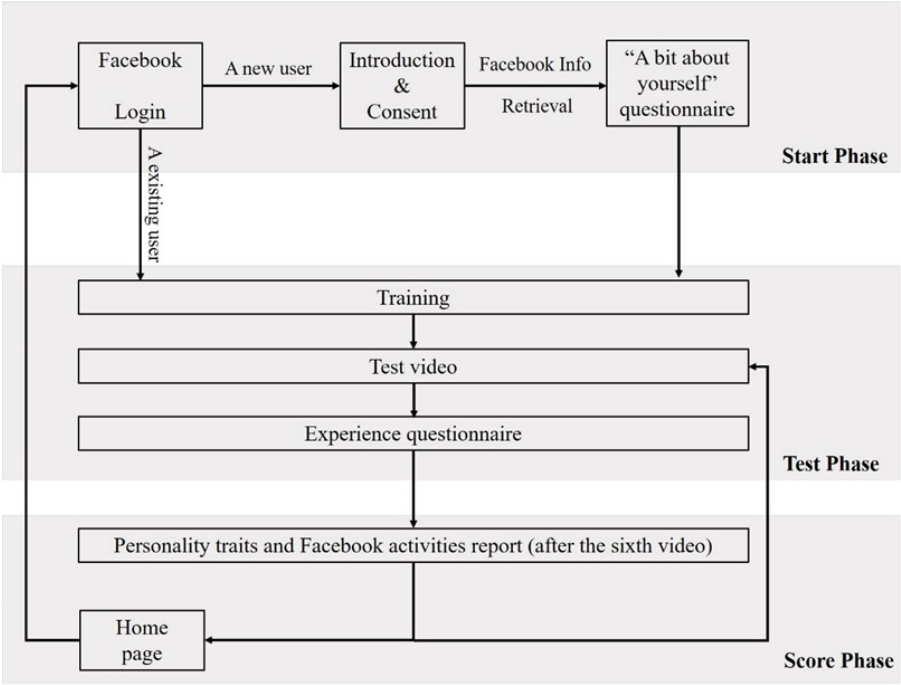

Figure 4.1: The overall design of YouQ

In the *Test phase*, users perform the actual QoE experiment. In the case of YouQ, it consists of video QoE rating. When the user starts this phase from the home page, and s/he is already registered in the database, s/he is asked to perform a short training before the actual test phase takes place. In this training phase, the user has to watch two sample videos at two different distortion levels (to get familiar with the perceptual quality range spanned by the test stimuli) and perform QoE evaluation as it would be done during the test. Then, the actual experiment takes place and the user watches a video with a given bitrate/quality. Next, s/he is asked to rate the viewing experience, in our case through

a questionnaire (explained in detail in Section 4.3.4). In order to ensure the experiment will not be affected by some unwanted impairments (e.g., buffering), it is necessary to pre-cache all data required during the experiment [46]. In other words, a video will not be played if it is not fully downloaded in the experiment. Once the user finished the evaluation, the data, i.e., the name of the video watched and the questionnaire answers are stored and the user is redirected to the next phase.

In the *Score phase*, the user first gets the summary of his/her personality traits and is then asked how s/he wants to proceed. Note that the user is not required to evaluate six (or more) videos. After each evaluation and score phase, the user can choose either to evaluate more videos or to quit the test and application. If the user chooses to score other videos, s/he will be redirected to the test phase, where s/he will be scoring videos directly, without going through the training phase again.

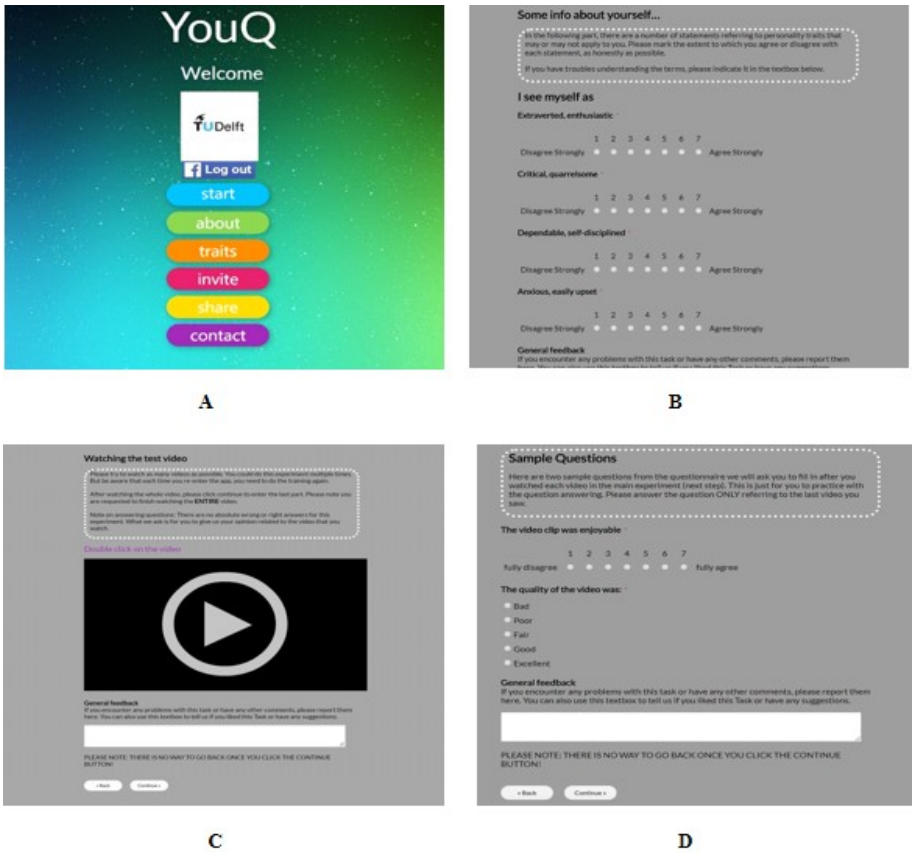

Figure 4.2: Screenshots of YouQ. (A) represents the welcome page of YouQ; (B) represents partially the questionnaire collecting personal information; (C) represents the test page of YouQ; (D) represents a sample questionnaire regarding individual QoE.

4.3.2. USER INFORMATION COLLECTED IN YOUQ

Users volunteer to share their Facebook information as part of the YouQ application (note that the user retains the right to decide what to share and what not to). YouQ asks for eight types of Facebook information. Requested data are the information in one's Facebook profile, including gender, age, as well as the user's list of friends. In addition, YouQ gets access to the list of all pages and Open Graph objects that a user has liked, the photos and videos a user has uploaded and been tagged in, and the posts on a user's timeline, including his/her own and others posts. Note that YouQ can access this information once a user agrees to give YouQ the access. Such personal information is shown to capture many individual differences (i.e., personality [60, 61, 80, 81], demographics [67], health [79], socioeconomic status [68] etc.) and consequently might be predictive of individual QoE as well. For example, the total number of Facebook friends, uploaded photos/videos, likes, posts are proven to be significantly correlated with personality [60, 61], and therefore, are collected by YouQ.

4.3.3. RELIABILITY CONTROL MECHANISMS

We introduce 5 types of reliability checks – before, during and after the experiment as suggested in [54]:

1. When a user logs in with his/her Facebook account and agrees to share profile information, YouQ collects his/her age/gender/ nationality. It then checks whether such information matches the information s/he has given in the initial questionnaire (from the start phase). Those who do not provide consistent information are not considered in the analysis.

2. In the initial questionnaire, YouQ includes one reliability question to check whether the user can understand English. The user will be asked to indicate the resolution of their display. For those who don't know the answer, it is also fine to just type in "DK". Those who do not provide a resolution value or "DK", demonstrating they either don't have sufficient English proficiency or didn't pay attention to the question, are not considered in the analysis.

3. Consuming media with other tasks is an observed behavior in many people [69] e.g., the study has shown that around 75% users are multitasking while watching television [69]. To make sure users from the online experiment pay attention to the video and are able to understand the video, we included a very simple content question (a multiple choice, e.g., a man telling a joke for Figure 4.3, video 1) after each video, checking whether participants paid attention to the video they just watched. Data from participants who do not answer this question correctly is not considered in the analysis.

4. *YouQ* tracks the time a user is staying on each page by collecting the timestamps. If a user stays on a page far more than it normally takes (two times longer/shorter than the reference time), the user's response is discarded in the analysis, as it may indicate that the user has been multi-tasking while evaluating the stimuli [54]. Note that the reference time varies depending on the task and the length of the test video. In this work, the reference time is set based on a pilot experiment which involved five users. The average time spent on the task is defined as the reference time.

5. The questionnaires are answered on a Likert scale built out of radio buttons. All questions must be answered, otherwise the user cannot continue.

4.3.4. Questionnaires used in YouQ

YouQ has been designed to collect user information automatically (as explained in Section 4.3.1) together with subjective QoE ratings. Specifically, YouQ measures the user's viewing experience in terms of perceptual quality and enjoyment, as it was proven in previous research that these two aspects are core to characterizing Quality of Experience [70]. For perceptual quality, YouQ asks the user to indicate the perceived overall quality of a video s/he just watched on an ACR (i.e., Absolute Category Rating) scale [6]. Besides perceptual quality, the measurement of QoE should also be complemented by a measurement of the level of enjoyment of the experience, which reflects how much happiness or fun a user gets from the video [70]. As a result, YouQ asks users to indicate their level of enjoyment for the video they just watched on a 7-point Likert scale. The app is also configurable to add further measurements of the viewing experience evaluation, such as endurability, satisfaction, or involvement [11].

In addition, for this pilot phase of the app, we introduced a short introductory questionnaire (see "start phase" in Section 3.1), where we collected self-reports of user data to cross-check with the Facebook data collected automatically. The introductory questionnaire asks participants to indicate their age, gender, nationality, and levels of interest in different video genres (on a 7-point Likert scale). The "a bit about yourself" personality questionnaire, mainly adapted from [71], is then presented, consisting of 10 items to be scored on a 7-point Likert scale. The aggregated questionnaire results provide a personality profile along the "Big-5" traits [71], including Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

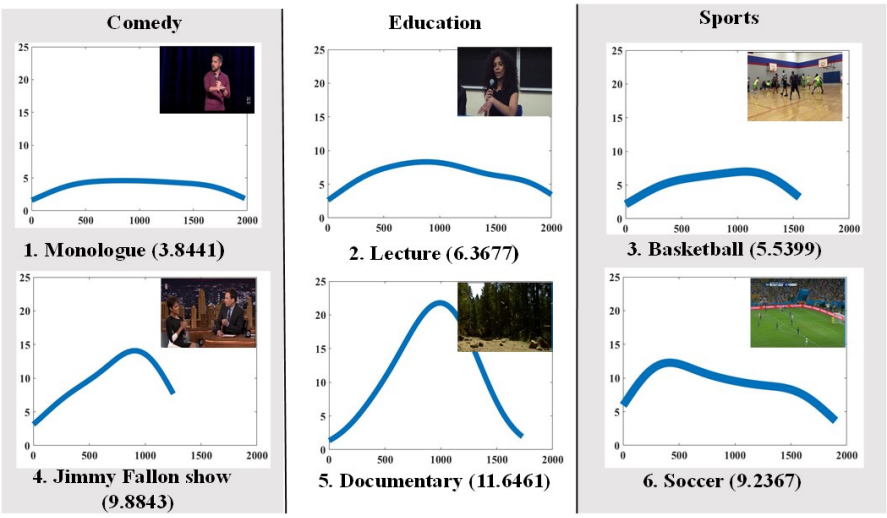

Figure 4.3: The videos with their arousal curve used in this experiment. Here, the Y-axis represents the arousal value (the intensity of emotion the video represents) whereas the X-axis represents the frame number. Video 1, 2, 3 represent videos with a low excitement level whereas video 4, 5, 6 represent videos with a high excitement level. The number between brackets represents the mean value of the arousal curve.

4.4. EXPERIMENTAL SETUP

THE main goal of this chapter is to explore the suitability of Facebook to function as an experimental platform for video individual QoE research. More concretely, the goal is to create a Facebook experimental environment that is able to mimic users' viewing experience in real life, reflecting users' natural behavior. As a first step, we intended to verify whether the app we designed would provide reliable results when launched "in the wild", i.e., openly on Facebook without supervision and controlled settings, as compared to a classic laboratory setting. As a second step, we wanted to check that the results obtained were also in line with the existing literature on individual QoE [11, 12, 84–86]. To this end, we designed an experiment with a setup similar to that used for two recent studies in the lab environment [11, 12], investigating the effect of personality and demographics on perceptual quality and enjoyment of videos of varying bitrate, genre, and affective content [11, 12].

We performed this experiment by using YouQ in two different sessions. In the first one, participants were invited to use the app in a controlled laboratory setting, with an experimenter supporting them throughout the task. In the second session, the app was launched on Facebook and heavily advertised throughout the personal networks of the authors. In the following, we describe the experimental setup for both sessions in more detail.

4.4.1. STIMULI

We used six video contents from three genres (two contents per genre) in this experiment (shown in Figure 4.3), and chose them according to the indications of [11]. All video contents were originally from YouTube and in English. These videos were chosen to match content associated with day-to-day user consumption [73]. In the traditional video quality evaluation, users usually are asked to watch short samples (i.e., around 10 seconds video) with different visual impairments in them. Such short videos usually fail to reach user engagement and attention close to real-life video viewing [72]. Authors of [72] suggested at least 1-minute long videos to be able to engage users. In addition, users from an online experiment can easily leave the experiment if it is too long. Authors of [14] suggested that online experiments should not be longer than 10 minutes to keep users motivated. Therefore, the duration of all test videos was limited to between 1 and 2 minutes to make sure users can engage in the video while being able to finish the whole experiment in 10 minutes.

Videos were varied in terms of their affective content. The affective content of a video (i.e., the intensity of emotion that is expected to arouse in the user while watching that video [66, 87]) is suggested to influence individual viewing experience, and has been previously investigated within individual QoE in [11] and [87]. For this reason, in our experiment, we selected per each genre two content types, at high and low excitement level, respectively. To determine the excitement level of a video, we took the mean value of the arousal curve (as shown in Figure 4.3) throughout the whole video, calculated according to [73]. During our lab experiment, we asked users to rate the affective response for each test video on a 7-point Likert scale (i.e., reflecting how exciting the video that the user just watched was). After collecting the data, a linear mix model was used to validate the manipulation of affective content. Here, the affective ratings were considered as

dependent variable, whereas the excitement level, bitrate, and genre were considered as independent variables. The result shows that the test videos with high excitement level were rated significantly higher than those with low excitement level ($F = 60.284$, $p < .001$).

All videos were encoded in H.264/AVC, 30 frames per second. To represent different video quality conditions, we encoded each video at two bitrate levels, i.e., 2000kbps and 500kbps, according to what was done in [11]. As a result, a total number of 12 videos were used in this study. The audio configuration of the videos was kept constant, in the AAC format with a bit rate of 112kps, to avoid any effect of audio quality on overall QoE. The video resolution was set to 896 × 504 pixels, which is the default size of a YouTube video player, to accommodate for different screen resolutions.

Finally, two 10s video samples were used for training. One sample was encoded at 500kbps (representing lowest quality) while the other one was encoded at 2000kbps (representing highest quality). By doing so, we intended to make the users acquainted with the range of perceptual quality of the videos we had. The media configuration of these samples is the same as for the test videos (30fps, 896 × 504, H.264 AVC). All videos used in this study were encoded and stored on our server.

4.4.2. PROCEDURE

We performed this experiment via YouQ both in the lab and in the wild in order to compare the results.

Lab Experiment. In the lab setting, users sat in front of a 27-inch display. The environmental setup, including the lighting settings, were strictly followed the ITU recommendations [6]. Users were greeted and explained the purpose of the experiment. They were then asked to log into YouQ (displayed in a Google Chrome browser) with their Facebook account, and to give us the permission to retrieve their information from Facebook. They then reached a welcome page with a description of the experiment followed by a short consent form. After giving consent, the user started the experimental session, going through the three phases described in Section 4.3. First, the user filled in the personal information requested by the starting questionnaire described in Section 4.3.4. Then, s/he performed the training session to understand the task and visualize the different quality levels presented in the experiment. The test phase would then start, with the user watching all 12 test videos after the training, presented in a random order to avoid fatigue and memory effects. Note that in the lab experiment all users watched all 12 videos in a within subjects setting, given that there were no constraints on the brevity of the task. The experiment took about 40 minutes to finish, and users got on their respective personality traits after watching the first 5 videos.

In-the-Wild Experiment. For the in-the-wild session of the experiment, YouQ was openly published on Facebook, and users were recruited indirectly through advertisement. Also, in this case, when logging into the app for the first time, the user has to give consent for us to access his/her data and had to fill in the questionnaire on personal data. After that, the test phase (see Section 4.3) started, and users completed the evaluation of at least one video in a given bitrate. The test video was randomly selected from the video pool. After finishing the first test video (and receiving the first bit of information on their personality traits), the users could choose to quit the application or to continue and evaluate another test video, until they saw at most one version of each of the six video

contents. As a result, the within-subject design adopted in the lab was mostly dismissed in this phase, mainly to ensure that the task was kept short and engaging. It took around 10 minutes to finish the questionnaire plus the first test video.

4.5. RESULTS

4.5.1. YOUQ IN THE WILD AND IN THE LAB

Table 4.1 presents an overview of the participation in the two experimental sessions. In the lab experiment, we had 20 users (5 Female and 15 Male), all of them students from Delft University of Technology, aging from 20 to 32 with a mean age 27.8 years. Two hundred and twenty-five (225) users joined the online experiment, but 89 users failed to pass the reliability checks (e.g., inconsistency between self-report information and automatic Facebook information retrieval). As a result, 136 users (66 Female, 70 Male) were recruited in the online experiment, aging from 15 to 70 with a mean age 31, and coming from 17 countries (mainly from the USA, Canada and Netherlands). Each user rated on average 1.55 videos and, on average, each test video received more than 15 ratings. In order to check whether the online experiment could collect results as reliable as the lab experiment, we built a linear mixed model where Bitrate level, Excitement level (i.e., mean value of arousal curve), genre and experiment condition (i.e., in-the-wild or lab) were considered as fixed factors, and user was considered as one random factor. Model parameters were determined with the restricted maximum likelihood method (REML). The perceptual quality ratings (here, PQ indicates the overall perceptual quality of a user with a video with the Absolute Category Rating [6]) and Enjoyment were used as dependent variables. As rationale behind this type of analysis is that, if the in-the-wild experiment was not reliable, we would obtain different effects of the fixed factors on Enjoyment and Perceptual Quality and the factor Experiment condition would have a significant effect on the dependent variables. It should also be noted that, although in previous comparisons between lab- and crowdsourcing-based experiments [19, 54] result reliability was checked by, e.g., computing the correlation between the MOS obtained in the lab and through crowdsourcing, for our purposes (i.e., measuring individual QoE), this had little meaning as we were especially interested in the individual differences. The mixed model we use allowed us instead to still check whether the main effects were consistent throughout experimental conditions (i.e., between lab and in-the-wild), yet accounting for individual differences (modeled through the random factor user).

Table 4.1: Demography of the Users in the Online and Lab Experiment

	online	lab
No. of users	136	20
Age in year	Min: 15, Max: 70, Mean: 31	Min: 20, Max: 32, Mean: 27.8
Gender	66 Female, 70 Male	5 Female, 15 Male
Resident	17 countries	Netherlands

As shown in Table 4.2, the excitement level ($F = 43.31$, $p < .001$) has a significant impact on enjoyment whereas the bitrate level ($F = 9.298$, $p < .001$) has a significant impact on PQ. There is no significant difference between the evaluations of enjoyment and PQ

Table 4.2: The Linear Mixed Models for Enjoyment and PQ (indicating the Overall Perceptual Quality of a video rated with the Absolute Category Rating)

	Enjoyment		PQ	
Source	F	Sig. (p)	F	Sig. (p)
Experiment Condition	.058	.810	2.289	.131
Intercept	2236.96	.000	3634.08	.000
Bitrate	0.082	.774	9.298	.000
Excitement	43.31	.000	.032	.859
Genre	0.099	.906	.225	.799

*Sig. stands for significance

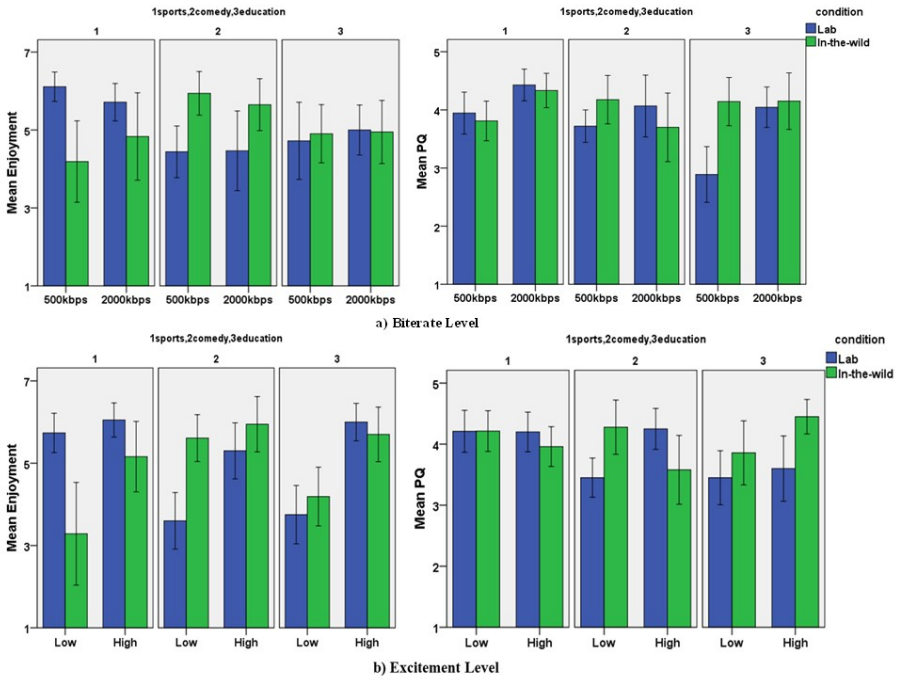

Figure 4.4: Enjoyment and perceptual quality ratings obtained in the lab (blue columns) and in-the-wild (green columns). In Figure 4.4(a), the Y-axis represents the mean value of enjoyment and PQ, whereas the X-axis represents the two bitrate levels. In Figure 4.4(b), the Y-axis represents the mean value of enjoyment and PQ whereas the X-axis represents the two excitement levels.

performed in two experimental conditions (i.e., in the lab and in the wild). For enjoyment, larger confidence intervals were found in the wild condition, for various bitrate levels and excitement levels (as shown in Figure 4.4). This might due to the fact that enjoyment is highly related to the user's interest [11] and personal taste. The users from the online experiment might be more divergent in terms of personal interest as compared to those from the lab experiment, who were mostly from the Computer Science department in a technical university. On the other hand, similar confidence intervals were found in both conditions on perceptual quality suggesting all users had a similar interpretation of video quality regardless the exprimental condition the user was in.

Table 4.3: A Summary of the Experimental Settings of CP-QAE-I and i_QoE

		CP-QAE-I	i_QoE
Users	Number	114	59
	Nationalities	18	13
	Gender	81Male/33Female	27Female/32Male
Outcomes		Enjoyment, Perceptual Quality	Enjoyment, Perceptual Quality
Input Questionnaires		BFI-10, Hofstede Culture Questionnaire, Gender, Nationality	BFI-10, Interest Gender
Stimuli	Videos	12	6
	Number of Videos	144 clips with varying quality conditions	12 clips with two bitrate levels

4.5.2. A SYSTEMATIC COMPARISON BETWEEN TWO RECENT STUDIES AND YOUQ ON INDIVIDUAL QOE

The next step in verifying the suitability for Facebook to function as a platform for the collection of individual QoE ratings is the comparison between the results from our online experiment with two previous independent studies (i.e., i_QoE [4] and CP-QAE-I [12]).

TWO SYSTEMATIC STUDIES ON INDIVIDUAL QOE

To the best of our knowledge, only two studies, i_QoE [11] and CP-QAE-I [12] have investigated systematically the influence of user factors on individual Quality of Experience. These studies have targeted not only perceptual quality but also and explicitly enjoyment of the experience, to obtain a more holistic view on what Multimedia (in this case, video) experiences are and how user factors impact them. Table 4.3 summarizes the experimental settings of the two studies. i_QoE found a weak correlation between perceptual quality and enjoyment. Videos delivered at a low bitrate were reported as having significantly lower perceptual quality than videos with maximum bitrate (about 2000Kbps). A corresponding drop in enjoyment was, however, not recorded. Looking at user factors, an influence of cultural background was found: users from Asia (e.g., China, India)

tended to rate their enjoyment higher than those from western countries (e.g., UK, the Netherlands). With respect to demographics, gender was shown to affect the user's level of involvement. In particular, males were easier involved with the video content than females in this study, and this trend was independent of the video genre. Concerning affective aspects, the level of interest that the user had in the genre of the video s/he was about to watch was found to be positively correlated with enjoyment ratings. Finally, a significant correlation between enjoyment and one of the big 5 personality traits, i.e., extraversion, was observed. CP-QAE-I also showed that lower system parameters (i.e., frame rate, bitrate, and resolution) to lower perception of video quality, but not necessarily lower enjoyment. Lower frame rate had a negative effect on user's level of enjoyment whereas lower bitrate and resolution did not. Especially for content videos with objectionable content (graphic murders etc.), no significant correlation between quality and enjoyment was found, which may be related to the level of interest users had in them (although in this study interest was not measured). Enjoyment was found to be significantly correlated with a subset of positive affect traits [12], i.e., joy, satisfaction, and interest (note that, here, interest refers to user's response to the video they just watched, which is different from the definition of "interest" used in the i_QoE study). Finally, neurotic users tend to enjoy videos with negative emotions (e.g., sad, fearful, or ashamed) more.

COMPARISON

We compare these two studies with the results from our online experiment by using an incremental approach. We use linear mixed-effects models with repeated measures, with model parameters determined with REML, following the analysis method used in our previous study [12]. We only consider common user factors that are available for all three studies (i.e., gender and five personality traits) as input for the analysis. Perceptual Quality and Enjoyment are treated as dependent variables separately. We first analyze the outcomes of three studies separately, and then we outline similarities and differences to understand how user factors play a role in individual QoE.

We first build baseline models which only consider common system factors of three studies, i.e., bitrate and video content (considered as two fixed factors). The results for Enjoyment and Perceptual Quality are shown in Tables 4.4 and 4.5, respectively. It can be seen that both bitrate and video content have a significant effect on the Perceptual Quality in all three studies. With regard to Enjoyment, video content has shown a significant impact whereas bitrate has not for all three studies.

Secondly, we build optimistic models to provide an estimate of the proportion of residual variance which could be reasonably attributed to user factors [12]. The optimistic models consider each user as a random factor, estimating intercept and slope of the linear model per user separately. This allows the model to explain more variance by taking into account individual differences. The optimistic models are then compared with the baseline models on the Mean Squared Residuals (MSR). A reduction in the MSR by the optimistic model with respect to the baseline one is attributed to the presence of the user random factor, and indicates that by incorporating knowledge on individual differences, more accurate models can be devised.

Table 4.4: Baseline Model for Enjoyment on YouQ, i_QoE and CP-QAE-I

Enjoyment		YouQ			i_QoE			CP-QAE-I		
	Source	De.df	F	Sig.	De.df	F	Sig.	De.df	F	Sig.
Baseline	Intercept	179.5	2071.9	.000	319	2960	.000	1183	7157	.000
	Bitrate	177.31	.920	.433	304	1.564	.212	1163	.463	.496
	Content	62.665	14.277	.000	107	17.56	.000	187	39.21	.000
Optimistic	Intercept	159.89	2028.8	.000	129.5	2493	.000	209.1	3066	.000
	Bitrate	137.59	.829	.480	111.9	1.388	.241	209.3	0.038	.846
	Content	58.9	14.398	.000	113.9	19.92	.000	184.2	47.811	.000

*De. df. stands for Denominator degrees of freedom, Sig. stands for significance.

Table 4.5: Baseline Model for Perceptual Quality on YouQ, i_QoE and CP-QAE-I

PQ		YouQ			i_QoE			CP-QAE-I		
	Source	De.df	F	Sig.	De.df	F	Sig.	De.df	F	Sig.
Baseline	Intercept	202	3822.9	.000	259	8394	.000	1159	8559	.000
	Bitrate	102	6.748	.000	237	52	.000	1186	10.78	.001
	Content	202	3.936	.002	133	11	.000	209	9.021	.000
Optimistic	Intercept	163.19	3833.4	.000	140	5943	.000	217.85	4087	.000
	Bitrate	148.70	6.991	.000	96.769	29.49	.000	213.25	4.698	.031
	Content	71.533	4.883	.001	105.55	15.83	.000	179.06	10.98	.000

*De. df. stands for Denominator degrees of freedom, Sig. stands for significance.

Table 4.6: Extended model for Perceptual Quality on YouQ, i_QoE and CP-QAE-I

	YouQ		i_QoE		CP-QAE-I	
Source	F	Sig.	F	Sig.	F	Sig.
Intercept	67.436	.000	154.632	.000	67.315	.000
Bitrate	7.038	.000	51.137	.000	11.932	.001
Content	4.593	.001	11.982	.000	9.045	.000
Gender	1.159	.284	3.738	.054	.568	.451
Extraversion	.000	.994	.035	.852	.001	.976
Agreeableness	.009	.925	7.190	.008	3.069	.080
Conscientiousness	.169	.682	.467	.495	5.382	.021
Neuroticism	.122	.728	2.504	.115	5.934	.015
Openness	.303	.583	1.138	.287	.788	.375

* Sig. stands for significance.

YouQ. For enjoyment, the MSR in the baseline model is 2.42. The optimistic model reduces the MSR to 0.77 implying that 68% of the predicted variance is attributable to individual differences. For Perceptual Quality, the MSR is reduced from 0.82 to 0.63 from the baseline to optimistic model, indicating that an additional 24% of the variance can be explained by the optimistic model, and hence by user factors.

i_QoE. For enjoyment, the MSR in the baseline model is 36.95. The optimistic model reduces the MSR to 28 with an increase of 24% of the overall variance predicted. For Perceptual Quality, the MSR is reduced from 0.76 to 0.61 from the baseline to optimistic model, representing an increase in predicted variance of 20% attributable to users.

CP-QAE-I. For enjoyment, the MSR in the baseline model is 1.39. The optimistic model reduces the MSR to 0.97 with an increase 30% of the overall variance predicted. For Perceptual Quality, the MSR is reduced from 1.35 to 0.99 from the baseline to optimistic model, hence the user factor allows an increase of 27% in the variance explained by the model.

The results above illustrate the importance of user factors since a large proportion of variance can be explained by considering users as a "random effect", especially on the results of YouQ. This suggests that the use of online testing may reveal even more about individual differences in QoE, yet preserving reliable results with respect to the effect of fixed factors on QoE ratings.

Modeling users as a random factor is useful to quantify the impact of user factors on QoE, but has limited applicability when the models are to be deployed in real time to steer individual QoE management, as it is impossible to know a-priori which intercept and slope will correspond to each user. In that sense, it is more convenient to identify individual traits that characterize the user and impact individual QoE, to be used as fixed factors and co-variates in fully defined models. To this end, the next step in our analysis is to identify the most informative user factors in predicting the experience of Enjoyment and/or Perceptual Quality. Note that we only consider the common factors of the three studies (YouQ, i_QoE, CP-QAE-I) as covariates with direct effects (i.e, five personality traits and gender). Again, bitrate and video content are considered as two fixed factors.

Table 4.7: Extended model for Enjoyment on YouQ, i_QoE and CP-QAE-I

	YouQ		i_QoE		CP-QAE-I	
Source	F	Sig.	F	Sig.	F	Sig.
Intercept	8.599	.004	18.330	.000	62.918	.000
Bitrate	1.170	.324	1.739	.188	.433	.510
Content	12.683	.000	18.894	.000	38.718	.000
Gender	.174	.677	.077	.781	9.425	.002
Extraversion	7.600	.007	11.933	.001	.934	.334
Agreeableness	.873	.352	.135	.714	3.990	.046
Conscientiousness	.770	.382	2.507	.114	2.808	.094
Neuroticism	.269	.605	.000	.984	.000	.984
Openness	4.134	.044	.211	.646	3.543	.060

* Sig. stands for significance.

The rest of the setup is kept the same as for the baseline models.

The results of the extended models are shown in Table 4.6 and 4.7. All significant effects are marked in red. For YouQ, an extroverted (F=7.6, p=0.007) and open (F=4.134, p=0.044) user significantly enjoyed the content more. On the other hand, no personality trait significantly affected Perceptual Quality. For i_QoE, an extroverted user (F=11.933, p=0.001) is also found to significantly enjoy the content more. A user who has a more agreeable personality tend to rate the Perceptual quality significantly higher (F=7.190, p=0.008). For CP-QAE-I, female user (F=9.425, p=0.002) and agreeable users (F=3.990, p=0.046) enjoy the video more. A user who tend to have a conscientious (F=5.382, p=0.021) and/or nervous personality (F=5.934, p=0.015) rates Perceptual quality of a video significantly higher.

DISCUSSION

The previous analysis is, to the best of our knowledge, the first meta-analysis across the existing systematic studies of the influence of user factors on individual Quality of Experience. From the results above, we find three major points:

1. *The investigated system factors* do not influence user's level of Enjoyment, though they have a clear negative effect on user's Perceptual Quality, as expected from literature. Lower bitrate corresponds to lower Perceptual Quality, but that does not impact on Enjoyment, implying users may still be satisfied with the experience, even if bit rate decreases a bit. As Enjoyment is strongly influenced by individual differences, user factors should play a more prominent role in future video delivery optimization mechanisms.

2. *User factors* represent a large proportion of variance for both Enjoyment and Perceptual Quality. The results of YouQ indicate that user factors play a major role in determining individual experience. However, no single user factor is found to have a significant impact on either Enjoyment or Perceptual Quality across the three studies. This may be due to the fact that the three studies have only a limited number of user factors in common, and that those are not the most informative ones. For example, i-QoE found a strong impact of interest on Enjoyment, but this factor was not considered in CP-QAE-I, and thus not included in the analysis. In fact, interest is significantly correlated to Enjoy-

Table 4.8: Correlations between Facebook Information and Enjoyment and PQ on YouQ (the labels in the first column are the information collected via YouQ as explained in Section 4.3.2)

	Enjoyment		PQ	
Source	Coeff.*	Sig.	Coeff.	Sig.
user_Friends	.098	.156	.092	.186
user_PhotosUpload	.054	.437	.040	.566
user_PhotosTagged	-0.026	.711	-0.088	.204
user_Status	.142	.043	.111	.114
user_Likes	0.146	.035	.086	.214

* Coeff. stands for Coefficients. Sig. stands for significance.

ment for YouQ (r = 0.295, p = 0.01) suggesting users are interested in a certain genre (e.g., comedy) enjoy the video of that genre more. The result implies that more user factors should be collected/considered in future QoE studies, and also that a large QoE dataset with user factors is badly needed.

3. *Content matters.* Video content itself has been found to have a significant influence on individual QoE in all studies. Except for the above-mentioned relationship between genre and amount of movement in the video which may impact perceptual quality and enjoyment as a consequence, the differences in enjoyment between different video content may relate to user factors (e.g., personality). For instance, users with a certain type of personality (e.g., neurotic) tend to enjoy a certain type of video (e.g., sad or fearful).

4.5.3. THE IMPACT OF USERS' FACEBOOK PROFILE ON BOTH ENJOYMENT AND PERCEPTUAL QUALITY

As introduced in Section 4.3.2, user information was collected automatically from the Facebook profiles of users of YouQ during the Start phase in our experiment. This data could be used to further improve our individual QoE models, having the advantage that it can be collected automatically once the user has given consent. To show the usefulness of this Facebook information, as a first attempt, we performed a spearman rank order correlation analysis between the Facebook information and two individual QoE measures. Here, only the data from the online setting was considered. Since users tend to like the content they enjoy, ratings on Enjoyment are found to have a positive significant correlation with the number of likes users gave (Table 4.8). However, there is no significant correlation between PQ and the Facebook information.

Additionally, all types of Facebook information were normalized and put into a linear model. Here, PQ and Enjoyment were considered as two dependent variables separately, whereas bitrate and video content were considered as two fixed factors; age, gender, total numbers of photo uploaded, photo tagged, likes, friends and posts were considered as covariates. The results are shown in Table 4.9: the number of user likes was found to have significant impact on Enjoyment (F = 4.172, p = 0.043), none of the information has a significant effect on PQ. These results indicate that Facebook information may have an impact on individual viewing experience (i.e., Enjoyment) and should be further inves-

Table 4.9: Linear Model for Enjoyment and PQ on the Importance of Facebook Information

	Enjoyment		PQ	
Source	F	Sig.	F	Sig.
age	2.040	.156	.000	.983
gender	.127	.722	.275	.601
user_Friends	.845	.360	2.297	.132
user_PhotosUpload	.859	.357	.000	.982
user_PhotosTagged	2.167	.145	3.140	.079
user_Status	.200	.656	2.676	.104
user_Likes	4.172	.043	.033	.856

* Sig. stands for significance.

tigated. However, it is important to remark that these results are preliminary and more data needs to be collected to further substantiate them. The use of non-linear models for the prediction of Enjoyment and PQ may be beneficial to uncover complex interactions among the different types of user information.

4.6. CONCLUSION

THERE is no such thing as an "average" user. The next generation of multimedia services has to be geared towards personalization and multimedia delivery needs to start taking individual differences and preferences into account. The associated costs of doing this, in terms of efforts and resources to conduct subjective tests, is, however, a challenge. From a modeling standpoint, it also implies the adoptions of methods which automatically sense user properties and incorporate them in existing quality metrics and models.

In this chapter, we have furthered the understanding of individual Quality of Experience by reporting on the results of a study involving Facebook-based subjective experiments. We designed YouQ, a Facebook application created to allow QoE evaluations of video, which can be configured to collect both self-reports of user information as well as personal data enclosed in the user's Facebook profile. We have shown three important outcomes. First, when launched on Facebook, YouQ allows us to collect data as reliably as in classic lab experiments. In addition, it allows us to reach out to a much more diverse set of users (with a higher gender balance and a greater diversity in cultural backgrounds) with comparatively lower effort (for instance, the experimenter does not need to be present during the experiments), and in a timespan comparable to that of classic lab experiments.

Second, through the first meta-analysis of studies on individual QoE, we showed that individual QoE evaluations collected through YouQ are consistent with those reported in the literature [11, 12]; moreover, they also capture additional individual differences, with a large part of the data variance explained by modeling the user as a random factor. This may, in part, be explained by the fact that every user who accessed the app performed the experiment in a different environment and different device. On the other hand, crowdsourcing studies have shown that the impact of different viewing and envi-

ronmental settings on QoE ratings may be minimal. Hence, we can hypothesize that the greater diversity in user responses is also due to the greater diversity in user background and demographics, allowed by the use of the Facebook platform. We are still actively recruiting participants and expect to have a follow-up study on a much larger cohort. The reason that we want YouQ to go public is for more people in the field to see the value of a Facebook-based experiment and build a great QoE test platform together with us.

Finally, we showed that the personal user information that YouQ automatically acquires from the user profiles can be useful in modeling individual QoE more accurately. For instance, the number of likes users gave is significantly correlated to their Enjoyment. However, it should be noted that, given the importance of online privacy and data protection, Facebook has currently restricted privacy regulations for its apps and it is rather difficult to get approval to collect user information via Facebook apps, as users have the rights to refuse giving their information while using our app. Therefore, more fun and gamification elements should be added in YouQ to attract an even greater number of users, concomitantly ensuring they are willing to share their Facebook information.

Our work raises the potential for some worthwhile future endeavors. Although clearly pointing out the importance of investigating individual differences in QoE, our results have also highlighted that the community is still far from having a clear picture of which user factors are relevant in QoE optimization, and to what extent. As a first priority, the community should aim at shedding some light on this issue. An inventory of potential factors should be created, and their influence systematically investigated across systems and services through empirical research and subjective studies.

Assessing the influence of these user factors on QoE will require extensive empirical research and involve a large number of users in order to have a diverse enough (in terms of individual differences in the factor under observation) sample population. In this sense, traditional lab-based subjective experiments may be limiting. Besides the well-known limitations in terms of ecological validity [76], lab-based subjective experiments are often designed to minimize, rather than embrace, individual differences. Our study proved that Facebook may be a valuable tool towards this end.

Of course, our ultimate goal is to achieve user-centered multimedia delivery optimization and adaptation. For accomplishing this goal, user information needs to be obtained and the user modeling community should distinguish between explicit and implicit methods for inferring user characteristics [13]. Explicit methods require the user to provide information directly, e.g., via web forms or questionnaires. Netflix, for example, asks new users for an indication of their movie preferences with a star-based voting system. When it comes to online optimization, for example for HTTP base adaptive video streaming, explicit user data collection is inconvenient, especially when dynamic user factors are of interest. Implicit, or unobtrusive, methods would be preferable in this case and the study reported in this chapter has indeed shown that social media can be used as a powerful research tool for future QoE evaluation. Nonetheless, more research needs to be done to establish the right mix of methods to collect user information, striking a balance between unobtrusiveness as well as user privacy concerns. We trust that the community will pick up this challenge and, in so doing, go some way towards providing a truly personalized user multimedia experience.

4.7. REFERENCE

[1] Cisco Visual Networking Index. 2013. The zettabyte era–trends and analysis. Cisco white paper (2013).

[2] Conviva. 2015. Experience report. Retrieved from http://www.conviva.com/conviva-viewer-experience-report/ vxr-2015/.

[3] P. Le Callet, S. Möller, and A. Perkis. 2012. Qualinet White Paper on definitions of quality of experience. ENQEMSS (COST Action IC 1003) (2012).

[4] Y. Zhu, A. Hanjalic, and J. A. Redi. 2016. QoE prediction for enriched assessment of individual video viewing experience. ACM Mutimedia, Amsterdam.

[5] J. A. Redi, Y. Zhu, H. de Ridder, and I. Heynderickx. 2015. How passive image viewers became active multimedia users. Springer.

[6] ITU-R BT.500-11, Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, Geneva, Switzerland, 4 (2002), 2.

[7] W. Lin and C.-C. Jay Kuo. 2011. Perceptual visual quality metrics: A survey. JVCI 22, 4 (2011), 297–312.

[8] R. Hu and P. Pu. 2011. Enhancing collaborative filtering systems with personality information. ACM Recsys.

[9] W. Wu, L. Chen, and L. He. 2013. Using personality to adjust diversity in recommender systems. ACM Hypertext.

[10] M. Swan. 2012. Health 2050: The realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen. J. Personalized Medicine 2, 3 (2012), 93–118.

[11] Y. Zhu, I. Heynderickx, and J. A. Redi. 2015. Understanding the role of social context and user factors in video quality of experience. Computers in Human Behavior 49 (2015), 412–426.

[12] M. J. Scott, S. C. Guntuku, Y. Huan, W. Lin, and G. Ghinea. 2015. October. Modelling human factors in perceptual multimedia quality: On the role of personality and culture. ACM Multimedia.

[13] S. Möller and A. Raake. 2014. Quality of experience - advanced concepts, applications and methods. Springer International Publishing Switzerland 2014. https://doi.org/10.1007/978-3-319-02681-7

[14] T. Hoßfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia. 2014. Best practices for QoE crowdtesting: QoE assessment with crowdsourcing. IEEE TMM 16, 2 (2014), 541–558.

[15] S. D. Gosling and W. Mason. 2015. Internet research in psychology. AR Psychology 66 (2015), 877–902.

[16] L. J. Skitka and E. G. Sargis. 2006. The internet as psychological laboratory. AR Psychology 57 (2006), 529–555.

[17] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell. 2015. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations,

and practical guidelines. American Psychologist 70, 6 (2015), 543.

[18] A. Doan, R. Ramakrishnan, and A. Y. Halevy. 2011. Crowdsourcing systems on the world-wide web. Communications of the ACM 54, 4 (2011), 86–96.

[19] J. Redi, E. Siahaan, P. Korshunov, J. Habigt, and T. Hossfeld. 2015. When the crowd challenges the lab: lessons learnt from subjective studies on image aesthetic appeal. CrowdMM 2015.

[20] K. Casler, L. Bickel, and E. Hackett. 2013. Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. CHB 29, 6 (2013), 2156–2160.

[21] R. E. Wilson, S. D. Gosling, and L. T. Graham. 2012. A review of Facebook research in the social sciences. Perspectives on Psychological Science 7, 3 (2012), 203–220.

[22] F. Delogu, M. Franetovic, and L. Shamir. 2015. Keep me posted!: Human and machine learning analysis of facebook updates. Mobile MM Communications, 2015.

[23] M. D. Back, J. M. Stopfer, S. Vazire, S. Gaddis, S. C. Schmukle, B. Egloff, and S. D. Gosling. 2010. Facebook profiles reflect actual personality, not self-idealization. Psychological Science (2010).

[24] L. Gou, M. X. Zhou, and H. Yang. 2014. KnowMe and shareme: Understanding automatically discovered personality traits from social media and user sharing preferences. ACM CHI, 2014.

[25] N. Seitz. 2003. ITU-T QoS standards for IP-based networks. IEEE Communications Magazine (2003), 82–89.

[26] P. Brooks and B. Hestnes. 2010. User measures of quality of experience: Why being objective and quantitative is important. Network, IEEE 24, 2 (2010), 8–13.

[27] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang. 2011. Understanding the impact of video quality on user engagement. ACM Computer Communication Review (2011), 362–373.v [28] G. Ghinea and J. Thomas. 2005. Quality of perception: user quality of service in multimedia presentations. IEEE Transactions on Multimedia 7, 4 (2005), 786–789.

[29] S. Ickin, K. Wac, M. Fiedler, L. Janowski, J.-H. Hong, and A. K. Dey. 2012. Factors influencing quality of experience of commonly used mobile applications. IEEE Communications Magazine 50, 4 (2012), 48–56.

[30] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang. 2012. A quest for an internet video qualityof-experience metric. In Proceedings of the 11th ACM Workshop on Hot Topics in Networks. 2012.

[31] Q. Huynh-Thu and M. Ghanbari. 2008. Temporal aspect of perceived quality in mobile video broadcasting. IEEE Transactions on Broadcasting 54, 3 (2008), 641–651.

[32] J. H. Westerink and J. A. Roufs. 1989. Subjective image quality as a function of viewing distance, resolution, and picture size. SMPTE Journal 98, 2 (1989), 113–119.

[33] N. Staelens, S. Moens, W. Van den Broeck, I. Marien, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester. 2010. Assessing quality of experience of IPTV and video on demand services in real-life environments. IEEE Transactions on Broadcasting 56, 4

(2010), 458–466.

[34] K. Yamori and Y. Tanaka. 2004. Relation between willingness to pay and guaranteed minimum bandwidth in multiplepriority service. Communications, 2004.

[35] C. Owsley, R. Sekuler, and D. Siemsen. 1983. Contrast sensitivity throughout adulthood. Vision Research, 1983.

[36] M. S. El-Nasr and S. Yan. 2006. 3dVisual attention in 3D video games. Workshop on CST, 2006.

[37] P. Orero and A. Remael. 2007. Media for all: Subtitling for the deaf, audio description, and sign language. Rodopi, 2007.

[38] R. E. Nisbett and Y. Miyamoto. 2005. The influence of culture: holistic versus analytic perception. Trends in Cognitive Sciences 9, 10 (2005), 467–473.

[39] E. Balcetis and G. D. Lassiter. 2010. Social Psychology of Visual Perception. Psychology Press, 2010.

[40] A. Acar, T. Taura, E. Yamamoto, and N. F. M. Yusof. 2011. Object vs. relation: Understanding the link between culture and cognition with the help of wordnet. Int. J. of Asian Lang. Proc 21, 4 (2011), 199–208.

[41] I. Wechsung, M. Schulz, K.-P. Engelbrecht, J. Niemann, and S. Möller. 2011. All users are (not) equal-the influence of user characteristics on perceived quality, modality choice and performance. In Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop. Springer, New York, NY.

[42] A. B. Naumann, I. Wechsung, and J. Hurtienne. 2010. Multimodal interaction: A suitable strategy for including older users? Interacting with Computers 22, 6 (2010), 465–474.

[43] P. J. Silvia. 2008. Interest—The curious emotion. Curr. Dir. Psychological Science 17, 1 (2008), 57–60.

[44] H. L. O'Brien and E. G. Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. JAIST 59, 6 (2008), 938–955.

[45] P. Kortum and M. Sullivan. 2010. The effect of content desirability on subjective video quality ratings. Human Factors: The Journal of the Human Factors and Ergonomics Society 52, 1 (2010), 105–118.

[46] J. Palhais, R. S. Cruz, and M. S. Nunes. 2011. Quality of experience assessment in internet TV. In International Conference on Mobile Networks and Management. Springer, Berlin, Heidelberg, 261–274.

[47] A. Sackl, R. Schatz, and A. Raake. 2017. More than I ever wanted or just good enough? User expectations and subjective quality perception in the context of networked multimedia services. QUEx 2, 1 (2017), 3.

[48] A. Sackl and R. Schatz. 2014. Got what you want? Modeling expectations to enhance web QoE prediction. Quality of Multimedia Experience (QoMEX), IEEE, 2014.

[49] A. Sackl and R. Schatz. 2014. Evaluating the influence of expectations, price and content selection on video quality perception. Quality of Multimedia Experience (QoMEX), IEEE, 2014.

[50] P. A. Kara, L. Bokor, A. Sackl, and M. Mourão. 2015. What your phone makes you see:

Investigation of the effect of end-user devices on the assessment of perceived multimedia quality. Quality of Multimedia Experience (QoMEX), IEEE, 2015.

[51] P. G. Engeldrum. 2000. Psychometric Scaling: A Toolkit for Imaging Systems Development. Imcotek Press, 2000.

[52] A. G. Greenwald, D. E. McGhee, and J. L. Schwartz. 1998. Measuring individual differences in implicit cognition: the implicit association test. Journal of Personality and Social Psychology 74, 6 (1998), 1464.

[53] C. Keimel, J. Habigt, and K. Diepold. 2012. Challenges in crowd-based video quality assessment. Quality of Multimedia Experience (QoMEX), IEEE, 2012.

[54] T. Hoßfeld, M. Hirth, J. Redi, F. Mazza, P. Korshunov, B. Naderi, M. Seufert, B. Gardlo, S. Egger, and C. Keimel. 2014. Best practices and recommendations for crowdsourced QoE-lessons learned from the qualinet task force. Crowdsourcing (2014).

[55] M. Weisbuch, Z. Ivcevic, and N. Ambady. 2009. On being liked on the web and in the "real world": Consistency in first impressions across personal webpages and spontaneous behavior. JESP 45, 3 (2009), 573–576.

[56] M. Burke, C. Marlow, and T. Lento. 2009. Feed me: Motivating newcomer contribution in social network sites. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2009.

[57] H. Li, H. Wang, J. Liu, and K. Xu. 2012. Video sharing in online social networks: measurement and analysis. In Proceedings of the International Workshop on Network and Operating System Support for Digital Audio and Video. ACM, 2012.

[58] L. T. Graham, C. J. Sandy, and S. D. Gosling. 2011. Manifestations of individual differences in physical and virtual environments. The Wiley-Blackwell Handbook of Individual Differences (2011), 773–800.

[59] D. Garcia and S. Sikström. 2014. The dark side of facebook: Semantic representations of status updates predict the dark triad of personality. Personality and Individual Differences 67 (2014), 92–96.

[60] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell. 2012. Personality and patterns of Facebook usage. In Proceedings of the 4th Annual ACM Web Science Conference. ACM, 2012.

[61] A. Eftekhar, C. Fullwood, and N. Morris. 2014. Capturing personality from facebook photos and photo-related activities: How much exposure do you need? Computers in Human Behavior 37 (2014), 162–170.

[62] M. Kosinski, D. Stillwell, and T. Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences 110, 15 (2013), 5802–5805.

[63] M. Park, M. Naaman, and J. Berger. 2016. A data-driven study of view duration on Youtube. arXiv:1603.08308 (2016).

[64] X. Yi, L. Hong, E. Zhong, N. N. Liu, and S. Rajan. 2014. Beyond clicks: Dwell time for personalization. In Proceedings of the 8th ACM Conference on Recommender Systems. ACM, 2014.

[65] B. Gardlo, M. Ries, T. Hossfeld, and R. Schatz. 2012. Microworkers vs. facebook: The impact of crowdsourcing platform choice on experimental results. Quality of Multimedia Experience (QoMEX), IEEE.

[66] D. J. Stillwell and M. Kosinski. 2004. Mypersonality project: Example of successful utilization of online social networks for large-scale social research. American Psychologist 59, 2 (2004), 93–104.

[67] H. A. Schwartz, J. Eichstaedt, and M. E. Seligman. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. PloS one 8, 9 (2013), e73791.

[68] D. Preoţiuc-Pietro, S. Volkova, V. Lampos, Y. Bachrach, and N. Aletras. 2015. Studying user income through language, behaviour and affect in social media. PloS one 10, 9 (2015), e0138717.

[69] L. Zwarun and A. Hall. 2014. What's going on? Age, distraction, and multitasking during online survey taking. Computers in Human Behavior 41 (2014), 236–244.

5

Conclusion

1.This chapter consists of material adapted from J. Redi, Y. Zhu, H. de Ridder, I. Heynderickx. *How Passive Image Viewers Became Active Multimedia Users*, in C. Deng, L. Ma, W. Lin, K.N. Ngan: Visual Signal Quality Assessment pp 31-72, Springer, 2015.

5.1. MAIN CONTRIBUTIONS

THE objective of the research reported in this thesis is to develop a feasible method for predicting individual video viewing experience that considers multiple Quality of Experience (QoE) aspects and multiple factors influencing these aspects. To this end, three major steps were taken. The first and the second step investigated the relationship between multiple QoE aspects and their influencing factors on the subjective level (Chapter 2) and objective level (Chapter 3), respectively. These first two studies showed that user factors, reflecting characteristics of a human user, play a critical role in QoE prediction. The third step addressed the issue of reliable data collection for the research on predicting individual QoE. We designed a Facebook-based application for conducting subjective QoE experiments, while automatically collecting information about user factors at the same time. The data collected through this application can be used to train the new model we proposed in Chapter 3 for individual QoE prediction. As a result, a complete, feasible method for individual QoE prediction is presented in this thesis.

More specifically, the contributions that this thesis delivers to the QoE research community can be summarized as follows:

1. We show that QoE is a multifaceted quantity, and that enjoyment and perceptual quality clearly reflect different aspects of QoE, both contributing to the user's final QoE judgement. Enjoyment is shown to be strongly related to the level of satisfaction a user has with a video. We argue that enjoyment should be also considered as a primary aspect in QoE assessment.

2. We show that QoE can be better predicted/characterized when user information (user factors) is taken into account: combining information from video and user as input for a QoE prediction model does a better job in predicting QoE than only using information about the video. We argue that injecting user factors into QoE prediction models enables further in-depth investigation of how user experience differs over demographics or personality, hence enabling a personalized and accurate optimization of video experience.

3. We introduce one of the very first publicly available datasets for individual QoE analysis: i_QoE. The creation of this dataset is motivated by the lack of existing video QoE datasets targeting other aspects of QoE beside perceived quality. i_QoE provides individual QoE ratings on different QoE aspects, e.g., enjoyment, perceived quality, as well as corresponding individual user information (e.g., gender, personality), as collected through extensive user studies (Chapter 2). As such, it can be a valuable tool for QoE researchers investigating the impact of user factors on individual QoE and automatic (individual) QoE prediction.

4. We designed and developed a Facebook-based experimental platform for individual QoE assessment, named YouQ. The objective was to perform subjective experiments for video QoE evaluation while automatically collecting user information via a user's Facebook profile. YouQ is shown to reach out to a more diverse user demographic with little effort as compared to subjective experiments conducted in a controlled lab environment. We experimentally confirm that QoE ratings collected though YouQ are coherent with those reported in traditional laboratory-based experiments, making YouQ and social media-based experimentation an interesting option for individual QoE prediction research.

5.2. PRACTICAL IMPLICATIONS

SERVICE providers know that the more users are satisfied, the longer they will remain customers. Continuously offering satisfying viewing experience is not a static process, it is a dynamic practice that service providers must stay on top of in order to receive optimal results. Currently, most video services are "best-effort" services, i.e., bandwidth is shared amongst all users, and therefore a user gets the video service based on his/her current network conditions. Consequently, a viewing experience measurement relies on network monitoring (e.g., delay) in order to estimate the quality perceived by the users [1, 2].

This thesis challenges current methods in use. Our findings in Chapter 2 show how service providers need to nail down the measurement of "enjoyment" which is closely related to user satisfaction, to be able to properly characterize, and therefore optimize, visual experience. In fact, the leading video service providers start embracing the concept of "enjoyment". In 2017, Netflix has replaced their classic 5-stars rating system with a "thumbs up, thumbs down" system. Neil Hunt, the CPO of Netflix claims the problem of five point system is that "*people subconsciously try to be critics. When they rate a movie or show from one to five stars, they fall into trying to objectively assess the "quality," instead of basing the stars on how much enjoyment they got out of it.* [1]" Such binary rating system happens to hold the same view as our proposed QoE model in Chapter 3, i.e., focusing on "enjoyed" and "not enjoyed" experience.

Moreover, our proposed model in Chapter 3 allows service providers to make decisions regarding user enjoyment and/or to optimize resources (e.g., bandwidth) accordingly. It does so by (1) accounting for different QoE aspects (including enjoyment) and (2) including the individual user perspective on video experience. As such, it could radically change the way streaming resources are optimized, yet guaranteeing high customer satisfaction. For example, instead of using a passive approach and waiting for a user to click on a video before transmission, service providers can estimate individual enjoyment and preload a part of the video that the user might enjoy in order to reduce the long loading time. Moreover, our model can be used to determine the user's minimal perceptual quality requirement to guarantee an enjoyable experience. Based on the output of such model, another possible usage could be a smart personalized subscription plan by providing multiple video quality levels for users to choose from. For example, given a limited monthly data a user has access to, such smart plan can help the user to wisely spend limited data on videos ensuring his/her enjoyment. With an estimation of user enjoyment, service providers may also proactively reshape their bandwidth to better fit in the limited network capacity. For example, instead of lowering perceptual quality for all active users when high network traffic load comes in, a smart dynamic traffic distribution strategy may be developed to better allocate network capacity for individual levels of enjoyment across active users.

This thesis shed light on individual experience prediction as well. Our results show that it is possible to reliably predict the level of enjoyable experience in a particular viewing section, given the available data (both from video and user). Unlike most traditional QoE models, our model aims at predicting the single experience a specific user has with

[1] Please find more information here: https://www.businessinsider.com/netflix-ditches-5-star-ratings-2017-3

a specific video. Service providers need to determine what information they need to collect based on its importance to the goal, e.g., user enjoyment and/or best video quality. Based on our findings in Chapter 4, we suggest social media as a valuable source of information for improving the accuracy of individual experience prediction if the user agrees to share his/her information. Currently, video service providers, such as Netflix or Hulu, allow users to login with their social media accounts, e.g., Facebook and Twitter. However, it is still unclear what the most relevant user information is for individual experience prediction. Therefore, video service providers intend to collect as much information from a user's social media account as possible. Sometimes they collect information simply because they can, without any apparent reason. On the one hand, building an accurate individual QoE model requires a combination of user information from social media and information from the video platform, e,g., viewing time or viewing history, to properly characterize user factors. But deeply digging into user's social media information leads to privacy issues. We think the problem here is that there is a gap between what user information a video service provider is actually using, and what the user thinks the service provider needs. For example, when a user uses Netflix service with his/her Facebook account, he/she thinks it is just for convenience and may not be aware that Netflix may also be looking into his/her Facebook likes and shared pages to improve the user's experience. We suggest to service providers to develop a tool/feature within their service platforms, helping each user to understand where his/her information is used in the service and to see when this usage takes place (e.g., an information usage notification in the platform telling users which information is used and when for improving their experience).

5.3. ANSWERS TO THE RESEARCH QUESTIONS

IN the following, we summarize the findings for each chapter and give answers to the research questions.

1.a. Which user factors influence enjoyment and perceptual quality?

We answered this research question mainly in Chapter 2. We performed an empirical study to investigate the role of user factors (i.e., gender, age, interest, culture background) in combination with system factors (i.e., the bitrate level of a video and video genre) and contextual factors (i.e., the presence/absence of co-viewers). Based on the analysis, we determined that 1) user interest in the video content was positively correlated with that user's enjoyment with the video being watched; 2) no significant impact of any user factor was found for perceptual quality in our empirical study; 3) in addition, male users were found to be less involved in the viewing experience than female users, suggesting that gender influences viewing experience as well; 4) users with different cultural background had different interpretation of the video content resulting in different QoE ratings. In our study, Asian users rated their QoE significantly higher than Western users. Our further analysis in Chapter 3 showed that personality significantly influenced both enjoyment and perceptual quality, though different QoE aspects were influenced by different personality traits.

1.b. How do user factors interact with each other and with context and system factors in forming the final QoE impression?

We answered this research question by showing that 1) low bitrate levels did not af-

fect user's level of enjoyment when watching videos with friends, even though the users could clearly spot the visual artifacts present in the video (Chapter 2); 2) although user interest affected QoE in general, this effect might be suppressed by having company during the viewing experience (Chapter 2); 3) video content was found to have a significant impact on QoE together with user factors (Chapter 4). For example, users with a certain type of personality (e.g., neurotic) were more likely to have an enjoyable experience with a certain type of video (e.g., sad or fearful). We concluded that user factors would influence QoE, particularly in enjoyment. User factors (such as interest and personality) were interrelated and interacted with system/contextual factors, reflecting whether a user had an enjoyable experience or not.

2. Can we design an objective quality model that, by processing user, system, and context information, is able to predict individual QoE?

This research question was mainly addressed in Chapter 3. We proposed a new model by incorporating information related to both system and user factors. We validated our model based on two publicly available QoE datasets. We showed that combining user factors with information from videos (i.e., the perceptual characteristics and the affective charge) could achieve better performance than considering only one type of input information. More in depth, we showed that enjoyment could be better predicted by adding user information related to user interest and personality. Perceived Quality prediction could be improved as well by adding user information related to personality and gender. We concluded that user factors could help improve the performance of QoE prediction in general, but predicting Enjoyment and Perceived Quality required different sets of user information.

3.a. Can we use social media as a platform to perform online experiments aimed at collecting reliable subjective assessments of QoE?

3.b. Can social media-based experimentation support the reliable and unobtrusive collection of user factor data?

To answer this research question, we developed YouQ, an online platform to conduct subjective QoE experiments. Via YouQ, we were able to collect user information both from self-reports and user's Facebook profile. In particular, the information from the Facebook profile can be collected automatically with user's permission. Our analysis showed that the user information collected from Facebook profile was useful for QoE prediction. We tested YouQ both on Facebook and in a controlled lab experiment. In addition, the data collected in Facebook were further compared with two previous independent studies. We found that Facebook-based experiments could acquire data as reliably as a lab experiment, meanwhile capturing more individual differences. These individual differences were due to the diverse demographics YouQ could reach through Facebook. These findings encouraged us to consider Facebook as an appealing source for future individual QoE research.

5.4. LESSONS LEARNED AND FUTURE DIRECTIONS

We conclude this thesis with an overview of the limitations we identified and highlight some possible directions for further research. As it should be clear by now, user factors play a crucial role in predicting individual QoE. Individual differences should be taken into account in the production and distribution of multimedia content. How-

ever, individual QoE prediction is a rather challenging problem, and this thesis is just an initial step towards solving it. A number of actions still need to be taken to inject knowledge in the QoE community. Below we elaborate on each action in more detail.

(1) There are still many unexplored user factors

Although clearly pointing out the importance of investigating individual differences in QoE, this thesis has only investigated a limited number of user factors (i.e., interest, gender, culture background and personality). We believe the QoE community should aim at creating a taxonomy of user factors and their impact on multimedia experiences.

There are certainly many other user factors that potentially influence Quality of Experience. It is commonly agreed that user factors can have a relatively stable character (i.e., invariant through time) or rather a variable one [3]. In fact, this thesis only investigated several stable factors (e.g., personality, culture and interest), since they are relatively easy to quantify and embed into individual QoE prediction models. Previous experience is one stable factor that we did not investigate but that quite clearly may influence QoE. Depending on previous experiences, users form expectations regarding enjoyment and perceptual quality that may differ considerably on an individual basis [4]. In [5], authors suggested to take user's previous experience into account in QoE evaluation. Most users now watch, rate (like/dislike), comment and share videos online, which will be logged on social media and/or in the databases of the video providers. This information can be considered as a representation of a user's previous experience. In fact, the leading service providers (e.g., Netflix, YouTube) have used such information to recommend new videos to their users [6, 7]. For example, Netflix recommends video that a user may like to watch based on his/her previously watched videos [6]. It would be interesting to take one step further and see whether this information can predict individual QoE.

Variable factors, such as curiosity, are also important to investigate. Users intend to seek elements that are new or unusual in one's environment [8]: novel elements attract curious users and bring out enjoyable experiences [9, 10]. User's current emotional state is another variable user factor. Emotion is a reaction towards a specific object or event [11] and users may therefore experience positive or negative emotional states during the interaction with online video services, influencing their (judgment on the) experience. In a similar way, affective states pre-existing the experience may influence it too [12]. For example, negative emotional states, such as frustration, anxiety, and boredom, may lead to a lower appreciation of the experience [13]. We see some early attempts trying to incorporate a user's emotional state into QoE models [14-17]. However, variable user factors change rapidly over time, making them difficult to quantify. Effort is needed in order to find ways to accurately track these variable user factors. While most research in this direction is now using accurate implicit physiological response (EEG, Skin conductance, etc.,), the devices necessary to collect this information are still quite obtrusive and may not be usable for predicting everyday experiences of paying consumers [16, 17]. We see that more convenient and smart devices (e.g., Microsoft Band [2] or Apple Watch[2]) equipped with physiological sensors (measuring e.g., heart rate, or galvanic skin response) have been released on the market. A few studies have tried to use such devices to measure multimedia experience [18]. However, these commercial devices usually sac-

[2]Microsoft Band: https://www.microsoft.com/en-us/band
[2]Apple watch: https://www.apple.com/lae/watch/

rifice accuracy and performance in order to improve comfort and lower the price [17]. To this end, it is necessary to validate the reliability of commercial devices as compared to the traditional ones used in the lab and select a promising set of physiological responses that are meaningful to viewing experience. At last, it is interesting to check whether considering these implicit physiological responses as additional input can improve the performance of our individual QoE prediction model.

It is important to note that the abovementioned user factors are not a complete list. More physiological, demographic, affective and socio-cultural factors exist that may be worth exploring. Furthermore, it is worth remarking that there are some user factors (e.g., socio-economic or educational background) that are also considered to be part of the contextual factors, because these factors are difficult to disentangle and categorize [11]. Creating an inventory of influencing user factors on QoE is quite challenging but will have profound significance to the QoE community.

(2) A large-scale QoE empirical experiment requires a well-designed, popular social media-based application

Studying user factors will require extensive empirical experiments. We recommend to use Facebook as the primary social media experimental platform for video QoE research since other social media platforms are not particularly good for video sharing. For example, Instagram is ideally suited for image sharing. This thesis investigated Facebook as a platform to run extensive empirical studies ensuring sufficient diversity across users. However, there are still challenges ahead in this research direction. First, we found that the use of social media for empirical research was less straightforward than expected. In particular, similar to what was observed for crowdsourcing-based experimentation [19], short tasks are mandatory for social media–based experiments in order to limit the risk of disengagement of users. This poses limitations when investigating user factors e.g. on personal traits (e.g. personality, intelligence, creativity), which are typically quantified through extensive self-report questionnaires. Further research is needed to transform these self-report questionnaires into sufficiently short versions to be included in 5-minutes QoE assessment tasks. Alternatively, proxy measures, e.g., the analysis of user behavior on social media or comment analysis, should be found allowing us to assess user factors in a short time.

In addition, we have to look for new ways to attract users to join the Facebook-based experiment. YouQ is still not comparable with those apps attracting millions of active users. According to [20], the most effective way to attract users on Facebook is by so-called “snowball sampling”, i.e., convincing the existing users to invite their friends to join. For example, MyPersonality [21], one of the most popular Facebook experimental applications, was originally shared with 150 Facebook users, and finally attracted over 6 million users in 4 years. However, making a Facebook experimental platform go viral requires great efforts in aesthetic design, appealing gamification elements and socially supercharged functions [20]. As a result, developing such application needs a large number of people combining different expertise, as well as the necessary funding. That is why we decided to open source our YouQ application. We hope the QoE community can see the value of Facebook-based experimentation and will work together towards improving this experimental platform.

(3) The QoE community needs to compile a code for ethical treatment of user data

In this thesis, we showed that user factors are important in predicting individual viewing experience and Facebook has made it easier than ever before to collect data on user factors in an unobtrusive fashion. At the time of writing this thesis, Facebook provides 55 sources of user information for its application, ranging from self-provided information (e.g., user's name, age, and/or favorite movies) to user activity information (e.g., user's likes, attended events, and/or installed applications). Although this user information has great value for QoE research, it introduces great ethical challenges (i.e., privacy preservation) as well. How to provide personalized experiences without invading privacy?

Users are usually more worried about their privacy as compared to the benefits they receive from personalized services [22]. The research community should treat privacy with extra care when collecting user information unobtrusively (even more if it comes from social media, where the information was shared with different intent than that of the research). Based on our experience in designing Facebook-based experimentation, we would give the following suggestions:

1) Researchers should always inform the users about the intention of the experiment they are about to join, make sure the users are aware of the potential privacy risk, i.e., let users know what kind of personal information the experiment is about to collect, how it will be stored, protected, and for how long it will be kept.

2) Users should be able to stop sharing their information at any time they want. This is also suggested by Facebook, because it allows users to be in control over their own data.

3) Researchers should remove from the data analysis and storage username, profile pictures or any other sensitive information which may potentially trace back to a specific user [23].

4) Researchers should only use/collect the user information related to the experiment. They should not blindly collect all sorts of user information from social media only because it is technically possible.

5) Researchers should be aware that any model making automatic decisions (e.g., loan, health insurance) may produce discriminatory and unfair outcomes, especially when it comes to individual prediction. Therefore, we would suggest to have an ethical review process before publishing papers related to those topics. More importantly, when deploying any of such models in the market, it should be transparent to its user, letting the user understand how the decision has been made and what type of personal information has been used as input.

On 25 May 2018, the General Data Protection Regulation (GDPR) [4] is implemented in the EU to protect online privacy for all individuals within Europe. However, until now, only very few documents provided guidelines for internet-based experimentation [24, 25], the latest one was even before Facebook was founded. Recently, authors from [23] proposed a protocol for health research using social media. We argue that a standard protocol is urgently needed for Facebook-based QoE research, i.e., a guideline for designing experiments, storing data, and analyzing results without violating individual privacy. Such clear protocol can help researchers achieve better ethical clearance and share their results with the QoE community.

[4] More information at: https://www.eugdpr.org/

5.5. REFERENCE

[1] Skorin-Kapov, Lea, et al. "A Survey of Emerging Concepts and Challenges for QoE Management of Multimedia Services." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 14.2s: 29, 2018.

[2] Robitza, Werner, et al. "Challenges of future multimedia QoE monitoring for internet service providers." Multimedia Tools and Applications 76.21: 22243-22266, 2017.

[3] Brunnström, et al. Qualinet white paper on definitions of quality of experience, 2013.

[4] Sharot, Tali, and Elizabeth A. Phelps. How arousal modulates memory: Disentangling the effects of attention and retention. Cognitive, Affective, & Behavioral Neuroscience 4.3: 294-306, 2004.

[5] Bampis, Christos G., and Alan C. Bovik. Learning to predict streaming video QoE: Distortions, rebuffering and memory. arXiv preprint arXiv:1703.00633, 2017.

[6] Gomez-Uribe, Carlos A., and Neil Hunt. "The netflix recommender system: Algorithms, business value, and innovation." ACM Transactions on Management Information Systems (TMIS) 6.4: 13, 2016.

[7] Walker, Russell, et al. "Netflix leading with data: The emergence of data-driven video." Kellogg School of Management Cases: 1-19, 2017.

[8] Huang, Ming-Hui. Designing website attributes to induce experiential encounters. computers in Human Behavior 19.4 : 425-442, 2003.

[9] Zhang, Chenyan, et al. Quality of Alternate Reality Experience and Its QoE Influencing Factors. Proceedings of the 2nd International Workshop on Multimedia Alternate Realities. ACM, 2017.

[10] Isikman, Elif, et al. The effects of curiosity-evoking events on activity enjoyment. Journal of experimental psychology: applied 22.3: 319, 2016.

[11] Reiter, U., et al, A. Factors influencing quality of experience. In Quality of Experience (pp. 55-72). Springer International Publishing. 2014.

[12] Redi, J. A., et al, I. How passive image viewers became active multimedia users. In Visual Signal Quality Assessment (pp. 31-72). Springer International Publishing, 2015.

[13] O'Brien, H. L. and Toms, E. G. What is user engagement? A conceptual framework for defining user engagement with technology. Journal of the American Society for Information Science and Technology, 59, 6, 938-955, 2008.

[14] Chen, Min, et al. User intent-oriented video QoE with emotion detection networking. Global Communications Conference (GLOBECOM), IEEE, 2016.

[15] Moldovan, Arghir-Nicolae, et al. Analysis of Learner Interest, QoE and EEG-Based Affective States in Multimedia Mobile Learning. Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on. IEEE, 2017.

[16] Engelke, Ulrich, et al. Psychophysiology-based QoE assessment: a survey. IEEE Journal of Selected Topics in Signal Processing 11.1: 6-21, 2017.

[17] Moon, Seong-Eun, and Jong-Seok Lee. Implicit Analysis of Perceptual Multimedia Experience Based on Physiological Response: A Review. IEEE Transactions on Multimedia 19.2: 340-353, 2017.

[18] Röggla, Thomas, et al. "The Club of The Future: Participatory Clubbing Experiences." Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, 2017.

[19] Hoßfeld, Tobias, et al. "Best Practices and Recommendations for Crowdsourced QoE-Lessons learned from the Qualinet Task Force" Crowdsourcing"." 2014.

[20] Kosinski, Michal, et al. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. American Psychologist 70.6: 543, 2015.

[21] Bachrach, Yoram, et al. Your digital image: factors behind demographic and psychometric predictions from social network profiles. Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

[22] Kang, Ruogu, et al. my data just goes everywhere: user mental models of the internet and implications for privacy and security. Symposium on Usable Privacy and Security (SOUPS). Berkeley, CA: USENIX Association, 2015.

[23] Benton, Adrian, et al.. Ethical research protocols for social media health research. Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. 2017.

[24] Hewson, Claire. Conducting research on the internet. PSYCHOLOGIST-LEICESTER-16.6: 290-293, 2003.

[25] Kraut, Robert, et al. Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. American psychologist 59.2: 105, 2004.

ACKNOWLEDGEMENT

This four-year PhD life is the most precious experience in my life so far. Beside the fun of living in a completely new environment, the PhD life enhanced my scientific competence, broadened my view, and made me more international. I would not enjoy the PhD life as much as I do, without the help from the people around me.

The beginning of my PhD life was difficult. Everything was new, the place, the people, and the weather, when I first arrived in this "low land" (Netherlands). I am grateful to my co-promotor Judith Redi; she was patient to explain things to me. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor for my Ph.D study. Dear Judith, I don't have enough words to express my gratitude for everything you did for me.

My thanks are intended to my promoters, Ingrid Heynderickx and Alan Hanjalic, for their advices, time invested in my thesis. Without their guidance and constant feedback this PhD would not have been achievable. I would also thank you for your trust in me, which helped me in my scientific endeavor and being a source of inspiration.

My gratitude is intended for all the committee members, which helped me through their valuable comments to improve this thesis in many ways. Many thanks also to Wei Lei who convinced me during our many discussions back in Southeast University that I should pursue my doctoral degree and who made it possible for me to obtain a scholarship from CSC. I also would like to thank Weisi Lin who offered me a position in Nanyang Technological University, Singapore.

I also would like to thank Willem-Paul Brinkman for sharing his knowledge of statistical analysis. Acknowledgement is also given to the people from Pi-Lab, Huib de Ridder, Sylvia Pont, Harold Nefs, Tatiana Kartashova, Tingting Zhang, Ling Xia, Fan Zhang, Ernestasia Siahaan, thank you for broadening my view on the area of image quality perception.

In addition, I would like to thank all my colleagues, Martha Larson, Cynthia Liem, Huijuan Wang, Soude Fazeli, Raynor Vliegendhart, Alessio Bazzica, Christina Katsimerou, Xinchao Li, Bo Qu, Bcabak Loni, Karthik Yadati, Jeahun Kim and Xiuxiu Zhan, in my group. I miss the days we had our "weekly meeting" to report our progress in research and discuss other work plans. You also gave me different kinds of support in my research such as giving advice on building my experimental setups, being the participants of my experiments.

Academic activities in our group are well supported. I appreciate the help from Ruud de Jong, Bart Vastenhouw and Saskia Peters. I give great thanks to Ruud and Bart for helping me set up my experiments, and to Saskia for all the arrangements.

I want to express my sincere thanks to my friends, Junchao Xu, Kang Ni, Chao Qu, Yun Ling, Tingting Zhang, Chang Wang, Changyun Wei, Wenxin Wang, Bin Shi, Haozheng Zhu, Jie Li, and so on, my PhD life would not have went so smoothly without you.

Last but not least, I want to thank my parents for your endless love and encouragement from the other side of the earth, China. I give great thanks to my wife Qian Liu. She spent much time taking care of our little boy, Oscar Zhu. Of course, I have to thank my son, a special gift during my PhD life. He brought us quite a lot of joy and prolonged my PhD life.

YI ZHU
Nanjing, June 2019