QoE Prediction for Enriched Assessment of Individual Video Viewing Experience

Yi Zhu
TU Delft
Delft, the Netherlands
Y.Zhu-1@tudelft.nl

Alan Hanjalic
TU Delft
Delft, the Netherlands
A.Hanjalic@tudelft.nl

Judith A. Redi
TU Delft
Delft, the Netherlands
J.A.Redi@tudelft.nl

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 perceived quality (related to coding or network-induced artifacts), but also other – more hedonic - aspects, like enjoyment. In this paper, we propose an expanded model which aims to assess QoE of a given video, not only in terms of perceived 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 perceived 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 perceived quality and enjoyment.

Keywords

QoE; User Factors; Objective Quality Assessment.

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.

MM '16, October 15-19, 2016, Amsterdam, Netherlands

DOI: http://dx.doi.org/10.1145/2964284.2964330

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 perceived (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.

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 paper as specified in the following items:

Contribution 1: We expand the traditional QoE assessment paradigm (Figure 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 1.a) are illustrated in Figure 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 perceived 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 and the experimental setup for its validation in Section 4. The results are presented in Section 5, while the discussion in Section 6 summarizes the main insights and sets targets for future research.

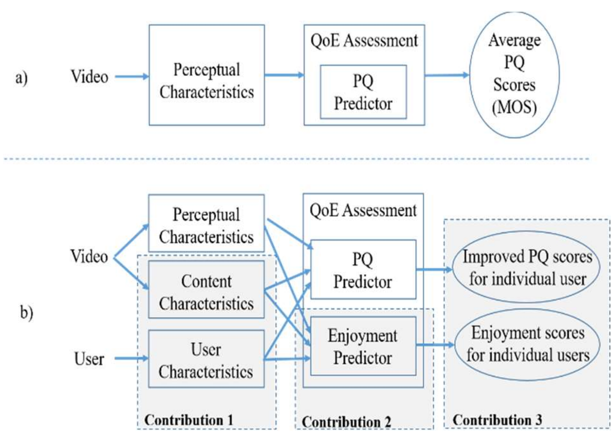

Figure 1 a) The traditional approach to assessing video QoE, b) The proposed expanded QoE assessment model targeting individual scores representing two selected QoE aspects, PQ and enjoyment.

2. RELATED WORK

2.1 Perceived 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 measure is popular since it can express the perceived 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 paper we abandon the MOS-based PQ assessment paradigm and explore possibilities for individual PQ assessment.

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 paper.

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, however, 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 paper.

3. THE PROPOSED QoE MODEL

As illustrated in Figure 1a, the prediction of QoE in a video viewing scenario has been approached so far according to the following three principles:

- QoE was simplified to PQ only,
- PQ was assessed using a limited set of perceptual characteristics related to the visibility of artifacts and other video impairments,
- 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 paper, 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 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 paper, 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 paper 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 a future work.

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 traditional approach (Figure 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, …M) and are denoted as $PC_k(m)$, where k = 1,2,…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:

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

where the gamma function is defined as $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt,\ a > 0$ 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 γ, β_l, β_r:

$$f(x;\ \gamma, \beta_l, \beta_r) = \begin{cases} \frac{\gamma}{(\beta_l+\beta_r)\Gamma(\frac{1}{\gamma})} \exp\left(-\left(\frac{-x}{\beta_l}\right)^{\gamma}\right) & \forall x < 0 \\ \frac{\gamma}{(\beta_l+\beta_r)\Gamma(\frac{1}{\gamma})} \exp\left(-\left(\frac{-x}{\beta_r}\right)^{\gamma}\right) & \forall x \geq 0 \end{cases} \quad (2)$$

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

$$\eta = (\beta_r - \beta_l) \frac{\Gamma(\frac{2}{\gamma})}{\Gamma(\frac{1}{\gamma})} \quad (3)$$

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 α, β) 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,…36, that feed out 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].

3.2 Content characteristics

As described in Section 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 paper 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 paper.

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

$$MA = \frac{1}{M}\sum_{m=0}^{M-1}\frac{100}{N_m|\vec{v}_{MAX}(m)|}\left(\sum_{i=1}^{N_m}|\vec{v}_i\,(m)|\right) \tag{4}$$

In (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:

$$\mathrm{HR} = \sum_{m=1}^{M}\frac{G_m}{H_m}/M \tag{5}$$

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

$$\mathrm{S} = \frac{\sum_{m=1}^{M}100e^{\wedge}((1-(m_n-m_p))/\delta)}{M} \tag{6}$$

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

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.

3.4 The Prediction modules

The proposed model, depicted in Figure 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 the 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 prediction process may be needed depending on the selected model [8]. In this paper, 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 4.2.

4. EXPERIMENTAL 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 4.1), predictor implementation (Section 4.2), and the evaluation procedure (Section 4.3).

4.1 Dataset description

To the best of our knowledge, only two public datasets, namely i_QoE (available at http://ii.tudelft.nl/iqlab/iQOE.html) and CP-QAE-I (Available at: 1drv.ms/1M1bnwU), meet our requirements, i.e., include user characteristics and individual QoE (PQ and enjoyment) ratings. These two datasets were derived from two independent user studies conducted in [40] and [29], respectively.

4.1.1 i_QoE

As shown in table 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. 59 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 perceived video quality on a 5-point ACR scale, according to [16]. In total, 354 ratings (59x6 videos) on these two QoE aspects were collected.

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*.

Table 1 Description of the core characteristics of the i_QoE and CP-QAE-I datasets (F = female, M = male)

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

4.1.2 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 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.

4.1.3 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 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 of each dataset in our study, with values being normalized between 0 and 1. In addition, as shown in Table 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.

4.2 Prediction module implementation

Both datasets report discrete (ordinal) ratings of enjoyment and PQ. The rating distributions are shown in Figure 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, perceived 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 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 perceived quality of the video unacceptable. The CP-QAE-I dataset

Table 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	154UE, 200E(56.4%)	130UQ, 224AQ (63.28%)
CP-QAE-I	771NE, 461E(62.58%)	788UQ, 444AQ (63.96%)

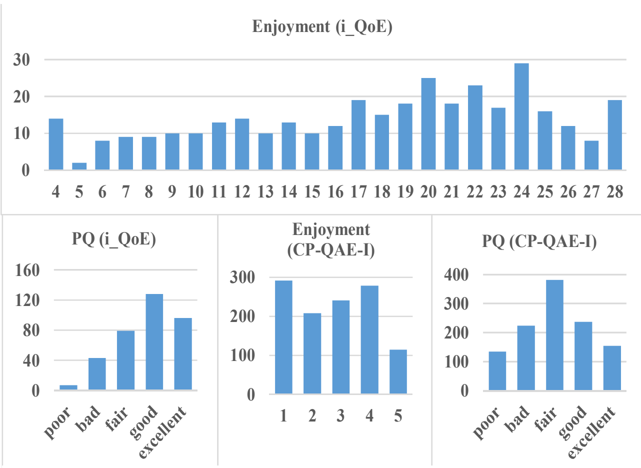

Figure 2. Histograms showing the distribution of individual ratings for the two datasets considered in the experiments. Here, the X-axis represents points on the rating scale, whereas the Y-axis represents the number of instances for each score

includes 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 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 *UCs* 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.

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:

$$\text{Accuracy} = 1 - \sum_{r=1}^{R} MisRate_r \,/R \qquad (13)$$

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 II, the prior probability for a user to find a video enjoyable in i_QoE is calculated as:

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

The prior probabilities per QoE aspect and dataset are indicated in Table 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].

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.

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.

5.1.1 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 and 4 respectively, whereas the MCC values and the confusion matrices are reported in Table 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

Table 3. i_QoE：The performance of LDC and SVM on Enjoyment and Perceived Quality, based on all characteristics, the three characteristic categories and selected indicators

	Enjoyment (Baseline 56.4%)					Perceived Quality (Baseline 63.28%)				
Predictor	All	CC	PC	UC	Selected	All	CC	PC	UC	Selected
LDC	67.79%	63.02%	59.59%	66.11%	69.51%	78.24%	64.94%	80.22%	60.14%	78.48%
SVM	69.51%	63.02%	60.15%	65.30%	69.51%	80.22%	64.94%	80.22%	63.27%	80.22%

Table 4. CP-QAE-I：The performance of LDC and SVM on Enjoyment and Perceived Quality, based on all characteristics, the three characteristic categories and selected indicators

	Enjoyment (Baseline 62.58%)					Perceived Quality (Baseline 63.96%)				
Predictor	All	CC	PC	UC	Selected	All	CC	PC	UC	Selected
LDC	71.27%	70.45%	65.42%	61.85%	63.80%	65.50%	64.44%	66.07%	63.96%	64.77%
SVM	73.13%	70.37%	68.67%	65.75%	68.53%	68.18%	65.42%	67.20%	69.57%	66.23%

Table 5. The confusion matrices and MCC of LDC and SVM for Enjoyment and Perceived 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

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

Table 6. Key influencing characteristics for LDC on i_QoE

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

Table 7. Key influencing characteristics for LDC on CP-QAE-I

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

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).

5.1.2 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 and 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 80.22% 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 around 70.45%. 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.

5.1.3 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.

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. Then, a series of independent sample t-tests were performed between corresponding values from using either the key or all characteristics in order to

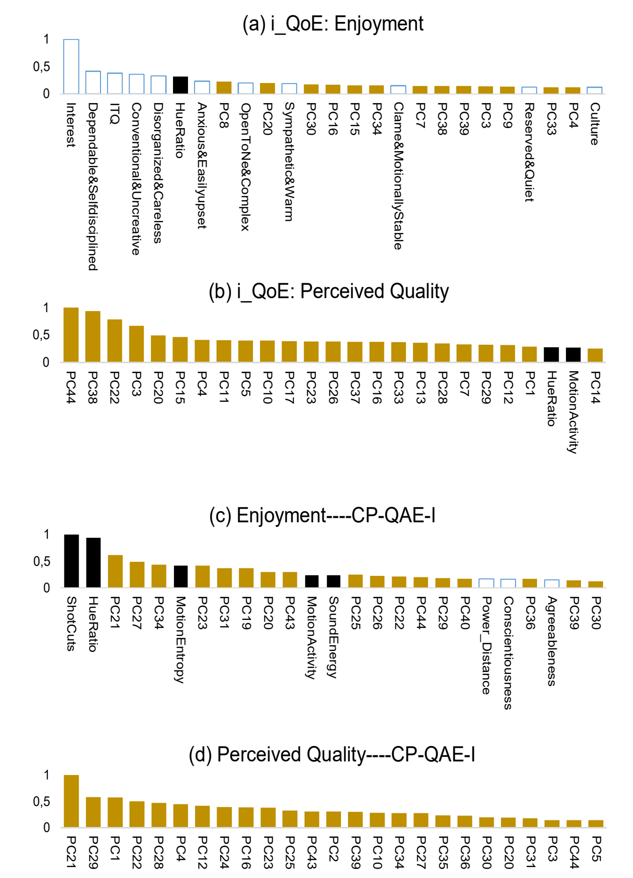

Figure 3. Relevance (y-axis) of the key influencing indicators per QoE aspect and dataset, by using SVM. Yellow bars indicate perceptual characteristics, black bars indicate content characteristics, and white bars concern user characteristics

check whether there is a statistically significant difference between the performances in both cases. Tables 6 and 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. 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 personal preferences on video content and genres (e.g., possibly via social media by tracking user 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.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.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 5.1.2) 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.b, 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 and 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.

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 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 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

Table 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 9. The confusion matrices, Accuracy (Acc) and MCC values of SVM for validation on merged datasets

	Enjoyment Acc:67.91%, MCC:0.33	
Pred./ Truth	NE	E
NE	731	194
E	315	346

	Perceived Quality Acc:68.91% MCC:0.35	
Pred./ Truth	UQ	AQ
UQ	805	113
AQ	380	288

perceived quality prediction in this experiment was 58.32% 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 the 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 model here were trained on a smaller number of common indicators (especially *UC* characteristics), missing essential characteristics (such as, e.g., *interest*) and possibly under-fitting the data.

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 1.

Regarding the 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 in 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 perceived 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).

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.