Computers in Human Behavior 49 (2015) 412–426

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Understanding the role of social context and user factors in video Quality of Experience

Yi Zhu [a,*], Ingrid Heynderickx [b], Judith A. Redi [a]

[a] Delft University of Technology, Mekelweg 4, Delft 2628 CD, The Netherlands
[b] Eindhoven University of Technology, Postbus 513, Eindhoven 5600 MB, The Netherlands

ARTICLE INFO

Article history:
Available online 28 March 2015

Keywords:
Video quality assessment
Quality of Experience
Social context
User factors

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 paper aims to identify the role of social context 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 toward building an accurate model of user QoE appreciation, to be deployed in future multimedia systems to optimize the user experience.

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 (Moore, 2011). These figures are forecasted to further grow in the coming years (Cisco, 2012; Moore, 2011). 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 (Pérez, Macías, Ruiz, & García, 2011; Wang, Speranza, Vincent, Martin, & Blanchfield, 2003). 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 (Naumann, Wechsung, & Hurtienne, 2010; Yamori & Tanaka, 2004). 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 (Le Callet, Möller, & Perkis, 2012), 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 (Serral-Gracià et al., 2010). 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 (Asghar, Le Faucheur, & Hood, 2009), 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 (Dobrian et al., 2011), 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 (Hemami & Reibman, 2010; Lin & Jay Kuo, 2011). Artifact visibility was considered to be inversely related to perceptual quality, and therefore to user satisfaction (Chikkerur, Sundaram, Reisslein, & Karam, 2011). In

* Corresponding author.
E-mail address: Y.Zhu-1@tudelft.nl (Y. Zhu).

http://dx.doi.org/10.1016/j.chb.2015.02.054
0747-5632/

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 (Le Callet et al., 2012; Zhu, Heynderickx, & Redi, 2014). For example, recent studies claimed that QoE should also be considered from a user perspective (De Pessemier, De Moor, Joseph, De Marez, & Martens, 2013): evidence has been provided that user's interest (Kortum & Sullivan, 2010) and personality (Wechsung, Schulz, Engelbrecht, Niemann, & Möller, 2011) 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 context factors (Le Callet et al., 2012). *System* factors concern the technical aspects of a multimedia system (e.g., network parameters, media genre and media configuration). *User* factors refer to individual characteristics of the user who is experiencing the video (e.g., demographics, personal interest or personality). *Context* 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 context 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 (Tercek, 2011). 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 context 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 perceived video quality, enjoyment, endurability, satisfaction, involvement and information assimilation. The outcomes should support building an accurate objective model for QoE on the longer term.

The paper continues by presenting the related work in Section 2, which we reviewed to define the hypotheses for the empirical study as described in Section 3. We then outline our experimental methodology in Section 4, followed by the analysis of the results in Section 5. We discuss our findings in Section 6, leading to the most important conclusions in Section 7.

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" (ITU-T, 1994). 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 (Brooks & Hestnes, 2010). 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 (2007). 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 (Hemami & Reibman, 2010; Lin & Jay Kuo, 2011).

Recently, the Qualinet White Paper (Le Callet et al., 2012) 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 (Le Callet et al., 2012) 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 2.1), we use the Qualinet definition as operational definition of QoE throughout this paper.

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 1, these factors can be arranged into in three categories, namely system factors, user factors and context factors (Le Callet et al., 2012).

System factors refer to the system, application and media "properties and characteristics that determine the technically produced quality of an application or service" (Jumisko-Pyykkö, 2011). Within video delivery services, system factors can influence QoE by altering the perceptual quality of the video (Serral-Gracià et al., 2010). 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 (Dobrian et al., 2011), and the media configuration (Gulliver & Ghinea, 2006) 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 session time, including playing plus buffering) is inversely related to QoE (Dobrian et al., 2011), 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 (Ickin et al., 2012; Mok, Chan, & Chang, 2011). Besides the signal/network factors, user's QoE with video also may be influenced by the nature of the video content itself (Balachandran et al., 2012). Different genres

Table 1
Factors influencing QoE discussed in this section.

System factors	User factors	Context factors
Signal and network parameters (Dobrian et al., 2011; Gulliver & Ghinea, 2006; Serral-Gracià et al., 2010)	Interest (Kortum & Sullivan, 2010; O'Brien & Toms, 2008)	Physical environment (Westerink & Roufs, 1989)
	Personality (Wechsung et al., 2011)	Economic aspects (Kilkki, 2008)
Genre (Rainer et al., 2012)	Demographics (Naumann et al., 2010; Wolters et al., 2010)	Social situation (Chorianopoulos & Lekakos, 2008; Kubey & Csikszentmihalyi, 2013; Lee & Lee, 1995)

(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 (Huynh-Thu & Ghanbari, 2008). 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 (Rainer et al., 2012; See-To, Papagiannidis, & Cho, 2012).

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 (O'Brien & Toms, 2008) and that personal interest in video content significantly influences user's QoE judgment (Kortum & Sullivan, 2010). 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 (Palhais, Cruz, & Nunes, 2012). 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 (Wechsung et al., 2011). 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 (Wolters et al., 2010). However, another study demonstrates the opposite trend: older users tend to rate video quality more positively than younger users, although the performance is worse (Naumann et al., 2010). Similar scattered results exist for gender (Bracken, 2005; Hyder, Crespi, Haun, & Hoene, 2012; Murray, Qiao, Lee, Muntean, & Karunakar, 2013), and no systematic investigation has been carried out, to the best of our knowledge, to clarify the role of demographic characteristics in QoE appreciation.

Context 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 (Westerink & Roufs, 1989). 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. (Kilkki, 2008) 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 paper we mainly focus on one particular context 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 (Scheinkman, 2008), 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 (Morrison & Krugman, 2001) may increase user's overall satisfaction with the program (Oehlberg, Ducheneaut, Thornton, Moore, & Nickell, 2006). The social element of the viewing experience stretches even further, with users recording their favorite programs and sharing them with families and friends (Kubey & Csikszentmihalyi, 2013), or with people using their viewing experience as a conversation topic (Lee & Lee, 1995). Recently, a concept of "social TV" has emerged: it provides multiple viewers with a joint TV-watching experience by adding communication features (Chorianopoulos & Lekakos, 2008). 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 (Fröhlich, Baillie, & Schatz, 2006). All these results point toward 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. Existing approaches to measure QoE

QoE has been historically measured in two ways: objectively and subjectively (Redi, Zhu, de Ridder, & Heynderickx, 2015). 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., (Fiedler, Hossfeld, & Tran-Gia, 2010; Hemami & Reibman, 2010; Kim et al., 2008; Lin & Jay Kuo, 2011). 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 context factors to QoE (Redi, 2013; Staelens et al., 2010). To design objective metrics that can properly reflect the influence of context 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 (Engeldrum, 2000) 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 specific aspects 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 perceived overall quality of a video (ITU-R, 2002).

In fact, many studies are based on the analysis of MOS of perceived video quality (PVQ) 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 (Ghinea & Thomas, 2005; Gulliver & Ghinea, 2004). Gulliver and Ghinea (2004) 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 (O'Brien & Toms, 2008). 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 (O'Brien & Toms, 2008). 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 (Antons, Arndt, Schleicher, & Möller, 2014; De Moor et al., 2014). 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 (O'Brien & Toms, 2008).

The measures of QoE, as discussed above, have been administered in different forms, i.e., by means of self-report (De Moor, Quintero, Strohmeier, & Raake, 2013; See-To et al., 2012), via interviews (Desmet, 2005) or with physiological measurements, such as facial expression, galvanic skin response and EEG (Antons et al., 2014; De Moor et al., 2014). It should be pointed out here, that besides for PVQ, 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 perceived visual quality, enjoyment, satisfaction, endurability, involvement and information assimilation.

3. Research questions and hypotheses

Based on the literature overview given in Section 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: perceived visual 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 (Oehlberg et al., 2006), we formulate our first hypothesis as:

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

It is generally known that video bitrate affects perceived video quality (Ickin et al., 2012; Mok et al., 2011), and so, also QoE; the lower the bitrate level, the lower PVQ, 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 presented 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 (Palhais et al., 2012), 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 (Witmer & Singer, 1998). 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 (O'Brien & Toms, 2008), 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 (Bradley, Codispoti, Sabatinelli, & Lang, 2001) and have different perception of olfactory and visual media synchronization (Murray et al., 2013). Some impact of age on QoE has been demonstrated, though not all reported results in literature were consistent (Naumann et al., 2010; Wolters et al., 2010). Finally, users with a different cultural background usually have a different understanding of experience, and thus may perform differently toward a same task (Marcus, 2006). 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 measured QoE along six dimensions, and characterized the participants along the user factors interest, immersive tendency and demographics.

4. Experimental set-up

4.1. Experimental design

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 Fig. 1a) 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 Fig. 1b) 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.

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). 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.

4.2. Participants

Sixty participants (i.e., 27 females and 33 males) 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 Delft University of Technology, 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.

4.3. 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. A clear preference (as in gathering over 50% of the choices) was found for two genres, as shown in Fig. 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 Fig. 2c) was also used in our study.

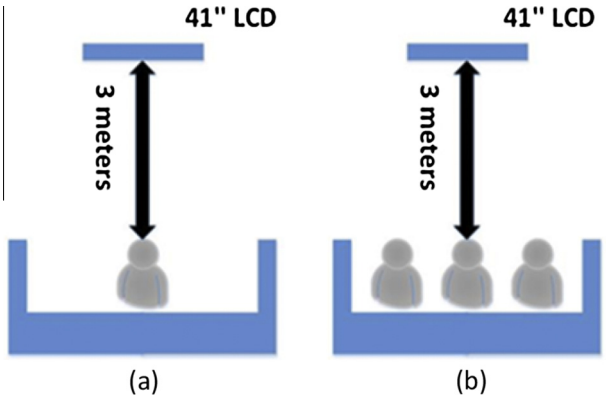

Fig. 1. The two different social contexts investigated in the experiment. In viewing situation (a), a single participant watched videos on a 41″ screen from a couch 3 m away. In viewing situation (b), three participants, friends, watched the same video together in the same environmental conditions.

Table 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.

	Genre 1		Genre 2		Genre 3	
	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

We selected 2 different clips from each of above three genres (screenshots are given in Fig. 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 min and originated from YouTube. They were encoded with H.264/AVC, which is the most commonly used codec for online videos (Schwarz, Marpe, & Wiegand, 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 PVQ) 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, 2006) with a bit rate of 112 kbps to avoid any effect of the sound on QoE.

Finally, three 10 s-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

[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.

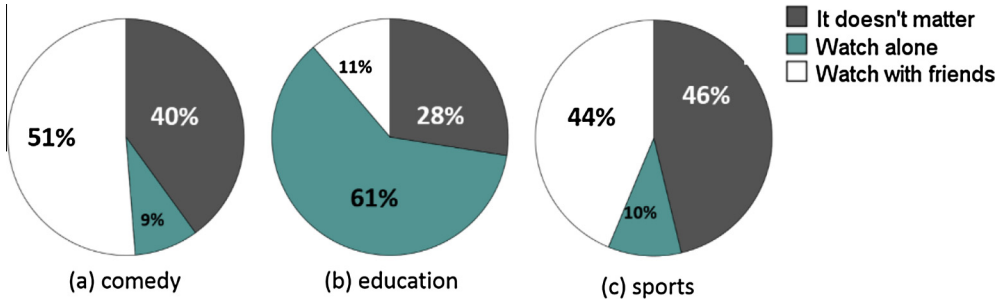

Fig. 2. Results of the online survey on preferred social context for watching (a) comedy, (b) education and (c) sports videos. The white area in the pie plots indicates preference for watching alone, the greenish area indicates preference for watching with friends, and the gray area indicates that it does not matter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

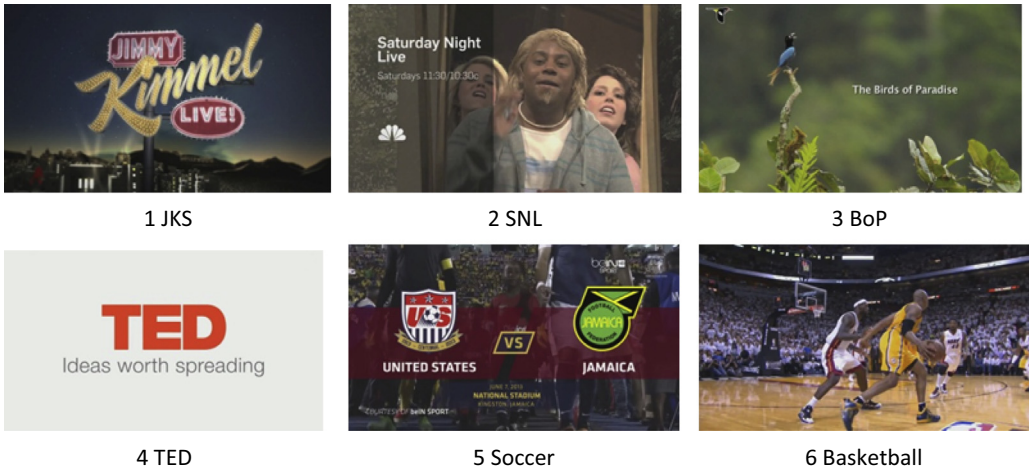

Fig. 3. Screenshots of the six video clips tested: The Jimmy Kimmel Show (JKS) and Saturday Night Live (SNL) for comedy, The Birds of Paradise (BoP) and a TED talk (TED) for education, and Soccer and Basketball for sports.

Table 3
An overview of the three questionnaires used in our study.

Questionnaire	Aspect	Number of Questions	Measure scale	Dependent (D)/Independent (I)
UF questionnaire	Demographics	9	–	I
	Interest on genre	3	7-point Likert	I
	Immersive Tendency	18 (Witmer & Singer, 1998)	7-point Likert	I
QoE questionnaire	Enjoyment	4 (See-To et al., 2012)	7-point Likert	D
	Endurability	4 (O'Brien & Toms, 2008)	7-point Likert	D
	Satisfaction	4 (See-To et al., 2012),	7-point Likert	D
	Involvement	4 (Schubert et al., 2001)	7-point Likert	D
	Perceived Visual Quality	2 (ITU-R, 2002)	5-point ACR	D
IA questionnaire	Information Assimilation	24 (Gulliver & Ghinea, 2004)	True–False	D

low). These samples were used to let participants get acquainted with the range of video quality used in the experiment.

4.4. 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 m) in order to satisfy the preferred viewing distance (ITU-R, 2002). 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.

4.5. 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 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), (Witmer & Singer, 1998). 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 perceived visual quality (PVQ). Information assimilation was measured with a separate questionnaire. Satisfaction and enjoyment were measured through subsets of the questionnaire proposed by *See-To (2012)*. Involvement was measured through an adapted version of the Igroup Presence Questionnaire (Schubert, Friedmann, & Regenbrecht, 2001). Items from *O'Brien's* questionnaire *(2008)* were used to measure endurability. Each of these four aspects was quantified through 4 items to be rated on a 7-point Likert scale. Perceived video 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 (2002)* 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 were not. 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.

4.6. 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-s 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.

5. Results

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 were not 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.

5.1.1. 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 (Witmer & Singer, 1998). 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 (Ling, Nefs, Brinkman, Qu, & Heynderickx, 2013). 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 two social situations ($U = 13{,}770$, $p = 0.49$), 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 ($U = 15{,}153$, $p = 0.593$), suggesting that the distribution of user interest was similar for the two groups of participants.

5.1.2. QoE questionnaire

The QoE questionnaire consisted of five sub-questionnaires (see Table 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) (Cortina, 1993). 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 4, for four of the five sub-questionnaires, α was higher than 0.82, indicating high internal consistency. The α value of PVQ, on the other hand, was lower than 0.8, indicating that the two items included in the sub-questionnaire investigated a slightly different concept.

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 PVQ, we instead decided to analyze the Opinion Scores (OS) separately for the two aspects. As a results, PVQ 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 between them. As shown in Table 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 ($0.58 < r < 0.62$). The two PVQ items (OS1 and OS2 in Table 5) were also moderately correlated with each other ($r = 0.63$), 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

Table 4
Internal consistency measured by Cronbach's alpha among items in each aspect of the questionnaire.

Aspect	Enjoyment	Involvement	Satisfaction	Endurability	PVQ
Chronbach's α	0.917	0.82	0.961	0.939	0.773

Table 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 perceived video 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**

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

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.

5.1.3. 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 Fig. 4, genre had a significant impact on IAS ($\chi^2 = 22.093$, $p < 0.001$) 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.

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 Fig. 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 (Wobbrock, Findlater, Gergle, & Higgins, 2011). 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 PVQ 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 Fig. 5a and 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

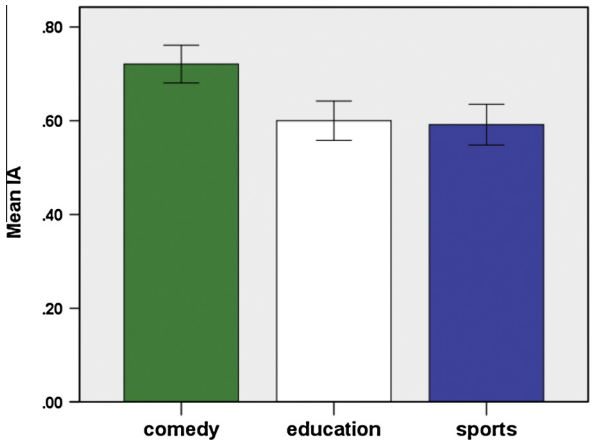

Fig. 4. The information assimilation for the three different genres. Here the *Y*-axis represents the mean percentage of correct answers for all participants. Comedy videos were on average better understood than education and sports videos.

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 Fig. 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 did not 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 did not 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 PVQ 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 Fig. 5e and 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 PVQ 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 Fig. 6a and b, participants clearly recognized the 600 kbps videos as having lower quality and more visible artifacts than the 2000 kps videos with the same, 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 did not find an effect of presence of co-viewers on involvement, satisfaction, PVQ and information assimilation. We have to reject hypothesis H2a, since the presence of co-viewers did not affect the tolerance to artifacts, i.e., the PVQ aspect of QoE. Finally, genre only affected the impact of social context on involvement, but not in the hypothesized way. We did not 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.

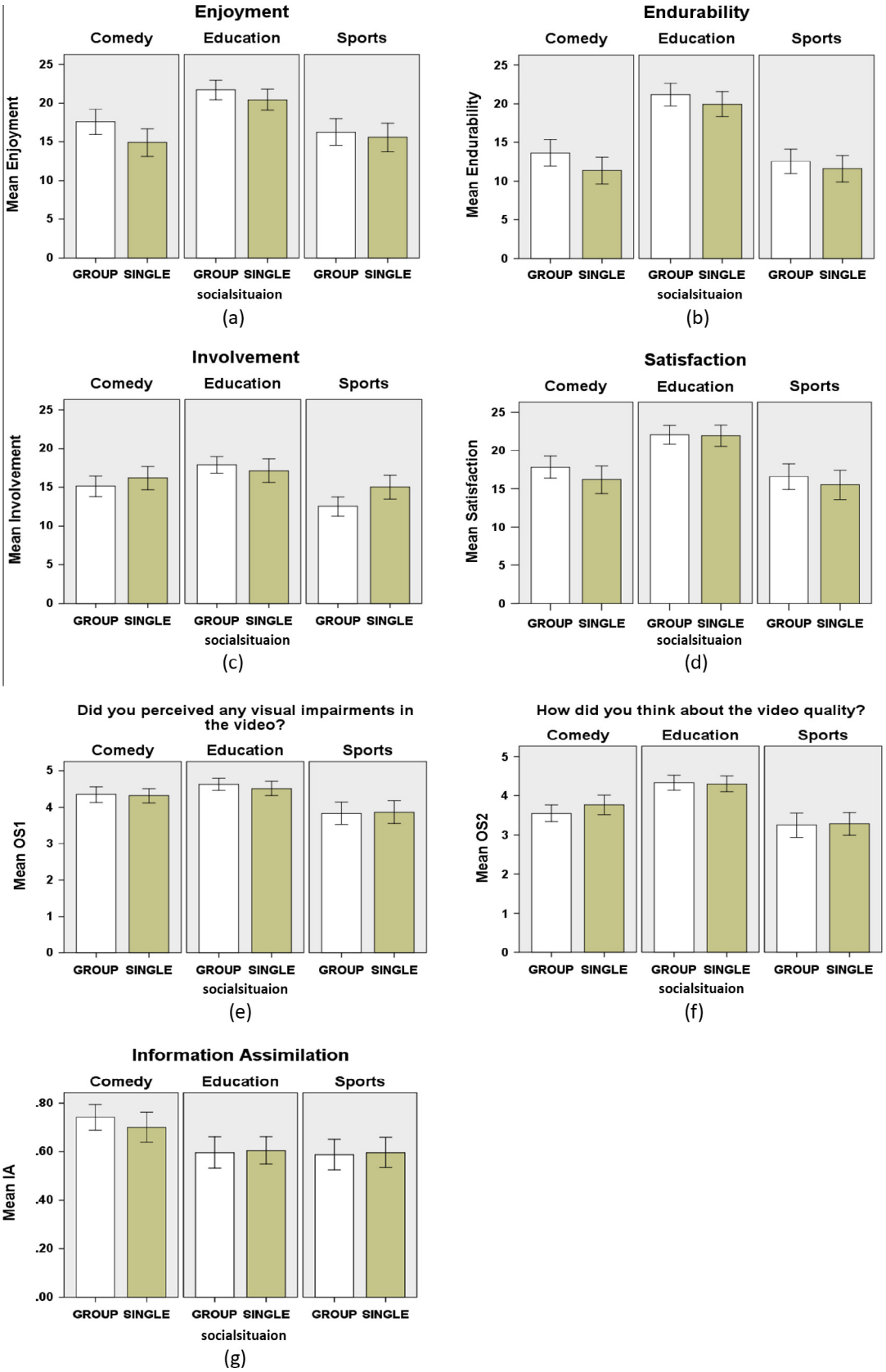

Fig. 5. The mean AS, OS and IAS for the two social situations and three video genres: (a) the mean AS of enjoyment, (b) the mean AS of endurability, (c) the mean AS of involvement, (d) the mean AS of satisfaction, (e and f) the mean Opinion Scores of the two PVQ items, and (g) the mean IA. The white 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

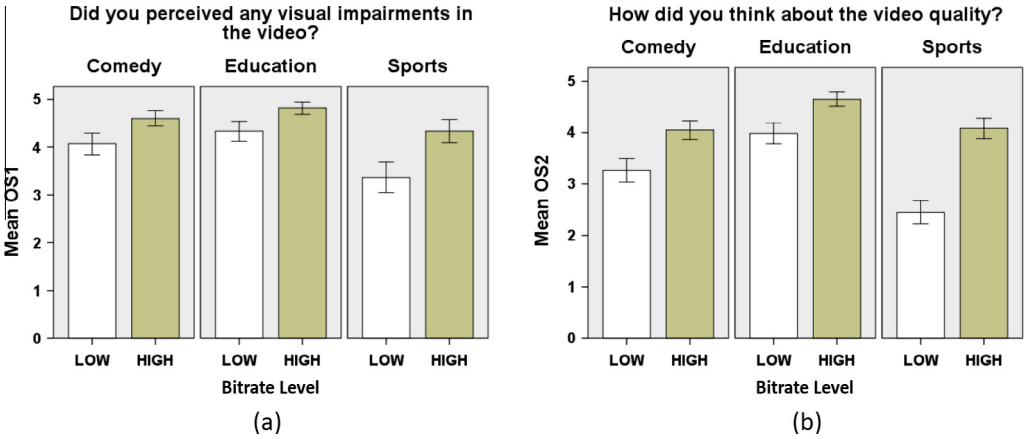

Fig. 6. The mean OS for the two bitrate levels (600 kpbs and 2000 kbps) split up for the three video genres. The white bars indicate the low bitrate level, while the yellow bars indicate the high bitrate level. The mean OS per video genre (i.e., comedy, education and sport) are shown in three separated columns.

Surprisingly, PVQ was the only QoE aspect affected by the bitrate level: we did not 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.

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

5.3.1. 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 6, satisfaction and endurability 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 PVQ 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 interest and IA sores, 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 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.

5.3.2. 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.

5.3.3. 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 did not 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 Fig. 7 this trend was independent on the video genre. No significant influence was found on any other QoE aspect.

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 ($U = 12{,}979$, $p = 0.008$) and endurability ($U = 13357.5$, $p = 0.023$). As shown in Fig. 8, Asian participants tended to rate

Table 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
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
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
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
Education				
		OS1	OS2	
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).

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 did not find a significant difference between the two cultural backgrounds for involvement, the two PVQ 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.

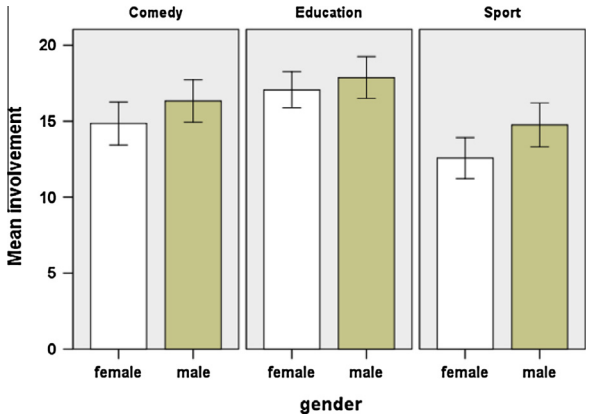

Fig. 7. The effect of gender on involvement for the three video genres separately. Here the white bars represent the involvement scores for the female participants, while the colored bars represent the scores for the male participants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 perceived visual quality (along two separate dimensions of (1) artifact visibility and (2) overall quality), enjoyment, satisfaction, endurability, involvement and information assimilation. Measurement scales for perceived visual 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 toward 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 do not 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 perceived visual quality, enjoyment, endurability and satisfaction as equally important aspects of QoE, but obviously more research toward a validated questionnaire is needed.

Multiple factors that potentially influence QoE have been identified in literature (Le Callet et al., 2012), 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 perceived 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

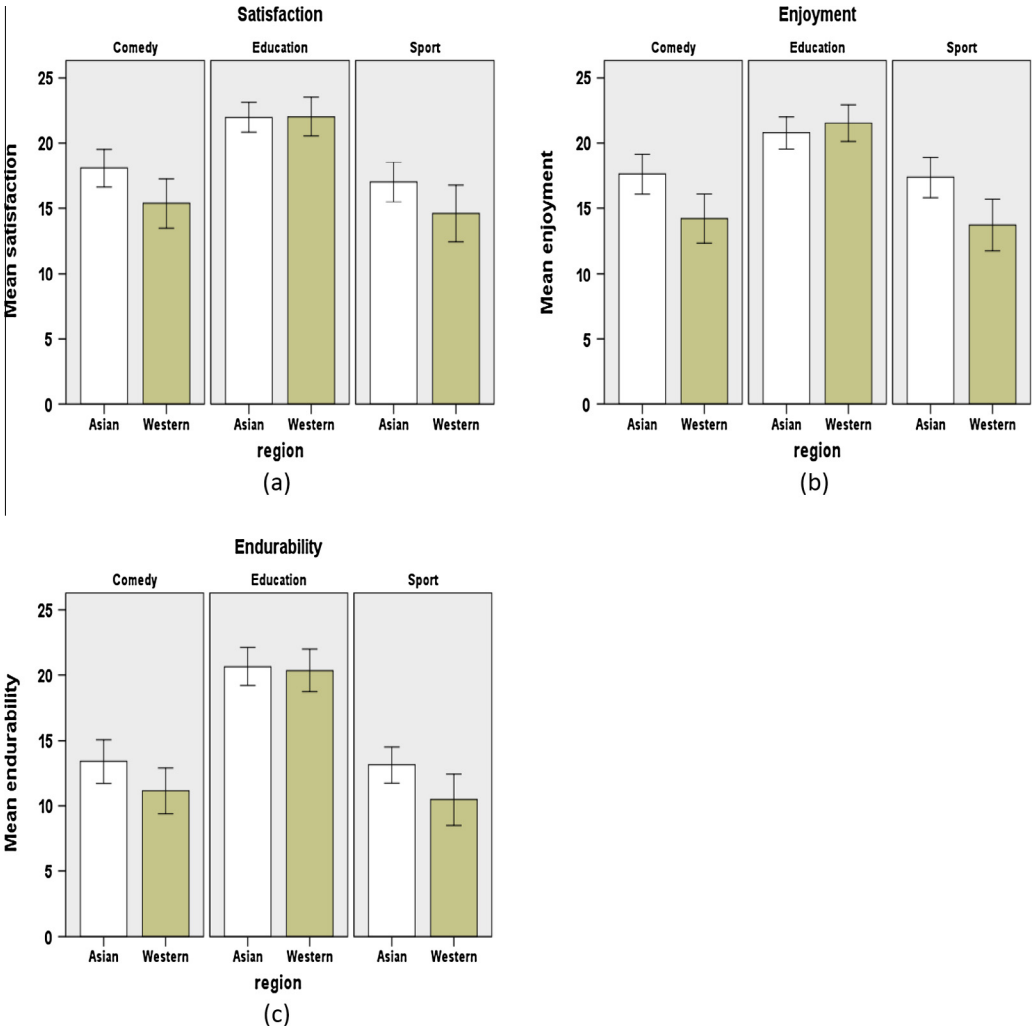

Fig. 8. The mean score on satisfaction (a), enjoyment (b) and endurability (c) for the two cultural backgrounds and the three video genres. The white bars represent the Asian participants, while the yellow bars represent the Western participants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 context 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 from 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 (Palhais et al., 2012). Yet, interest did not correlate with perceived video quality, contradicting previous results which showed a link between content desirability and PVQ (Kortum & Sullivan, 2010). 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 (Ling et al., 2013). 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 (Gardlo, Ries, Hossfeld, & Schatz, 2012). 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 ($F(4,353) = 21.727$, $r = 0.447$, $R^2 = 0.199$, $p < 0.001$), 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, Zunino, and Redi (2013)).

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.

7. Conclusions

In this paper, 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.

Admittedly, we considered limited test conditions (i.e., two bitrate levels and three video genres) and number of participants in this study. To ensure generalization over video contents, a much larger pool of videos, covering possibly more artifacts and genres, will have to be evaluated. Furthermore, more users, with more diverse demographics, should be involved in the evaluation of video experiences. In addition, we only considered one context factor (i.e., the presence/absence of co-viewers) and a few user factors (i.e., gender, culture background, interest and immersive tendency). In order to extend the validity of our findings and to design an overarching model for video QoE appreciation, future investigations should target the impact of other user factors (i.e., personality, mood), other types of social contexts (e.g., presence of strangers, or impact of online communities) and other context factors (e.g., environmental conditions, screen size and type, payment scheme for the video service). In addition, how these potential influencing factors interact with known system factors other than bitrate (e.g., type of compression, buffer ratio) should to be explored.

Clearly, covering such a vast amount of factors and interactions would require the collection of a huge amount of data. For this reason, the prosecution of this work based on controlled lab experiments seems rather unfeasible. One possible solution to this is to consider performing online experiments using crowdsourcing (e.g., via platforms such as Mechanical Turk or Microworkers[2]). These platforms allow to reach out to a large amount of users by asking them to accomplish a small task in exchange of a symbolic payment (Redi et al., 2013). It has been shown that large amounts of subjective QoE data can be collected via crowdsourcing within few days and with a high reliability (Hoßfeld et al., 2014). This would enable the exploration of more video genres, video artifacts, network parameters, and simulated contextual schemes in the crowd-based experiment. More importantly, crowdsourcing would give us the opportunity to collect additional user factors from populations with very diverse demographics (e.g., participants form diverse culture and different age range).

The long-term goal of this work is to define an overarching model that, given a user watching a video through a given system and in a given context, is able to predict user's appreciation for the viewing experience based on user, context and system characteristics. Such model is expected not only to be able to estimate user's perceptual quality (as most existing video quality metrics do) but also other QoE aspects (e.g., enjoyment or endurability). Such model is expected to allow tackling video delivery bottlenecks in a more user-oriented way. In the case of bandwidth scarcity, for example, existing protocols envision an equal decrease in bitrate for all users. Our model would instead allow to allocate bandwidth per user in a smarter way, e.g. ensuring that all users have a similar level of endurability (or enjoyment, or satisfaction, depending on service provider's requirement), based on their tolerance for visual artifacts.

Acknowledgements

This work is supported in part by the scholarship from China Scholarship Council (CSC) under the Grant CSC No. 201206090028. This work is also partially supported by the NWO Veni Grant 639.021.230.

References

Antons, J.-N., Arndt, S., Schleicher, R., & Möller, S. (2014). *Brain activity correlates of quality of experience. Quality of experience*. Springer, pp. 109–119.

Asghar, J., Le Faucheur, F., & Hood, I. (2009). Preserving video quality in IPTV networks. *IEEE Transactions on Broadcasting, 55*, 386–395.

Balachandran, A., Sekar, V., Akella, A., Seshan, S., Stoica, I., & 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, pp. 97–102.

Bracken, C. C. (2005). Presence and image quality: The case of high-definition television. *Media Psychology, 7*, 191–205.

[2] Mechanical Turk: https://www.mturk.com/mturk/welcome; Microworkers: https://microworkers.com/.

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion, 1*, 300.
Brooks, P., & Hestnes, B. (2010). User measures of quality of experience: Why being objective and quantitative is important. *Network, IEEE, 24*, 8–13.
Chikkerur, S., Sundaram, V., Reisslein, M., & Karam, L. J. (2011). Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE Transactions on Broadcasting, 57*, 165–182.
Chorianopoulos, K., & Lekakos, G. (2008). Introduction to social TV: Enhancing the shared experience with interactive TV. *International Journal of Human–Computer Interaction, 24*, 113–120.
Cisco, I. (2012). Cisco visual networking index: Forecast and methodology, 2011–2016. CISCO White paper, 2011–2016.
Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98.
De Moor, K., Quintero, M. R., Strohmeier, D., & Raake, A. (2013). *Evaluating QoE by means of traditional and alternative subjective measures: An exploratory'living room lab'study on IPTV*, Vienna, Austria.
De Moor, K., Mazza, F., Hupont, I., Quintero, M. R., Mäki, T., & Varela, M. (2014). Chamber QoE: A multi-instrumental approach to explore affective aspects in relation to quality of experience, IS&T/SPIE Electronic Imaging. *International Society for Optics and Photonics*, pp. 90140U-90140U-90114.
De Pessemier, T., De Moor, K., Joseph, W., De Marez, L., & Martens, L. (2013). Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching. *IEEE Transactions on Broadcasting, 59*, 47–61.
Desmet, P. (2005). *Measuring emotion: Development and application of an instrument to measure emotional responses to products. Funology*. Springer, pp. 111–123.
Dobrian, F., Sekar, V., Awan, A., Stoica, I., Joseph, D., Ganjam, A., et al. (2011). Understanding the impact of video quality on user engagement. *ACM SIGCOMM Computer Communication Review, 41*, 362–373.
Engeldrum, P. G. (2000). *Psychometric scaling: A toolkit for imaging systems development*. Imcotek Press.
Fiedler, M., Hossfeld, T., & Tran-Gia, P. (2010). A generic quantitative relationship between quality of experience and quality of service. *Network, IEEE, 24*, 36–41.
Fröhlich, P., Baillie, L., & Schatz, R. (2006). Exploring the Joint iTV Experience. FTW Technical Report, FTW-TR-2006-005.
Gardlo, B., Ries, M., Hossfeld, T., & Schatz, R. (2012). *Microworkers vs. facebook: The impact of crowdsourcing platform choice on experimental results. In 2012 Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, pp. 35–36.
Gastaldo, P., Zunino, R., & Redi, J. (2013). Supporting visual quality assessment with machine learning. *EURASIP Journal on Image and Video Processing, 2013*, 1–15.
Ghinea, G., & Thomas, J. (2005). Quality of perception: User quality of service in multimedia presentations. *IEEE Transactions on Multimedia, 7*, 786–789.
Gulliver, S. R., & Ghinea, G. (2004). Stars in their eyes: What eye-tracking reveals about multimedia perceptual quality. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 34*, 472–482.
Gulliver, S. R., & Ghinea, G. (2006). Defining user perception of distributed multimedia quality. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2*, 241–257.
Hemami, S. S., & Reibman, A. R. (2010). No-reference image and video quality estimation: Applications and human-motivated design. *Signal Processing: Image Communication, 25*, 469–481.
Hoßfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., et al. (2014). Best practices for QoE crowdtesting: QoE assessment with crowdsourcing. *IEEE Transactions on Multimedia, 16*, 541–558.
Huynh-Thu, Q., & Ghanbari, M. (2008). Temporal aspect of perceived quality in mobile video broadcasting. *IEEE Transactions on Broadcasting, 54*, 641–651.
Hyder, M., Crespi, N., Haun, M., & 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. In *Emerging trends and applications in information communication technologies* (pp. 200–209). Springer.
Ickin, S., Wac, K., Fiedler, M., Janowski, L., Hong, J.-H., & Dey, A. K. (2012). Factors influencing quality of experience of commonly used mobile applications. *Communications Magazine, IEEE, 50*, 48–56.
ISO/IEC, 2006. ISO/IEC 14496-3: 2005. Amd.
ITU-R (2002). BT.500-11, Methodology for the subjective assessment of the quality of television pictures. In *International Telecommunication Union*, Geneva, Switzerland 4, 2.
ITU-T (1994). Recommendation E. 800: Terms and definitions related to quality of service and network performance including dependability. ITU-T August 1994.
ITU-T (2007). Definition of Quality of Experience. TD 109rev2 (PLEN/12), Geneva, Switzerland (pp. 16–25).
Jumisko-Pyykkö, S. (2011). User-centered quality of experience and its evaluation methods for mobile television. Doctoral thesis, Tampere University of Technology, Tampere.
Kilkki, K. (2008). Quality of experience in communications ecosystem. *Journal of Universal Computer Science, 14*, 615–624.
Kim, H.-J., Lee, D. H, Lee, J. M, Lee, K.-H., Lyu, W., & Choi, S.-G. (2008). The QoE evaluation method through the QoS–QoE correlation model. In *Fourth international conference on networked computing and advanced information management, 2008. NCM'08* (pp. 719–725). IEEE.
Kortum, P., & 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*, 105–118.
Kubey, R., & Csikszentmihalyi, M. (2013). *Television and the quality of life: How viewing shapes everyday experience*. Routledge.
Le Callet, P., Möller, S., & Perkis, A. (2012). Qualinet white paper on definitions of quality of experience. In *European network on quality of experience in multimedia systems and services (COST Action IC 1003)*.
Lee, B., & Lee, R. S. (1995). How and why people watch TV: Implications for the future of interactive television. *Journal of Advertising Research, 35*, 9–18.
Lin, W., & Jay Kuo, C.-C. (2011). Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation, 22*, 297–312.
Ling, Y., Nefs, H. T., Brinkman, W.-P., Qu, C., & Heynderickx, I. (2013). The relationship between individual characteristics and experienced presence. *Computers in Human Behavior, 29*, 1519–1530.
Marcus, A. (2006). Cross-cultural user-experience design. In *Diagrammatic representation and inference* (pp. 16–24). Springer.
Mok, R. K., Chan, E. W., & Chang, R. K. (2011). Measuring the quality of experience of HTTP video streaming. In *2011 IFIP/IEEE international symposium on integrated network management (IM)* (pp. 485–492). IEEE.
Moore, K. (2011). 71% of online adults now use video-sharing sites. Pew Internet and American Life Project.
Morrison, M., & 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 and Electronic Media, 45*, 135–161.
Murray, N., Qiao, Y., Lee, B., Muntean, G.-M., & Karunakar, A. (2013). Age and gender influence on perceived olfactory & visual media synchronization. In *2013 IEEE international conference on multimedia and expo (ICME)* (pp. 1–6). IEEE.
Naumann, A. B., Wechsung, I., & Hurtienne, J. (2010). Multimodal interaction: A suitable strategy for including older users? *Interacting with Computers, 22*, 465–474.
O'Brien, H. L., & 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*, 938–955.
Oehlberg, L., Ducheneaut, N., Thornton, J. D., Moore, R. J., & Nickell, E., 2006. Social TV: Designing for distributed, sociable television viewing. In *Proc. EuroITV* (pp. 25–26).
Palhais, J., Cruz, R. S., & Nunes, M. S. (2012). Quality of experience assessment in internet TV. In *Mobile networks and management* (pp. 261–274). Springer.
Pérez, P., Macías, J., Ruiz, J. J., & García, N. (2011). Effect of packet loss in video quality of experience. *Bell Labs Technical Journal, 16*, 91–104.
Rainer, B., Waltl, M., Cheng, E., Shujau, M., Timmerer, C., Davis, S., et al. (2012). Investigating the impact of sensory effects on the quality of experience and emotional response in web videos. In *2012 fourth international workshop on quality of multimedia experience (QoMEX)* (pp. 278–283). IEEE.
Redi, J. A (2013). Visual quality beyond artifact visibility. In *IS&T/SPIE electronic imaging*. International Society for Optics and Photonics (pp. 86510N–86510N-86511).
Redi, J. A., Hoßfeld, T., Korshunov, P., Mazza, F., Povoa, I., & Keimel, C. (2013). Crowdsourcing-based multimedia subjective evaluations: A case study on image recognizability and aesthetic appeal. In *Proceedings of the 2nd ACM international workshop on crowdsourcing for multimedia* (pp. 29–34). ACM.
Redi, J. A., Zhu, Y., de Ridder, H., & Heynderickx, I. (2015). *How passive image viewers became active multimedia users. Visual signal quality assessment*. Springer.
Scheinkman, J. A. (2008). Social interactions. *The New Palgrave Dictionary of Economics*.
Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence, 10*, 266–281.
Schwarz, H., Marpe, D., & Wiegand, T. (2007). Overview of the scalable video coding extension of the H. 264/AVC standard. *IEEE Transactions on Circuits and Systems for Video Technology, 17*, 1103–1120.
See-To, E. W., Papagiannidis, S., & 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*, 1484–1494.
Serral-Gracià, R., Cerqueira, E., Curado, M., Yannuzzi, M., Monteiro, E., & Masip-Bruin, X. (2010). An overview of quality of experience measurement challenges for video applications in IP networks. In *Wired/Wireless internet communications* (pp. 252–263). Springer.
Staelens, N., Moens, S., Van den Broeck, W., Marien, I., Vermeulen, B., Lambert, P., et al. (2010). Assessing quality of experience of IPTV and video on demand services in real-life environments. *IEEE Transactions on Broadcasting, 56*, 458–466.
Tercek, R. (2011). Trends in Multichannel TV and Online Video in the United States. <http://www.omdc.on.ca/Assets/Research/Research+Reports/Trends+in+Multichannel+TV+and+Online+Video/Trends+in+Multichannel+TV+and+Online+Video_en.pdf>.
Wang, D., Speranza, F., Vincent, A., Martin, T., & 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. In *Visual communications and image processing* (pp. 198–209). International Society for Optics and Photonics.
Wechsung, I., Schulz, M., Engelbrecht, K.-P., Niemann, J., & Möller, S. (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* (pp. 175–186). Springer.
Westerink, J. H., & Roufs, J. A. (1989). Subjective image quality as a function of viewing distance, resolution, and picture size. *SMPTE Journal, 98*, 113–119.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*, 225–240.

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 143–146). ACM.

Wolters, K. M., Engelbrecht, K.-P., Gödde, F., Möller, S., Naumann, A., & 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*, 311–325.

Yamori, K., & Tanaka, Y. (2004). Relation between willingness to pay and guaranteed minimum bandwidth in multiple-priority service. In *The 2004 joint conference of the 10th Asia-Pacific conference on communications, 2004 and the 5th international symposium on multi-dimensional mobile communications proceedings* (pp. 113–117). IEEE.

Zhu, Y., Heynderickx, I., & Redi, J. A. (2014). Alone or together: Measuring users' viewing experience in different social contexts. In *IS&T/SPIE electronic imaging*. International Society for Optics and Photonics, pp. 90140W-90140W-90111.