Impact of Technical and Content Quality on Overall Experience of OTT video

Weiwei Li *, Petros Spachos §, Mark Chignell*, Alberto Leon-Garcia*, Leon Zucherman‡ and Jie Jiang‡
*Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada
§School of Engineering, University of Guelph, Guelph, ON, Canada
‡Technology Strategy and Operations, TELUS Communications Company, Toronto, Canada

***Abstract*—Quality of Experience (QoE) is a crucial guiding factor for network management of an end-to-end service session. The network provider can control the resources allocated to sessions and in doing so, influence the Technical Quality (TQ), which covers the technical aspects of signal quality during the session. On the other hand, the network provider has no control over the Content Quality (CQ), which pertains to the user's level of interest in a particular video. Together TQ and CQ influence the Overall eXperience (OX) in a session. In this paper, we present results from a user subjective study in which the impact of TQ and CQ on OX was investigated for Over-The-Top (OTT) video sessions from the perspective of a network provider. This perspective places a focus on those elements of QoE that can be controlled by the provider. Various studies have shown that very high interest in a content can strongly influence QoE independent of other factors, so our study uses videos that are neutral with respect to content. We assess the TQ, CQ and OX for video sessions that contain Integrity impairments (in the form of image freezing) and failures in terms of session Accessibility and Retainability. Our findings indicate that TQ and CQ have a strong impact on OX in the presence of impairments, but no failures. On the other hand, TQ is the main determinant of OX when failures are present.**

I. Introduction

Mobile communications is experiencing explosive growth, with the proliferation of a new wave of Internet mobile services. These new services introduce new technological challenges. The first challenge involves accommodating growth in data traffic for mobile services in the face of limited available wireless spectrum. The second challenge involves accommodating the expectations for consistent and high quality that customers demand from novel mobile services. In order to meet these challenges and adjust service characteristics appropriately, we need to have accurate estimates of service quality based on customer experience.

Quality of Experience (QoE) has been defined as a subjective measure of a customer's experience with a telecom service. The concept of QoE proposed by QUALINET [1], [2] 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."

The growth in mobile services underscores the importance of QoE for network evaluation, control and management. Overall viewing eXperience (OX) as assessed by QoE, is a key indicator of the level of customer's satisfaction with a service, and hence must be tracked carefully by network and service providers. OX is influenced by Technical Quality (TQ) in the delivery of a service as well as by the level of the user's interest in the content, the Content Quality (CQ), especially in video streaming. For this reason, in this paper we examine the impact of TQ and CQ on OX.

Traditional Quality of Service (QoS) is focused on network performance, and its measurement involves network-centric metrics for service assessment. QoE is a customer-centric measure, and so QoE evaluation requires experimental assessment of the subjective perception of quality from users. Therefore, QoE and QoS are complementary but distinct measures of service quality. Objective QoS metrics are important in assessing network performance. However, network performance cannot directly represent the users' perception of quality about services. Many other factors, such as psychological aspects, end-to-end service processes and context, should be considered in QoE evaluation.

The customer experience with a telecom service is determined by all the interactions of the customer with the service during the entire session. For this reason we have proposed that the scope of measurement of customer experience cover the entire session [3]. Hence the assessment of the user's experience for a service should cover the entire life cycle of a video session, from invocation of the video, through play, and on to the conclusion of the video session. We consider three service components that impact user experience during a session: Accessibility, Retainability, and Integrity.

The rest of this paper is organized as follows: Related work is reviewed in Section II. Session-oriented QoE is described in Section III. Section IV gives presents the experiment design, and results are provided in Section V. Section VI presents conclusions and discusses future work.

II. Related Work

There has been considerable recent research in QoE evaluation, especially in developing QoE models for OTT video applications [4]–[7]. An early indication of the need to assess QoE for an entire session is in [8]. Human factors that influence QoE, such as context, human memory and attention effects, are investigated in [5], [8]. Moorthy et.al. have implemented studies including subjective testing, subjective opinion evaluation and objective algorithm performance analysis [7]. Their QoE evaluation emphasizes the impact of rate adaptation. Pessemier et. al. have proposed a new area for QoE assessment [4]. They have explored the thresholds at which TQ becomes unacceptable for users. They have focused on QoE and the (un)acceptability of rebuffering interruptions.

978-1-4673-9292-1/16/$31.00 ©2016 IEEE

A QoE measurement for mobile, online YouTube video streaming has been proposed in [9]. Their aim was to bridge the gap between objective technical aspects and subjective user-related dimensions impacting QoE. Oyman et.al. consider how to develop performance metrics to accurately evaluate QoE for adaptive streaming services [10]. They find that rebuffering is the single most influential impairment relating to QoE and they used rebuffering percentage to estimate user satisfaction. This provides an objective assessment model for video quality.

Other assessment methods have been discussed for QoE. Machine learning was introduced into the video QoE assessment framework in [11]. Results showed that the percentage of user quality ratings assigned accurately was about 70%. The ITU-T Study Group is active in developing standards for video QoE evaluation [12]–[15]. QUALINET, a European Network on QoE, developed systematical methodologies for QoE on multimedia applications [1], [2].

In our previous research, we have proposed session-oriented QoE in [3] and investigated the impact of Retainability failures on session-oriented QoE in [16], [17]. In [18], we examine the impact of TQ and CQ only on video acceptability. In this paper, we further examine the impact of TQ and CQ on OX.

III. TQ, CQ, and OX in Session-Oriented QoE

Previously, we proposed that QoE for a service is determined by the entire interaction during the session of a customer with a service [3]. Therefore, the scope of customer experience measurement should necessarily cover the whole life cycle of a session. A session-oriented QoE typically includes three components: Accessibility, Retainability and Integrity.

- *Accessibility* is the successful starting of the session.
- *Retainability* is the capability to continue the session (with or without impairments) until its completion, or until it is interrupted by user action.
- *Integrity* refers to the degree to which the session unfolds without excessive impairment. The traditional assessment of QoE considers Integrity only.

The life cycle of an OTT video service from the user perspective has three parts. At the beginning, the user attempts to initiate a session (Accessibility). After a certain delay, the session will either start successfully or fail to start. The latter event is an Accessibility failure. As long as a session starts, it may continue until it naturally stops or may terminate prematurely due to a failure. This event is a Retainability failure. Finally, during the session, all the service-specific impairments impacting customer perception are referred as Integrity impairments.

QoE is determined by many factors including servers, networks and devices, as well as context, emotion of the user, the content type and quality. Our previous research focused on TQ, which denotes all technical features of the signal perceived by the users during the video session [16], [17]. However, it is well known that extreme levels of interest or disgust with content can bias a user's assessment of experience [5], [8], [19]. Therefore, to understand this possible bias, we investigated the impact of TQ and CQ on OX in the research reported in this paper.

IV. Experiment Design

In this section, we present an experiment that collected subjective evaluations about TQ, CQ, and OX.

A. Experiment Overview

Three parts constitute the subjective experiment.

In the first part, subjects were presented with questionnaires called pre-questionnaires. These pre-questionnaires were designed to collect information about the subjects' demographics, video viewing habits, and video quality preferences. Furthermore, the pre-questionnaires assessed subjects' patience levels and personality.

In the second part, subjects were required to watch videos and answered video quality evaluation questionnaires. A training session was used to get them familiar with the procedure, the terminology and the video evaluation. Their responses during the training session were not used in any analysis.

The Absolute Category Rating (ACR) method, as was recommended in [15], was used for video evaluation. In the ACR method, the subject answers a few questions based on the video he/she has just watched. In our experiment, there are four questions for each video. The questions are similar to [4], while in some of them, we increase the scale, following our finding from previous experiments [16], [17]. The questions along with the corresponding scales are shown in Table I.

Question 1 refers to the video acceptability. Question 2 is related to user's perception of TQ, question 3 is the evaluation of CQ and question 4 asked subject's overall viewing experience, OX. Each subject evaluated 30 videos in total, with the study taking about 90 minutes for each participant.

The set of videos used in the experiment was selected to avoid videos that are too interesting or too boring. We had found in previous experiments that extreme levels of interest in a video can strongly influence the assessment of TQ [16], [17]. Our network provider perspective leads us to avoid such videos in our QoE assessment because the provider cannot improve said assessment through control of network resources.

The labels associated with the 5-point scale and the 6-point scales are listed in Table I. We note the TQ and OX use a 6-point scale that adds the category "Terrible" at the bottom. The reason is in our finding in [17] that the traditional 5-point scale is not enough to measure the range of QoE that is

TABLE I. Video Questionnaire And Possible Answers

No.	Rating Criterion/Question	Possible Answers
1	Is the technical quality of this video acceptable?	Yes/No
2	Your evaluation of the technical quality in the video is:	Excellent/Good/Fair/ Poor/Bad/Terrible
3	The content of the video is:	Very interesting/ Interesting/Neutral/ Boring/Very boring
4	Your overall viewing experience (Content + Technical quality) during the video play back is:	Excellent/Good/Fair/ Poor/Bad/Terrible

TABLE II. THE RATING SCALE FOR CQ, TQ, AND OX

CQ		TQ and OX	
Quality	Score	Quality	Score
Very interesting	5	Excellent	5
Interesting	4	Good	4
Neutral	3	Fair	3
Boring	2	Poor	2
Very Boring	1	Bad	1
		Terrible	0

experienced in the presence of Accessibility and Retainability failures. The traditional 5-point scale was employed for CQ. Users responded to assessment questions using the word labels. The numerical values corresponding to word labels shown in Table II were used in the calculation of MOS scores.

B. Experiment Materials

Each video in the experiment had a resolution of 512×288 pixels and a frame-rate of 30 frames/second (fps). The videos were between 80 and 123 seconds in length. The videos consisted of 22 short movie trailers (teaser-trailers) and 8 short movies. Thirty subjects completed the experiment. All subjects were over 18 years of age and had normal or corrected-to-normal vision. Additional details about the experiment can be found in [16].

C. Design of Impairments and Failures

In the evaluation portion of the experiment, subjects watched videos with different types of impairments and failures. The 11 types of impairments and failures that were used are listed in the first column of Table III; the final row shows the one instance of non-accessibility where the video never starts. The set of test videos was generated based on 30 unimpaired videos and using the different types of impairments and failures, as listed in the second column of Table III: Accessibility, Integrity and Retainability. The Integrity video types are prefixed with an "I". Since frame freezing represents the dominant impairment in streaming video, our Integrity impaired videos have zero to three 10-second temporary interruptions during the course of the video session. In previous experiments we had found that the MOS for these impairments cover the range of possible values [16], [17]. The Retainability video types are prefixed with "R". These videos have zero to three temporary interruptions followed by a Retainability failure.

The description of each impairment/ failure is shown in the third column of Table III. The fourth and fifth columns give the corresponding objective parameters (rebuffering numbers and viewing ratio) for each impairment/ failure. The sixth column shows the number of times the particular impairment/ failure was shown in the entire experiment.

V. RESULTS AND ANALYSIS

In this section, we examine the dependence of OX on TQ and CQ as demonstrated by the results in our experiment. The analysis is based on three groupings of videos: i) Unimpaired (I0) videos only; ii) Videos with Integrity impairments (I1-I3); and iii) Videos with Retainability failures (R failures). The

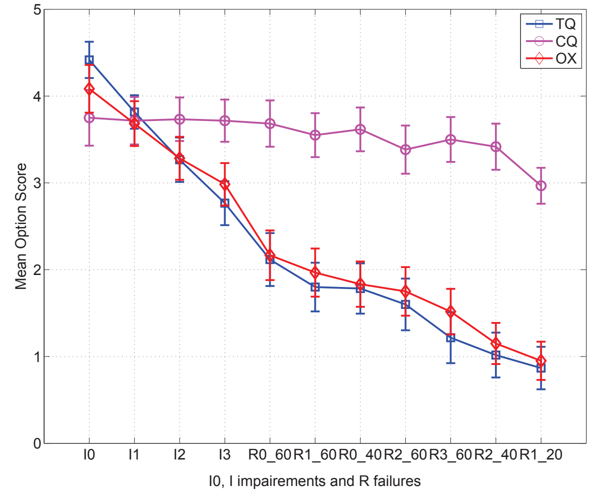

Fig. 1. Mean Opinion Scores (MOS) of TQ, CQ, and OX under *I* impairment and *R* failure.

Accessibility failure was not included in the analysis because CQ cannot be assessed when no video is shown. In total, there were 660 rating samples as listed in Table III.

Figure 1 shows Mean Opinion Scores (MOS) values of TQ, CQ, and OX for the various types of impairments and failures. It is clear that impairments and failures impact the scores of TQ and OX. The ordering of impairments and failures in the abscissa of the graph clearly correspond to progressively lower levels of quality. The mean TQ and OX ratings remain close to each other across all the impairment/ failure conditions in the experiment.

In Fig. 1, the MOS values of CQ remain relatively constant and drop only for retainability and availability failures. This result is consistent with the fact that the set of videos were selected to be neutral, neither too interesting nor too boring. The largest drop in MOS is $R1\_20$, which corresponds to a video, which terminates after only 22% of the clip has been viewed. In order to understand the impact of TQ and CQ on OX, subsequent analysis focused on the following two questions:

Q1. How does the relationship between TQ and CQ vary according to the type of impairments and failures?

Q2. How do TQ and CQ affect overall experience and does this relationship change depending on the presence of impairments or failures?

A. Relationship between CQ and TQ

We calculated the Pearson correlation coefficient between CQ and TQ separately for the three different groupings of impairments and failures. The results in Table IV indicate that CQ and TQ were not significantly correlated within the separate I0, Integrity, and Retainability groupings. However, there is a small, but statistically significant correlation of CQ and TQ over the entire set of videos. This suggests that as lower values of TQ are associated with lower values of CQ.

Impact of impairments/ failures on TQ and CQ. Next we evaluated the Pearson correlation between CQ (or TQ) and

TABLE III. DESCRIPTION OF VIDEO IMPAIRMENTS AND FAILURES

Impairment /Failure set	Category	Description	Rebuffering No.	Viewing Ratio*(%)	Total No. of Ratings from 30 Subjects
I0	Integrity	Unimpaired	0	100	60
I1	Integrity	Single temporary interruption of 10s duration happening at 10s of video content time	1	100	60
I2	Integrity	Two 10s temporary interruptions happening at 20s and 30s of video content time	2	100	60
I3	Integrity	Three 10s temporary interruptions happening at 10s, 20s and 40s of video content time	3	100	60
$R0\_40$	Retainability	One Retainability failure (permanent interruption) happening at 40s of video content time	0	44.44	60
$R0\_60$	Retainability	One Retainability failure (permanent interruption) happening at 60s of video content time	0	66.67	60
$R1\_20$	Retainability	A temporary interruption of 10s duration happening at 10s of the video content time and a Retainability failure happening at 20s of the video content time.	1	22.22	60
$R1\_60$	Retainability	A temporary interruption of 10s duration happening at 10s (of the video content time) and a Retainability failure happening at 60s (of the video content time.)	1	66.67	60
$R2\_40$	Retainability	Two 10s temporary interruptions happening at 20s and 30s and a Retainability failure happening at 40s	2	44.44	60
$R2\_60$	Retainability	Two10s temporary interruptions happening at 20s and 30s and a Retainability failure happening at 60s	2	66.67	60
$R3\_60$	Retainability	Three 10s temporary interruptions happening at 10s, 20s, and 40s and a Retainability failure happening at 60s	3	66.67	60
A	Accessibility	Video never starts. Video player displays a message indicating a failure after 10s temporary interruption at the beginning.	1	0	30

$*$Viewing Ratio = watched content length /average video length(90s)

TABLE IV. PEARSON CORRELATION COEFFICIENT BETWEEN CQ AND TQ

	I0	I impairments	R failures	I0-I3 and R failures
corr(CQ, TQ)	0.0885	0.1080	0.0492	0.1364^*

$^*: p < .05$

objective parameter values associated with Integrity impairments (Rebuffering No.) and Retainability failures (viewing ratio). The first three rows of Table III show the correlations for TQ and Rebufferings or Viewing ratio. The next three rows show the correlations for CQ.

The results in Table V show a relatively strong negative correlation between rebuffering times and TQ when there are I impairments only. In contrast, viewing ratio has a positive correlation with TQ, with more content viewed leading to higher technical quality values. As expected, these results show that TQ increases with less freezing and with greater viewing ratio.

For CQ, Table V shows that CQ has low correlation with rebuffering time. CQ has slightly higher, but still small correlation with viewing ratio. These results show a general insensitivity of the CQ subject ratings to the objective properties of impairments and failures.

Impact of CQ on TQ. Since rebuffering number and viewing ratio were correlated with TQ, we carried out hierarchical regression contrasting the effects of CQ and the objective measures, respectively. Rebuffering number and viewing ratio were entered in the first step of the hierarchical regression, and CQ was added as an additional predictor in the second step. Our goal was to see if CQ would provide additional predictive power, over and above the prediction provided by rebuffering number and viewing ratio.

TABLE V. PEARSON CORRELATION COEFFICIENT BETWEEN OBJECTIVE PARAMETERS AND SUBJECTIVE EVALUATIONS

		Rebuffering No.	Viewing Ratio
TQ	I impairments	-0.4280^*	-
	R failures	-0.2170^*	0.2375^*
	I0-I3 and R failures	-0.1936^*	0.6457^*
CQ	I impairments	0	-
	R failures	-0.0539	0.1590^*
	I0-I3 and R failures	-0.0229	0.1767^*

$^*: p < .05$

The predictor obtained in the first step (with raw regression weights reported here and elsewhere in this paper) was:

$$TQ = -0.0669 - 0.4395 \cdot B + 4.1137 \cdot VR, \qquad (1)$$

where B denotes rebuffering number, and VR is Viewing Ratio. The corresponding goodness of fit was $R^2 = 0.511$.

When CQ was added to the model there was no significant improvement in prediction showing that after accounting for the effects of rebuffering number and viewing ratio, CQ has no significant impact on TQ, in accordance with the experiment design.

Another way of viewing the relationship between CQ and TQ is shown in Fig. 2 where the mean values of CQ and TQ for different impairments and failures are plotted. Clearly, the mean values of TQ are different under different impairments/ failures, while CQ mean values look similar.

One-way Analysis of variance (ANOVA) was conducted to compare the means of CQ under different groups of impairments/ failures as shown in Fig. 2. The null hypothesis

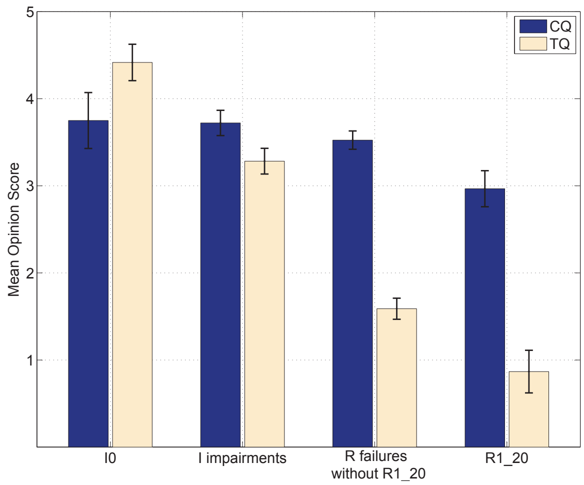

Fig. 2. The mean values of CQ by given case

TABLE VI. PEARSON CORRELATION COEFFICIENT BETWEEN OBJECTIVE PARAMETERS AND OX

		Rebuffering No.	Viewing Ratio
OX	I impairments	-0.2850^*	-
	R failures	-0.1670^*	0.2883^*
	I0-I3 and R failures	-0.1282^*	0.6458^*

is that the CQ ratings for the four groups are drawn from populations with the same mean. The ANOVA gave a small p-value ($p < 0.0001$) which indicates that the differences between groups are significant. Next, post hoc comparisons using the Tukey's range test indicated that CQ with $R1\_20$ was the only group whose mean value is significantly different from the mean values of CQ of the other groups. Therefore, we conducted another one-way ANOVA to compare the means of CQ with I0, I impairments, and R failure without $R1\_20$. The p-value is 0.0594. Hence, it accepts the null hypothesis that all ratings with I0, I impairments and R failures without $R1\_20$ are drawn from populations with the same mean.

B. Dependence of OX on TQ and CQ

Impact of impairments/ failures and CQ on OX. The Pearson correlations between the objective parameters and OX are shown in Table VI. The correlations between these parameters and OX are not as strong as they are in TQ, especially in the I impairments only case.

To see whether the objective parameters provide powerful predictors and whether CQ impacts OX, we first used objective parameters as predictors for OX. Based on all instances from I0-I3 and R failures, we obtained:

$$OX = 0.0738 - 0.3214 \cdot B + 3.7768 \cdot VR, \quad (2)$$

The goodness of fit was $R^2 = 0.475$. Adding CQ, we obtain:

$$OX = -0.8682 - 0.3055 \cdot B + 3.5492 \cdot VR + 0.3047 \cdot CQ, \quad (3)$$

The corresponding goodness of fit increased to $R^2 = 0.523$ which was a small but significant improvement to the fit.

Joint impact of TQ and CQ on OX. To study the joint impact of TQ and CQ on OX, we started by plotting the occurrence matrix for pairs of TQ and OX in Table VII, which demonstrates a linear dependence between TQ and OX.

TABLE VII. OCCURRENCE TIMES UNDER GIVEN TQ AND OX

	OX=0	OX=1	OX=2	OX=3	OX=4	OX=5
TQ= 0	66	29	5	0	0	0
TQ = 1	5	92	34	4	0	0
TQ = 2	0	16	79	37	4	0
TQ = 3	0	4	22	82	26	1
TQ = 4	0	1	6	24	63	8
TQ = 5	0	1	1	6	11	33

Next we calculated the Pearson correlations between CQ and OX, and TQ and OX, respectively, as shown in Table VIII across different impairment/ failure types (columns). It can be seen that the correlations of OX with TQ are generally much stronger than corresponding correlations with CQ, except in the case of I0 where there were no impairments or failures.

We then attempted to characterize the joint impact of TQ and CQ on OX using regression analysis. It can be seen in Table VIII that CQ tends to drive overall experience more when there are no impairments whereas TQ tends to drive overall experience more when there are impairments and/or failures. Given this pattern of results hierarchical regressions was used to determine the relative contributions of the predictors. Analysis was carried out separately for four cases: I0, I impairments only, R failures only, and I0-I3 and R failures.

Case 1 (I0 only): First we use CQ in the predictor with OX as the dependent variable:

$$OX = 1.7717 + 0.6164 \cdot CQ, \quad (4)$$

The regression model had an R^2 value of 0.521.

Adding TQ in the predictor give:

$$OX = -0.8269 + 0.5808 \cdot CQ + 0.6186 \cdot TQ, \quad (5)$$

The goodness of fit of the regression model is improved significantly with $R^2 = 0.741$. Eq. (4) and (5) shows that CQ and TQ have strong effect on OX, in the I0 case. Thus CQ has a strong effect on OX, for the I0 case, and TQ adds significant additional predictive value.

Case 2 (I impairments only): Since both CQ and TQ had fairly strong correlations with OX in this case (Table VIII), we added them both as predictors in a one step regression model with the following equation resulting:

$$OX = -0.6568 + 0.4832 \cdot CQ + 0.6624 \cdot TQ, \quad (6)$$

In this case both predictors were significant, but TQ had a higher weighting in the model (consistent with its higher correlation with OX). The goodness of fit for the model was $R^2 = 0.730$.

Case 3 (R failures only): Based on the results shown in Table VIII, we input TQ as the predictor in first step:

$$OX = 0.4890 + 0.7606 \cdot TQ, \quad (7)$$

with $R^2 = 0.668$.

TABLE VIII. PEARSON CORRELATION COEFFICIENT BETWEEN TQ AND OX, AND CQ AND OX

	I0	I impairments	R failures	I0-I3 and R failures
$corr(TQ, OX)$	0.5311*	0.7129*	0.8170*	0.8732*
$corr(CQ, OX)$	0.7216*	0.5455*	0.2018*	0.3421*

When we added CQ in the second step there was only a minor (but statistically significant) improvement in prediction with R^2 rising slightly to 0.694. The corresponding regression model was dominated by TQ.

$$OX = -0.1019 + 0.1747 \cdot CQ + 0.7532 \cdot TQ, \quad (8)$$

Thus in the presence of retainability failures, OX was mainly determined by TQ.

Case 4 (I0-I3 and R failures): In this last case (with data pooled across all the impairment and failure conditions), we first used TQ as the predictor for OX:

$$OX = 0.4790 + 0.8148 \cdot TQ, \quad (9)$$

with $R^2 = 0.763$.

Adding CQ to the predictor gives:

$$OX = -0.5590 + 0.3108 \cdot CQ + 0.7859 \cdot TQ, \quad (10)$$

with R^2 increasing significantly to 0.813.

The above four cases show that OX is dependent on both CQ and TQ. However, the dependence changes depending on the type of impairment/ failure. For pristine videos (Case 1), CQ and TQ both have a strong impact on OX. Similarly, when only I impairments are present (Case 2), then both CQ and TQ have an impact on the value of OX. When R failures are present (Cases 3 and 4), then TQ becomes the much stronger predictor of OX.

VI. CONCLUSIONS

We have examined the impacts of technical quality and content quality on overall experience, under different impairment/ failure conditions. In particular we considered the relative strength of TQ and CQ in predicting OX. In the absence of failures, CQ and TQ are both determinants of OX. When retainability failures are added, TQ is the overwhelming determinant of OX.

Our study intentionally selected video material to control the interest levels to be neither too high nor too low. Our motivation was that the network provider is limited to addressing issues that impact technical quality, but it has no recourse to address content quality since the content is provided by a third party. Therefore our findings on the relative impact of TQ and CQ on OX are valid for this restricted set of videos. In future work we will investigate the dependence of OX on TQ and CQ for videos spanning a broader range of interest levels.

ACKNOWLEDGMENT

This research was supported by a grant from TELUS and a matching grant from NSERC/CRD.

REFERENCES

[1] A. Raake and S. Egger, "Quality and quality of experience," in *Quality of Experience: Advanced Concepts, Applications and Methods*, S. Moeller and A. Raake, Eds. Springer, 2014, ch. 2.

[2] I. Wechsung and K. D. Moor, "Quality of experience versus user experience," in *Quality of Experience: Advanced Concepts, Applications and Methods*, S. Moeller and A. Raake, Eds. Springer, 2014, ch. 3.

[3] A. Leon-Garcia and L. Zucherman, "Generalizing mos to assess technical quality for end-to-end telecom session," in *Globecom Workshops (GC Wkshps), 2014*, Dec. 2014, pp. 681–687.

[4] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, "Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching," *IEEE Transactions on Broadcasting*, vol. 59, no. 1, pp. 47–61, March 2013.

[5] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the impact of video quality on user engagement," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 362–373, Aug. 2011.

[6] R. Mok, E. Chan, and R. Chang, "Measuring the quality of experience of HTTP video streaming," *2011 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pp. 485–492, May 2011.

[7] A. Moorthy, L. K. Choi, A. Bovik, and G. De Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652–671, Oct. 2012.

[8] M. Söderlund, "Behind the satisfaction facade: An exploration of customer frustration," in *32nd European Marketing Academy Conference*, 2003.

[9] I. Ketykó, K. De Moor, T. De Pessemier, A. J. Verdejo, K. Vanhecke, W. Joseph, L. Martens, and L. de Marez, "QoE measurement of mobile YouTube video streaming," *Proceedings of the 3rd Workshop on Mobile Video Delivery*, pp. 27–32, 2010.

[10] O. Oyman and S. Singh, "Quality of experience for HTTP adaptive streaming services," *IEEE Communications Magazine,*, vol. 50, no. 4, pp. 20–27, April 2012.

[11] A. Dalal, A. Bouchard, S. Cantor, Y. Guo, and A. Johnson, "Assessing QoE of on-demand TCP video streams in real time," *2012 IEEE International Conference on Communications (ICC)*, pp. 1165–1170, June 2012.

[12] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation BT. 500-13, Recommendations of the ITU, Radiocommunicaion Sector*, 2012.

[13] ITU-T, "Telephone transmission quality, telephone installations, local line networks," *Recommendation Series P, Telecommuniation standardization sector of ITU*, p. P.10/G.100 (2006) /Amd.1, Jan. 2007.

[14] ITU-T, "Quality of experience requirements for IPTV services," *Recommendation G.1080, Telecommunication standardization sector of ITU*, Oct. 2009.

[15] ITU-T, "Subjective video quality assessment methods for multimedia applications," *Recommendation P.910, Telecommuniation standardization sector of ITU*, Sep, 2009.

[16] W. Li, H.-U. Rehman, M. Chignell, A. Leon-Garcia, J. Jiang, and L. Zucherman, "Impact of retainability failures on video quality of experience," in *1st Workshop on Quality of Multimedia Services at SITIS'14*, Morocco, Dec. 2014.

[17] W. Li, H.-U. Rehman, M. Chignell, A. Leon-Garcia, J. Jiang, and L. Zucherman, "Video quality of experience in the presence of accessibility and retainability failures," in *10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*, Greece, Aug. 2014.

[18] P. Spachos, W. Li, M. Chignell, A. Leon-Garcia, J. Jiang, and L. Zucherman, "Acceptability and Quality of Experience in Over The Top Video," in *IEEE International Conference on Communications Workshops (ICC), 2015*, Jun. 2015.

[19] M. Mirkovic, P. Vrgovic, D. Culibrk, D. Stefanovic, and A. Anderla, "Evaluating the role of content in subjective video quality assessment," in *The Scientific World Journal*, vol. 2014, 2014.