Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 111 (2017) 95–104

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Conference on Advances in Information Technology, IAIT2016, 19-22 December 2016, Macau, China

A case study in visualizing disruptions to service quality

Mark Chignell[a,*], Ryan Kealey[a], Chelsea DeGuzman[a], Leon Zucherman[a], Jie Jiang[b]

[a]*Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada*
[b]*TELUS Communications Company, Scarborough, Canada*

Abstract

This study measured Session Quality of Experience (SQE) by letting people look at visual proxies of a video play bar without having to watch the actual video, with its associated impairments and/or failures. We supplemented a laboratory study of 10 participants with two further online samples of 14, and 20, participants respectively, obtained using online surveys. Cluster analysis was carried out on the entire data set of 44 participants and clusters were obtained representing differences in sensitivity to pricing and to types of disruption. Crucially, there was no significant dependency between samples and cluster demonstrating that visualization can be used in a crowdsourcing, as well as a laboratory, context. This study demonstrates the combined use of crowdsourcing, visualizations of service experience, and cluster analysis to create more efficient, and nuanced, interpretations of SQE.

Peer-review under responsibility of the organizing committee of the 8th International Conference on Advances in Information Technology

Keywords: Quality of experience; crowdsoucring; visualized experience; human factors; clustering; paired comparisons; online video.

1. Introduction

Evaluation of online video Session Quality of Experience (SQE) using ratings collected after watching videos can be difficult and time consuming. A typical experiment might involve a group of representative people watching online videos that vary in the frequency, type, and timing of disruptions (e.g., a video that freezes for a few seconds

* Corresponding author. Tel.: +1 (647) 389-8951
E-mail address: chignell@mie.utoronto.ca

1877-0509
Peer-review under responsibility of the organizing committee of the 8th International Conference on Advances in Information Technology
10.1016/j.procs.2017.06.015

and then starts playing again) or failures (instances where a video fails to start playing or fails to play all the way through to completion). SQE research could progress faster if more efficient, but still valid, methods of data collection were available.

An innovative method was developed for visualizing video impairments and failures as marked-up regions indicating when the video didn't play, on a video play bar[1]. That research found that visualization methodology could be used to get ratings of Technical Quality (TQ) that were statistically indistinguishable from corresponding ratings made after having actually watched video with the corresponding disruptions (rather than making the rating based on a visualization). In addition to providing greater efficiency in collecting data relating to the TQ of online video, the method of TQ rating based on the visualized playbar also has the advantage that it is unaffected by the content quality of the video, or other factors such as the resolution or bandwidth. A further advantage that we exploit in the research reported below is that, with visualizations, remote participants can make TQ ratings using an online survey platform. The case study reported in this paper examines whether crowdsourced online ratings of TQ are similar to corresponding ratings collected within a university research laboratory under more controlled conditions. The study will also examine whether TQ ratings based on visualized (rather than actual) experience can provide reasonable data even when there is no training other than brief instructions of how disruptions to video are represented in the visualized playbars.

The main objective of the current research was to build on our earlier finding that participants are able to imagine or visualize what they would have experienced if they had actually seen the impairments and failures indicated in a play bar visualization[1]. We are specifically interested in how robust this effect is, and if it can be made even more efficient by collecting data using online surveys rather than laboratory studies. Thus the research questions that we addressed in the study reported in this paper were: (a) How much training do people need in order to provide reliable ratings of TQ experience, and (b) do visualization studies have to be carried out in the lab, or can they be done online?

After briefly reviewing relevant research literature we describe and report on the study, where a complete set of 66 possible pairwise combinations of 12 stimuli (four combinations of video impairments/failures crossed with three payment brands, i.e., pay-per-view, monthly paid service, free to view) were rated in terms of preference. Data for 44 participants making the 66 pairwise preference comparisons were collected across three different samples (in-person, online without screening, online with screening). Mean preference ratings for each stimulus within each participant were then subjected to feature extraction, cluster analysis, and analysis of variance to assess the impact of payment, and disruption, type on preferences for different scenarios. The results were also interpreted in terms of the reliability of online vs. in person data collection for measuring quality of experience through visualizations of disrupted video.

1.1. The visualization method

A previous study[1] provided an initial validation of the use of the visualization method as a proxy for actually viewing videos when rating TQ of the video experience. The study had two conditions (experimental and control), and participants in each condition completed three sessions. In Session 1, participants watched actual video clips and after each clip were asked to provide a Technical Quality (TQ) rating for that video, on a scale from 1 (Bad) to 5 (Excellent). In the experimental condition, a play bar was visible underneath the video, which showed the visualization of successful video playback, and disruptions, as they occurred. For example, as the video was playing, the red section would extend across the play bar (similar to how video duration is indicated on YouTube) and when there was a disruption, a grey section would appear and extend across the bar until the video started playing again (at which point the red section would resume). There was no playbar in the control condition.

Like Session 1, Session 2 also required participants to watch actual video and provide TQ ratings, but in Session 2, neither group saw the play bar. Additionally, after rating the TQ for each video, participants were also required to generate a play bar that represented the video they had just watched, using a software tool specially developed for this task. In the experimental condition, participants received feedback when generating visualizations, whereas participants in the control condition did not receive feedback. In Session 3, participants did not watch any videos – they rated TQ based on visualizations only.