

Towards Automated Subjective Quality Assessments

Maximilian Krumholz¹, Benjamin Rainer¹, and Christian Timmerer^{1,2}

¹Institute of Information Technology (ITEC)

Alpen-Adria-Universität Klagenfurt, Universitätsstrasse 65-67, 9020 Klagenfurt

Email: {*firstname.lastname*}@itec.aau.at

²Bitmovin GmbH

Schleppe Platz 7, 9020 Klagenfurt

Email: christian.timmerer@bitmovin.com

Abstract—In this paper we describe a first step towards the automation of subjective quality assessments. Current frameworks require a manual configuration for each individual evaluation which is only suitable for experts with a programming background. In order to overcome this limitation we propose a Web-based management framework to setup and conduct subjective quality assessments by utilizing open source Web tools and frameworks.

Keywords—*Quality of Experience; Subjective Quality Assessment; Crowdsourcing; Web tools and frameworks*

I. INTRODUCTION

Subjective quality assessments (SQAs) are used as a vital tool within the Quality of Experience (QoE) research community and beyond. It provides reliable results but are considered as cost-intensive when conducted within controlled laboratory environments. Crowdsourcing has been proposed as an alternative to reduce costs but various aspects need to be considered in order to get the same reliable results as within laboratory environments [1]. In the past, various frameworks have been proposed leveraging crowdsourcing platforms to conduct SQAs which come with a different feature set and functionalities [2]. However, a common shortcoming of these frameworks is that they require manual configuration and setup for each SQA which makes it difficult to use. Therefore, we propose a Web-based management framework that shall i) enable easy and simple configuration of SQAs, ii) provide means to conduct SQAs using existing crowdsourcing platforms, and iii) allow for evaluating the results.

In this paper we describe *WESP+* which is based on the **WE**b based **S**ubjective evaluation **P**latform (**WESP**) [3] and utilizes Web frameworks and several tools to enable this kind of functionality. The goal of our approach is to provide the possibility to select a preferred evaluation method via a Web interface. Moreover, we aim for a framework that fulfils all ITU recommendations for subjective evaluations of multimedia applications and television pictures [4], [5]. The framework will be available at [6].

II. DESIGN AND ARCHITECTURE

WESP+ adopts the open source Symfony PHP framework v2.8 as a basis. Symfony is a PHP Framework that follows the Model-View-Controller (MVC) pattern. It comes with the Doctrine ORM database bundle v2.5.2 which we used as object-relational mapper for the MySQL database. We

also used two popular bundles called *FOSUserBundle* and *EasyAdminBundle* v1.11.8. The *FOSUserBundle* adds support for a database-backend user system and the *EasyAdminBundle* provides means to create administration backends.

The principal design of the desired functionality is shown in Figure 1 with the following major objectives: a) select or add the content which should be evaluated, b) select the preferred ITU recommendation, c) add pre- and/or post-questionnaires, d) select the metrics to be collected during the evaluation, e) run the evaluation using a set of users – from a given crowdsourcing platform –, and f) save the results for offline analysis. We deliberately excluded the evaluation of the results from *WESP+* as there are many (static) tools which could be used for this purpose much better and more efficiently.

The main goal was to develop a platform, that can easily be adopted for every evaluation purpose.

A. Select/Add Content

WESP+ allows the evaluation host to select existing or add new content to a database of test sequences. In principle, it should also be possible to add or connect to other databased such as [7]. Currently, two types of content are supported, i.e., images and videos.

B. Evaluation Methodology

Users of the platform can select the preferred evaluation methodology according to existing ITU recommendations and currently we support those defined in BT.500 [4] and P.910 [5]. Depending on the selection of the evaluation methodology, the rendering of the content and evaluation will take place.

C. Questions

WESP+ supports the possibility to add questions and related answers or answer sets. Prior defined questions can be reused and new ones can be added. Newly added questions can have a type and a group assigned to them. Type options are related to most of the HTML5 input types like radio-box, selection-box, drop-down, simple text or date form elements. Additionally, a slider element is available to support range related questionnaires.

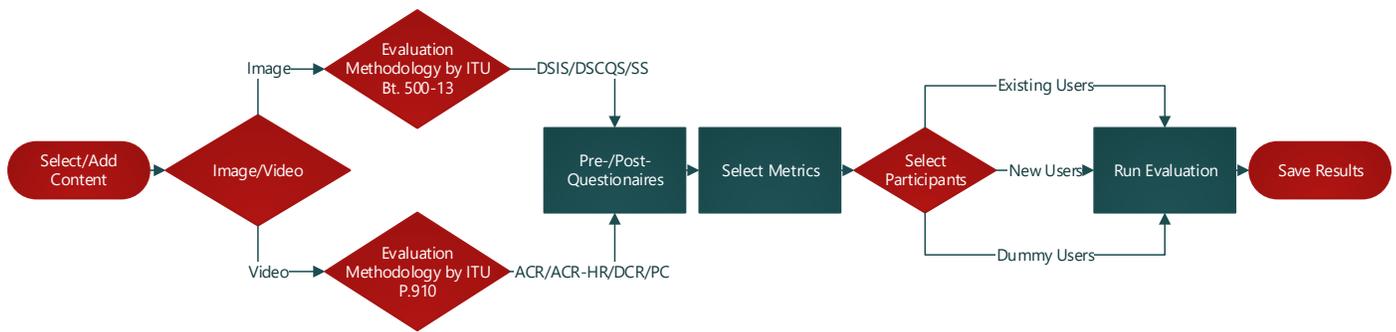


Fig. 1. Principle High-Level Design and Workflow of WESP+.

D. Metrics

After defining questions, WESP+ provides the possibility to collect content-related metrics during the evaluation. Currently, we support the following set of metrics: startup time, stalls, pauses, buffer state, representation bitrate and guessed bandwidth (the main focus is on adaptive streaming but it could be also used for other purposes). Note that the usage of user-defined metrics are subject to future work. Furthermore, it is possible to check the behavior of the participants by logging the browsers focus, browsers fingerprint, and full screen state.

E. Participants

There are options to add already existing participants from the database, provide the possibility to let participants create accounts or choose the option to create dummy users for every new participant. New users can add information about their external accounts from, e.g., crowdsourcing platforms, which allows the evaluation host to approve their work at the corresponding platform.

F. Run Evaluation

The selected users get the related content rendered as described in the ITU recommendations. For example, ACR [5] will render sequences subsequently, asking the user to rate each sequence independently within a given time frame, while DCR [5] will show two sequences before asking the user to rate the impairment of the second stimulus in relation to the first. For the moment, evaluations that require simultaneous, synchronized playback of multiple sequences must be prepared in one video stream. Another use case would be SSCQE [5] where a slider next to the stimulus provides continuous evaluation during the stimulus playback.

G. Save Results

Results will be stored in the MySQL database according to the evaluation, the content, each user's result, and the

collected metrics. Additionally, they can be exported into different formats which allows for offline analysis utilizing external tools.

III. CONCLUSIONS AND FUTURE WORK

In this paper we have briefly described a Web-based management framework to setup and conduct subjective quality assessments. It overcomes limitations of existing approaches by utilizing open source Web tools and frameworks. In the future, we would like to add support for a preliminary interpretation of the collected results using various graphs. Additionally, we plan to add color blindness tests and the support for APIs to directly connect to several crowdsourcing platforms like Microworkers.

REFERENCES

- [1] T. Hoßfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best Practices for QoE Crowdttesting: QoE Assessment With Crowdsourcing," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 541–558, Feb 2014.
- [2] T. Hoßfeld, M. Hirth, P. Korshunov, P. Hanhart, B. Gardlo, C. Keimel, and C. Timmerer, "Survey of Web-based Crowdsourcing Frameworks for Subjective Quality Assessment," in *2014 IEEE International Workshop on Multimedia Signal Processing (MMSP 2014)*, S. Rahardja, Z. Zhang, F. Pereira, and A. Loui, Eds. Piscataway, N.J. 08854, U.S.A.: IEEE, sep 2014, p. 6.
- [3] B. Rainer, M. Watzl, and C. Timmerer, "A Web based Subjective Evaluation Platform," in *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*. IEEE, 2013, pp. 24–25.
- [4] I.-R. Recommendation, "BT. 500-13 (2012-01). Methodology for the subjective assessment of the quality of television pictures. ITU, 2012," 2014.
- [5] I.-T. Recommendation, "P. 910: Subjective video quality assessment methods for multimedia applications," *International Telecommunication Union, Geneva*, 2008.
- [6] "Sensory Experience Lab," <http://selab.itec.aau.at>.
- [7] K. Fliegel and C. Timmerer, "WG4 Databases White Paper v1.5: QUALINET Multimedia Database enabling QoE Evaluations and Benchmarking," *Prague/Klagenfurt, Czech Republic/Austria, Version 1.5*, 2013.