AVRate: An open source modular Audio/Visual subjective evaluation test interface

Pierre Lebreton[1], Marie-Neige Garcia[2], Alexander Raake[1,2]
[1]Audio Visual Technology - Institute of Media Technology, Ilmenau University of Technology
Helmhotzplatz 2, 98693 Ilmenau, Germany
[2]Assessment of IP-Based Applications, Telekom Innovation Laboratories, TU Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, Germany

***Abstract*—This paper presents an open source modular tool for performing audio, video and audiovisual user tests. The tool can be used with a large variety of video players. It covers standardized test methods and rating scales. The rating scale characteristics can easily be configured by editing a simple XML file, therefore enabling researchers to focus on the test methods and stimulus generation instead of the implementation of the test interface.**

I. Introduction

In this paper a modular tool enabling to collect subjective data is presented. The process of conducting a subjective experiment is well defined in ITU-R and ITU-T Recommendations (-R) BT.500, (-T) P.910, (-T) P.913. In these recommendations various subjective assessment methods such as Absolute Category Rating (ACR) with or without reference, Degradation Category Rating (DCR), Pair Comparison, Double Stimulus Comparison Scale (DSCS), Subjective Assessment Methodology for Video Quality (SAMVIQ), etc. are described. In each methodology, a well-defined flow of stimulus and rating scale presentation needs to be followed. It is a familiar task to researchers in QoE to define a test interface that handles the playback of the sequences in the right order, lets the participants choose between different alternatives such as in the SAMVIQ methodology, possibly lets participants repeat videos, and finally presents the rating scales with the preferred characteristics: discrete, continuous, forced choice. To enable researchers to focus on the fundamental question of methods and design of test stimuli and let the implementation of the test interface aside, this paper provides a customizable open-source tool to conduct subjective experiments. Using this tool, it is possible to apply the previously mentioned methods (and others) without any need to adjust a line of code. The only task is to configure the test interface in a comprehensive setting file. In the next section existing solutions will be presented and the motivation of the proposed software will be explained. Section 3 provides examples of configurations for different rating/evaluation tasks. The various features of the tool will be described in Section 4 before concluding in Section 5.

II. Related work and motivation

Very few software solutions enabling subjective evaluation tasks have been released to the scientific community. One of them is the MSU Perceptual Video Quality tool [1] which supports the following testing methodologies: Double Stimulus Impairment Scale (DSIS), Double Stimulus Continuous Quality Scale (DSCQS), Stimulus Comparison Adjectival Categorical Judgement (SCACJ), SAMVIQ and MSU Continuous Quality Evaluation (MSUCQE). One of the key limitations of this tool is its video player. Indeed, the MSU uses its own video player to play back the videos. This was found impractical as the player did not always reach the required performance, yielding jerky playback with high resolution videos. In addition, it requires a specific video format to playback the video. Providing the ability to use any kind of player such as the VideoLAN media player ("VLC" [1]), "mplayer" [2], "stereoscopic player"[3], or professional server-based video players was therefore needed. Having personalized players also enables the experimenter to record statistics about the playback performance such as unexpected/unwanted freeze or frame-drop events. Thus AVRate was designed to work with any video player. A second main advantage is to let the experimenter select the question(s) to be asked to the test participants and the characteristics of the rating scale(s). It may also be required to use multiple scales such as for comfort, quality, degree of immersion or the like. An issue frequently observed with continuous rating scales based on sliders is the default position of the cursor on the continuous line: it should not have a default value as it may affect the user's decision.

The Video Expert Quality Group (VQEG) provides different tools to perform subjective evaluation tasks. The AcrVQWin [2] designed by ACREO enables to perform ACR tests with user-defined questions and labels. The rating interface is currently limited to a single five-point grade scale, and the playback of the video is performed by an embed player. The randomization of the sequences can be performed automatically, and also supports pauses in the middle of the test. The VQEGplayer [3] is an evolution of AcrVQWin. It adds support for the Paired Comparison method. It also benefits of a precise monitoring of the video playback performed by the embed player and adds support for 3D. It is limited to the 5-point ACR-scale or Paired Comparisons in standalone mode, but this limitation can be solved using the "remote" mode, and the test interface can be deported to a distant computer running MATLAB. Compared to these tools "AVRate" provides more flexibility in terms of scales definition without the dependency to MATLAB. Finally, many experiments have been using script languages such as MATLAB or Python, or compiled languages such as Java to design test interface. This requires

[1]http://www.videolan.org/vlc/
[2]http://mplayerhq.hu/
[3]http://www.3dtv.at/products/player

programming skills, adjusting code for each experiment, and may eventually result in bugs. It is then proposed to only adjust an XML configuration file enabling to test different aspects without requiring modification of the code. In the next section, various examples of configuration will be presented illustrating the features of AVRate.

III. Configuration Examples

All parameters described in the following can be found in a single XML file which is used to configure the evaluation tool. One of the key points addressed previously is the ability to choose the video player. To set this option, a parameter "localplayer" is defined. Different players are supported by default: windows media player, mplayer, VLC and stereoscopic player. Alternatively, it is possible to use any kind of command-line-based video player. Listing 1 illustrates how such player is called with a specific argument. "%1" is a placeholder for the path to the video to be played.

Listing 1: "Selection of a video player"

```
<localplayer>command_line</localplayer>
<playerpath>C:\some_player.exe</playerpath>
<playerarguments>-i %1</playerarguments>
```

The design of the test interface is flexible. Two types of input are possible: groups of buttons letting the test participant choose between different options, or sliders letting participants choose a position on a scale with discrete or continuous values. For example, Listing 2 illustrates two different scales: one is a continuous scale with labels and the other one is a forced choice between two options ("yes" or "no"). Figure 1 depicts the rendering of this configuration. The number of scales is not limited. It can be noted that the slider defines an option "fixed" (see Listing 2). If set to true, the slider becomes a discrete scale. In that case, the test participant can only click on the exact position of the ticks.

Listing 2: "Definition of the scales"

```
<slider>
  <name>Please rate the quality of the picture</
      name>
  <label>Bad</label>
  <label>Poor</label>
  <label>Fair</label>
  <label>Good</label>
  <label>Excellent</label>
  <min>0</min>
  <max>10</max>
  <ticks>11</ticks>
  <fixed>false</fixed>
  <shownumbers>true</shownumbers>
</slider>
<buttons>
  <name>Is it acceptable?</name>
  <label>Yes</label>
  <label>No</label>
</buttons>
```

With the proposed tool, video sequences are played one at a time and following the order defined in a playlist file.

To perform a double stimulus test such as DCR, for which two video sequences are sequentially shown to the test participant, the playlist is adjusted as shown in Listing 3. This results in the test procedure shown in Fig. 2 and for which the test participant provides a rating after viewing two video sequences in a row.

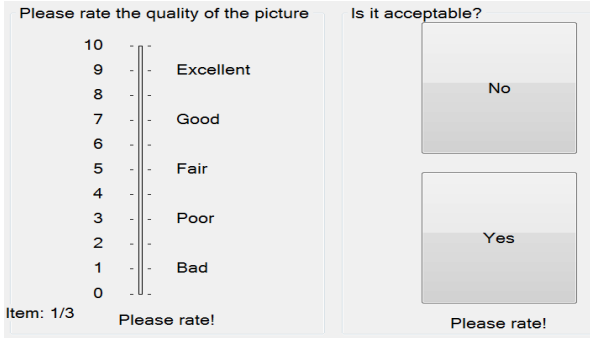

Fig. 1: Example of rating interface

Listing 3: "Example of playlist file for double stimulus"

```
video1_A.avi video1_B.avi
video2_A.avi video2_B.avi
video3_A.avi video3_B.avi
```

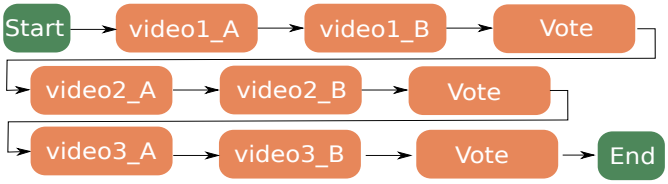

Fig. 2: Flow of a subjective experiment with double stimulus

IV. Capabilities

As previously mentioned, the test methods ACR, DCR, Paired Comparison, Two or Three Alternative Forced Choice Paired Comparison, and SAMVIQ can be directly configured in AVRate. The tool also supports hardware sliders as input instead of the graphical interface.

V. Conclusion

This paper presents an overview of an open-source subjective test software with its flexible- and easy-configurable features. The code and builds of the application can be found on GitHub[4] and DOI[5] under the license of GNU GPL v3.

References

[1] MSU Graphics Media Lab (Video Group), "MSU perceptual video quality tool, program for counting mean opinion score," http://www.compression.ru/video/quality_measure/perceptual_video_quality_tool_en.html, [Online; accessed 20-april-2016].

[2] J. Jonsson and K. Brunnström, "AcrVQWin (version 1.0)," https://www.acreo.se/acrvqwin, 2007, [Online; accessed 20-april-2016].

[3] K. Brunnström, R. Cousseau, J. Jonsson, Y. Koudota, V. Bagazov, and M. Barkowsky, "VQEGPlayer," http://vqegjeg.intec.ugent.be/wiki/index.php/VQEGplayer-main, [Online; accessed 20-april-2016].

[4] https://github.com/Telecommunication-Telemedia-Assessment/AVRate
[5] https://zenodo.org/record/51403