MIRO360: A Tool for Subjective Assessment of 360 Degree Video for ITU-T P.360-VR

Pablo Pérez
Nokia Bell Labs, Madrid, Spain
pablo.perez@nokia-bell-labs.com

Javier Escobar
Universidad Politécnica de Madrid, Spain
jer@gti.ssr.upm.es

***Abstract*—This paper describes MIRO360, a tool which implements the diverse set of methodologies that are being evaluated by ITU-T Study Group 12 and the Video Quality Experts Group in the development of the future ITU-T Recommendation P.360-VR for the subjective assessment of 360 video with Head Mounted Displays (HMDs). MIRO360 is an application for Android-based HMDs, with a flexible configuration that allows the simple creation of tests for short sequences (either single or double reference), as well as for long ones, including in-sequence continuous or discrete evaluation, as well as simulator sickness and spatial presence.**

***Index Terms*—360 video, subjective assessment, Quality of Experience**

I. Introduction

With the recent popularization of Virtual Reality (VR) devices and applications, the research on the Quality of Experience of VR systems and, in particular, of 360 video, has also become increasingly common. However, there is still a lack of standardized methodologies for the subjective assessment of VR applications or 360 video. For the latter, most researchers basically adapt existing recommendations for 2D video (e.g. ITU-T P.910 [1]) in the way they consider more appropriate.

Recently, ITU-T Study Group 12 and the Video Quality Experts Group (VQEG) have started a joint work to develop a new Recommendation for the subjective assessment of 360 video in HMDs, provisionally named P.360-VR [2]. The work plan implies a set of subjective assessment tests to validate the methodology, for which a testing interface is required.

Some tools are already available for subjective assessment tests in VR. For instance, AVTrack360, which records head tracking for desktop environment [3], or VRate, which provides an asset to integrate subjective assessment in Unity3d scenarios [4]. However, none of them, to the authors' knowledge, is a full-fledged application with all the functionality supported by the current test plan for P.360-VR and, in particular, none implements in-sequence voting methodologies.

To overcome these limitations we have developed MIRO360[1], an Android Virtual Reality application to assess subjective quality of 360 video following the methodologies under study for P.360-VR.

This work has been partially funded by project VINEDO IDI-20180015, from Spain's Centre for the Development of Industrial Technology (CDTI).

[1]Software can be accessed at https://git.gti.ssr.upm.es/jer/miro360 (user: "Reviewer", password: "qOmex2k19"). Select "standard user" for access. Code will be made public upon paper acceptance.

II. Features

The application has been developed for Android using Samsung Gear VR framework, and therefore it is prepared to run on a cell phone attached to an HMD. Android-based standalone HMDs can be used as well. The scoring process is performed with a gaze cursor: the user has to fix the gaze in the specific response, and then click the main button of the remote control. This click can also be done by a single tap on the touchpad included on Samsung VR HMD, in case the remote control is not available.

The user experience is simple. On start-up, a background picture is displayed to the user, with a text box displaying "click to start", to prevent the test from starting before the user is wearing the full setup (HMD, headphones, etc.). The test is then composed of a series of video sequences, each one admitting in-sequence and/or post-sequence assessment. Once all the sequences have been played, and their assessment questions answered, the test finishes, which is also indicated to the user with a text box.

A. In-sequence assessment

The tool supports performing assessment tests during the playout of the sequences. Two different methods are provided: Single-Stimulus Continuous Quality Evaluation (SSCQE) and Single-Stimulus Discrete Quality Evaluation (SSDQE).

SSCQE implementation is based on the Recommendation ITU-R BT.500 [5]. The quality is assessed in a continuous scale, and the user can vary the assessed quality level at any time. ITU-R BT.500 proposes using a desk-mounted 10 cm physical slider. As this would limit the user motion, which is needed in immersive video, MIRO360 implements a virtual slider instead (Fig. 1.a): the current level of quality is displayed numerically in front of the user, in an integer scale from 00 to 99. The user can vary the quality level with a scrolling gesture on the tactile interface of the remote control or, alternatively, on the HMD touch panel.

Besides the numerical quality value, the corresponding categorical word is also displayed (e.g. *Bad* for 0 to 19, *Poor* for 20 to 39, and so on). Absolute Category Rating (ACR) is used by default, but others can be selected: Degradation Category Rating (DCR) and two different cybersickness scales [6], [7]. Custom 5-level categorical scales can also be added.

As an alternative to SSCQE, MIRO360 also implements the methodology proposed in [8], which has been labeled

(a) (b) (c)

Fig. 1. Screenshots of MIRO360: a) SSCQE, b) SSDQE, c) post-sequence questions.

TABLE I
MINI-MEC-SPQ QUESTIONNAIRE

Factor	Question
Attention Allocation	I devoted my whole attention to the video.
Spatial Situation	I was able to imagine the arrangement of the spaces presented in the video very well.
Self-Location	I felt like I was actually there in the environment of the presentation.
Possible Actions	I had the impression that I could be active in the environment of the presentation.
Cognitive Involvement	The video presentation activated my thinking.
Suspension of Disbelief	I didn't really pay attention to the existence of errors or inconsistencies in the video.

as Single Stimulus Discrete Quality Evaluation (SSDQE) in the context of this tool[2]. In this methodology, instead of continuously recording the quality level, it is periodically requested to the user while the video keeps playing on the background (Fig. 1.b). The video sequence is divided in periods of potentially impaired video (the Processed Video Sequence, PVS, under test), followed by (shorter) periods where the user is requested to rate the quality of the PVS, while the content keeps playing in the background. SSDQE supports the same categorical scales already mentioned for SSCQE, including user-defined ones. In the tool, the length of the PVS and the voting periods are configurable.

Either SSCQE and SSDQE, but not both, can be configured individually for each sequence. It is also possible not to use any in-sequence assessment method. Besides, head orientation (unit vector of the gaze direction) is recorded each time the HMD renders a frame.

B. Post-sequence assessment

After each sequence, zero, one or several questions can be asked to the users. There is full flexibility with the number of questions, to cover scenarios such as standard ACR test (one ACR question after each sequence), DCR (no question after reference, DCR question after PVS), or multi-item questionnaire after a longer sequence (e.g. to assess presence). Questions are displayed in front of a static picture as background. There is no time limit to answer them.

Several scales are included in the application: ACR, DCR, two simulator sicknes scales (Vertigo [6] and Dizziness [7]), and two 8-item presence scales: Spatial-Presence Experience Scale (SPES) [9] and mini-MEC-SPQ. Mini-MEC-SPQ is an aggressive subsampling of the MEC Spatial Presence Scale (MEC-SPQ) [10], using a single item for each of the 8 measured factors of Spatial Presence, instead of 4, 6 or 8 as in the original publication. The order of the questions in these multi-time questionnaires is randomized for each sequence. It is also possible to configure custom single and multiple item scales.

C. Configuration

The application is configured by a JSON playlist as shown in Fig. 2. Each sequence can have an independent URI, which can be any file playable by Android MediaPlayer. In-sequence and post-sequence methodologies can be individually selected for each sequence. More details on the potential parameters and their values can be found in the software documentation.

D. Output format

Scores are stored on a local file on the phone in csv format. Each line of the csv file represents one event and includes: timestamp, uri, PVS number, test section ("idle", "video", or "questionnaire", depending on what it is being displayed to the user), event type and event value. Events represent the different kinds of measures available in the application (Table II).

III. TESTS

The application has been tested on Samsung VR framework (Samsung Galaxy S8 and Gear VR HMD) and Google Daydream (Google Pixel 2 and Daydream HMD). Fig. 3 shows an example of a sequence tested with MIRO360. It is a three-minute video recorded with Nokia OZO camera and encoded in AVC with variable QP along time. Seven subjects scored

[2]This is the tentative name given to the methodology in ITU-T SG12 and VQEG joint test plan.

```
{
  "items": [
    {
      "uri": "Movies/training.mp4",
      "in_seq_method": "sscqe",
      "post_seq_questions": ["acr"]
    },
    {
      "uri": "Movies/summer_xc.ts",
      "in_seq_method": "sscqe",
      "post_seq_questions": ["acr", "vertigo", "mec"]
    },
    {
      "uri": "Movies/madrid_xc.ts",
      "in_seq_method": "sscqe",
      "post_seq_questions": ["acr", "vertigo", "mec"]
    },
    {
      "uri": "Movies/aleppo_xc.ts",
      "in_seq_method": "sscqe",
      "post_seq_questions": ["acr", "vertigo", "mec"]
    }
  ]
}
```

Fig. 2. Example of configuration file

TABLE II
OUTPUT EVENTS

Event	Value
START	Start of test session
STATE_CHANGE	Start of each phase (*idle*, *video*, *questionnaire*)
END	End of test session
LOOK_AT	Gaze direction unit vector (x, y, z)
SSCQE	SSCQE level (1-99)
SSDQE	Score (1-5) [8]
ACR	Absolute Category Rating
DCR	Degradation Category Rating
VERTIGO	Cybersickness scale [6]
DIZZY	Cybersickness [7]
MEC.X[1]	Mini-MEC-STD score
SPES.X[1]	Spatial Presence Experience Scale

[1]There is a different ”X” event for each item in the scale.

the sequence, first in SSCQE and secondly in SSDQE. Even though the test scope is very limited, it shows how the tool can capture the quality shape, both in continuous and discrete evaluation.

IV. Conclusions and Roadmap

We have presented MIRO360, an Android Virtual Reality application to assess subjective quality of 360 VR video using Samsung Gear and Google Daydream HMDs. The tool implements the different methodologies that are under study for the development of ITU-T P.360-VR, the future ITU-T recommendation for such kind of subjective tests.

Two lines of evolution are foreseen for MIRO360. On the one hand, the tool will be adapted to the final version of ITU-T P.360-VR, iteratively adopting the scales and protocols proposed for evaluation. On the other, some features are planned to ease the preparation and handling of tests. The most relevant are: configuration and event reporting via http,

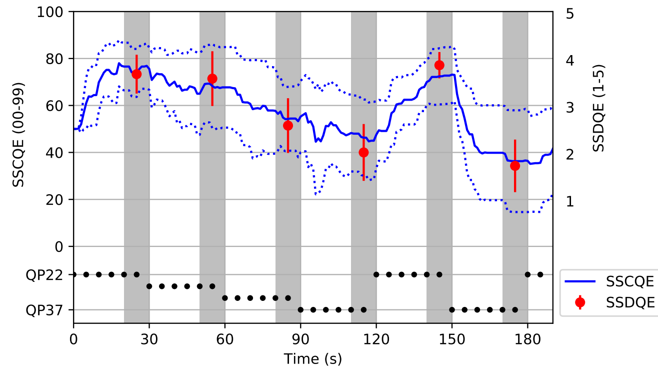

Fig. 3. Example of results from a test. Mean and confidence intervals are shown for SSCQE and SSDQE. Scoring periods for SSDQE are shown in grey.

recording of scores using voice recognition, and development of external tool to generate test configuration files and help with the analysis of the results.

References

[1] “ITU-T P.910. Subjective video quality assessment methods for multimedia applications,” International Telecommunication Union, Recommendation, Apr. 2008.

[2] “Subjective test methodologies for 360 degree video on HMD (P.360-VR),” International Telecommunication Union, ITU-T Recommendation (Under Study).

[3] S. Fremerey, A. Singla, K. Meseberg, and A. Raake, “AVTrack360: An open dataset and software recording people’s head rotations watching 360° contents on an HMD,” pp. 1–6, 2018.

[4] G. Regal, R. Schatz, J. Schrammel, and S. Suette, “VRate: a Unity3D asset for integrating subjective assessment questionnaires in virtual environments,” in *Proc. 10th Int. Conf. Quality of Multimedia Experience (QoMEX)*, 2018, pp. 1–3.

[5] “ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures,” International Telecommunication Union, Recommendation, Jan. 2012.

[6] P. Pérez, N. Oyaga, J. J. Ruiz, and A. Villegas, “Towards systematic analysis of cybersickness in high motion omnidirectional video,” in *Proc. 10th Int. Conf. Quality of Multimedia Experience (QoMEX)*, 2018, pp. 1–3.

[7] H. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, “A subjective study on QoE of 360 video for VR communication,” in *Proc. IEEE 19th Int. Workshop Multimedia Signal Processing (MMSP)*, 2017, pp. 1–6.

[8] J. Gutiérrez, P. Pérez, F. Jaureguizar, J. Cabrera, and N. García, “Validation of a novel approach to subjective quality evaluation of conventional and 3d broadcasted video services,” in *Proc. 4th Int. Workshop Quality of Multimedia Experience (QoMEX)*, 2012, pp. 230–235.

[9] T. Hartmann, W. Wirth, H. Schramm, C. Klimmt, P. Vorderer, A. Gysbers, S. Bocking, N. Ravaja, J. Laarni, T. Saari, F. Gouveia, and A. M. Sacau, “The Spatial Presence Experience Scale (SPES): A short self-report measure for diverse media settings,” *Journal of Media Psychology*, vol. 28, no. 1, pp. 1–15, 2016.

[10] P. Vorderer, W. Wirth, F. R. Gouveia, F. Biocca, T. Saari, F. Jancke, S. Bocking, H. Schramm, A. Gysbers, T. Hartmann, C. Klimmt, J. Laarni, N. Ravaja, A. Sacau, T. Baumgartner, and P. Jancke, “Development of the MEC Spatial Presence Questionnaire (MEC SPQ),” 2004, Unpublished report to the European Community, Project Presence: MEC (IST-2001-37661).