TEST EQUIPMENT OF TIME-VARIANT SUBJECTIVE PERCEPTUAL VIDEO QUALITY IN MOBILE TERMINALS

Olivia Nemethova, Michal Ries, Antitza Dantcheva, Stefan Fikar, and Markus Rupp

Vienna University of Technology, Institute of Communications and RF Engineering
Gusshausstrasse 25/389, 1040 Vienna, Austria
{onemeth, mries, mrupp}@nt.tuwien.ac.at

Abstract

In this article we discuss the suitability of various equipment types used to capture the user's subjective perceptual quality evaluation for video streams with time-variant quality. We present four basic types of testing equipment (mouse, joystick, throttle and sliding bar). Since method and equipment used for testing must not influence the test results, we analyzed the requirements for the testing equipment in general and in particular its usability and design features. Furthermore, we performed and evaluated an assessment of video quality using all presented tools. The evaluation allows us to propose the most appropriate testing tool for time-variant video quality.

KEY WORDS

Usability, time-variant video quality, perceptual quality evaluation

1 INTRODUCTION

Transport of real-time video streams over an error prone environment results in packet loss. This leads to various kinds of error artifacts and their possible spatial and temporal propagation in the video stream rendered at the user terminal. The character of the artifacts depends on the used compression method, the position of the error and the decoder implementation (i.e. using error concealment). The sensibility of the user to different types of artifacts can be measured by means of the mean opinion score (MOS) according to [1, 2]. These MOS values can be captured continuously while the sequence is being displayed which is especially suitable for evaluating of time-variant video quality. However, performing the MOS measurements is demanding time and human resources. Therefore, there are many proposals for video quality metrics supposed to correspond to human perception and their proper design, see for example [3, 4, 6]. In order to find such metrics, a reliable set of MOS measurements is of utmost importance. There are several methods how continuous feedback from the users during the evaluation process can be captured. Oral feedback from users [5] is rather ambiguous and difficult to evaluate. Therefore, different kinds of electronic equipment are utilised in praxis [4, 6], usage of which can more or less influence the test results.

Intention of this article is to analyze the differences between four basic types of capturing equipment both theoretically and by means of a survey. Section 2 contains the analysis of the requirements on the test equipment from the design point of view. In Section 3 the evaluated test equipment is described. In Section 4 the test method for presented types of equipment is described, while in Section 5 the results are presented and evaluated. Section 6 contains conclusions.

2 USABILITY OF THE TEST EQUIPMENT

It is known from literature [7, 8] that humans can consciously perform one action at a time only. As a test person has to concentrate on both the video quality assessment as well as the evaluating process, it is necessary to ensure that the evaluating process is as simple as possible and thus does not influence the video quality survey results. This requirement is closely connected to the usability and design of the test equipment. From [8, 9] we collected the following common requirements for the usable design and interpreted them with respect to the test equipment design:

- Common understanding, knowledge, natural constraints and habits should be taken into account (i.e. the way of using the equipment should be the same for testing as for the purposes usually used for).
- The structure of the task to be performed by the equipment should be as simple and easy as possible (i.e. equipment handling should not cause any physical or psychical fatigue).
- The purpose of the equipment's functional units should be as visible and understandable as possible without any need for further explanation (i.e.

start/stop button position and the way of evaluation should be clearly visible).

- The user should be provided a continuous feedback about the equipment state (i.e. instantaneous position).

- The mapping of the desired action (i.e. to give higher grade for better quality) onto the performed equipment manipulation (i.e. moving the equipment) should be as obvious as possible (i.e. moving up).

- Design of the equipment should be robust against human error (i.e. unambiguous buttons).

To evaluate the usability of different test capturing tools according to the above design rules we formulated a questionnaire and performed the video evaluation tests as described in the following sections.

3 EQUIPMENT DESCRIPTION

We implemented four types of test equipment (joystick, sliding bar, throttle and a mouse with control software). Pictures of these tools can be seen in Figure 1. Each of these tools can be controlled by the same PC software allowing for starting and stopping the data capturing, saving or resetting the captured data and changing the sampling frequency. The software also contains a vertical bar showing the instantaneous value of the evaluation. However, if the equipment were properly designed, the bar would not be needed, and the information would be provided by the test equipment itself. Looking at the bar as well as looking at the test equipment can distract the test person's attention to the video sequence. This is especially critical for the tests with mobile terminals as it is not possible to have the feedback bar together with video on the same screen.

3.1 Mouse

The simplest and also widely used method to capture the user feedback is using the mouse. In Figure 1 the snapshot of the software slider moved using the mouse is shown.
The slider can be either moved by the mouse wheel or by keeping the left mouse button pressed upon the slider and moving the mouse up or down. Starting and stopping of the capturing has to be performed via the PC software (clicking on the button). There is no notifcation upon achieving the lower or upper range value. This tool keeps its position constant if there is no force to change it.

Figure 1. User interfaces to capture the evaluation: joystick, sliding bar, throttle, screenshot of a program controlled by a standard mouse or any of the mentioned devices.

3.2 Joystick

More comfortable device than the mouse is a joysick. It has an extra button to start and stop the capturing so that the test person does not need to interact with the PC software unless the captured results have to be saved. The joystick is in the upright position at the beginning and falls back to this position if not affected by any other force. During the evaluation a force is necessary to keep the joystick in the constant position other than the center of the range. A joystick can be used equally by left and right handed people.

3.3 Throttle

It is a tool originally designed for flight simulators. It contains several buttons, one of them controlling starting and stopping the data capturing. A test person does not have to use the PC software unless the data have to be saved. We used a left hand throttle to allow the test person holding the mobile terminal displaying the video stream in the right hand. The throttle keeps the last set position. It is held by the whole hand; evaluating is performed by moving it up and down mainly in the horizontal plane.

3.4 Sliding Bar

We also constructed a hardware sliding bar. To connect linear 10cm long potentiometer to the USB port we used an 8-bit microcontroller PIC16C745 by Microchip. Our sliding bar contains one button to start and stop the capturing process. The sliding bar can be used equally by left and right handed people. It also keeps the last position. Evaluating is performed by sliding the potentiometer up and down in the horizontal plane. The force needed for that is minimal - it can be moved by one finger, the hand resting on the sliding bar box.

4 EQUIPMENT AND VIDEO TESTING

We performed the tests with 14 unpaid test subjects. We chose the group to be homogenous in age and education to obtain consistent evaluation. All test persons were 20-30 years old students of the Vienna University of Technology. We had two left-handed persons in the test group. However, from our equipment the only asymmetrical one was the left hand throttle to allow people to evaluate with the left hand while keeping the terminal in their right hand as usual. Each of the test persons was asked to evaluate the overall quality of one two-minutes long video sequence using mouse, joystick, throttle and sliding bar. The video sequence had QCIF (144×176) resolution and was displayed at the screen of a UMTS mobile terminal from Sony Ericsson (Z1010). The two minutes long video sequence contained artifacts corresponding to a 10% uniformly distributed packet loss at the Internet Protocol (IP) layer. We chose such extreme packet loss to obtain a lot of impairments that need to be evaluated correspondingly. An H.263 codec was used to compress the sequence to a 128 kbit/s RTP stream. The same sequence was shown to all the test persons, who evaluated it using all equipment introduced in Section 3. It contained a soccer game with several scene cuts. This sequence was chosen as the people are extremely critical to such content [10].

At the beginning, the test person was informed how to handle the equipment. We used single stimulus continuous quality evaluation (SSCQE) as described in [2]. After each evaluation, the test subjects were asked to evaluate following equipment features using a five-grade scale (one being the worst and five being the best note):

- user friendliness
- simplicity of use
- visibility and clarity of equipment's functional units
- feeling for the position within the evaluation scale (feedback).

We did not specially ask to evaluate the immediacy of feedback as it was the same for all of the methods. They were furthermore asked to answer the following questions by 'yes' or 'no' and if needed to comment the answer:

- Was there any previous knowledge about the equipment necessary to perform the evaluation?
- Did the usage of the equipment cause any fatigue?
- Did you feel swamped using the equipment?
- Did the equipment mislead to the wrong evaluation due to its construction or the difficulties to manipulate it?

After testing with all the equipment, the test subjects were asked to choose their favourite for the final test of video quality.

5 RESULTS AND EVALUATION

5.1 Equipment Test Results

The results for user friendliness, simplicity, position feedback and visibility of functions can be seen in the Table 1.

Tool	User friendly	Simple	Position feedback	Visible functions
mouse	2.50	3.21	2.57	4.21
slider	4.21	4.71	4.50	4.57
joystick	3.21	4.00	3.00	4.14
throttle	3.93	4.36	3.93	3.86

Table 1. Average user evaluation for different capturing tools. Scale: 1-bad, 2-poor, 3-fair, 4-good, 5-excellent.

Not surprisingly, the mouse reached the lowest value for the user friendliness, simplicity and the position feedback. According to the comments this was caused by the lack of feedback. As the mouse wheel does not have any hard limits, the test person had to check regularly the state of the slider at the PC screen which on the other hand misled the attention from the video displayed on the mobile terminal. The simplicity was evaluated low because it was necessary to simultaneously start the video sequence at the mobile terminal and start the capturing by clicking the button on the PC screen. The Mouse is a frequently used tool for the evaluation of time-variant video quality [4]. Its disadvantages are not so apparent if the video is displayed on the PC screen. Anyhow, the peripheral visual feedback when watching the video can

also mislead the attention from the artifacts and thus influence the evaluation results.

More details on what notes the user gave to different test equipment are listed in Figure 2. For the

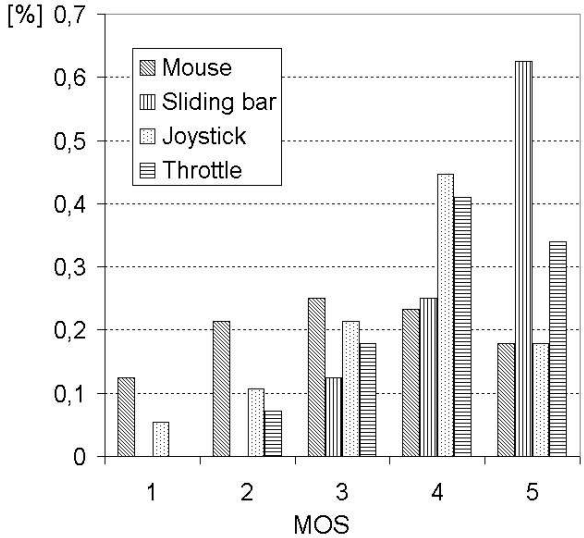

Figure 2. Histogram of values given to mouse, slider, joystick and throttle together for user friendliness, simplicity, position feedback and functions visibility.

visibility of the functions, the throttle obtained the lowest value. This is caused by a rather high number of buttons in its top part, which some users considered to be confusing and not self-explaining.

The joystick has lost the users' preferences mainly for its property to fall back in the upright position if there is no other force. This can be clearly seen in Table 2.

Tool	Prev. knowl.	Fatigue	Swamped	Wrong decis.
mouse	21.43	50.00	7.14	57.14
slider	14.29	0	0	7.14
joystick	14.29	42.86	14.29	71.43
throttle	21.43	14.29	7.14	14.29

Table 2. Percent of people that needed previous knowledge, felt fatigue and swamped, and believed that the equipment can lead to wrong evaluation for different equipment types.

Test persons complained to get tired by using the joystick for a longer time, as they had to concentrate on holding it in the constant position. Forgetting to do so led to wrong evaluations. Some people got tired also when using a mouse, mainly because of the necessity to watch both PC and terminal simultaneously. Previous knowledge was needed to start evaluating by mouse and by throttle, the joystick is a well-known tool, people did not have problems with, while the sliding bar seems to be simple and obvious enough. However, more than 71% of the test persons consider the joystick to misleading the evaluation and more than 57% of the people do not trust the mouse. We asked the people to choose one tool for the final test to see their real preference. In Figure 3 the ratio of the chosen tools can be seen.

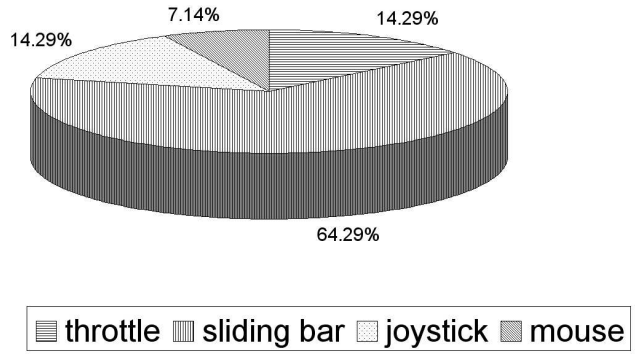

Figure 3. Percent of the people that chose sliding bar, mouse, joystick or throttle for their final test.

The definite winner was the sliding bar. There were no problems with the feedback feeling, so that the test person could fully concentrate on the displayed video sequence.

5.2 Video Test Results

Different capturing tools can also influence the video test results. The effect is not identical for each of the users. In Figure 4 a typical evaluation of one user by all four capturing tools can be seen.

All curves follow the same gross shape. However it can be seen, that the difference between the maxima and minima (used range) is largest for the sliding bar and smallest for the mouse. The difference between the peak value evaluation by two different tools is more than 50%. The curve obtained by the joystick contains more narrow peaks, probably caused by the back falling effect. This in turn could fuzz the effect of the error propagation on the user perceptual quality. The curves obtained by the throttle seem to decay slower than the others. This can be caused by the fact that the movement performed by the throttle is of larger scale compared to other tools. The reason can also be using the left hand that is slower for right handed people.

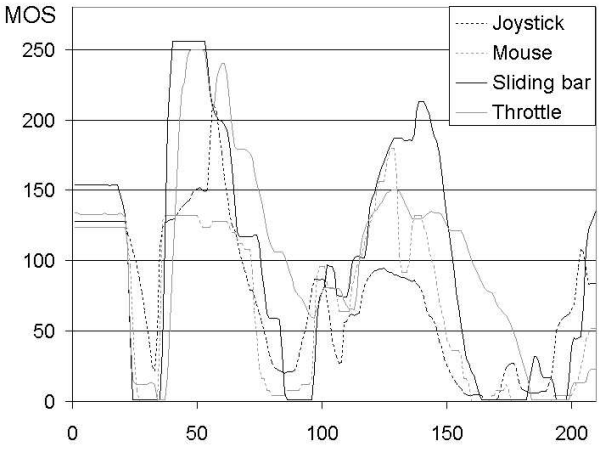

Figure 4. Typical evaluation of one user.

6 CONCLUSIONS

According to our tests the sliding bar tool seems to be the most appropriate for time-variant video quality testing. This is given by its simple design, obvious functionality and comfortable size. The test persons showed clear preferences for the sliding bar tool that in turn also best corresponds to the design and usability rules. Investigating the video evaluation, the difference between the results obtained by different tools is considerable. Therefore, we conclude that constructing an appropriate capturing device is surely worth while if a reliable set of measurements needs to be obtained. When proposing such equipment, it is necessary to follow the design rules for better usability of such device as video quality evaluation is a task that requires high concentration. People get easily annoyed by inappropriate designs.

The features that contribute at most to the usability of the devices used for video quality testing are according to our studies: simplicity, learnability and the feedback. This is caused by the fact, that for the testing of the video quality it is very important, that the test person remains steadily concentrated to the screen with the video sequence, as some of the impairments can occur for a short time only. This requires the test equipment to provide sufficient feedback on the position of the capturing device preferably by feeling, without looking on the computer screen containing the position bar.

To achieve this, the capturing device needs to be very simple, so that the test person is able to learn the necessary movements within the short time given by the trial evaluation runs, that are not taken into account when evaluating the video quality tests. Thus, suitable devices for video quality evaluation should only contain separated start/end button and the moving part. The moving part should be easy to handle and should not have any preferred state (like the middle position for joystick) to avoid wrong evaluation and fatigue annoying the test persons. Furthermore, the moving part should be of size corresponding to the comfortable range of the movement of the human hand fixed for instance on the table to get the feeling for the position - feedback. If possible, the device should allow only the movement in one direction to avoid confusing position feeling (for instance by moving the mouse or joystick left or right) and necessity to check the visual position bar on the screen.

7 ACKNOWLEDGEMENT

The authors would like to thank mobilkom austria AG&Co KG for supporting their research. The views expressed in this paper are those of the authors and do not necessarily reflect the views within mobilkom austria AG&Co KG.

References

[1] ITU-T Recommendation P.910, *Subjective video quality assessment methods for multimedia applications* (ITU, 1998).

[2] ITU-R Recommendation BT.500, *Methodology for the subjective assessment of the quality of television pictures* (ITU, 2002).

[3] VQEG, *Final report from the video quality experts group on the validation of objective models of video quality assessment* (VQEG 2000, available at http://www.vqeg.org).

[4] S. Winkler, F. Dufaux, Video Quality Evaluation for Mobile Applications. Proc. of *Visual Communication and Image Processing Conference (VCIP2003)*, Lugano, Schwitzerland, 2003.

[5] J.D. McCarthy, M.A. Sasse, D. Miras, Evaluating Mobile Video Quality. Proc. of *13th IST Mobile and Wireless Communications Summit*, Lyon, France, 2004.

[6] M.A. Masry, S.S. Hemami, CVQE: A Metric for Continuous Video Quality Evaluation at Low Bit Rates. Proc. of *SPIE Conf. on Human Vision and Electronic Imaging*, Jan. 2002.

[7] A.S. Reber, *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious* (Oxford Psychology Series no. 19, Oxford University Press, Oxford, 1993).

[8] D.A. Norman, *The Design of Everyday Things* (Currency and Doubleday, New York, 1990).

[9] H. Beyer, K. Holtzblatt, *Contextual Design: Defining Customer-Centered Systems* (Morgan Kaufmann Publishers, Academic Press, San Diego, USA, 1990).

[10] O. Nemethova, M. Ries, E. Siffel, M. Rupp, Quality Assessment for H.264 Coded Low-Rate and low-Resolution Video Sequences. Proc. of *IASTED CIIT2004*, US Virgin Islands, November 2004.