AVHD Project

Subjective Test Method

NOTE: Nothing in this document has been fully decided on yet!

Version 0.0

June 27, 2014

|  |  |
| --- | --- |
| Contacts: |  |
| Christian Schmidmer, OPTICOM GmbH | Email: cs@opticom.de |
| [tab to enter email] | [tab to add a line] |

Editorial History

|  |  |  |
| --- | --- | --- |
| Version | Date | Nature of the modification |
| 0.0 | June 27, 2014 | Initial Draft, created by Chris Schmidmer |
|  |  |  |

Contents

[1 . Introduction 1](#_Toc392751632)

[2 . Relation to other documents 2](#_Toc392751633)

[3 . General Subjective Test Method, P.913 3](#_Toc392751634)

[3.1 Method of Choice: ACR with Hidden Reference 3](#_Toc392751635)

[3.2 Viewing Distance, Number of Viewers per Monitor, and Viewer Position 3](#_Toc392751636)

[3.3 Display Specification and Setup 4](#_Toc392751637)

[3.3.1 Viewing Conditions 6](#_Toc392751638)

[3.4 Audio Setup and Listening Environment 6](#_Toc392751639)

[3.5 Subjective Test Playback 7](#_Toc392751640)

[3.6 Evaluators (Viewers) 7](#_Toc392751641)

[3.6.1 Test Data Collection 8](#_Toc392751642)

[3.7 Results Data Format 8](#_Toc392751643)

# . Introduction

This document describes the subjective test method which shall be used to create databases for the validation of audio visual quality metrics.

# . Relation to other documents

Things defined by other documents…

# . General Subjective Test Method, P.913

In general the subjective test procedure should follow the guidelines given in P.913 if not otherwise stated in this document. P.913 leaves the experimenter a lot of freedom but forces them to take many design decisions. This section explains the details which should be chosen for the AVHD project, e.g. exact test method, sequence length, report format etc.

## Method of Choice: ACR with Hidden Reference

The method to be used is the ACR method with hidden reference, as explained in P.913. The absolute category rating scale (ACR) from ITU-T Rec. P.913 will be used for collecting subjective judgments.

Only one question (e.g. rate the video quality) and one type of stimulus (e.g. video only) is allowed per session.

*Open questions:*

* *Rate AV quality only?*
  + *The overhead for scoring A only and V only is not necessarily so much.*
* *Rate audio only, video only without seeing/hearing the other component, or rate them in the presence of the not rated component.*

*Preliminary decision: Main text: See/hear/score one component only, but leave the option for some labs to also score a only and v only in the presence of both components.*

* *How to deal with Lip Synch problems in objective models?* 
  + *-> Excluded for now*
* *How to deal with content dependencies*
  + *Source content should be balanced and the selected content type should classified by type of sound type.*

*🡪 Editing stopped here!*

The test presentation order must be randomized according to standard procedures (e.g., Latin or Graeco-Latin square or via computer).

## Viewing Distance, Number of Viewers per Monitor, and Viewer Position

The test instructions request evaluators to maintain a specified viewing distance from the display device. The viewing distance is 3H, with H=Picture Height (picture is defined as the size of the video window).

Preferably, each test viewer will have his/her own video display. For those parameters that are not specified in this test plan, the subjective test will conform to ITU-T Rec. P.910 requirements.

It is recommended that viewers be seated facing the center of the video display at the specified viewing distance. That means that viewer's eyes are positioned opposite to the video display's center (i.e. if possible, centered both vertically and horizontally). If two or three viewers are run simultaneously using a single display, then the viewer’s eyes, if possible, are centered vertically, and viewers should be centered evenly in front of the monitor.

## Display Specification and Setup

**Note** that in all subjective tests 1 pixel of video will be displayed as 1 pixel native display. No up-sampling or down-sampling of the video is allowed at the player.

If interlaced video will be evaluated on a monitor that does not accept interlaced content but meets all other necessary specifications, then the interlaced SRC and PVS may be de-interlaced separately (e.g., using software) prior to playing the content to the monitor. Preferably, it is desirable that the lab who can display interlaced signals should be assigned interlaced experiments.

All subjective experiments will use LCD monitors. Only high-end consumer TV (Full HD) or professional grade monitors should be used. LCD PC monitors may be used, provided that the monitor meets the other specifications (below) and is color calibrated for video.

Given that the subjective tests will use different HD display technologies, it is necessary to ensure that each test laboratory selects an appropriate display and that common setup techniques are employed. Due to the fact that most consumer grade displays employ some kind of display processing that will be difficult to account for in the models, all subjective facilities doing testing for AVHD shall use a full resolution display.

All labs that will run viewers must post to the AVHD reflector information about the monitor model to be used. If a proponent or ILG has serious technical objections to a monitor, the proponent or ILG should post the objection with detailed explanation within two weeks. The decision to use the monitor will be decided by a majority vote among proponents and ILGs.

**Input requirements**

* HDMI (player) to HDMI (display); or DVI (player) to DVI (display)
* HD-SDI (player) to HD-SDI (display)
* Missing: Displayport!

Conversion (HDMI to HD-SDI or vice versa) should be transparent

If possible, a professional LCD monitor should be used. The monitor should have as little post-processing as possible. Preferably, the monitor should make available a description of the post-processing performed.

If the native display of the monitor is progressive and thus performs de-interlacing, then if 1080i SRC are used, the monitor will do the de-interlacing. Any artifacts resulting from the monitor’s de-interlacing are expected to have a negligible impact on the subjective quality ratings, especially in the presence of other degradations.

The smallest monitor that can be used is a 24” LCD.

A valid HDTV monitor should support the full-HD resolution (1920 by 1080). In other words, when the HDTV monitor is used as a PC monitor, its native resolution should be 1920 by 1080. On the other hand, most TV monitors support overscan. Consequently, the HDTV monitor may crop boundaries (e.g, 3-5% from top, bottom, two sides) and display enlarged pictures (see Figure 10.2). Thus, it is possible that the monitor may not display whole pictures, which is allowed.

The valid monitor should be LCD types. The monitor should be a high-end product, which provides adequate motion blur reduction techniques and post-processing which includes deinterlacing.



Figure 10.2. An Example of Overscan

### Viewing Conditions

Viewing conditions should comply with those described in International Telecommunications Union Recommendation ITU-T Recommendation P.910, 1999.

## Audio Setup and Listening Environment

All Audio equipment should be studio quality grade as far as possible. Digital signals should have a sample rate of at least 48 kHz and should be represented in linear PCM format with at least 16 bits resolution. Standalone (pre-) amplifiers are generally very good these days, but care must be taken with sound cards and consumer devices for PCs which are often of poor quality especially with regard to noise and RFI. Wherever possible either optical or symmetrical connections should be used. Mobile phones and other HF sources should be kept as far away from the listening setup as possible.

Listening should be performed using headphones. The headphones used should be high quality, neutral sounding and ideally be diffuse field equalized (e.g. Beyerdynamic DT770 Pro, around EUR 160,-).

The listening level should be adjusted to a comfortable level on the analog amplifier. Since the dynamic range of movie sound may be extremely high it is recommended that subjects may listen to the test clips with highest and lowest volume before the test and to correct the level for each subject individually to avoid any discomfort. Once the level is adjusted and the test has been started it should not be modified gain.

No level adjustments should be performed in the digital domain, e.g. using the volume controls of a Laptop since this increases the quantization noise.

The listening environment should be reasonably quiet (e.g. no office or traffic noise, no air conditioning hum). If possible, conditions laid out in BS.1116 shall be met.

## Subjective Test Playback

All subjective tests will where possible be run using the same software package, provided by Acreo. The software package will include the following components:

* Entry system for evaluator details (e.g. name, age, gender)
* Test screens (prompts to users, grey panel, ACR scale, response input, data capture, data storage)
* Timing control
* Correct video play-out check
* Video player

No additional visual impairments must be introduced by the subjective playback system.

## Evaluators (Viewers)

Exactly 24 valid viewers per experiment will be used for data analysis.

Different subjective experiments will be conducted by several test laboratories. A valid viewer means a viewer whose ratings are accepted after post-experiment results screening. Post-experiment results screening is necessary to discard viewers who are suspected to have voted randomly. The rejection criteria verify the level of consistency of the scores of one viewer according to the mean score of all observers over the entire experiment. The method for post-experiment results screening is described in Annex IV. Only scores from valid viewers will be reported in the results spreadsheets[[1]](#footnote-1).

It is preferred that each viewer be given a different randomized order of video sequences where possible. Otherwise, the viewers will be assigned to sub-groups, which will see the test sessions in different randomized orders. A maximum of 6 viewers may be presented with the same ordering of test sequences per subjective test.

Each viewer can only participate in 1 experiment (i.e. one experiment at one image resolution).

Only non-expert viewers will participate. The term non-expert is used in the sense that the viewers’ work does not involve video picture quality and they are not experienced assessors. They must not have participated in a subjective quality test over a period of six months.

Prior to a session, the observers should usually be screened for normal visual acuity or corrected-to-normal acuity and for normal color vision. Acuity will be checked according to the method specified in ITU-T P.910 or ITU-R Rec. 500, which is as follows. Concerning acuity, no errors on the 20/30 line of a standard eye chart[[2]](#footnote-2) should be made. The chart should be scaled for the test viewing distance and the acuity test performed at the same location where the video images will be viewed (i.e. lean the eye chart up against the monitor) and have the evaluators seated. Ishihara or Pseudo Isochromatic plates may be used for colour screening. When using either colour test please refer to usage guidelines when determining whether evaluators have passed (e.g. standard definition of normal colour vision in the Ishihara test is considered to be 17 plates correct out of a 38 plate test; ITU-T Rec. P.910 states that no more than 2 plates may be failed in a 12 plate test. Evaluators should also have sufficient familiarity with the language to comprehend instructions and to provide valid responses using the semantic judgment terms expressed in that language.

### Test Data Collection

The responsibility for the collection and organization of the data files containing the votes will be shared by the ILG Co-Chairs and the proponents. The collection of data will be supervised by the ILG and distributed to test participants for verification.

## Results Data Format

The following format is designed to facilitate data analysis of the subjective data results file.

The subjective data will be stored in a Microsoft Excel 2010 (i.e., \*.xls) spreadsheet. Each spreadsheet will contain all of the data for one experiment. The top row of this file will be a header. Each row below the header will contain one video sequence. The columns are as follows, in this order: experiment number, SRC number, HRC number, file name, subject #1’s ACR score, subject #2’s ACR score, … subject #24’s ACR score.

Missing ACR values will be left blank.

Figure 10.3 contains an example, showing 12 of the 24 subjects’ scores, and only six PVS.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Experiment | SRC Num | HRC Num | File | SUBJECT'S RESULTS | | | | | | | | | | | |
| 1 | 1 | 1 | hybrid1\_s01\_hrc01.avi | 2 | 3 | 1 | 2 | 2 | 1 | 3 | 1 | 3 | 2 | 2 | 3 |
| 1 | 1 | 2 | hybrid1\_s01\_hrc02.avi | 2 | 2 | 1 | 2 | 1 | 2 | 3 | 2 | 3 | 3 | 1 | 2 |
| 1 | 1 | 3 | hybrid1\_s01\_hrc03.avi | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 3 | 1 | 1 | 1 |
| 1 | 1 | 4 | hybrid1\_s01\_hrc04.avi | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 5 | hybrid1\_s01\_hrc05.avi | 2 | 2 | 2 | 2 | 2 | 1 | 3 | 2 | 3 | 2 | 1 | 1 |

Figure 10.3. Format for subjective data spreadsheet.

1. Test laboratories can keep data from invalid viewers if they consider this to be of valuable information to them but they must not include them in the VQEG data. [↑](#footnote-ref-1)
2. Grahm-Field Catalogue Number 13-1240. [↑](#footnote-ref-2)