|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| D:\usr\campos\TSB-Reference\Logos\ITU\sigleITU.gif | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | | SG12-C.n | |
| **STUDY GROUP 12** | |
| **Original: English** | |
| **Question(s):** | | Q13/12 | Krakow, 28-29 Nov 2017 | | |
| **CONTRIBUTION** | | | | | |
| **Source:** | | Huawei Technologies Co. Ltd. | | | |
| **Title:** | | Huawei subjective tests for 360 degree HMD video | | | |
| **Purpose:** | | [Choose a purpose from the dropdown list] | | | |
| **Contact:** | | Rachel Huang Huawei Technologies Co. Ltd. China | | | Tel: +86 25 56623633 Fax:  E-mail: rachel.huang@huawei.com |
| **Contact:** | | Haoping Yu Futurewei Technologies Inc. USA | | | Tel: +1 317 965 9895 Fax:  E-mail: haoping.yu@huawei.com |
| **Contact:** | | Paul Coverdale Huawei Technologies Co. Ltd. Canada | | | Tel:  Fax:  E-mail: coverdale@sympatico.ca |

|  |  |
| --- | --- |
| **Keywords:** | 360 degree, subjective test, FOV, HDM, UHD |
| **Abstract:** | This contribution provides more details of the subjective tests carried out by Huawei for 360 degree HMD video. |

**Discussion**

As discussed during the September SG12 meeting, it was suggested to submit more detailed information (than given in SG12-C127) for 360 degree HMD video. This contribution is for this purpose.

**Proposal**

Huawei subjective tests for 360 degree HMD video

# Tools for VR-MOS Subjective Test

## VR Device

The HMD used for the test is HTC Vive (market value: 6888 CNY), with a screen resolution of 2160x1200, a monocular resolution of 1080x1200, and an FOV of 110°.

## PC

Intel Core CPU i7-4790k @4 GHz CPU, 64 GB RAM (DDR4), NVIDIA GeForce GTX 1080 GPU, and business grade PCIe SSD Drive (1400MB/s write, 1000MB/s read)

## 360 Video Player

The domestically made player is based on OpenVR and OpenGL. The player supports a list of functions: playing 360° 4K UHD video in the HMD; adding additional visual delay or audio delay to the device's inherent delay while playing; player control functions such as play/pause, initial buffering duration, stalling durations and playing duration, etc.

## The Rating Tool

The rating tool is developed based on Unity. It utilizes HTC Vive's controller as the user input device. A user can simply pull the trigger button on HTC Vive controller to perform grading operations.

## The Audio/Video Encoder

The audio/video encoder is developed based on FFmpeg audio/video encoder (i.e. libfdk\_aac, libx264).

# Test Environment

The test is carried out in a noise isolation room, as noise is the major environmental factor that would affect the test. Environmental factors other than noise are shielded out by the HMD and thus do not impact the subjects. The noise isolation room is set up with a PC, an LCD display, a spatial audio speaker, and two HTC Vive detecting cameras. The two cameras are placed in line with the PC, with a 1.5 meter distance apart from each other. During the test, subjects should be sitting 1 meter away from the PC and facing it. The two detecting cameras should be shooting the subjects. The whole setup of the test environment is showing in the following pictures.



# Participants

25 post-graduate students from Xidian University (14 male, 11 female) are recruited as participants. The age range of the participants is from 20 to 30, with an average of 23. All subjects have normal or correct-to-normal vision.

# Subjective Testing Methods

## P.913

The existing P.913 subjective testing standard uses SSM (Single Stimulus Methods). When using SSM, test sequences are displayed to subjects in a random and separated manner, and subjects rate only the relevant aspect of the test sequences. In SSM, the same test sequence is only displayed to a subject once. There are intervals between displaying successive test sequence. Grading is done by subjects during the intervals. The following figure illustrates the procedure:



Unlike using a fixed interval between two test sequences to 10 seconds, our test let subjects control when to continue to next test sequence. After a grading operation, a subject can click a "Next" button to continue to next test sequence.

The grading tool adopts ACR (Absolute Category Rating) of P.913 standard for rating, showing below:

|  |  |
| --- | --- |
| 5 | Excellent |
| 4 | Good |
| 3 | Fair |
| 2 | Poor |
| 1 | Bad |

## Test design

1. There is a preparation phase for the test. In that phase, subjects are familiarized with some high rating and low rating video samples. Those video samples are selected from among the actual test sequences. Therefore, it is possible that a video sequence used in the preparation phrase will be given to the same subjects for grading. How much video is given to subjects in the preparation phrase can be decided by the instructor before the test. The rule is every subject should only be prepared once, and the suggested time for the preparation phase is 15 minutes.
2. When the test begins, subjects are asked to read the instructions and wear the HDM. Once the HMD is worn, the knobs on each side of the HMD can be used to obtain a clear vision.
3. If the above steps are done, a computer program (a .bat file) is launched to execute a test sequence. Once a video test sequence is finished, the subject is asked to grade the content. Having done the rating, the subject can move forward to the next video test sequence. The operations of grading and moving forward are performed by using the HTC Vive Controller. The process continues until all the video test sequence are graded.
4. The two HTC Vive detecting cameras only support an FOV range of 110°. The subject should be careful about not looking outside of the FOV range.
5. The displayed video content on HMD has limited field of view, as it doesn’t have visual information at the top and bottom angles. So the subject should be instructed not to look at the top or bottom angle:





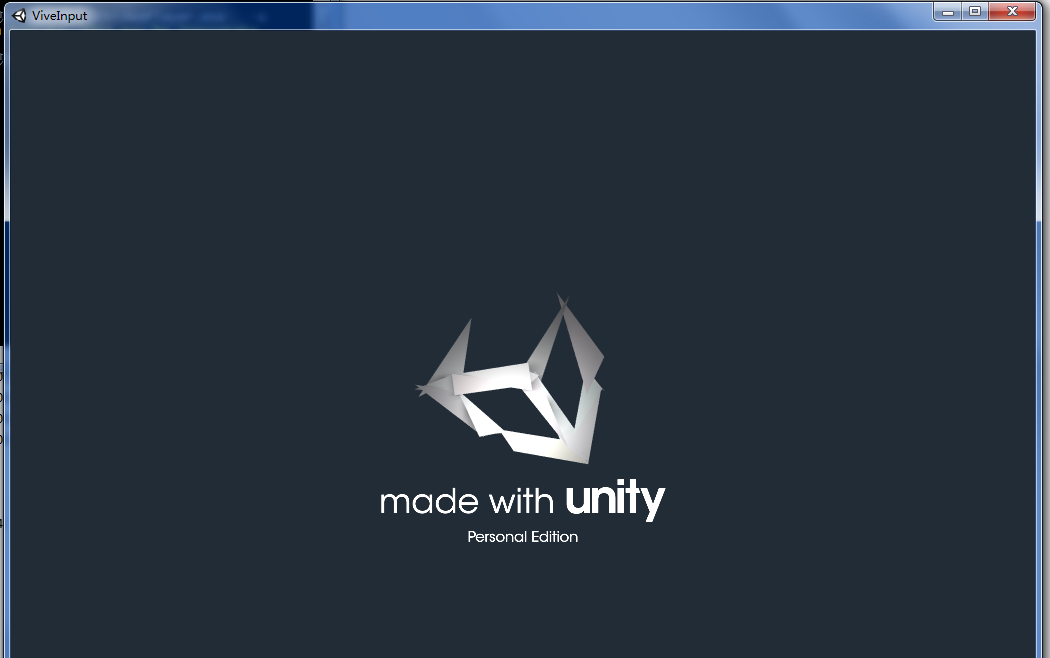
## Test Procedure

1. An assigned instructor prepares the subjects for the test.
2. The subjects are informed about the procedure, equipped with the HMD, and set in ready position.
3. The instructor launches a computer program which is a Windows batch processing file (with a .bat suffix). The computer program starts the player and plays the first test sequence. The content is binocular and projected to the display screen, meanwhile filling the frame buffer for OpenVR. The frame buffer is illustrated as below:

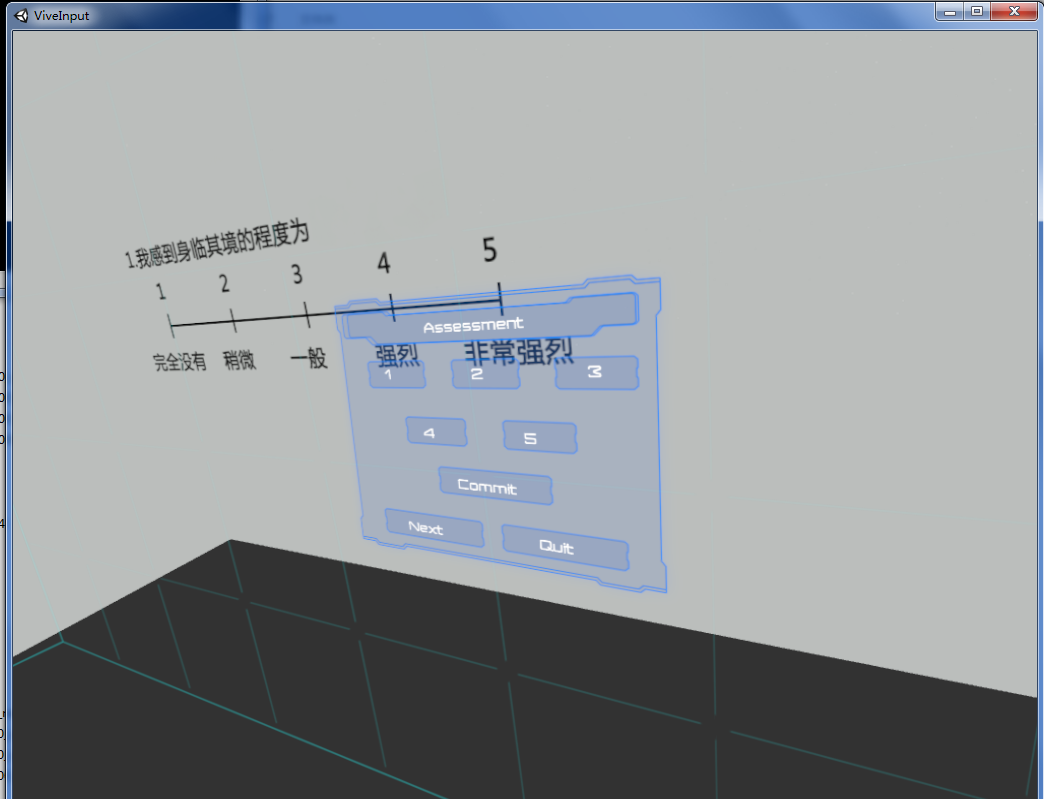


The player reads the content in the frame buffer (through the API provided by OpenVR's SDK), and sends it to the HMD for viewing.

1. After a subject has finished viewing the content through the HMD, the grading tool is launched, which takes about 5 seconds to complete. Here is the launch screen of the grading tool:



When the rating tool is completely launched, the rating interface will be shown to subjects through the HMD. Then the subject can give the rating score by using the HMD controller. The following figure is a snapshot of the rating interface.



During the rating process, subjects will be asked one or more questions. For each question, they should first choose the rating (1-5), then click "Commit". They should click "Next" to advance to the next question if any. When all questions are completed, "Quit" is clicked to finish the grading process. Thereafter, subjects can instruct the player to start playing the next test sequence. The rating process takes about 20 seconds.

For each test sequence, there is a rating given by an expert group. The ratings given by subjects will be compared against the ratings given by the expert group, and yields a result of PCC (Pearson Correlation Coefficient). If the PCC is under 0.75 (as the threshold suggested by P.910), this round of test is deemed as invalid. The subject will be requested to repeat the test after a certain time of rest.

# Subjective Test Tasks

**Task 1**：Video Quality assessment. Assess video with different resolutions and different frame rate.

**Task 2**：Audio Quality assessment. Together with task 4.

**Task 3**：Visual Fidelity assessment, which evaluates the fidelity of the video sensed by the subject while watching the 360-degree video. This assessment results in two rating values for the same video content, one for the video quality, another for the video fidelity.

**Task 4**：Acoustic Fidelity assessment which evaluates the fidelity of the audio sensed by the subject while watching the 360-degree video. This assessment results in two rating values for the same video content, one for the audio quality, another for the acoustic fidelity.

**Task 5**：Interaction Quality assessment. The interaction quality sensed by the subject after the subject performs some operation, while the subject is watching the 360-degree video.

**Task 6**：Sense of Presence assessment. The presence sensed by the subject while watching the 360-degree video.

# Instructions to the Subjects

|  |  |
| --- | --- |
| Number | Question examples |
| Task 1 | What is your opinion of the video quality? |
| Task 3 | What is your opinion of the video comparing to the real-world? |
| Task 4 | 1. What is your opinion of the audio quality? 2. What is your opinion of the audio comparing to the real-world? |
| Task 5 | How do you feel the responsiveness comparing to your movement in the real-world? |
| Task 6 | What is your opinion of the “be there” experience? |

# Subjective Test Sequence

## Task 1

Test Resource: 8 video sequences, each of which contains 4K resolution (3840x1920) monoscopic 360-degree YUV videos, 3 of them last for 20 seconds, 5 of them last for 10 seconds. These video sequences come with no audio. Therefore suitable soundtracks are played with them while testing.

Test Sequence: 8 YUV video sequences, each of which is classified into 4 resolutions. Each resolution is encoded in 5 different bitrates, resulting in 160 test sequences.

|  |  |
| --- | --- |
| Resolution | bitrate（Mbps） 30fps |
| 4K (3840×1920) | 6.6, 11.1, 19.9, 26.5, 50.9 |
| 2K (2160×1080) | 2.1, 3.5, 6.3, 8.4, 16.1 |
| 1080P (1920×960) | 1.7, 2.8, 5.0, 6.6, 12.7 |
| 720P (1280×640) | 0.7, 1.2, 2.2, 2.9, 5.7 |

## Task 3

Test Resource: 3 video sequences, all last for 20 seconds and contain 4K resolution (3840x1920 and 3440x1720) stereoscopic 360-degree MP4 video of 30fps framerate. Each video sequence comes with AAC stereo audio and 8 channel WAV spatial audio.

Test Sequence: 3 video sequences, each of which is encoded in 4 different bitrates, among which the monoscopic bitrates are 0.8/1.2/3/12Mbps and stereoscopic bitrates are 1.2/2.0/6.0/18Mbps. With the HMD's FOV value configured to 60, 90° and 110°, the resulting number of test sequences is 72.

## Task 4

Test Resource: Use the same resource as Task 3, but select the video with highest encoding quality.

Test Sequence: Selected audio code rates are 16/32/64/128kbps, with 16 bit sampling bit and 44.1 KHz sampling frequency. The audios are encoded using AAC-LC. The number of stereo audio sequences and spatial audio sequences are both 12.

## Task 5

Test Resource: Use the same resource as in Task3, monoscopic videos of highest encoding quality are played to subjects

Test Sequence: The player's video delay is set to 0, 20, 60, 100, 200, 300, or 500 (in millisecond) separately. The FOV is set to 110. There are 21 test sequence in total.

## Task 6

6.1~6.2：

Test Resource: Use the same resource as in Task3 and Task5, stereoscopic videos of varied quality are played to subjects.

Test Sequence: Use the same setting as Task 3 and Task 5.

6.3~6.5：

High quality stereo 360-degree video obtained from the Internet.

# Issues and Limitations

1. The test material is limited. A majority of the number of tests use only 3 distinct video sequences. This could result in unreliable conclusion. Moreover, video content used in the preparation phrase and the rating phrase could be same, this would have an impact on the rating, since the subject has watched the same video before.
2. The provided spatial audios are low in quality. Actually it was hard for the subjects to judge from the played spatial audios where the sounds come from. Therefore, the evaluation of the spatial audios is impaired.
3. The typical durations of the video sequences last from 10 seconds to 20 seconds. Longer sequences may be worth considering.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_