|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **itu-old** | INTERNATIONAL TELECOMMUNICATION UNION | | | | **COM 9 – C xx – E** |
| **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2009-2012 | | | **December 2012** | |
| **English only**  **Original: English** | |
| **Question(s):** | | 12/9 |  | | |
| **STUDY GROUP 9 – CONTRIBUTION xx** | | | | | |
| **Source:** | | Ericsson AB | | | |
| **Title:** | | Subjective assessment of 3D video quality | | | |

# 1 Introduction

The introduction of stereoscopic 3D video has brought new experiences to viewers which may also induce new problems such as visual discomfort. Subjective assessment can help us to understand these new aspects of the 3D and their impact on the video quality and QoE. Therefore it is of high importance not only to be able to conduct subjective tests properly but also to understand what 3D video quality properties or aspects that were measured subjectively.

The purpose of the test described here was to evaluate a subjective test methodology regarding assessment of stereoscopic 3D video. Three voting scales: “Depth Naturalness”, “Video Quality” and “Visual Discomfort” were used in the test. It was conducted by a group of experts. The results of this test are consistent and indicate that it is advantageous to use more than one quality scale for 3D video quality assessment when it is expected that 3D related properties of the video varies.

This study continues with naïve subjects but the naïve group results are not included here.

# 2 Subjective study outline

This study comprises three different groups of conditions:

* Group 1 – uncompressed and encoded 2D video in full resolution and anamorphic
* Group 2 – uncompressed conditions with different levels of 3D quality.
* Group 3 – compressed conditions encoded in Side-by-Side (SbS) format at different bitrates.

One purpose of the experiments was to verify whether 3D and 2D related properties can be assessed in the same subjective test and provide consistent results.

Three voting scales were used in this test for assessment of: “Depth Naturalness” (DN), “Video Quality” (VQ) and “Visual Discomfort (feeling of…)”. The “Depth Naturalness” and “Video Quality” were evaluated according to an absolute category scale and the “Visual Discomfort” according to an impairment scale (see table below). It was expected to be easier to evaluate visual discomfort in negative terms. The “Visual Discomfort” test results were transformed and used as “Visual Comfort” (VC) later in this document for comparison beside the DN and VQ scales.

Table 1: Voting scales.

|  |  |  |
| --- | --- | --- |
| **Score** | **Absolute Category Rating scale** | **Impairment Category Rating scales** |
| 5 | Excellent | Imperceptible |
| 4 | Good | Perceptible but not annoying |
| 3 | Fair | Slightly annoying |
| 2 | Poor | Annoying |
| 1 | Bad | Very annoying |

An example of the voting GUI is shown in the figure below.

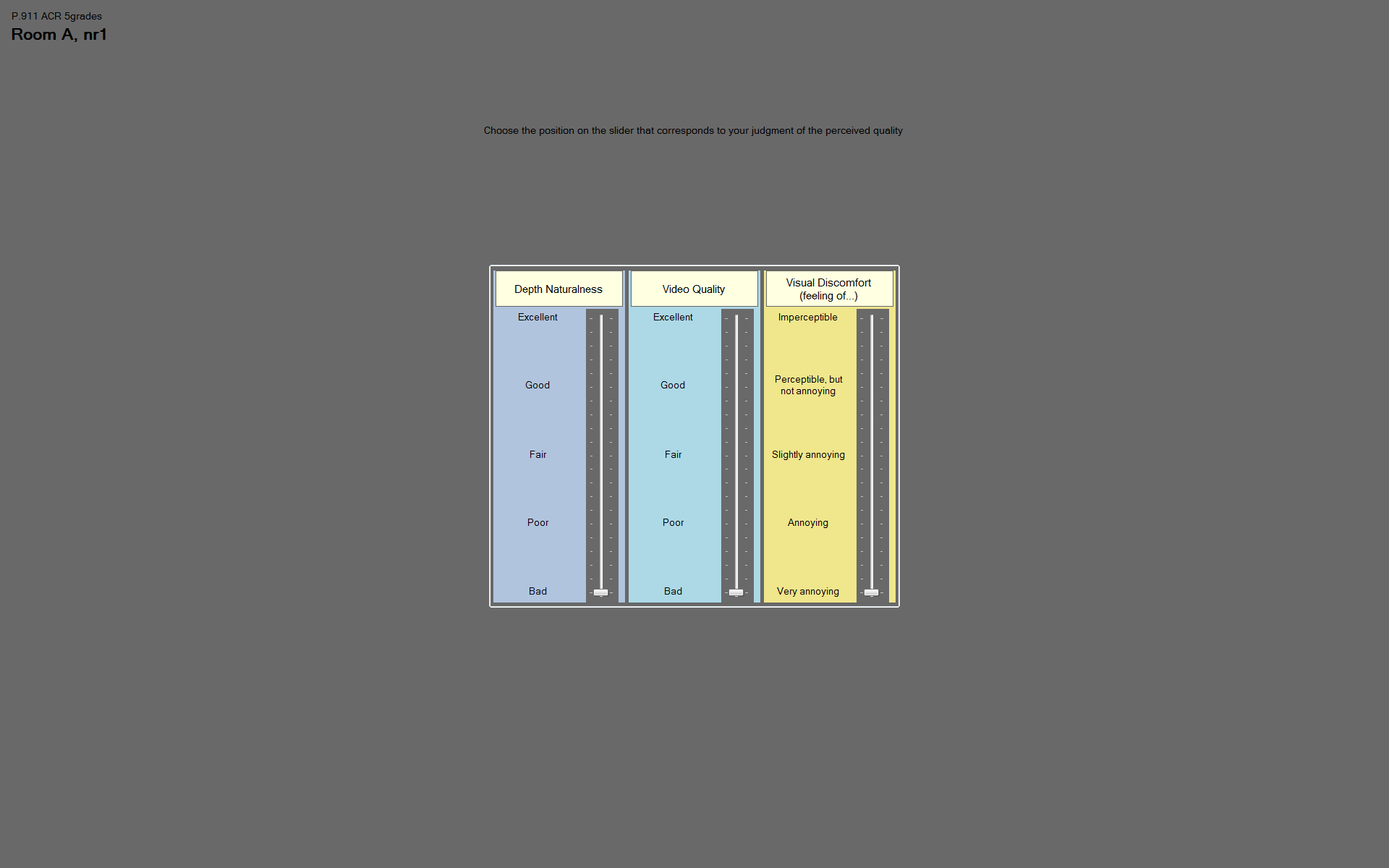


Figure 1: Voting GUI example.

# 3 Test subjects

This test was performed using 8 experts working with 3D video implementations and/or quality evaluation. All test subjects vision was tested using per Snellen visual acuity and Randot stereovision test before the test execution. A test subject was allowed to perform the test independently of the vision test results. One expert was removed from the final analysis due to bad results for stereo vision.

## 4 Viewing conditions

The test persons were placed at four screen heights (4H) distance from the screen.

4H distance was used since the Hyundai S465D display has line-interleaved pattern for separation of the left and right views and due to this the black horizontal strips are visible at 3H viewing distance for a subject with visual acuity 20/20 or better. The display was put into factory default mode by resetting the settings mentioned below. The refresh rate was set to 50Hz. The default settings for the Hyundai S465D display are: Brightness – 70%, Contrast 80% and Color – 50%. The brightness of the screen was increased to 90% i.e. 205 cd/m2 peak level

The room lighting was switched off due to use of passive glasses that limits the luminance (~40%) and hence the contrast property of the screen. The light reduction level by the passive 3D glasses was measured per eye. It was 62% for Ral3D and MasterImage and 60% for EX3D.

The lighting level from the screen at 4H distance was measured using a LX-101 lux meter and a Hagner LC-1 luminance meter was used for luminance measurements. The lighting level reaching eyes was ~8 cd/m2 (~23 lux), at 4H distance, before glasses.

Two subjects at a time performed the test, sitting beside each other.

# 5 Description of the test conditions

## 5.1 Source content

The S3D video material was extracted from available commercial blu-ray discs: one documentary and three movies. All source video was 1080p, 24/1.001 fps. It was played in 25Hz frame rate in the test.

The extracted source material was chosen with purpose to avoid scene changes and can be divided into three types:

1. Content 1 (c1) – recorded using still camera and contain small amount of motion.
2. Content 2 (c2) – recorded using still camera and contain a moderate amount of motion.
3. Content 3 (c3) – recorded using zoom and/or moving camera and contain from moderate to large amount of motion.

The properties of the selected content are summarized in the table below. Note that content c1.5 was used only for 720p25 test conditions.

Table 2: Test content types and properties.

| **№** | **Content** | **Content scene** | **Shooting camera** | **Source video format** | **Video format used in the test** |
| --- | --- | --- | --- | --- | --- |
| 1 | c1.1 | Two people standing in a bar | Almost static | 1080p24 | 1080p25 |
| 2 | c1.2 | A standing men | Static | 1080p24 | 1080p25 |
| 3 | c1.3 | Group of people sitting | Static | 1080p24 | 1080p25 |
| 4 | c1.4 | Two men (sitting and standing) | Static | 1080p24 | 1080p25 |
| 5 | c1.5 | One standing man | Static | 1080p24 | 1424x800p25 |
|  | | | | | |
| 6 | c2.1 | A group on a terrace | Handheld, some zoom at the end | 1080p24 | 1080p25 |
| 7 | c2.2 | Two standing men and a car passing by | Static | 1080p24 | 1080p25 |
| 8 | c2.3 | Two men walking towards camera | Static | 1080p24 | 1080p25 |
| 9 | c2.4 | Group of people inside, a man going towards camera | Static | 1080p24 | 1080p25 |
|  | | | | | |
| 10 | c3.1 | One man at a table | Moving (dolly) | 1080p24 | 1080p25 |
| 11 | c3.2 | Two men walking towards the camera | Moving and panning around at the end | 1080p24 | 1080p25 |
| 12 | c3.3 | One man sitting outside | Zooming in | 1080p24 | 1080p25 |
| 13 | c3.4 | Group of people (some people standing and some moving) | Handheld | 1080p24 | 1080p25 |

## 5.2 Test cases

It was decided to use 3D processing cases that produce obvious and detectable impact on the 3D scenes. Following processing cases were decided to use for this test (see Ref.[1] for more details):

1. 2D-to-3D conversion (fake 3D). Displayed video resolutions 1920x540 and 1280x720 up-scaled to 1920x540 per eye.
2. Depth gradient in vertical direction (using horizontal shear distortion). This processing resulted in different amount of depth on the upper and lower parts of the video frame i.e. object or part located in the upper part of the frame were perceived further away while objects or parts in the lower part of the frame were shifted closer. Displayed video resolutions 1920x540 per eye.
3. One frame temporal mismatch between left and right views. Displayed video resolutions 1920x540 per eye.
4. 2D video using the left view with the same resolution as for 3D depending on the test case. Displayed video resolutions 1920x540 and 720p uspcaled to 1920x540 per eye.
5. Video quality related properties were tested using compressed SbS 3D video at different bitrates (denoted in this document in increasing order as r01, r02, r03, r04, r05. The encoding was done using H.264 encoder implementation with constant bitrate control. Displayed video resolution 640x720 up-scaled to 1920x540 per eye.
6. The subsampling of the vertical vs. vertical and horizontal resolution. Video resolutions 1920x1080, 960x1080 and 640x720 were displayed as 1920x540 per eye.

# 6 Description of simulation of the 3D test conditions

This chapter describes the 3D related processing using open source video editing framework “Avisynth” and its plugins. Note that the final row-interleaving was done outside of the Hyundai display processing. The Hyundai display was used only as PC monitor.

## 6.1 2D-to-3D conversion by geometric distortion “3DF”

This simulation is denoted in this document as “3DF”.

A simple simulation of 2D-to-3D conversion was done by using a geometric distortion. The perception of depth was achieved by simulating a presence of disparity using two processing steps: up-scaling horizontally and cropping different parts of the up-scaled frame: right part for the left and left part for the right view. The disparity between views and hence 3D feeling is created due to horizontal geometric distortion coming from horizontal up-scaling.

A processing example is depicted in the figure below.



Figure 1: 2D-to-3D conversion using geometric distortion.

## 6.2 Uneven depth in vertical direction “3DHS”

This simulation is denoted in this document as “3DHS”.

This type of depth unevenness was simulated by creating a gradient geometrical distortion of the left and right views. The processing steps are showed in the figure below. The horizontal shear distortion was applied to both left and right views by shifting pixels in opposite directions. Pixels in the upper half of the left view frame were shifted to left while pixels in the lower half were shifted to the right. The opposite shifting was performed for pixels in the right view. Thus, the left view was distorted counter clockwise while the right view was distorted clockwise.

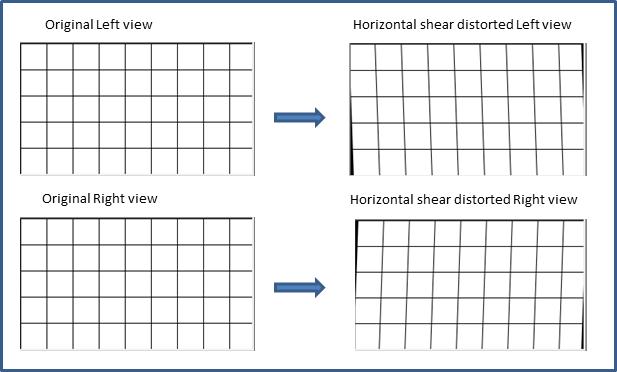


Figure 2: Horizontal shear distortions applied to the left and right views.

Two views, displayed together, create a disparity gradient resulting in different depth perception in the lower and upper parts of the 3D frame. The upper part of the displayed 3D scene was perceived to be further away than the lower part. The horizontal disparity gradient is shown in the next figure below.

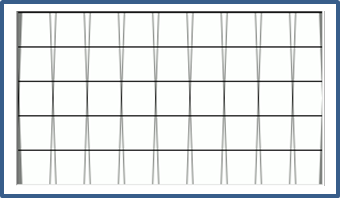


Figure 2: Disparity gradient created by opposite horizontal shear distortions of the left and right views.

## 6.3 Temporal mismatch “3DP”

This simulation is denoted in this document as “3DP”.

Temporal mismatch is a temporal misalignment between left and right views. This type of misalignment is usually very small during capture but can arise during or after depending on the post-processing of the 3D video or network loss. Typical and expectable cause is packet loss due to e.g. transmission and subsequent error concealment techniques that e.g. missed that one frame should be deleted from one view when the corresponding frame in another view is lost or damaged.

The temporal mismatch between left and right views was achieved by removing one frame at the beginning of the left view. The temporal mismatch enhances the motion parallax (called also the Pulfrich’s phenomenon) and can create the depth illusion in the video scenes. It was also observed during the test preparation that motion present locally (e.g. hand or eye movements) can be interpreted as video quality distortion.

# 7 Test results

## 7.1 2D video test conditions

The results for 2D video conditions are showed in the figure below. It was expected that only video quality would vary depending on the resolution and applied encoding, which also was the result.

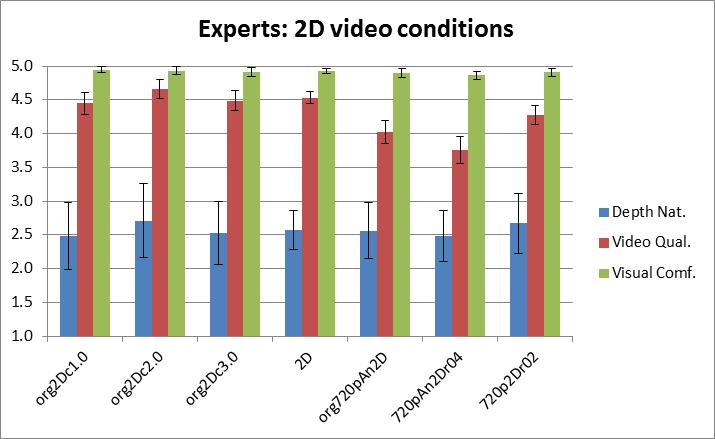


Figure 3: MOS results for 2D video.

Following test conditions results are presented in the figure above from left to right: MOS for different content types (“org2Dc1.0”, “org2Dc2.0” and “org2Dc3.0”), the average MOS over all content types “2D”, uncompressed anamorphic 720p “org720pAn2D”, the same anamorphic 720p but encoded at bitrate r04 “org720pAn2Dr04” and full 720p encoded at bitrate r02 (r02 < r04). Note that “org2Dc1.0”, “org2Dc2.0”, “org2Dc3.0” and “2D” had resolution 1080p and was displayed on the Hyundai as 1920x540 per eye while 720p videos were upscaled to 1080p and also displayed as 1920x540 per eye on the screen.

One can observe in the expert results that “Visual Comfort” (VC) and “Depth Naturalness” (DN) were voted to be rather constant as it was expected. The video quality was judged be lower for anamorphic “org720pAn2D”, compressed anamorphic “720pAn2Dr04” and compressed plain 2D “720p2Dr02”. The anamorphic compressed video got the lower MOS than the encoded 720p despite the fact that bitrate r04 was higher than r02 due to higher video resolution.

One can state based on the results described above that 2D video could be evaluated using only one scale “Video Quality” as it was expected.

## 7.2 Uncompressed 3D video test conditions

The test results for the uncompressed 3D video conditions are showed in the figure below.

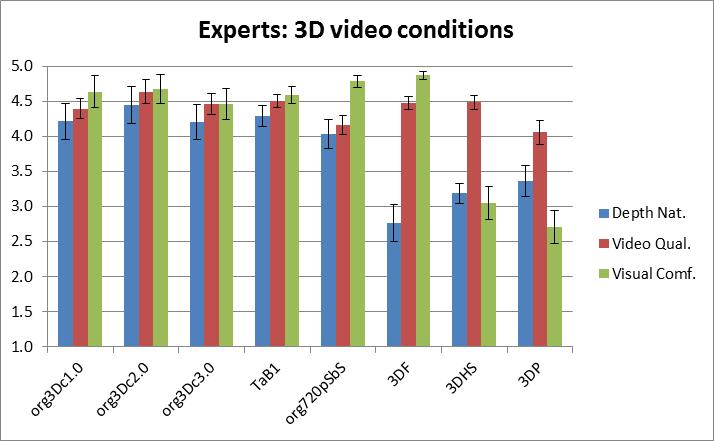


Figure 4: MOS results for the uncompressed 3D video.

Following test conditions are presented in the figure above from left to right: MOS for the content types (“org3Dc1.0”, “org3Dc2.0” and “org3Dc3.0”), the average over all content types “TAB1”, uncompressed 720p SbS “org720pSbS”, the 3D simulation “3DF”, depth distortion simulation “3DHS” and temporal mismatch simulation between left and right view “3DP” (left view one frame ahead of the right view). Note that all conditions except one had resolution 1080p and were displayed on the Hyundai as 1920x540 per eye while 720p SbS videos were up-scaled to 1080p and also displayed as 1920x540 per eye on the screen.

It was expected that video quality would reflect the lower video resolution for SbS conditions and the temporal mismatch simulation. It was observed during the temporal mismatch preparation that motion present locally in the 3D video scene (e.g. eyes and hand movements) can be perceived also as video quality degradation.

It can be observed in Figure 4 above that simulated 3D was detected and resulted in the lowest DN MOS. The depth distortions present in “3DHS” and “3DP” conditions were reflected by low MOS for both DN and VC quality scales. The “3DP” simulation resulted also in lower VQ MOS than e.g. undistorted “TaB1” as it was expected. The 3D video condition “org720pSbS” with lower native resolution was also judged with lower VQ MOS than “TaB1” that had higher horizontal resolution as one can expect. Note that “org720pSbS” MOS comprises smaller number of sequences than “TaB1” MOS.

## 7.3 Encoded SbS 3D video

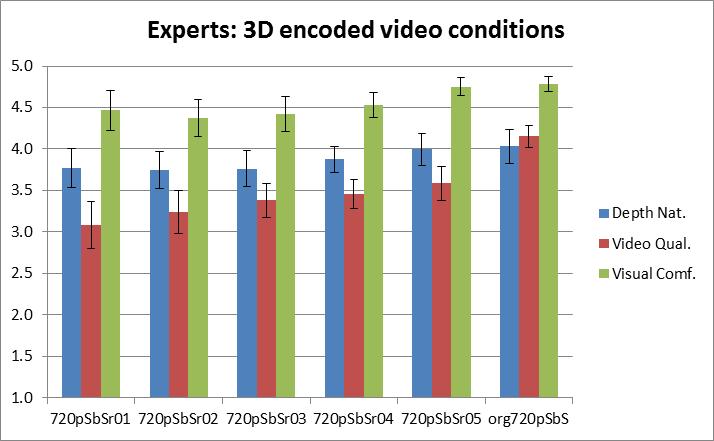


Figure 5: MOS results for compressed SbS 3D video.

Following test conditions are presented in the figure above from left to right: encoded SbS 3D videos at bit rates r01, r02, r03, r04, r05 where r01 is the lowest and r05 is the highest bit rate (“720pSbSr01”, “720pSbSr02”, “720pSbSr03”, “720pSbSr04” and “720pSbS2.5”) and uncompressed 720p SbS “org720pSbS”. Note that all 640x720 (720p SbS) resolutions were up-scaled to 1920x1080 and displayed as 1920x540 per eye on the screen.

The up-scaling magnifies the compression and downscaling artefacts. Therefore it was expected that test conditions would be differentiated using the “Video Quality” (VQ) and reflected by VQ MOS than “Depth Naturalness” (DN) and “Visual Comfort” (VC) MOS. One can notice that different quality conditions were reflected mostly by VQ MOS but also by DN and VC MOS for the highest bitrate (r05) and uncompressed (“org720pSbS”) conditions. That indicated that higher video quality enhance 3D perception as one can expect.

## 7.4 MOS variations and voting scales correlation

The standard deviations of the MOS per scale and Pearson correlation between scales were calculated for MOS results shown in the previous chapters. The calculated results for all 2D and encoded 3D video test conditions are shown in the figure below.

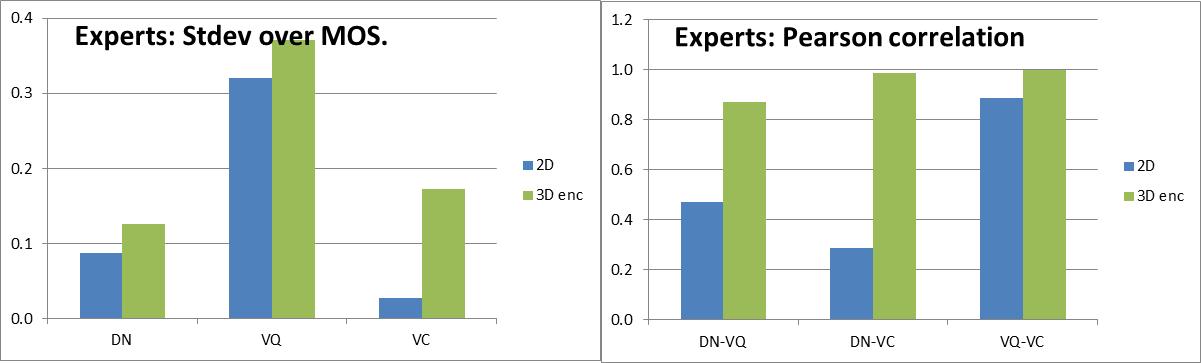


Figure 6: Standard deviation and Pearson correlation coefficient over MOS per quality scale.

For 2D video, the Pearson correlation is low for DN-VQ and DN-VC combinations and rather high between VQ and VC. At the same time one can see in Figure 6 above that standard deviation for VC MOS is very low (0.03) which indicates that VC was almost constant taking into account 95% confidence intervals that were between 0.03 and 0.07. Therefore, one can state that the expert MOS results were exactly as one would expect when only video quality varied which was reflected mostly by VQ MOS results.

For the encoded 3D video, the highest correlation for VQ-VC is similar to the 2D video groups’ correlation and this similarity can be explained by presence of the compression artefacts in the encoded 3D video more typical for 2D. At the same time the standard deviation for VC MOS is small (0.17) similarly to 2D video. The confidence intervals for VC were between 0.09 and 0.24 for expert viewers. The experts have also very high correlation between DN and VC. It seems that the experts combined the DN and VC. At the same time the standard deviation for DN MOS values are rather low (0.13) and smaller than confidence intervals obtained for MOS values on this scale. The experts have CI for DN between 0.16 and 0.21. This indicates that DN can be considered rather constant as one would expect since no depth distortions simulation were applied for the discussed conditions. Thus the DN and VC varied but in a very small range. The larger standard deviations were obtained for VQ MOS.

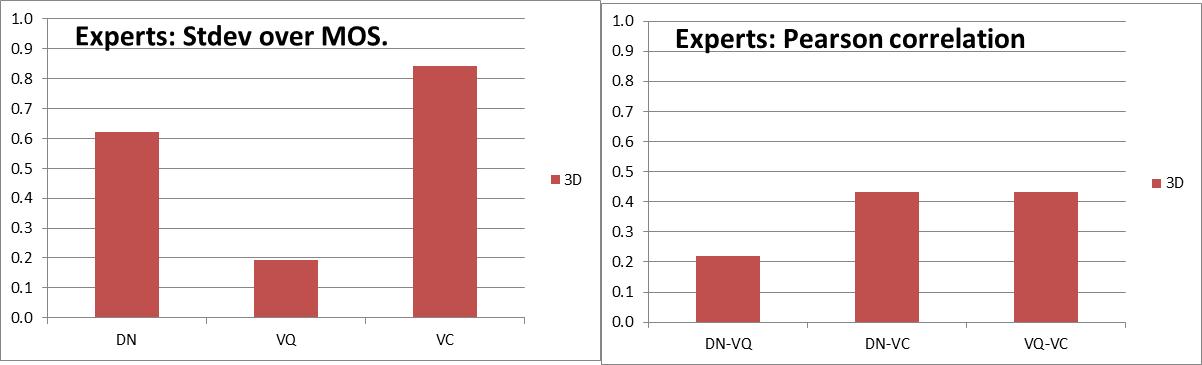


Figure 7: Standard deviation and Pearson correlation coefficient over MOS per quality scale.

For the uncompressed 3D video, the Pearson correlations are low between all scales which indicated that experts used all three scales during the quality assessment mostly independently from each other as one would expect. One can state that they separated the DN, VQ and VC quality concepts. The standard deviations for MOS per voting scale are rather large (larger than 95% confidence interval). The smallest standard deviation was obtained for VQ scale (0.19) while the confidence intervals for VQ were between 0.09 and 0.18. The video quality varied due to use of “3DP” simulation and SbS 3D video with lower resolution than “TaB1”. The DN and VC MOS varied in larger range than VQ as it was expected.

# 8 Conclusions

It was expected that different quality properties of the test conditions would be graded differently on the different scales used in the test. The video quality related to the traditional 2D video quality concept was expected to be reflected by VQ MOS results. The DN MOS results were expected to be related to the 3D video simulated depth distortions.

The test results analysis revealed that variation of the video quality in 2D and 3D video was reflected by VQ MOS results as it was expected. It was detected that VQ and VC scales had higher correlation when the simulated depth distortions were not present in the stimuli. The experts used exclusively the VQ scale for such conditions. This test showed that the VQ scale would be enough for 2D video quality assessment. At the same time it was detected that degradation of the video quality in 3D videos due to compression was reflected on VC and DN though the VQ MOS varied more for different bitrates. Therefore one can state that all three scales are needed for 3D video quality assessment since the possible impact of the degradations on the quality of 3D video is not always easy to predict. Artefacts in 3D videos sometimes cause different types of degradations simultaneously, both depth and video quality related.

Video compression, which is usually considered as video quality related degradations typical for 2D, can also affect the 3D video perception by creating disparities not typical for a presented 3D scene.

The simulation of the depth distortions were reflected exclusively by DN and VC MOS results as it was expected. The 2D and simulated 3D videos were recognized by the expert subjects and reflected by low DN but high VC MOS values.

The use of 2D and simulated 3D can be considered as useful anchors for 3D video quality assessment.

The study continues with verification of these results with naïve subjects.