# Testing QoE in Different 3D HDTV Technologies

*Martin SLANINA, Tomáš KRATOCHVÍL, Václav ŘÍČNÝ,*
*Libor BOLEČEK, Ondřej KALLER, Ladislav POLÁK*

Dept. of Radio Electronics, Brno University of Technology, CZ-61200 Brno, Czech Republic

slaninam@feec.vutbr.cz

**Abstract.** *The three dimensional (3D) display technology has started flooding the consumer television market. There are a number of different systems available with different marketing strategies and different advertised advantages. The main goal of the experiment described in this paper is to compare the systems in terms of achievable Quality of Experience (QoE) in different situations. The display systems considered are the liquid crystal display using polarized light and passive lightweight glasses for the separation of the left- and right-eye images, a plasma display with time multiplexed images and active shutter glasses and a projection system with time multiplexed images and active shutter glasses. As no standardized test methodology has been defined for testing of stereoscopic systems, we develop our own approach to testing different aspects of QoE on different systems without reference using semantic differential scales. We present an analysis of scores with respect to different phenomena under study and define which of the tested aspects can really express a difference in the performance of the considered display technologies.*

## Keywords

3D, stereoscopic display, Quality of Experience, subjective test, reference-free methodology.

## 1. Introduction

A strong motivation for the recent video quality research was the approach of digital processing, compression, storage and transmission systems, which introduce completely different phenomena compared to analog systems. In the area of video delivery and display systems, the common term Quality of Service (QoS) does not suffice to get a good enough description of the system performance. QoS is understood as a technical measure related to objective performance of a system. As the consumer of the video system output is a human observer, QoS obviously lacks sufficient evidence regarding the user's quality perception [1]. In contrary, Quality of Experience (QoE) is a measure that involves the subjective factors of the user. As such, QoE is the only correct measure of the real system performance. We can find two definitions if QoE in literature [2] and [3]: QoE is "the overall acceptability of an application or service, as perceived subjectively by the end-user" [2]; QoE is "A measure of user performance based on both objective and subjective psychological measures of using an ICT service or product" [3].

As such, there is more than just the technical properties that has to be taken into account. In order to address QoE in different aspects, the standardized procedures for measuring performance of television and multimedia systems as found in [4], [5] and [6] are not sufficient – representing the quality with one single rating value is simply too constraining. For this reason, a different test methodology has been developed and used in the experiment described in this paper. The methodology aims at exploring several aspects related to QoE in 3D: 3D effect intensity, depth, presence, visibility of impairments, sharpness, discomfort, and ambient light effects for three different display systems using different types of displayed content.

The recently published research papers address mostly the impact of impairments, transmission errors or degradations on the QoE in 3D scenarios – e.g. [7], [8], [9]. A very important factor influencing the Quality of Experience in viewing stereoscopic movie content is, however, also the display system that is used for the presentation. One of the most important parameters of 3D display systems is their ability to suppress crosstalk between the left- and the right-eye image [10]. Still, there is a number of other phenomena that objectively influence the perceived 3D sensation. The most important technical prerequisites are mentioned in [11] and [12].

The display technology has gone a long way in recent years and there are a number of different products available in the market utilizing different tools to provide stereoscopic viewing sensation. Whereas for the two dimensional television scenario, several parameters defining the reference setup of a display monitor suffice for an overall acceptable definition of display part of the system, this is not the case for the stereoscopic viewing scenario. No standard methodology has been defined for the stereoscopic quality testing yet. The Video Quality Experts Group is working towards performing tests including measurement of crosstalk influence on user QoE. In the ITU-T Work Programme [13], we can find that ITU-T works towards standardization of stereoscopic quality assessment in two directions: Firstly, the Study Group (SG) 9 under Question 9 woks on the “Display requirements for 3D video quality assessment” which should result in an ITU-T Recommendation in 2012. Under Question 12, the same SG intends to define “Subjective assessment methods for 3D video quality” in 2012. The details of the progress are, however, not available to the public.

The contribution of this paper can be seen in several levels: Firstly, we define a *test methodology* for assessment of Quality of Experience in the 3D scenario with the motivation to address several aspects of the viewing experience. Secondly, we *compare the different display technologies* according to the test results for each of them. Further, we provide a set of diverse *stereoscopic video sequences*. Finally, we provide a short analysis of how *changing ambient illumination affects the results* for each of the display technologies.

The rest of the paper is organized as follows. In Section 2, we introduce the test methodology used during our experiment, including the organization of the whole test, the presentation scheme, the content, the playback setup as well as the observers who took part in the experiment. Section 3 provides an analysis of the results. The paper concludes in Section 4.

# 2. The Test Methodology

During our tests, we used a methodology based on post-presentation rating of several aspects in the sequences under test. The presentation scheme itself was based on the standard procedures as found in e.g. ITU-R BT.500 or ITU-T P.910 [4], [5]. Both of these Recommendations describe methodologies for full reference (double stimulus, comparative) assessment of video/image quality and no reference (single stimulus) assessment. The full reference approach is, however, not appropriate in our situation. Since the purpose is to compare whole systems including the displays, one comes to difficulties in constructing the full reference test sessions - to ensure proper viewing angles, distance, ambient illumination and to prevent distractions, it is very inconvenient to assess the perceived quality on two systems in fast succession provided the systems require special (and different) stereoscopic glasses. In addition, the ability of revealing small differences through double stimulus tests is paid by lengthiness of the test sessions. The no reference approaches are thus the basis of the test methodology used in our experiment.

According to the above mentioned recommendations, short video clips shall be presented sequentially. After an (approximately) ten seconds long sequence, a mid-gray frame shall be displayed for several seconds. During this interval, the user ratings are recorded. We intentionally do not present the exact lengths at this point as they differ in the two recommendations and need tuning for the specific test purpose.

## 2.1 Organization of the Test

At the beginning of the testing session, the observers were instructed about how to provide their answers. Firstly, they were asked to fill in a one page questionnaire regarding their personal data – such as age, the amount of previous experience with 3D display technology, any known vision defects, etc. We did not test the subjects for vision defects; in our previous studies (e.g. [14]) we found no difference between the results of such tests and vision defects reported directly by the observers in interviews. After this, the observers were asked to fill in a short questionnaire after viewing a set of eight randomized video sequences. Consequently, another set of eight sequences was played back to them, and so on. Four sets (eight sequences per set) were played back in total. At the end, the observers answered three questions regarding the test as a whole. The organization of the test is shown in Fig. 1. When finishing the whole test, the subjects were asked to repead the same for a different visualization system (LCD/plasma/projection). The order of the systems was kept pseudo random in order to avoid systematic influence of the ratings.

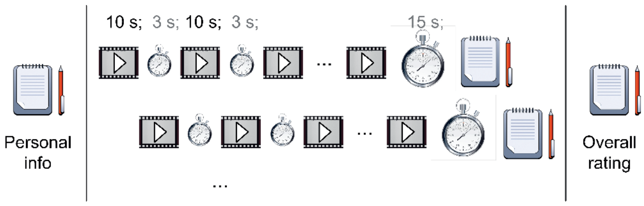



**Fig. 1.** The test session.

There were several motivations to using questionnaires after the sequences have been played back: Firstly, we decided to avoid gathering scores during the playback of each sequence as we found such results biased in our previous work [14]. Secondly, we aimed at gathering more specific ratings than the standard methodologies describe – instead of describing the overall quality (or level of impairments) in one score, the users were asked several questions regarding different qualities of the presented material. As the number of questions to answer increases, it becomes difficult to provide ratings for each of the sequences separately. The scores are thus collected for each source content type (Blu-ray, satellite, camcorder, static images). A similar approach of post-presentation questionnaires was used by e.g. Staelens et al. in [15].

As we needed to judge several aspects of the stereoscopic video presentation, we refrained from using the classical continuous or categorical scales as used in [4] or [5], for instance. Instead, we used 7-point *semantic differential scales* [16]. The semantic differential scales employ pairs of opposite adjectives at either end of a scale. In such technique, a very important issue is to select the appropriate terms that are truly opposite and reflect the property that should be addressed. Usually, the scales are 7-point or 5-point. We selected the 7-point scales in order to increase the discriminative power of the methodology. It is known that the users often avoid using the extreme values of the scales, thus having a 7-point scale has a better ability to capture the nuances between slightly different opinion scores.

The scales used throughout our experiment are summarized in Tab. 1. In order to minimize bias in the ratings, the questions were organized in such a manner that for every following question, the answer corresponding to the best system performance was found on the opposite side of the scale (except for question 2). As all the observers were all fluent in Czech and most of them were native Czech speakers, the questions were in Czech language; Tab. 1 provides the English translation of the questions and the differential scale terms. In question 4, the observers could also specify the type of impairments (blur, overlapping images, false contours, opaque areas or other). Similarly, in question 6, the users could specify the uncomfortable feelings they experienced (if any) - dizziness, headache, eyestrain, other. In this paper, these specific selections are not considered.

| | Question | from | to |
|---|---|---|---|
| 1 | How intensive is the 3D effect? | imperceptible | very intensive |
| 2 | Judge the depth of the scene. | too low | too high |
| 3 | Did you feel like being a part of the scene? | absolutely not | intensively |
| 4 | Did you notice impairments / artifacts in the scene? | absolutely not | strong impairment |
| 5 | What is the sharpness of the scene? | very low | very high |
| 6 | Did you experience any uncomfortable feelings? | absolutely not | strongly |
| 7 | Did you feel disturbed by ambient light? | intensively | not at all |

**Tab. 1.** Semantic differential scale terms used in the test.

For successful completion of the rating session, also the graphical layout of the questionnaire is important. Fig. 2 shows a part of the layout (translated to English).

1. How intensive is the 3d effect?

imperceptible ❑ ❑ ❑ ❑ ❑ ❑ ❑ very intensive

2. Judge the depth of the scene (natural = center).

too low ❑ ❑ ❑ ❑ ❑ ❑ ❑ too high

**Fig. 1.** Layout of the questionnaire.

## 2.2 The Video Content

In order to cover a variety of different source formats, we used four different sources of video sequences throughout the test:

(1) Blu-ray disc records. In the original format of the 3D Blu-ray disc records, each video frame is composed of two full resolution frames at the decoder output – this is the highest spatial and temporal resolution available. Originally, the sequences were encoded using the multi-view extension of H.264 - MVC. The original resolution was 1920x1080, progressive, 24 frames per second. We have created several short uncompressed sequences covering different content with different spatial and temporal dynamics.

(2) Satellite reception. The original sequences were recorded from the Astra 3D DVB-S2 demo channel. Compared to Blu-ray, the format is very different: the multiplex contains interlaced video with resolution 1920x1080, 25 frames per second. Each frame is composed of two sub-frames: the left half represents the left-eye picture while the right half represents the right-eye picture (the composition is called the side by side format). In such manner, one complete frame with full HD resolution covers both images and it is up to the display to separate the two images and display them in correctly. The real resolution is thus reduced in both the vertical (interlacing) and the horizontal direction (reduction of image width in the side by side format).

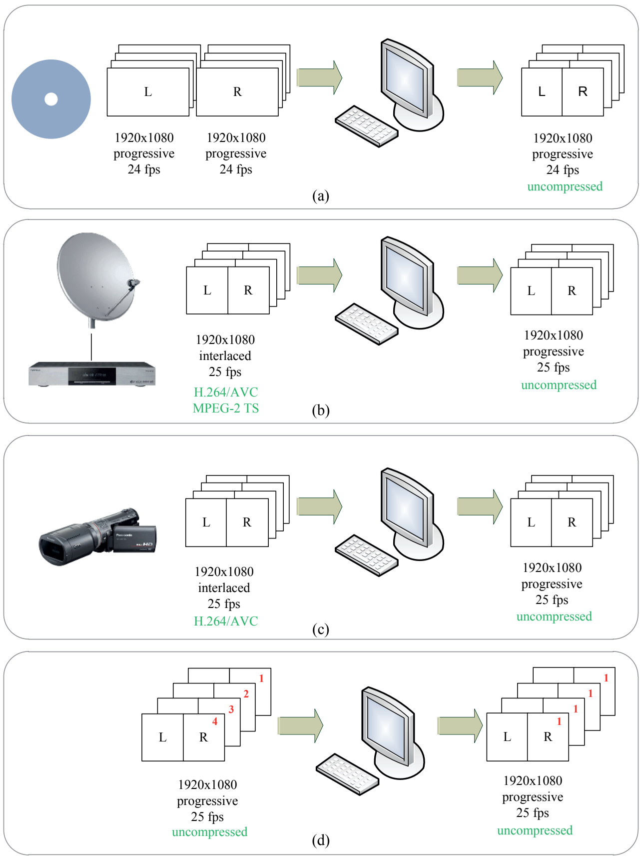



**Fig. 3** The conversion of different source formats to a common uncompressed video format for playback during the test session: (a) Blu-ray disc records; (b) satellite reception; (c) camcorder shots; (d) static images.


**Fig. 4.** Video sequences used throughout the test: : (a) Blu-ray disc records; (b) satellite reception; (c) camcorder shots; (d) static images.

(3) Camcorder recordings. This group of sequences was created using a consumer-level stereoscopic camcorder. This solution extends an ordinary camcorder by adding a lens that assures that each the right-eye and the left-eye images are captured by distinct parts of the camera chip. In principle, this reduces the usable resolution by one half, similarly to the satellite format. In addition, to account for imperfect mounting of the stereoscopic lens, the camcorder inserts borders around each of the images (approx. 5 % the image size). The resulting format is side by side video with resolution 1920x1080, 25 fps. The camcorder, as well as the satellite broadcasting, uses interlaced video, so the video frame resolution suffers in the vertical direction (due to interlacing) as well as in the horizontal direction (due to side by side format).

(4) Still images. Out of the three aforementioned content types, we created a set of still image video sequences just by keeping a single frame of a 3D video sequence. To assure fluent playback of the video content, all the sequences were converted to side by side format, progressive, 1920x1080. As a consequence, the quality of Blu-ray sequences was degraded as their horizontal resolution was downsampled to one half. We kept the frame rate mismatch (24 vs 25 fps) unattended as such difference is unlikely to be noticed by the observers. The conversion process is summarized in Fig. 3. Frames representing the final sequences in each set (only the left image in each stereo pair) are shown in Fig. 4.

We have tried to cover video content of different characteristics especially in terms of distance of the captured objects, the amount of motion and scene brightness. We assume these parameters play an important role in overall stereoscopic perception. Among the Blu-ray disc records (Fig. 4a), we can find two scenes from the movie Avatar – one is very dynamic with a helicopter flying over a lake while the other is a dialogue with artificial surroundings and moderate lighting and a low amount of motion. Another two scenes originate in the FIFA 2010 movie, from which we have selected one scene with distant view of the crowd of fans and a close view of an offensive action. Especially the latter scene contains a lot of motion. The two following scenes are an example of a feature film – they were recorded from the movie Green Hornet. One scene is a dialogue while the other is a car chase with extremely high amount of motion with close camera views and moderate lighting. The last two Blu-ray sequences come from the documentary film Ocean Wonderland – the first one displays a cloud of very small fish on dark background while the latter shows a number of fish around a coral reef.

Fig. 4b displays the first frames of the sequences recorded from satellite broadcasting, namely from the Astra 3D demo channel. They contain sequences with different lighting levels (low lighting in theatre play and concert scenes, higher lighting levels in the scene with children playing in front of a windmill, a look over rooftops of a city and a view of a sculpture. The camcorder sequences in Fig. 4c represent a typical set of scenes a tourist shoots on holiday; we have included exterior scenes with architectural views, a tourist bus passing, and a shopping street. The interiors are represented by a marketplace and hotel foyer. The last sequence is a close view of people in a swimming pool. Overall the satellite sequences have quite low amount of motion. The final set of sequences, the static images, is shown in Fig. 4d. These scenes are created from single frames of the sequences described above– they include three Blu-ray scenes, three satellite scenes and two camcorder scenes.

All the sequences that are not subject to copyright issues can be found in the DEIMOS (Database of Images – Open Source) [17]. All other sequences are accessible upon registration and request to the database administrator.

## 2.3 The Playback Setup

As mentioned above, the display properties for quality testing of stereoscopic systems are not yet standardized. In our setup, we used the total of three different displays: Firstly, a LCD display using circular polarization of light and passive glasses for separation of the left- and right-eye images (LG 32LW570S). The second type of display was a plasma display using temporal multiplexing and synchronized active shutter glasses for the separation of both images (Panasonic TX-P42GT20E). Finally, the third solution utilized a projector with high enough refresh rate (Benq W710ST) and nVidia 3D Vision system with time multiplexing active shutter glasses.

As the displays had different screen sizes (LCD 32” diagonal, plasma 42”diagonal, projection 80”diagonal), instead of fixing the absolute distance to the display, we chose to control the viewing distance as four times the image height. There were two devices used for playback. A home media centre (xStreamer Ultra) equipped with fast solid state hard drive fed the LCD and Plasma television screens. The projector was fed by a personal computer with stereoscopic player and nVidia 3D Vision transmitter. A schematic diagram of the playback setup is given in Fig. 5. The luminance of the LCD and plasma displays was adjusted to 200 cd/m$^2$ (peak luminance when white area is displayed). The peak luminance at the projection screen was adjusted to 150 cd/m$^2$, i.e. the highest practically achievable value on the projection screen that we used.

One of the aims of the test was to examine the behavior of the different display technologies in different levels of ambient illumination. Throughout the test, we used various levels of ambient illumination between 20 lux and 200 lux. For the ambient light, fluorescent lamps were used.

## 2.4 The Observers

To properly assess the performance of the different stereoscopic display systems, it is advisable that the users are able to concentrate on the presentation extremely well and have no vision problems. To account for this, the observers for our test were mainly recruited from university students. They had no prior experience in video quality tests and can be considered naive observers. Another, much smaller, group of observers consisted of elderly people taking part in the University of the Third Age program. Unfortunately, this group provided only few results.

The first part of the questionnaire that each participant filled in asked about personal data, such as age, gender, TV and 3DTV experience, and vision deficiencies, whether corrected or not. After disqualifying the observers who reported serious vision faults (e.g. low color sensitivity, amblyopia, etc.), we gathered the total of 126 questionnaires that were used for further analysis of the results.

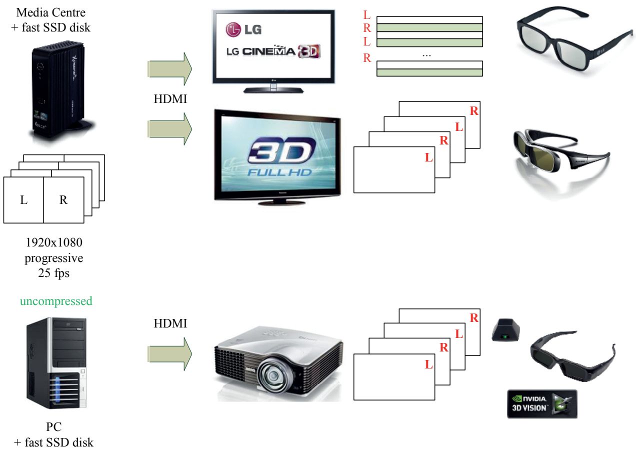
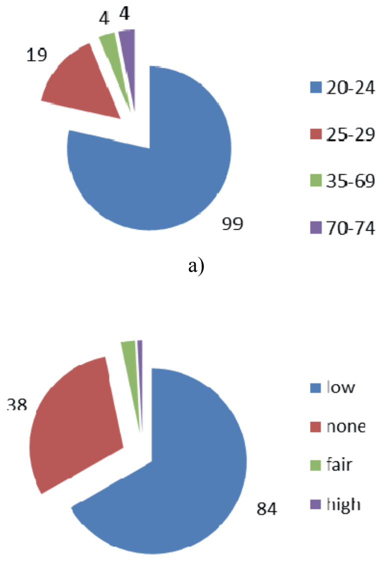



**Fig. 5** The setup of playback during the test session.




a)

b)

**Fig. 6** a) The age composition of the observers and b) their previous experience with the 3D display technology (bottom). Low experience means that the observers have only seen a stereoscopic image several times but do not own a 3D TV set. Fair experience corresponds to the users who own a 3D TV set but rarely watch 3D content.

The age composition of the observers is displayed in a pie chart in Fig. 4. We also provide the reported amount of previous experience with the 3D display technology. We can see that the stereoscopic display technology, although currently massively offered in the market, is not widespread among the tested sample of observers – only one observer reported high experience with the technology (owned a 3D TV set and was used to watching 3D content) while just 3 observers reported that they had fair experience with 3D television technology.

# 3. Analysis of Results

In this section, the received scores will be analyzed in order to provide a comparison of the display technology used. We are going to describe the results of each question that the observers were asked and visualize the results in a bar graph. All the results have been adjusted in such a manner that the higher the value, the better the performance of the system. One exception is question 2, where the ideal performance corresponds to the value 4 (the perceived depth is natural).

## 3.1 Mean Opinion Scores

In order to test for the significance in differences among the user ratings in different situations, we can use the two-sample Student t-test; such procedure tests the hypothesis that the difference between two sets of measurements in two populations is insignificant (the mean value of the two populations is equal).

The 7-point scales that we used in the questionnaires are not easily understandable to the scientific community of multimedia quality as all the relevant recommendations use 5-point scales (or, seldom, 10- or 11- point scales). To account for this, the plots have secondary scales in the right part together with a dashed grid which makes representation of the data clearer for professionals used to 5-point scales. The linear mapping function transforming the scores into the 5-point scale is defined as

$$S_5 = \frac{2}{3} S_7 + \frac{1}{3}$$

where $S_7$ is the score at the 7-point scale and $S_5$ is the score mapped to the 5-point scale. It should be noted that such transformation only makes sense for mean opinion scores as they are represented by real numbers in the given interval while the original scores collected from the observers have categorical values.

Fig. 7 shows the bar graph representation of the mean opinion scores calculated for question no. 1: How intensive is the 3D effect. The differences between the respective display technologies are very small and the two sample t-test shows that there is no statistically significant difference between the means at the 95 % significance level. What we can observe is that there is a much bigger difference between the source types of the video sequences that we used. It could be expected that the Blu-ray and satellite sequences have the potential of providing the most intensive 3D effect as they come from professional production. Even the static images, which are just cut from the video sequences, represent a more intensive 3D sensation than the consumer level camcorder records.

Fig. 8 presents the mean opinion scores corresponding to question 2: Judge the depth of the scene. In the ideal case, the scores should be around 4 as this value means that the depth of the scene is natural. Since the natural depth is at the zero value, we can clearly see that the performance of Blu-ray records is very close to ideal. Another interesting fact is that for all the systems and sequences under study the depth was either natural or low, but never unnaturally high. For this question, the t-test revealed no difference among the display systems under test. Still, we can see that the camcorder video sequences suffer from low perceived depth compared to all other content. The reason for such low scores of the camcorder sequences in these first two questions is probably caused by the construction of the camcorder Panasonic HDC SDT 750 that was used to capture the sequences, especially the stereo base; the camcorder uses a special lens that assures that each of the images of the stereo pair is captured by one half of the chip. The stereo base of the lens is, however, only 12.5 mm.

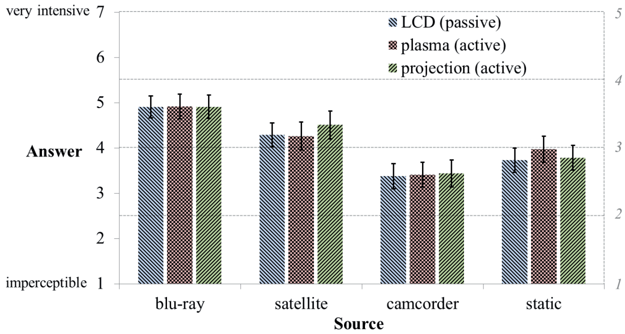



**Fig. 7.** "How intensive is the 3D effect?" The mean scores and 95 % mean confidence intervals for different displays and different source content types.

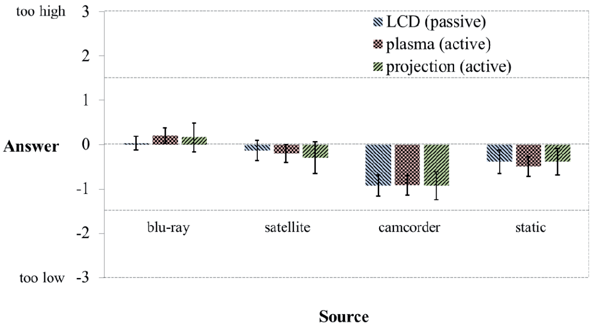



**Fig. 8.** "Judge the depth of the scene" The mean scores and 95 % mean confidence intervals for different displays and different source content types.

In Fig. 9 the mean opinion scores representing the answers to question 3: Did you feel like being part of the scene, are plotted. Overall, the performance of all systems is below average. Applying the two sample t-test on the results for different display technologies, no difference between the mean scores is found at the 95 % significance level. Even at this point the camcorder video sequences show the worst performance among all the source content types.

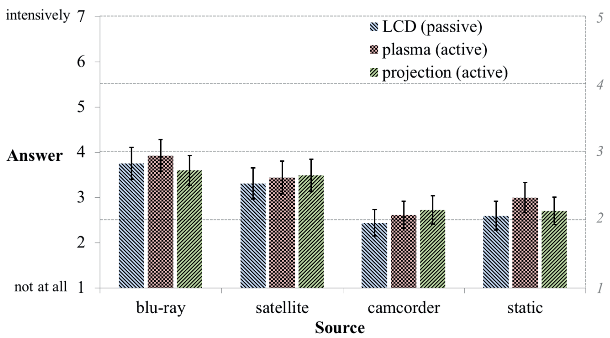



**Fig. 9.** "Did you feel like being part of the scene?" The mean scores and 95 % mean confidence intervals for different displays and different source content types.

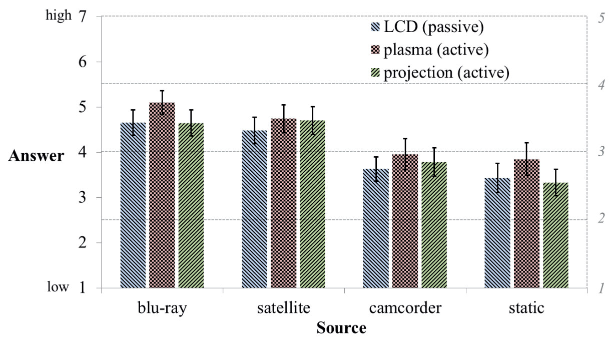



**Fig. 11.** "What is the sharpness of the scene?" The mean scores and 95 % mean confidence intervals for different displays and different source content types.

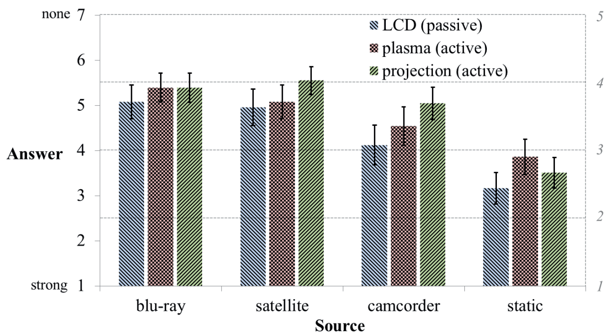



**Fig. 10.** "Did you notice any impairments?" The mean scores and 95 % mean confidence intervals for different displays and different source content types.

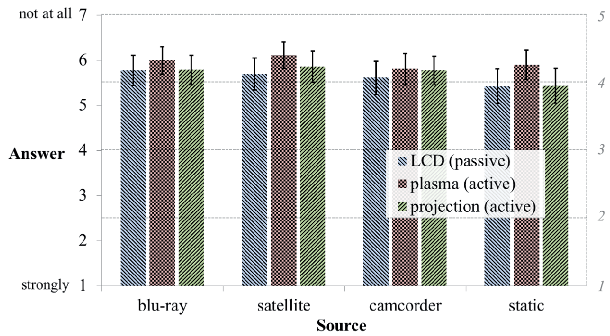



**Fig. 12.** "Did you experience any uncomfortable feelings?" The mean scores and 95 % mean confidence intervals for different displays and different source content types.

We can observe a different situation compared to the previous questions while looking at the plots of mean opinion scores corresponding to question 4: Did you notice any impairments, see Fig. 10. Performing a two sample t-test for results corresponding to the three display technologies, we find out that difference between the means is proved. At the 95 % significance level, the scores corresponding to the LCD display system with passive polarized glasses exhibit lower mean compared to the other two system. No difference between plasma display and the projection system is proved. Please note that even though the mean confidence intervals overlap especially in case of LCD and plasma display, the t-statistic using two sample t-test reveals difference in the expected mean (for explanation, see [18]). One explanation of these results could be that the passive display suffers from serious degradations of the image when the observers are not in the ideal viewing axis. During the test, the observers were within 15 degrees of the normal to the display in both the vertical and horizontal direction which is very close to the limits especially for the LCD display.

Fig. 11 displays the mean opinion scores for question 5: What is the sharpness of the scene. The t-test shows that the results for plasma display with active shutter glasses are significantly better than for the other two systems. In fact, the plasma display allows for displaying the highest resolution of the video frames among the three systems under test. The LCD display is handicapped in case the side by side format is used for the source video sequences. Such format natively reduces the effective resolution to one half in the horizontal direction. What is more, the polarized LCD display devotes one half of the display lines to the left-eye image and the other half to the right-eye image. As such, the effective resolution is just quarter full HD. The plasma display does not suffer from reduction in the vertical direction, thus the effective resolution is one half of full HD. Finally, the projection system that was used is very sensitive to proper adjustment and projection screen quality.

The answers to question 6: Did you experience any uncomfortable feelings, are summarized in Fig. 12. The plasma display exhibits the best results in this case, which is also proved by the t-test. Although the active technology using shutter glasses is likely to introduce flicker through shutting and opening the passage of light in a time multiplexed manner, no negative effect of this technique is proved by our experiment.

The last two bar plots (Fig. 13 and Fig. 14) represent the extent to which the observers reported to be disturbed by ambient light. As mentioned in Sec. 2.3., the ambient light levels were changed between 20 lux and 200 lux. Fig. 13 shows the mean opinion scores for the lower values

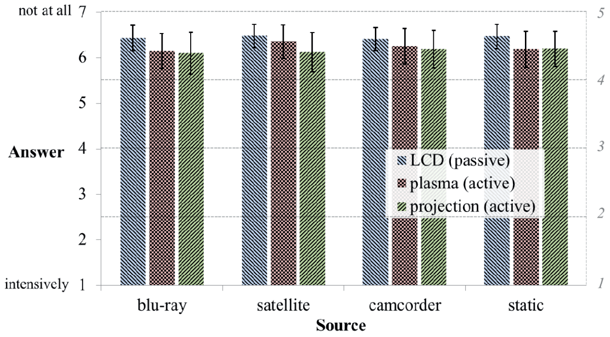



**Fig. 13.** "Did you feel disturbed by ambient light?" The mean scores and 95 % mean confidence intervals for different displays and different source content types. The results correspond to ambient illumination below 40 lux.

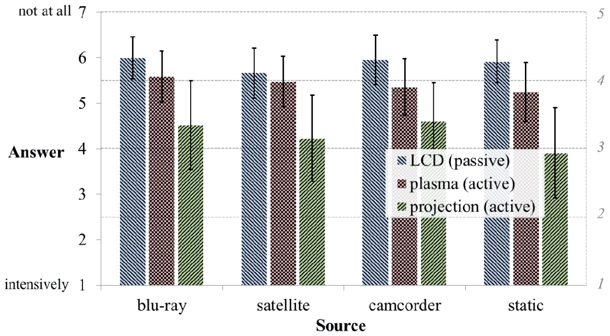



**Fig. 14.** "Did you feel disturbed by ambient light?" The mean scores and 95 % mean confidence intervals for different displays and different source content types. The results correspond to ambient illumination over 100 lux.

of ambient illumination (up to 40 lux). In this case, the users were almost never disturbed by ambient light and all the systems perform reasonably well. However, when we increase the ambient illumination level to over 100 lux, we can see rapid decrease of the ratings especially for the projection systems (quite naturally) and for the active shutter glass plasma display. The question that arises is why the passive polarized LCD 3D performs better than the active shutter glass plasma technology. Even though the active shutter glasses do not cause significant uncomfortable feelings, the higher levels of ambient light are intrusive. During the experiment, fluorescent lamps have been used as a source of ambient light. Such lamps may produce flicker at double the mains frequency (100 Hz). As this frequency is actually close to the refresh frequency of the active shutter glasses, the lower scores for plasma display at higher level of ambient illumination are likely to be caused by the perceived flicker.

## 3.2 Correlation Analysis

To allow for better understanding of the dependencies between the questions asked throughout the test, the results of correlation analysis are graphically presented in Fig. 15. The data represent the Pearson's correlation coefficient calculated over all ratings gathered from all observers.

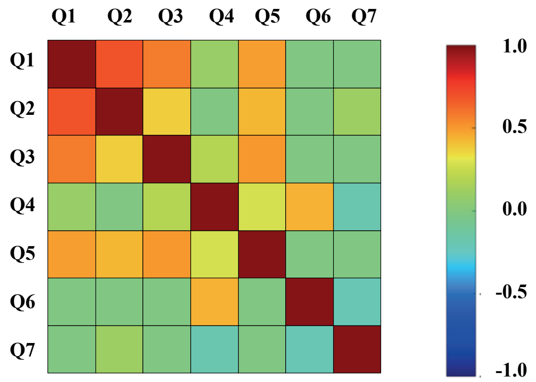



**Fig. 15.** Visualization of correlation coefficient values among all pairs of evaluated questions. The question indices 1 to 7 correspond to numbers in Tab. 1

There is a strong correlation between 0.4 and 0.7 among the first three questions. Looking back at the results reported in Figs. 7 to 9, we can recall that no significant difference was found in scores awarded to the systems, but there was an observable difference among the different contents. Consequently, the first three questions address the different properties of video sequences more than the differences among the display technologies. Also question 5 can be included in this group. In contrary, the results for questions 4, 6 and 7 (Figs. 11 - 14), exhibit very low correlation with the Pearson's correlation coefficient close to zero. These questions thus describe independent properties which we can award to the different display systems.

# 4. Conclusions

In this paper, we have analyzed several aspects of Quality of Experience reported by users viewing stereoscopic video sequences of different types displayed on equipment making use of different techniques for the separation of the left- and right-eye images.

The main conclusive remarks have been expressed in Section 3. To sum up, we have proved that the performance of all the display technologies under study is comparable in terms of the observed intensity of 3D effect, depth of the scene and user involvement in the displayed scene. However, differences can be observed for the other phenomena under study.

As the video sequences themselves have a strong impact on the different aspects of perceived QoE, the motivation of further research on the topic will be to reveal which properties of the video sequences are the strongest contributor to the perceived quality.

# Acknowledgements

This work was supported by the Czech science foundation under project number P102/10/1320. The research published in this submission was financially supported by the project CZ.1.07/2.3.00/20.0007 WICOMT of the operational program Education for competitiveness. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 230126. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation.

The authors also acknowledge all the observers who took part in the experiments and the anonymous reviewers who helped to largely improve the paper.

# References

[1] REICHL, P., FABINI, J., HAPPENHOFER, M., EGGER, C. From QoS to QoX: A charging perspective. In *Proceedings of the 18th ITC Specialist Seminar on Quality of Experience*. Karlskrona (Sweden): Blekinge Insitute of Technology, 2008, p. 35 - 44.

[2] ITU-T Recommendation P.10/G.100 (incl. Amendment 2). Vocabulary for performance and quality of service. July 2006 (2008).

[3] ETSI TR 102 643 v. 1.0.1. Human factors (HF); Quality of Experience requirements for real-time communication services. 2009.

[4] ITU-R Recommendation BT.500-11. *Methodology for the Subjective Assessment of the Quality of Television Pictures*. Geneva: ITU, 2002.

[5] ITU-T Recommendation P.910. *Subjective Video Quality Assessment Methods for Multimedia Applications*. Geneva: ITU, 2008.

[6] ITU-R Recommendation BT.1788 *Methodology for the Subjective Assessment of Video Quality in Multimedia Applications*. Geneva: ITU, 2007.

[7] BARKOWSKY, M., WANG, K., COUSSEAU, R., BRUNNSTRÖM, K., OLSSON, R., LE CALLET, P. Subjective quality assessment of error concealment strategies for 3DTV in the presence of asymmetric transmission errors. In *Proceedings of IEEE 18th International Packet Video Workshop.* Hong Kong (China), 2010, p. 193 - 200.

[8] UMAR, A. S., SWASH, R. M., SADKA, A. H. Subjective assessment of 3D videos. In *Africon 2011*, Livingstone (Zambia), 2011, p. 1 – 6.

[9] XING, L., YOU, J., EBRAHIMI, T., PERKIS, A. Estimating quality of experience on stereoscopic images. In *International Symposium on Intelligent Signal Processing and Communication Systems ISPACS 2010.* Cheng Du (China), 2010.

[10] BULAT, J., JANOWSKI, L., JUZSKA, D., SOCHA, M., GREGA, M., PAPIR, Z. Evaluation of crosstalk metrics for 3D display technologies with respect to temporal luminance analysis. In *IEEE International Symposium on Multimedia, ISM 2011.* Dana Point (CA, USA), 2011, p. 512 - 517.

[11] QUAN HUYNH-THU, BARKOWSKY, M., LE CALLET, P. The importance of visual attention in improving the 3D-TV viewing experience: Overview and new perspectives. *IEEE Transactions on Broadcasting*, 2011, vol. 57, no. 2, p. 421 - 431.

[12] QUAN HUYNH-THU, LE CALLET, P., BARKOWSKY, M. Video quality assessment: From 2D to 3D – challenges and future trends. In *Proceedings of IEEE 17th Conference on Image Processing.* Hong Kong (China), 2010, p. 4025 - 4028.

[13] ITU, Switzerland. *ITU-T Work Programme* [Online] 2012. [Cited 2012-02-14] Available at: http://www.itu.int/ITU-T/workprog/wp_search.aspx?isn_sp=545&isn_sg=549

[14] SLANINA, M., KRATOCHVIL, T., POLAK, L., RICNY, V. Analysis of temporal effects in quality assessment of high definition video. *Radioengineering*, 2012, vol. 21, no. 1.

[15] STAELENS, N., *et al*. Assessing Quality of Experience of IPTV and video on demand services in real-life environments. *IEEE Transactions on Broadcasting*, 2010, vol. 56, no. 4, p. 458 – 466.

[16] TULLIS, T., ALBERT, B. *Measuring the User Experience; Collecting, Analyzing and Presenting Usablility Metrics*. Burlington (MA, USA): Morgan Kaufmann, 2008.

[17] KLIMA, M. *et al.* DEIMOS – an open source image database. *Radioengineering*, 2011, vol. 20, no. 4, p. 1016 - 1023.

[18] KNAZEVIC, A. Overlapping confidence intervals and statistical significance. *StatNews #73*. Ithaca (NY, USA): Cornell University, 2008. [Online] Cited 2011-03-10. Available at: http://www.cscu.cornell.edu/news/statnews/stnews73.pdf.

# About Authors …

**Martin SLANINA** was born in 1982. He received the Master’s degree in 2005 and the Ph.D. degree in 2009, both in Electronics and Communication from the Brno University of Technology (BUT), Czech Republic. Currently he is an assistant professor at the Department of Radio Electronic, BUT. His research interests are in no reference video quality metric design, subjective video quality methodologies and video technology. He serves as a GOLD AG chair in the IEEE Czechoslovakia Section.

**Tomáš KRATOCHVÍL** was born in Brno, Czech Republic in 1976. He received the M.Sc. degree in 1999, Ph.D. degree in 2006 and Assoc. Prof. position in 2009, all in the Electronics and Communications program from the Brno University of Technology. He is currently an associated professor at the Department of Radio Electronics, Brno University of Technology. His research interests include digital television and audio broadcasting, its standardization and video and multimedia transmission including video image quality evaluation. He has been an IEEE member since 2001.

**Václav ŘÍČNÝ** was born in 1937. He is emeritus professor at the Faculty of Electrical Engineering and Communication, Brno Technical University. His research interest includes, in particular, video and television technology, analogue and digital signal processing and measurement.

**Libor BOLEČEK** was born in Šternberk, Czech Republic in April 1985. He graduated from the Faculty of Electrical Engineering and Communication (FEEC), Brno University of Technology (BUT), in 2010. The field of his interest is imaging processing, quality evaluation and photostereometric systems.

**Ondřej KALLER** was born in 1986. He graduated from the Brno University of Technology in 2010. His master's thesis topic was Transmission Analysis and Measurements in Digital Television DVB-S/S2. He is currently a PhD student at the same university. His fields of interest include 3D digital television systems, 3D video capturing, transmission, interpretation and evaluation.

**Ladislav POLÁK** was born in Štúrovo, Slovakia in 1984. He received the M.Sc. degree in Electronics and Communications 2009 in from the Brno University of Technology. He is currently a Ph.D. student at the Department of Radio Electronics, Brno University of Technology. His research interests are the Digital Video Broadcasting (DVB) and Digital Audio Broadcasting (DAB) systems, theory of digital television and audio, communication systems, mobile systems and multimedia transmission, including video image quality evaluation. He has been an IEEE student member since 2010.