

Influence of viewing device and soundtrack in HDTV on subjective video quality

Arne Redl, Christian Keimel and Klaus Diepold

Institute for Data Processing, Technische Universität München
Arcisstr. 21, 80333 München, Germany

ABSTRACT

Today in many homes big TV screens and hifi systems are common. But is the perception of subjective video quality under professional test conditions the same as in home use? For this two things are examined: How large is the influence of the presentation device but also the influence of the soundtrack, both in HDTV (1080p50).

Previous work has shown that a difference is noticeable, but there were no studies with consumer devices. It was also shown that there is an influence of the soundtrack, but only in SD or lower resolutions. Therefore we conducted subjective video tests: One test with different devices, a 23-inch-reference monitor, a high quality 56-inch-LCD-TV and an HD-projector, and one test in which we presented a soundtrack on a 7.1-channel hifi system in addition to the HD-projector.

The results show two things: First the test subjects had a higher quality of experience with the consumer devices than with the reference monitor, although the video quality of the reference monitor itself was rated better in an additional questionnaire and the mean opinion score (MOS). The second result is that there is no significant difference in the MOS between showing the videos on the projector with or without sound.

Keywords: Subjective video quality, presentation device, soundtrack

1. INTRODUCTION

Quality of experience is defined by the ITU with “The overall acceptability of an application or service, as perceived subjectively by the end-user.”¹ From this and subjective video quality interpreted as quality of experience it is possible that not only the codec has an impact on the perceived video quality, but also the form of presentation.

Today’s large, high resolution TVs with multi-channel hifi systems are widespread. This raises the question whether the subjective video quality under professional test conditions matches the viewing habits of the test subjects. In this work two things are examined: The influence of both the presentation device and the soundtrack on the perceived video quality.

There have already been some studies in this research area: The influence of a soundtrack has been proofed, but all studies were conducted using standard or less resolution and a stereo loudspeaker system. There has not been any verification that the same applies to high-definition television (HDTV), here 1920×1080 . Joly et al. proved that a good audio quality has an influence on video quality in standard definition television (SDTV).² Beerends and Caluwe confirmed this in previous work also for SDTV.³ Winkler and Faller have examined the audiovisual quality for low bit rate videos.⁴ Also there were studies about the influence of the viewing device: For example Keimel and Diepold showed that the subjective video quality diverges with different monitor types of the same size⁵ or Ardito et al. showed the influence of display parameters,⁶ both in HDTV. But there were no studies using consumer device in different sizes.

Therefore we wanted to examine these two possible impacts on the subjective video quality in HDTV. First we checked the influence of the presentation device on the mean opinion score. For this, we presented the

Further author information: (Send correspondence to Arne Redl)

Arne Redl	E-mail: redl@tum.de	Telephone: +48 89 289 23601
Christian Keimel	E-mail: christian.keimel@tum.de	Telephone: +49 89 289 23629
Klaus Diepold:	E-mail: kldi@tum.de	Telephone: +49 89 289 23602

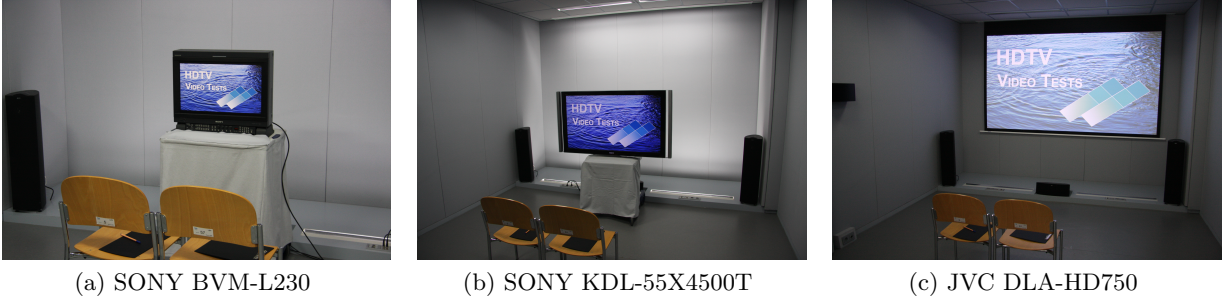


Figure 1. Presentation devices in the laboratory environment.

videos on three different devices: A 23-inch studio reference monitor, a big 56-inch-LCD-TV with background LED illumination and an HD-projector with a width of the projection screen of 2.60 m. The reference display represents the standard testing conditions. The other two consumer devices represent ambitious, but common home viewing conditions. All devices support the used video format of 1080p50. Additionally, we checked if a soundtrack, played simultaneously on a 7.1 channel hifi system, would have an influence on the mean opinion score. To present a wide range of video qualities we encoded the well-known SVT test set with an H.264/AVC encoder using four different bit rates, representing a quality range from bad to very good. We tested the possible influence in two ways: On one hand we conducted four traditional single stimulus tests with a discrete eleven-point rating scale. On the other hand we designed a questionnaire to determine the quality of experience. This contribution is organized as follows: First we introduce the test environment and all used devices, then we present and discuss all results and finally we conclude with a short summary.

2. TEST SEUTUP

2.1 Environment

All tests were conducted in the video laboratory of the Institute for Data Processing at the Technische Universität München. This laboratory allows testing under all specifications given by the ITU Recommendation BT.500-12.⁷ All components support the selected video format of 1080p50 and can provide both professional and consumer display devices.

2.2 Testing devices

To cover a wide range of display types we selected three different presentation devices, which all provide the used video format of 1080p50:

- A Class A LCD video monitor with 23" (SONY BVM-L230):
This display represents a standard test environment, because this class of devices is an often used device for subjective video tests. Surely this device performs best from a professional point of view.
- A LCD Digital Color TV with 56" (SONY KDL-55X4500):
Over the last years more and more high definition LCD-TVs have become popular. In Germany the number of households with a flat screen rose from 5% in 2006 to 26% in 2009. So this LCD-TV with a LED-background illumination should be a representative of an ambitious, but common home-environment. To provide this device with the video signal we had to convert the signal from DualLink-HD-SDI to HDMI.
- A D-ILA Projector with a width of the projection screen of 2.60 m (JVC DLA-HD750):
Projectors become more popular too, especially in home-cinemas. So we decided to have one representative of this class of devices. This high level consumer device also needs a HDMI signal.

All devices as used in this test are shown in Fig.1 and summarized in Tab.1. To achieve a nearly equal presentation of videos all devices are calibrated by the X-Rite's EyeOne Pro. To present the soundtrack we used

SONY BVM-L230	LCD Video Monitor (Class A)	23"	Reference monitor
SONY KDL-55X4500	LCD Digital Colour TV	56"	large LCD-TV
JVC DLA-HD750	D-ILA Projector	2,60 m (width of the projection screen)	Projection

Table 1. Used presentation devices

No.	Sequence	Loudness	
		Maximum L_{max} [dBA]	Average L_{eq} [dBA]
1	CrowdRun	74,4	70,4
2	TreeTilt	71,9	66,4
3	PrincessRun	71,9	68,6
4	DanceKiss	74,1	72,2
5	FlagShoot	79,6	69,0

Table 2. Maximum and average loudness.

a permanently installed 7.1-hifi-system, consisting of an AV-Receiver, two front-speakers, one center-speaker, four dipole loudspeakers and one subwoofer. The sound-system is a high-level consumer system. We decided not to use professional sound equipment, because a consumer device would represent a realistic home environment and the feeling of a movie-soundtrack much better. To obtain a good surround feeling the seats were positioned with line the two side speakers according to the Dolby Home Theater recommendation.⁸ So the installed projection screen and the installed hifi-system result in a viewing distance of two heights of the screen. To avoid an influence of the viewing distance on the results for all test cases the same distance was chosen.

In previous tests by Ardito et. al⁶ and Zonja et al.⁹ a loudness of about 80 dBA is used. So we adjusted the sound system that a maximum of 80 dBA was reached. The average and the maximum loudness of every scene are shown in Tab. 2. As silence noise level, meaning the noise level with all devices running but without playing a sound, we measured a level below 30 dBA.

2.3 Stimuli and Codec

In order to get stimuli with different visual quality we encoded five scenes of 10seconds from the well-known SVT multi format test set with the H.264/AVC codec. We applied this codec because it is the most common codec in HDTV as it is used in BluRay or DVB-S2. The scenes were selected to contain a wide range of coding difficulties and most of the criteria proposed by the Video Quality Expert Group.¹⁰

We selected the following five scenes from the test set (Fig.2), thereof four of the proposed cut offs: CrowdRun, TreeTilt, PrincessRun, DanceKiss and FlagShoot. The scene FlagShoot is not one of the proposed clips but is still a scene with an interesting soundtrack. To examine the influence of the audio and the presentation device a wide range of quality from poor to very good quality close to the original of each scene should be presented. So we produced four quality grades of each scene. In previous works¹¹ for the SVT test set the same bit rates were chosen for all classes of coding complexity. Unfortunately, scenes with an easy coding complexity got almost the same rating for all bit rates. An analysis of the PSNR (Fig. 6) and MSE values confirm this result.

Therefore the lower and the higher bit rate for each scene were examined step by step with the help of a person which is not familiar with video coding quality. We used a wide range of different bit rates (Tab.3) for the four different quality levels to get visible differences in each scene.

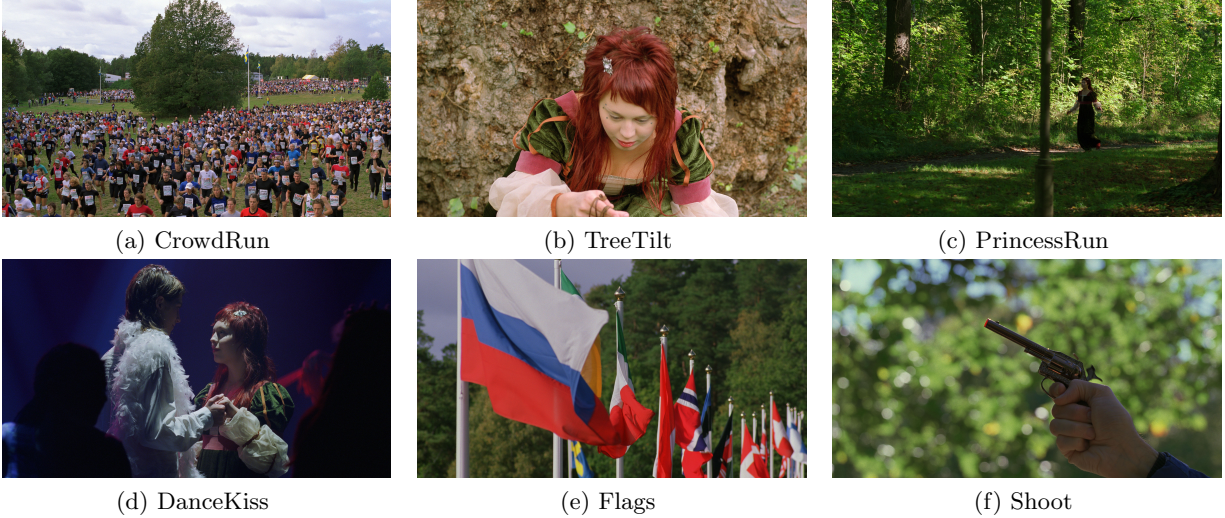


Figure 2. Test sequences from the SVT test set.

		1	2	3	4
1	CrowdRun	8 Mbit/s	20 Mbit/s	30 Mbit/s	40 Mbit/s
2	TreeTilt	2 Mbit/s	3 Mbit/s	6 Mbit/s	10 Mbit/s
3	PrincessRun	8 Mbit/s	20 Mbit/s	30 Mbit/s	40 Mbit/s
4	DanceKiss	1,5 Mbit/s	2 Mbit/s	3 Mbit/s	4 Mbit/s
5	FlagsShoot	2 Mbit/s	3 Mbit/s	6 Mbit/s	10 Mbit/s

Table 3. Selected bit rates ordered by scene and quality level (1 –4).

2.4 Testing methodology

Primarily two methods for evaluating subjective video quality are common: The Single Stimulus and the Double Stimulus method. For our subject we decided to use the Single Stimulus test method with a discrete eleven-point scale where each encoded 10 sec sequence is shown without reference, mainly for one reason: Under the assumption that an influence of the presentation device or the soundtrack would influence both the encoded video and the reference in the same way, this influence would be canceled out.

Only for the scenes CrowdRun and PrincesRun the references were additionally shown as “hidden references” to verify that the upper bound of the bit rate had not been chosen too low. The Video Quality Expert Group has proposed this method in its test plan for HDTV.¹⁰

Additionally a questionnaire with questions regarding the quality of experience of each whole test scenario was developed. Of course the most of the participants had no idea of the correct definition of quality of experience. Thus the questions in Tab. 4 were asked to get an assessment of the perceived overall video quality, the quality of experience and the overall impression.

Question	Scale
How do you feel about the overall impression?	Eleven-point scale
How do you evaluate the overall impression regarding to a home cinema?	Eleven-point scale
How do you evaluate the overall video quality?	Eleven-point scale

Table 4. Question of the additional questionnaire.

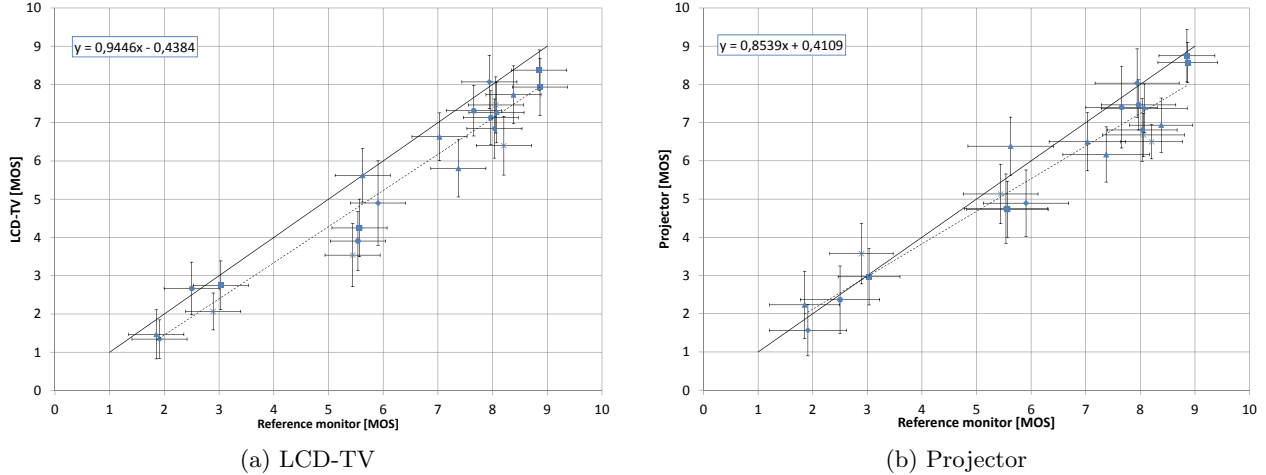


Figure 3. Comparison between reference monitor and LCD-TV respectively projector including 95 % confidence intervals and linear regression line (♦ CrowdRun; ■ TreeTilt; ▲ PrincessRun; × DanceKiss; ● FlagShoot).

2.5 Implementation of the test

Altogether five scenes with four quality grades and in addition the hidden references for two scenes were shown. The single stimulus test with two presentations of each sequence and a stabilization phase contained 48 sequences or 12 min. for each test setup. The order for each test setup – except for the stabilization phase – was generated by a random generator but without repeating two equal scenes. Thus the presentation of two setups could be conducted to one session with less than 30 min of testing time according to the ITU Recommendation BT.500-12.⁷ Between two sessions there was a break of at least 30 min.

For each test session there should be a minimum of 15 valid votes.¹² To get these minimum requirement 21 subjects took part with an age from 16 to 27, two of them female. All participants were tested for normal visual acuity and normal color vision with Snellen- and Ishihara tests.¹⁰ To introduce every participant in the methods of subjective video tests an introduction with a short training was performed.

For the decision weather single votes or even all votes of one participant are valid the vote of the repetition may deviate from the first vote by a maximum of three scale units. Furthermore the average of both votes may deviate by a maximum of three scale units from the average of all other participants. All votes of a participant were removed, if he had more than 15 % of invalid votes, in this case a minimum of three invalid votes.

The mean opinion score (MOS) is built by the average of all valid votes. For a better statistical significance we give the 95 % confidential interval for each MOS based on the Student's t-distribution.

3. RESULTS

In Fig. 7 and Tab. 5 the results for all four test scenarios are shown ordered by the scenes. Basically we have a increasing tendency of the mean opinion score for increasing bit rates. The values for the different test setups are close together, mostly the confidence intervals are overlapping. Hereafter we compare the test setups with each other.

Fig. 3 compare the results of the presentation on the projector and the LCD-TV, both without soundtrack, with the results of the reference monitor. However the video quality is, against the assumption, perceived rather worse, especially on the LCD-TV, than on the reference monitor. Hence there is an offset of 0.8 of MOS at the LCD-TV. By the way this confirms the evaluation of the questionnaire (Fig. 5), where the participants quoted the quality of the reference monitor as the best, even though they perceived the quality of experience and the overall impression better on the consumer devices. This indicated that the video quality is not the only factor for the quality of experience.

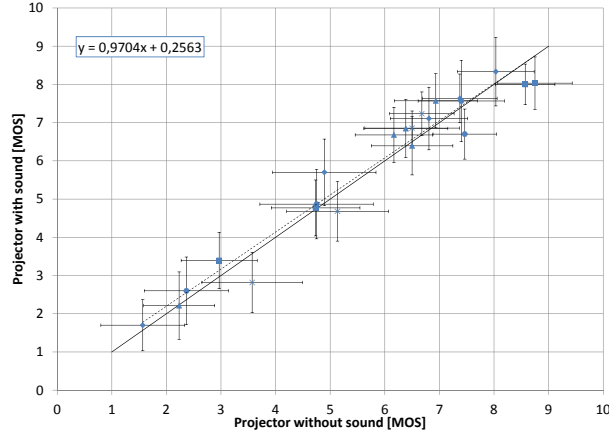


Figure 4. Comparison between the projector with and without sound including 95% confidence intervals and linear regression line (♦ CrowdRun; ■ TreeTilt; ▲ PrincessRun; × DanceKiss; • FlagShoot).

Sequence	Reference monitor	LCD-TV	Projector with sound	Projector without sound
CrowdRun (8 Mbit/s)	1,912	1,344	1,567	1,700
CrowdRun (20 Mbit/s)	5,906	4,900	4,893	5,700
CrowdRun (30 Mbit/s)	8,033	6,846	6,808	7,107
CrowdRun (40 Mbit/s)	8,071	7,267	7,375	7,633
CrowdRun (Reference)	7,941	8,067	8,033	8,333
TreeTilt (2 Mbit/s)	3,031	2,750	2,967	3,393
TreeTilt (3 Mbit/s)	5,567	4,250	4,733	4,769
TreeTilt (6 Mbit/s)	8,867	7,933	8,571	8,000
TreeTilt (10 Mbit/s)	8,853	8,375	8,750	8,033
PrincessRun (8 Mbit/s)	1,853	1,469	2,233	2,214
PrincessRun (20 Mbit/s)	5,625	5,625	6,385	6,846
PrincessRun (30 Mbit/s)	7,375	5,808	6,167	6,679
PrincessRun (40 Mbit/s)	7,029	6,633	6,500	6,393
PrincessRun (Reference)	8,382	7,733	6,933	7,571
DanceKiss (1,5 Mbit/s)	2,893	2,067	3,571	2,818
DanceKiss (2 Mbit/s)	5,441	3,538	5,133	4,679
DanceKiss (3 Mbit/s)	8,206	6,400	6,500	6,857
DanceKiss (4 Mbit/s)	8,059	7,464	6,679	7,233
FlagsShoot (2 Mbit/s)	2,500	2,667	2,367	2,600
FlagsShoot (3 Mbit/s)	5,538	3,906	4,750	4,867
FlagsShoot (6 Mbit/s)	7,967	7,133	7,464	6,700
FlagsShoot (10 Mbit/s)	7,656	7,313	7,400	7,567

Table 5. Summary of the results ordered by sequence and test case.

Furthermore the participants could detect differences between the quality levels depending on the presentation device. For example in CrowdRun there are still rising mean opinion scores for the upper bit rates for the consumer device in contrast to the reference monitor.

Fig.4 contrasts the presentation on the projector without sound with the presentation with sound. This figure shows that most confidence intervals are clipping the centerline. Also the regression line almost lies on the centerline. Only in the questionnaire the video quality and the overall impression is rated a little bit, but not significant, better for the presentation with soundtrack. Thus the soundtrack has in HDTV in our test nearly no influence on the perceived video quality.

Altogether this is confirmed by the high correlation coefficients of the mean opinion scores (Tab.6). Thus

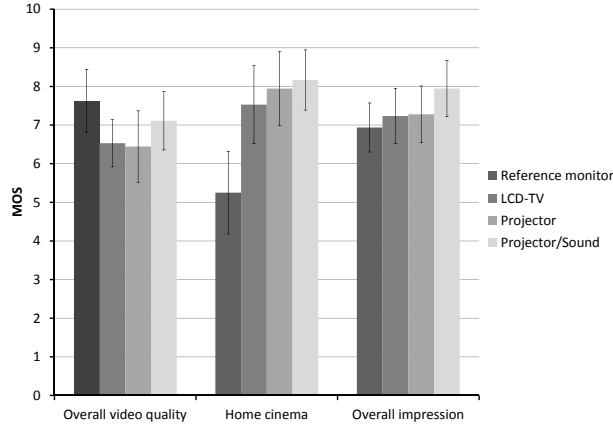


Figure 5. Evaluation of the questionnaire.

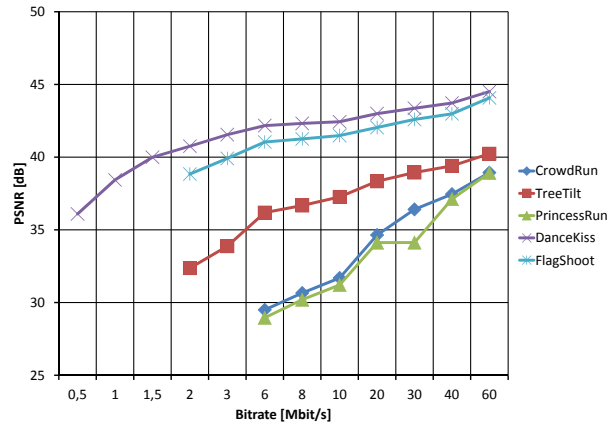


Figure 6. PSNR values for a wide range of bit rates for each scene.

the various forms of presentation have almost no influence on mean opinion score although the participants have rated the overall video quality better for the consumer devices.

4. CONCLUSIONS

The results show two things: First- and this is not surprising - the subjects prefer the audiovisual test setup for a home environment, closely followed by the setup with the projector without sound and the big TV-screen. Secondly, more surprising, for each quality grade, the video quality is perceived better on the reference display than on all other. This is the result of both the mean opinion score of the single stimulus test and the

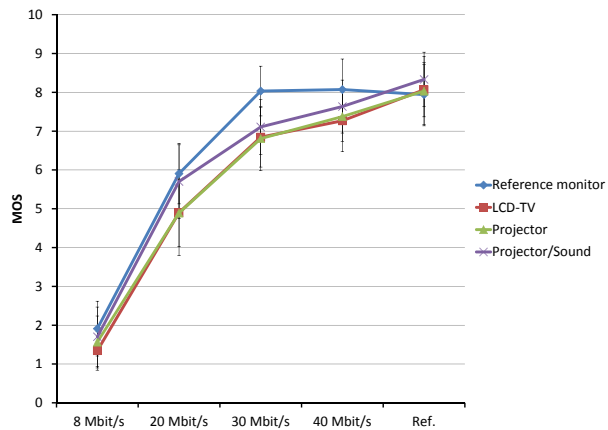
	Reference monitor	LCD-TV	Projector without sound	Projector with sound
Referenz-Monitor	1,000	0,968	0,962	0,969
Sony Bravia		1,000	0,968	0,980
JVC-Beamer			1,000	0,975
AV-Test				1,000

Table 6. Correlation coefficients for the different test cases compared each other.

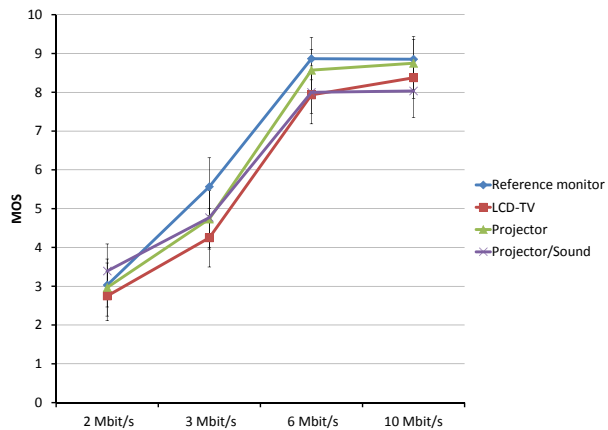
questionnaire, in which the participants were to rate the overall video quality. The biggest quality difference showed between the reference monitor and the LCD-TV, although it was a constant offset in all quality grades. It is also interesting that presentation on the beamer was rated equally regardless of whether sound was played or not. To summarize we can confirm that despite an offset, all forms of presentation are highly correlated. Thus common professional test setups have still a high degree of significance for today's home viewing.

REFERENCES

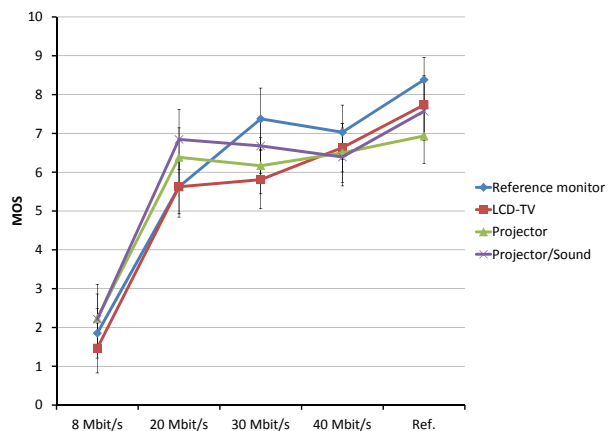
- [1] Focus Group on IPTV Standardization, "Definition of Quality of Experience," (2007).
- [2] Joly, A., Montard, N., and Buttin, M., "Audio-visual quality and interactions between television audio and video," *Signal Processing and its Applications, Sixth International, Symposium on. 2001* **2001**, 438–441 (2001).
- [3] Beerends, J. G. and Caluwe, F. E. d., "The Influence of Video Quality on Perceived Audio Quality and Vice Versa," *J. Audio Eng. Soc* **47**(5), 355–362 (1999).
- [4] Winkler, S. and Faller, C., eds., [*Audiovisual quality evaluation of low-bitrate video: Proc. SPIE Human Vision and Electronic Imaging*], **5666** (2005/January 16-20).
- [5] Keimel, C. and Diepold, K., "On the use of reference monitors in subjective testing for HDTV," *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on* **2010**, 35–40 (2010).
- [6] Ardito, M., Gunetti, M., and Visca, M., "Influence of display parameters on perceived HDTV quality," *IEEE Transactions on Consumer Electronics* **42**(1), 145–155 (1996).
- [7] ITU-R (Radiocommunication Sector of ITU), "Recommendation ITU-R BT.500-12 (09/2009): Methodology for the subjective assessment of the quality of television pictures," (2009).
- [8] Dolby Europe Licensing Limited, "Home Theater Speaker Guide 7.1 Channel," (26.01.2010).
- [9] Zonja, S., Livun, N., and Jambrosic, K., "Audio - Visual Interaction: Multimedia Applications," *Multimedia Signal Processing and Communications, 48th International Symposium ELMAR-2006 focused on* **2**, 143–146 (2006).
- [10] VQEG HDTV Group, "Test Plan for Evaluation of Video Quality: Models for Use with High Definition TV Content," (2009).
- [11] Keimel, C., Habigt, J., Habigt, T., Rothbacher, M., and Diepold, K., "Visual quality of current coding technologies at high definition IPTV bitrates," *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on* **2010**, 390–393 (2010).
- [12] Winkler, S., "On the properties of subjective ratings in video quality experiments," *Quality of Multimedia Experience, 2009. QoMEX 2009. International Workshop on* **2009**, 139–144 (2009).



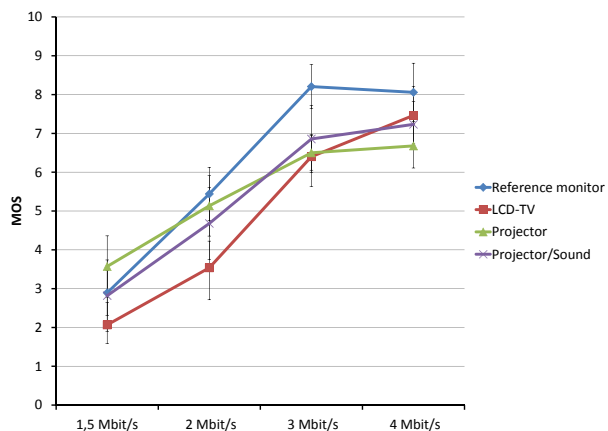
(a) CrowdRun



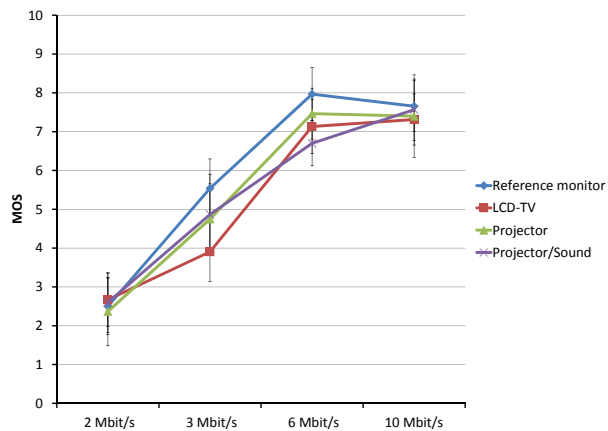
(b) TreeTilt



(c) PrincessRun



(d) DanceKiss



(e) FlagShoot

Figure 7. Mean opinion score ordered by scene including 95 % confidence intervals.