Comparison of HDTV formats in a consumer environment

Christian Keimel, Arne Redl and Klaus Diepold

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

ABSTRACT

High definition television (HDTV) has become quite common in many homes. Still, there are two different formats used currently in commercial broadcasting: one interlaced format, 1080i50/60, and one progressive format, 720p50/60.There have already been quite a few contributions comparing the visual quality of these formats subjectively under common standard conditions. These conditions, however, dont necessarily represent the viewing conditions in the real-life consumer environment. In this contribution we therefore decided to do a comparison under conditions more representative of the consumer environment with respect to display and viewing conditions. Furthermore, we decided to select not specially prepared test sequences, but real-life content and coding conditions. As we were not interested in the influence of the transmission errors, we captured the sequences directly in the play-out centre of a cable network provider in both 1080i50 and 720p50. Also we captured for comparison the same content in digital PAL-SDTV. We conducted extensive subjective tests with overall 25 test subjects and a modified SSIS method. The results show that both HDTV formats outperform SDTV significantly. Although 720p50 is perceived to have a better quality than 1080i50, this difference is not significant in a statistical sense. This supports the validity of previous contributions results, gained in standard conditions, also for the real-life consumer environment.

Keywords: HDTV formats, 720p, 1080i, subjective testing, visual quality, HDTV format comparison

1. INTRODUCTION

High definition television (HDTV) is no longer something new, but has already become a fixture in many homes. Yet, there are two competing formats used in commercial broadcasting: one interlaced format with a resolution of 1920 × 1080 pixels at 50/60 fields per second (1080i50/60) and one progressive format with 1280 × 720 pixels at 50/60 frames per second (720p50/60).In order to evaluate the advantages and disadvantages of each format, subjective tests according to standards e.g. ITU-R BT.500[1] are conducted in many contributions. These standards, however, dont necessarily represent the viewing conditions in a real-life consumer environment.

We therefore decided to do a comparison in a more realistic consumer environment. Instead of a calibrated reference monitor, we selected a middle class 40 inch consumer LCD display. We did neither color calibrate it, nor did we reduce its brightness and the background illumination to the recommended levels, as this is usually not done in the consumer environment. Secondly, we chose not the usual viewing distance of 3H for HDTV, but rather a fixed distance of 3m, representing the usual distance for standard definition television (SDTV) in many households.

Also we did not chose specially prepared test sequences, but real-life content and coding conditions. An opportunity arose as a sporting event was broadcast simultaneously in different formats. Hence the same content was available in different formats. As we were not interested in the influence of the transmission errors, we captured the sequences directly in the play-out centre of a cable network provider in both 1080i50 and 720p50 coded with H.264/AVC. Note, that while there may be different coding parameter for both formats, these are inherent in common broadcasting conditions. Also we captured for comparison the same content in digital PAL-SDTV coded with MPEG-2. In order to compare our results to a worst case scenario, we captured additionally

Further author information: (Send correspondence to Christian Keimel)

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

Image Quality and System Performance VIII, edited by Susan P. Farnand, Frans Gaykema, Proc. of SPIE-IS&T Electronic Imaging,
SPIE Vol. 7867, 786716 · © 2011 SPIE-IS&T · CCC code: 0277-786X/11/$18 · doi: 10.1117/12.876781

the same content at an analogue PAL-SDTV consumer cable network connection point with a significant amount of noise. Because we were interested in the differences of the formats and not the signal processing capabilities of the used display, we decided to convert all formats for presentation on the display into 1080p50. This pre-processing was done using simple software tools. We selected bob de-interlacing and bicubic interpolation for de-interlacing and up-scaling, respectively. While these methods may not be the most sophisticated techniques, they provide a lower bound of the signal processing to be expected in consumer devices.

This contribution is organized as follows: we review related work before introducing the test sequences, the test setup and the subjective testing. Then we present the results of our subjective test, before concluding with a short summary.

2. RELATED WORK

Contributions so far conducted subjective tests in standard environments that not necessarily correspond to the the real-life consumer environment. In these tests, the progressive resolution of 1920×1080 at 50/60 frames per second (1080p50/60) usually shows the best quality and the progressive 720p formats are found to be superior to the interlaced 1080i formats.

Hoffmann et.al.[2–4] compared different HDTV format using the Triple Stimulus Continuous Evaluation Scale (TSCES) method, enabling the test subject not only to asses the video sequence under test, but also provides a low and high quality anchor to the test subjects. In another contribution, Hoffmann et.al.[5] use a more traditional setup with the Double Stimulus Impairment Scale (DSIS) method according to ITU-R BT.500.[1] In these contributions the results show that 720p50 is judged to have a better perceived visual quality by the test subjects.

Meenowa et.al.[6] and Loncaric et.al.[7] showed that for the same bitrates, 720p50 delivers a better perceived quality compared to 1080i50. In a comparison of different display types, SVT[8] showed that 720p is preferred to 1080i.

3. TEST SEQUENCES

Our aim in this contribution is to compare common HDTV formats in a real-life consumer environment. Therefore we decided to use test sequences captured from actual HDTV and SDTV broadcasts, representing typical content that was processed in the transmission chain from broadcasting stations to the play-out center of a (cable) network provider.

One problem with this approach is, that usually the same content is not broadcast simultaneously in the different HDTV formats. As in most cases the same content is only shown on one channel and each channel only provides one specific HDTV format, it is difficult to capture the same content for test proposes in different formats. During a large athletics tournament, however, many events were simulcast in 720p and 1080i, but also in the PAL-SDTV format 576i by different broadcasting stations in Germany.

We selected in total seven video sequences, providing similar, but still different content. An overview of the different sequences is given in Table 1 [*].For the first four sequences *Hurdle1-Hurdle4*, the same content was available in both HDTV formats and SDTV; for the last three sequences *Hammer1-Hammer3* the same content was available for both HDTV formats, but a different content had to be used for SDTV, as this content was not simulcast in HDTV and SDTV. We took, however, great care, that even though the content was not identical, it was still similar to the HDTV content. As our focus in this contribution is on the comparison of different HDTV formats and SDTV results are just provided for reference, this limitation is, in our opinion, acceptable in the context of this contribution.

Both the H.264/AVC encoded HDTV and the MPEG-2 encoded (digital) SDTV sequences were extracted from the captured transport stream and decoded. The transport streams were captured in the play-out centre of a cable network provider. Assuming the content delivery from the broadcasters to the cable network's play out centre is error free, the (digital) sequences used in this test did not contain distortion introduced by transmission

[*]Due to copyright issues we are unfortunately not able to provide screen shots of the video sequences.

Table 1: Video sequences

Video sequence	Description	Format 576i50 analogue	576i50 digital	720p50	1080i50
Hurdles1	hurdle race start camera pan & zoom slow movement	X	X	X	X
Hurdles2	hurdle race; camera pan slow movement detailed background	X	X	X	X
Hurdles3	hurdle race camera pan slow movement	X	X	X	X
Hurdles4	hurdle race finish line camera pan slow movement	X	X	X	X
Hammer1 HDTV	hammer throwing flying hammer & impact fast movement & zoom			X	X
Hammer1 SDTV	(same as HDTV)	X	X		
Hammer2 HDTV	hammer throwing launch fast movement			X	X
Hammer2 SDTV	(same as HDTV)	X	X		
Hammer3 HDTV	hammer throwing thrower approaches circle slow movement & zoom			X	X
Hammer3 SDTV	(same as HDTV)	X	X		

errors. Clearly, the coding structure of all three digital signals is different. Nevertheless, this represents a realistic application scenario, where different formats will be encoded with different parameters or, as in the case of SDTV, even different encoding technologies.

Additionally, we captured and digitalized the SDTV content at an analogue consumer cable network connection point with significant noise. Obviously, the visual quality of these sequences will have a rather bad visual quality. Hence the analogue SDTV version should be viewed as an extremely low quality anchor. Still, we decided to include this format in our subjective tests in order to get a measure of the overall improvement of digital content delivery in comparison to a existing worst case in the consumer environment.

4. TEST SETUP

The test setup we used reflects a typical consumer environment with regard to display device, viewing distance and background illumination.

Instead of a reference display as recommended in standards for subjective video quality testing e.g. ITU-R BT.500,[1] we used a middle class consumer so-called *Full HD* LCD display with a native resolution of 1920×1080 and a screen diagonal of 40 inch, as well as a cold cathode fluorescent lamp (CCFL) backlight (Samsung LE40B650). We chose a LCD display instead of a reference class CRT as required by most standards, as most displays sold nowadays are LCD displays and are thus more representative of the consumer environment. We have already shown in previous contributions,[9] that reference displays may not always be needed in subjective testing. Secondly, we chose a display size and quality representative of neither low-cost nor high-end market segments, but of the sensible market segment in the middle with a good value for money. Furthermore, the display was not calibrated and the factory default settings were used.

As we were interested mainly in the difference between the formats, we applied a preprocessing and provide the display only with a 1080p50 input signal. The reasoning is, that in consumer electronics quite often vendor dependent algorithms are used to convert the input signals into the display's native resolution. In order to avoid this unknown preprocessing, that could possibly be different for each input format presented in this test, we decided to use only one input signal. Hence the internal, unknown signal processing is applied in exactly the same way to all formats under test. We used the common open source software tools *VirtualDub* and *AviSynth* for de-interlacing (576i, 1080i) and/or scaling (576i, 720p). For de-interlacing we used bob de-interlacing and for scaling bicubic interpolation. Although display-integrated algorithms may very well be more sophisticated, the chosen methods still represent effective solutions for de-interlacing and scaling. Our preprocessing can therefore be considered to be a lower bound for current state-of-the-art signal conversion. An overview of the preprocessing chain is given in Fig. 1.

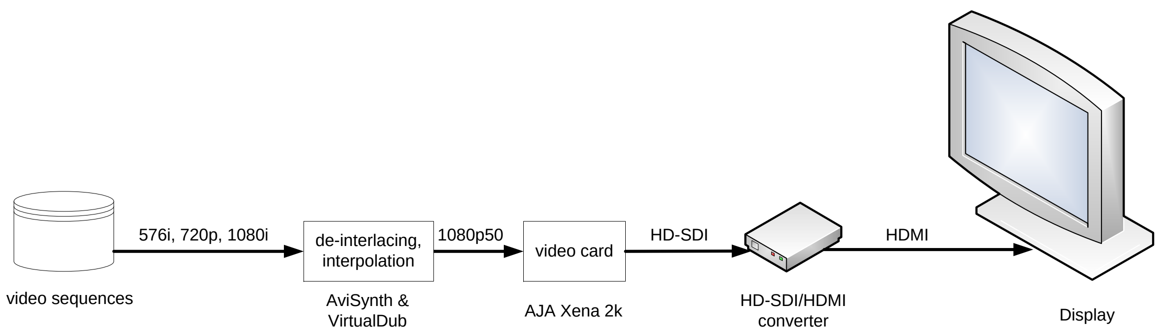

Figure 1: Preprocessing of the different formats

In addition to using a consumer display, we increased the viewing distance to one more representative for the consumer environment. Relevant standards e.g. ITU-R BT.500[1] usually suggest a viewing distance of three times the display height (3H) for HDTV. For the display used in this test, this would result in a recommend viewing distance of roughly $1.5\,m$. A survey by the BBC,[10] however, has shown that the median absolute


Figure 2: Test room

distance for wide screen displays is $2.7\,m$ or, in our case, closer to five times the display height (5H). Although HDTV was not considered explicitly in this survey, it seems likely that even if consumers upgrade from SDTV to HDTV displays, they would not necessarily rearrange their viewing conditions and therefore significantly reduce the viewing distance to 3H. Thus we decided to use a viewing distance of $3\,m$ in our test setup. Also we increased the background illumination to levels more common in the consumer environment compared to those recommended in e.g. ITU-R BT.500.[1] The test took place in the video quality evaluation laboratory of the Institute for Data Processing at the Technische Universität München. The setup can be seen in Fig. 2.

5. SUBJECTIVE TESTING

Obviously, we do not have a uncoded reference available in our tests, as we did not encode video sequences ourselves, but rather captured the sequences from real-life broadcasts. Therefore we selected a single stimulus method for the subjective testing. We used a variation of the well known Single Stimulus Impairment Scale (SSIS)[1] method: the *Single Stimulus MultiMedia* (SSMM) method. Instead of an impairment scale as in SSIS, SSMM uses a scale that directly evaluates the quality perceived by the test subject. This method has been used extensively in the MPEG standardization of H.264/AVC[11] and its extension SVC.[12] The structure of a SSMM basic test cell is shown in Fig. 3.

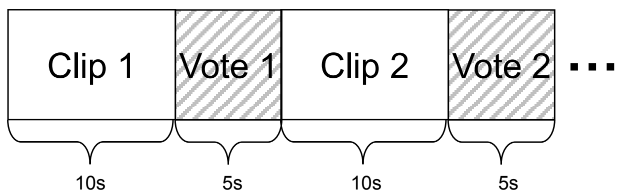

Figure 3: SSMM basic test cell

To allow the test subjects to differentiate between relatively small quality differences, a discrete voting scale from very bad to very good with eleven grades ranging from 0 to 10 was used. Before the test itself, a short training was conducted with three sequences of different content to the test, but similar visual quality, resulting in a training session of ten sequences. During this training, the test subjects had the opportunity to ask questions regarding the testing procedure. The test subjects were also instructed to ignore possible on screen caption e.g. *Digital* or *HD*. We also included a stabilization phase of six sequences representing the whole quality range at the beginning of the test, so that the test subjects are able to get an impression of the quality range that will be encountered during the test. The results of this stabilization phase are discarded during the processing of the votes. One problem in single stimulus methods can be the influence of the context in which the test sequences are shown: the visual quality of the previous sequence influences the perceived quality of the current test case. In order to avoid context effects, every test case was therefore shown twice in a different context. Both votes of each test case are then averaged during processing of the result.

Processing of outlier votes was done according to.[1] During the processing of the votes, votes were rejected if the difference between the two votes for one test case was larger than three scale units. Always both votes were removed. If more than 15% of all votes of a subject had to be removed, the subject and its votes were removed completely, as this strongly indicates that the test subject is not able to reproduce its quality estimation. We then determined the mean opinion score (MOS) by averaging all valid votes for each test case.

6. RESULTS

The test subjects were mostly students between 20 and 30, with no or very little experience in video coding. All test subjects were screened for visual acuity and color blindness. In total 25 test subjects participated in the subjective tests. The votes of three subjects were removed completely as they were unable to reproduce their results in more than 15% of all test cases. For the remaining 22 subjects single outliers were removed. Overall less than 3% of all votes were removed as outliers.

The results of the subjective tests are shown in Fig. 4 and Table 2. We can see clearly that –as expected– both HDTV formats achieve a consistently higher visual quality than SDTV for all video sequences: both in the sequences with identical content, *Hurdle1-4*, but also for the sequences with only similar content *Hammer1-Hammer3*. Also not surprisingly the noisy analogue SDTV signal performs worse than all digital signals, except for *Hammer3*. This outlier can be explained by exceptionally strong blocking in the digital SDTV version.

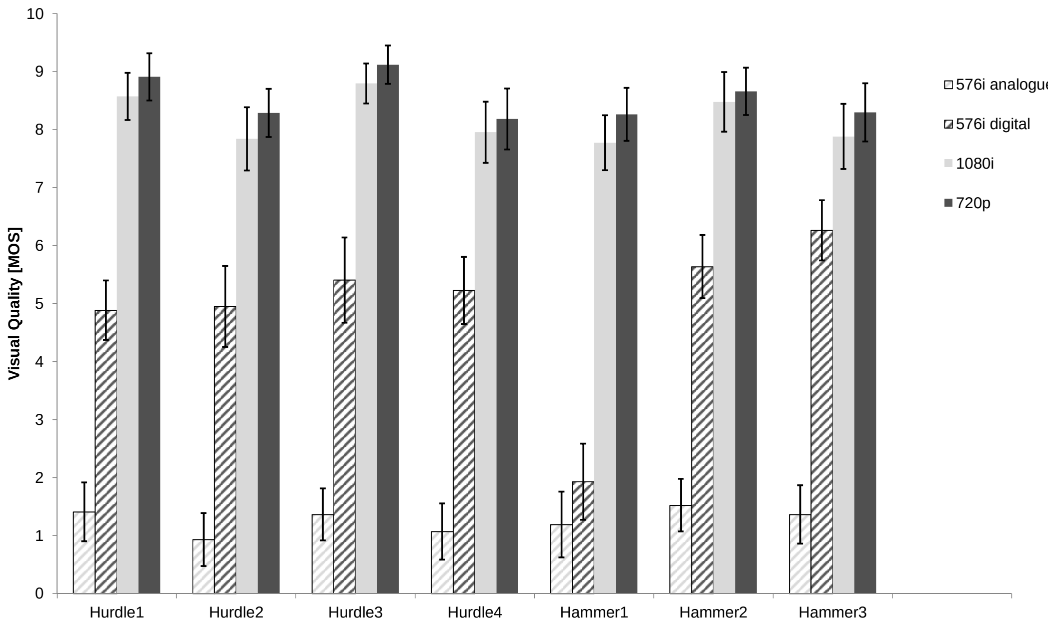

Figure 4: Visual quality and the corresponding 95% confidence intervals

In the comparison between both HDTV formats, 720p outperforms 1080i for all sequences. This is consistent with the result of previous contributions in Section 2. It should be noted, however, that this result is not necessarily statistically significant, as demonstrated by the overlapping confidence intervals in Fig. 4 .

7. CONCLUSION

We have conducted extensive subjective tests in a real-life consumer environment by using video sequences captured from actual broadcasts and displayed them on a consumer grade display in a realistic viewing environment with respect to viewing distance and background illumination.

Our results confirm previous results on the preference of 720p over 1080i by viewers. Moreover, we show that this is also valid in a real-life consumer environment with vastly different conditions to the more theoretical standard environments considered so far.

REFERENCES

[1] "ITU-R BT.500 methodology for the subjective assessment of the quality for television pictures," (June 2002).

[2] Hoffmann, H., Itagaki, T., and Wood, D., "Quest for finding the right HD format : A new psychophysical method for subjective HDTV assessment," *SMPTE Motion Imaging Journal* **17**(3), 30–39 (2008).

[3] Hoffmann, H., Itagaki, T., Wood, D., Hinz, T., and Wiegand, T., "A novel method for subjective picture quality assessment and further studies of HDTV formats," *Broadcasting, IEEE Transactions on* **54**(1), 1 –13 (2008).

[4] Hoffmann, H., "EBU format comparisons at IBC 2006," *EBU TECHNICAL REVIEW* **308** (Oct. 2008).

[5] Hoffmann, H., Itagaki, T., Wood, D., and Bock, A., "Studies on the bit rate requirements for a HDTV format with 1920 x 1080 pixel resolution, progressive scanning at 50 hz frame rate targeting large flat panel displays," *Broadcasting, IEEE Transactions on* **52**(4), 420 –434 (2006).

[6] Meenowa, J., Hands, D. S., Young, R., and Bayart, D., "Subjective assessment of HDTV content: comparison of quality across HDTV formats," *Human Vision and Electronic Imaging XV* **7527**(1), 75270I, SPIE (2010).

[7] Loncaric, M., Tralic, D., Brzica, M., Vukovic, J., Lovrinic, J., Dumic, E., and Grgic, S., "Testing picture quality in HDTV systems," in [*ELMAR, 2008. 50th International Symposium*], **1**, 5 –8 (2008).

[8] SVT, "Overall-quality assessment when targeting Wide-XGA flat panel displays," tech. rep., SVT Corporate Technology (2002).

[9] Keimel, C. and Diepold, K., "On the use of reference monitors in subjective testing for HDTV," in [*Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*], 35 –40 (2010).

[10] Tanton, N., "Results of a survey on television viewing distance," Tech. Rep. WHP 090, BBC Research & Development (July 2004).

[11] MPEG Test Subgroup, "Report of the formal verification tests on AVC (ISO/IEC 14496-10 ITU-T Rec. H.264)," Tech. Rep. N6231, ISO/IEC JTC1/SC29/WG11 (Dec. 2003).

[12] Oelbaum, T., Schwarz, H., Wien, M., and Wiegand, T., "Subjective performance evaluation of the SVC extension of H.264/AVC," in [*Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*], 2772 –2775 (2008).

Table 2: Subjective test results

Video sequence	Format	Visual quality [MOS]	Standard deviation	95% confidence interval	valid votes
Hurdles1	576i analogue	1.41	1.21	0.51	22
	576i digital	4.89	1.22	0.51	22
	720p	8.91	0.97	0.41	22
	1080i	8.57	0.95	0.41	21
Hurdles2	576i analogue	0.93	1.09	0.46	22
	576i digital	4.95	1.59	0.70	20
	720p	8.29	0.97	0.41	21
	1080i	7.84	1.30	0.54	22
Hurdles3	576i analogue	1.36	1.07	0.45	22
	576i digital	5.40	1.71	0.73	21
	720p	9.12	0.77	0.33	21
	1080i	8.80	0.83	0.35	22
Hurdles4	576i analogue	1.07	1.16	0.48	22
	576i digital	5.23	1.39	0.58	22
	720p	8.18	1.26	0.53	2
	1080i	7.95	1.26	0.53	22
Hammer1	576i analogue	1.19	1.33	0.57	21
	576i digital	1.93	1.54	0.66	21.
	720p	8.26	1.07	0.46	21
	1080i	7.77	1.13	0.47	22
Hammer2	576i analogue	1.52	1.09	0.45	22
	576i digital	5.46	1.30	0.54	22
	720p	8.66	0.98	0.41	22
	1080i	8.48	1.23	0.51	22
Hammer3	576i analogue	1.36	1.21	0.50	22
	576i digital	6.26	1.21	0.52	21
	720p	8.30	1.20	0.50	22
	1080i	7.88	1.31	0.56	21