Subjective Quality Assessment of H.265 versus H.264 Video Coding for High-Definition Video Systems

Emil Dumić, Sonja Grgić
Department of Wireless Communications
University of Zagreb, Faculty of Electrical Engineering and Computing
Zagreb, Croatia
emil.dumic@fer.hr, sonja.grgic@fer.hr

Krešimir Šakić
Radio Communication Department
Croatian Regulatory Authority for Network Industries (HAKOM)
Zagreb, Croatia
kresimir.sakic@hakom.hr

Domagoj Frank
HP Production LLC
Zagreb, Croatia
domagoj.frank@hpprodukcija.hr

***Abstract*—This paper gives our findings on subjective assessment of H.265 versus H.264 Video Coding for High-Definition Video Systems. For the purposes of our research a database consisting of 4 original HD video sequences was prepared with 30 degraded HD video sequences each, with various compression steps both in H.265/HEVC and H.264/AVC. The subjective assessment was conducted in one research laboratory in Croatia. The aim of this paper is to compare results obtained for H.265 versus H.264 Video Coding and to assess the performance of both standards. This assessment will help in the future decision on the coding standard that is going to be used in DVB-T2 networks in Croatia. Additionally, a status of the preparation activities for the allocation of the second digital dividend band in Croatia is given and further developments are described.**

Keywords—H.265; H.264; subjective assessment; high-definition; video quality

I. INTRODUCTION

Croatia is currently conducting preparation activities which are needed before the second digital dividend band (694 MHz to 790 MHz) is allocated to mobile services. The allocation of the second digital dividend has great impact on digital terrestrial television broadcasting in the UHF (470 – 790 MHz) band. As a big part (96 MHz) of the UHF band will be allocated to mobile services there will be a lot less spectrum for digital terrestrial television.

In order to sustain the number of services and competitive position of the digital terrestrial television broadcasting platform, a transition to a newer and more efficient system is needed – DVB-T2 [1]. The choice of the coding standard which is going to be used in the DVB-T2 system is also important. There are two choices. One is the proven and developed H.264/AVC [2] and the other is the new H.265/HEVC [3] coding standard. The choice of the coding system depends greatly on the chosen date for the transition to the DVB-T2 system. In [4] and [5] authors have tested different coding standards, including HEVC standard.

Countries that have decided for early allocation (i.e. 2016-2017) of the second digital dividend band to mobile services could choose the H.264/AVC coding standard, and those that have decided for a later date (i.e. 2018-2022) could incline towards H.265/HEVC. It is expected that the next year or two could be the turnaround when the price of receivers supporting H.265/HEVC falls substantially. One of the catalysts for this is Germany-s decision to use H.265/HEVC in its DVB-T2 networks [6]. However, Croatian and German DTT markets differ in great extent, where German market Pay TV saturation is close to 100% resulting in low significance of DTT, while DTT in Croatia is a dominant way of viewing television. Coding efficiency is only one of the aspects that need to be taken into account when deciding about optimal DVB-T to DVB-T2 migration strategy where specific Croatian technical, economic, market and social perspectives have to be taken into account [7].

Another opportunity of the transition to the DVB-T2 system is better picture quality. Currently there is no HD-quality free to air services in the DVB-T network in Croatia. The Croatian Regulatory Authority for Network Industries has conducted a Survey regarding picture quality in the DVB-T networks in Croatia and the majority of the broadcasters have stated that they plan to implement HD-quality in the production of their programmes in the next one to five years [8].

It needs to be recognized that most international HDTV events are made in either 1080i/25 or 1080i/29.98 (for 60 Hz countries). Nowadays HD television production is limited by the use of 1.485 Gbit/s HD-SDI (SMPTE 292M) infrastructures. It is supports 720p/50 format with 0,98 Gbit/s net transport rate (4:2:2 10 bit resolution) and 1080i/25 format with 1,036 Gbit/s net rate (4:2:2 10 bit). 1080p/50 format with the same resolution has 2,072 Gbit/s net rate, which requires dual link HD-SDI (SMPTE 372M) infrastructure or new 3 Gbit/s HD-SDI (SMPTE 424M). Studio and broadcast camera sensors operate already with at least 1920x1080 pixel progressive scanning at 24, 35, 50, 60 or more frames per second, but usually provide a down sampled output (1080i/25, 720p/50, 1080p/25). Cameras with dual link HD-SDI output and native 1080p/50 are still rare in television production. While some video recorders and servers can record 1080p/50, having end-to-end TV production is a big challenge and finally requires developments and significant investments. All of the above results that 70% of televisions broadcast their program using 1080i/25 format, 30% 720p/50 format [9], but there is still no broadcast in 1080p format.

For content produced in 1080i/25, 720p/50 broadcasters need to deinterlace and resize the signal before emission. The final image quality depends on the algorithms employed and the processing power of the deinterlacer used, but some motion artefacts are inevitable. 1080i/25 broadcasters have to accept that their deinterlacing is carried out in the consumer display. Although older consumer displays tended not to do this well, recent, higher-end displays can have reasonably good deinterlacers. On the other hand, for 720p/50 signal, consumer displays with 1920x1080 screen matrix need to upscale the picture.

All of these aspects have created a need for assessment of the performance of H.264/AVC and H.265/HEVC coding standards in the whole content production and delivery chain. This and future assessments will aid in the decision on the coding standard that is going to be used in DVB-T2 networks in Croatia.

This paper is organized as follows. Section II gives an overview of subjective and objective measures and related work. Section III presents the database construction which was designed and used for this subjective assessment. Section IV describes our subjective assessment and section V contains the results from the subjective assessment. Finally section VI describes our conclusions.

II. Overview of Subjective and Objective Measures

Subjective video quality assessment is known to be the most accurate reflection of user experience. Users experience is a combination of colour, motion, texture, audio and context. In a typical subjective assessment scenario, test subjects watch a number of original and/or degraded video sequences and rate their quality on a numeric scale. Subjective quality is often expressed as MOS (Mean Opinion Score) that represents a score from standard observer for a given sequence.

One of the traditional subjective video quality assessment methods is described in ITU-R BT.500-13 [10]. According to this recommendation the test methods have been divided in:

- double stimulus impairment scale (DSIS);
- double stimulus continuous quality scale (DSCQS);
- single stimulus methods;
- stimulus-comparison methods;
- single stimulus continuous quality evaluation (SSCQE);
- simultaneous double stimulus for continuous evaluation method (SDSCE).

Choosing reliable test subjects has great impact on the accuracy of subjective assessment. The test subject motivation needs to be kept at a high level. Various reimbursement schemes help in maintaining motivation levels, but also raise costs. Even with some reward scheme in place the number of test repetitions is limited as the test subjects soon develop biases and expectations which then lead to inaccurate results. All of these setbacks in conducting subjective video quality assessment are the reason why objective measures are often used in system and algorithmic optimization, although so far there is no universally accepted objective measure.

Objective video quality assessment methods are often used during the course of designing a video communication system and in other applications where there is a constant need for assessment of various algorithmic optimizations and content variations. Objective quality measures can be generally divided into three categories according to the reference information they use: full reference, reduced-reference and no-reference quality measures. Objective measures such as PSNR give a measure of how accurately an encoder can represent encoded video pixels. It uses all video pixels in order to asses a quality without taking into account whether the user perceived all of the pixels or not. Because of that, different image (e.g. Structural Similarity index, SSIM, [11]) and video (e.g. Video Quality Measure, VQM, [12]) quality measures have been developed. Their goal is to approximate the human quality perception (or Human Visual System, HVS) as much as possible, and consequently to correlate well with subjective measures (Mean Opinion Score, MOS).

To be able to compare some of the objective measures, usually MOS scores are processed to obtain DMOS (Difference MOS) grades which quantify the difference in subjective quality between original and degraded sequences (as full reference and reduced reference objective measures do).

III. Database construction

We used publicly available 4 full HD sequences [13], [14], to build a database which we used in our subjective assessment. According to ITU-R BT.500-13 [10], 4 different sequences (with different spatial and temporal characteristics) should be used for subjective evaluation. Sequences were compressed using x264 encoder (H.264/AVC) and x265 encoder (H.265/HEVC) using ffmpeg x64 downloaded from [15]. 4 sequences were used: CrowdRun, TreeTilt, PrincessRun and DanceKiss, each with duration of 10 seconds and with 50 fps, in YUV 4:2:0 format.

First frame of each sequence is shown on Fig. 1. The dynamic characteristics of the four reference sequences measured by spatial and temporal activity indices were computed for the left and right view according to the procedure defined in ITU-T recommendation P.910 [16]. The sequences' activity indices are plotted in Fig. 2, whose analysis shows that the sequences are diverse in terms of their dynamic characteristics and suitable for later subjective assessment.

(a)

(b)

(c)

(d)

Fig. 1. First frame from uncompressed sequences: a) CrowdRun, b) TreeTilt, c) PrincessRun and d) DanceKiss

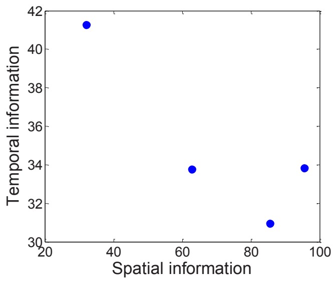
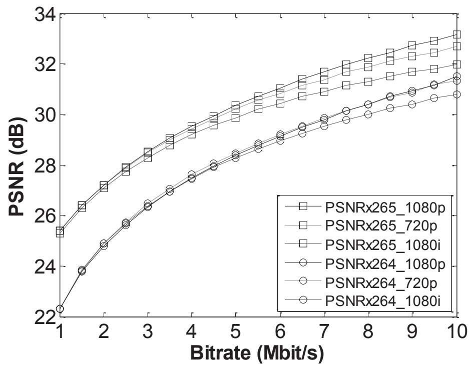


Fig. 2. Spatial versus temporal information

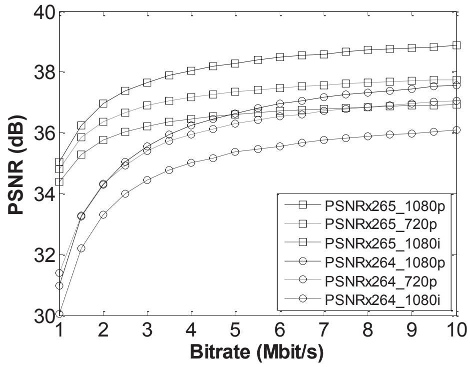
We have prepared sequences for testing with 3 spatial resolutions: 1080p/50 (resolution 1920x1080 pixels with 50 fps), 1080i/25 (resolution 1920x1080 pixels with 25 fps) and 720p/50 (resolution 1280x720 pixels with 50 fps). In order to fairly compare 1080i/25 and 720p/50 spatial resolutions without monitor (or software player) preprocessing, we have rescaled resolutions 720p/50 and 1080i/25 back to 1080p/50 (YUV 4:2:0 format) prior compression. For resolution 720p, we used lanczos3 interpolation for both downscaling and upscaling, while for resolution 1080i we used avisynth [17] for interlacing and "tdeint" filter [18] (doubling input frame rate) with avisynth for deinterlacing resulting in three uncompressed 1080p/50 (YUV 4:2:0) files per sequence: original (1080p), downscale-upscale degraded (720p) and interlace-framerate-deniterlace degraded (1080i). Otherwise, 720p resolution should be up-sampled by the full HD monitor itself (or software player), which would affect final PSNR as well as subjective assessment. In addition, interlaced content should be compressed which could also lower final PSNR (x264 uses MBAFF – Macro Block Adaptive Field Frame, while x265 for now has strictly experimental interlaced encoding) because it will depend on decoder quality and deinterlacing filter in monitor (or software player). It should be noted that, using this approach, we neglect potential differences in encoder efficiency related to handling input resolutions, because encoders effectively handle only 1080p/50 input format for compressing all test resolutions.

Each resolution was compressed with two-pass encoding, "placebo" setting (slowest, which means the best quality for given bitrate), with average bitrate ranging from 1 -10 Mbit/s with step 0.5 Mbit/s. Two-pass encoding is suitable for offline content, which produces higher quality of the content for the same average bitrate in comparison with single pass encoding. However, two-pass encoding cannot be used in live streaming. 1080p/50 was used as an output format for all tested spatial resolutions (1080p, 720p and 1080i) in order to cut out the effect of HD monitor and software player.

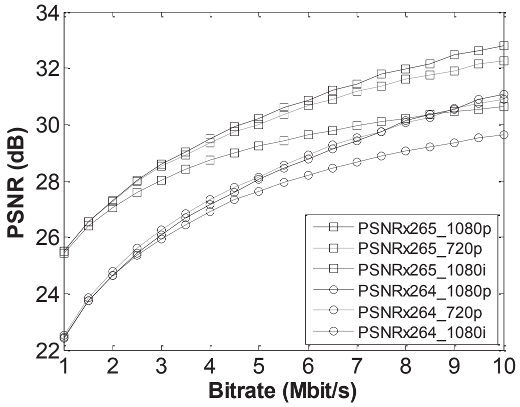
All sequences were firstly compared with original sequences (uncompressed 1080p) using PSNR and SSIM measures. Results for PSNR are presented in Fig. 3. Results for SSIM are very similar to the PSNR results (when comparing different resolutions and encoders), so those results are not shown.



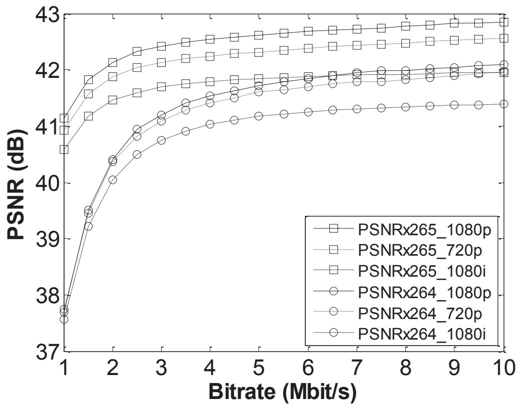
(a)



(b)



(c)



(d)

Fig. 3. PSNR for tested sequences: a) CrowdRun, b) TreeTilt, c) PrincessRun and d) DanceKiss

From Fig. 3 it can be concluded that for lower bitrates (up to 5 Mbit/s), x265 encoder outperforms x264 encoder for all resolutions. Also, 1080i resolution has lowest PSNR, while 720p has more similar PSNR with 1080p (which has the highest PSNR), when encoded with the same encoder. For higher bitrates, in some cases x264 with 720p and 1080p resolutions has similar PSNR like x265 with 1080i resolution. It should be also noted that higher PSNR or SSIM values could be obtained, by tuning x264 and x265 parameters for tested objective measures (separately for PSNR or SSIM). However, we omitted those tuning because of the later subjective assessment. Is should be noted that x264 and x265 encoders use "Psy RDO" measures to improve subjective quality, although this can have negative influence on PSNR or SSIM values.

IV. Subjective assessment

For subjective assessment we used a subset from the earlier explained sequences: x265 encoded with bitrates 2, 2.5, 3, 3.5, 4, 5 and 7 Mbit/s and x264 encoded with bitrates 3, 5 and 7 Mbit/s. Those bitrates were also interesting to test from the Fig. 3, as those bitrates represent higher changes between PSNR values. Also, those values represent usual transmitting bitrates of a TV channel to the viewer. Original sequences were also tested to be able to calculate DMOS (Difference Mean Opinion Score). Those sequences were compressed using x264 encoder in lossless mode (CRF, Constant Rate factor 0). We tested all mentioned resolutions (720p, 1080i and 1080p) giving overall 31 MOS results per video sequence. We used DMOS scores instead of MOS to minimize differences between MOS results of original sequences. MOS scores have in our test range 1-5, while DMOS scores have range 0-100 (lower DMOS means better quality).

The subjective assessment was divided in 3 parts, randomizing 120 degraded sequences. Every subset consisted of 40 degraded sequences (combined from all three resolutions and two compressions) together with 4 original sequences. Two sequences were added at the beginning of each test, representing one with the best and the one with the worst quality. These two sequences were used as an introduction to the test procedure. Those scores were later discarded. The grading scale consisted of scale from 1 to 5 with a step of 1 according to the [10]. Grade 1 represents bad, while 5 represents excellent. The subjective quality of the displayed video is graded on an Absolute Category Rating (ACR) with hidden reference (ACR-HR) scale. In ACR-HR, each original unimpaired video stimulus is included in the experiment but not identified as such. The quality ratings for these unimpaired stimuli are removed from the scores of the associated processed video sequences during data processing [16].

The monitor type used in the subjective assessments is a Samsung UE32H6400. Monitor settings used were the factory defaults. The computer used in the work was equipped an Intel i7-4790k processor, Solid State Drive (SSD), 16 GB of RAM and running Windows 8.1. Subjective assessment was made using MPC-HC media player 1.7.8 64-bit [19] with default values.

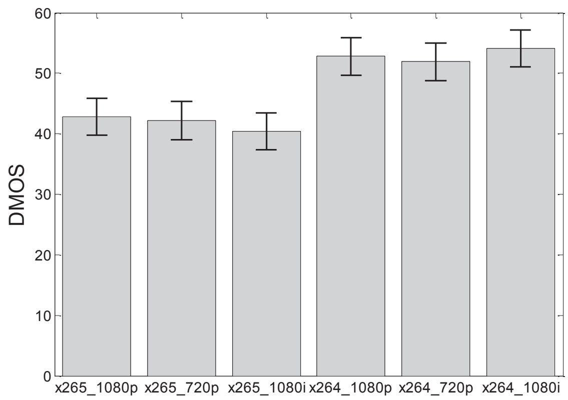
Overall we gathered 60 observations, resulting an average of 60/3=20 grades per video sequence, before elimination of outliers. The Lab was organized in the premises of Croatian Post, where observers participated in the subjective assessment on voluntary bases. Observers were primarily Croatian Post's employees of various age, gender and education levels.

DMOS scores were calculated from observers' scores using algorithm described in [20]. Firstly, each residual score (difference between the reference and degraded video sequence grades for the same observer) was converted to a *z*-score [20] to account for any differences in the use of the quality scale (differences in the location and range of values used by the observer). Afterwards, screening of the observers' scores was performed according to the ITU-R BT.500-13 [10] to discard scores from observers who differ too much from the average value. We used a fixed threshold for the number of outliers, namely 3, giving the threshold 3/40=7.5%. Using this threshold, we removed 5 observers from the pool. After outlier removal there were 17-19 grades per video sequence. Afterwards, z-scores for every observer were rescaled to the full range of 0-100. Finally, an average DMOS grade was calculated for each of the distorted video sequence as an arithmetic mean of all grades for that sequence.

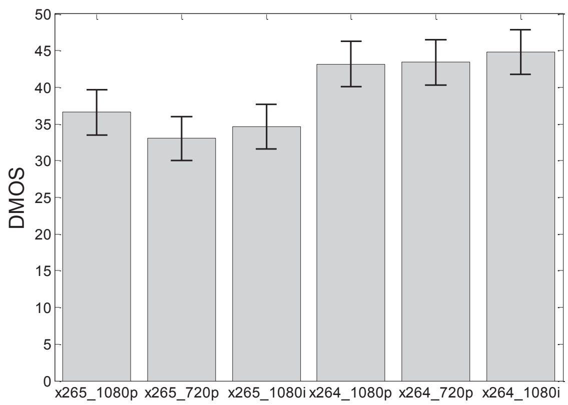
Original and compressed sequences, as well as DMOS results can be downloaded from [21].

V. RESULTS

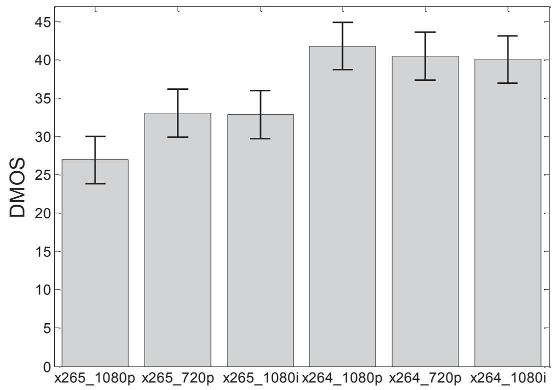
Average DMOS scores for all 4 tested sequences are shown on Fig. 4, for 3 bitrates: 3 Mbit/s, 5 Mbit/s and 7 Mbit/s. Both encoders, x264 and x265, were tested for mentioned bitrates, in all three spatial resolutions. Average DMOS (from all 4 sequences) are presented in Fig. 4, as well as associated confidence interval (according to Fisher's least significant difference procedure).



(a)



(b)



(c)

Fig. 4. Average DMOS for tested sequences: a) 3Mbit/s, b) 5 Mbit/s, c) 7 Mbit/s

We also compared average DMOS scores of all 4 sequences (in same spatial format) for x265 and x264 in TABLE I. For example when comparing DMOS scores for resolution 1080p to DMOS scores from x264 it can be seen that:

- x264, 3 Mbit/s (average DMOS 52.53) - x265 has lower DMOS on 2 Mbit/s (average DMOS 43.89) (we did not test bitrates lower than 2 Mbit/s)
- x264, 5 Mbit/s (average DMOS 43.87) - x265 has similar DMOS on 2 Mbit/s (average DMOS 43.89)
- x264, 7 Mbit/s (average DMOS 41.95) - x265 has similar DMOS on 2.5 Mbit/s (average DMOS 41.02)

We have further tested DMOS scores between x264 and x265 encoders for the next cases: x265_1080p, x265_720p, x265_1080i, x264_1080p, x264_720p, x264_1080i and bitrates 3 Mbit/s, 5 Mbit/s and 7 Mbit/s, using two sample t-test, with 2.5% significance level, one-tailed test (5% two-tailed test). This test has the null hypothesis that data in the tested samples (DMOS scores in our case) are independent random samples from normal distributions with equal means and equal but unknown variances, against the alternative that the means are not equal. Results are shown in TABLE II. '0' means that means are equal, '-' means that case in related row has statistically significant lower mean (e.g. better DMOS) than case in related column, while '+' means that case in related row has statistically significant higher mean (e.g. worse DMOS) than case in related column. Because t-test assumes that samples have normal distribution, we also tested each case using chi-square goodness of fit test against normal distribution, with 5% significance level. Results are shown in the last column of the TABLE II. 'NOK' means that samples do not have normal distribution, while 'OK' means that samples have normal distribution. It can be seen that some samples do not have normal distribution, so t-test could give unreliable results in those cases.

TABLE II. shows same conclusions which we obtained from Fig. 4: the x265 encoder nearly always outperformed x264 for the same bitrate. Exception is x265_1080p, 5 Mbit/s which has similar DMOS like x264_1080p, 5 Mbit/s.

The x264 encoder has similar DMOS across all tested spatial resolutions, for the same bitrate. The x265 encoder has similar DMOS across nearly all tested spatial resolutions, for the same bitrate. Exception is case x265_1080p, 7 Mbit/s, which outperformed all other tested cases (both x264 in all resolutions and x265 in other resolutions).

Also, the x265 encoder with 3 Mbit/s has similar DMOS like x264 with 5 Mbit/s and 7 Mbit/s. This shows that x265 has similar subjective score with half the bitrate (or less) of x264 encoder.

TABLE I. AVERAGE DMOS SCORES

	x265, 1080p	x265, 720p	x265, 1080i	x264, 1080p	x264, 720p	x264, 1080i
2 Mbit/s	43.89	50.56	46.57	-	-	-
2.5 Mbit/s	41.02	42.37	43.46	-	-	-
3 Mbit/s	43.52	41.66	40.38	52.53	51.33	53.59
3.5 Mbit/s	37.78	38.77	41.62	-	-	-
4 Mbit/s	36.51	38.18	38.51	-	-	-
5 Mbit/s	36.63	33.02	34.66	43.87	42.96	44.77
7 Mbit/s	26.79	33.23	33.03	41.95	39.91	40.44

In order to minimize the influence of display and player, we introduced limitations to our measurement that have to be taken into account for further interpretation of results:

1. Input (uncompressed) files were normalized to 1080p/50 format regardless of tested special resolution (1080p, 720p or 1080i), which neglects potential differences between x264 and x265 encoders in handling input 720p/50 and 1080i/25 formats.
2. Output (compressed) files were also normalized to 1080p/50 and, as a result, we have always encoded 1080p/50 uncompressed to 1080p/50 compressed format. Therefore, our measurements do not take into account H.264 and H.265 encoder efficiencies in handling interlaced content.

Further limitations stem from lab environment. We have used available open source codecs in a configuration suitable for off-line encoding. For the purpose of simulating television program distribution, HD-SDI interface with real-time encoder should be used.

TABLE II. TWO SAMPLE T-TEST BETWEEN DMOS SCORES FOR CASES: X265_1080P, X265_720P, X265_1080I, X264_1080P, X264_720P, X264_1080I AND BITRATES 3 MBIT/S, 5 MBIT/S AND 7 MBIT/S; x^2 TEST AGAINST NORMAL DISTRIBUTION

		3 Mbit/s						5 Mbit/s						7 Mbit/s						
		x265			x264			x265			x264			x265			x264			χ^2
		1080p	720p	1080i	1080p	720p	1080i	1080p	720p	1080i	1080p	720p	1080i	1080p	720p	1080i	1080p	720p	1080i	
x265 3Mbit/s	1080p	0	0	0	-	-	-	0	+	+	0	0	0	+	+	+	0	0	0	NOK
	720p	0	0	0	-	-	-	0	+	+	0	0	0	+	+	+	0	0	0	NOK
	1080i	0	0	0	-	-	-	0	+	0	0	0	0	+	+	+	0	0	0	NOK
x264 3Mbit/s	1080p	+	+	+	0	0	0	+	+	+	+	+	+	+	+	+	+	+	+	NOK
	720p	+	+	+	0	0	0	+	+	+	+	+	+	+	+	+	+	+	+	OK
	1080i	+	+	+	0	0	0	+	+	+	+	+	+	+	+	+	+	+	+	NOK
x265 5Mbit/s	1080p	0	0	0	-	-	-	0	0	0	0	-	-	+	0	0	0	0	0	OK
	720p	-	-	-	-	-	-	0	0	0	-	-	-	+	0	0	-	-	-	NOK
	1080i	-	-	0	-	-	-	0	0	0	-	-	-	+	0	0	-	-	-	NOK
x264 5Mbit/s	1080p	0	0	0	-	-	-	0	+	+	0	0	0	+	+	+	0	0	0	NOK
	720p	0	0	0	-	-	-	+	+	+	0	0	0	+	+	+	0	0	0	OK
	1080i	0	0	0	-	-	-	+	+	+	0	0	0	+	+	+	0	0	0	NOK
x265 7Mbit/s	1080p	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	OK
	720p	-	-	-	-	-	-	0	0	0	-	-	-	+	0	0	-	-	-	OK
	1080i	-	-	-	-	-	-	0	0	0	-	-	-	+	0	0	-	-	-	OK
x264 7Mbit/s	1080p	0	0	0	-	-	-	0	+	+	0	0	0	+	+	+	0	0	0	OK
	720p	0	0	0	-	-	-	0	+	+	0	0	0	+	+	+	0	0	0	OK
	1080i	0	0	0	-	-	-	0	+	+	0	0	0	+	+	+	0	0	0	NOK

VI. CONCLUSION

In this paper we have presented our results from the subjective assessment of H.265 versus H.264 Video Coding for High-Definition Video Systems. We have conducted our research on a database consisting of 4 original HD video sequences and 30 degraded HD video for each original video sequence.

The results have shown that x265 has similar subjective score with half the bitrate (or less) of x264 encoder. We also compared different spatial resolutions using the same encoder. Average DMOS scores were similar for x264 encoder and nearly similar for x265 encoder. This means that final spatial resolution (1080p, 1080i or 720p) of the broadcasting video stream can be chosen by the broadcasters, depending on their equipment. The results of this subjective assessment will help in the future decision on the coding standard that is going to be used in DVB-T2 networks in Croatia.

Further research is needed on the availability, compatibility and performance of equipment supporting H.265 in the whole content production and delivery chain, as well as performance of real time H.264 and H.265 encoders. Different network parameters could be also incorporated in future research, such as the influence of packet losses on final video quality.

REFERENCES

[1] EN 302 755 V.1.3.1 "Digital Video Broadcasting, Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting sytem (DVB-T2)", ETSI, 2012

[2] Recommendation ITU-T H.264 (02/2014) "Advanced video coding for generic audiovisual services", ITU, 2014

[3] Recommendation ITU-T H.265 (10/2014) "High efficiency video coding", ITU, 2014

[4] P. Hanhart, M. Rerabek, F. De Simone, T. Ebrahimi, "Subjective quality evaluation of the upcoming HEVC video compression standard", SPIE Optical Engineering + Applications, Proceedings of SPIE (ISSN: 1996-756X), vol. 8499, 2012.

[5] Ohm, J-R., Sullivan, G.J., Schwarz, H., Thiow Keng Tan, Wiegand, T., "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)." IEEE Transactions on Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1669-1684, 2012

[6] https://www.dvb.org/news/tender-for-dvb_t2-in-germany/country/germany, available on 27 March 2015

[7] D. Frank, E. Dumic, "Planning the Migration of Digital Terrestrial Broadcasting in Croatia to DVB-T2 Standard", Media Research, Vol. 20, No. 2, pp. 193-210, 2014.

[8] HAKOM presentation from the second forum ond the future of the UHF band, http://www.hakom.hr/UserDocsImages/2015/dokumenti/Prezentacija-Regulatorni%20okvir%20priprema%20za%20DD2_20150223.pdf, available on 27 March 2015

[9] P. Larbier, "HEVC & Broadcast Content", Ateme Whitepaper, Dec. 2013.

[10] ITU-R BT.500-13 "Methodology for the subjective assessment of the quality of television pictures", International Telecommunication Union/ITU Radiocommunication Sector, 2012.

[11] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. on Image Proc., Vol. 13, No. 4, pp. 600-612., 2004.

[12] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality", IEEE Trans. Broadcast, vol. 50, no. 3, pp. 312–322, Sep. 2004

[13] C. Keimel, A.Redl and K. Diepold, "The TUM High Definition Video Data Sets", Fourth International Workshop on Quality of Multimedia Experience (QoMEX 2012), pp. 97 - 102, 2012

[14] ftp://ftp.ldv.ei.tum.de/videolab/public/TUM_1080p50_Data_Set/, available on 27 March 2015

[15] http://ffmpeg.zeranoe.com/builds/, available on 27 March 2015

[16] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications", 2008.

[17] http://sourceforge.net/projects/avisynth2/, available on 27 March 2015

[18] http://avisynth.org.ru/docs/english/externalfilters/tdeint.htm, available on 27 March 2015

[19] http://mpc-hc.org/, available on 27 March 2015

[20] A.Zaric, N.Tatalovic, N.Brajkovic, H.Hlevnjak, M.Loncaric, E.Dumic, S.Grgic, "VCL@FER Image Quality Assessment Database", AUTOMATIKA Vol. 53, No. 4, pp. 344–354, 2012

[21] http://goo.gl/wSEkHd