Subjective Video Quality Assessment of H.265 Compression Standard for Full HD Resolution

Miroslav UHRINA, Juraj BIENIK, Martin VACULIK

Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Univerzitna 8215/1, 010 07 Zilina, Slovakia

miroslav.uhrina@fel.uniza.sk, juraj.bienik@fel.uniza.sk, martin.vaculik@fel.uniza.sk

DOI: 10.15598/aeee.v13i5.1503

Abstract. *Recently increasing interest in multimedia services leads to requirements for quality assessment, especially in the video domain. There are many factors that influence the video quality. Compression technology and transmission link imperfection can be considered as the main ones. This paper deals with the assessment of the impact of H.265/HEVC compression standard on the video quality using subjective metrics. The evaluation is done for two types of sequences with Full HD resolution depending on content. The paper is divided as follows. In the first part of the article, a short characteristic of the H.265/HEVC compression standard is written. In the second part, the subjective video quality methods used in our experiments are described. The last part of this article deals with the measurements and experimental results. They showed that quality of sequences coded between 5 and 7 Mbps is for observers sufficient, so there is no need for providers to use higher bitrates in streaming than this threshold. These results are part of a new model that is still being created and will be used for predicting the video quality in networks based on IP.*

Keywords

ACR, DSCQS, DSIS, H.265/HEVC, subjective assessment, video quality.

1. Introduction

In the last years the demand for the multimedia services, which stand for broadcasting, transmission and receiving the video, audio and other data in one stream - the multimedia stream has been increased. The video quality is affected by many factors, especially by compression technology and transmission link imperfection. Nowadays many new compression standards are being developed, e.g. H.265/HEVC, VP9 or DAALA. Because of that fact, video quality assessment became and still plays an important role in the research area.

2. State of the Art

Because of mentioned growth of services in video domain a need to create a new model that will predict the quality occurs. In papers [1], [2], [3], [4], [5], [6], [7], [8], [9] the impact of bitrate on the video quality using objective metrics and in the paper [10] using subjective metrics is researched. The measurements in these works show the dependence of resolution, bitrate and type of content on video quality. References [11], [12], [13] focus on the degradation of quality caused by delay and packet loss. Work [14] analyses in detail the impact of network utilization and set policies on variable component of total delay. Since the final delay and packet loss are factors depending on full network utilization and QoS policy applied to prioritized data flow processing by routers, it is necessary to consider this link as well. What is still missing is the model used for predicting the video quality that is influenced by compression and network impact. It is necessary to mention that such model should correlate well with subjective perception. Due to this fact it is needed to do many subjective tests. This paper focuses on video quality evaluation of the newest compression standard H.265/HEVC for Full HD resolution using subjective methods.

3. H.265/HEVC Compression Standard

H.265/HEVC (High Efficiency Video Coding) compression standard, a new generation successor of

H.264/MPEG-4 AVC standard, expands a family of MPEG standards. This standard has been developed in cooperation with ISO / IEC Moving Picture Experts Group (MPEG) and ITU-T video Coding Experts Group (VCEG), working together in a partnership known as the Joint Collaborative Team on Video Coding (JCT-VC). It has been developed with a priority for streaming HDTV and UHDTV signal via a network or for storage such signal on storage media. HEVC algorithms have been designed to improve substantially coding efficiency compared to H.264/MPEG-4 AVC, i.e. to reduce bitrate requirements by a half with comparable image quality at the expense of increased computational complexity. It has been proposed with the goal of allowing video content to have a data compression ratio of up to 1000:1. The video coding layer of H.265/HEVC employs the same hybrid approach (inter-/intrapicture prediction and 2-D transform coding) used in all video compression standards since H.261. However, H.265/HEVC contains many incremental improvements such as:

- More flexible partitioning, from large to small partition sizes.
- Greater flexibility in prediction modes and transform block sizes.
- More sophisticated interpolation and deblocking filters.
- More sophisticated prediction and signaling of modes and motion vectors
- Features to support efficient parallel processing.

H.265/HEVC is the standard of the future. Nowadays it is used mostly for testing and just rarely on websites. Because of high computational complexity and necessity of the implementation of the new decoder to devices, communication market and ordinary users have to wait for a few years to use it [15], [16], [17].

4. Subjective Video Quality Assessment

The video quality evaluation can be divided into the objective and subjective assessment. The subjective assessment consists of the use of human observers (people) who watch the sequences and score the video quality. It is the most reliable way how to determine the video quality and should not be replaced with objective assessment. The disadvantage of this method is that it is time-consuming and human resources are needed. Owing to this fact, the objective methods are mostly preferred and used.

The well-known subjective methods are:

- Double Stimulus Impairment Scale (DSIS) also known as Degradation Category Rating (DCR).
- Double Stimulus Continuous Quality Scale (DSCQS).
- Single Stimulus Continuous Quality Evaluation (SSCQE).
- Absolute Category Rating (ACR) also known as Single Stimulus (SS).
- Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) [18], [19].

To achieve reliable results, minimum 15 observers should be used. They should be non-experts, in the sense that they are not directly concerned with television picture quality as a part of their normal work and they are not experienced assessors. The number of assessors needed for the tests depends upon the sensitivity and reliability of the test procedure adopted and upon the anticipated size of the effect sought. Before the test session, assessors should be introduced to many factors as for instance the method of assessment, the types of impairments, the grading scale, the sequence, the timing (the reference, the test sequence time duration, the time duration for voting) and so on.

The presentation structure of a test session is shown in the Fig. 1.

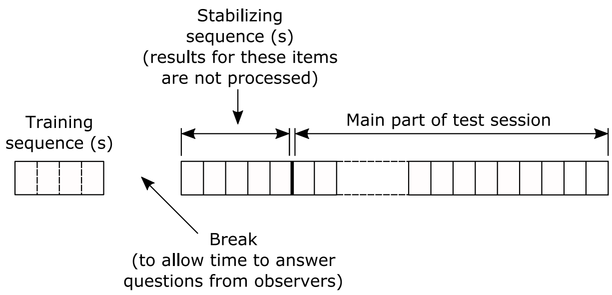

Fig. 1: The presentation structure of the test session.

The whole session should last up to 30 minutes. At the beginning of the first session, some sequences (from three to five) should be shown to stabilize the observers' opinion. The data obtained from these presentations must not be taken into account in the results of the test. A random order should be used for the presentations, but the test condition order should be arranged so that any effects on the grading of tiredness or adaptation are balanced out from session to session. Some of the presentations can be repeated from session to session to check coherence.

Finally after the test session the calculation of the mean score (MOS) is done:

$$\bar{u}_{jkr} = \frac{1}{N}\sum_{i=1}^{N} u_{ijkar}, \quad (1)$$

where u_{ijkrs} is the score of observer i for test condition j, sequence k, repetition r and N stands for a number of observers.

Also, the 95 % confidence interval, which is derived from the standard deviation and size of each sample is calculated. It is given by:

$$\delta_{jkr} = 1.96 \cdot \frac{S_{jkr}}{\sqrt{N}}, \quad (2)$$

where:

$$\delta_{jkr} = \sqrt{\sum_{i=1}^{N} \frac{(u_{jkr} - u_{ijkr})^2}{(N-1)}}, \quad (3)$$

[18], [19], [20], [21].

In our experiments, DSIS, DSCQS and ACR methods were used.

4.1. The Double-Stimulus Impairment Scale Method (DSIS)

In this method two sequences are shown to the assessor - the unimpaired (the reference) sequence and the same sequence impaired (the test one). The reference sequence is shown before the test one (Fig. 2), and the viewer knows which one is the reference and which one the test.

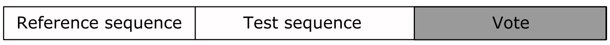

Fig. 2: The presentation structure of the test material.

After watching both sequences, the assessor is asked to rate the second, keeping in the mind the first. The five-grade impairment scale is used:

- 5 imperceptible.
- 4 perceptible, but not annoying.
- 3 slightly annoying.
- 2 annoying.
- 1 very annoying [18], [19], [20].

4.2. The Double-Stimulus Continuous Quality-Scale Method (DSCQS)

By this method also two sequences are shown to the assessor - the unimpaired (the reference) sequence and the same sequence impaired (the test one) but the viewer is not informed which one is the reference and which one is the test (Fig. 3). The position of the reference sequence is changed in a pseudo-random fashion.

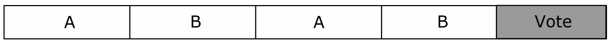

Fig. 3: The presentation structure of the test material.

After watching both sequences, the assessor is asked to rate the second, keeping in the mind the first. The five-grade impairment scale is used:

- Excellent (80–100 = 5).
- Good (60–79 = 4).
- Fair (40–59 = 3).
- Poor (20–39 = 2).
- 1 very annoying [18], [19], [20].

4.3. The Absolute Category Rating Method (ACR)

This method is also called single stimulus method (SS). In this method, only the impaired (the test) sequence is shown to the assessor (Fig. 4) so the viewer does not know which quality is the reference sequence.

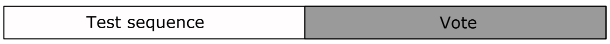

Fig. 4: The presentation structure of the test material.

The assessor is asked to rate the quality of the test sequence based on the level of the quality he has in his opinion for it after watching it. The five-level grading scale is used:

- 5 excellente.
- 4 good.
- 3 fair.
- 2 poor.
- 1 bad [18], [19], [20], [21].

5. Measuremens

In our experiments two types of test source sequences (SRCs) depending on content were used:

- One with a dynamic scene called "Basketball" (Fig. 5).

- One with slow motion called "Cactus" (Fig. 6).

Both sequences were in the Full HD resolution, i.e. 1920×1080 pixels and 16:9 aspect ratio with 50fps (frames per second). The length of these sequences was 500 frames, i.e. 10 seconds. The measurement procedure consists of four steps:

- First, both sequences were downloaded from [22] in the uncompressed format (*.yuv) and used as the reference sequences.
- Afterwards, they were encoded to H.265/HEVC compression standard using x265 tool [23]. The sequences were coded to 1, 3, 5, 7, 9 Mbps, which means 10 hypothetical reference circuits (HRCs) were used - for each SRC five HRCs restricted by maximum bitrate. The parameters of the encoded sequences were set to Main Profile, Level 4. The GOP parameter was set to $N = 12$ and $M = 3$ which means the GOP length was 12 and two B frames between two successive P frames were stored. The command line settings of x265 tool is shown in Tab. 1.
- Then, the sequences using the same x265 tool were decoded back to the format *.yuv.
- Finally, the video quality was evaluated. It was done using people (observers) who scored the video quality using subjective metrics DSIS, DSCQS, ACR.

Fig. 5: One frame of the "Basketball" test sequence.

The whole process of measuring of both sequences is shown in Fig. 7.

Fig. 6: One frame of the "Cactus" test sequence.

Tab. 1: Command line settings of x265 tool.

Command line options	x265 command line settings
Input options	–input Basketball_1920x1080_50fps _420_8bit.yuv –input-res 1920x1080 –input-csp i420 –input-depth 8 –fps 50
Slice decision options	–no-open-gop –keyint 12 –min-keyint 12 –no-scenecut –bframes 2 –b-adapt 0
Quality rate control and rate distortion options	–bitrate 3000 –vbv-maxrate 3000 –vbv-bufsize 3000
Debugging options	–recon Basketball_1920x1080_50fps_420 _8bit_GOP12-BF2_3M_x265.yuv –recon Basketball_1920x1080_50fps_420 _8bit_GOP12-BF2_3M_x265.mp4

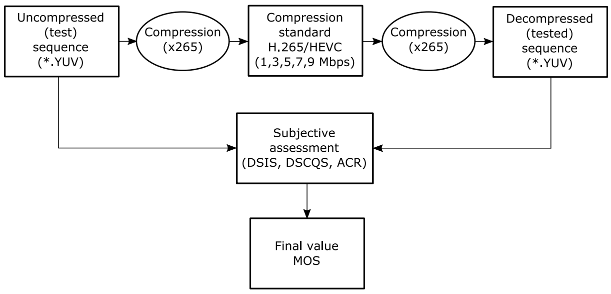

Fig. 7: The process of measuring the impact of H.265/HEVC compression standard on the video quality.

6. Experimental Results

By subjective assessment 30 assessors were used (24 men and 6 women). Their age was in the range from 20 to 28 years, the average age was 22 years. All figures below (from 8 to 12) show the impact of bitrate on the quality of experience for both test sequences coded into H.265/HEVC compression standard.

As it can be seen from the graphs – the observers recognized the differences in the quality of both test sequences, especially in lower bitrates. It can be said that people assessed the quality of both sequences coded to bitrate between 5 and 7 Mbps as a good one. After this threshold, the observers did not evaluate the quality of sequence coded to 9 Mbps with too higher marks.

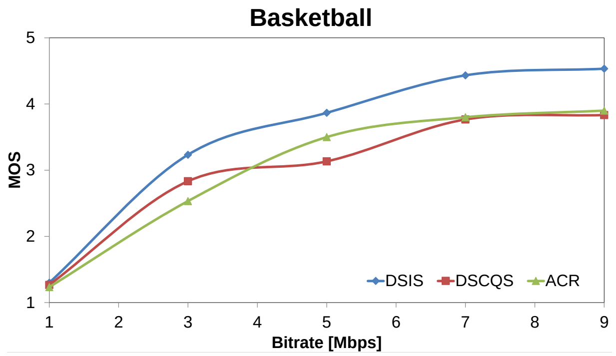

Fig. 8: The relationship between video quality (MOS) measured by all subjective metrics and bitrate for "Basketball" test sequence.

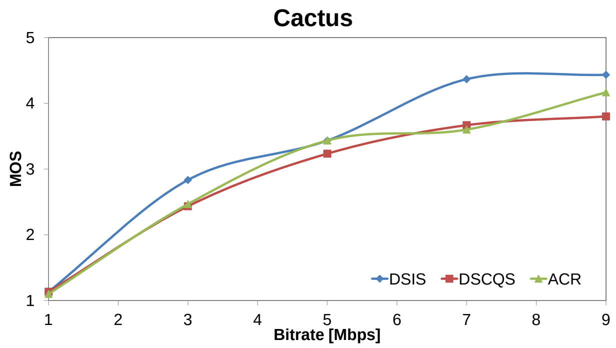

Fig. 9: The relationship between video quality (MOS) measured by all subjective metrics and bitrate for "Cactus" test sequence.

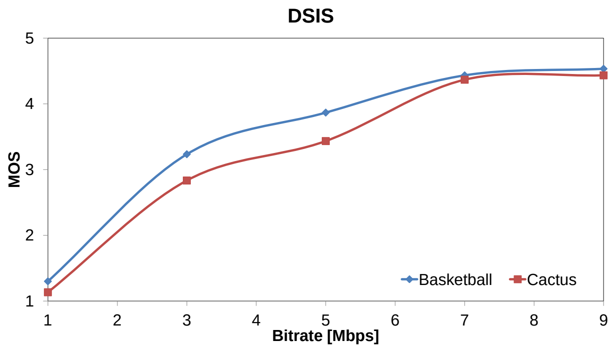

Fig. 10: The relationship between video quality (MOS) measured by DSIS subjective metric and bitrate for both test sequences.

This fact means that there is no need for providers to use higher bitrates in streaming than this threshold, so they can save space in transmission chain and use it for other channels or services. The plots also show that the assessors did not rate the video quality with extremities which confirms the fact that people mostly do not like to give the extreme values.

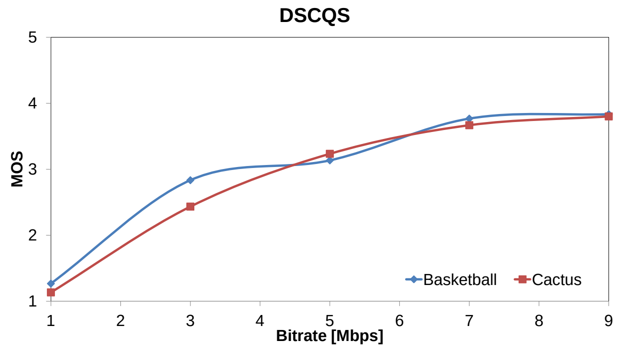

Fig. 11: The relationship between video quality (MOS) measured by DSCQS subjective metric and bitrate for both test sequences.

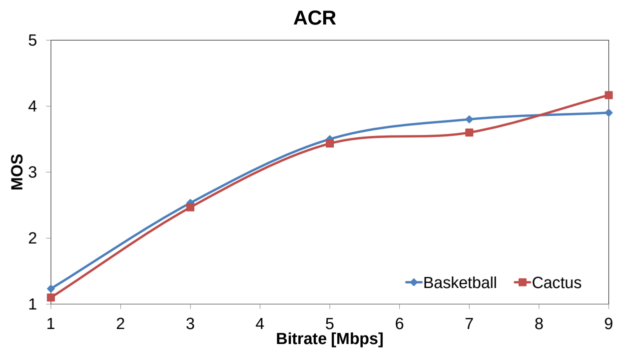

Fig. 12: The relationship between video quality (MOS) measured by ACR subjective metric and bitrate for both test sequences.

7. Conclusions

This paper dealt with assessment of the impact of H.265/HEVC compression standard on the video quality using selected subjective metrics. The aim of this paper was to research how the people evaluate the video quality affected by the bitrate. The assessment was done for two types of Full HD sequences depending on content. The results showed that the observers evaluated the quality of both sequences coded to bitrate between 5 and 7 Mbps as a good one. This fact leads to the conclusion that there is no need for providers to use higher bitrates in streaming than this threshold, so they can save space in transmission chain and use it for other channels or services. The results were subsequently used for computing correlation coefficients between objective and subjective methods. All results are part of a new model that is still being created and will be used for predicting the video quality in networks based on IP. The next step should be an analysis of the impact of H.265/HEVC compression standard with 4K resolution on video quality using subjective metrics.

References

[1] UHRINA, M., J. HLUBIK and M. VACULIK. Impact of Compression on the Video Quality. *Advances in Electrical and Electronic Engineering*. 2012, vol. 10, no. 4, pp. 251–258. ISSN 1336-1376. DOI: 10.15598/aeee.v10i4.737.

[2] UHRINA, M., J. FRNDA, L. SEVCIK and M. VACULIK. Impact of H.264/AVC and H.265/HEVC compression standards on the video quality for 4K resolution. *Advances in Electrical and Electronic Engineering*. 2014, vol. 12, no. 4, pp. 368–376. ISSN 1336-1376. DOI: 10.15598/aeee.v12i4.1216.

[3] UHRINA, M., J. FRNDA, L. SEVCIK and M. VACULIK. Impact of H.265 and VP9 Compression Standards on the Video Quality for 4K Resolution. In: *22nd Telecommunications Forum TELFOR*. Belgrade: IEEE, 2014, pp. 905–908. ISBN 978-1-4799-6190-0. DOI: 10.1109/TELFOR.2014.7034551.

[4] GROIS, D., D. MARPE, A. MULAYOFF and O. HADAR. Performance Comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC Encoders. In: *Picture Coding Symposium (PCS)*. San Jose: IEEE, 2013, pp. 394–397. ISBN 978-1-4799-0292-7. DOI: 10.1109/PCS.2013.6737766.

[5] RAO, K. Video coding standards: AVS China, H.264/MPEG-4 PART 10, HEVC,VP9, DIRAC and VC-1. In: *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. Poznan: IEEE, 2013, pp. 1–11. ISBN 2326-0262.

[6] KIM, K.-I., S. LEE, Y. PIAO and J. CHEN. Coding Efficiency Comparison of New Video Coding Standards: HEVC vs VP9 vs AVS2 Video. In: *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. Chengdu: IEEE, 2014, pp. 1–6. ISBN 1945-7871. DOI: 10.1109/ICMEW.2014.6890700.

[7] RAMZAN, N., Z. PERVEZ and A. AMIRA. Quality of Experience Evaluation of H.265/MPEG-HEVC and VP9 Comparison Efficiency. In: *26th International Conference on Microelectronics (ICM)*. Doha: IEEE, 2014, pp. 220–223. ISBN 978-1-4799- 8153-3. DOI: 10.1109/ICM.2014.7071846.

[8] KUFA, J. and T. KRATOCHVIL. Comparison of H.265 and VP9 Coding Efficiency for Full HDTV and Ultra HDTV Applications. In: *25th International Conference Radioelektronika*. Pardubice: IEEE, 2015, pp. 168–171. ISBN 978-1-4799-8117-5. DOI: 10.1109/RADIOELEK.2015.7128999.

[9] RERABEK, M., P. HANHART, P. KORSHUNOV and T. EBRAHIMI. Quality Evaluation of HEVC and VP9 Video Compression in Real-Time Applications. In: *7th International Workshop on Quality of Multimedia Experience (QoMEX)*. Pylos-Nestoras: IEEE, 2015, pp. 1–6. ISBN 978-1-4799-8958-4. DOI: 10.1109/QoMEX.2015.7148088.

[10] UHRINA, M., J. HLUBIK and M. VACULIK. The comparison of MPEG compression standards using different subjective video quality methods. In: *11th International Conference on Knowledge in Telecommunication Technologies and Optics – KTTO*. Szczyrk: VSB–Technical University of Ostrava, 2011, pp. 56–59. ISBN 978-80-248-2399-7.

[11] VOZNAK, M. E-model modification for case of cascade codecs arrangement. *International Journal of Mathematical Models and Methods in Applied Sciences*. 2011, vol. 5, iss. 8, pp. 1439–1447. ISSN 1998-0140.

[12] FRNDA, J., L. SEVCIK, M. UHRINA and M. VOZNAK. Network Degradation Effects on Different Codec Types and Characteristics of Video Streaming. *Advances in Electrical and Electronic Engineering*. 2014, vol. 12, no. 4, pp. 377–383. ISSN 1336-1376. DOI: 10.15598/aeee.v12i4.1218.

[13] FRNDA, J., M. VOZNAK and L. SEVCIK. Network Performance QoS Prediction. *Advances in Intelligent Systems and Computing AISC*. 2015, vol. 297, no. 1, pp. 165–174. ISSN 2194-5357. DOI: 10.1007/978-3-319-07776-5_18.

[14] FRNDA, J., M. VOZNAK, J. ROZHON and M. MEHIC. Prediction Model of QoS for Triple Play Services. In: *21st Telecommunications Forum TELFOR*. Belgrade: IEEE, 2013, pp. 733–736. ISBN 978-1-4799-1419-7. DOI: 10.1109/TELFOR.2013.6716334.

[15] SULLIVAN, G. J., J.-R. OHM, W.-J. HAN and T. WIEGAND. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Transactions on Circuits and Systems for Video technology*. 2012, vol. 22, no. 12, pp. 1649–1668. ISSN 1051-8215. DOI: 10.1109/TCSVT.2012.2221191.

[16] WINKEN, M., P. HELLE, D. MARPE, H. SCHWARZ and T WIEGAND. Transform Coding in the HEVC Test Model. In: *18th IEEE International Conference on Image Processing*. Brussels: IEEE, 2011, pp. 3693–3696. ISBN 978-1-4577-1304-0. DOI: 10.1109/ICIP.2011.6116521.

[17] RICHARDSON, I. HEVC: An Introduction to High Efficiency Video Coding. *vcodex* [online].

2013. Available at: http://www.vcodex.com/images/uploaded/342512928230717.pdf.

[18] WINKLER, S. *Digital Video Quality: Vision Models and Metrics*. Hoboken: John Wiley and Sons Ltd., 2005. ISBN 0-470-02404-6.

[19] WU, H. R. and K. R. Rao. *Digital Video Image Quality and Perceptual Coding*. London: Taylor and Francis Group LLC, 2006. ISBN 0-8247-2777-0.

[20] RECOMMENDATION ITU-R BT.500-13. *Methodology for the subjective assessment of the quality of television pictures*. Geneva: ITU-T, 2013.

[21] RECOMMENDATION ITU-T P.910. *Subjective video quality assessment methods for multimedia applications*. Geneva: ITU-T, 2008.

[22] Test Sequence. *Ultra Video Group* [online]. 2014. Available at: http://ultravideo.cs.tut.fi/#testsequences.

[23] H.265/HEVC Encoder, x265 tool. *VideoLAN Organization* [online]. 2014. Available at: http://www.videolan.org/developers/x265.html. 2013.

About Authors

Miroslav UHRINA was born in 1984 in Zilina, Slovakia. He received his M.Sc. and Ph.D. degrees in Telecommunications at the Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, at the University of Zilina in 2008 and 2012, respectively. Nowadays he is an assistant professor at the same department. His research interests include audio and video compression, video quality assessment, TV broadcasting and IP networks.

Juraj BIENIK was born in 1987 in Zilina, Slovakia. He received his M.Sc. in Telecommunications at the Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, at the University of Zilina in 2012. Nowadays he is a Ph.D. student at the same department. His research interests include audio and video signal processing, funkcionality and optimalisation of networks and video quality assessment.

Martin VACULIK was born in 1951. He received his M.Sc. and Ph.D. in Telecommunications at the University of Zilina, Slovakia in 1976 and 1987 respectively. In 2001 he was habilitated as associate professor of the Faculty of Electrical Engineering at the University of Zilina in the field of Telecommunications. Currently he works as a head of Department of Telecommunications and Multimedia. His interests cover switching and access networks, communication network architecture, audio and video applications.