RIGA TECHNICAL UNIVERSITY
Faculty of Electronics and Telecommunications
Institute of Telecommunications

Romass PAULIKS
Student of the Doctoral Study Program “Telecommunication and Computer Network”

RESEARCH OF VIDEO TRANSMISSION SERVICE QUALITY IN PACKET NETWORKS

Summary of the Doctoral Thesis

Scientific supervisor
Dr. sc. ing., Professor
I. SLAIDIŅŠ

RTU Press
Riga 2016

Pauliks R. Research of Video Transmission Service Quality in Packet Networks. Summary of the Doctoral Thesis. – R.: RTU Press, 2016.– 71 pp.

Printed in accordance with the decision of the ETF Promotion Board “RTU P-08” No. 33 of March 17, 2016.

ISBN 978-9934-10-819-8

THE DOCTORAL THESIS
PROPOSED TO RIGA TECHNICAL UNIVERSITY FOR THE PROMOTION TO THE SCIENTIFIC DEGREE OF DOCTOR OF ENGINEERING SCIENCES

The defence of the thesis submitted for the promotion to the scientific degree of Doctor of Engineering Sciences will take place at an open session on June 30, 2016 on Āzenes Street 12, auditorium 38, Riga Technical University, Faculty of Electronics and Telecommunications.

OFFICIAL REVIEWERS

Professor, *Dr.habil.sc.ing.* Ernests Petersons, Riga Technical University, Promotion Council, Latvia

Professor, *Dr.sc.ing.* Aleksandrs Grakovskis, Transport and Telecommunication Institute, Latvia

Professor, *Dr.habil.phys* Eligijus Sakalauskas, Kaunas University of Technology, Lithuania

DECLARATION OF ACADEMIC INTEGRITY

I hereby declare that the Doctoral Thesis submitted for the review to Riga Technical University for the promotion to the scientific degree of Doctor of Economics is my own and does not contain any unacknowledged material from any source. I confirm that this Thesis has not been submitted to any other university for the promotion to any other scientific degree.

Romass Pauliks ……………………………. (Signature)

Date: ………………………

The Doctoral Thesis has been written in Latvian and includes introduction, 4 chapters, conclusions, bibliography with 236 reference sources, and 20 appendices. It has been illustrated by 40 figures. The total volume of the Thesis is 190 pages.

GENERAL DESCRIPTION OF THE THESIS

The Topicality of the Research

Every year on November 21, the Worldwide Day of Television established in 1996 by General Assembly (UN) is celebrated. During the period of several years, the annual statistical data of *TNS Latvia* [1] prove that within a year the inhabitants of Latvia – watch TV programs on average 4–5 hours a day in various ways, including while sitting, sleeping, driving a car, using a computer, tablet *PC* or a mobile phone.

Nowadays there are many technologies of content broadcasting by television, video and multimedia which are characterized, compared and classified in the Thesis, but the main research is devoted to real time video broadcasting technologies using IP protocol. In the 21st century, the number of Internet users has increased rapidly, especially the number of mobile Internet users and it is going to increase within next five years even faster [2]. The content variety offered by the Internet is also going to increase where during the past decade a very essential role has been played by multimedia data stream. Every year multimedia data stream takes larger and larger amounts of data [2], including images and videos placed by users themselves in different social networks (*draugiem.lv*, *odnoklasniki.ru*), video calls (*skype*), video clips (*youtube.com*, *vimeo.com*), short films, feature video films, and classical *TV* broadcasting channels. In recent years, a significant increase of data stream amount on the Internet has been created by cloud computing [3], i.e. technologies of data storage and computing which provide users' data, including multimedia data stream placement in the Internet data centres. Demanded by market conditions the consumers' wish to use higher quality video content grows essentially as a result of which the following correlations should be taken into consideration:

- The higher the quality of video, the higher the user's evaluation;
- The higher the quality of video, the larger amount of data is necessary;
- The larger the amount of data, the faster data transmission channels are necessary;
- The higher the quality of video, the higher performance data storage, computing and data transmission equipment is necessary.

– It could be concluded that increase of video quality is a catalyst for fast development of television, video and multimedia technologies. Fast development of television, video and multimedia technologies facilitates the development of other technologies, for example, the demand for fast performance Internet, high performance video and network equipment as well as for larger amounts of data storage.

Historically since the analogue black-and-white television was changed to colour *TV* the development of technologies rapidly changed towards the increase of the picture quality. While transferring from analogue to digital television, the necessity appeared to decrease the amount of transmitted information. It can be implemented in both ways with special compression methods (*lossless*), decreasing the amount of transmitted data without any losses at the same time not decreasing the quality of the image, and using the peculiarities of the human sight and perception. Therefore, such compression methods (*lossy*) can be called picture quality limitation by losses. The development of new compression technologies has initiated a new stage towards the increase of quality of the digital television picture. The main task of new compression technologies is to increase or ensure unchanged quality of the image by reducing the amount of coded data, thus, creating the possibility to transmit several television programs instead of one program in one channel (6, 7 or 8 MHz).

During the past decade, the usual group of cables, earth, mobile and satellite television technologies of *DVB* (*Digital Video Broadcasting*) [4] has been supplemented by Internet Protocol Television IPTV [5], where to broadcast the television channels packet commutation channels are used with *QoS* (*Quality of Service)* [6] usually guaranteed within a single operator network. However, in the recent years more and more popularity has been gained by *OTT* (*Over-the-top* or *Over-the-Internet*) video [7] and *i-TV* (*Internet TV*), which could be mentioned as *IPTV* successor because in both cases the *IP* protocol is used. *OTT* and *i-TV* are not connected to the operator's network and for the transmission of television channels Internet is used without *QoS* guarantees. *IPTV*, *OTT* and *i-TV* can ensure the service of *ITV* (*Interactive Television)*. Video broadcast, while using the Internet or packet commutation network technologies with non-guaranteed (*best-effort)* data stream supply, should take into account packet delays, jitter, burst, as well as the overflow of network node buffers and packet loss which may essentially lower the quality of the image. The compression limits the quality of images, but packet losses which appear as a result of broadcast lower the quality of the picture. Different streaming protocols can be used for *i-TV* and *OTT* solutions, the most popular of which are streaming protocols of *HTTP* (*Hypertext Transfer Protocol),* which ensure dynamic increase or decrease of the video quality depending on available network resources [8]. Over the past decade, the technologies towards the image quality increase have developed like successive waves where newly created compression technologies can both increase and limit the quality of the picture, where individual broadcast technology solutions can also deteriorate the quality of the picture.

To conclude, compression and broadcast essentially influence the quality of the video image of television channels.

Nowadays a slow transition takes place from the technology-centred *QoS*, e.g. *IntServ, DiffServ* [9] and *SLA (Service Level Agreement)* [10] to the user-centred *QoE* (*Quality of Experience)* quality assessment [11], [12]. In this case, the assessment of technology-centred [13] quality or *QoS* should be understood as performance parameters of objectively measurable network and service [14], for example, delay, jitter, transmission speed, packet loss, etc. User-centred quality assessment is used to be called subjective or user's perception quality [15], [16]. *QoE* is the user-centred quality assessment criterion [17] that is created for assessment not only of video image, but also of multimedia content (video, audio, text), as well as for the assessment of the quality of the website usage. It could also be regarded as one of the reasons why even nowadays there is no unequivocal definition of *QoE* regarding different multimedia content application, although international scientific project workshop group [18] has carried out a lot of research on *QoE* [19].

QoE includes both the subjective and the objective quality assessment methods. Subjective *QoE* is specialized quality assessment experiments where people (respondents) participate. Non-experts are employed as respondents [20], who do not possess specific knowledge of video quality assessment procedure, including specific operation principles of video coding and transmission technologies. Only in specific cases, expert assessments are used, for example, for preparing experiments for video quality assessment, in the process of video code development, for tuning of specific video coding parameters. However, in all other cases the results of the non-experts are used to assess video quality. The subjective quality assessment is usually carried out in controlled laboratory conditions according to standardized procedures [21] and [22], which demand considerable investment. As the alternative for the controlled laboratory conditions there are true-to-life living conditions (at home on sofa, at work by the desk, in the cafe, in the park, on the bus, etc.) [23] and [24], where the exact recording of technical parameters is not possible and the experimental data of quality assessment most often are acquired using the Internet (specially prepared experiments at the website). Such a method is called *crowdsourcing* [25], [26], [27]. *MOS* (*Mean Opinion Score)* [28] is most often used as the subjective quality indicator or characterising parameter which could be considered *QoE de-facto* standard; however, it is not the only parameter of statistics [29]. *MOS* reflects the average quality assessment of specific coding or transmission system users, e.g. the average assessment of the picture quality [30]. Although the subjective *QoE* assessment is regarded as the only safe method for acquiring reliable results, it is a time–consuming, complicated and expensive process which is necessary to be replaced by computerized quality assessment, called objective *QoE* assessment or forecasting that is calculated from objective parameter measurements. Therefore, intensive research is still taking place on the development and improvement of subjective and objective quality assessment methods. In the given Doctoral Thesis *MOS* is used as the subjective *QoE* indicator.

In practice, there are various objective quality evaluation and forecasting models, but most of them have nothing to do with the subjective quality or *MOS* forecasting. However, some of them can be used as a network planning tool, for instance, Internet telephony network quality planning according to [31] or video telephony transmission arrangements in accordance with [32]. These ITU-T models should be used for the quality assessment of telecommunication operator and video telephony service quality. However, to date such quality assessment is not put into practice, because to calculate the possible video quality of service by using more than ten different parameters is very difficult and time consuming, one could even say that in practice it is almost impossible to accomplish. Models proposed by ITU-T have to be simplified; parameters and a number of measurements have to be significantly reduced. For example, [32] video parameter limit values and factor related limit values are defined vaguely and are generalized, because all the limit values of video parameters have not been researched and validated yet. As a result, it can be concluded that the model [32] is not fully usable in the multimedia quality evaluation, strict limit values for parameters that are associated with *MOS* have to be defined. It has also motivated to carry out research aimed at finding parameter limit values for the compression and transmission impact. The author is convinced that two parameters would be sufficient for the television and video transmission quality assessment: packet loss and video scene content (synthetic and natural video, *SA* and *TA* values), where each is necessary to define clearly understandable coefficients with experimentally firmly established thresholds.

The most important and the most complicated issue is the linking quality assessment results of the service users with the objectively measurable parameters of coding (coding blocks, blurred image blocks) and transmission (packet delay, jitter, and loss), which could be used when developing and improving the objective *QoE* methods (algorithms or metrics). This is the main task of the performed research in the given Doctoral Thesis.

The main task of the objective quality metrics is automated and highly accurate forecast how a human with visual perception will assess the image quality, which is called a model of human visual perception or *HVS* (*Human Visual System*). In order to acquire data about human visual perception, a series of scrupulously well–considered subjective quality experiments should be carried out as a result of which it is possible to calculate *MOS*. In order to model human visual perception with the help of objective quality metrics, the objectively measurable quality parameters should ensure as close correlation with *MOS* as possible. Any technical quantity that refers to *QoS* could be regarded as objectively measurable quality parameters, e.g. the speed, delay, or jitter of data transmission and packet loss.

Nevertheless, the objective quality metrics have many advantages; a few existing drawbacks may turn out to be very essential because objective quality assessment metrics (*QoE* metrics) consider only a limited number of human visual perception peculiarities; thus, inaccurate results could be obtained. The bigger risk to obtain inaccurate and incomprehensible results happens when the performance of metrics is tested under unusual conditions [33].

Therefore, *QoE* metrics should be developed for clearly defined application, e.g., for assessment of motionless image quality, then its accuracy should be checked and assessed with the help of various quality assessment experiments. It can be concluded that the research on linking the subjective human perception peculiarities with objective parameters will continue until the moment when almost identical results are obtained by *QoE* metrics calculations and realistic quality assessment experiments.

Video quality measurement and value can ascertainable on several levels, according to the ISO / *OSI* model, this can be done on – bit level (*SNR* and *BER*), packet-level (packet loss, jitter and delay) or element of video image level (*MPEG I*, *P* and *B*). In these cases, a specific video transmission or coding system period is assessed, for example, from HeadEnd station (*HE*) to the decoder (*STB*) or from the decoder (*STB*) to the screen (*TV*). In order to measure the last period of the video signal from the screen to the viewer, it is necessary to use specialized video quality assessment experiments. Experiments are made in order to estimate video transmission quality which depends on the transmission in various channels and technical parameters of the signal.

The main research of the Thesis is devoted to video broadcast using the *IP* protocol, and operation principles of *DVB* broadcasting are only dealt with the extent to which they have common features with video streaming using the IP protocol.

The Goal and Objectives of the Thesis

The aim of the Doctoral Thesis is to define the dependence of the video quality services on the technical parameters of the transmission channel and the signal.

To achieve the aim, the following **objectives** have been set:

1. To perform the literature survey on human visual perception main elements:
   1.1. To find out the most important parameters of human visual perception which influence the assessment of video quality.
2. To perform the literature survey on operational principles of video signal transmission and compression technologies, their main elements and their parameters which can influence the assessment of video quality:
   2.1. To find out the most important parameters of video signal transmission and compression which influence the assessment of video quality;
   2.2. To create the testbed for carrying out the experiments of video quality assessment;
   2.3. To develop the methodology for carrying out the experiments of video quality assessment with the help of which technical parameters of video coding and transmission could be modelled;
   2.4. To develop the guidelines for the improvement of video quality methods in order to obtain two-dimensional quality assessment (*MOS*);
   2.5. To carry out tests on respondents' visual acuity (*Snellen*), colour vision (*Ishihara*), and contrast sensitivity (*Pelli-Robson*).
3. To perform the literature survey on video quality assessment methods:
   3.1. To find out which of the methods is the easiest to perceive and comprehend;
   3.2. To compare one-screen and two-screen methods;
   3.3. To compare the scales of 5, 9, 11, and 100 points;
   3.4. To develop recommendations for choice of video quality assessment methods and to select the most important criteria for the choice of method;
   3.5. To develop guidelines for the video quality assessment method with the help of which the assessment of two-dimensional video quality and acceptance can be obtained.
4. To carry out the selection and the assessment of video scenes which will be used in the research experiments:
   4.1. To carry out the selection of the video scenes according to their content (nature, animals, sports, news, etc.);
   4.2. To carry out the selection of the video scenes according to space (*SA*) and space-time (*TA*) measurements;
   4.3. To carry out the comparison of the original and damaged (damage caused by coding or transmission) video scenes according to *MSE* and *PSNR* measurements.

5. To carry out the comparison experiments of various video signals and interface video quality, and the processing of statistical data:
   5.1. To carry out the assessment of video signal and interface quality: component digital (*YCbCr* 1080i), component analogue (*YPbPr* 720p), composite (*YIQ* 576p);
   5.2. To compare the results and examine whether there are statistically significant differences in video quality assessment while using *PC* (one-screen method) and *SCJAC* (two-screen method);
   5.3. To carry out the assessment of synthetic video scene quality for the video signals 1080i and 576p, using the video interface of the component *YCbCr* and the composite *YIQ*;
   5.4. To make the calculations of average assessment (MOS), standard deviation, validity intervals, and correlation coefficient.
6. To carry out the assessment experiments of packet loss influence on video quality, and the processing of statistical data:
   6.1. To accomplish packet rejection for variable speed video (*VBR*) stream;
      6.1.1. To reject packets 0.25 %, 1 %, 2 % and 5 % from the total number of video scenes;
      6.1.2. For packet rejection to use Poisons, Gaussian and uniform probability distribution.
   6.2. To perform video quality assessment:
      6.2.1. To carry out video quality assessment using *ACR – HR* (for 0.25 % and 1 % packet loss);
      6.2.2. To carry out video quality assessment according to *MSE* and *PSNR* (for 0.25 %, 1 %, 2 %, and 5 % packet loss).
   6.3. To perform the measurements of the *GOP* shots (*I* and *P* shots) of damaged video scenes.
   6.4. To make the calculations of average assessment (*MOS*), standard deviation, validity intervals, and correlation coefficient.

Research Methodology

The fulfilment of the set objectives can be divided into three stages:

1. During the theoretical studies literature analysis has been performed on video streaming technologies with the aim to find out the most essential parameters which can influence the

video quality assessment; the comparison of quality assessment methods has been carried out and the choice of the methods has been made as well as principles for video scene choice have been defined.

2. Experimental part includes the creation of the testbed and many experiments have been carried out with the aim to define how various technical parameters of video coding and transmission influence quality assessment.
3. In the calculations and analysis part, data processing has been performed with the aim to find out relationship between the quality assessment experiments and the objectively measurable technical parameters of video signals and broadcasting channels.

Scientific Results and Main Conclusions of the Thesis

Scientific Novelty

1. The methodology of video quality assessment has been developed and approbated:
   1.1. The selection of the respondents according to age, gender, education, as well as specific tests of vision acuity (*Snellen)*, colour vision (*Ishihara)*, and contrast sensitivity (*Pelli-Robson*);
   1.2. Statistical processing of respondents' data (calculations of average assessment, standard deviations, validity interval, correlation, etc.);
   1.3. The selection guidelines used for video scene selection in experiments as well as their assessment according to space (*SA*) and space-time (*TA*) measurements;
   1.4. Comparison and quality and quantity assessment of the original and damaged video scenes according to *MSE*, *PSNR*, *SSIM*, *VQM*, as well as other metrics.
2. The recommendations have been developed for the choice of video quality assessment methods.
3. The guidelines have been developed for video quality assessment method with the help of which two-dimensional video quality and acceptance assessment (*2D-MOS*) can be obtained.
4. The video streaming solution has been developed with the help of which emulations of video coding can be carried out and artefacts are seen as a result of broadcasting — block effect (*blocking*) and blurred, indefinite, picture fragment with reduced resolution (*blurring*), etc.:
   4.1. The emulations of video coding and video interface (*YCbCr*, *YPbPr*, *YIQ* and *SDI*, *ASI*),

4.2. The emulations of video broadcast (delays, jitter, and packet loss), using three different rules of probability distribution (Poisons, Gaussian, and uniform).

The Main Conclusions of the Thesis

1. It is experimentally concluded that video quality assessment (*MOS*) is essentially dependant on video scene content ($r = 0.75$). In order to compare mutually how essential the differences of those various video scene content assessment types are it is advisable to use two-dimensional video scene assessment (*2D – MOS*), where for each video scene both the quality and the acceptance assessments are available.
2. The greater the temporal activity (*TA*) value of the video scene and higher technical quality of the video scene, the worse the video scene assessment (*MOS*), which confirms that this correlation is advisable to use as one of the parameters in the forecasting of synthetic video scene quality. Video quality assessment is inversely proportional to logarithm of temporal activity value ($r = –0.85$).
3. The choice of respondents can be considered the primary and most significant selection criterion of video quality assessment method, which the respondents have acknowledged as the most simple – easy to perceive and comprehend. Giving an optional choice to choose a one- or two-screen method, more than 70 % of respondents would choose a two-screen method although in the Thesis it is found that it is not useful to use a two-screen method, because applying a one-screen method statistically equivalent results ($r = 0.78$) can be obtained. For a one-screen method 5, 9, 11, and 100 points scales are used, which are mutually linearly transformable, but for a two-screen method 7 points scale assessments are not transformable to 5, 9, 11, or 100 points scales. It is advisable to choose a one-screen method because it in a simplified way ensures mutual comparison of the results obtained by different methods and scales as well as it is more easy to implement.
4. It is important to notice that for the generation of self-similarity data flow the variable speed video (*VBR*) is used in order to ensure maximum realistic *IP* broadcast network conditions because only in such conditions it is possible to improve and develop objective video quality assessment metrics as well as acquire reliable results using video quality assessment methods. However, for statistical emulation of packet loss, in self-similarity data flow it is advisable to use several rules, the Poisson, the Gaussian, and the uniform probability distribution rules, because in video quality assessment (*MOS*) among these rules of

probability distribution there are not any essential statistical differences (r = 0.5 – 0.7) observed.

5. Packet loss is not desirable for real time video streaming solutions, or it should be to the maximum extent small, not higher than 0.25 %. The amount of packet loss that is higher than 0.25 %, incl., 1, 2, and 5 % is high and is not acceptable in video streaming. Experimentally obtained packet loss of 0.25 % corresponds to MOS = 2.4, that is very close to MOS = 2.5, which is considered to be the threshold of quality acceptance.

The Practical Significance of the Thesis

1. The results of the research have been used to implement 3 international scientific research projects.
   1.1. Part of the research results are used for Ltd. Lattelecom service "Interactive Television" *i – TV* and *IPTV* user's final equipment *STB* video quality assessment;
   1.2. The research results can serve as recommendations to Public Service Regulation Commission (Regulator) when developing and adapting the methodology and television broadcasting dissemination service quality measurement system.
2. Recommendations developed during the research are meant for testing and rooting of the existing video broadcast network elements as well as for planning and research of new solutions.
3. The **testbed** has been created as well as the **methodology** has been developed and approbated, which can be used:
   3.1. For video subjective and objective quality assessment, incl., for **testing** experimental developments and commercial products;
   3.2. For the **research** of video coding and video quality of the broadcast system elements, and for the **research** of interference persistence and performance.
4. The Doctoral Thesis conforms with the **Service Quality Calculation Methods** of the Department of **Electronics and Telecommunications** of the Latvian Council of Science and sub-sector of **Telecommunications Networks**

The Thesis Statements to Be Defended

1. The synthetic video quality MOS score is inversely proportional to the logarithm of temporal activity (TA) values (r = –0.85).

2. In the video quality evaluation a single screen and dual screen methods give statistically equivalent results ($r = 0.78$), but, when given the choice, 70 % of evaluators preferred the two-screen method.
3. When packet loss is emulated with the Poisson, Gaussian or uniform probability distribution, the video quality evaluations *MOS* differences are not statistically significant.
4. The video streaming packet losses cannot exceed 0.25 %, as it corresponds to video quality assessment *MOS* = 2.4, which is very close to *MOS* = 2.5, which is commonly referred to as the quality acceptance threshold.

The Approbation of the Doctoral Thesis

The main results of the Doctoral Thesis have been discussed at eight international scientific conferences and several conferences of business and technologies as well as reflected in seven publications and scientific editions. Within the framework of the Doctoral Thesis, the text book has been developed.

<u>International scientific and technologies conferences:</u>

1. ELECTRONICS’2012, Lithuania, Palanga, June 18–20.
2. BMSB’2012 – IEEE International Symposium on Broadband Multimedia Systems and Broadcasting 2012, Korea, Seula, June 27–29.
3. QoMEX’2012 – International Workshop on Quality of Multimedia Experience 2012, Australia, Melbourne, July 5–7.
4. ICIP’2012 – IEEE International Conference on Image Processing 2012, USA, Florida, Orlando, September 30 – October 3.
5. ICMMP’2013 – International Conference on Multimedia Processing 2013, Tunisia, Tunisia, June 22–24.
6. ICIIP’2013 – IEEE International Conference on Image Information Processing 2013, India, Shimla, December 9–11.
7. BMSB’2014 – IEEE International Symposium on Broadband Multimedia Systems and Broadcasting 2014, Beijing, China, June 25–27.
8. APMediaCast’2015 – IEEE Asia Pacific Conference on Multimedia and Broadcasting, 2015, Bali, Indonesia, April 23–25.

9. Separate parts of the research have been presented in several business and technologies conferences, for example, Broadband Forum 2011, IPTV World Forum 2012, IBC–International Broadcasting Convention 2013, etc.

Publications and Scientific Editions

1. Pauliks R. Objektīvās un subjektīvās kvalitātes vērtēšanas metodes Interneta televīzijas pakalpojuma nodrošināšanai, Saturs un zināšanas – ērti un interaktīvi lietojumi. *Zinātniski praktiskās konferences materiālu krājums*. Ventspils: 2008. 145–158 lpp.
2. Pauliks R., Slaidins I. Impact of Video Content and Technical Specifications on Subjective Quality Assessment, *Electronics and Electrical Engineering*. Kaunas: Technologija, 2012, No. 6(122), pp. 91–96. Citation indexed: Scopus, IEEE Xplore.
3. Pauliks R., Belahs K., Tretjaks K. A survey on some measurement methods for subjective video quality assessment. *Journal of Multimedia Processing and Technologies*, Vol. 3, No. 2, 2013, p. 113–123. Citation indexed: Scopus, IEEE Xplore, Google Scholar.
4. Pauliks R., Belahs K., Tretjaks K. Subjective Video Quality Assessment Methods, *Space Research Review*, Vol. 2, 2013, pp. 25–33.
5. Pauliks R., Slaidins I., Quality Evaluation of Synthetic Video In Simultaneous Double Stimulus Environment, *Proceedings of the 2013 IEEE International Conference on Image Information Processing (ICIIP–2013)*, India 2013, pp. 170–175. Citation indexed: Scopus, IEEE Xplore.
6. Schneps–Schneppe M. A., Pauliks R. On the Role of Subjective Assessments in IPTV Quality Configuration, *Automatic Control and Computer Sciences*, 2014, Vol. 48, No. 1, pp. 25–36. © Allerton Press, Inc., 2014. Citation indexed: Scopus, Springer.
7. Pauliks R., Slaidins I.,, Krauze A., Tretjaks K. Assessment of IP packet loss influence on perceptual quality of streaming video, *Proceedings of the 2015 Asia Pacific Conference on Multimedia and Broadcasting (APMediaCast'2015)*, Indonesia 2015. Citation indexed: Scopus, IEEE Xplore.

During the development of the Doctoral Thesis, the textbook has been written: Pauliks R., Krauze A. Broadcast and Streaming Video Quality Measurements and Assessments. A practical Engineering Guide. GlobeEdit. OmniScriptum GmbH&Co.KG, 2015. 560 p.

The results of the Doctoral Thesis have been used to implement several projects:

1. LR IZM Market – oriented research project "Research and Applications of Internet Television (IPTV) System Quality Measurement Algorithms for the Development of Programming Tools", No. TOP08–12, (2009, 2010, 2011).
2. LATLIT project "Use of multimedia and interactive television to improve effectiveness of education and training (Interactive TV)", No. LLIV–343, (2012, 2013, 2014).
3. Leonardo da Vinci project "Education Course of Digital TV Technologies for Vocational Educational Schools (DigTV)", No. 2013–1–LV1–LEO05–05127, (2013, 2014, 2015).

The Volume and the Structure of the Thesis

The volume of the Doctoral Thesis is 190 pages. The Thesis consists of the introduction, four chapters, conclusions, the list of references and appendices.

The introduction describes the main aspects and problems of video quality assessment as well as substantiates the topicality of the research and defines the research trend.

The first chapter is devoted to video compression and broadcasting technology review. The first sub-chapter is devoted to the description of the human visual perception – human visual acuity and resolution, light and colour vision as well optical illusions. The second sub-chapter is devoted to a video compression technology report, the most popular video compression technology and *MPEG* (Moving Pictures Expert Group), including review of video signal acquiring, formatting, coding, video stream packaging and the preparation for transportation is also listed and described here. The third sub-chapter includes a short video and television technology development report. The fourth sub-chapter includes characteristics of the most popular television broadcasting technologies as well it includes a thorough review of video broadcasting technologies using *IP* protocol. Video signal broadcasting technologies and digital television technologies that ensure transportation from the video signal source to the spectator, including video signal transportation – commuting and routing, video transmission interfaces, decoding, different methods of video signal transformation and displaying on the screen are described here. The fifth sub-chapter is devoted to the transmission of a video data flow model. The sixth sub-chapter describes specific transmission quality measurements for different *TV* technologies. The chapter ends with a summary and conclusions.

The second chapter includes a report of video quality assessment methods. The first sub-chapter contains information about video quality assessment methods, scales, their signs and a variety of applications, also describes and compares the various interrelated video quality assessment methods, provides advice and describes the developed solutions for new methods. The second sub-chapter summarises the objective video quality assessment methods and the variety of ways to apply them, including *MSE*, *PSNR* and *SSIM*. The third sub-chapter summarises the objective video quality assessment using *SA* and *TA*. The fourth sub-chapter summarises the video quality forecasting using objective video quality parameters. The chapter ends with a summary and conclusions.

The third chapter is devoted to the video quality assessment methodology. The first three sub-chapters contain information about the pilot studies used for creating conditions for the test environment, including information about the technical characteristics of compression and transmission equipment, monitor calibration as well about the testing structure and the micro-measurements. The fourth sub-chapter describes the selection principles of respondents for the experiment, including the selection of the respondents by gender, age, occupation, also the usage of tests for testing visual acuity, colour vision, contrast sensitivity. The fifth sub-chapter describes the requirements for the preparation and selection of the video material, including video scene quality assessment of the *SA* and *TA*, as well as video scene quality assessment according to the *MSE* and *PSNR*. The sixth sub-chapter describes the video quality assessment – statistical data processing. The chapter ends with a summary and conclusions.

The fourth chapter is devoted to the description of video quality assessment experiments, which identified the relationship of subjective and objective parameters of the video image quality. The first sub-chapter includes a description of the variety of video quality assessment scales as well as the screen evaluation experiment. The second sub-chapter includes a description of video compression and interface impact evaluation experiment. The third sub-chapter includes a description regarding the effects of packet loss in a quality evaluation experiment. The chapter ends with a summary and conclusions.

The conclusions summarise and substantiate the main results and conclusions of the research as well as indicate the further directions of the research.

Appendices include technical information corresponding to each chapter description and technical specifications of the testbed equipment and examination premises.

While developing the Doctoral Thesis, all *SMART* (*Specific, Measurable, Appropriate, Result-oriented, and Trackable*) basic principles have been taken into account: as a result of literature survey, the topicality of the research has been identified and the field of the research specified; within the research experiments have been carried out, measurement data analysed, loopholes identified, measurements repeated specifying and supplementing experiments resulting in the development of methodology with the help of which it is possible to repeat the whole cycle of the research.

AN OUTLINE OF SEPARATE CHAPTERS OF THE THESIS

1. Technologies of Video Compression and Transmission

The first chapter is devoted to video compression and broadcasting technology review. The first sub-chapter is devoted to the description of the human visual perception – human visual acuity and resolution, light and colour vision as well optical illusions. The second sub-chapter is devoted to a report of video compression technology, the most popular video compression technology and *MPEG* (Moving Pictures Expert Group) including review of video signal acquiring, formatting, coding, video stream packaging and the preparation for transportation is also listed and described here. The third sub-chapter includes a short video and television technology development report. The fourth sub-chapter includes characteristics of the most popular television broadcasting technologies as well it includes a thorough review of video broadcasting technologies using *IP* protocol. Video signal broadcasting technologies and digital television technologies that ensure transportation from the video signal source to the spectator, including video signal transportation – commuting and routing, video transmission interfaces, decoding, different methods of video signal transformation and displaying on the screen are described here. The fifth sub-chapter is devoted to the transmission of a video data flow model. The sixth sub-chapter describes specific transmission quality measurements for different *TV* technologies. The chapter ends with a summary and conclusions.

Video Compression

The aim of the video signal compression is to decrease to the minimum the amount of transmittable data, at the same time influencing the video image quality as little as possible. The main task of the newly-developed compression algorithms is to ensure by constant video quality the possibly effective compression, e.g., *H264 AVC P10* compared to *MPEG – 2 P2*, ensures up to 50 % more effective compression [34]. Similar correlation is found when *H265 HEVC* is compared to *H264 AVC P10* [35]. The higher the stage compression is, the more complicated decoding algorithms are and the higher the computational capacity is necessary [36]. Decrease in the amount of data can be achieved in two ways: lossless and lossy. Redundant data lossless compression or coding and renovation of the original signal at the receiving part do not influence video image quality.

The compression effectiveness of video coder can be increased taking into account human visual perception peculiarities and video image building principles. Therefore, video quality assessment or *MOS* depends on video scene, coder chosen and coding parameters. The higher the degree of compression is, the greater the probability is to notice video impairment or artefacts, especially noticeable in those video scene areas where the movement is. Similar correlation could be noticed in the packet loss instance; however, the form of impairments or the form of visually noticeable artefacts is significantly different.

Within the framework of capacious experimental research, the results of which are shown in [37] and [38] the evaluation of both the video interfaces and the video coder quality have been carried out (see Chapter 4).

With the help of *PSNR* (*Peak Signal to Noise Ratio*), *SNR* (*Signal to Noise Ratio*) correlation can be calculated between two pictures and expressed in decibels. This ratio very often is used as quality measurement between the original and compressed images.

To obtain *PSNR*, first of all *MSE* is calculated:

$$MSE = \frac{\sum_{M,N}[I_1(m,n) - I_2(m,n)]^2}{M \times N}, \tag{1.1}$$

where *M* and *N* are the rows and columns of bits of the picture.

MSE (*Mean Square Error*) and *PSNR* are two error calculation formulas to compare the degree of compression of the picture. Using *MSE* the mean square error is calculated between the original and the compressed picture, but with the help of *PSNR* the volume of maximum error is calculated. The lower the *MSE* value, the smaller the error.

Then *PSNR* is calculated:

$$PSNR = 10\log_{10}\left(\frac{R^2}{MSE}\right), \tag{1.2}$$

where *R* is a maximum value of bits, e.g. for 8 bits picture *R* = 255.

To assess the colour changes of the original and coded video scene picture it is advisable to use a rather simple method (*Delta*). This method most often is used for testing video coding and the quality of video filters.

$$Delta(X,Y) = \frac{\sum_{i=1,j=1}^{m,n}\left(X_{i,j} - Y_{i,j}\right)}{mn}, \tag{1.3}$$

where *Xi,j* is original and *Yi,j* coded picture pixels, *m* and *n* are rows and columns of picture bits.

To calculate *PSNR* for colour images, various methods can be used, basically performing conversions from one colour space to the other, e.g. *RGB* conversion to *YCbCr* colour space because human vision is more sensitive to luminance *Y*, comparatively less towards colours *CbCr* (*chrominance*). For the above-mentioned reasons *PSNR* very often is calculated only for luminance *Y* component.

In research [39] on the influence of the packet loss on *MOS*, rather poor correlation with *PSNR* is stated (r = 0.28 – 0.60) (see Chapter 4).

Video Transmission

In *IPTV* or Internet *TV* and *OTT* (*Over the Top Technology*) network solutions, three elements can be distinguished: head station *HE* (*HeadEnd*), decoder *STB* (*Set-top-Box*) and *TV* display. *HE* will always be physically separated equipment (software and hardware), but *STB* and *TV* can be both two separate pieces of equipment and the united one piece of equipment, e.g., tablet *PC* or mobile device with appropriate software for video decoding and display; thus, the decoder and display can be united into one device.

Interfaces of video equipment can be divided into two relative groups: the so-called consumer grade and professional video interfaces. Consumer grade interfaces can be divided into several subgroups: analogue, digital interfaces; interfaces designated for one or several video signal simultaneous transmission; common interface used for audio and video signal transmission, or separate interface used for each signal; as well as according to application – for computer monitors or *TV* sets. However, in modern products this strict division according to application begins to disappear, and technical specifications of interfaces become rather equivalent. Table 1.1 summarises the most popular interfaces of consumer grade and professional video, but several solutions patented by manufacturers that have not gained popularity are not mentioned. Until nowadays almost all analogue and digital video interfaces of consumer grade are still integrated into majority of modern products.

This explains why during the research experiments the quality assessment tests of the most popular video interfaces have been carried out.

Table 1.1

Video Interfaces (Consumer Grade and Professional)

Set-top-box and TV/monitors		HeadEnd un TV studio
Analogue interfaces	Digital interfaces	Digital interfaces
DVI (RGB) analogue, digital or combined (DVI – A, DVI – D or DVI – I)		**SDI** (uncompressed raw data video YCbCr, RGB, sRGB, xvYCC)
Composite CVBS/CCVS (YIQ)	**HDMI** (YCbCr, RGB, sRGB, xvYCC)	**ASI** (MPEG – TS packets
S–Video (Y/C)	**DP** (RGB, xvYCC, sRGB, scRGB, Adobe RGB)	
SCART (RGB, Y/C, YIQ)		
VGA (RGB)		
Component (YPbPr)		

Within the framework of ample experimental research the results of which are shown in [37] and [38], quality assessment tests of both video interfaces and video coders are carried out (see Chapter 4).

In the transmission systems, there are packet loss and jitter which to a certain extent can be compensated by a buffer memory. However, the break of communication channels as well as overflow or underflow of the buffer memory creates the packet loss. The jitter and the delay which cannot be compensated with the buffer memory will create the packet loss. The mentioned facts have served as the basis for performing the research experiments testing exactly the influence of the packet loss but not the packet delay or jitter.

STB buffer size should be large enough to be able to compensate the jitter as well as the overflow or underflow of the buffer memory packets. In the cases of buffer memory overflow or underflow [40], there may be packet loss; for the visual purpose refer to Fig. 1.1.

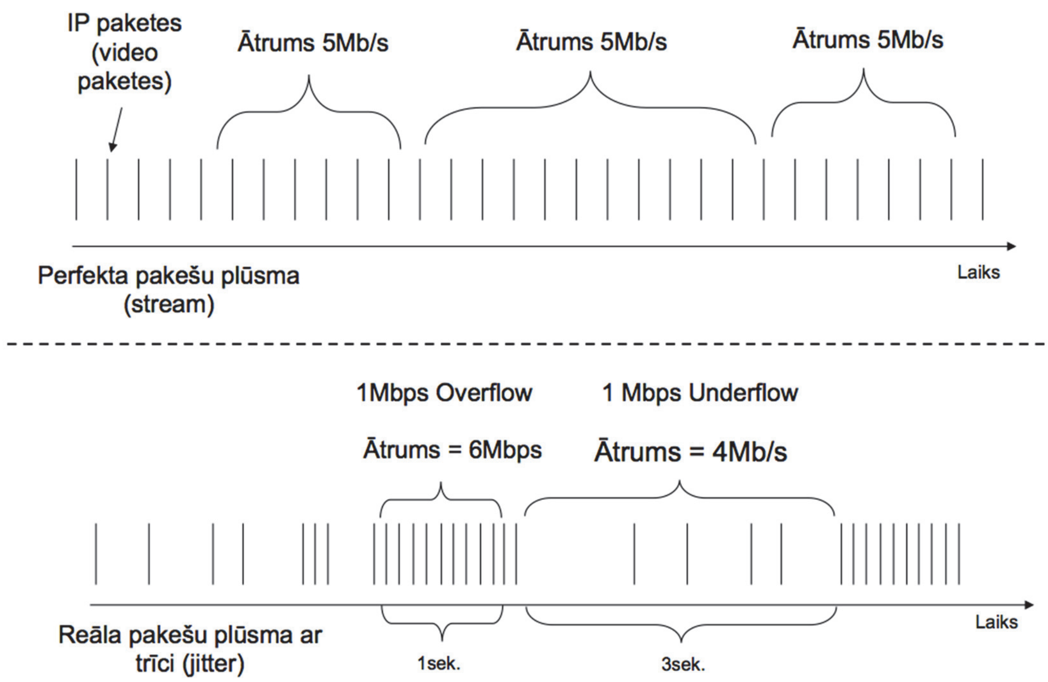


Fig. 1.1. Packet broadcasting process with delays and jitter.

Fig. 1.1 demonstrates a video packet flow with constant delay and jitter as well as a real packet flow with variable jitter.

Classical *TV* broadcasting technologies, such as *DVB – T*, *DVB – C* and *DVB – S*, use the principles of channel commutation network technologies, but *TV* streaming technologies, e.g., *IPTV* or *OTT* use packet commutation networks. In order to ensure guaranteed broadcasting quality, packet commutation technologies have to perform prioritization and reservation of the network resources.

While looking for correlation between packet loss amount and its distribution influence on video quality assessment in the research experiments *VBR* [41], [42] coded video were used. The amount of the packet loss was chosen as percentage from the total data flow within limits of 0.25 % to 5 %. Individually for each packet loss different probability distributions were adjusted according to which the packets were abandoned, e.g., regular (Gauss), Poisson uniform distribution, as well as determined (periodical) packet loss [43], [44]. In order to emulate self-resemblance data flow in *IP* network the *VBR* video flow was used [39]. The mentioned probability distribution laws are used for packet loss emulations, but the developed laboratory and methodology allow using them not only for packet loss but also for packet delay and jitter parameter emulations.

It has been found that in majority of studies the principles of packet loss emulation are not described as well as the parameters of video coding, video streaming and network are poorly defined. In the author's opinion, in all packet loss experiments parameters should be defined according to Table 1.2.

Table 1.2

The Conditions of the Packet Loss Experiment

Video coder	RAW YUV 4:2:0, H.264/MPEG – 4 AVC P10, Main@L4.1, 1280 × 720, 16:9, 25 fps
Video streaming	*VBR* video 30 Mbps (10–50 Mbps burst), MPEG – TS, 7 TS transports 1 IP packet
Packet loss emulations	The amount of packet loss (0.25 %, 1, 2 and 5 %), probability distributions used in packet abandonment (Poisson, Gaussian and uniform)
Parameters of IP broadcasting network	Jitter buffer memory 1000 ms, CAS (*Conditional Access*), repeated broadcasting, error correction, re-routing, fragmenting of the packets, packet duplicates, as well as other packet damage is not used

In the majority of studies, the amount of packet loss is defined which is usually expressed in percentage but the packet abandonment principles are not defined or they are described very poorly. In the research most frequently uniform distribution is used [45], [46], including ITU – T [32] in the recommendation. In separate works, quite simplified packet row management mechanism built in routers [47] or the network filtration tools of the operator systems (Linux, Unix) are used in order to ensure packet abandonment emulations. In some studies, Poisson distribution is used [48], [49], as well as any other occasional distribution very often is not even described. In some studies, in addition to packet abandonment distribution dependence parameters of burst and long-term packet loss are used [50], [51], [52], but in this study [53] restrictions of Poisson's distribution law usage are mentioned. However, in research [46] it is found that the amount of burst does not significantly affect the assessment of the video quality. The mentioned fact has served as the basis for using Poisson's and uniform distributions for the packet loss emulations as well as examining the influence of Gaussian or uniform distribution. Gaussian distribution depends on several parameters, which could be equalled to packet loss that may appear in the network depending on many factors, e.g., overflow of the router buffer memory, overload of network equipment, restriction of throughput capacity of connection, breakage of connection, one and the same data flow broadcast using different routes, re-routing as a result of which big delays and jitter appear.

It is found that in some studies in order to simulate visually noticeable artefacts – blurred elements of the image (blurring) or image blocking – which appear as a result of *IP* packet loss, incorrect methods are usually used, arbitrarily cutting out image elements that do not have any connection with *MPEG GOP* coding principles and transmission of the coded data using IP protocol. These studies [54], [55] compare different *MPEG GOP I*, *P* and *B* shot impairment (impairment and loss of shot block, macro-block and slice) effects on video quality assessment. It is known that packet loss from *IP* network [56] can significantly influence the quality of video streaming [57], because quality degradation of the video image in the form of visually noticeable artefacts depends on IP packet encapsulation protocol stack and on *MPEG* coder *GOP I*, *P* and *B* structure of frames [58]. *DVB* (*Digital Video Broadcasting*) (*DVB – T/C/S*) *MPEG – TS* packet loss is not the same as *IP* packet loss; usually one *IP* packet transports 7 *MPEG – TS* packets. The *IP* packet loss can be considered 7 *MPEG – TS* packet burst loss. Difference is not only in the amount of the lost data, but also in the structure and placement of visually noticeable artefacts in the image.

While performing the analysis of literature, it has been found that in general there are many of studies available where the statistics of packet loss is described. A considerable number of studies refer to modelling of *TCP/IP* data flow, e.g., [56], [59]. However, there is less research on modelling of real time video flow [60], [61].

It has been found that in various studies and also in ITU recommendations, the limit values of the packet loss are not clearly defined as well as their connection with *MOS* is not determined and quality experiments are not approbated. The limit values of the amount of packet loss and the related coefficients are defined in an unclear and generalized way, e.g., in case of ITU – T [32] packet loss should be less than 10 %. However, in research [62] the amount of packet loss is defined from 1 % to 20 %. In the following studies [63], [64] it has been found that 1 % packet loss is equal to *MOS* = 2.5 but in other studies for the same *MOS* value the level of packet loss can be ten times bigger or smaller. It is possible that these differences are connected with essentially different conditions of experiment performance which in most of the cases are poorly defined. Very often conditions of the experiment performance are not clearly defined, where some of them are summarised in Table 1.2, and sometimes it is not clear what is being tested during the experiment: the influence of the packet loss, the algorithm of error correction, the capacity of decoder or router, etc. Video quality assessment laboratories and the parameters of testing premises are described inaccurately. Probably the testing conditions do not correspond to the demands mentioned in recommendations ITU – R [21] and ITU – T [65],

[66], i.e. video scenes are incorrectly prepared, the calibration of monitor is not carried out, as well as other measurements of lighting luminance and colour are not made (see Chapter 3).

The above-mentioned aspects have given motivation to look for relevance between those various probability distributions (Poisson's, Gaussian and uniform) that have been used to describe the statistics of packet abandonment. In the given research correlation has been searched and found between limit values of packet loss amount and *MOS* or quality assessment (see Chapter 4).

In the IP broadcasting network, the typical packet loss is rather low (≤1 %), which can be regarded as rare instance without memory because packet loss does not depend on previously lost packet, which is typical of the process in Poisson's distribution. The Poisson's probability distribution process is used to describe the statistics of packet loss [67], [53]:

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \tag{1.4}$$

where *P(k)* is probability that k packets will be lost at a certain interval of time and λ is dispersion and average number of lost packets during this interval of time.

If it could be assumed that packet loss is influenced by many independent processes, then describing the statistics of packet loss uniform or Gaussian distribution could be used:

$$p(k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-\lambda)^2}{2\sigma^2}}, \tag{1.5}$$

where *P(k)* is probability that k packets will be lost at a certain interval of time and λ is an average number of lost packets during this interval of time and σ^2 is the dispersion of the number of lost packets.

Discreet uniform distribution is a typical generator of sample random value, which is used to describe the statistics of packet loss [68], [53]. It means that some minimal *a* and maximal *b* number of packets are defined, which can be lost at a certain interval of time. For uniform distribution, any number k of packets can be lost at interval $k \ni [a, b]$ with equal probability. Uniform distribution:

$$P(k) = \begin{cases} \frac{1}{a-b} & a < k < b \\ 0 & k < a \\ 0 & k > b \end{cases} \quad (1.6)$$

In research [39] about the packet loss influence on *MOS*, the difference among Poisson's, Gaussian and uniform distributions is not observed and they are used as incidental distributions when abandoning packets.

2. Video Quality Assessment Methods

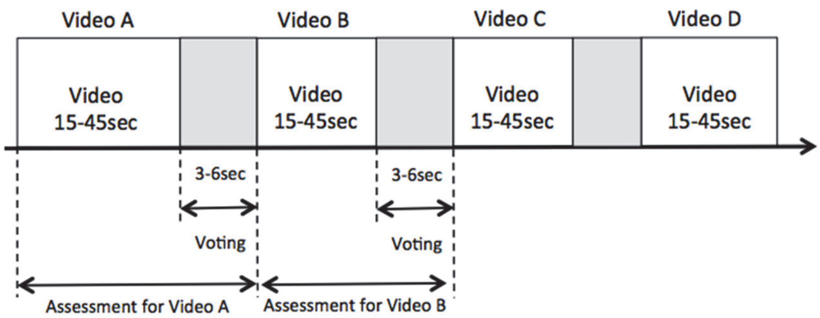
The second chapter includes a report of video quality assessment methods. The first sub-chapter contains information about video quality assessment methods, scales, their signs and a variety of applications, also describes and compares the various interrelated video quality assessment methods, provides advice and describes the developed solutions for new methods. The second sub-chapter summarises the objective video quality assessment methods, the variety of ways to apply them, including *MSE*, *PSNR* and *SSIM*. The third sub-chapter summarises the objective video quality assessment using *SA* and *TA*. The fourth sub-chapter summarises the video quality forecasting by using objective video quality parameters. The chapter ends with a summary and conclusions.

Nowadays there are a comparatively great number of video quality assessment methods, such as the ones the recognized by international organization recommendations [21], [22] and the methods suggested by the results of different experiments and scientific research, for example, [69] or *EBU* (*European Broadcasting Union*) *SAMVIQ* [70].

For video quality assessment, single screen *SS* (S*ingle Stimulus*) and double screen *DS* (*Double Stimulus*) methods are used, but in double screen simultaneous application instances the method called *SDS* (*Simultaneous Double Stimulus*) is used. In *DS* methods to assess the quality pairs of video fragments (original and defective) are used, which are showed consecutively one after another on one screen. However, in *SDS* method the comparable video fragments are shown simultaneously on two identical screens. In *DS* and *SDS* methods both the general quality assessment and the brought-in damage assessment scales are used. However, in the *SS* method the general quality assessment scales are used, where mutually non-linked video fragments are shown on the screen one after another. Most of the *DS* and *SDS* methods use *FR* (*Full Reference*) when testing because there is the signal of the original video available, but *SS* can use in the tests both the *RR* (*Reduced Reference*), and the *NR* or *ZR* (*No Reference* or *Zero Reference*) in cases when there is no signal of the original video available.

There are many methods available, but the most popular ones are *ACR*, *DCR*, *PC*, *DSCQS*, *SSCQS*, etc. Each method has its own peculiarities and application, e.g., to assess coding or broadcast damage, to assess the quality of general or brought-in damage is rather complicated and time-consuming; thus, accurate methods, methods of lower accuracy are used in the home environment, whereas high accuracy methods are applicable to the laboratory environment, etc. Further each method is discussed in more detail as well as the assessment of the methods and their comparison are given.

ACR (Absolute Category Rating) [22] is known as *SS* (*Single Stimulus*) method which is used to assess video quality according to the five-point scale (5 – *Excellent*, 4 – *Good*, 3 – *Fair*, 2 – *Poor*, 1 – *Bad*), similarly for the assessment of audio quality respectively [66], and its assessment is expressed by general quality *MOS* (*Mean Opinion Score*). In Fig. 2.1 video fragments of different quality, including the originals, are shown consecutively one after another in mixed sequence, when after each video fragment a grey screen is displayed with invitation to assess the quality according to a general quality assessment scale.



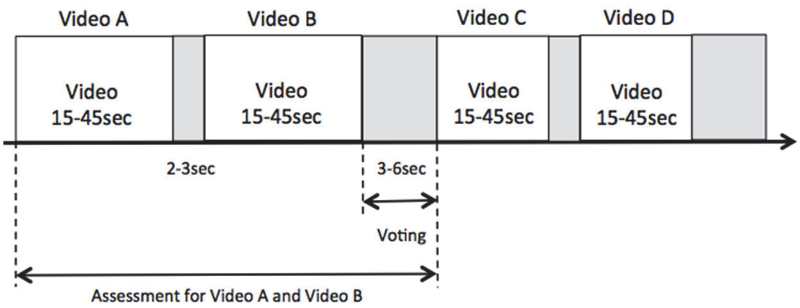
Score	Assessment
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Fig. 2.1. Process of *ACR* video fragment display.

Sometimes *MOS* with evaluation of 2.5 is considered the lowest acceptable quality limit (*acceptance level).* To prevent the influence of assessment of different video content, *ACR – HR* method is used.

ACR – HR (*Hidden Reference*) is one of the easiest comprehensible methods, which can be used by non-experts to assess the absolute video quality of both the original and the defective video fragments.

PC (*Pair Comparison*) [22] is applied to two video fragment pair qualities to assess the differences according to the seven-point comparison scale ("−3" – *Much worse*, "−2" – *Worse*, "−1" – *Slightly worse*, "0" – *About the same*,"+1" – *Slightly better*, "+2" – *Better*, "+3" – *Much* *better*). The mentioned 7-point assessment scale is called *CCR* (*Comparison Category Rating*) according to [66] or *DSCS* (*Double Stimulus Comparison Scale*) according to [21].



Score	Assessment
-3	Much worse
-2	Worse
-1	Slightly worse
0	About the same
+1	Slightly better
+2	Better
+3	Much better

Fig. 2.2. Process of *PC* video fragment display.

Fig. 2.2 displays two video fragments of different quality, after each pair of video fragments a grey screen is shown with invitation to assess the quality according to a general quality assessment scale. The results of *PC* assessment depend on the sequence of video fragment display; therefore, video quality assessment results, which are obtained with this method, are rather relative, they are not absolute, and more time is needed for testing in comparison with *ACR* or *DCR*. However, when using *PC* method, it is possible to assess very delicate video quality differences between two video fragments in both cases, with the single screen PC method, and with the double screen method, for example, *SCACJ*.

In the *PC* case, the comparable video fragments are displayed on a single screen consecutively one after another, but in *SCACJ* (*Stimulus Comparison Adjectival Categorical Judgement*) and in *SDSCE* (*Simultaneous Double Stimulus for Continuous Evaluation*) [21] cases both comparable video fragments are simultaneously displayed on two identical screens where "Video A" is shown on the left screen and "Video B" – on the right screen. This method is called *SDS* (*Simultaneous Double Stimulus*).

Primarily the method is chosen by the following:

- What has to be assessed – coding or broadcasting process influence on quality;
- How it should be assessed – general quality or quality of damage influence (*impairment*);
- Which of the methods is more comprehensible for the respondents.

Secondarily it should be taken into account:

- Which of the methods is faster;
- Which of the methods is more accurate.

In research *MOS* is recognised as de-facto general video quality and impairment assessment metrics but there are alternatives, which are reflected in Table 2.1.

Table 2.1

MOS Alternatives

Assessment metrics		Quality	Impairment	Acceptability
MOS	*Mean Opinion Score*	Yes	Yes	
JND	*Just Noticeable Difference*		Yes	
GoB un PoW	*Good or Better un Poor or Worse*			Yes
QoP	*Quality of Perception*	Yes		Yes

According to use, the assessment methods can be divided into general quality assessment (*quality*), assessment of the damage influence (*impairment*), and assessment of service level acceptance (*acceptance* or *acceptability*).

MOS measures service satisfaction level [28], but *GoP* and *PoW* are used to assess service acceptance [71]. So far *GoP* and *PoW* are used in telephoning voice broadcast quality assessment, but that is not an obstacle in future to adjust it for video broadcasting quality assessment. However, in the case of *QoP* the level of satisfaction and acceptance is assessed separately [72]. *QoP* is user-oriented [73] multimedia (video, audio, and text) quality assessment, which is obtained after complicated and time-consuming procedure – questions and answers on video quality and scene [74].

JND (*Just Noticeable Difference*) is used in experimental psychology where it is called the measure of human senses and perception. The experiment of Weber–Fechner law (Weber contrast, constant) can be considered to be the first *JND* measure applications in practice. *JND* can be used for detection of any kind of interference comparing the original with impaired signal. The higher the *JND* value, the greater the quality differences [75]. *JND* is the subjective quality measure although in practice there is also *JND* realization available in objective quality assessment models [76], as well as in commercial products of video quality measurement equipment [77].

Although in the research different alternatives of *MOS* metrics are discussed (Table 2.1), *MOS* already for more than ten years can be called de-facto measure in video quality assessment. Further in the Thesis, the methods of obtaining and calculation of *MOS* are discussed.

Table 2.2 classifies the most popular assessment methods by their application (coding or broadcasting), original (*reference video*) video signal necessity (*FR*, *RR* or *NR/ZR*), single screen or double screen, as well as by approximate accuracy and operation speed assessment.

The first different testing methods were developed and tested for motionless picture *JPEG* (*Joint Photographic Experts Group*) quality assessment [78], [79], [80]; however, during the last decade more attention was paid to the assessment of motion images or video quality and to the development and testing of appropriate methods. Video quality assessment methods are further discussed according to Table 2.2.

In research [81], *ACR*, *DSIS*, *DSCQS*, and *SAMVIQ* methods are tested. Comparing the results obtained by the methods, high correlation has been found ($r = 0.96 – 0.99$) which serves as the indicator that essential differences are not observed. In [82], two methods *ACR* and *SAMVIQ* are compared, and within the framework of the research it has been found that both methods ensure equivalent accuracy, in the case of *SAMVIQ* they use 1/3 less number of respondents. Having the equal number of respondents, *SAMVIQ* method will have a higher accuracy. In research [83], a review is made about *ACR*, *DSCQS* and *SSCQE*. However, study [84] compares two methods *ACR – HR* and *DSCQS*. It has been found that both methods ensure equivalent results but with *ACR – HR* the same test can be performed almost in four times shorter period. Pair comparison methods ensure higher accuracy [85]; therefore, they are suitable for video coding parameter tuning or quality assessment. In research [86], *SSCQE*, *DSCQS* and *DSCS* are compared, where the obtained test results of all three methods show high correlation. In the same research, it is found that the duration of human short-term memory is about 15 seconds. However, in [87] it is found that the duration of short-term memory varies from 15–30 seconds. In another research [88], memory duration is assessed to be 18–20 seconds, but according to Miller's research [89] the results of which are based on several unrelated experiments, it has been found that human short-term memory is limited to 7 ± 2 elements. It can be concluded that video fragment duration used for picture quality testing needs to be no longer than 30 seconds and also the impairment number in the picture needs to be no more than 9. In research [90], it has been found that exactly in short video fragment quality assessment respondents are less tolerant to observed interference. It means that short video fragments (15–30 seconds) should not be saturated with a large number of interference (7 ± 2).

Table 2.2

Scales of Assessment Methods

Assessment methods		Scales	
		Discreet	Continuous
ACR	*Absolute Category Rating*	5, 9, 11	
ACR – HR	*ACR Hidden Reference*	5, 9, 11	
DCR (DSIS)	*Degradation Category Rating (Double Stimulus Impairment Scale)*	5*, 9*	
DSCQS (T1 un T2)	*Double-Stimulus Continuous Quality Scale (T1 DS un T2 SDS)*		5 vai 5*, 9*
PC	*Pair Compare*	7	
SCJAC	*Stimulus Comparison Adjectival Categorical Judgement*	7	
SAMVIQ	*Subjective Assessment of Multimedia Video Quality*		100**
SSMR	*Single Stimulus with Multiple Repetitions*	5	
SSCQE	*Single Stimulus Continuous Quality Evaluation*		100**
SDSCE	*Simultaneous Double Stimulus for Continuous Evaluation*		100**

* impairment assessment scale is used

** physical equipment with slider is used

For the methods shown in Table 2.2, the suggested scales are indicated, but it should not be considered a limitation or obstacle to use other scales, e.g. instead of 5-point discreet scale 9- and 11-point discreet scales can be used. Theoretically, the *MOS* 5 point results have to be with greater standard deviation [91] when compared with the *MOS* results of 9, 11, or 100 points, although in [92] it has been found that these differences are statistically insignificant.

Table 2.3 summarises the most popular quality assessment methods and enumerates the most important conditions for the choice of the method as well as outlines the author's recommendations in favour of the choice of one or the other assessment method.

Table 2.3

Guidelines for the Choice of Assessment Methods

Method	Type of quality assessment		Duration of testing	Respondents	
	Coding	Broadcast		Non-experts	Experts
ACR	***	*****	***	*****	*****
ACR – HR	**	*****	*****	*****	*****
DCR	****	****	***	***	*****
DSCQS	****	****	***	***	*****
PC	****	**	***	****	*****
SCJAC	*****	*	****	*****	*****
SAMVIQ	*****	*	*	**	*****
	***** very well suitable **** well suitable *** fairly suitable ** poorly suitable * very badly suitable		***** very short **** short *** normal ** long * very long	***** very well suitable **** well suitable *** fairly suitable ** poorly suitable * very badly suitable	

The choice of quality assessment type (coding or broadcasting influence quality assessment) can be considered the first criterion in the choice of method. The duration of the testing of the given method can serve as the second criterion because the time spent by respondent during the test is essential. None of the methods limits the number of the video fragments used in testing; therefore, the total length of the test can take from some minutes to several hours. Respondents have no objections against 5–15-minute tests, 20–25-minute- test is already considered tiresome, but 40–60-minute test for most of the respondents is not acceptable and is considered to be very exhausting and irritating. The third criterion of the choice is the suitability of each given method for the given audience of respondents either experts or non-experts. Methods, which are regarded as accurate and complicated, for example, *SAMVIQ*, are advisable to be used by non-experts. If the respondents involved in the testing do not comprehend, misunderstand or during the test forget the principles of the method, the obtained results may be inaccurate and questionable. Methods which are suitable very well for quality assessment of coding influence not always will be suitable for quality assessment of broadcasting influence. For video with large packet loss as a result of broadcasting (≤ 1 %) in quality assessment it is not advisable to use the methods which are suitable for quality assessment of coding influence. As a result of coding when video image has got insignificant lowering of quality, it is not advisable to use the methods which are suitable for broadcasting quality influence assessment. In the first case, the quality assessment could be worse than it actually is, but in the second case in complete contrast – the assessment will be better than it actually is.

The author of the given Thesis outlines the recommendation and offers his solution for improvement of video quality assessment methods. The recommendation only applies to those methods which use the general quality of 5, 9, 11 point discreet assessment scales *ACR* and *ACR – HR*. Describing the operational principles of the method, the applied example of 5-point discreet scale will be reviewed.

After each video fragment display, the respondents assess not only the general quality of video according to a 5-point scale, but they also assess acceptance quality of the same video according to a 5-point scale on which there are not any values marked (see Table 2.4), but three evaluations that are expressed by words are given: acceptable, neutral, or unacceptable quality.

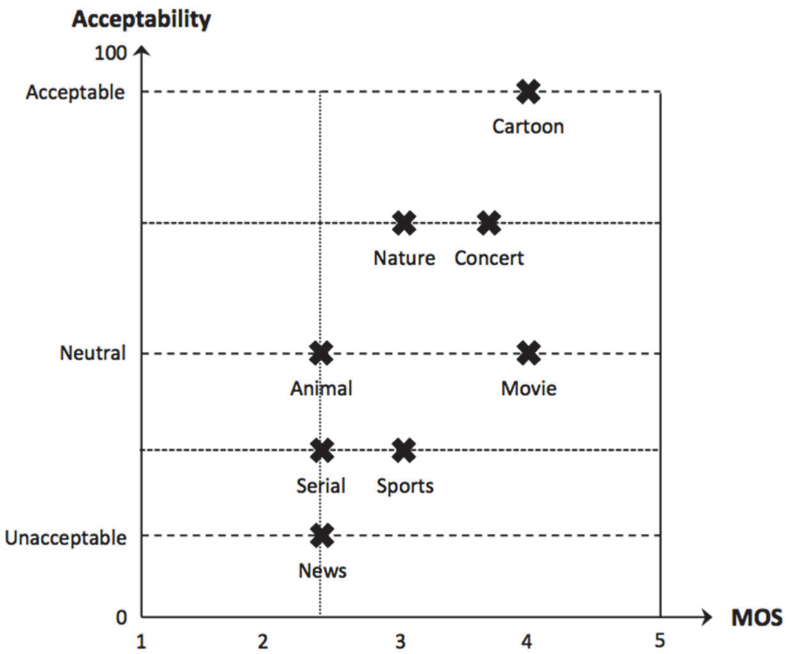
Table 2.4

The Scales of Quality and Acceptance Assessment

Value	Quality scale	Value	Acceptability scale
1	*Bad*	X	*Acceptable*
2	*Poor*	X	
3	*Fair*	X	*Neutral*
4	*Good*	X	
5	*Excellent*	X	*Unacceptable*

Denomination "acceptable quality" corresponds to 5, denomination "neutral quality" corresponds to 3, and denomination "unacceptable quality" corresponds to 1. The relative assessment 4, which is between "acceptable" and "neutral" quality, is chosen by those respondents who cannot choose between those two denominations. Similar situation refers to the relative assessment 2, which is between "neutral" and "unacceptable" quality and which is chosen by those respondents who cannot choose between those two denominations. It is essential that this double assessment very insignificantly prolongs the total testing time. It means that instead of single dimension *MOS* assessment, double two-dimension general quality and acceptance assessment can be obtained.

Figure 2.3 demonstrates the provisional graph of the results, where two-dimensional *MOS* quality and acceptance assessment are shown. If necessary a 5-point marked quality scale as well as a 5-point non-marked acceptance scale can be linearly transformed into 0–100 assessment scale [93], where 1 corresponds to 10 and 5 corresponds to 90.



Acceptability	5 grade
Acceptable	5
	4
Neutral	3
	2
Unacceptable	1

Fig. 2.3. Provisional graph of the results – two–dimensional *MOS*.

In the given research [94] an attempt has been made to analyse various multi-dimensional video quality assessment methods where some of them are complicated and time-consuming. Complicated method is not suitable for non-specialists.

The method combines marked and non-marked quality assessment scales because in some research it is found [95] that in the instance of marked scale the respondents have tendency to avoid the assessment with the final (minimum and maximum) evaluations marked on the assessment scale.

In experimental research [92], a very strong linear correlation is shown between 5, 9-point discreet and 5, 11-point continuous assessment scales. Insignificant statistical differences were found among the video quality assessments which were obtained using the mentioned assessment scales.

In its turn, in research [96] it has been found out that the majority of respondents consider a 5-point scale easier to perceive and comprehend. Table 2.7 shows that a similar result has also been obtained in research [81], where *ACR* method with a 5- and 11-point scale is recognised as the most comprehensible (see Chapter 4).

The chapter analyses different quality assessment methods, but one of the most important results is the recommendation developed by the author of the Thesis, with a corresponding solution to the improvement of video quality assessment methods. With the help of the developed method, it is possible to obtain two-dimensional *MOS* quality and acceptance assessment.

Nowadays *MOS* can be considered de-facto metric or measure when assessing the multimedia quality [28]. To obtain plausible conclusions guided by calculated *MOS*, a rather large number of respondents should be ensured, where according to [22] there have to be from 6 to 40 respondents, but according to [21] no less than 15. Data processing has to be carried out before calculations with the aim to find and delete erroneous or false data [21].

First of all, MOS_i *i*s calculated or the average assessment of all respondents $\overline{X_i}$, then standard deviation S_i and intervals of validity μ;

$$MOS_i = \frac{\sum_{j=1}^{N} X_{ij}}{N}, \quad (2.1)$$

where $MOS_i = \overline{X_{\square}}$.

$$S_i = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_{ij} - MOS_i\right)^2} \quad (2.2)$$

$$MOS_i - t_{\propto,N}\frac{s_i}{\sqrt{N}} < \mu < MOS_i + t_{\propto,N}\frac{s_i}{\sqrt{N}}, \quad (2.3)$$

where $1-\propto = 0.95$.

In separate cases correlation is calculated (Pearson correlation coefficient) r:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 (y-\bar{y})^2}} \quad (2.4)$$

or Spearman's correlation ρ:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \quad (2.5)$$

where $d_i = X_i - Y_i$ is difference between ranks.

The above-mentioned calculations are used for data processing of video quality assessment experiments, including the obtaining of *MOS* value. *MOS* is the mean arithmetical indicator (one-dimensional indicator), which cannot be considered clearly accurate objective measurement [29], because it is one of the several statistical measures such as standard deviation, validity interval, variations, asymmetries, excess coefficient, median, mode, calculation of correlations, etc.

3. The Methodology of Video Quality Assessment

The third chapter is devoted to the video quality assessment methodology. The first three sub-chapters contain information about the pilot studies used for creating conditions for the test environment, including information about the technical characteristics of compression and transmission equipment, monitor calibration as well about the testing structure and the micro-measurements. The fourth sub-chapter describes the selection principles of respondents for the experiment, including the selection of the respondents by gender, age, occupation, also the usage of tests for testing visual acuity, colour vision, contrast sensitivity. The fifth sub-chapter describes the requirements for the preparation and selection of the video material, including video scene quality assessment of the *SA* and *TA*, as well as video scene quality assessment according to the *MSE* and *PSNR*. The sixth sub-chapter describes the video quality assessment – statistical data processing. The chapter ends with a summary and conclusions.

The technical parameters of video compression and broadcasting can essentially influence the video quality assessment; therefore, to ensure reliable experiment results the appropriate laboratory and methodology for video quality assessment have been created. First, to carry out the experiments the laboratory has been created, which ensured identical testing conditions for all respondents as well as controllable and unchanged technical parameters. Second, the methodology has been developed that describes specific laboratory parameters, their calibration and measurements, recommendations for quality assessment methods to be used in the experiments as well as the procedure of the acquired data processing and analysis.

The methodology for video quality assessment can be divided into several stages:

- The laboratory of video quality testing;
- The selection of the respondents according to Ishihara, Snellen and Pelli-Robson;
- The selection of the respondents according to age, gender, education;
- Recommendations for video scene selection (synthetic, typical video scenes);
- Video scene assessment according to *SA* and *TA*;

- Video scene assessment according to *MSE* and *PSNR*;
- Recommendations for the choice of video quality assessment methods;
- Recommendations for obtaining two-dimensional video quality and acceptance assessment (*2D-MOS*);
- Statistical data processing;
- Emulations of packet loss, jitter and delay.

The laboratory has been equipped according to specified parameters in the guidelines [21] and [22]. During the testing the room of 25 m^2, approximately 5m x 5m has been used. To detain natural light and regulate the light intensity, opaque textile roller-type blinds have been used, the level of light in closed position is 0 lx. For artificial light 2700–3200 K colour temperature fluorescent lamps have been used, whose intensity of lights is adjustable to 50–800 lx, measuring perpendicularly to monitors at the height of the desktop surface using a lux-meter and a colorimeter. The colorimeter was used to measure the room light temperatures (K, *CCT* according to *X* and *Y*), the dominating wave length and colour of the light sources as well as to tune in the temperature of the room, and to isolate the dominant colours. The examiner's seat is arranged directly opposite the monitors. It is also important to take into account, probably, troublesome factors that are not mentioned in the guidelines, but which may influence the testing results, for example, the level of background noise [66], temperature, humidity, lack of ventilation, etc. All respondents were ensured with measurable, controllable and as much as possible equivalent testing conditions (see Table 3.1). Background noise of 30–35 dB, air temperature of 20–22 ^{0}C, relative humidity of 40–60 %, draught, speed of air motion of 0.1–0.2 m/s, and the level of CO_2 were regularly controlled and their adjustment was ensured by ventilation, in case of necessity it was compensated by airing. Table 3.1 demonstrates the most important micro-climate indicators of the testing room and their values.

Table 3.1

Micro-climate of the Testing Room

Indicators	Values
Level of lighting	200 lx
Temperature of lighting (CCT)	3000 K
Air temperature	20–22 °C
Humidity	40–60 %
Level of background noise	30–35 dB
Speed of air motion	0.1–0.2 m/s

For the room lighting measurements lighting meter (*illuminance meter, lux)* was used and for the lighting colour measurements colorimeter (c*hrominance meter, XYZ*, *xyY*) was used. The laboratory room equipment ensured all respondents with controllable and almost identical testing conditions.

In the research the equipment and software of several manufacturers were used, including monitors of different classes, decoders, linking interfaces, head-stations and network equipment, and as the basis there were a specially equipped testing room, lighting and colour measurement equipment, *LCD* monitors of studio quality, raw data video signal interface cards, computer manufactured by *Apple MacPro* for video server (*HeadEnd)* and customer (*Set-top-Box*) simulations, real time network equipment tools for emulation and broadcasting process simulation: switchboards, made by *Cisco*, routers, specialized emulation software as well as a whole range of software tools and additional equipment.

A monitor screen 1920 × 1080 (4 : 2 : 2), RGB 10 – bits, video interfaces: *SDI*, *CVBS*, RGB/*YPbPr*, HDMI/*YCbCr*, adjustment and measures of screen luminance (cd/m^2, 0–500), ratio of contrast (1300:1), built-in calibration using spectroscope, *AWB* (*Automatic White Balance*) adjustment, colour temperature adjustment (K, 4000–13000, *CCT*, *X*, *Y* = 0.27–0.35) and measurements, colour intensity vectroscope and measurements of the signal amplitude (*IRE*, 0 or 7.5–100), etc. are only some controllable parameters of the monitor. Table 3.2 demonstrates the most important characteristics, which are used to calibrate monitors used in experiments.

Table 3.2

Monitor Parameters and Calibration

Indicators	Value
Level of luminance (peak values)	200 cd/m^2
Temperature of lighting (*CCT*)	6500 K
White Colour Point (*CCT*)	D65
CIE xyY (*CCT*)	x = 0.3128, y = 0.3292
Display resolution and proportion	1920 × 1080, 16 : 9
Digital video signal (raw data/*RAW*)	*SDI* 1920 x 1080p
Viewing distance and angle	*3H*, 30°
Other technical characteristics according to [21], [65] and [66]	

In the research on video coding and video interface quality assessment [38], the following circuit scheme is used (Fig. 3.1). In the testing scheme *FR* (*Full Reference*) method is used.

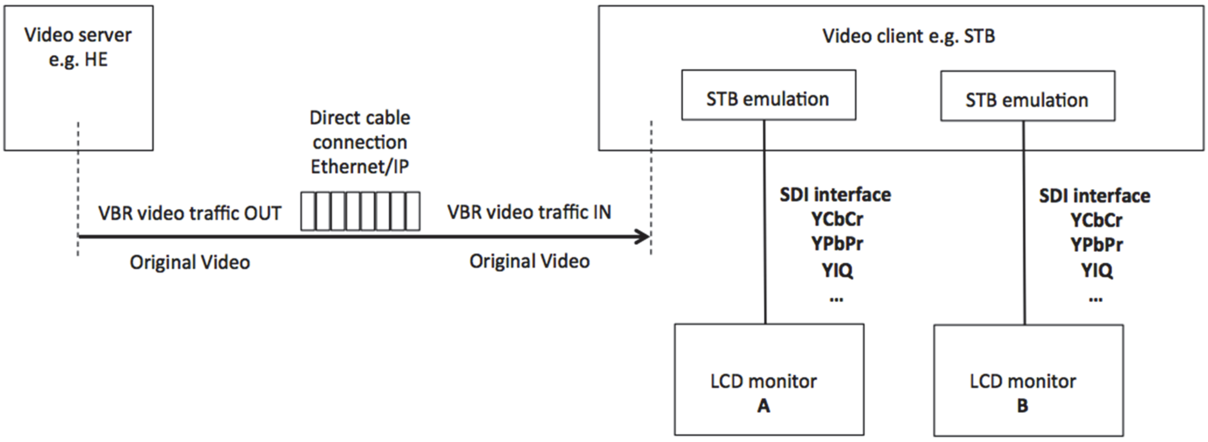


Fig. 3.1. The block diagram of video coding quality assessment.

In the research on packet loss influence on video broadcasting quality assessment [39] there is used the following circuit scheme Fig. 3.2. Likewise in the tests of coding, also in the tests of broadcasting channel quality assessment *FR* (*Full Reference*) method is used, where the original and received / impaired video pictures are compared pixel-by-pixel, thus performing time-consuming and voluminous computational operations.

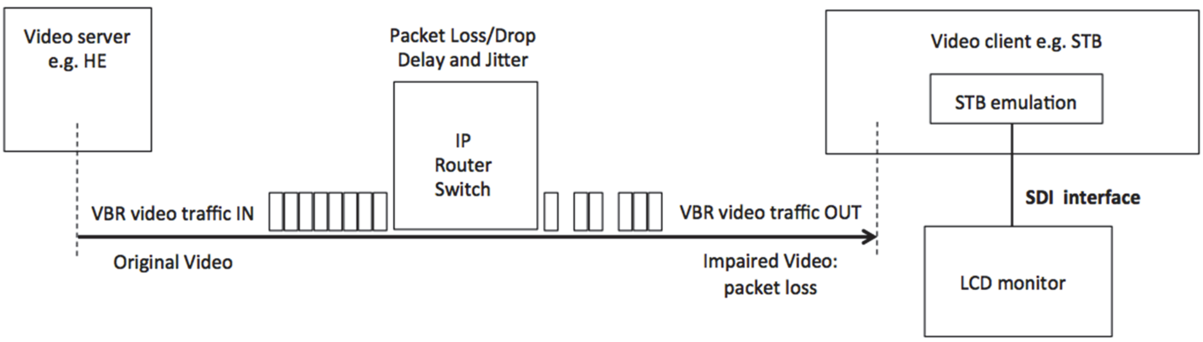


Fig. 3.2. The block diagram of video broadcasting channel quality assessment.

One of the most important factors is the calibration of the monitors applied in the experiments, using vectroscope and waveform monitor, as well as measurements and tuning of luminance level, *AWB* (*Automatic White Balance*) and colour temperatures. It is also important to perform measurements and tuning of the lighting level (*luxometer*) and colour tone (*colorimeter*) in the testing room that is used in the experiments.

To measure the luminance of the monitors in the room, luminance meter was used, and to measure the lighting of the testing room illuminance meter was used, lighting colour measurements were performed by colorimeter (*chrominance meter* or *colorimeter*). For packet loss, delay, and jitter emulations simulators were used, e.g. *ns2*, *netem*, *anue*, as well as various *Unix/Linux* packet filtration tools. For video processing a series of software tools were used from *WinOS* and *MacOS*, e.g. *Final Cut Pro X*, *Commpressor*, *Broadcaster*, *Darvin*, *Matlab/Simulink*, *AJA TV*, *h264*, *x264*, *ffmpeg*, *VLC*, etc.

The video material chosen for video quality testing essentially influences the obtained results. Therefore, when selecting the video material, recommendations from the research [97] were taken into account as well as separate suggestions from the information available in public video material databases were also analysed in the research [98], including individual research about each video material data repository. Video material used in testing was selected according to various video file characteristics [99]: parameters of video coding, *SA* and *TA* measurements of video fragments, *MSE*, *PSNR* or *SSIM* measurements of damage quantity brought in by original and impaired video files. For the needs of experiments, mainly publicly available raw data/ uncompressed video material was used, for example, [100], such type of publicly available materials is rather limited in number. In some experiments for video testing synthetically created video materials were used from commercial video data repositories [101]. It was especially complicated to find and select video material of wide spectrum content (news, serials, films, cartoons, nature, animals, sports, etc.) with desirable video file parameters (*raw* data, coding format, resolution, shot replacement speed, *I*, *P* and *B* structure of shots, etc.).

In the experimental part of the research, basically video fragments scrupulously selected and prepared by the author were used. The given fragments were used in the research about the packet loss influence on video quality assessment [39].

In the video fragment selection process, the first stage is to eliminate video fragments which can essentially influence each individual's content assessment [99]: violence, horror films, politics, religion, pornography, erotic, medical operations, etc. Insulting, controversial or irritating video content can essentially interfere with respondent's concentration to complete the task, in general average *MOS* (*Mean Opinion Score*) decreases, and dispersion increases between two substantively different video fragments; therefore, it is advisable to choose as neutral topics as possible.

It is particularly important, while comparing motion scenes or video with static pictures, to assess the amount and the type of movement information of every chosen video fragment. During the past twenty years, a simple and strong model of human visual perception movement assessment has been created, which is discussed in a whole range of studies. First physiological research and experiments about the human visual speed perception in connection with video quality assessment were started by Stocker and Simoncelli [102]. For example, if the background of video fragment has very high movement intensity, the human perception cannot obtain as accurate structural information about the object seen in the video fragment as from the static picture. However, increasing the contrast of video fragment, objects are perceived with less uncertainty [103]. The main aim of video quality assessment algorithms is to forecast human behaviour during the video quality assessment. Therefore, it is important to take into consideration also the movement information of the video fragment, which can leave essential impact on respondents' video quality assessment.

One more very important issue, which follows the video fragment selection, is preparation of video fragments. In testing video fragment *HRC* (*Hypothetical Reference Circuits*) is used, which is obtained from the selected (original) video fragments *SRC* (*Source Reference Channel*) or *Circuits* (see Fig. 3.3).

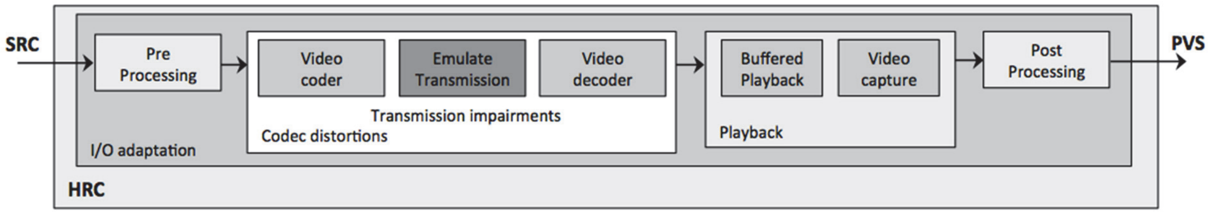


Fig. 3.3. HRC video acquisition from SRC video fragments.

HRC is obtained from *SRC* but to respondents on the screen a series of testing video fragments is shown, which is called *PVS* (*Processed Video Sequences)*. In the case of *FR* (*Full Reference)* method, the following two video scenes should be mutually compared: original video (*SRC*) and impaired video (*PVS*). It is important to note that *PVS* is the resulting video scene obtained from the *SRC* video and which includes the reflection in hypothetical *HRC* channel that influences the coding or transmission process. The playback duration, number and sequence (original and impaired scenes) of video fragments are determined by the chosen video quality testing assessment method (see Chapter 2).

If it is planned to use in testing video material, which should be filmed, photographed and processed by oneself, two essential aspects should be taken into account: the first is optical illusions and the second is optical effects and defects caused during the video shooting process.

In video quality assessment tests, the selection of the chosen video scenes is very important. Spatial *SA* (*Spatial Activity*) and temporal *TA* (*Temporal Activity*) parameters of video scene [104] [105] [106], as well the content of the video scenes [107] is also important. There are several methods available in practice for video scenes *SA* and *TA* parameter calculations, but in the given Thesis the method from the recommendation [22] is used.

SA value indicates the informative saturation of the image; the more delicate elements are in the picture, the higher the SA value.

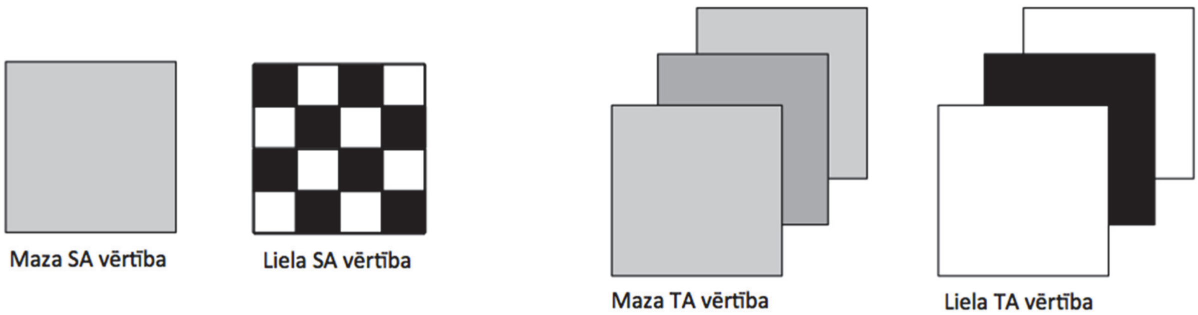


Fig. 3.4. An illustrative depiction of *SA* and *TA* values.

Separate video picture quality impairments/ artefacts can be linked with *SA* and *TA* calculated values of the video fragment and their changes over the time. For example, *SA* impairments (artefacts) can be linked with the changes of picture contrast, luminance, and colours, visible noise in the picture, blocking effect, and blurred, indistinct picture fragment with decreased resolution (*blurring*), etc. *TA* impairments/ artefacts can be linked with picture freeze, jerkiness, flickering, inadequacy of object movement compensation, fluctuation of stationary objects, etc.

SA calculations can be performed using Sobel filter, where to detect the edges of each video shot (*Y* luminance component) in the period of time (*Fn*) [Sobel(F_n)] filtration is applied. For the filtered result in the shot standard deviation (std_{space}) is calculated pixel by pixel. The mentioned calculations are carried out for all shots of the video fragment. The maximum acquired value reflects the SA of the video fragment, which is calculated as follows:

$$SA = std_{space}[Sobel(F_n)] \tag{3.1}$$

TA is temporal function of the movement measure. The more movements there are between two consecutive video shots, the higher the *TA* value. If the comparable video shots are identical, then *TA* = 0.

TA value is found by calculating residual $M_n(i, j)$ between two consecutive video frames:

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j) \ , \tag{3.2}$$

where $F_n(i, j)$ is *n-th* pixel of the shot in *i-th* row and *j-th* column. *TA* is calculated as follows:

$$TA = std_{space}[M_n(i,j)] \tag{3.3}$$

TA and *SA* parameters are possible to apply for video quality monitoring, e.g., in picture freeze instance $TA = 0$, in lost shot instance both the $TA = 0$ and the $SA = 0$.

Synthetic video is used in studies [108], [109] and regular television broadcasting video is used in research [39]. In the same research synthetic video scenes were found to be negative logarithmic dependence on *TA* ($r = –0.85$).

The sample of the video fragments used in testing is available at www.youtube.com (in search write: Romass Pauliks video quality testing).

The chosen video material for video quality testing can exert substantial influence on video quality assessment or *MOS*; therefore, while selecting the video material several essential preconditions have to be observed. First of all, video fragments are selected by the content (nature, sports, animals, etc.), then it is necessary to exclude insulting, controversial or irritating video scenes (violence, horror films, politics, religion, pornography, etc.), which can essentially interfere with the respondent's concentration to complete the task, as a result of which *MOS* will decrease and validity interval limits will increase. Therefore neutral topics should be chosen as much as possible. Secondly, video scenes should be selected according to technical parameters, for each individual video scene spatial time *SA* and temporal time *TA* values should be calculated. The more delicate elements in the picture, the higher the *SA* value, the more movements in the video scene, the higher the *TA* value. Therefore, in the video quality assessment tests a wide spectrum of video scenes with different *SA* and *TA* values should be used.

During the first stage, respondents were selected by gender, age, level of education, occupation or business field [20]. The research compares the responses of different respondents selected by the following characteristics: gender, age, education, profession, as well as professionalism or experience working with digital video technologies (these respondents are called “experts”). The research concludes that the quality assessment is influenced by all characteristic features of the respondents, but only two characteristics can be regarded statistically important: the age of the respondent and the professionalism or knowledge and experience working with digital video technologies. It has been observed that younger people and respondents selected as experts assess the quality more critically in comparison with elderly people and all other respondents selected as non-specialists or non-experts. The mentioned criteria of respondent selection have been observed in all experiments. In all experiments respondents were invited who were not experts. It is important to mention that for coding algorithm improvement and development non-specialist tests of quality assessment are used. In studies [110] and [111], it has been established that quality assessment of experts and non-specialists may differ, because non-specialists when assessing the quality pay more attention to luminance changes of the picture, but experts pay more attention to element outline seen in the picture and to the general luminosity of texture.

During the second stage in order to assess the general level of vision of the selected respondents, a test of vision, colour resolution, and contrast sensitivity was carried out. To collect the above-mentioned information about the respondents, a special questionnaire was developed.

To test the respondents' colour resolution, a colour perception test was used, which allowed in a simplified way to find out whether the respondents had colour vision defects (daltonism / colour blindness). According to general statistics, about 8 % of men and 0.5 % of women have colour vision defects, where green and red colour vision defects are especially pronounced. That is why for the test there *Ishihara Colour Test* was used, which was first published in 1917 and named after its developer Prof. Dr. Shinobu Ishihara. *Ishihara* test allows identifying the problem in a simple and fast way.

To examine the respondents' colour vision, electronic pictures of 24 *Ishihara* cards were used, which were shown on the monitor consecutively one after another. Each card was shown on the screen for 3 seconds, during the card exchange a grey screen was shown with instruction how to fill the questionnaire. Respondents should fill in the questionnaire with 24 empty boxes in 96 seconds.

To examine the respondents' general vision, the electronic picture of *Snellen Chart* card was used, which was shown on the monitor. The card was shown on the screen as a static picture, during the display of the card the respondents should fill in the questionnaire. The respondents should complete the questionnaire with 29 empty boxes in 30 seconds.

However, in individual experiments to examine contrast sensitivity the electronic picture [112] of the (*Pelli–Robson Chart*) method [113] card was used, which was displayed on the monitor. The card was displayed as a static picture, during the display questionnaire had to be completed. The respondents should complete questionnaire with 48 empty letter boxes in 60 minutes.

4. Experiments of Video Quality Assessment

The fourth chapter is devoted to the description of video quality assessment experiments, which identified the relationship of subjective and objective parameters of the video image quality. The first sub-chapter includes a description of the variety of video quality assessment scales and of the screen evaluation experiment. The second sub-chapter includes a description of video compression and interface impact evaluation experiment. The third sub-chapter includes a description regarding the effects of packet loss in a quality evaluation experiment. The chapter ends with a summary and conclusions.

Impact of Video Quality Assessment Methods (Scales and Screen) on Quality Evaluation

With reference to the literature review (see Chapter 2), it is evident that in practice a lot of different video quality assessment methods are used, such as *ACR*, *DCR*, *PC*, *SCJAC*, etc. and each of them has its own specific test execution conditions – *FR*, *RR* or *NR*, one or two screens, 5, 7, 9, 11 or 100–point scale. It has also served as a major motivator for the research, which has to confirm which of the video quality testing methods are more easily perceived and understood.

Aim of the experiment:

1. To perform the literature survey on video quality assessment methods:
    - 2.1.1. To find out which of the methods is the easiest to perceive and comprehend;
    - 2.1.2. To compare one-screen and two-screen methods;
    - 2.1.3. To compare the scales of 5, 9, 11, and 100 points.

As an alternative to the controlled test bed conditions, actual living conditions were selected in this experiment (on the couch at home, at work at the desk, cafe, park, bus, etc.), where accurate fixation of technical parameters is not possible and where the quality evaluation experimental data are most often acquired through the Internet (specially prepared for experiments on the web – a web site), and through their distribution in social networks, such as *researchgate.net*, *facebook.com*, *linkedin.com*, *twitter.com*, *draugiem.lv*, *vk.com*, *ok.ru*, etc. This technique is called *crowdsourcing*.

The web site had a specialized site with an 8-question survey *skypromo.lv/iptv* included in it. In order to fill in the questionnaire a computer with an Internet connection and a web browser was needed, and completion took no more than 2–3 minutes. By using these 8 questions it was possible to find out which of the video quality testing methods (single or dual screen) and scales (5, 9, 11 or 100) were easily perceived and understood.

More than 1300 respondents participated in the experiment, 1265 respondent questionnaires were validated, 738 women and 527 men aged 11–85 years.

These tests were used in one and two-screen techniques, 5, 9, 11 and 100-point scale.

In research [96], it has been found that a majority of respondents consider a 5-point scale easier to perceive and comprehend. Table 4.1 shows that a similar result has also been obtained in research [81], where *ACR* method with a 5- and 11-point scale is recognised as the most comprehensible one.

Table 4.1

Choice of Assessment Scales

Assessment scale	Choice of respondents
5-point scale	399
9-point scale	97
11-point scale	117
100-point scale	280
5, 9, 11 or 100	372
Total	1265

In the same research, insignificant statistical differences were shown between single screen and double screen methods [96], although majority of respondents (70 %) during testing would prefer using a double screen method. In a 5-point scale, the difference between *DS* and *SDS* methods is statistically insignificant ($r = 0.78$) (see Fig. 4.1).

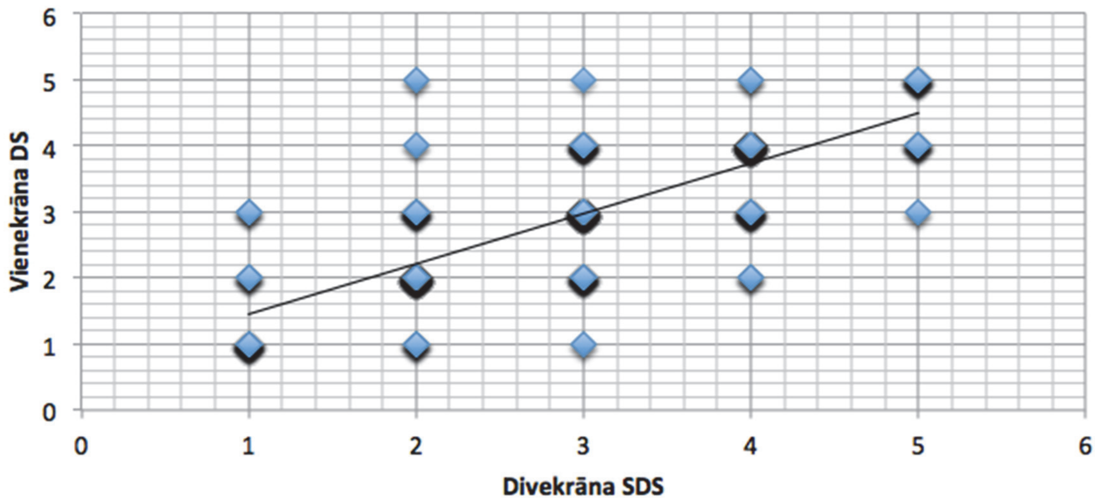


Fig. 4.1. Choice of single screen and double screen methods.

Thus, for the assessment of coding and video interface parameter influence, *PC* method (*DS* and *SDS*) is used, and for packet loss influence assessment *ACR – HR* method is used.

Main results of the experiment:

1. In the video quality evaluation, single screen and dual screen methods give statistically equivalent results ($r = 0.78$), but, when given the choice, 70 % of evaluators preferred the two-screen method.
2. It has been found that a majority of respondents consider a 5-point scale easier to perceive and comprehend.

Video Compression and Interface Impact on Quality Evaluation

With reference to the literature review (see Chapter 1), it is evident that in practice various video interfaces are used – analogue, digital, standard or high definition, each has its own specific parameters of video compression. It has also served as a major motivator for research, resulting in finding out the quality rank of the video interface, and the most important video compression parameters that affect video quality assessment.

Aim of the experiment:

1. To carry out the comparison experiments of various video signals and interface video quality, and the processing of statistical data:

1.1. To carry out the assessment of video signal and interface quality: component digital (*YCbCr* 1080i), component analogue (*YPbPr* 720p), composite (*YIQ* 576p);
1.2. To compare the results and examine whether there are statistically significant differences in video quality assessment while using *PC* (one-screen method) and *SCJAC* (two-screen method);
1.3. To carry out the assessment of synthetic video scene quality for the video signals 1080i and 576p, using the video interface of the component *YCbCr* and the composite *YIQ*;
1.4. To make the calculations of average assessment (*MOS*), standard deviation, validity intervals, and correlation coefficient.
2. If synthetic video scene quality *MOS* depends on *SA* and *TA* parameters.
3. To carry out the selection and the assessment of video scenes which will be used in the research experiments:
3.1. To carry out the selection of the video scenes according to their content (synthetic video scenes);
3.2. To carry out the selection of the video scenes according to spatial (*SA*) and temporal (*TA*) measurements.

For most research of video quality evaluation, standard video scenes are used (Nature, News, Sport, Etc.), which reflect the modern television broadcasting. In most cases, they are static video scenes, although nowadays dynamic video stories are becoming more common, mostly special effects with synthetic objects. Such films generated by computer graphic programs are imbued with various graphic special effects which do not occur in real life, so they are referred to as synthetic video clips, for example, in films such as “Avatar”, “The Lord of the Rings 3”, as well in video games.

In the research of video encoding and video interface quality assessment [114], the scheme has been used (Fig. 3.1).

Video quality assessment tests were carried out in two test groups: *DS* – Double stimulus (2 tests) and *SS* – Single stimulus (3 tests), a total of 58 video quality assessments (tests *SS* 3x12 + *DS* tests 2x11). The total test time was 60 minutes.

More than 30 respondents participated in the experiment, 25 respondent questionnaires were validated. Out of the 25 respondents, 28 % were women and 72 % men; 90 % – 20–30 years old and 10% – 30–50 years old. 28 % in their daily life wear glasses or lenses. 48 % work in engineering, 28 % in social sciences and 24 % in humanities.

In these tests *PCs* and *SCJAC* methods were used, 5- and 7-point scale.

Within the framework of ample experimental research, the results of which are shown in [37] and [38], quality assessment tests of both video interfaces and video coders are carried out.

Three video interfaces are mutually compared, where *SS1* is an analogue component signal *YPbPr* with 720p video, *SS2* is a digital HDMI signal *YCbCr* with 1080i video, but *SS3* is a composite signal *YIQ* with 576p video. The calculations of *SA* and *TA* of synthetic and natural video scenes are presented in Chapter 1.

The results of quality assessment are shown in Table 4.2 and Fig. 4.2. For each video interface, *MOS* (5 point scale, *ACR*) and validity intervals are calculated. When comparing the results of SS1 and SS2, a tight correlation (r = 0.91) is observed, but the results of *SS3* significantly differ (it is 1 to 2.5 points lower in comparison with *SS1* and *SS2*). It can be concluded that in the case of the single screen method the quality assessment differences between the analogue *YPbPr* 720p and digital *YCbCr* 1080i video interfaces are statistically insignificant.

Table 4.2

Comparison of Video Interfaces (Single Screen)

Video scenes	*SS1 (YPbPr)* 720p		*SS2 (YCbCr)* 1080i		*SS3 (YIQ)* 576p	
	MOS	*Std.*	*MOS*	*Std.*	*MOS*	*Std.*
MIS	4.28	0.25	4.56	0.27	1.84	0.31
SC	4.20	0.27	4.40	0.32	2.44	0.43
FO	4.16	0.26	4.24	0.38	1.56	0.27
RCG	3.80	0.34	4.00	0.32	1.76	0.38
CTP	3.48	0.34	3.88	0.27	1.88	0.27
FC	3.56	0.36	3.80	0.29	2.76	0.36
ColP	3.68	0.35	3.84	0.28	1.76	0.25
ConP	3.20	0.43	3.72	0.35	2.64	0.39
ST	3.20	0.34	3.28	0.35	1.48	0.24
SP	3.36	0.36	3.28	0.40	2.24	0.30
LCG	3.12	0.38	2.80	0.38	1.68	0.31

Video quality assessment depends on video scene, e.g., in the case of *SS1* interface *MIS* = 4.28 and *LCG* = 3.12 or in the case of *SS2 MIS* = 4.56 and *LCG* = 2.80. The assessment of the video scene is influenced by Temporal Activity *TA*. The higher the *TA* value, the lower video scene quality assessment *MOS*. However, the value of the Spatial Activity *SA* does not influence synthetic video scene assessment; quality assessment differences are not observed either in big or small *SA* value instances.

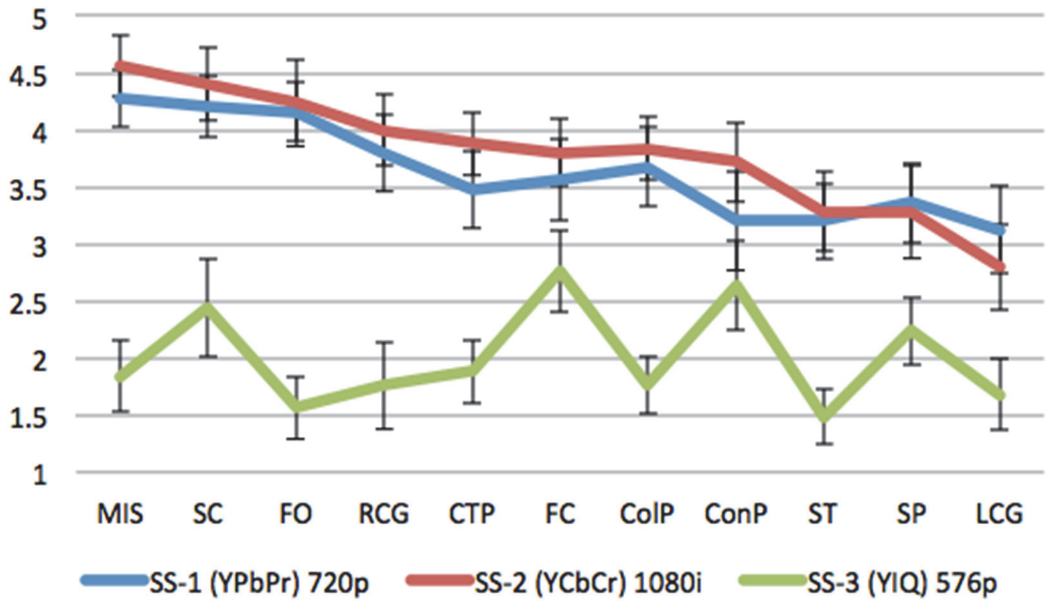


Fig. 4.2. Single screen *MOS* values for 720p, 1080i and 576p video signals.

It can be concluded that between 720p and 1080i coded videos, there are statistically insignificant differences. For 720p video broadcasting and storage, it is necessary to have 10–15 % less broadcasting speed and volume if compared to 1080i video. To save the resources, it is advisable to choose 720p video. No significant quality differences are observed between the analogue (*YPbPr* 720p) and the digital (*YCbCr* 1080i) component video. However, there are significant quality differences when compared to the analogue composite (*YIQ* 576p) video. It is advisable to choose either analogue or digital component video.

In order to mutually compare two different video interface specifications, two video signals were compared on two screens simultaneously using one and the same digital video interface, where *DS1* is HDMI signal *YCbCr* with 1080i video and *DS2* is HDMI signal *YCbCr* with 576p video.

Table 4.3

Comparison of Video Coders (Two Screens)

Video scenes	*DS1 (YCbCr vs YIQ) 576p*		*DS2 (YCbCr vs YIQ) 1080i*	
	MOS	*Std.*	*MOS*	*Std.*
MIS	–0.76	0.77	–0.96	0.83
SC	1.04	0.64	0.88	0.51
FO	–1.04	0.69	–1.44	0.71
RCG	–1.88	0.58	–1.88	0.56
CTP	–0.44	0.56	0.52	0.60
FC	0.00	0.49	–0.68	0.60
ColP	–0.84	0.58	–1.24	0.52
ConP	1.48	0.46	1.28	0.63
ST	–1.28	0.71	–2.24	0.42
SP	–0.64	0.37	–1.52	0.46
LCG	0.36	0.59	–0.28	0.55

The results of quality assessment are shown in Table 4.3 and Fig. 4.3. For each video interface *MOS* (7 point scale, *SCJAC*) and validity intervals have been calculated. Between *DS1* and *DS2* a tight correlation (r = 0.89) can be observed. It means there are statistically insignificant differences between coded videos 576p and 1080i, where in both cases one and the same video interface was used on two independent screens. In *DS1* and *DS2* tests (Table 4.3 and Fig. 4.3), there is exactly the same video quality assessment dependence on video scene as in *SS1*, *SS2* and *SS3* (Table 4.2 and Fig. 4.2). In the case of synthetic video scenes, there is a negative logarithmic dependence on space-time activity *TA* (*Temporal Activity*), for example, *DS2* against lg(TA) r = –0.85. The calculations of video *SA* and *TA* are presented in Chapter 1.

For video scenes with high temporal activity *TA* video quality assessment in case of video 1080i was lower than for video 576p. The higher *TA* of the video scene and the higher quality video (1080i) is used, the greater the probability to observe visually seen impairments (artefacts). However, for video scenes with high spatial activity *SA* video quality assessment in case of 1080i video was higher than for video 576p.

The assessment of synthetic video scene is influenced by Temporal Activity *TA*, the higher the *TA* value, the lower the video scene quality assessment *MOS*. However, Spatial Activity *SA* value does not influence the synthetic video scene assessment: there are not quality assessment differences in cases with either high or low *SA* values.

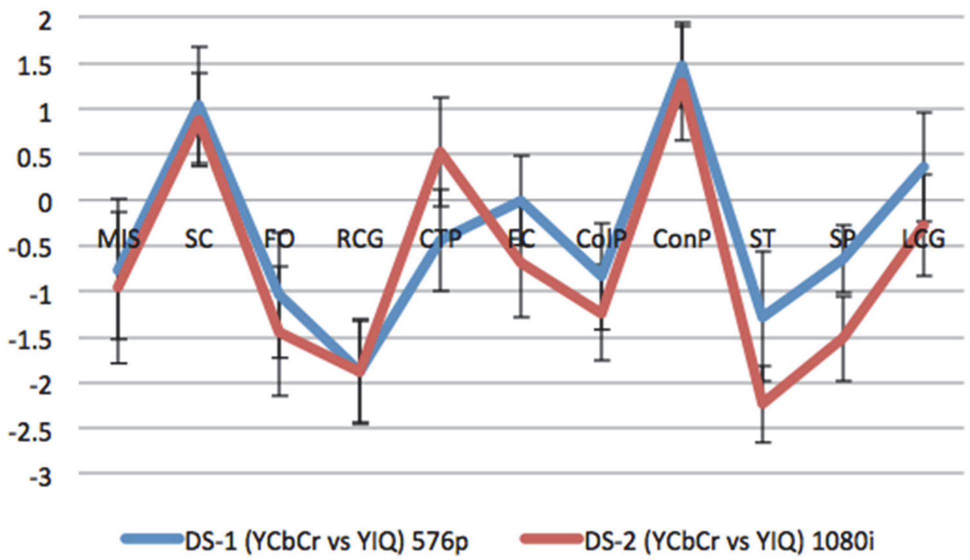


Fig. 4.3. Two-screen *MOS* values of 576p and 1080i video signals.

Table 4.4 demonstrates *TA* and *SA* calculations. Synthetic video is used in [115], [116]. In the same studies, synthetic video scenes have been found to have negative logarithmic dependence on *TA* (r = -0.85).

Table 4.4

SA and *TA* Values of Synthetic Video

Video	*SA*	*TA*	*T (s)*
MIS	64.92	8.23	47
SC	2.23	1.08	47
FO	27.27	7.31	73
RCG	3.43	23.71	47
CTP	31.05	2.77	37
FC	22.89	15.7	47
ColP	14.59	5.86	71
ConP	24.56	2.38	43
ST	52.04	46.16	79
SP	30.23	47.53	55
LCG	11.8	10.57	67

The first column of Table 4.4 provides synthetic movie abbreviations (*MIS*, *SC*, *FO*, *RCG*, *CTP*...), *TA* (*Temporal Activity*) computed value, *SA* (*Spatial Activity*) computed value and video fragment duration in seconds *T (s)*.

Main results of the experiment:

1. The synthetic video quality *MOS* score is inversely proportional to the logarithm of temporal activity (*TA*) values (r = -0.85).

Impact of Video Transmission Packet Loss on Quality Evaluation

With reference to the literature review (see Chapter 1), the quality of the video transmission over IP networks is affected by the packet delay, tremors and loss. Large packet delays and packet losses cause tremors; therefore, packet loss ratio is used in the experiments. It has also served as a major motivator for research, resulting in finding out how the amount of packet loss and distribution affect video quality assessment.

Aim of the experiment:

1. To carry out the assessment experiments of packet loss influence on video quality, and the processing of statistical data:
   1.1. To accomplish packet rejection for variable speed video (*VBR*) stream;
      1.1.1. To reject packets 0.25 %, 1 %, 2 % and 5 % from the total number of video scenes;
      1.1.2. For packet rejection to use Poisons, Gaussian and uniform probability distribution.
   1.2. To perform video quality assessment:
      1.2.1. To carry out video quality assessment using *ACR – HR* (for 0.25 % and 1 % packet loss);
      1.2.2. To carry out video quality assessment according to *MSE* and *PSNR* (for 0.25 %, 1 %, 2 %, and 5 % packet loss);
   1.3. To perform the measurements of the *GOP* shots (*I* and *P* shots) of damaged video scenes;
   1.4. To make the calculations of average assessment (*MOS*), standard deviation, validity intervals, and correlation coefficient.
2. To carry out the selection and the assessment of video scenes which will be used in the research experiments:
   2.1. To carry out the selection of the video scenes according to their content (nature, animals, sports, news, etc.);

2.2. To carry out the selection of the video scenes according to space (*SA*) and space-time (*TA*) measurements;
2.3. To make the comparison of the original and damaged (damage caused by coding or transmission) video scenes according to *MSE* and *PSNR* measurements.

It is known that in IP networks video streaming is affected by packet delay, tremors and loss. Large packet delays and packet losses cause tremors, which the viewer sees on the screen as damage (artefacts), ultimately decreasing the video quality score (*MOS*). The visually apparent damage nature and extent are closely related to the video packet encapsulations and compression technologies. In this experiment, it was required to find out how the amount of packet loss (0.25 %, 1 %, 2 % and 5 %) and the loss distribution (Poisson, Gaussian and evenly) affect the video quality assessment (*MOS*). As a result of the research and literature analysis [39], it has been found out that in some studies a 1 % packet loss amount corresponds to *MOS* = 2.5, but in the other research the packet loss amounts to a ten times higher or lower number. It is possible that these huge differences in outcomes are related to the poorly defined experimental conditions, and it is often unclear what are the conditions of the experiment, and what is being tested, for example, the impact of packet loss, error correction algorithm performance, buffering performance, etc.

In the research regarding the impact of packet loss on the video transmission quality assessment [39], the scheme has been used (Fig. 3.2). Like in the encoder, also in the transmission channel quality assessment tests *FR* (*Full Reference*) technique was used, where pixel by pixel (pixel-by-pixel) the original and perceived / damaged (impaired) video images are compared by performing time-consuming and bulky computing operations.

Video quality evaluation tests were carried out in three test groups: respondent vision tests, reading tests and the key study. The total test time was 45 minutes.

More than 50 respondents participated in the experiment, 40 respondent questionnaires were validated. 90 % of those were 20–25 years old.

In these tests a 5-point scale *ACR-HR* method was used.

The above-mentioned aspects have given motivation to look for relevance between those various probability distributions (Poisson's, Gaussian and uniform) that have been used to describe the statistics of packet abandonment. In the given research, the correlation has been searched and found between limit values of packet loss amount and *MOS* or video quality assessment.

In research [39] dealing with 0.25 % packet loss influence on *MOS*, there is not any difference observed between the Poisson's, Gaussian and uniform distribution (Fig. 4.4) used as incidental distributions when abandoning packets.

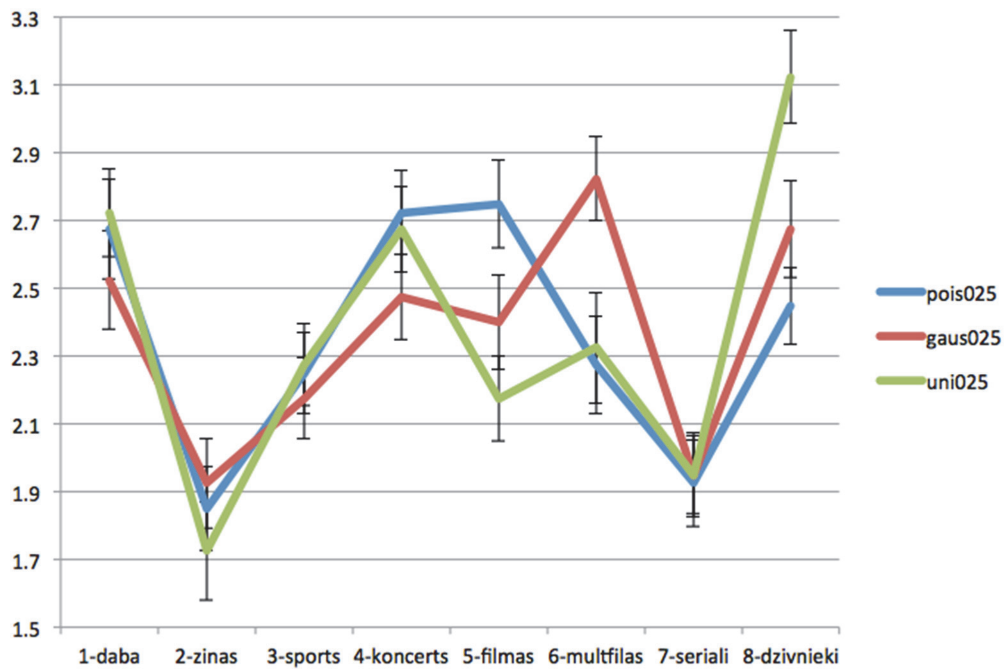


Fig. 4.4. *MOS* values for Poisson's, Gaussian, and uniform distributions.

In research [39] about the influence of packet loss on *MOS*, a significant difference can be observed between 0.25 % and 1 % packet loss according to Poisson's, Gaussian, and uniform distributions depending on video scenes (see Fig. 4.5).

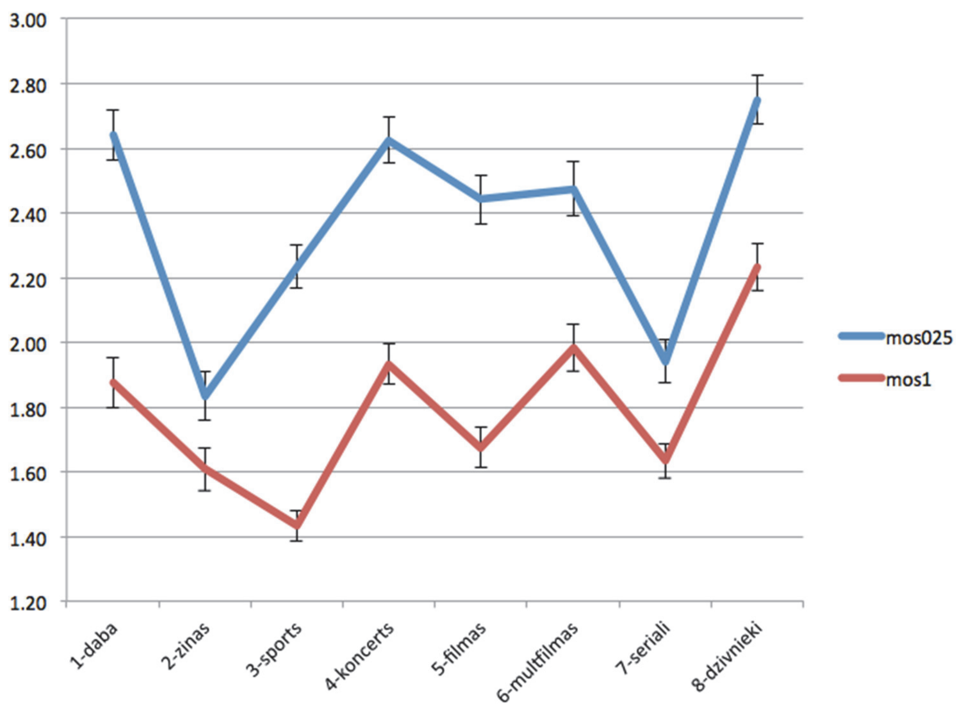


Fig. 4.5. *MOS* values for 0.25 % and 1 % packet loss.

In research [39] dealing with the influence of the packet loss on *MOS*, a rather poor correlation is stated with *PSNR* (r = 0.28 – 0.60) (see Table 4.5).

Table 4.5

MOS (0.25 % and 1 % Packet Loss) Compared to *MSE* and *PSNR*

Scenes	*MSE*		*PSNR* (dB)		*MOS*	
	0.25 %	1 %	0.25 %	1 %	0.25 %	1 %
Nature	222.5	261.6	25.28	23.88	2.64	1.88
News	133.5	263.8	27.55	23.39	1.83	1.61
Sports	205.4	447.5	25.28	22.12	2.23	1.43
Concerts	161.7	464.4	26.52	21.96	2.63	1.93
Films	130.8	188.4	27.68	25.81	2.44	1.68
Cartoons	94.9	223.8	30.05	24.50	2.48	1.98
Serials	177.5	496.1	25.60	22.17	1.94	1.63
Animals	72.9	141.2	29.22	26.80	2.75	2.23

Although *PSNR* rather poorly correlates with *MOS* (0.25 % r = 0.28 and 1 % r = 0.60 for packet loss), it can serve as a general indicator of assessment of impairment amount. Figure 4.6 shows *PSNR* calculations for 0.25 %, 1 %, 2 % and 5 % packet loss.

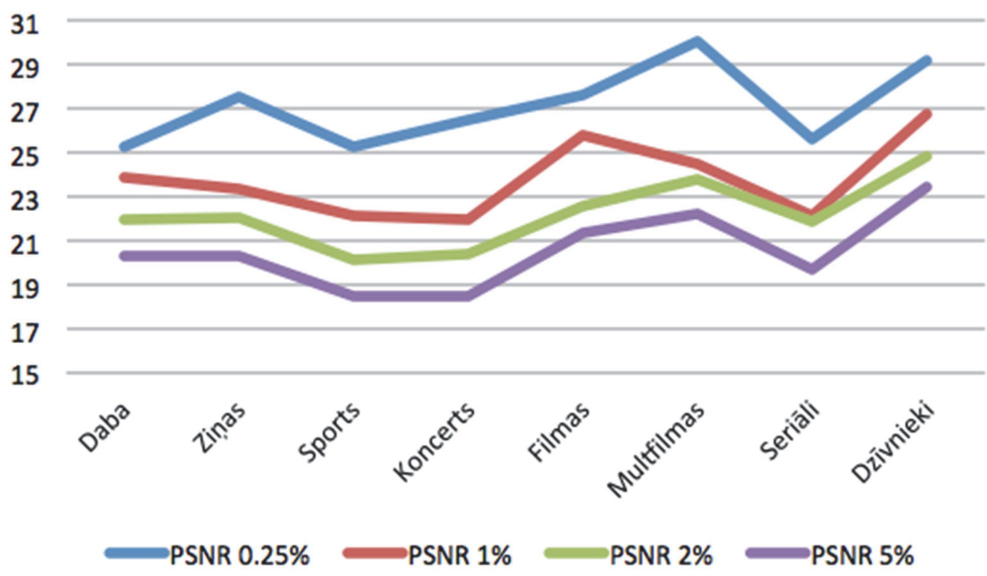


Fig. 4.6. *PSNR* depending on video scene.

Although video quality assessment (*MOS*) poorly correlates with objective video quality assessment (*PSNR* and *lgMSE*) for packet loss influence assessment (r = 0.28 – 0.60), in cases of large packet loss (≥1 %), *PSNR* and *lgMSE* can be used to forecast *MOS*.

Main results of the experiment:

1. When packet loss is emulated with the Poisson, Gaussian or uniform probability distribution, the video quality evaluation *MOS* differences are not statistically significant.
2. The video streaming packet losses cannot exceed 0.25 %, as it corresponds to video quality assessment *MOS* = 2.4, which is very close to *MOS* = 2.5, which is commonly referred to as the quality acceptance threshold.

THE MAIN RESULTS OF THE DOCTORAL THESIS

The author of the present Doctoral Thesis has found several correlations of video image quality subjective and objective parameters, developed and approbated the methodology of video quality assessment, worked out recommendations for improvement of video quality assessment methods, and also established the testbed for testing and research of video coding, and element quality, interference resistance, and performance of video transmission system. Methodology and testbed have been the most important preconditions in the experimental research accomplishment.

The aim of the research has been achieved and the tasks accomplished, which is proved by the reached results:

- The literature review has been made on human visual perception, their main elements and their parameters influencing video quality assessment; the main human visual perception characteristics have been identified that affect video quality assessment;
  - Human visual perception sensitivity to light, contrast and colours has been taken into account;
  - Vision acuity (*Snellen*), colour vision (*Ishihara*), and contrast sensitivity (*Pelli-Robson*) tests have been carried out.

One of the most important factors is the calibration of the monitors applied in the experiments, using vectroscope and waveform monitor, as well as measurements and tuning of luminance level, *AWB* (Automatic White Balance) and colour temperatures.

For a simplified way in order to assess the general level of vision of the selected respondents, a simplified examination test of human visual perception has been performed – vision acuity (*Snellen*), colour vision (*Ishihara*), and contrast sensitivity (*Pelli-Robson*) tests.

- The literature review has been made on video signal transmission and operational principles of compression technologies, their main elements and their parameters influencing video quality assessment. The main video signal interfaces and transmission parameters have been identified that affect video quality assessment;
  - Video interfaces (analogue, digital, *YIQ*, *YPbPr*, *YCbCr*), video signal encoding and conversion;
  - Video transmission parameters: packet loss, delay and jitter.
- The literature review has been made on video signal transmission and operational principles of compression technologies, their main elements and their parameters influencing video quality assessment. The main video signal compression parameters have been identified that affect video quality assessment;
  - Video signal compression parameters: 576p and 1080i video signal encoding and conversion;
  - Video signal compression parameters: 720p and 1080i video signal encoding and conversion;

It has been necessary to carry out experiments on video encoding, including the impact of video interfaces and video signal compression parameters on the quality, and video transmission, including video transmission loss impact on quality.

In transmission systems, there are packet delays and jitter, which to a certain extent can be compensated with buffer memory. However, communication channel breakage as well as overflow and underflow of buffer memory create packet loss. Jitter and delay that cannot be compensated by buffer memory create packet loss. The above-mentioned aspects serve as the basis for the fact that during the research experiments exactly the packet loss has been tested but not the packet delay or jitter influence.

Poisson, normal and uniform distributions of the process are designed to be used under various conditions, but none of them can be used to describe the packet loss statistics. Therefore, all distribution processes used in emulated packet loss statistics have been experimentally verified.

The aim of the video signal compression is to ensure as effective compression as possible, e.g. *H264 AVC P10*, by the constant video quality assessment. Compared to *MPEG – 2 P2*, it ensures even up to 50 % more effective compression. Similar correlation is found for *H265 HEVC* compared with *H264 AVC P10*. The higher the degree of compression, the more complicated the decoding algorithms and the bigger calculation capacity are necessary. The effectiveness of video coder compression can be increased taking into account the peculiarities of human visual perception and video scene construction principles. Therefore, video quality assessment or *MOS* depends on video scene, coder chosen, and coding parameters. The higher the degree of compression, the greater the probability to observe video damage or artefacts, especially they are noticeable in those video areas where there is movement. Similar correlation could be observed in the case of packet loss; however, the shape of impairments or visually noticeable artefacts was significantly different.

The overload of transmission network creates packet jitter, which can be compensated by the buffer memory. However, communication channel breakage or buffer memory overflow will create packet loss. Packet loss influences any network traffic, but packet jitter can significantly influence the real time data flow services, e.g., *VoD* (*Video on Demand*) unicasts video supplying service where *PVR* (*Personal Video Recorder*) type control commands are used (record, delete, play, pause, fast playback forwards, fast playback backwards), because each command is performed by the user and expects immediate reaction. High average value of packet delay and huge jitter directly influence video image quality, while it is possible to compensate them with buffer memory. For real time interactive services the average value of packet delay and the big jitter can significantly influence the total service quality assessment. However, it cannot directly influence and it is not connected with video quality assessment. As a result of the packet loss, significant video quality assessment differences may appear between various encapsulation methods and streaming protocols, and also between *MPEG* coder *I*, *P* and *B* shot data distribution among those packets. One *IP MPEG* packet loss is equal to packet loss of seven *DVB MPEG*. Theoretically, losing *I* or *P* frames, finally there should be greater video quality fall than losing *B* frames, but in real broadcast systems complete loss of *I* and *P* frames is very rare because each of these frames is divided into many independent *IP* packets. During broadcasting, there are lost packets, which are part of one or several *I* and *P* frames. Therefore, *MOS* is only partly dependent on the number of *I* and *P* frame damage, rather much more it depends on the size of damage in each definite *I* and *P* frame as well as on the place the damaged *I* and *P* frame are shown on the screen. However, the place of impairment on the screen depends on video scene and its coding system *GOP* (*I*, *P*, *B* frames). The depiction of the movement vector or movement changes in video scene directly depends on P frames. P shot damage as a result of packet loss is shown on the screen as visually noticeable artefacts exactly in those video scene places where there is movement, so the spectator pays more attention to it. In its turn, in scene areas with static images, damage of *I* and *P* frames are more difficult to notice as the spectator pays less attention to them.

- The review of literature has been made on video quality assessment methods; as a result, appropriate methods have been chosen for the research experimental tests:
    - The recommendations have been developed for the choice of video quality assessment methods. The recommended criteria for the choice of method are as follows: what should be assessed for coding or broadcasting process influence on video quality; how to assess general quality or the quality of impairment influence; which of the methods is more easily perceivable and comprehensible for respondents, as well as which method is the quickest, which is the most accurate;
    - The guidelines have been developed for improvement of video quality assessment methods, where both the marked and the non-marked quality assessment scales are simultaneously used, thus obtaining data about the general and acceptance quality of the assessable fragment. The developed method gives a possibility to obtain the assessment of two-dimensional *MOS* quality and acceptance;
    - It has been found that in recommendation [32] there are unclear and generalized definitions of various video parameters and related coefficients of limit values, e.g. packet loss should be less than 10 %. However, as a result of the author's experimental research, it has been found that if packet loss is above 1 % then *MOS* is unacceptably high *MOS* = 1.4 – 2.2;
    - It has been found that in recommendations [32] and [31] in order to assess quality, many quality assessment parameters and related coefficients are

developed. In practice, to carry out such a quality assessment model is complicated and time-consuming; it could be said that it is almost impossible. The number of parameters should be significantly decreased and the limit values of coefficients, which are approbated experimentally with video quality assessment tests, should be accurately defined.

Primarily, the choice of video quality assessment method should be guided by the fact what should be assessed – compression or broadcasting process influence on quality; how to assess: the general quality or the influence of the damage (degradation) on quality. Secondly, it is important to take into account which of the methods is more perceivable and comprehensible for respondents; which method is faster, which is more accurate.

ACR and *ACR – HR* can be considered the most popular methods of general quality assessment, which are usually used to assess video signal broadcasting influence. *ACR* and *ACR – HR* are advisable to use for assessment of packet loss influence. However, *DCR* is known as brought-in damage or degradation level assessment, which can be used for assessment of both the compression and the broadcasting influence. However, *PC* and *SCJAC* are general quality assessment methods, which are advisable to use for video compression influence assessment. For tuning of different compression parameters, where high accuracy is demanded, it is advisable to use *PC* or *SCJAC* methods.

In practice, there are various methods to assess video image quality, e.g. *FR* (*Full Reference*), *RR* (*Reduced Reference*), *NR* or *ZR* (*No Reference* or *Zero Reference*). *FR* is a popular and rather simple method for subjective and objective video image quality assessment. Comparably insignificant research activities are devoted to *RR* and *NR/ZR* methods [117]. In the experiments of the present Doctoral Thesis, *FR* subjective and objective video quality assessment methods have been used.

The duration of the human short-term memory lasts for 15 till 30 seconds, but the size of memory is limited by 7 ± 2 elements. Therefore, the duration of video fragment used for image quality testing should not be longer than 30 seconds, as well as the number of impairments noticeable in the image should not be more than 9.

MOS is usually used as the indicator of the general video quality or brought-in damage assessment, which is called one-dimensional quality assessment because it does not include all peculiarities of the human vision perception. However, *MOS* for more than ten years can be called “de facto” measurement in video quality assessment.

For *MOS* data acquisition and calculation, various video quality assessment scales are used. Although theoretically 9, 11, and 100 point scales should have smaller standard deviation if compared to a 5-point scale, in practice these differences are statistically insignificant. Therefore, one of the main criteria for choice of quality assessment scale can be the scale chosen by respondents, which is recognized as the most easily perceivable and comprehensible.

The correct choice of respondents and their selection is significant. First, the respondents are selected by gender, age, education, as well as other social economic indicators. Second, all selected respondents using special tests have been assessed by general vision acuity (*Snellen*) and colour vision (*Ishihara*) tests. In order to consider the video quality assessment test results plausible, the number of respondents should be larger than 15; however, in various literature sources this number fluctuates from 6 to 40.

- The testbed has been created for carrying out video quality assessment experiments:
  - The laboratory room has been created according to technical parameters [21], [22] and [66], including lighting of the room, measurement equipment for measuring lighting colours and micro-climate;
  - Equipment has been developed for laboratory video coding and broadcasting network, including *LCD* displays, hardware and software for video server (*HE*), video client (*STB*) and broadcasting network (*Ethernet*, *IP*, *MPEG*) emulation.
- The methodology for carrying out video quality assessment experiments has been developed with the help of which video coding and transmission technical parameters can be modelled:
  - In order to model different packet loss scenarios (delay, jitter and packet loss) in IP networks, including different probability distributions of packet abandonment (Poisson's, Gaussian, and uniform), packet abandonment emulations have been created;
  - The emulations of video coding, and of impairment caused by packet loss have been developed in order to model visually noticeable artefacts on the screen (damage of image blocks and slices, blurred image elements, etc.);
  - The vision acuity (Snellen), colour vision (Ishihara), and contrast sensitivity (Pelli-Robson) tests have been carried out.

Technical parameters of video compression and broadcasting can significantly influence the video quality assessment and, therefore, to ensure plausible experiment results the appropriate laboratory and methodology have been created. First, for carrying out experiments the laboratory has been created, which ensured identical conditions for all respondents and controllable and constant technical parameters. Second, the methodology has been developed, which describes specific laboratory parameters, their calibration and measurements, guidelines for choice of methods applicable to the experiments of quality assessment, as well as the procedure of processing and analysis of the obtained data.

During the research experiments while selecting and classifying respondents, at first visual screening (vision quality assessment) has been carried out. In vision screening, the vision acuity has been assessed applying *Snellen* method, but for colour vision testing *Ishihara* method has been used. Vision screening has been set as the criterion when selecting the data in order to separate plausible results from doubtful ones.

The video material chosen for video quality testing can exert significant influence on video quality assessment or *MOS*; therefore, when selecting the video material several significant preconditions should be taken into account. First of all, video fragments are selected by the content (nature, sports, animals, etc.), then it is necessary to exclude insulting, controversial or irritating video scenes (violence, horror films, politics, religion, pornography, etc.), which can essentially interfere with the respondent's concentration to complete the task, as a result of which *MOS* will decrease and validity interval limits will increase. Therefore neutral topics should be chosen as much as possible. Secondly, video scenes should be selected according to technical parameters, for each individual video scene spatial time *SA* and temporal time *TA* values should be calculated. The more delicate elements in the picture, the higher the *SA* value, the more movements in the video scene, the higher the *TA* value. Therefore, in the video quality assessment tests a wide spectrum of video scenes with different *SA* and *TA* values should be used.

If in the testing process, videos with different technical parameter *HRC* are used, then for the mutual comparison of those different *HRC* video fragments and for comparison with *SRC* original video *PSNR* can be used. Although *PSNR* rather poorly correlates with *MOS*, this is the method which can be used a general quantitative assessment indicator of the compression level and the impairment amount.

Respondents have been selected by gender, age, level of education, profession or occupation areas. It has been found out that the quality of assessment affects all characteristics of respondents, but only two features are statistically significant – the respondent's age and professional skills or knowledge and experience of working with digital video technologies (experts). It is important to mention that the improvement and development of compression algorithm are used for evaluation of quality tests by non-experts rather than experts. Thus, respondents who are not experts have been invited to take part in all experiments.

- The assessment experiments of various video quality assessment methods have been carried out and statistical data processing has been performed;
  - During the video quality assessment using single screen and double screen methods, the obtained results (*MOS*) are statistically equivalent ($r = 0.78$), but giving the opportunity to choose, 70 % of respondents would choose the double screen method;
  - It has been found out that out of 1265 respondents 399 respondents consider the questionnaire according to a 5-point scale more easily perceivable and comprehensible, 97 respondents prefer a 9-point scale –, 117 respondents would rather use a 11-point scale, while 280 consider a 100-point scale to be appropriate. However, 372 respondents have recognised that any scale is equally perceivable and comprehensible.

- The assessment experiments of various video signals and interfaces have been carried out and statistical data processing has been performed:
  - Statistically insignificant differences ($r = 0.91$) are between high resolution component digital video (*YCbCr* 1080i) and component analogue video (*YPbPr* 720p), but the results of the composite analogue video (*YIQ* 576p) significantly differ (for 1 to 2.5 *MOS* points lower than if compared with (*YCbCr* and *YPbPr*);
  - The quality assessment of (*MOS*) of synthetic video scenes is inversely proportional to logarithm of corresponding spatial time (*TA*) measurements of video scenes ($r = -0.85$);
  - Between 576p and 1080i coded, where in both cases one and the same video interface has been used on two independent screens, there are insignificant differences observed ($r = 0.89$).

- The experiments of packet loss influence on video quality assessment have been carried out and statistical data processing has been performed;
  - In general, quality assessment (*MOS*) of video scenes does not differ using various incidental processes (Poisson’s, Gaussian, or uniformed) for packet abandonment;
  - The amount of packet loss of 0.25 % according to Poisson’s, Gaussian, or uniformed case distribution corresponds to average quality assessment *MOS* = 2.4 which is the margin between average and poor video quality. The obtained result is close to *MOS* 2.5, which is considered to be an acceptable quality marginal value;

- Although *PSNR* rather poorly correlates with *MOS* (0.25 % $r = 0.28$ and 1 % $r = 0.60$ for packet loss), it can serve as a general indicator of assessment of the size of impairment;
  - No remarkable connection has been found between *PSNR* and *I*, *P* shot damage, as well as there is no correlation between *MOS* and *I*, *P* shot damage;
  - Although *PSNR* rather poorly correlates with *MOS*, it can serve as a compression and packet loss indicator of assessment of the size of impairment.
- The selection and assessment of video scenes have been carried out;
  - The synthetic video scene assessment is influenced by temporal activity (*TA*), the greater the *TA* value, the lower the video scene quality assessment *MOS*;
  - The higher the *TA* of the video scene and the higher quality video is used (1080i against 576p), the greater the probability to observe visually noticeable impairments (artefacts);
  - For synthetic video scenes with large spatial activity (*SA*), the assessment of video quality in 1080i video instance is higher than that of 576p video;
  - The higher the degree of compression, the lower the *PSNR* value, and the greater the packet loss, the lower the *PSNR* value.

Video quality assessment is dependent on the video scene, due to packet loss or excessive compression result visual artefacts appear directly in the video scene, in place of movement, so that the viewer pays close attention to it. By contrast, static scenes of damage in video are less visible, because the viewer is not focused on the image movements.

It can be concluded that between 720p and 1080i coded videos there are statistically insignificant quality differences. For broadcasting and storage of 720p video, it is necessary to have about 10-–15% less transmission speed and size if compared to 1080i video. In order to save resources, it is advisable to choose 720p video. There are not significant quality differences observed between the analogue (*YPbPr* 720p) and digital (*YCbCr* 1080i) component video, but there are significant quality differences if compared to analogue composite (*YIQ* 576p) video. It is advisable to choose analogue or digital component video.

For statistics of packet loss, delay and jitter various distribution processes have been used. In cases of small packet loss (≤ 1 %) which can be considered a rare event without memory because packet loss does not depend on previously lost packet, Poisson's distribution is used. If it can be assumed that packet loss is influenced by many independent processes then for packet loss statistics regular or Gaussian distribution can be used. However, discreet uniform distribution is a typical generator of sample random value, which sometimes is also used to describe the statistics of packet loss.

It has been found that the Internet and *VBR* video data flow have self-equivalent characteristics independently of the content of the video material (video conference, video call, feature film, cartoon, sports broadcast, etc.). The afore-mentioned aspects substantiate the use of exactly VBR video data flow rather than *CBR*.

Although Poison's, regular and uniformed distribution processes are envisaged to be used in various conditions, any of them can be used to describe the statistics of packet loss because there are no significant differences observed among those distribution processes.

Packet loss (≥ 0.25 %), probably, is usable in video streaming. *IP* packet 0.25 % loss corresponds to the average quality assessment *MOS* = 2.37, which is very close to MOS = 2.5, which is also called a marginal value of service level acceptance (*acceptance* or *acceptability*). *MOS* = 2.5 is considered to be the marginal value of one-dimensional general quality assessment, which cannot be regarded as the only and truthful measurement for the determination of marginal value of service level acceptance.

Although the video quality assessment (*MOS*) poorly correlates with the objective video quality assessment (*PSNR* and *lgMSE*) to assess the influence of packet loss ($r = 0.28 - 0.60$), in the cases of big packet loss (≥1%) *PSNR* and *lgMSE* can be used to forecast *MOS*.

The methodology and the laboratory created within the framework of the Thesis can be used for both the scientific and the commercial activities.

Future research:

- To perform video quality assessment experiments of packet loss influence where the amount of packet loss is less than 0.25 %. The task is to find marginal values of the packet loss, which can be connected to minimal acceptable video quality and acceptance assessment applying a two-dimensional *MOS* method;
- To equip the laboratory with three-dimensional (*3DTV*) video monitors and other transmission channels of communication system, e.g. *WiFi*, *3G/4G* and *DVB*;
- To elaborate the methodology using subjective and objective video quality assessment methods suitable for *3DTV*.

The research on linking the human perception characteristics with objective parameters will continue, until identical results are obtained by video quality tests and *QoE* metrics.

BIBLIOGRAPHY

[1] TNS Latvia. 21 November 2013. [Online]. Availabe at: http://www.tns.lv/?lang=lv&category=showproduct&id=tvm. [Accessed on 21 November 2013].

[2] Cisco VNI. 29 May 2013. [Online]. Availabe at: http://www.cisco.com/en/US/netsol/ns827/networking_solutions_solution_category.html. [Accessed on 14 November 2013].

[3] Cisco GCI. 1 October 2013. [Online]. Availabe at: http://www.cisco.com/en/US/netsol/ns1175/networking_solutions_solution_category.html. [Accessed on 14 November 2013].

[4] dvb.org. 4 March2015. [Online]. Availabe at: https://www.dvb.org/. [Accessed on 4 March 2015].

[5] atis.org. 5 March 2015. [Online]. Availabe at: http://www.atis.org/. [Accessed on 5 March 2015].

[6] E.800, Definitions of terms related to quality of service. ITU–T Rec., Geneva, Switzerland, 2008.

[7] Bonastre, O. M., Montpetit, M.-J., Cesar, P. Advances in IPTV technologies. *Signal Processing: Image Communication,* vol. 26, No. 7, 2011, pp. 325–326.

[8] Stockhammer, T. Dynamic adaptive streaming over HTTP – Standards and design principles. *MMSys'11 – Proceedings of the 2011 ACM Multimedia Systems Conference*, 2011.

[9] Gozdecki, J., Jajszczyk, A., Stankiewicz, R. Quality of service terminology in IP networks. vol. 41, No. 3, 2003, pp. 153–159.

[10] Duan, Z., Zhang, Z.L., Hou, Y.T. Service Overlay Networks: SLAs, QoS, and Bandwidth Provisioning. *IEEE/ACM Transactions on Networking,* vol. 11, No. 6, 2003, pp. 870–883.

[11] Schatz, R., Hoßfeld, T., Janowski, L., Egger, S. *From packets to people: Quality of experience as a new measurement challenge,* vol. 7754, 2013, pp. 219–263.

[12] Brooks, P., Hestnes, B. User measures of quality of experience: Why being objective and quantitative is important. *IEEE Network,* vol. 24, No. 2, 2010, pp. 8–13.

[13] Reichl, P. From Charging for Quality-of-Service to Charging for Quality-of-Experience. *Annals of Telecommunications, special issue on "Quality of Experience and Socio-Economic Issues of Network–Based Services",* 2010, pp. 189–199.

[14] ATIS-0800008. QoS Metrics for Linear (Broadcast) IPTV, 2011.

[15] Hands, D., Wilkins, M. A Study of the Impact of Network Loss and Burst Size on Video Streaming Quality and Acceptability. *in Michel Diaz; Philippe Owezarski & Patrick Sénac, ed., IDMS,* Springer, 1999, pp. 45–57.

[16] Reichl, P., Tuffin, B., Schatz, R. Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience. *Telecommunication Systems,* 2011, pp. 1–14.

[17] Jumisko-Pyykkö, S., Häkkinen, J., Nyman, G. Experienced quality factors – Qualitative evaluation approach to audiovisual quality. *Proceedings of SPIE – The International Society for Optical Engineering*. San Jose, CA; United States, 2007.

[18] QUALINET. 12 February 2015. [Online]. Availabe at: http://www.qualinet.eu. [Accessed on 12 February 2015].

[19] Le Callet, P., Möller, S., Perkis, A. Qualinet White Paper on Definitions of Quality of Experience. European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), QoMEX, Lausanne, Switzerland, 2013.

[20] Jumisko-Pyykkö, S., Häkkinen, J. Profiles of the evaluators – Impact of psychographic variables on the consumer-oriented quality assessment of mobile television, *Proceedings of SPIE – The International Society for Optical Engineering*, San Jose, CA; United States, 2008.

[21] BT.500, Methodology for the subjective assessment of the quality of television pictures. ITU–R Rec, Geneva, Switzerland, 2012.

[22] P.910. Subjective video quality assessment methods for multimedia applications. ITU–T Rec, Geneva, Switzerland, 2008.

[23] De Moor, K., Ketyko, I., Joseph, W., Deryckere, T., De Marez, L., Martens, L., Verleye, G. Proposed framework for evaluating quality of experience in a mobile, testbed–oriented living lab setting. *Mobile Networks and Applications,* vol. 15, No. 3, 2010, pp. 378–391.

[24] Catellier, A., Pinson, M., Ingram, W., Webster, A. Impact of mobile devices and usage location on perceived multimedia quality. *4th International Workshop on Quality of Multimedia Experience, QoMEX 2012*, Melbourne, VIC; Australia, 2012.

[25] Chen, K.T., Wu, C.-C., Chang, Y.-C., Lei, C.-L. A crowdsourceable QoE evaluation framework for multimedia content. *MM'09 – Proceedings of the 2009 ACM Multimedia Conference, with Co-located Workshops and Symposiums*, Beijing; China, 2009.

[26] Chen, K.-T., Chang, C.-J., Wu, C.-C., Chang, Y.-C., Lei, C.-L. Quadrant of euphoria: A crowdsourcing platform for QoE assessment. *IEEE Network,* vol. 24, No. 2, 2010, pp. 28–35.

[27] Hossfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., Tran-Gia, P. Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing. *IEEE Transactions on Multimedia,* vol. 16, No. 2, 2014, pp. 541–558.

[28] Streijl, R.C., Winkler, S., Hands, D.S. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. *Multimedia Systems,* 2014*,* pp. 2–16.

[29] Xu J., Xing, L., Perkis, A., Jiang, Y. On the properties of mean opinion scores for quality of experience management. *Proceedings – 2011 IEEE International Symposium on Multimedia, ISM 2011*, Dana Point, CA, US, 2011.

[30] P.10/G.100. Vocabulary and effects of transmission parameters on customer opinion of transmission quality. ITU-T Rec., Geneva, Switzerland, 2006.

[31] G.107. The E-model: a computational model for use in transmission planning. ITU-T Rec., Geneva, Switzerland, 2015.

[32] G.1070. Opinion model for video-telephony applications. ITU-T Rec, Geneva, Switzerland, 2012.

[33] Moller, S. Quality Engineering: Qualitat Kommunikationstechnischer Systeme, Springer, 2010, p. 212.

[34] Ostermann, J., Bormans, J., List, P., Marpe, D., Narroschke, M., Pereira, F., Stockhammer, T., Wedi T. Video coding with H.264/AVC: Tools, performance, and complexity. *IEEE Circuits and Systems Magazine,* vol. 4, No. 1, 2004, pp. 7–28.

[35] Sullivan, G.J., Ohm, J.–R., Han. W.-J., Wiegand, T. Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 22, No. 12, 2012, pp. 1649–1668.

[36] Bossen, F., Bross, B., Suhring, K., Flynn, D. HEVC complexity and implementation analysis. *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 22, No. 12, 2012, pp. 1685–1696.

[37] Pauliks, R., Slaidins, I. Impact of Video Content and Technical Specifications on Subjective Quality Assessment. *Electronics and Electrical Engineering,* vol. 122, No. 6, 2012, pp. 91–96.

[38] Pauliks, R., Slaidins, I. Quality Evaluation of Synthetic Video in Simultaneous Double Stimulus Environment. *IEEE International Conference on Image Information Processing (ICIIP)*, 2013.

[39] Pauliks, R., Slaidins, I., Krauze., A., Tretjaks, K. Assessment of IP packet loss influence on perceptual quality of streaming video, 2015.

[40] Ma, S., Gao, W., Lu, Y. Rate–Distortion Analysis for H.264/AVC Video Coding and its Application to Rate Control. *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 15, No. 12, 2005, pp. 1533–1543.

[41] Lakshman, T.V., Ortega, A., Reibman, A.R. VBR Video: Tradeoffs and Potentials. *Proceedings of the IEEE,* vol. 85, No. 5, 1998, pp. 952–972.

[42] Vetro, A., Christopoulos, C., Sun, H. Video transcoding architectures and techniques: An overview. *IEEE Signal Processing Magazine,* vol. 20, No. 2, 2003, pp. 18–29.

[43] Gonick, L., Smith, W. *The Cartoon Guide to Statistics*. New York: HarperPerennial A Divission of Harper Collins Publishers, 1993.

[44] Spiegel, M., Stephens, L. *Schaums Outline of Statistics*, Fourth Edition (Schaum's Outline Series). McGraw-Hill; 4 edition, 2011, p. 600.

[45] Lin, C-H., Tainan, Ke, C-H., Shieh, C-K., Naeen C. The Packet Loss Effect on MPEG Video Transmission in Wireless Networks. *Advanced Information Networking and Applications, AINA*, Vienna, 2006.

[46] Joskowicz, J., Sotelo, R. A Model for Video Quality Assessment Considering Packet Loss for Broadcast Digital Television Coded in H.264. *International Journal of Digital Multimedia Broadcasting,* 2014, p. 11.

[47] Shihab, E., Wan, F., Lin, C., Gulliver, A., Tin, N. Performance Analysis of IPTV Traffic in Home Networks. *Global Telecommunications Conference, GLOBECOM '07*, Washington, DC, 2007.

[48] Zhang, Y., Duffield, N., Paxson, V. Shenker, S. On the Constancy of Internet Path Properties. *Proceedings of the ACM SIGCOMM Internet Measurement Workshop*, 2001.

[49] Win, M.Z., Pinto, P.C., Shepp, L.A. A mathematical theory of network interference and its applications. *Proceedings IEEE*, 2009.

[50] Cidon, I., Khamisy, A., Sidi, M. Analysis of packet loss processes in high-speed networks. *IEEE Transactions on Information Theory*, 1993.

[51] Paxson, V. End-to-end Internet packet dynamics. *EEE/ACM Transactions on Networking*, 1999.

[52] Yajnik, M., Moon, S., Kurose, J., Towsley, D. Measurement and modelling of the temporal dependence in packet loss. *Proceedings IEEE INFOCOM*, 1999.

[53] Beran, J., Sherman, R., Taqqu, M.S., Willinger, W. Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications,* vol. 43, No. 3, 1995, pp. 1566–1579.

[54] Kanumuri, S., Cosman, P.C., Vaishampayan, V.A., Reibman, A.R. Modeling Packet-Loss Visibility in MPEG-2 Video. *IEEE Transactions on Multimedia*, 2006.

[55] Liang, Y.J., Apostolopoulos, J.G., Girod, B. Analysis of packet loss for compressed video: Effect of burst losses and correlation between error frames. *IEEE Transactions on Circuits and Systems for Video Technology*, 2008.

[56] Lakshman, T.V., Madhow, U. The performance of TCP/IP for networks with high bandwidth-delay products and random loss. *IEEE/ACM Transactions on Networking*, 1997.

[57] Boyce, J.M., Gaglianello, R.D. Packet loss effects on MPEG video sent over the public Internet. *Proceedings of the 6th ACM International Conference on Multimedia, MULTIMEDIA*, 1998.

[58] Greengrass, J., Evans, J., Begen, A.C. Not all packets are equal, Part 2: The impact of network packet loss on video quality. *IEEE Internet Computing*, 2009.

[59] Mathis, M., Semke, J., Mahdavi, J., Ott, T. The macroscopic behavior of the TCP congestion avoidance algorithm. *Computer Communication Review*, 1997.

[60] Izquierdo, M.R., Reeves, D.S. Survey of statistical source models for variable-bit-rate compressed video. *Multimedia Systems*, 1999.

[61] Sen, P., Maglaris, B., Rikli, N.-E., Anastassiou, D. Models for packet switching of variable-bit-rate video sources. *IEEE Journal on Selected Areas in Communications*, 1989.

[62] Khan, A., Sun, L., Ifeachor, E. Content–Based Video Quality Prediction for MPEG4 Video Streaming over Wireless Networks. *Journal of Multimedia,* vol. 4, No. 4, 2009, pp. 228–239.

[63] Shmueli, R., Hadar, O., Huber, R., Maltz, M., Huber, M. Effects of an encoding scheme on perceived video quality transmitted over lossy internet protocol networks. *IEEE Transactions on Broadcasting*, 2008.

[64] Yamagishi, K., Hayashi, T. Parametric Packet–Layer Model for Monitoring Video Quality of IPTV Services. *IEEE International Conference on Communications 2008 (ICC 08)*, Beijing, 2008.

[65] P.911. Subjective audiovisual quality assessment methods for multimedia applications. ITU–T Rec, Geneva, Switzerland, 1998.

[66] P.800. Methods for subjective determination of transmission quality. ITU–T Rec, Geneva, Switzerland, 1996.

[67] Sommers, J., Barford, P., Duffield, N., Ron, A. A geometric approach to improving active packet loss measurement. *IEEE/ACM Transactions on Networking,* vol. 16, No. 2, 2008, pp. 307–320.

[68] Liang, Y.J., Apostolopoulos, J.G., Girod, B. Analysis of packet loss for compressed video: Effect of burst losses and correlation between error frames. *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 18, No. 7, 2008, pp. 861–874.

[69] MSU Graphics & Media Lab (Video Group), compression.ru, 25 August 2011. [Online]. Availabe at: http://www.compression.ru/video/quality_measure/subjective_metrics_info_en.html. [Accessed on 6 November 2013].

[70] Kozamernik, F., Steinmann, V., Paola, S., Wyckens, E. SAMVIQ—A New EBU Methodology for Video Quality Evaluations in Multimedia. *SMPTE Motion Imaging Journal,* 2005, pp. 152–160.

[71] G.107. The E-model: a computational model for use in transmission planning. ITU-T Rec., Geneva, Switzerland, 2014.

[72] Jumisko-Pyykkö, S., Malamal Vadakital, V. K., Hannuksela, M. M., Acceptance Threshold: A Bidimensional Research Method for User-Oriented Quality Evaluation Studies. *International Journal of Digital Multimedia Broadcasting,* 2008, p. 20.

[73] Ghinea, G., Thomas, J.P. QoS Impact on User Perception and Understanding of Multimedia Video Clips. *Proceedings of the 6th ACM International Conference on Multimedia*, Bristol, UK, 1998.

[74] Gulliver, S.R., Serif, T., Ghinea, G. Pervasive and standalone computing: the perceptual effects of variable multimedia quality. *International Journal of Human Computer-Studies,* vol. 60, No. 5–6, 2004, pp. 640–665.

[75] Watson, A.B., Kreslake, L. Measurement of visual impairment scales for digital video. *Proceedings of SPIE Human Vision and Electronic Imaging*, Los Altos, CA, 2001.

[76] Jia, Y., Lin, W., Kassim, A.A. Estimating just–noticeable distortion for video. *IEEE Transactions: Circuits and Systems for Video Technology,* vol. 16, No. 7, 2006, pp. 820–829.

[77] Tektronix/Sarnoff PQR Picture Quality Metric, Tektronix, 21 February 2015. [Online]. Availabe at: www.tektronix.com/PQA500. [Accessed on 21 February 2015].

[78] Ridder, H., Majoor, G.M. Numerical category scaling: an efficient method for assessing digital image coding impairments. *Proc. SPIE: Human Vision and Electronic Imaging: Models, Methods, and Applications,* B. E. Rogowitz and J. P. Allebach, Eds., Santa Clara, CA, 1990.

[79] Dijk, A.M., Martens, J.B. Perceptual image quality: concept and measurement. *Philips Journal of Research,* vol. 47, 1992, pp. 35–62.

[80] Dijk, A.M., Martens, J.B. Subjective quality assessment of compressed images. *Signal Processing,* vol. 58, 1997, pp. 235–252.

[81] Tominaga, T., Hayashi, T., Okamoto, J., Takahashi, A. Performance comparisons of subjective quality assessment methods for mobile video. *2nd International Workshop on Quality of Multimedia Experience, QoMEX 2010*, Trondheim, Norway, 2010.

[82] Rouse, D.M., Pepion, R., Le Callet, P., Hemami, S.S. Tradeoffs in subjective testing methods for image and video quality assessment. *Proceedings of SPIE – The International Society for Optical Engineering*, Nantes, France, 2010.

[83] Lee, C., Choi, C., Lee, E., Lee, S. Choe, J. Comparison of various subjective video quality assessment methods. *Image Quality and System Performance III*, San Jose, CA, 2006.

[84] Huynh-Thu, Q., Ghanbari, M. A comparison of subjective video quality assessment methods for low-bit rate and low-resolution video. *IASTED International Conference on Signal and Image Processing, SIP 2005*, Honolulu, United States, 2005.

[85] Keelan, B. *Handbook of Image Quality: Characterization and Prediction*. CRC Press, 2002, p. 544.

[86] Pinson, M., Wolf, S. Comparing subjective video quality testing methodologies. *Proceedings of SPIE – The International Society for Optical Engineering*, Lugano, Switzerland, 2003.

[87] Atkinson, R.C., Shiffrin, R.M. *The Control Process of Short Term Memory*. Standord University, Stanford, California, US, 1971.

[88] Peterson, L.R., Peterson, M.J. Short–term retention of individual verbal items. *Journal of Experimental Psychology,* vol. 58, No. 3, 1959, pp. 193–198.

[89] Miller, G. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review,* vol. 63, No. 2, 1956, pp. 81–97.

[90] Staelens, N., Moens, S., Van Den Broeck, W., Mariën, I., Vermeulen, B., Lambert, P., Van De Walle, R., Demeester, P. Assessing quality of experience of IPTV and video on demand services in real–life environments. *IEEE Transactions on Broadcasting,* vol. 56, No. 4, 2010, pp. 458–466.

[91] Winkler, S. On the properties of subjective ratings in video quality experiments. *International Workshop on Quality of Multimedia Experience, QoMEx 2009*, San Diego, CA, US, 2009.

[92] Quan, H.-T., Garcia, M.-N., Speranza, F., Corriveau, P., Raake, A. Study of rating scales for subjective quality assessment of high-definition video. *IEEE Transactions on Broadcasting,* vol. 57, No. 1, 2011, pp. 1–14.

[93] Rouse, D.M., Pepion, R., Le Callet, P., Hemami, S.S. Tradeoffs in subjective testing methods for image and video quality assessment. *Human Vision and Electronic Imaging XV*, San Jose, CA, 2010.

[94] Strohmeier, D., Jumisko-Pyykkö, S., Kunze, K. Open Profiling of Quality: A Mixed Method Approach to Understanding Multimodal Quality Perception. *Advances in Multimedia,* 2010, p. 28.

[95] De Koning, T.C.M., Veldhoven, P., Knoche, H., Kooij, R.E. Of MOS and men: bridging the gap between objective and subjective quality measurements in mobile TV. *SPIE*, 2007.

[96] Pauliks, R., Belahs, K.. Tretjaks, K. A Survey on Some Measurement Methods for Subjective Video Quality Assessment. *Journal of Multimedia Processing and Technologies (JMPT),* vol. 3, No. 2, 2013, pp. 113–123.

[97] Winkler, S. Analysis of public image and video databases for quality assessment. *IEEE Journal on Selected Topics in Signal Processing,* vol. 6, No. 6, 2012, pp. 616–625.

[98] Winkler, S., stefan.winkler.net, 5 June 2013. [Online]. Availabe at: http://stefan.winkler.net/resources.html#Video. [Accessed on 27 December 2013].

[99] Pinson, M.H., Sue Boyd, K., Hooker, J., Muntean, K. How To Choose Video Sequences For Video Quality Assessment. *7th International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Arizona, 2013.

[100] Pinson, M.H., Barkowsky, M., Le Callet, P. Selecting scenes for 2D and 3D subjective video quality tests. *EURASIP Journal on Image and Video Processing,* 2013, pp. 2–12.

[101] CDVL, The Consumer Digital Video Library, 13 March 2014. [Online]. Availabe at: http://www.cdvl.org. [Accessed on 13 March 2014].

[102] equasys, equasys GmbH, 13 March 2014. [Online]. Availabe at: http://download.equasys.de/downloadvideo.html. [Accessed on 21 January 2012].

[103] Stocker, A.A., Simoncelli, E.P. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience,* vol. 9, No. 4, 2006, pp. 578–585.

[104] Wang, Z., Li, Q. Video quality assessment using a statistical model of human visual speed perception. *The Journal of the Optical Society of America A (JOSA A),* vol. 24, No. 12, 2007, pp. B61–B69.

[105] Fischer, W. *Digital Video and Audio Broadcasting Technology*. Springer-Verlag Berlin Heidelberg: Springer, 2008.

[106] Wolf, S., Pinson, M.H. Spatial–temporal distortion metrics for in–service quality monitoring of any digital video system. *Proceedings of SPIE – The International Society for Optical Engineering*, Boston, MA, USA, 1999.

[107] Seshadrinathan, K., Bovik, A.C. Motion tuned spatio–temporal quality assessment of natural videos. *IEEE Transactions on Image Processing,* vol. 19, No. 2, 2010, pp. 335–350.

[108] Kortum, P., Sullivan, M. The effect of content desirability on subjective video quality ratings. vol. 52, No. 1, 2010, pp. 105–118.

[109] Pauliks, R., Slaidins, I. Impact of Video Content and Technical Specifications on Subjective Quality Assessment. *Electronics and Electrical Engineering,* vol. 122, No. 6, 2012, pp. 91–96.

[110] Pauliks, R., Slaidins, I. Quality Evaluation of Synthetic Video in Simultaneous Double Stimulus Environment. *IEEE International Conference on Image Information Processing (ICIIP)*, 2013.

[111] Deffner, G., Yuasa, M., McKeon, M., Arndt, D. Evaluation of Display-Image Quality: Experts vs. Non-Experts. *SID Digest,* 1994, pp. 475–478.

[112] Heynderickx, I.E., Soren, B. Image quality assessment by expert and non-expert viewers. *Human Vision and Electronic Imaging VII,* 2002.

[113] Eye charts, 10 October 2013. [Online]. Availabe at: http://www.psych.nyu.edu/pelli/charts.html. [Accessed on 21 May 2015].

[114] Pelli, D.G., Robson, J.G., Wilkins, A.J. The design of new letter chart for measuring contrast sensitivity. *Clin. Vision Sci.,* vol. 2, No. 3, 1987, pp. 187–199.

[115] Pauliks, R., Slaidins, I. Quality Evaluation of Synthetic Video in Simultaneous Double Stimulus Environment. *IEEE International Conference on Image Information Processing (ICIIP)*, 2013.

[116] Pauliks, R., Slaidins, I. Impact of Video Content and Technical Specifications on Subjective Quality Assessment. *Electronics and Electrical Engineering,* vol. 122, No. 6, 2012, pp. 91–96.

[117] Shahid, M., Rossholm, A., Lovstrom, B., Zepernik, H.-J. No-reference image and video quality assessment: a classification and review of recent approaches. *EURASIP Journal on Image and Video Processing 2014,* p. 32.