Prediction of Quality of User Experience for Video Streaming over IP Networks

Martin Kapa, Lucia Happe and Frantisek Jakab

***Abstract*—This paper deals with the video quality evaluation of mobile video delivery. It analyzes the current trends in video quality evaluation while trying to shift video quality measurement from service-centered to more user-centered. The attention is in defining the causalities among specific video and network parameters, and perceived video quality. As a part of this target, vast subjective video quality survey has been conducted to define these causalities. These causalities were particularly important in designing of our video quality prediction model based on Bayesian network. Proposed solution and prediction results were then experimentally evaluated compared to video quality results of independent subjective video quality assessment.**

***Index Terms*—QoE, QoS, Bayesian networks, video quality, prediction model.**

I. Introduction

In the current Information and Communication Technology (ICT) environment, technology serves as a synergistic part of our daily life. The evolution of the broadband and wireless communication enables the technology to dive even more into our day-to-day interaction with the world. The importance of the requirements regarding user expectations in any new product/service development is undeniable. Therefore, quality measurement techniques have to move towards user centered designs of assessment. In which the user expectations are predicted and considered during the whole phase of the development and the usage of a particular multimedia service. We also have to alter the view on the service quality assessment where we have to understand that gaining the knowledge about user experience is a journey rather than a destination. Needs and expectations are influenced by several factors, so what is needed is a continuous and synergic process. The overall steps should consist of several interaction runs with users. The idea of multiple interactions is supported by a range of authors [1] [2]. The process of gaining insight data can be divided into following phases such as prior-to-development and prior-to-launch, post-development and prior-to launch and post-development and post-launch. The goal is to gather enough insight information to accurately predict the expected user experience and to provide video quality framework that can be used as a mean for service providers to potentially measure how they meet the requirements defined in their service level agreements (SLA). Therefore, future success of any quality assessment technique lies in its accuracy to anticipate ever changing customers expectations. In order to achieve this, a multidisciplinary approach is called for, including both technical and user aspects, where the most important success factor is the optimal match between quality of experience (QoE) and quality of service (QoS). To succeed in this challenge, objective technical QoS metrics need to be strongly linked and correlated to more subjective QoE measures such as potential usability, user expectations and user experience. In order to achieve this goal we firstly need to consider which parameters have influence on degradation of user experience and quality of multimedia content. This has been partially done with signal-to-noise ration (SNR), peak signal-to-noise ration (PSNR) or bit error rate (BER). However, the measurements have shown that they do not correlate well with quality perceived by an end-user [3]. Therefore, concepts based on QoS and QoE [4] [5] have been introduced. However, most of the current approaches are oriented to one specific video content type, specific application or scenario, which is not enough. Video quality metrics need to be more cross-content to provide better correlation with subjective ratings that are really important for appropriate decisions on a suitable optimization method for video streaming. For example, todays methods of quality assessment do not accurately reflect the change in video parameters such as bite rate, video resolution, frame rate, codec, delay, jitter, etc. on resulted value of perceived quality. To measure these correlations, the future framework for video quality evaluation has to include subjective quality measurement methods [6] into its design.

II. Bayesian networks in quality assessment

Machine learning (ML) is considered as a subfield to artificial intelligence. This scientific discipline is dealing with the design process and development of various algorithms that allow computer systems to optimize theirs behaviors based on empirical data acquired through sensor data or databases. The particular learner is then able to take advantage of this data to capture and understand characteristics through probability distribution. In other words, data is used to observe the relationships between variables of our interest. The major task of machine learning is to be able automatically learn and recognize complex patterns and intelligently react on acquired data. In our solution, machine learning will play the most significant role in novel approach of understanding and defining of relationships among video, network parameters and resulted level of user's experience. These dependencies are going to be derived from data gained through sets of subjective video quality assessments. However, machine learning, like all other subjects in artificial intelligence, requires cross-disciplinary proficiency in areas such as probability theory, mathematic statistics, pattern recognition, data mining, adaptive control, computational neuroscience and theoretical computer science.

A. Fundamental knowledge about probability theory

The probability of seeing a particular outcome connected to a specific experiment can be describe as a relative frequency of seeing this particular outcome in all of the experiment performed [7]. The set of possible outcome is then referred as a *sample space (S)* of the experiment. Where the experiment is any process where the outcome is uncertain. It is assumed that sample space contains all the possible outcomes of an experiment, and that each outcome is mutually exclusive. This is an assurance that the experiment is guaranteed to end up in one of the specified outcomes. To measure a degree of uncertainty of an experiment, it is required to assign a probability *P(X)* to each outcome $X \subseteq S$. The probability of this outcome must be nonnegative and bellow or equal to 1.

Let's have two events; event A and event B. If two events A and B are disjoint, then the probability of the combined event is the sum of two individual events [7].

$$A \subseteq S, B \subseteq S \; and \; A \cap B = 0, \qquad (1)$$
$$then \; P(A \cup B) = P(A) + P(B)$$

On the other hand, when those events are not disjoint.

$$A \subseteq S, B \subseteq S \; and \; A \cap B \neq 0, \qquad (2)$$
$$then \; P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Conditional probability

When we are talking about a probability of an event, there are always given conditions on other factors. It can be illustrated on an example of calculating of the probability of the die turning 6. In this example it is an unsaid condition that it is a fair die, or it assumed that it is a fair die. In this manner, every statement on probabilities is a statement conditioned on what else is known. These types of probabilities are called conditional probabilities. The notation for the statement is following *P(A|B) = p* and should be read as, if B is true then the probability of A is p.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (3)$$

However, in the world of conditional probabilities we can also condition on more than one event, in which case the formula is generalized as:

$$P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} \qquad (4)$$

The initial formula for conditional probability can be rewritten to so-called *fundamental rule* for probability.

$$P(A|B)P(B) = P(A \cap B) \qquad (5)$$

Through fundamental rule we are able to measure the probability of seeing both A and B when we know the probability of A given B and probability of B.

Fundamental rule helps us to get to the ***Bayes' rule*** [7].

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (6)$$

Bayes' rule describes a method for updating our beliefs about an event A given that we get information about another event B. For this reason P(A) is called the prior probability, whereas *P(A|B)* is called the posterior probability of A given B and the probability *P(B|A)* is then called the likelihood of *A* given *B*.

Furthermore, as for the conditional probability we can also state Bayes' rule in terms of C [7].

$$P(A|B,C) = \frac{P(B|A,C)P(A|C)}{P(B|C)} \qquad (7)$$

Conditional independence

Is defining when events are independent by given evidence about another event. In following formula we see that events A and B are conditionally independent given the event C when [7]:

$$P(A|B \cap C) = P(A|C) \qquad (8)$$

B. Definition of Bayesian networks

Bayesian networks (BNs) are known as belief networks and they belong to the group of probabilistic graphical models [8]. These models are mostly used to represent knowledge about a specific uncertain domain. The graphical representation of BNs consists of nodes and each node in the graph resents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables.

A Bayesian network consists of following parts:

- Set of variables and set of directed edges between particular variables
- Each variable from the set has a finite number of mutually exclusive states
- The variables together with the directed edges create an acyclic directed graph (DAG), directed graph is acyclic when there is no directed path $A_1 \rightarrow \ldots \rightarrow A_n$ so that $A_1 = A_n$
- To each variable *A* with parents $B_1 \ldots B_n$ a conditional probability table $P(A|B_1, \ldots, B_n)$ is attached

Whenever a variable *A* has no parents, then the table of conditional probability is reduced to an unconditional probability table *P(A)*, which will be representing prior probabilities of variable *A*.

The definition of Bayesian network does not require that the links between variables represent causal impact of this connection. That means, when we are designing a model based on Bayesian network we do not have to insist on having the links in a causal direction. However, we have to check the model's *d-separation* properties whether they correspond with our perception of the conditional independence properties.

Most Common Usage of BNs:

- to model and explain a domain
- to update beliefs about states of certain variables

- to find the most probable configurations of variables
- to support decision making under uncertainty
- to find good strategies for solving tasks in a domain with uncertainty

Bayesian network can be considered as a complete model that consists of variables and relationships among them. Therefore, it is mostly used to answer probabilistic queries about the variables.

III. Proposed framework for subjective video quality assessment

The integration of human experience into measurement of video quality requires support of well-designed evaluation environment that assists researchers or network operators in gathering information facts about network quality, video parameters and user perception, and in adjusting the video delivery services to the required quality. Therefore, the evaluation tool must be based on a set of requirements allowing long-term measurements:

- The measurement tool must be an integrated part of video delivery solution, all of which results in the fact that end-users should not feel any influence of the tool during its usage.
- The measurement tool must accommodate measurements based on different kind of dimensions such as contextual, social, application, network and device.
- The measurement tool should support wide range of IP-based services.
- The measurement tool should be remotely manageable and should allow users in particular roles to automatically schedule a wide range of tests.

In (Fig. 1) a high level overview of the video quality measuring tool can be seen. It consists of the video delivery solutions on which the video quality is evaluated. The subjective assessment is done by the usage of handset-based approach [9], where the measurements are realized on the device, while the data is processed in the back-end. This architecture allows us

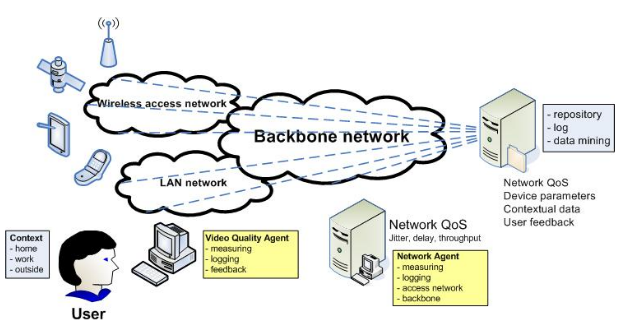

Fig. 1. Architecture of the video quality measuring tool

to measure and understand the influence of cross-contextual effects and the importance of different video and network parameters. The intention is to develop a basic algorithmic prediction model of overall quality. To achieve these goals there are two main approaches that can be used:

- Neural Network-based evaluation: In current research [10] [11] it is very popular to use neural networks to simulate and to solve problem in the field of video quality measurement. However, the high processing demand is one of the main drawbacks of this approach.
- Statistical model-base evaluation: Is another approach that is based on the statistical models and their following evaluation. In [12] it is described how the data is post-processed after quality evaluation and emphasize the importance of the statistical confidence during the measurements.

In our prediction model, the second approach was chosen because of its high correlation with the subjective perception of the end user on the overall video quality. Moreover, the usage of statistical models has been shown as understandable for future users and more importantly service providers for which requirements this tool is designed. To be able to create a statistical model for overall video quality prediction, it is firstly required to perform several series of subjective assessments that, in this framework, are done through *video quality agent*.

A. Implementation of the mobile agent

The central component of the subjective assessment is the mobile agent that is used and installed on the user's end-device. The resulting quality of any given video service is strongly dependant on a various range of factors, because the audiovisual data is processed by vast number of applications and the data delivery depends on the quality of a distributing channel. Typical phases in video delivery are following:

- Storing the input signal - the quality of particular video is strongly dependant on the conditions of the environment where the video is recorded (e.g. bad light, noise, quality of the recording device etc.)
- Coding of the input data stream - the compressing algorithm is processing audio and video data input by decreasing their data flow. On the client side is then a decompressor that transform the data back to the initial audio and video format. The resulting video quality is highly dependent on the usage of a particular compressing algorithm.
- Streaming - streaming server is sending a stream of audio and video data into distribution channel. Based on the hardware and software performance of the server the video quality is changed.
- Distribution - the quality of the distribution channel is the key characteristics to assure the quality of real-time video delivery. When the bandwidth is low or the network can not deliver data in requested manner, the video can be distorted or blurry and the sound can come out.
- Video playback - playback of the received video data is done by a specific application that receives the streamed flow of audiovisual data, decodes them and displays them on particular user interface. The performance of the application is also a factor that has high impact on the perceived video quality.

B. Selection of the investigated parameters

The testing environment is saving characteristics of every multimedia session used for testing purposes. Before any testing session can be initiated, multimedia session must be registered in the system and the system has to have all the required information. These characteristics are firmly set and system gets them from the multimedia file. In other words, the characteristics are video and audio parameters, which are indicated by a compression of the file. These parameters are helping service providers to identify particular sequences and estimate their requirements on the testing system, their structures and constraints. The gathering of the parameters is integrated in the system, all of which emphasis the higher lucidity, flexibility and simplicity of the system.

Implementation of the testing environment is using dynamic library *MediaInfo.dll* that is programmed in C++. This library is used as the lowest interface for gathering specific parameters of the multimedia file. It is compatible with various different programming languages, which are compatible with standard dynamic libraries (C++, C#, J#, Visual Basic, Delphi). The testing environment is using C++ for its high performance and simplicity.

Information that we can collect from the file of the video sequence and video playback application can be divided into following parts:

- General information - general characteristics about a multimedia file [Tab. I].

TABLE I
GENERAL INFORMATION ABOUT THE MULTIMEDIA FILE

Name of the parameter in API	Description
FileName	Name of the multimedia file
Format	Used format
FormatInfo	Information about used format
FormatProfile	Information about profile of the used format
OveralBitRate	Average bitrate (bit/s)
OverallBitRate_Minimum	Minimal value of the bitrate (bit/s)
OverallBitRate_Nominal	Nominal value of the bitrate (bit/s)
OverallBitRate_Maximum	Maximal value of the bitrate (bit/s)
Duration	Duration of the multimedia file
OverallBitRate_Mode	Bitrate mode (VBR - variable, CBR - constant)

- Video flow - general characteristics about every coded video flow [Tab. II].

TABLE II
INFORMATION ABOUT VIDEO PARAMETERS OF THE MULTIMEDIA FILE

Name of the parameter in API	Description
Format	Used video format
Format_Version	Version of the video format
Format_Profile	Profile of the video format
Format_Setting_Matrix	Format for the decoder
CodecID	ID of the used codec
Duration	Duration of the video sequence
BitRate_Mode	Bitrate mode - VBR or CBR
BitRate	Bitrate of the video sequence
BitRate_Minimum	Minimal value of the bitrate
BitRate_Nominal	Nominal value of the bitrate
BitRate_Maximum	Maximal value of the bitrate
Width	Width of the video sequence
Height	Height of the video sequence
DisplayAspectRation	Ratio of the displayed video sequence (No. x No.)
FrameRate_Mode	Framerate mode - VFR or CFR
FrameRate	Number of frames per second
FrameRate_Minimal	Minimal value of the framerate
FrameRate_Nominal	Nominal value of the framerate
FrameRate_Maximum	Maximal value of the framerate
Standard	Standard type (NTSC, PAL)
ColorSpace	Color system (RGB, CMYK, etc.)
ChromaSubsampling	(i:j:k)
BitDepth	Bit depth - 16/24/32 bit
ScanType	Type of the scan - progressive or overlaying

- Audio flow - general characteristics about every coded audio flow [Tab. III].

TABLE III
AUDIO INFORMATION ABOUT THE MULTIMEDIA FILE

Name of the parameter in API	Description
Format	Used format
FormatInfo	Information about used format
CodecID	ID of the used codec
Duration	Duration of the audio sequence
BitRate_Mode	Bitrate mode - VBR or CBR
BitRate	Bitrate of the sequence
Channels	Number of channels
SamplingRate	Sampling frequency (Hz)
BitDepth	Bit depth - 16/24/32 bit
StreamSize	Size of the stream (B)

C. Realization of the subjective video quality assessment

During the subjective video quality assessment a multimedia sequence, stored on the data server, is closely monitored. The actual measurement consists of mutually dependent steps that have strong influence on each other:

- Initial creation of a data stream - service provider initiate a data streaming based on specific protocol such as RTP or UDP. The following data stream is then delivered through the network using IP multicast. On the other hand, for the video-on-demand technologies, a service provider creates a separate data stream for every request, which is then delivered through HTTP protocol.
- Creation of a playback instance - playback from a perspective of subjective video quality assessment is a unique data stream of video sequence that is used for quality measurement.
- Construction of a user profile - for every playback a genuine user profile is created. It means that for every playback a specific hardware and software characteristics of the end user are gained, together with average bandwidth availability towards the streaming server.
- Sstatistical data collection - during playback a background process is ran, which gathers a statistical data about the playback in a real-time manner.
- Video quality evaluation - end user can in any given time evaluate the tested multimedia sequence. Three different aspects of the playback are evaluated - video quality, audio quality and playback quality.
- Storing statistical data to the database server - in a moment when the end-user finish his evaluation all gathered data together with the evaluation are sent to a remote database. After this step the evaluation process of a particular end-user is complete.
- Analyse of the statistical data - consists of displaying the gathered data in a manner that is understandable and valuable for the requestor of the video quality assessment.

IV. USER EXPERIENCE SENSITIVITY ANALYSIS

Studying video quality from the perception of an end-user is strongly dependent on the usage environment, streamed content and displayed size of the multimedia content. Therefore, scenarios of the video quality assessment on the currently popular mobile devices such as smartphones and tablets are strictly different compared to classical TV broadcasting and IPTV services. The contemporary solutions and recommendations in the field of quality assessment are focusing on high resolution video delivery. Based on that, one of the initial research activities was to design a methodology for subjective video assessment with the focus on mobile video delivery and its impact on the perceived level of video quality.

The other reason why to speak about video quality is that it is not always necessary to stream high-resolution video content, because of the low display quality, displayed size or the required user experience. As such, when lower user experience is required (that could be the case with certain devices or certain types of video such as more static videos with less movement) we can optimize the stream quality to fit these requirements and thus save some resources (e.g. capacity of the network connection) and transfer them to other streamed video content where high quality is very important (e.g. remote surgery consultancy as it is done for the surgeries in Africa).

A. *Test methodology*

The standards in subjective video quality assessment such as ITU-T P.910 and ITU-T P.911 define various methodologies for quality assessment. One of the key differences among these methods is their usage of reference video sequences. Moreover, non-reference methods such as ACR (Absolute Category Rating) and PC (Pair Comparison) are not suitable for evaluation of the transparency of the video system or its trueness. On the other hand, reference methods are able to evaluate these characteristics which are very often important factors in the evaluation of quality of the systems. Among the reference methods, which can be used in this manner, is DCR (Degradation Category Rating) that has been mostly used for the assessment of the video quality specifically in videotelephony and videoconferencing. The DCR's scale is mostly valued for its comments of imperceptible/perceptible impairments. Therefore, when the measurement of system trueness is an important factor, DCR should be used. On the other hand, when the easiness of implementation is an important factor, ACR is a good way to start. In comparison to ACR, the merit of the PC method is its high discriminatory power that is particularly valuable when the test items are nearly equal in quality. But one of the disadvantages of this method is its lengthiness of testing. In such case an ACR or DCR test should be carried out first with limited number of observers, followed by a PC test solely on those items which have received about the same rating.

For the subjective testing of the video streaming on the mobile devices, together with the instruction of the Video Quality Expert Group and Laboratory for Image & Video Engineering[1] [13] [14] following conditions were defined:

- To simulate the real-life environment, viewers do not have access to the reference video sequence; therefore, a reference-free method of subjective evaluation was chosen.
- The test sequences were displayed on the mobile devices such as smartphones and tablets (Fig. 2 and 3).
- Several content types of video sequences were defined and used during the subjective testing.

[1]http://live.ece.utexas.edu/

Fig. 2. Illustrative picture of the smartphone

Fig. 3. Illustrative picture of the tablet

B. *Evaluation of the video quality*

The test methodology used to gain extensive subjective data is based on ITU-T P.910 [6]. For the purposes of this survey the most suitable method defined by ITU-T Recommendation was chosen. ACR is a reference-free method that is also called Single Stimulus Method where the test sequences are presented in a continuous form and rated independently one at a time. By implementing it, the real world scenario become a model, where the customers of mobile video services do not have access to original video sequences. However, ACR introduces higher variance in the acquired results, compared to other methods, mainly because the results do not only depend on the quality of a particular test sequence, but also on other factors such as the mental state or the quality of the test conditions. A description of recommended reference conditions and the ways how to produce them is described in Recommendation P.930 [15]. Particular recommendation advises to use LCD monitors as a displaying interface towards the user during the subjective assessments. However, the contemporary state of development in mobile devices offers a large amount of choices in the hardware and software used for delivering multimedia content. This range of standardized decoders, players and devices, puts the assessment of video quality for mobile devices into totally different perspective compare to the standard broadband video services where the system parameters do not vary so much. All of which makes it much harder to evaluate the final quality of multimedia content compared to other video services.

TABLE IV
TEST ENVIRONMENT

Viewing Distance	Viewing Angle	Illumination
20-30 cm	0	$\leq$ 20 lux

The test environment used during the subjective evaluation within this work is closely described in [Tab. IV]. It fulfills all the requirements set by ITU-T P.910. After each presentation the test subject was asked to evaluate the resulting quality of experience of the test sequence. To measure the perceived quality accurately, a subjective scaling method was required. However, this rating method is only meaningful when there is actual correlation between characteristics of the video sequence and experience that it causes to the user. Therefore, as a grading method we used mean opinion score (MOS). Five grade MOS scale, on one side, is well known to the test subjects because they use similar grading scale at school, and on the other side, it provides a good interpretability of the results.

C. Subjective testing of the video quality

The particular testing can be divided into three main parts. Firstly, each test starts with a trial run where three untested sequences are presented. The main intention is to offer every subject an initial understanding and experiences with subjective quality evaluation. The trial runs are not taken into account of the final test results. Afterwards, the test sequences are presented in a manner that two clips with the same content, though differently distorted, cannot appear after each other. This rule was used to diminish the possibility that the subjects would use the degradation rating instead of the absolute one. Duration of every of the test sequences is about 10 seconds. The inequality of the length is because the sequences were adjusted to keep the content consistent. The final phase is the actual voting. The length of the assigned voting time is set to 10 seconds as well. The fluency in the testing is assured by usage of the playlist that ensures that the subject will not interact with the device and will be fully concentrated on the actual testing. The usage of a playlist also assures the homogeneity of the viewing conditions among all the test subjects, because the presentations cannot be stopped and the voting time is fixed.

D. Content classes of the source materials

The source test sequences were formatted into two different resolutions - QCIF and CIF. The content of the video sequences differs with the intention to use the most frequent contents to define the impact of them on the user perception.

Within this paper following six content classes were defined:

1) **Content Class News/Debate [CC1]** - this content class includes test sequences with a small moving region with mostly a static background. The movement within this content class is mostly generated in the Region of Interests (ROI) by the moderator and his/her eyes, mouth and facial movements.
2) **Content Class Music Video Clip [CC2]** - this content class contains test sequences that cover a lot of global and local motion or fast scene changes.
3) **Content Class Sport Match [CC3]** - this content class includes test sequences with uniform camera movement where the camera is tracking the movement of players and a ball, while the background is mainly single colored (green).
4) **Content Class Action Scenes [CC4]** - in this content class object motion and changes in the scenes is dominant, with a lot of local and global movement.
5) **Content Class Cartoon [CC5]** - this content type includes test sequences where the background is usually static and the movement of the objects is dominant. However, the movement object has no natural character.
6) **Content Class Panorama [CC6]** - in this content type the test sequences contains wide angle panorama picture of an area. The camera movement is uniform and in only a single direction at a time.

To better understand the connection between a perception quality and the content of the video sequence, the spatial and temporal information is used to characterize a video sequence. The Spatial Information (SI) is reflecting the complexity of still images. SI measurement is based on the Sobel filter, that is applied to each luminance frame F_n at time instance *n*. Afterwards, the standard deviation over the pixel is computed and the maximum value throughout the whose sequence represents the spatial information.

$$SI = \max_{time_n} \left\{ std_{space_{i,j}} \left[Sobel \left(F_n \left(i,j \right) \right) \right] \right\} \tag{9}$$

On the other hand, temporal perceptual information is based on the motion changes among the pictures. Therefore, for every time instance *n*, the luminance pixel values difference has to be calculated.

$$M_n \left(i,j \right) = F_n \left(i,j \right) - F_{n-1} \left(i,j \right) \tag{10}$$

Temporal Information (TI) is then calculated as a maximum over time of the standard deviation over space.

$$TI = \max_{time_n} \left\{ std_{space_{i,j}} \left[M_n \left(i,j \right) \right] \right\} \tag{11}$$

In (Fig. 4) it is presented a comparison between SI and TI values of various test sequences based on the pre-defined content classes.

E. Results of the subjective video quality assessments

The subjective assessment of the video quality focused on the usage of current highly appreciated portable devices such as smart-phones and tablets. From the high level, the tests could be divided into two main groups based on the resolution of particular test sequences. These test sequences were encoded in H.264/AVC codec. During the testing we worked with 22 students from Technical University of Kosice, Slovakia. The age of the chosen subjects ranged from 22 to 24, with various levels of knowledge and experience in

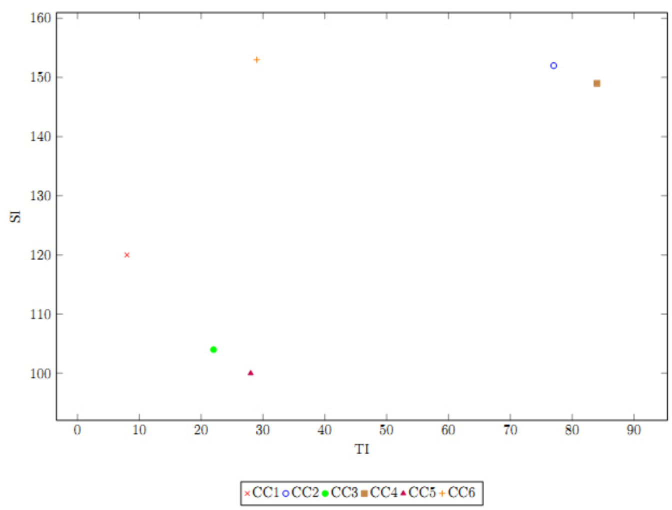

Fig. 4. Comparison of the spatial and temporal features of the test sequences

video quality evaluation. The main goal of the testing was to study impact of various video parameters on a perceived video quality. One of the most significant outcomes of the testing was the confirmation that human visual perception of video quality is strongly determined by the actual character of the observed sequence. It can be seen on (Fig. 5 and 6) that the measured subjective video quality is strongly content dependent especially at low bit rate and resolution.

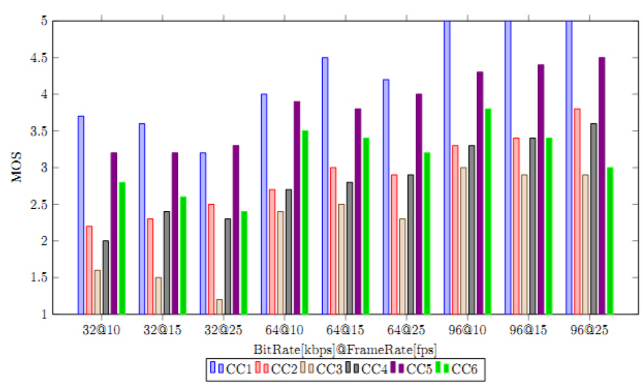

Fig. 5. Summary of the subjective tests for QCIF resolution

The content dependence is visible even in the grade differences that differed in up to 3.2 MOS grades for the QCIF resolution and up to 2 MOS grades for the CIF resolution. Moreover, the subjective assessment showed that the bit rate of 96kbps and frame rate of 10fps is probably one of the most moderate settings for the QCIF resolution. Besides, it can be seen that lower frame rates do not always result in the decrease of the subjective quality at the QCIF resolution. In this resolution the CC with the lowest MOS grades was the CC3. It was explained in previous section that it was mainly because this class requires high level of detail on objects such as ball, players and lines that define the borders of the pitch. On the other hand, CC1 was the class that obtained the highest grades mainly because there wasn't much movement in the sequences except the changes in the facial expressions of the moderator.

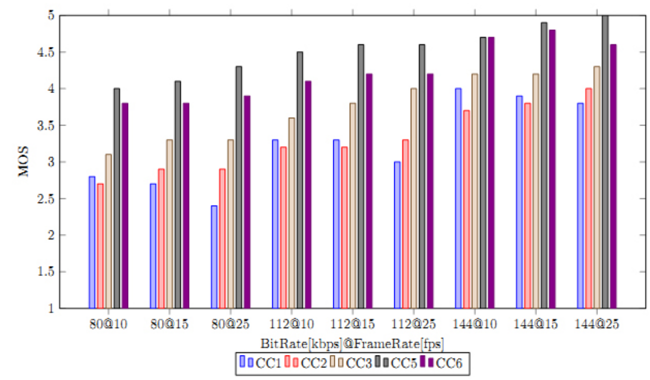

Fig. 6. Summary of the subjective tests for CIF resolution

The subjective assessment of the CIF resolution shows similar behavior as for the QCIF resolution. What is interesting in the results is the impact of higher bit rates and resolution on the perceived quality of the CC3, which obtained the highest increase compared to the results for QCIF resolution. Increase in the bit rate and resolution has always a positive effect on the perceived quality of sport sequences, which is in contrast with other more static types such as CC1. It can be deduced that in the sport sequences viewers prefer smoothness of the motion rather than static quality. Moreover, through the results of this quality survey it possible to better understands the various aspects that have impact on the perceived quality and by this understanding provide estimation for service providers which coding parameters should be used.

V. Proposed Prediction Model Based on Bayesian Networks

The quality perception is mostly considered as something with stochastic nature, what motivates us to treat overall quality and its attributes as probability distribution. One of the concepts we can use is Bayesian theory that provides natural tools for modeling and analyzing of hypothesis under uncertainty. Moreover, it is an attractive tool because it is represented by a probabilistic model that contains set of variables and their conditional interdependencies with a directed graph. The intention of this paper is to come up with a unified model representing and explaining the concept of overall quality and also a more practical tool that produces a single quality value defining the video quality of a streamed video sequence.

A. Main structure and variables

Bayesian networks are used for a decision making under uncertainty. Where the network is represented by nodes (*V - variable*) and edges (*E - oriented causality*) that together result in directed acyclic graph *G = (V, E)*. The first step in constructing of Bayesian network is to select the nodes that are representing random variables. Moreover, each node is represented by its states that, in this case of discrete nodes, are defined by a truth table. The basic structure of our Bayesian network is displayed on (Fig. 7), where the left side represents the objective variables that are measured and are the external input into the model. On the other hand, the right side is the overall quality that defines the output of the Bayesian network in the form of a probability distribution within the range of possible values.

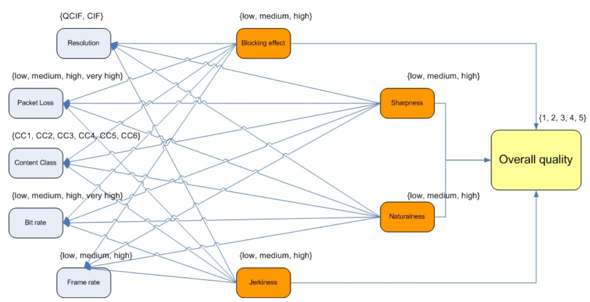

Fig. 7. QoE prediction model based on Bayesian network

Objective parameters:

The parameters that were used are based on the recommendations [13] [14] and were part of the subjective assessments survey described in previous section. In Bayesian networks every parameter/variable is represented by its states. Following table [Tab. V] defines names of the objective parameters and their states.

TABLE V
OBJECTIVE PARAMETERS USED IN BAYESIAN NETWORK

Name	States
Resolution	QCIF, CIF
Packet loss	low(<0.5%), medium(0.5-2.0%), high(2.0-4.0%), very high(>4.0%)
Bit rate	low(32-63kbps), medium(64-95kbps), high(96-111kpps), very high(112-144kpps)
Frame rate	low(10fps), medium(15fps), high(25fps)
Content class	CC1, CC2, CC3, CC4, CC5, CC6

Subjective parameters:

The parameters were chosen after the first run of subjective experiments, where by a questionaire were collected aspects that the viewers considered disturbing and had impact on their decision regarding the perceived video quality. Following table [Tab. VI] defines the names of the subjective parameters and their states.

TABLE VI
SUBJECTIVE PARAMETERS USED IN BAYESIAN NETWORK

Name	States
Blocking effect	low, medium, high
Sharpness	low, medium, high
Naturalness	low, medium, high
Jerkiness	low, medium, high

After defining the variables that are used in the Bayesian network, the relation or the edges among the variables of the Bayesian network had to be defined. This can be done manually, or the relationships can be learned by discovering techniques.

B. Training and using of subjective assessments

After defining the particular artifacts that are representing variables in Bayesian network, we had to define causal relationships among them. This was done through understanding the dependencies of the state changes and causal directions. Therefore, structured subjective assessments needed to take place. From high level perspective, the goal of the subjective assessments was to define the impact of changes in various video parameters and video content types on the overall quality perceived by end-user. Besides collecting the required data about bit rate, frame rate etc., the observers were asked about subjective attributes that formed his or her decision regarding the overall quality of the test sequence. In this work, multi-dimensional arrays were used to represent the conditional probability distribution. However, these arrays are only suitable when the data are discrete. The subjective parameters were discrete, but the value of objective parameters needed to be discretized into particular states of each parameter [Tab. V]. After collecting all the required data, the Bayesian network could be constructed and straightforwardly trained using the maximum likelihood parameter estimation. Due to the possibility that the acquired amount of data did not sufficiently describe the overall video quality, a *Laplace correction* was used to eliminate the potential problems. The *Laplace correction* is a technique that is used to deal with zero probability values or states that do not have represented outcome. There is a simple trick to avoid this problem. It can be assumed that the training set is so large that adding one value to each count that we need would make a negligible difference in the estimated probabilities.

There are several heuristic methods for structure learning; however, the gathering of subjective data in this case limits the availability of the training data. Therefore, techniques such as PC algorithm are not applicable. To solve this problem and to define a structure of video quality model, we formed a number of hypotheses that were supported by the subjective assessments that form the base of the causalities within the model itself. For example if the bit rate increate while the other parameters remain the same the quality should increase. On the other hand, when the packet loss increases, while the remaining parameters stay the same, the quality should be degraded. Using these heuristic hypotheses, it is possible to create a logically correct model, and test it by randomly pruning how well model behavior follows the hypotheses. What is even more important is to understand that the network should not be evaluated base on how well it represents the training data, but how well it represents the prior knowledge after the estimation with the training data. The prior knowledge acts as a regulator which enables the optimization process with a quite small number of data points. Within this paper following hypotheses [Tab. VII] were defined. They were tested by computing their marginal distribution of overall quality with selected values of objective parameters, gained through subjective assessments, that are consider as evidence followed by a change in single objective parameter and evaluating the change in marginal distribution. The outcome, or overall quality, is represented by an estimate of the mean opinion score (MOS).

TABLE VII
HYPOTHESES USED TO OPTIMISE THE STRUCTURE OF THE BAYESIAN NETWORK

Hypotheses	↑ of a variable	Effect on overall quality
1.	Packet loss	quality decrease
2.	Resolution	depends on a particular CC
3.	Bit rate	quality increase
4.	Frame rate	generally quality increase
5.	Content class	depends on a particular CC
6.	Blocking effect	quality decrease
7.	Sharpness	quality increase
8.	Naturalness	quality increase
9.	Jerkiness	quality decrease

In order to be able to construct the causality in the Bayesian network it was important to understand that when we are talking about relations in a directed graph, what causal network is, the usage of wording *family relations* is right in place. If there is a link from *A* to *B*, it can be said that *B* is a child of *A*, and *A* is a parent of *B*. There are three main types of connections between variables in Bayesian network. Choosing a connection type have a significant impact on the way how the rest of the variables are influenced by knowing a state of particular variable in the Bayesian network.

C. Causality evaluation and optimization

Testing hypotheses described in [Tab. VII] is very time consuming and testing all prior hypotheses with all possible input combinations is considered as illusive. It was advised to rather use random comparison that provides a statistical significance of results in the desirable area. This can be done by the help of *Central limit theorem.*

Let n represents a number of tests per hypothesis and X_i the result of the test (zero if the hypothesis is not supported and one if it is).

$$p_n^* = \sum_{i=1}^{n} \frac{X_i}{n} \tag{12}$$

The (Eq. 12) represents the number of tests that follow the hypothesis; the high of this value determines how the model is following the hypothesis. Now, let (Eq. 13) represents the positive value of p_n^* after all the possible combinations of inputs were tested.

$$p = E[X_i] \tag{13}$$

Lets select the sample size n (Eq. 14), where the p_n^* should not differ from p by more than ϵ with the probability P_ϵ.

$$P\left(|p_n^* - p| \leq \epsilon\right) \geq P_\epsilon \tag{14}$$

The *central limit theorem* defines p_n^* as:

$$p_n^* \approx N\left(p, \frac{p\,(1-p)}{n}\right) \tag{15}$$

$$P\left(|Z| \leq \frac{\epsilon}{\sqrt{\frac{p(1-p)}{n}}}\right) \geq P_\epsilon \tag{16}$$

It is clear that the sample size is dependent on the value of *p* and *p(p-1)*. The highest standard deviation of this normal distribution is when *p=1/2*. The sample size or number of tests per hypothesis is then calculated and its values are displayed in [Tab. VIII].

TABLE VIII
THE NUMBER OF TESTS PER HYPOTHESIS TO BE PERFORMED

P_ϵ	ϵ = 0.01	ϵ = 0.02	ϵ = 0.3	ϵ = 0.04	ϵ = 0.05	ϵ = 0.1
0.85	5181	1295	576	324	207	52
0.90	6764	1691	752	423	271	68
0.95	9604	2401	1067	600	384	96
0.99	16587	4417	1843	1037	663	166

The outcome of these runs is the confirmation that the initial hypotheses were correct and are able to become stepping stones for the establishment of particular causalities among variables of the Bayesian network.

D. Representation of the Bayesian network

Bayesian network illustrated in (Fig. 7) can be easily divided into three main layers - *IM - instrumental measures, SA - subjective attributes* and *OQ - overall quality*. Next, a joint probability is used to define the likelihood of numerous events occurring together and at the same time.

$$\begin{aligned} P(X_{IM}, X_{SA}, X_{OQ}) &= P(X_{SA}) \\ &\times P(X_{IM}|X_{SA}).P(X_{OQ}|X_{SA}) \end{aligned} \tag{17}$$

The joint probability of the proposed model is then defined in (Eq. 17). The variables within particular layers of the network are then represented by X_{IM}, X_{SA} and X_{OQ}.

- P(X_{SA}) - prior probability of X_{SA}
- P($X_{IM}|X_{SA}$) - conditional probability of X_{IM} given that X_{SA} has occurred
- P($X_{OQ}|X_{SA}$) - conditional probability of X_{OQ} given that X_{SA} has occurred

To describe particular conditional probability when an evidence on instrumental measurements is obtained, a Bayes formula has to be used, where the denominator *P(B)* was, by the usage of marginalization, derived to:

$$P(B) = \sum_{A} P(B|A)P(A) \tag{18}$$

therefore, the Bayes formula has changed to

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)} \tag{19}$$

Lets illustrate it on one variable within our Bayesian network, i.e. $X_{IM} = X^e{}_{IM}$:

$$P^e(X_{SA}|X^e_{IM}) = \frac{P(X^e_{IM}|X_{SA}).P(X_{SA})}{P(X^e_I M)} = \frac{P(X^e_{IM}|X_{SA}).P(X_{SA})}{\sum_{X_{SA}} P(X^e_{IM}|X_{SA}).P(X_{SA})} \quad (20)$$

The complexity of computing of particular conditional probabilities in Bayesian network is defined by the number of variables, their states and states of their parents. Moreover, in our case, to measure a conditional probability of particular state of the overall quality based on particular instrumental measures, it is required to define conditional probability distribution derived from Bayes formula:

$$P^e(X_{OQ}|X^e_{IM}) = \sum_{X_{SA}} P(X_{OQ}|X_{SA}).P^e(X_{SA}|X^e_{IM}) \quad (21)$$

Now we can move from layer description of the Bayesian network to actual description of prior and conditional probabilities of particular variables in each layer. The various states of *instrumental measures* and *subjective attributes* were defined in tables [Tab. V and VI]. Overal quality is defined by particular *MOS* values ranging from one to five.

The high level of the network was presented in (Fig. 7), where the states are assigned to particular variables and are clearly illustrated. The next step is to define the probability distributions of particular variables. The first layer we take is the *subjective attributes* layer. Variables in this layer, from Bayesian network point of view, do not have any parents; therefore, probability of these variables is called prior probability and is represented by a uniform distribution where all the states within a variable has equal chance of occurrence [Tab. IX].

TABLE IX
ILLUSTRATIVE CONDITIONAL PROBABILITY TABLE OF VARIABLE IN SUBJECTIVE ATTRIBUTES LAYER

Blocking Effect	P(Blocking Effect)
low	θlow
medium	θmedium
high	θhigh

On the other hand, the variables in instrumental measures layer are dependent on the state of variables in subjective attributes layer. Here it is required to calculate the conditional probabilities of variable in instrumental measures given the state of its parents or variables in subjective attributes layer. The complexity of the conditional probability tables is then calculated by (Eq. 22) where *M* is the number of states of the variable *X* and *N* is the number of states per parent *pa(X)*. Complexity represents the number of conditional probabilities that needs to be calculated per variable *X*.

$$Complexity(X) = \sum_{pa(X)} M.N \quad (22)$$

The illustrative conditional probability table of the variable in instrumental measures layer is displayed in [Tab. X]. Where blocking effect was abbreviated to BE, sharpness to SH, naturalness to NT and jerkiness to JK.

TABLE X
ILLUSTRATIVE CONDITIONAL PROBABILITY TABLE OF VARIABLE IN INSTRUMENTAL MEASURES LAYER

Resolution BE SH NT JK	P(Resolution|BE, SH, NT, JK)
QCIF,BE_{low},SH_{low},NT_{low},JK_{low}	θQCIF|BE_{low},SH_{low},NT_{low},JK_{low}
QCIF,BE_{low},SH_{low},NT_{low},JK_{medium}	θQCIF|BE_{low},SH_{low},NT_{low},JK_{medium}
QCIF,BE_{low},SH_{low},NT_{medium},JK_{low}	θQCIF|BE_{low},SH_{low},NT_{medium},JK_{low}
QCIF,BE_{low},SH_{medium},NT_{low},JK_{low}	θQCIF|BE_{low},SH_{medium},NT_{low},JK_{low}
QCIF,BE_{medium},SH_{low},NT_{low},JK_{low}	θQCIF|BE_{medium},SH_{low},NT_{low},JK_{low}
QCIF,BE_{medium},SH_{low},NT_{low},JK_{medium}	θQCIF|BE_{medium},SH_{low},NT_{low},JK_{medium}
⋮	⋮
CIF,BE_{high},SH_{high},NT_{medium},JK_{medium}	θQCIF|BE_{high},SH_{high},NT_{medium},JK_{medium}
CIF,BE_{high},SH_{high},NT_{high},JK_{medium}	θQCIF|BE_{high},SH_{high},NT_{high},JK_{medium}
CIF,BE_{high},SH_{high},NT_{high},JK_{high}	θQCIF|BE_{high},SH_{high},NT_{high},JK_{high}

Similarly, the *overall quality* layer is represented by the conditional probability of occurrence of one of the states of overall quality based on the knowledge about variables in subjective attributes layer. Therefore, it is required to calculate the conditional probabilities of overall quality given the state of its parents. These probabilities are then defined in the conditional probability table, which complexity is calculated by (Eq. 22).

TABLE XI
ILLUSTRATIVE CONDITIONAL PROBABILITY TABLE OF VARIABLE IN OVERALL QUALITY LAYER

Overall Quality BE SH NT JK	P(Overall Quality|BE, SH, NT, JK)
1,BE_{low},SH_{low},NT_{low},JK_{low}	$\theta 1$|BE_{low},SH_{low},NT_{low},JK_{low}
1,BE_{low},SH_{low},NT_{low},JK_{medium}	$\theta 1$|BE_{low},SH_{low},NT_{low},JK_{medium}
1,BE_{low},SH_{low},NT_{medium},JK_{low}	$\theta 1$|BE_{low},SH_{low},NT_{medium},JK_{low}
1,BE_{low},SH_{medium},NT_{low},JK_{low}	$\theta 1$|BE_{low},SH_{medium},NT_{low},JK_{low}
1,BE_{medium},SH_{low},NT_{low},JK_{low}	$\theta 1$|BE_{medium},SH_{low},NT_{low},JK_{low}
1,BE_{medium},SH_{low},NT_{low},JK_{medium}	$\theta 1$|BE_{medium},SH_{low},NT_{low},JK_{medium}
⋮	⋮
5,BE_{high},SH_{high},NT_{medium},JK_{medium}	$\theta 5$|BE_{high},SH_{high},NT_{medium},JK_{medium}
5,BE_{high},SH_{high},NT_{high},JK_{medium}	$\theta 5$|BE_{high},SH_{high},NT_{high},JK_{medium}
5,BE_{high},SH_{high},NT_{high},JK_{high}	$\theta 5$|BE_{high},SH_{high},NT_{high},JK_{high}

The particular conditional probabilities in tables [Tab. X and XI] were derived from the results of subjective assessments described in previous section. The prior probabilities in tables IX are in uniform distribution; therefore, they are equal.

VI. VALIDATION OF THE PREDICTION MODEL

The validation is based on two main goals: (i) to test the prediction results in comparison to results of an independent subjective video quality survey and to confirm that the model could be implemented as a video quality prediction tool, and (ii) to evaluate usability of the model. The validation is then structured based on these goals. These goals are evaluated based on different levels of validations for prediction models that are defined in [16]. In the first goal, we validate several important aspects such as the accuracy of the prediction model using results of an independent subjective video quality survey done by Laboratory for Image & Video Engineering[2] *Video Quality Expert Group (VQEG)* [13] [14] in comparison to actual prediction results of our analytical prediction model.

[2]http://live.ece.utexas.edu/

For the second goal, we are evaluating the extension of the model and describing a process of adding future subjective metrics and simplicity of the usage of the model from an end-user perspective.

A. Validation Type I: Accuracy Validation

The first level of validation [16] is comparing the prediction results of the analytical prediction model to the measured results of the quality survey done by VQEG. The studied property of the prediction approach is the accuracy of the prediction.

In the case of prediction model that focuses on user experience, additional aspects are important. The prediction approach is required to: deliver more accurate predictions with the usage of model based on understanding of subjective aspects that have impact on resulting level of video quality as without.

Type I: Prediction Accuracy

The goal of the validation was to evaluate whether the relationships among the measure attributes, subjective attributes and overall video quality were well defined and calibrated based on the subjective video quality assessment. The question we needed to answer for this purpose is:

Assumption 1: Supporting accurate video quality predictions by the completed model.

To validate the accuracy of the video quality prediction, we compared the results of our video quality prediction model and results of real measurements done by VQEG. The detailed description and results of the validation can be found in Section VII . In our studies, we demonstrated that the usage of Bayesian network is a meaningful concept for video quality prediction. However, we do not claim that our model is transferable to all other platforms, due to specific characteristics of each of these platforms. Nevertheless, this validation can be repeated for new platforms and metrics, and the model can be recalibrated.

Type I: Scalability of the prediction model

In this step of the evaluation, we evaluated the future possibilities of development of our prediction model in regard of adding of objective and subjective parameters and defining new causal relationships among them.

Assumption 2: Supporting extendibility by the completed prediction model.

By evaluating the *Assumption 1* it is possible to say whether the model is implementable for the platform of user experience evaluation. When it is, then by adding new variables we will increase the diversity of the platform where this model could be used. Within this validation we also algorithmically describe the process of adding new variable into the model with steps that needs to be completed.

B. Validation Type II: Applicability Validation

The second level of validation is trying to address the applicability and usability of the prediction approach based on Bayesian networks. The validation of applicability assesses the information that is required to apply the approach, to create the prediction model, to calibrate the model, to execute the model based on real-time data and to interpret the results.

Assumption 3: Supporting various levels of knowledge complexity based on particular roles in a usage of the completed model.

From the high level perspective, we can define two main roles - *administrator* of the model and its *user*. Administrator is someone who is responsible for definition of the prediction model, setting the range of variables and dependencies among them, and calculation of the conditional probabilities. This person has to have a vast knowledge about statistics and probability, Bayesian networks and about subjective video quality assessment. On the other hand user - who could be a service provider - is someone that is using this model for prediction of the video quality based on real-time instrumental measures in a computer network. Compare to administrator, this person does not need to have the same amount of qualification mainly because the expert-knowledge is hidden from the end-user within the actual model. Where the user sees it as only a model with particular inputs (real-time values of the instrumental measures) and output, which is the value of overall video quality.

The usage of Bayesian networks in this prediction model create a model, where from the end-user perspective the execution is fully automated and it represents an approach with minimal effort to get the resulting predicting value of video quality. Moreover, the representation of overall quality through MOS represents an intuitive grading scheme similarly used in various educational institutions.

C. Validation Type III: Cost/Benefit Validation

The last level of validation is named benefit validation and is focusing on the cost/benefit evaluation of a prediction method. In this type of evaluation the costs that results from usage of particular prediction approach are compared to the expected benefits, which - in this case - can be an improvement of video quality while decreasing the hardware requirements. The most common benefit of all prediction approaches is the reduction of effort in later development phases of particular approaches.

To validate them on this level, a controlled experiment is required during which whole quality assessment have to be executed - on the service delivery side - with and without using of the presented approach. Such validation is highly time and effort consuming and thus is rarely executed in practice. Due to the high effort, we cannot conduct this type of validation in scope of this thesis.

VII. Prediction accuracy validation

The main purpose of the prediction approach in the overall video quality prediction is to help service providers to understand the dependencies and causalities between measurable artifacts in the video delivery systems and resulting video quality. In this section, we present the validation settings for the prediction model and results for the validation goals specified in the previous section.

A. Process of validation

First question we are going to address is regarding the prediction accuracy:

Q1: Can completed model provide accurate video quality predictions?

As it was mentioned, the prediction approach based on Bayesian networks is defining the relationship between instrumental measures and overall quality through understanding of subjective attributes. The model consists of variables defining these various attributes and their states. The calibration of the causalities among them was done by calculation of conditional probabilities based on subjective video quality assessments described in Section IV. To evaluate whether the causalities were defined correctly and whether the prediction are with good accuracy it was desirable to validate the model based on independent quality assessment data, which were not included in the initial calibration of the model.

Therefore, in the validation we used the results collected within a *LIVE Video Quality Database* that contains the results of assessments conducted in co-operation with *VQEG*. The *LIVE Video Quality Datatbase* consists of ten uncompressed high-quality videos with a wide variety of content types. The testing was concluded by changing various video setting that created a set of 150 distorted videos where the settings varied in frame-rate, resolution, bit-rate and content type. All the video sequences were encoded in H.264 and MPEG-2 video codec. Each video in the database was then assessed by 38 human subjects in a single stimulus study with hidden reference video sequence. The results were then interpreted by the usage of MOS grading scheme where the resulting value was an average value of all the subjective evaluation runs of particular video sequence and its settings.

B. Results of the accuracy validation

As it was mentioned earlier in this section a validation of the model's accuracy is done by comparison of the predicted results derived from the model and results of quality assessment survey done by VQEG. The results are going to be presented in graphs (Fig. 8, 9 and 10) where one axis represents the predicted MOS values and the other one represents the MOS value collected in the VQEG's survey. The level of accuracy could then be illustrated in the distance of particular points from the graph representation of the linear function *f(x)=x*.

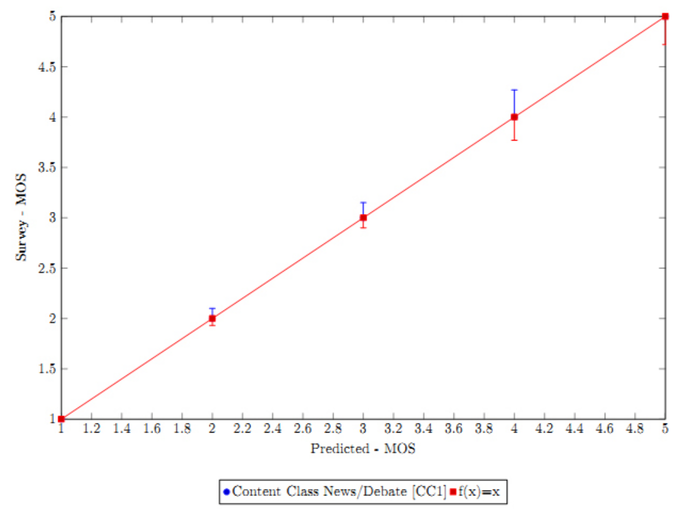

Fig. 8. Results of the validation of CC1

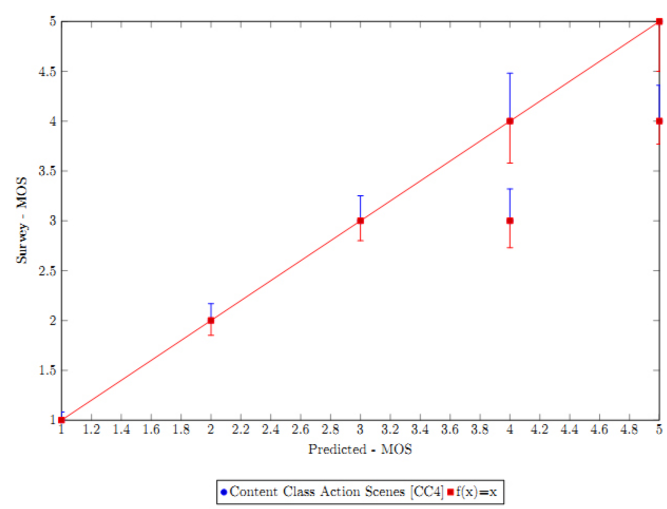

Fig. 9. Results of the validation of CC4

The outcome of the model that we are validating is represented by the state of the variable named overall quality. The states of this variable are mapping the various grades in the MOS, going from one to five. The final result of the overall quality is a state with the highest probability calculated based on conditional probabilities of its parents and entered values of instrumental measures. However, the results of the survey are represented as an average value of all runs - per video and per setting. Therefore, the validation described on is represented through error graphs where the level of inaccuracy between predicted and measured MOS values is clearly visible.

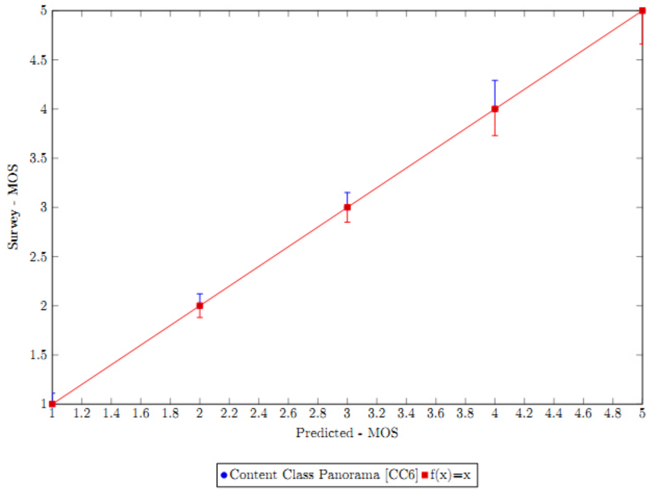

Fig. 10. Results of the validation of CC6

The predicted results showed high level of accuracy. However, the results also showed that the inaccuracy is highly dependent on the video content type described through particular content class. The highest inaccuracy resulted in CC4, which contains videos with the highest level of movements and changes in spatial representation. The error significantly increased with the surge in quality values where the predicted value differentiated in *δ=1.23* points compared to the measured value of MOS. The lowest inaccuracy was visible in CC1 where the error was *δ=0.28* points in the MOS scale. All of which showed that much vaster survey in the area of the subjective video quality assessment has to conducted for content classes with higher level of scene changes to predict the quality values more accurately. When it is not, then the model, in this thesis, tends to be much more positive towards the overall quality values of particular video sequence. Nevertheless, even though this model is not 100% accurate it still offers a very accurate prediction of perceived video quality.

VIII. Usability validation

In this section we are going to validate the usability of the prediction model. We illustrate the various roles that emerge from the usage of this approach and describe their specifics. Firstly, we address question *Q3* regarding the complexity of the model usage:

Q3: What is the complexity of the usage of the prediction model based on Bayesian networks?

By using Bayesian network as a base concept of prediction in our prediction model we are talking about statistical model-base evaluation of the video quality, where the statistical confidence during measurements is highly important. From high-level perspective we can define two main roles:

- End-user
- Administrator

End user is a person that uses the prediction model for predicting level of video quality defined by MOS. This person could represent a service provider or a researcher who is trying to evaluate its video delivery solution based on the instrumental measurements of video and network parameters. The advantage of this approach is that the end user does not need to be knowledgeable in areas of Bayesian networks, statistics and probability because from his or her perspective the concept is a black-box with inputs defined through values of instrumental measures and output is overall video quality represented by grading values of MOS. The hidden expert knowledge in this concept could be considered as one of the key aspects in the simplicity and high usability of this solution.

Administrator on the other hand is someone who is responsible for collecting subjective video quality assessments, defining and correcting the model, and future development of the model design. This is not something that a general end user should or could do; therefore, it is required that this position would be filled by someone highly knowledgeable. The person is expected to have a vast knowledge about the area of video quality assessment together video knowledge about various statistical concepts and models such as Bayesian networks. To assure that the model is providing accurate results it is required from administrator to perform regular subjective video quality surveys and accommodate the model design based on these results.

IX. Conclusion

The presented approach is motivated by the shift from a service-centered to much more user-centered development of new services in the area of multimedia delivery solutions. Developers of future multimedia solutions, have to understand contemporary trends in evaluation of user experience. In order to achieve this goal they have to focus more on modeling and assessing of user experience. The view on user experience assessment techniques has dramatically changed. Current approaches do not provide accurate results, all of which is caused by inadequate designing processes and misunderstanding of user experience. Therefore, to understand user experience we conducted several runs of subjective video quality assessment and the data we collected were the stepping stones for our prediction model based on Bayesian networks. The evaluation of the model showed high level of accuracy; however, the understanding of user experience is rather a journey then a destination so the evolution of this model will not stop and we have to continue to evaluate any new parameters and characteristics that have potential impact on perceived video quality in the future.

Acknowledgment

Paper is the result of the Project implementation: Competency Centre for Knowledge technologies applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] G. Bixler, "Extreme user centered design: Methodology for eliciting and ranking requirements in user-centered new product development: Case studies from honduras and the central african republic," in *Global Humanitarian Technology Conference (GHTC)*, Nov 2011.

[2] S. Humayoun, Y. Dubinsky, and T. Catarci, "Ueman: A tool to manage user evaluation in development environments," in *Software Engineering, 2009. ICSE 2009. IEEE 31st International Conference*, May 2009.

[3] S. Winkler, *Digital Video Quality: Vision Models and Metrics*. London: John Wiley & Sons, 2005.

[4] D. Soldani, M. Li, and R. Cuny, *QoS and QoE Management in UMTS Cellular Systems*. London: John Wiley & Sons, 2006.

[5] F. Pereira, "Sensations, perceptions and emotions towards quality of experience evaluation for consumer electronics video adaptations," in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, January 2005.

[6] ITU, *P.910 - Subjective video quality assessment methods for multimedia applications*, International Telecommunication Union, Sep. 1999.

[7] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*. New York: Springer, 2007.

[8] J. Guerin and J. Goldsmith, "Constructing a dynamic bayes net model of academic advising," in *Proceedings of the Uncertainty in Artificial Intelligence (UAI 2011) Workshop on Bayesian Modeling Applications (BMAW 2011)*, Jul 2011.

[9] H. T. Verkasalo, "Handset-based analysis of mobile service usage," Master's thesis, Helsinki University of Technology, 2009.

[10] H. Du, C. Guo, Y. Liu, and Y. Liu, "Research on relationship between qoe and qos based on bp neural network," in *Network Infrastructure and Digital Content, 2009. IC-NIDC 2009*, 2009.

[11] P. Kumar, V. Sehgal, and D. Chauhan, "Performance evaluation of decision tree versus artificial neural network based classifiers in diversity of datasets," in *Information and Communication Technologies (WICT)*, 2011.

[12] A. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms," in *Circuits and Systems for Video Technology*, April 2010.

[13] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," in *IEEE Transactions on Image Processing*, 2010.

[14] ——, "A subjective study to evaluate video quality assessment algorithms," in *SPIE Proceedings Human Vision and Electronic Imaging*, 2010.

[15] ITU, *P.930 - Principles of a reference impairment system for video*, International Telecommunication Union, 1996.

[16] R. Bohme and R. Reussner, "Validation of predictions with measurements," *Dependability Metrics*, vol. 4909, pp. 7–13, 2008.

Martin Kapa is a PhD candidate in the field of multimedia services and QoE evaluation at Technical University of Kosice, Slovakia. In May 2009 he graduated from a postgraduate program at the department of Computers and Informatics on the Faculty of Electrical Engineering and Informatics at Technical University in Kosice, Slovakia. His interests include video quality evaluation, video delivery and measurement of QoE.

Lucia Happe is a Senior Researcher at the Karlsruhe Institute of Informatik (KIT) in the group of Software Design and Quality since November 2011. She received her PhD in computer science in November 2011 from the KIT. In September 2008, she was awarded a PhD scholarship from the Deutscher Akademischer Austauschdienst (DAAD). She got her diploma in computer science in May 2006 from the University of Kosice in Slovakia. Her interests include software architectures, model-driven development methods, and model-based quality prediction.

Frantisek Jakab was born in 1959. He received the MSc. degree in Systemotechnic engineering from the St. Petersburg Electrotechnical Institute (Russia) in 1984 and the PhD. degree in 2005. He is employed as an assistant professor at the Dept. of Computers and Informatics, Technical university of Koice, Slovakia. He is a head of the Computer Engineering Group and Computer Networks Laboratory. His research interests include projecting of computer network, modelling, simulation and network management, new form of multimedia-based communication, QoS, telelearning systems, intelligent tutoring systems. He has been a coordinator of several large international e-learning oriented projects supported by EC. He is a coordinator of the Cisco Networking Academy Program for the Slovak Republic and head of the Application Section of the Communication Technology Forum Association in Slovak Republic.