Paper

QoS Conditions for VoIP and VoD

Przemysław Dymarski, Sławomir Kula, and Thanh Nguyen Huy

Institute of Telecommunications, Warsaw University of Technology, Warsaw, Poland

Abstract—This paper concerns quality evaluation of the telecommunication services: VoIP (representing the RT interactive class) and VoD (representing the MM streaming class). Subjective and objective methods and tools for perceived quality measurement are analyzed and compared. Subjective tests are performed for selected video sequences using the Double-Stimulus Impairment Scale (DSIS) method. Thus the objective algorithms (VQM and VQmon) are calibrated. Speech quality is measured using the objective methods: PESQ and POLQA. Threshold values for network parameters (packet loss rate, delay jitter) are set, that guarantee the acceptable service quality.

Keywords—delay jitter, packet loss rate, PESQ, POLQA, quality of service, VoD, VoIP, VQM, VQmon.

1. Introduction

Quality of telecommunication services grows in importance not only from user point of view. The services providers and operators take into account quality as an element of competition. Services can be delivered using different networks, protocols, devices etc. Quality of service depends on many factors, like the network transmission parameters such as throughput (bandwidth), bit error rate (BER), packet loss rate (PLR), delay, and delay jitter. Influence of these parameters on quality depends on a service. Thus, different end-to-end classes of service were introduced [1]. The examples of such classes are the real time (RT) services like VoIP or videoconference, the multimedia streaming like VoD or IPTV, the high throughput data like FTP and the standard services like email. The requirements concerning network parameters for each class of service were specified in the ITU-T and ETSI recommendations [2], [3], [4]. In DiffServ network [5] the threshold values of network parameters have to be defined, which guarantee the acceptable quality perceived by the end user. Differentiated services enable a scalable service discrimination and the potential users who may violate these threshold values are rejected.

In this paper two examples of telecommunication services are considered, namely the VoIP (representing the class of real time applications) and VoD (representing the class of multimedia streaming) – both accessed by the IP network. Acceptable quality of these services can be achieved by setting appropriate threshold values of network transmission parameters. According to our observations ([6], [7]) the requirements specified in [2], [4] do not always reflect user's preferences. Therefore we start with an analysis of norms and tools for video signal and speech signal quality mea-

surement (Subsections 2.1, 3.1). Subsections 2.3 and 3.1 are dedicated to a calibration of the selected video quality metrics, and setting up credibility conditions for speech quality metrics. Setup of the laboratory stands and methodology of testing the influence of the IP network transmission parameters on speech and video quality are presented in subsections 2.4 and 3.2. Results of these tests as well as the proposed threshold values of transmission parameters are discussed in Section 4.

2. QoS in Multimedia Streaming

2.1. Recommendations and Tools for Video Quality Evaluation

The metrics of video signal quality should reflect the opinion of the end user. Therefore the subjective tests are more credible than the objective quality measures. On the other hand, the subjective tests are more difficult to conduct, they involve a group of participants, require the special acoustic conditions, they are more costly and time-consuming. The subjective tests are described, e.g., in the ITU-T Recommendation P.910 [8] and the ITU-R Recommendation BT.500-12 [9]. In our work they are used to calibrate the selected objective measures.

We have applied the Double-Stimulus Impairment Scale (DSIS), due to relative simplicity of this test. Each participant observes first the reference video sequence and then the sequence to be assessed. Then he/she evaluates the loss of quality according to the mean opinion score (MOS) scale from 1 to 5 (where 5 – no degradation, 1 – unacceptable level of distortions).

The objective quality measures may use the reference video sequence (media based methods), the incoming packets (on line quality evaluation) or the information concerning network structure, codecs etc. (parametric methods). The media based methods may be divided into the following groups.

- The full reference methods (also called the intrusive methods) use the whole reference video sequence and compare it with the tested sequence.

- The reduced reference methods use only some parameters of the original video sequence.

- The no-reference methods (also called the non-intrusive methods) have no access to the original sequence.
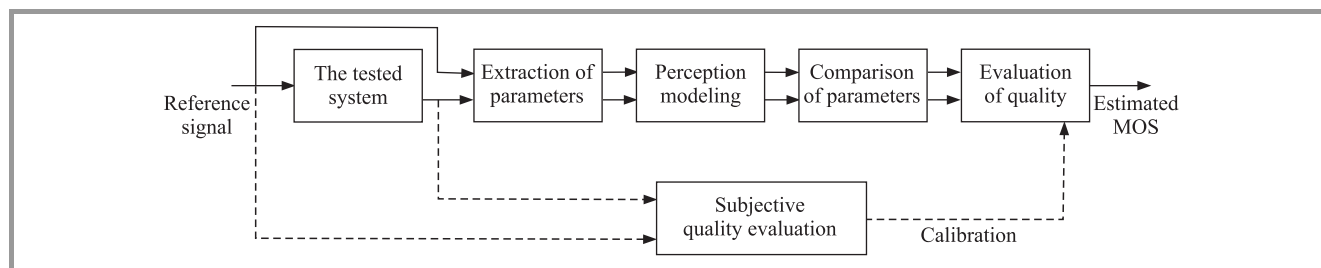
Fig. 1. General scheme of the objective full reference quality evaluation with calibration based on subjective tests.

The full reference quality evaluation is the most credible one – the selected algorithms of this kind are recommended by the ITU-T [10], [11]. The ITU-T Recommendation J.144 [10] presents a series of quality evaluation algorithms without pointing the best one. All the algorithms of this kind may have a similar structure shown in Fig. 1.

The algorithms described in this recommendation may be used for testing the video signals of a relatively high quality, e.g., the cable TV at the bit rates from 768 kbit/s to 5 Mbit/s. These algorithms were not thoroughly tested in presence of the channel errors (e.g., lost packets), therefore they are not recommended to quality evaluation of video sequences transmitted through the channels of low quality. Nevertheless, we have applied one of the J.144 algorithms, namely the Video Quality Metric (VQM), to test the VoD quality. This was possible due to the calibration described in the Subsection 2.3.

The VQM has been proposed by the Institute for Telecommunication Science (ITS), collaborating with the National Telecommunications and Information Administration (NTIA) [12]. It is based on the simplified human visual system model, particularly the spacial and temporal contrast perception.

In order to improve the accuracy and widen the application range of the objective video quality evaluation algorithms, the ITU-T started a new competition, in which the following institutions have taken part: NTT, OPTICOM, Psytechnics, Yonsei University and SwissQual. Finally, the ITU-T proposed:

1. As the full reference methods, recommend four algorithms: NTT, OPTICOM, Psytechnics and Yonsei University. These algorithms are described in the norm J.247 [11].

2. As the reduced reference method, recommend the Yonsei University algorithm. It is described in the norm J.246 [13].

3. Not recommend any of the no-reference methods despite of the relatively good results obtained by SwissQual.

The above mentioned algorithms may be used for evaluation of quality of video signals transmitted through channels of a low quality (packet loss, delay jitter etc.). They have sophisticated synchronization tools for alignment of both video sequences: the reference one and the tested one. The spacial alignment makes it possible to compare the cropped images with the full size images. After the temporal and spacial alignment a series of parameters is extracted from both sequences: they concern luminance, chrominance, edges, block effects etc. The human visual system models are used to compare these parameters in order to calculate the final quality measure using the MOS scale. The algorithm proposed by the Yonsei University is mainly based on the edges processing, therefore it does not require the full reference video sequence, only some information about edges (1 kbit/s do 128 kbit/s, depending on the video sequence). That is why it has been recommended as a reduced reference algorithm.

Unfortunately we had no access to the J.247 and J.246 algorithms, so we have decided to calibrate the VQM, being a part of the J.144 norm.

For the on line quality control, the no-reference methods are useful, particularly the methods based on the IP packets analysis. These algorithms use information concerning the lost packets (some of them identify the coder and analyze the influence of the lost packets on the image quality), the corrupted packets, delay jitter etc. An example of such algorithm is the VQmon/HD distributed by the Telchemy Inc. [14]. The VQmon/HD is used for monitoring of the IPTV, videoconference and VoD quality. It supports the RTP and UDP protocols and many video and audio coding schemes. Each packet is identified as the audio or video I, B or P packet and its influence on the audio/video quality is estimated. The following quality measures are calculated: MOS-A (audio), MOS-V (video) and MOS-AV (aggregated audio and video). The video quality metrics (MOS-V) are evaluated in a relative (mainly the transmission quality is considered) and absolute (not only the transmission but also codec parameters are considered) form. Moreover the instantaneous and average MOS values are delivered. Further comments concerning the VQmon/HD quality metrics will be presented in Subsection 2.3.

For the network planning purposes the parametric quality evaluation algorithms are used. They consider the codec parameters (coding scheme, bit rate) and the parameters of the communication link (bandwidth, packet loss rate, delay, delay jitter) and do not require any measurements. For telephony the E-model (ITU-T Recommendation G.107) and for multimedia the ITU-T Recommendation G.1070 is used.

2.2. Test Procedure to Determine Threshold Values of Transmission Parameters

Our purpose was to determine threshold values of transmission parameters adequate for acceptable perceptual quality of multimedia streaming. It could be done using existing networks and subjective methods of quality evaluation. Unfortunately, such procedure is complicated, very laborious, and time consuming. Because of this we have used a network emulator and objective quality evaluation methods. To test the influence of transmission parameters on perceptual quality of multimedia streaming we have decided to proceed as follows:

1. Calibrate of selected metrics for objective measurements by using the subjective tests.

2. Emulate network and perform multimedia streaming.

3. Using objective methods evaluate perceptual quality of perceived multimedia (video sequence with accompanying audio).

4. Determine threshold values of transmission parameters yielding the acceptable quality.

These steps will be described in the following subsections.

2.3. Calibration of Metrics for Objective Measurements

Because of unavailability of the attested software of the J.247 algorithms [11] we decided to use two objective tests, namely PSNR and VQM (the latter being a part of the J.144 norm). However, the calibration procedure was necessary, in order to express the quality estimates in a MOS scale. The calibration was performed in the following steps:

1. Selection of video material.

2. Using the network emulator (Netem) and streaming application (VLC) for preparation of the distorted video sequences.

3. Performing of the objective tests using the PSNR and VQM quality metrics.

4. Installing the appropriate video display software (MSU video quality measurement tool [15]) on six PCs and preparation of the quality evaluation tasks.

5. Selection of subjects (viewers who evaluate quality of video sequences).

6. Performing of the tests using the DSIS method.

7. Conversion of the objective quality measures to the MOS scale using the results of the objective tests.

As a test material five sequences, MPEG-2 – coded, of size 640×480, and bit rates form about 700 kbit/s to 7000 kbit/s

were selected. Laboratory stand for video on demand quality monitoring consisted of two workstations with VLC application to stream and receive video sequences. VLC media player [16] is a free of charge application, which supports various audio and video codecs (MPEG-1, MPEG-2, MPEG-4, DiviX, MP3, OGG Vorbis etc.) and transmission protocols like UDP and RTP. On the receiver workstation the Telchemy's VQmon [14] application (for on-line video evaluation) and video quality measurement tools [15] included VQM and PSNR metrics were installed.

Network and its parameters were emulated by Netem [17] which is a part of Linux system. In our work Netem was used to change the following transmission parameters: bit error rate, packet loss rate, bit rate, and delay jitter.

Subjects were 60 students of the Electronic and Information Technology Faculty (Warsaw University of Technology). Short instruction was given to subjects on arrival for their first visit in the laboratory. The subjects were informed on ideas of the subjective method and the objective methods, and on the measurement procedure.

The subjective method based on DSIS [8], [9] was used. In DSIS subjects watch two video sequences, the original one and the transmitted one. Subjects evaluate quality using the MOS scale recommended by ITU. The scale is given in Table 1.

Table 1
Viewing quality scale for DSIS method

MOS	Quality loss
5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Conversion of PSNR objective metric to MOS is based on finding a proper approximation function. It is relatively easy to find such function under assumption that the function is linear with the minimum value 1 and the maximum value 5. In Fig. 2 the subjective MOS values versus
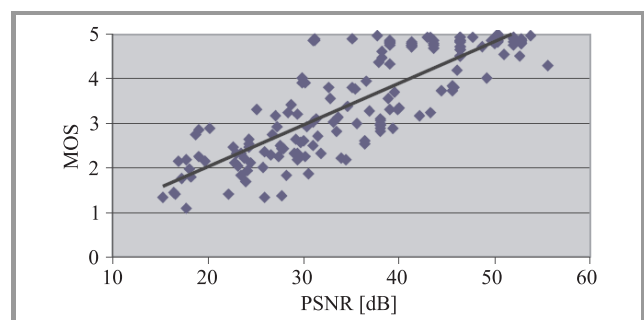

Fig. 2. Subjective MOS versus PSNR.

the measured PSNR [dB] values are given. Note that any point in this figure is an average of scores obtained by many subjects. Indeed, data presented in Fig. 2 suggest the linear

approximation. Using the least mean squares approach the following conversion function is obtained:

$$PSNR\_MOS = 0.0935\,PSNR + 0.152\,. \qquad (1)$$

In the case of VQM finding of an approximation function is more difficult because of nonlinearity. In Fig. 3 the subjective MOS versus the measured VQM values are given. Data presented in Fig. 3 suggest the logarithmic approximation curve. Using the least mean squares approach the following conversion function is obtained:

$$VQM\_MOS = -0.8634\,\ln(VQM) + 3.4854\,, \qquad (2)$$
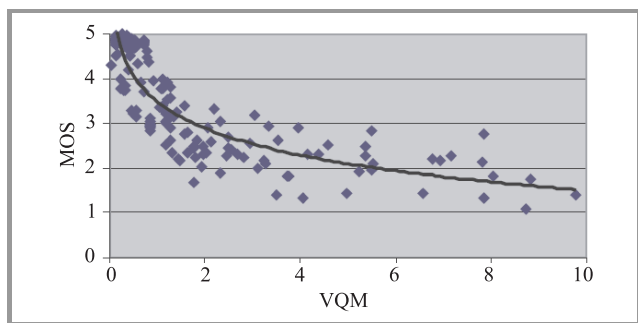
where ln – natural logarithm.

Fig. 3. Subjective MOS versus VQM.

Which estimate, PSNR_MOS or VQM_MOS, is better? In Fig. 4 and Fig. 5 are presented comparisons of the objective estimates: PSNR_MOS and VQM_MOS with the subjective MOS. The linear mapping ($y = x$) is also shown. To answer this question Pearson's correlation was calculated.
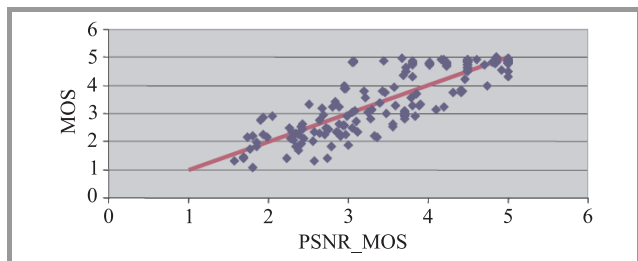

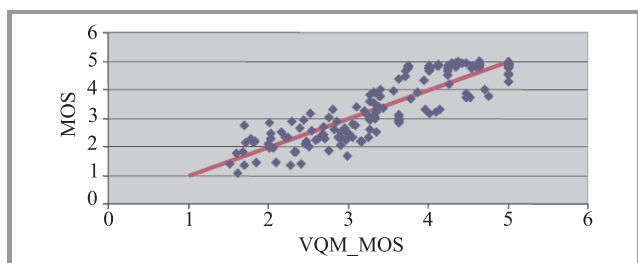**Fig. 4.** Subjective MOS versus PSNR_MOS.

Fig. 5. Subjective MOS versus VQM_MOS.

Calculation of Pearson's correlation needs centering of the sets x_i (e.g., PSNR_MOS values) and y_i (subjective MOS values) – thus the centered data $\hat{x}_i$ and $\hat{y}_i$ are obtained. Then, the correlation is calculated:

$$R_{xy} = \frac{\sum_i \hat{x}_i \hat{y}_i}{\sqrt{\sum_i (\hat{x}_i)^2 \; \sum_i (\hat{y}_i)^2}}\,. \qquad (3)$$

The Pearson's correlation values for PSNR_MOS and VQM_MOS equal 0.849 and 0.883, respectively, so better MOS approximation is obtained with the VQM_MOS.

For the on-line quality assessment, we find the VQmon distributed by the Telchemy Inc. [14] very useful. However, some measures must be taken, to obtain stable and credible results. VQmon delivers packets of results regularly, at a time interval set up by the user. In order to obtain stable results at low packet loss rates, longer measurement intervals should be used. Despite of this, there is sometimes an initial unstable phase, i.e., the first packets of results show lower MOS-V values than the subsequent ones. This concerns not only the instantaneous MOS-V values, but also the averaged values. We have used the measurement intervals of 30 s and we have ignored the initial packets of results, so we have obtained the credible results in most cases.

If a delay jitter causes a drop of quality, the MOS-V values may not reflect the image quality, because it depends on the size of the receiving buffer. VQmon is not informed about the buffer size, because in analyzes the incoming packets before buffering.

In case of corrupted (but not lost) packets, the quality drops, but MOS-V values are high, suggesting a good quality. This is probably because the VQmon analyzes mainly the packet headers, and is not sensitive to corruption od data.

2.4. Influence of IP Network Parameters on Video Signal Quality

For testing the influence of the network parameters on video signal quality, the same laboratory stand as for calibration of metrics was used, i.e., two workstations with VLC application and MSU video quality measurement tools [15] with VQM metric and a server with Netem network simulator. However the number, variety and length of video sequences were much higher. The tested sequences were divided into six categories. Their features are given in Table 2.

Each sequence was transmitted by network emulator. Transmission parameters (BER, PLR, channel throughput, delay jitter) were changed gradually. Quality of video sequence was measured using VQM for each transmission parameter separately. Results were converted to VQM_MOS and averaged for all the sequences – the example of is shown in Fig. 6. Negative influences on video were not being observed for the threshold values of transmission parameters. In Table 3 the obtained threshold values are given. These thresholds are the "worst case" values – they do not stem directly from the average results (like those presented in Fig. 6), but guarantee (according to our tests) lack of distortions in all of the observed video sequences. According to our results a service provider should guarantee that the transmission parameters are below (BER,

Table 2
Categories of tested video sequences

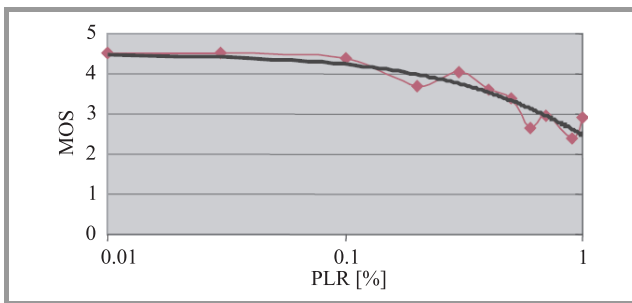| Cat.	Description	Content	Audio
A	Speaking people	Small and slow changes in picture	Speech
B	Publicity-graphic	Animated cartoons	Music and speech
C	Landscape	Slow movement of camera	Different
D	Pop music	Video clips	Song and music
E	News-reports	Speaker and short videos	Speech, background music
F	Sport	Dynamic changes of picture	Speech, background noise

Fig. 6. The average VQM_MOS versus PLR – measured and approximated values.

Table 3
Threshold values for transmission parameters

Parameter	Threshold
BER	0.03%
PLR	0.06%
Throughput	max bit rate*
Delay jitter	0.05 ms
* max. instantaneous video sequence bit rate.	

PLR, jitter) or above (throughput) threshold values given in Table 3. Note that the ITU-T and ETSI Recommendations [1], [2] set the PLR threshold at 1%, which, in our opinion, is too high. However, the Recommendation Y.1541 [3] defines the provisional QoS classes demanding PLR $< 10^{-5}$, but the base threshold is still at 1%. The Recommendation J.241 defines quality levels for videostreaming services. If the PLR is greater than 0.02% quality is referred as poor, the excellent quality is obtained for PLR $< 0.001\%$. Our result (PLR$<0.06\%$) is somewhat less strict, but we agree that some margin should be used, in order to guarantee very good quality of videostreaming services. For the xDSL networks much more strict conditions are formulated, e.g., PLR$< 10^{-6}$ for SD and PLR$< 10^{-7}$ for HD video transmission [18].

3. QoS in Voice over IP

3.1. Analysis of the Objective Measures of Speech Quality

For quality assessment of the telecommunication services based on speech transmission the media based methods are mainly used. The most popular full reference algorithm, Perceptual Evaluation Of Speech Quality (PESQ) is described in the ITU-T Recommendation P.862 [19]. The algorithm has access to the original speech phrase and the processed one. At the first stage time-domain synchronization of both phrases is accomplished. Then a series of speech parameters, which influence the human perception, are extracted from both signals. These parameters are defined in frequency domain (human ear is not sensitive to phase of the audio signal), using nonuniform scale (thus modeling the basilar membrane in the ear). Then the psychoacoustic representations of both signals are compared, using a human perception model (psychoacoustic model). Mainly the masking phenomena in time and frequency domain are considered in such model. The aggregated result of this comparison, called the Raw MOS, takes values from –0.5 (a big difference of both signals, suggesting a completely unacceptable quality of the tested phrase) to 4.5 (no perceptible difference between both phrases). At last, the Raw MOS is converted to the listening quality MOS (MOS-LQO) which takes values from 1.02 to 4.56 and maximizes correlation with the results of subjective tests.

In order to increase credibility of PESQ MOS-LQO values, ITU has specified conditions in which the measurements have to be performed. These conditions are described in the Recommendation P.862.3 [20]:

- Recommended phrase duration is $8-12$ s, accepted $3.2-30$ s, in any case it should not exceed million samples.

- In order to reduce the influence of speaker on the quality assessment results, phrases from 2 feminine and 2 masculine speakers should be used.

- Pure speech signal should take $40\%-80\%$ of the whole phrase (the rest contains initial, inter-word, and final silence), there should be at least 3.2 s of active speech.

- The initial and final silence should last from 0.5 to 2 s. In both phrases being compared, difference of duration of corresponding silences should not exceed 25%.

In the Institute of Telecommunications, Warsaw University of Technology, a series of experiments were performed, in

order to confront the PESQ MOS-LQO values with subjectively assessed quality [7], [21]. The greatest discrepancies were observed if a voice activity detector (VAD) was simulated, which substituted zero-valued samples for silent intervals of the phrase (Fig. 7). Despite of a slight cropping of the initial or final consonants of some words, speech quality was almost unchanged, according to informal listening tests. However, the PESQ_MOS values dropped considerably, almost achieving 2, thus suggesting annoying distortions. We concluded, that PESQ results are not credible if the VAD is applied.
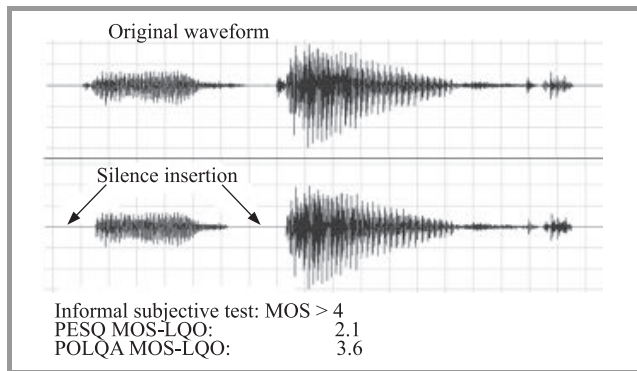

Fig. 7. Influence of the voice activity detector on the PESQ_MOS and POLQA_MOS [21].

The influence of the speaker and the phrase on the PESQ_MOS values is considerable: the results obtained for the same speech coder may differ by a unit on the MOS scale – see Fig. 8. Therefore it is necessary to increase number of speakers and phrases (in comparison with those recommended in [20]), in order to obtain credible average results. In our tests we have used 4 phrases and 4 speakers (2 men and 2 women) – in total 16 phrases.
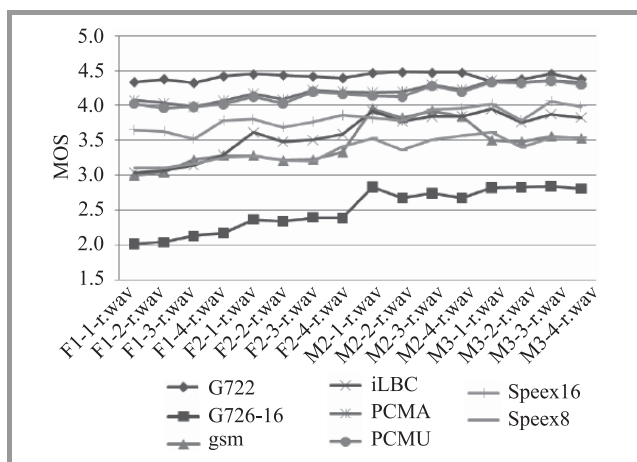

Fig. 8. Influence of the phrase and speaker on the PESQ_MOS [7].

In case of quality testing at low BER or PLR values, the number of phrases and their duration should be in-creased, because of the scatter of PESQ_MOS values due to random bit and packet loss process. This is illustrated in Fig. 9, where two series of tests were conducted, using the same coder (G.711 PCM) and PLR = 1%. This confirms our decision to use 16 phrases in our tests.
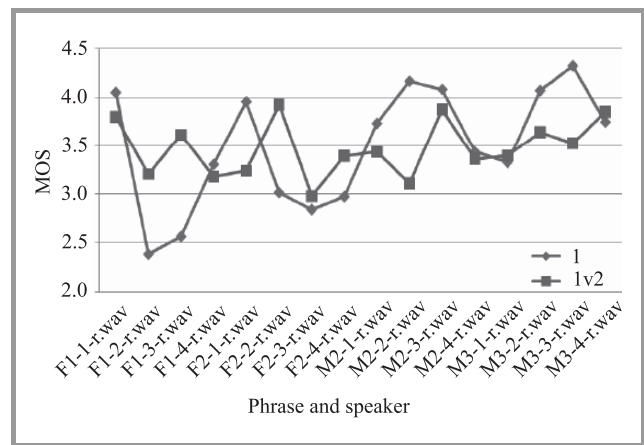

Fig. 9. PESQ_MOS for short speech phrases – PLR=1% [7].

We have also observed some synchronization problems: by increasing or decreasing the inter-word silent intervals the PESQ_MOS values changed despite of no change in subjectively assessed quality [7], [21].

The new algorithm for the objective speech quality evaluation, Perceptual Objective Listening Quality Analysis (POLQA) [22] is an improved version of PESQ. It may be used for quality measurement of speech signals of the bandwidth 4 kHz, 8 (or 7) kHz and 16 kHz. This method has an improved synchronization system and, unless the PESQ algorithm, may be used for the enhanced variable rate coders (EVRC) applied in CDMA systems. We have obtained a one-month license for the POLQA software from the Telchemy, Inc., and we observed, that the POLQA_MOS values exhibit greater correlation with the subjectively evaluated quality than the PESQ_MOS values. In particular, the quality assessment of phrases passed through the VAD simulator were much more realistic (POLQA_MOS = 3.6 while PESQ_MOS = 2.1 – see Fig. 7). Therefore we conclude that POLQA should be used instead of PESQ as a full reference algorithm.

For the on-line speech quality testing, ITU-T has recommended the 3SQM method [23]. It is a non-intrusive method, which does not require the original speech phrase. Speech quality assessment is based on the analysis of the processed phrase: the time-domain discontinuities, increased noise level, non-speech spectra are detected and then the dominant distortion source is found (the listener evaluates the speech quality out of this dominant distortion, usually ignoring the less annoying ones). Then the 3SQM_MOS value is calculated. Despite of relatively high correlation of 3SQM_MOS and PESQ_MOS, the intrusive methods, like PESQ and POLQA yield better accuracy of quality estimation. Therefore, these methods will be used for tests reported in Subsection 3.2.

3.2. Influence of IP Network Parameters on Speech Signal Quality

For testing the influence of the network parameters on speech signal quality, the server with Netem [17] network simulator and two workstations with Ekiga soft-phone applications were used. Ekiga [24] is a tool for VoIP and video-conference communication using SIP and H.323 protocols. It supports many speech coders, like G.711 PCM, Speex, iLBC, GSM-EFR and G.726 ADPCM (the latter with bit rates of 16, 24, 32 and 40 kbit/s). The wideband (speech bandwidth 7 kHz) speech coders Speex and G.722 are also supported.

The original phrase is read from the .wav file and sent directly (in digital form) to Ekiga. It is accomplished due to the application of the virtual audio cable (VAC) [25]. In a similar way, using the VAC, the received speech sig-
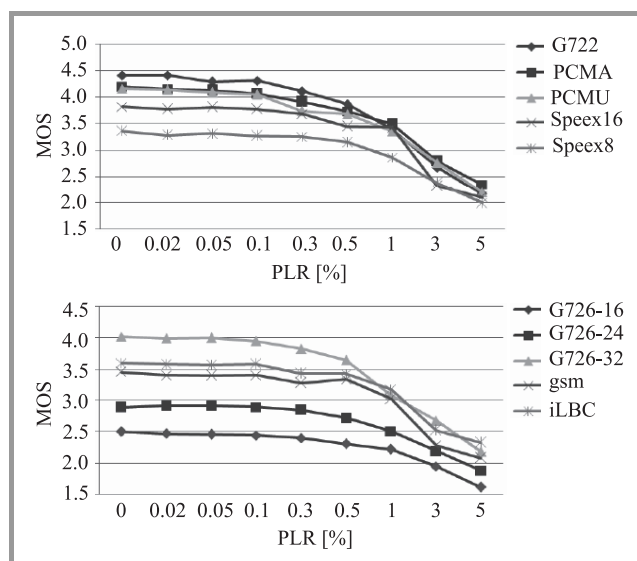

Fig. 10. PESQ_MOS versus PLR for different speech coders [7].
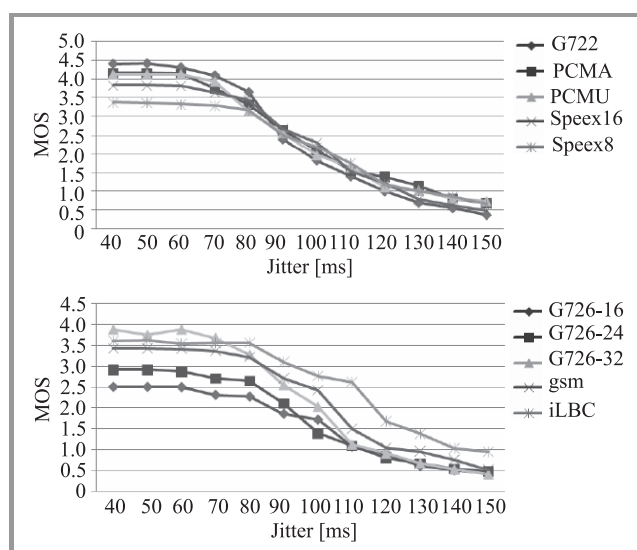

Fig. 11. PESQ_MOS versus delay jitter for different speech coders [21].

nal is written directly in the .wav file. Elimination of the A/D and D/A conversions is very important, because these operations cause a drop of the measured PESQ_MOS and POLQA_MOS values.

In our tests we have used 4 phrases and 4 speakers (2 men and 2 women) – in total 16 phrases concatenated in a single .wav file. In Figs. 10 and 11 the results of tests are shown. According to our tests, the impact of the packet loss on the speech quality is negligible if $PLR < 0.2\%$. It is much more restrictive condition than the threshold value specified for the conversational voice services in ITU-T Recommendation G.1010 [1], i.e., $PLR < 3\%$. However, in ETSI document [2] similar tests are reported, suggesting that PLR should be less than $0.2\% - 0.5\%$ (depending on speech coder), if $MOS > 4$ is to be maintained. The ITU-T Recommendation Y.1541 specifies more restrictive threshold value: $PLR < 0.1\%$. Our tests confirm this value.

The delay jitter may cause a drop of speech quality if it is greater than about 60 ms (Fig. 11). In ITU-T Recommendation G.1010 [1] a very restrictive threshold value is specified, namely 1 ms. However in ITU-T Recommendation Y.1541 delay jitter threshold is set at 50 ms, which is also confirmed by our results.

4. QoS Conditions for Selected Communication Services

In this paper two kind of problems are considered: credibility of tools for speech and video quality evaluation and QoS conditions for services based on speech and video signal transmission through the IP networks.

As a tool for the objective full reference speech quality evaluation the PESQ algorithm (ITU-T Recommendation P.862) [19] was examined in detail. According to our tests the credibility conditions specified for this algorithm in Recommendation P.862.3 [20] are not sufficient. In particular, PESQ delivers far too low quality estimation marks (PESQ_MOS values) if a voice activity detector (VAD) is applied. Moreover, the time domain synchronization of two phrases being compared is not perfect, which again yields too low PESQ_MOS values. The number of phrases should be greater than that specified in [20] – instead of 4 phrases we used 16 ones (4 phrases pronounced by 4 speakers). Our comparison of the PESQ algorithm and the newly introduced POLQA (ITU-T Recommendation P.863) [22] reveals that POLQA has better synchronization system and is not so sensitive to modifications introduced by VAD. Experiments illustrated with Fig. 7 show that erroneous drop of MOS is not so considerable as that of PESQ. Therefore we conclude that the POLQA_MOS is a more credible speech quality metric than PESQ_MOS. Our observations concerning the number of phrases and speakers still hold for POLQA algorithm.

PESQ as well as POLQA are intrusive algorithms – selected phrases, known at the receiver's side, must be transmitted through the network. For the on-line speech quality testing,

the 3SQM method [23] is recommended. As a no-reference algorithm, 3SQM is less accurate than PESQ, but a calibration procedure may increase the correlation between the 3SQM_MOS and PESQ_MOS values [26].

For testing the quality of video sequences, we have used a calibrated VQM metric. The calibration process was described in Subsection 2.3. The calibrated VQM values (VQM_MOS) exhibit relatively high correlation with the subjectively evaluated MOS values (Pearson's correlation 0.88). Because of unavailability of the software of the J.247 algorithms [11] we could not make any comparisons using this newly introduced recommendation.

For the on-line video quality evaluation, VQmon of the Telchemy Inc. [14] may be used. In our opinion however, additional conditions should be fulfilled in order to obtain credible results. Stable results are not always obtained at the beginning of the test, therefore it is better to increase the test duration – see Subsection 2.3. The results obtained in presence of packet delay jitter and corruption of the packets' content are not always accurate.

The QoS conditions for telecommunications services based on streaming od video signals were analyzed in terms of the BER, PLR, channel bandwidth (throughput) and delay jitter. The proposed thresholds for BER and PLR (Table 3) are more restrictive that these proposed in ITU Recommendation G.1010 [1], but slightly less demanding as those specified in the ITU-T Recommendation J.241 [4]. According to our tests, video transmission is very sensitive to packet delay jitter. This is due to the UDP protocol which performs no packet numbering and permutation of packets may occur. For the RTP protocol the corresponding threshold would be much higher. It should be considered, that the receiving buffer size may also influence the sensibility of transmission system to delay jitter. So as to the channel bandwidth, we have observed, that any value below the maximum bit rate of a video sequence (in the case of the variable bit rate coding) may cause visible distortions. Therefore we support the opinion, that the channel bandwidth should be greater than the maximum instantaneous bit rate of the transmitted video sequence.

The real time (RT) interactive services based on speech transmission (like VoIP) are less sensitive to BER and PLR than the services based on video transmission. According to our results PLR < 0.2% enables good speech quality, which is close to threshold value specified in the ITU-T Recommendation Y.1541 (PLR < 0.1%). The quality of the RT interactive services depends on the transmission delay, but in our tests we have skipped this parameter, because the acceptable delay values have been specified in ITU-T Recommendation G.114 [27] and they seem to be credible. According to this recommendation, the one-way delay should not exceed 150 ms. Values less than 250 ms may be accepted, but some users may perceive them as irritating. Our tests have shown, that the delay jitter should be less than 60 ms which is close to the value specified in the ITU-T Recommendation Y.1541 [3]. Note that this value is much greater as the corresponding threshold for video

transmission, but this is due to protocols which prevent the permutation of packets.

Comparison of speech coders (Figs. 10 and 11) shows the advantage of the G.722 coder, but it is paid with the relatively high bit rate (64 kbit/s). A good choice for the VoIP service would be the iLBC coder, yielding quite a good quality at bit rates 13 or 15 kbit/s. Note that at the PLR = 1% speech quality of most of the tested coders is similar (MOS about 3.5).

The results presented in this paper were obtained for typical VoIP and VoD hardware and software configurations and currently available tools for speech and image quality evaluation. The telecommunication services as well as tools for quality evaluation are still in phase of development [28]. Therefore the QoS conditions are still being reformulated and specified with greater accuracy.

Acknowledgement

References

[1] "End-user Multimedia QoS Categories", ITU-T Recommendation G.1010, 2001.

[2] "Speech and Multimedia Transmission Quality (STQ); Audiovisual QoS for Communication over IP Networks", ETSI ES 202 667, 2009.

[3] "Network Performance Objectives for IP-based Services", ITU-T Recommendation Y.1541, 2006.

[4] "Quality of Service ranking and measurement methods for digital video services delivered over broadband IP Networks", ITU-T Recommendation J.241, 2005.

[5] W. Burakowski, J. Śliwiński, H. Tarasiuk, A. Beben, P. Krawiec, S. Kula, P. Dymarski, S. Kaczmarek, M. Narloch, H. Gut-Mostowy, W. Latoszek, P. Pyda, T. Dalecki, E. Niewiadomska-Szynkiewicz, P. Arabas, M. Rotnicki, and T. Wiśniewski, "Specyfikacja Systemu IP QoS opartego na architekturze DiffServ" (in Polish), in *Proc. KSTiT, Przegl. Telekom.*, vol. 8-9, pp. 966–972, 2009.

[6] S. Kula, P. Dymarski and G. Szmyd, "Wpływ parametrów sieci na jakość sygnałów wideo (in Polish)", in *KSTiT, Przegl. Telekom.*, vol. 8-9, pp. 796–799, 2009.

[7] P. Dymarski, S. Kula and A. Sadowska, "PESQ jako narzędzie do oceny jakości sygnału VoIP (in Polish)", *KSTiT, Przegl. Telekom.*, vol. 8-9, pp. 1299–1308, 2010.

[8] "Subjective Video Quality Assessment for Multimedia Applications", ITU-T Recommendation P.910, 1999.

[9] "Methodology for the Subjective Assessment of the Quality of Television Pictures", ITU-R Recommendation BT.500-12, 2009.

[10] "Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference", ITU-T Recommendation J.144, 2004.

[11] "Objective Perceptual Multimedia Video Quality Measurement in the Presence of a full Reference", ITU-T Recommendation J.247, 2008.

[12] Y. Wang, "Survey of Objective Video Quality Measurements", Department of Computer Science, Worcester Polytechnic Institute, June 2006.

[13] "Perceptual Visual Quality Measurement Techniques for Multimedia Services over Digital Cable Television Networks in the Presence of a Reduced Bandwidth Reference", ITU-T Recommendation J.246, 2008.

[14] "TVQM video quality metrics", *Understanding IP Video Performance*, Feb. 2008, Telchemy application note.

[15] "MSU Video Quality Measurement tools", MSU Graphics and Media Lab (Video Group), May 2011 [Online]. Available: http://compression.ru/video/quality_measure/index_en.html

[16] "VLC Media Player", VideoLan Organization, May 2011 [Online]. Available: www.videolan.org/vlc

[17] "Netem", The Linux Foundation, May 2011 [Online]. Available: www.linuxfoundation.org/collaborate/workgroups/networking/netem

[18] DSL Forum, "Triple-play Services Quality of Experience (QoE) Requirements", Technical Report TR-126, 2006.

[19] "Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs", ITU-T Recommendation P.862, 2002.

[20] "Application Guide for Objective Quality Measurement Based on Recommendations P.862, P.862.1 and P.862.2", ITU-T Recommendation P.862.3, 2007.

[21] A. Sadowska, "Algorytm PESQ jako narzędzie do oceny jakości sygnału mowy (in Polish)", master's degree dissertation, Inst. of Telecommunications, Warsaw University of Technology, 2011.

[22] "POLQA – Perceptual Objective Listening Quality Analysis", ITU-T Recommendation P.863, 2010.

[23] "Single-ended Method for Objective Speech Quality Assessment in Narrow-Band Telephony Applications", ITU-T Recommendation P.563, 2004.

[24] "Ekiga", http://ekiga.org/

[25] E. Muzychenko, "Virtual audio cable" [Online]. Available: http://software.muzychenko.net/eng/vac.html

[26] Ł. Apiecionek, "Metoda oceny jakości transmisji głosowej w telefonii VoIP", Ph.D. dissertation, IPPT PAN, 2010.

[27] "International Telephone Connections and Circuits – General Recommendations on the Transmission Quality for an Entire International Telephone Connection – One Way Transmission Time", ITU-T Recommendation G.114, 2003.

[28] P. Dymarski and S. Kula, "Metody i standardy badania postrzeganej jakości sygnałów audio i wideo (in Polish)", in *proc. KSTiT, Przegl. Telekom.*, vol. 8-9, pp. 775–781, 2009.

includes various aspects of digital signal processing, particularly speech and audio compression for telecommunications and multimedia, audio watermarking and applications of Hidden Markov Models.
E-mail: dymarski@tele.pw.edu.pl
Institute of Telecommunications
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Sławomir Kula received the M.Sc. and Ph.D. degrees from the Faculty of Electronics, Warsaw University of Technology, in 1977 and 1982, respectively. In periods 1999–2002 and 2005–2008 he was Vice Dean of Faculty of Electronics and Information Technology. Since 2008 he is Deputy Director for Education at Institute of Telecommunications, Member of IEEE Communication Society (chairman of ComSoc Warsaw), Member of SIT (chairman of SIT-WUT). He is an author of books (in Polish): *Systemy teletransmisyjne*, *Przewodowe systemy dostępowe xDSL*, and *Systemy i sieci SDH* (also editor).
E-mail: skula@tele.pw.edu.pl
Institute of Telecommunications
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Przemysław Dymarski received the M.Sc. and Ph.D. degrees from the Wrocław University of Technology, Poland, in 1974 and 1983, respectively, both in Electrical Engineering. In 2004 he received the D.Sc. degree in Telecommunications from the Faculty of Electronics and Information Technology of the Warsaw University of Technology. Now he is with the Institute of Telecommunications, Warsaw University of Technology. His research

Thanh Nguyen Huy was born in Hanoi in 1986. Since 2005 he was a student at the Faculty of Electronics and Information Technology of the Warsaw University of Technology. In the years 2009/10 he participated in the LLP Erasmus Program, continuing his studies at St. Pölten University of Applied Science in Austria. His master's degree dissertation, *The impact of network parameters on perceived video quality* concerned the quality of selected telecommunications services, like VoD and IPTV. In 2011 he obtained master's degree in Telecommunications.
E-mail: T.NguyenHuy@stud.elka.pw.edu.pl
Institute of Telecommunications
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland