Impact of Different Active-Speech-Ratios on PESQ's Predictions in Case of Independent and Dependent Losses

Peter Počta, Helena Vlčková, Zuzana Polková

Dept. of Telecommunications and Multimedia, FEE, University of Žilina, Univerzitná 1, SK-01026, Žilina, Slovakia, pocta@fel.uniza.sk

Abstract

This paper deals with the investigation of *PESQ*'s behavior under independent and dependent loss conditions from an Active-Speech-Ratio perspective. This reference signal characteristic is defined very broadly by *ITU-T* Recommendation *P.862.3*. That is the reason to investigate an impact of this characteristic on speech quality prediction more in-depth. The *ITU-T G.729AB* encoding scheme is deployed in this study. We assess the variability of *PESQ*'s predictions with respect to Active-Speech-Ratios and loss conditions. Our results show that an increase in amount of speech in the reference signal (expressed by the Active-Speech-Ratio characteristic) may result in an increase of the reference signal sensitivity to packet loss change and also *PESQ*'s predictions accuracy improving. Predictions accuracy could be even improved by higher packet losses.

Keywords

Perceptual Evaluation of Speech Quality (*PESQ*), Speech quality, Intrusive measurement, Voice over Internet Protocol (*VoIP*), Reference signal characteristic, Active-Speech-Ratio.

1 Introduction

Voice over Internet Protocol (*VoIP*), the transmission of packetized voice over *IP* networks, has gained much attention in recent years. It is expected to carry more and more voice traffic for its cost-effective service. However, present-day Internet, which was originally designed for data communications, provides *best-effort* service only, posing several technical challenges for real time *VoIP* applications. Speech quality is impaired by packet loss, delay and jitter. Assessment of perceived speech quality in the *IP* networks becomes an imperative task to manufacturers as well as to service providers.

Speech quality is judged by human listeners and hence it is inherently subjective. The Mean Opinion Score (*MOS*) test, defined by *ITU-T* Recommendation *P.800* [1], is widely accepted as a norm for speech quality assessment. Subjective testing is expensive and time-consuming. That is the reason that subjective testing is impractical for the frequent testing such as routine network monitoring. Objective test methods have been developed in recent years. They can be classified into two categories: signal-based methods and parameter-based methods. Intrusive signal based methods use two signals as the input to the measurements, namely, a reference signal and a degraded signal, which is the output of the system under test. They identify the audible distortions based on the perceptual domain representation of two signals incorporating human auditory models. These methods include Perceptual Speech Quality Measure (*PSQM*) [2], Measuring Normalizing System (*MNB*) [3, 4], Perceptual Analysis Measurement System (*PAMS*) [5], and Perceptual Evaluation of Speech Quality (*PESQ*) [6, 7]. Among them, *PSQM* [8] and *PESQ* [9] were standardized by the *ITU-T* Recommendations such as *P.861* and *P.862* respectively. Parameter-based methods predict the speech quality through a computation model instead of using a real measurement. *E-model* is a typical model, defined by *ITU-T* Recommendation *G.107*. The *E-model* includes a set of parameters characterizing end-to-end voice transmission as its input, and the output (R-value) then can be transformed into the *MOS*-Listening Quality Estimated narrowband (*MOS-LQEn*) values.

The *PSQM* algorithm is based on comparison of the power spectrum of the corresponding sections of reference and degraded signals. The results of this algorithm more correlate with the results of listening tests, in comparison with *E-model*. At the present, this algorithm is no longer used due to a coarse time-alignment. Instead of it, the algorithm *PESQ* is rather used. *PESQ* combines merits of *PAMS* and *PSQM99* (an updated version *PSQM*), and adds new methods for transfer function equalization and averaging distortions over time. The algorithm *PESQ* facilitates with very fine time-alignment and one single interruption is also taken into account in the calculation of *MOS*. It can be used in wider range of network conditions, and gives higher correlation with subjective tests and the other objective algorithms [6-7, 9]. Unlike the conversational model, *PESQ* is a listening-only model; the degraded sample is time-aligned with the reference sample during pre-processing. The *PESQMOS* values do not reflect the effects of delay on speech quality. The disadvantages include impossibility to use it for codec's with data rate lower than 4 kbps and higher calculation load what is caused by recursions in the algorithm.

The characteristics of reference signals for objective speech quality measurements provided by *PESQ* are defined in Section 7 of the *ITU-T* Recommendation *P.862.3* [10]. Two reference signal characteristics are defined very broadly by this Recommendation from our point of view, namely the length of reference signal and Active-Speech-Ratio. The above-mentioned document recommends to use the reference signals in duration in the range from 8 seconds to 30 seconds for the purpose of *PESQ's* measurements. The speech activity in the reference signals, which can be measured according to *ITU-T* Recommendation *P.56* [11], should be between 40% and 80% of their length. We suppose that those two characteristics can have an impact on final *PESQ*'s predictions. The detailed investigation of both characteristics has been proposed in [12] from *PESQ*'s prediction perspective. Some very important issues raise from [12] especially in the case of Active-Speech-Ratio experiment. That is the reason for exhaustive investigation of the impact of different Active-Speech-Ratios on speech quality prediction provided by *PESQ* from dependent and independent losses perspective.

Some works have been carried out on study of *PESQ*'s behavior under single frame, uniform and dependent losses. In [13], the verification of *PESQ* performance in case of single frame losses has been conducted by means of formal listening only tests. The tests have proved that *PESQ* predicts the impact of single frame losses precisely. In [14], an investigation how subjects perceive bursty losses and how current objective measurement methods, such as *PSQM*, *MNB*, Enhanced Modified Bark Spectral Distance (*EMBSD*) and *PESQ,* correlate with subjective test results under burst loss conditions has been reported. Preliminary results have shown that *PESQ* displays an obvious sensitivity to bursty conditions compared to human subjects (it is more sensitive than subjects when loss burstiness is high and less sensitive when it is low). In [15], a study of *PESQ*'s behavior from networking perspective (dependent and uniform losses) has been presented. It seems that *PESQ* maintains reasonable correlation with subjective scores even when the network conditions are bad. Also, the deviations seem to be systematic from subjective scores, which suggest that a simple compensation factor might be found (for instance, derived from network conditions) and used to improve the results.

Here we focus on an impact of different Active-Speech-Ratios on speech quality prediction provided by *PESQ* in case of independent and dependent losses. The reference signals with Active-Speech-Ratios of 42, 62 and 82% are investigated and the *ITU-T G.729AB* encoding scheme is deployed in this study. We assess the variability of *PESQ*'s predictions with respect to Active-Speech-Ratios and loss conditions.

The rest of the paper is organized as follows: Section 2 introduces experimental scenario and experiments carried out in this study. In Section 3, the experimental results are presented and discussed. Section 4 concludes the paper and suggests some future studies.

2 Experiment description

2.1 Experimental scenario

One-way *VoIP* session was established between two hosts (*VoIP* Sender and *VoIP* Receiver), via the loss simulator (Figure 1). In case of loss simulator, two currently most widely used models have been deployed for the purpose of packet loss modeling, namely Bernoulli and Gilbert loss model. More details about loss models can be found in Section 2.2.

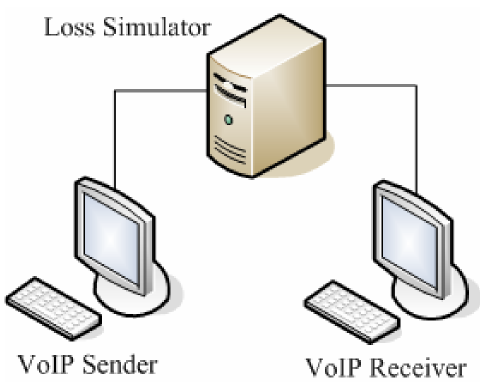

Fig.1 Experimental scenario

For this experiment the *ITU-T G.729AB* encoding scheme [16] was chosen. In the measurements, two frames were encapsulated into a single packet; thus corresponding to a packet size of 20 milliseconds. Adaptive jitter buffer, *G.729AB*'s native packet loss concealment (*PLC*), and Voice Activity Detection (*VAD*)/Discontinuous Transmission (*DTX*) were implemented in the *VoIP* clients used. The jitter buffer does not play any role in case of this experiment because of small constant jitter inserted by the loss simulator during the measurement. The Comfort Noise Generator (*CNG*) was disabled in case of this experiment.

The reference signals described in Section 2.3 were utilized for transmission through the given *VoIP* connection. Finally, speech quality was assessed by *PESQ* and then converted to *MOS*-Listening Quality Objective narrowband (*MOS-LQOn*) values by this equation:

$$y = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.4945 * x + 4.6607}} \qquad (1)$$

where *x* and *y* represent the raw *PESQ* score and the mapped *MOS-LQOn*, respectively. The equation mentioned is defined by *ITU-T* Recommendation *P.862.1* [17]. In case of *PESQ* score calculation, we used some batch data processing techniques proposed in [18].

2.2 Packet loss models

Packet loss is a major source of speech impairment in *VoIP*. Such a loss may be caused by discarding packets in the *IP* networks (network loss) or by dropping packets at the gateway/terminal due to late arrival (late loss).

Several models [19, 20] have been proposed for modelling network losses, the currently most widely used of them will be briefly discussed in the following subsections.

2.2.1 Bernoulli model

In the Bernoulli loss model, each packet loss is independent (memoryless), regardless of whether the previous packet is lost or not. In this case, there is only one parameter, the average packet loss rate, which is the number of lost packets divided by the total number of transmitted packets in a trace.

2.2.2 Gilbert model

Most research in *VoIP* networks uses a Gilbert model to represent packet loss characteristics [19-21]. In 2-state Gilbert model as shown in Figure 2, State 0 is for a packet received (no loss) and State 1 is for a packet dropped (loss). p is the probability that a packet will be dropped given that the previous packet was received. 1-q is the probability that a packet will be dropped given that the previous packet was dropped. 1-q is also referred to as the conditional loss probability (*clp*). The probability of being in State 1 is referred to as unconditional loss probability (*ulp*). The *ulp* provides a measure of the average packet loss rate and is given by:

$$ulp = \frac{p}{p + q} \qquad (2)$$

The *clp* and *ulp* are used in the paper to characterize the loss behavior of the network.

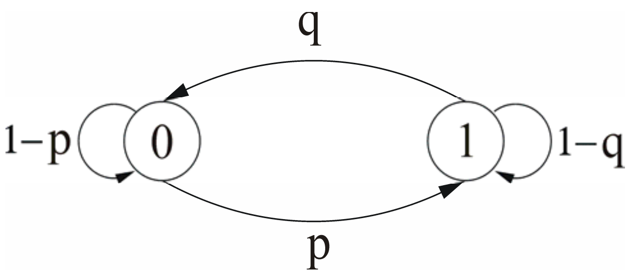

Fig.2 Gilbert model

Sixteen independent loss and dependent loss conditions were chosen to cover cases of interest. They consist of combinations of packet loss rate (from 0% to 15%) in case of independent losses and unconditional loss probability (*ulp*, 0%, 1.5%, 3%, 5%, 10% and 15%), conditional loss probability (*clp*, 15%, 30% and 50%) in case of dependent losses and 20 initial seeds to simulate different loss locations in both cases.

2.3 Reference signals

The reference signals selection should follow the criteria given by *ITU-T* Recommendations *P.830* [22] and *P.800* [1]. The reference signals should include talkspurts separated by silence periods, and are normally of 1-3 seconds long. They should also be active for 40-80% of their duration. The reference signals are composed of speech records. In our experiments, these speech records were taken from a Slovak speech database. In each set, two female and two male speech utterances were used. The reference signals were stored in 16-bit, 8000 Hz linear *PCM*. Background noise was not present.

Reference signals in length of 30 seconds with Active-Speech-Ratios of 42, 62 and 82% were applied. All reference signals used were spoken by the same people (as defined in Table 1), also for different Active-Speech-Ratios. The differences between reference signals used are only in case of number of talkspurts (sentences), resulting in different Active-Speech-Ratios. In case of higher Active-Speech-Ratios, the new sentences were added, as an extension. The decision about using reference signals in length of 30 seconds came from our previous published work [12]. The tests have proved that this length provides more accurate results in comparison with other investigated lengths therefore enables more precise investigation of an impact of different Active-Speech-Ratios on speech quality prediction, assessed by *PESQ*. The long reference signals usage for the speech quality assessment by *PESQ* has been also investigated in [23]. The experimental results have shown that for this purpose it is possible to use a longer reference signals and the author has proposed extending the maximum length of reference signals to 30 seconds. The results of this work have been included in *ITU-T* Recommendation *P.862.3*.

Table 1 Active-Speech-Ratios and numbers of talkspursts of the reference signals

Reference signal	Active-Speech-Ratio of 42%	Active-Speech-Ratio of 62%	Active-Speech-Ratio of 82%
Male1	40.183 (6)	60.333 (8)	82.516 (10)
Male2	43.249 (6)	63.861 (9)	80.612 (11)
Female1	43.677 (6)	62.065 (8)	84.180 (10)
Female2	41.780 (4)	62.054 (6)	82.121 (8)
Average value	**42.222 (5.5)**	**62.078 (7.75)**	**82.375 (9.75)**

The Active-Speech-Ratios and numbers of talkspurts (active speech periods) for each of the reference signals used are presented in Table 1. The Active-Speech-Ratio measurement process has to follow the criteria given by *ITU-T* Recommendation *P.56*. Those ratios were measured by means of *ITU-T* Recommendation *G.191*'s software tool [24], known as *sv56*.

3 Experimental results

The measurements were independently performed 80 times (20 different loss locations/patterns and 4 reference signals) under the same packet loss (independent losses) and the same pair of *ulp* and *clp* (dependent losses). The average *MOS-LQOn* score, 95% Confidence Interval (*CI*) and Mean Absolute Deviation (*MAD*) were calculated. The next subsections describe experimental results for the both examined types of losses in more details.

3.1 Experimental results for independent losses

Using a Bernoulli model gives us the possibility to analyze *PESQ*'s behavior only from two perspectives, namely packet loss and Active-Speech-Ratio. Figures 3 and 4 depict differences between investigated Active-Speech-Ratios in speech quality evaluation, provided by *PESQ*. It can be seen from above-mentioned figures that the difference in Active-Speech-Ratio has a significant impact on overall speech quality. This fact contributes our preliminary assumption that an increasing amount of speech (Active-Speech-Ratio) in reference signal has to result in increase of reference signal sensitivity to packet loss change. That may be explained by increase/decrease of information (speech) loss probability at the same packet loss ratio in the case of using higher/lower Active-Speech-Ratio. It is caused by a greater number of active speech periods in reference signals with higher Active-Speech-Ratio. The probability of information loss is greater if more periods are available. It means that it is possible to capture more impairments of speech quality in such a case. By capturing the majority of existing impairments, we are able to get a better insight about speech quality in investigated telecommunication network (especially in *VoIP* case) which turns to more reliable and accurate evaluation of investigated transmission line from this point of view.

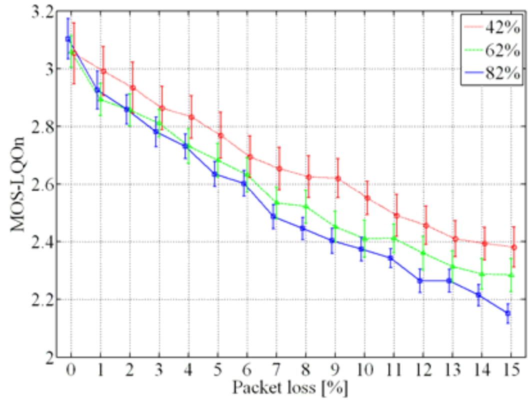

Fig.3 MOS-LQOn as a function of packet loss for different Active-Speech-Ratios in case of independent losses. The vertical bars show 95 % CI (derived from 80 measurements) for each loss.

The above-mentioned effect is depicted in Figure 3. In more detail, it can be seen in this figure that *MOS-LQOn* for higher Active-Speech-Ratio (82%) decreases faster in comparison with ratios 42% and 62%.

However, it can be seen from Figure 3, that one bias has been obtained in case of 1% packet loss. It could be explained by a bit more captured losses at active speech periods (effective losses) in case of 62% Active-Speech-Ratio. For instance, in the mentioned case we captured approximately 15 loss events and only one odd effective loss could have big impact on final *PESQ* score. Apparently, this effect is mainly occurred at lower packet losses, where the difference between captured effective losses is small.

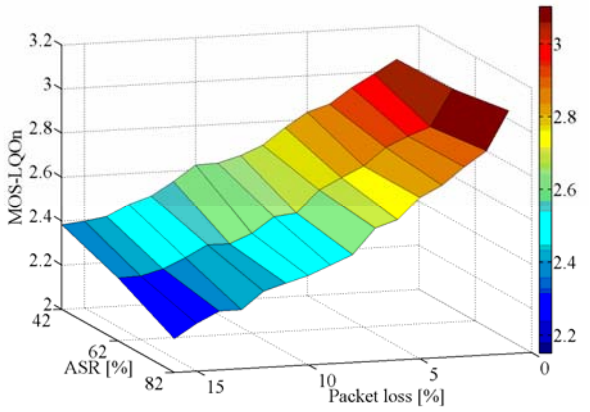

Fig.4 MOS-LQOn versus packet loss and Active-Speech-Ratio for independent losses.

Figure 5 shows *MAD* of *MOS-LQOn*'s, which has been obtained for this experiment. It can be seen from Figure 5 that the accuracy of predictions is much more in case of higher Active-Speech-Ratios, especially when network condition degrade (packet loss increase). First fact is related to sensitivity effect mentioned above and second one could be explained by higher probability of losses obtained at active speech intervals (effective loss probability) at higher packet losses. More effective losses may lead to small variation in *PESQ* score. If the higher Active-Speech-Ratio is used, this effect could be even more markedly achieved. This effect could be characterized as the sensitivity effect gain.

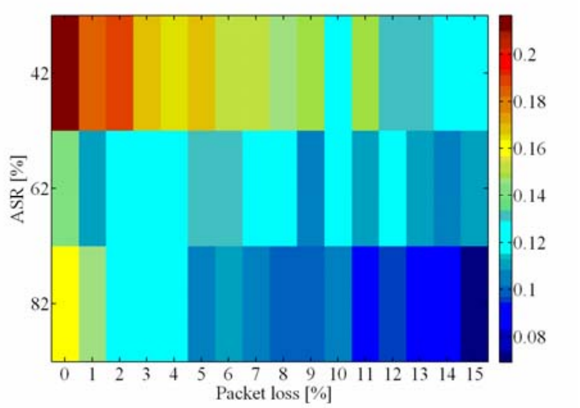

Fig.5 MAD of MOS-LQOn's at each point of loss space and Active-Speech-Ratio in case of independent losses.

A two-way analysis of variance (*ANOVA*) was conducted on *MOS-LQOn*'s using packet loss and Active-Speech-Ratio as fixed factors (Appendix 7.1, Table 2). We found clearly the highest F-ratio for the packet loss ($F = 456.38$, $p < 0.01$). Moreover, the Active-Speech-Ratio factor showed a little bit weaker effect on quality than packet loss, with $F = 153.87$, $p < 0.01$.

3.2 Experimental results for dependent losses

Using a Gilbert model extends our possibilities to investigate *PESQ*'s behavior to three perspectives, namely *ulp*, *clp* and naturally Active-Speech-Ratio. The experimental results for all investigated *clp*'s are depicted in Figures 6, 7 and 8. We can observe how speech quality drops, as expected, with both *clp* and *ulp*. Also, it is clear that the different Active-Speech-Ratios could seriously influence the quality in case of dependent losses. Obviously, we obtained same effect as in first case (independent losses). It means that using higher Active-Speech-Ratio leads to increase of reference signal sensitivity to packet loss change, also in case of dependent losses. One more bias in comparison with fist case has been achieved in case of 3% *ulp* (Figure 6). The reason for that is same as mentioned above.

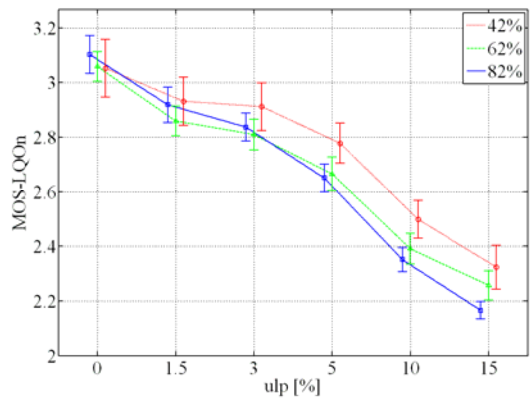

Fig.6 MOS-LQOn as a function of unconditional loss probability for different Active-Speech-Ratios in case of dependent losses (clp = 30%). Other detailed descriptions of Figure 3 apply appropriately.

In Figure 9, we can see the *MAD* of *MOS-LQOn*'s for 30% *clp*. Unsurprisingly, *PESQ*'s accuracy behavior is also similar as obtained in previous case. Interestingly, the highest deviation has been obtained at 0% packet loss. At this time, we have no theory that could explain this phenomenon. Naturally, that is a point for a future investigation because exhaustive study is needed to validate, and interpret this phenomenon. On basis of those results, we can pronounce that higher Active-Speech-Ratio usage may lead to *PESQ*'s predictions accuracy improving, especially in case of higher packet loss values.

Three two-way *ANOVA*'s were similarly carried out on *MOS-LQOn*'s for all investigated *clp*'s, using *ulp* and Active-Speech-Ratio as fixed factors (Appendix 7.2, Tables 3-5). We obtained similar results as in case of independent losses.

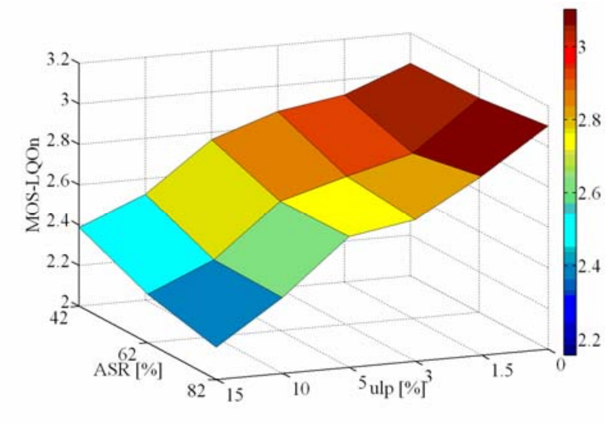

Fig.7 MOS-LQOn versus unconditional loss probability and Active-Speech-Ratio for dependent losses (clp = 50%).

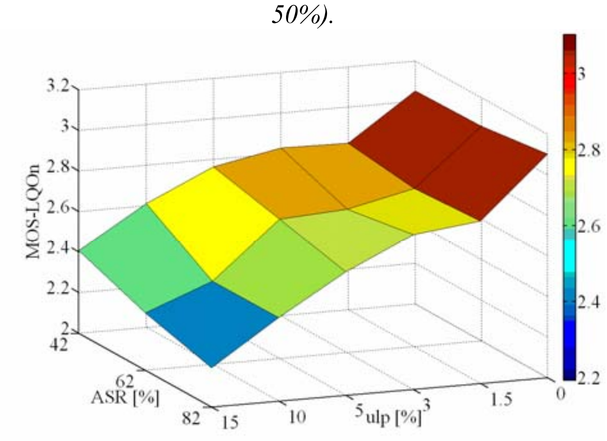

Fig.8 MOS-LQOn versus unconditional loss probability and Active-Speech-Ratio for dependent losses (clp = 15%).

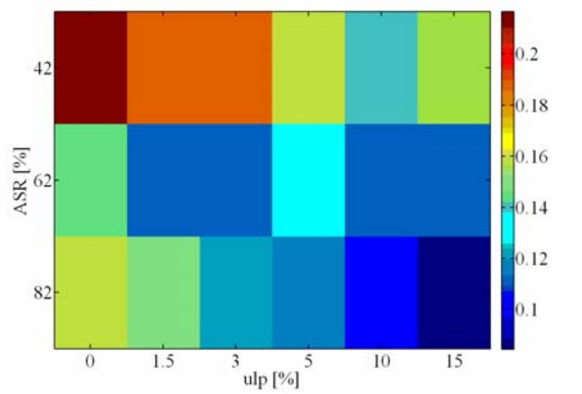

Fig.9 MAD of MOS-LQOn's at each point of loss space and Active-Speech-Ratio in case of dependent losses (clp = 30%).

The experimental results show that the change of Active-Speech-Ratio has a significant impact on overall speech quality in both investigated cases. This fact is our motivation for finding of the feasible average Active-Speech-Ratios for some languages or types of languages and conversational scenarios. Naturally, an issue of Active-Speech-Ratio setup with regards to different languages and conversational scenarios is also open for discussion. Average Active-Speech-Ratios adjustment might enable to provide an assessment of speech quality more reliably. Nowadays, such improved assessment of speech quality is demanded to be involved into Quality of Service in real *VoIP* scenarios to make comparison among network providers more feasible.

4 Conclusions and future work

This paper has investigated an impact of different Active-Speech-Ratios of an input reference signals in *PESQ* based speech quality prediction in case of dependent and independent losses. The main goal of this study is to gain a better understanding of behavior of the *PESQ*'s predictions under different Active-Speech-Ratios. The results presented in the paper have approved our hypothesis that an increase in amount of speech in the reference signal (expressed by the Active-Speech-Ratio characteristic) may result in an increase of the reference signal sensitivity to packet loss change and also *PESQ*'s predictions accuracy improving. Predictions accuracy could be even improved by higher packet losses.

A future work will focus towards the following issues. At first, we would like to verify our results by subjective tests. We are currently preparing the subjective tests in cooperation with *MESAQIN*'s laboratory in Prague (Czech Republic). Secondly, we will attempt to find out an appropriate average Active-Speech-Ratios for some languages or type of languages and conversational scenarios. Apparently, this point could be very interesting for other speech quality laboratories around the world. By this investigation, we might refine on the existing broadly recommended Active-Speech-Ratios (40% - 80%), defined by *ITU-T* Recommendation *P.862.3* and provide for more reliable speech quality assessment, provided by *PESQ*. Thirdly, we plan to exhaustively study the highest *MAD* at 0% packet loss and find out the reason for that.

5 Acknowledgements

This work has been partially supported by the Slovak VEGA grant agency, Project No.1/0313/08, "The investigation of the methods of detection of the critical conditions in telecommunication networks from the speech quality point of view" and the Slovak Research and Development Agency under the contract No.APVV-0369-07.

In addition, we would like to thank Joachim Pomy (*ETSI/STQ, ITU-T/SG12*) for valuable comments and help in preparation of this paper and last not at least the reviewers for their invaluable comments and suggestions.

6 References

[1] ITU-T Rec. P.800 "Methods for subjective determination of transmission quality" International Telecommunication Union, Geneva (Switzerland), 1996.

[2] Beerends, J. G., Stemerdink, J. A.: A Perceptual Speech Quality Measure Based on a Psychoacoustic Sound Representation, In *J. Audio Eng. Soc.*, vol. 42, pp. 115-123, 1994, ISSN 1549-4950.

[3] Voran, S.: Objective Estimation of Perceived Speech Quality - Part I: Development of the Measuring Normalizing Block Technique, In *IEEE Trans. on Speech and Audio Processing*, vol. 7, pp. 371-382, 1999, ISSN 1063-6676.

[4] Voran, S.: Objective Estimation of Perceived Speech Quality - Part II: Evaluation of the Measuring Normalizing Block Technique, In *IEEE Trans. on Speech and Audio Processing*, vol. 7, pp. 383-390, 1999, ISSN 1063-6676.

[5] Rix, A. W. and Hollier, M. P.: The perceptual analysis measurement system for robust end-to-end speech quality assessment. In *Proceedings of IEEE ICASSP 2000*, Istanbul (Turkey), June 2000.

[6] Rix, A. W., Hollier, M. P., Hekstra, A. P. Beerends, J. G.: PESQ, the new ITU standard for objective measurement of perceived speech quality, Part 1 – Time alignment, In *J. Audio Eng. Soc.*, vol. 50, pp. 755-764, 2002, ISSN 1549-4950.

[7] Beerends, J. G., Hekstra, A. P., Rix, A. W., Hollier, M. P.: PESQ, the new ITU standard for objective measurement of perceived speech quality, Part II – Perceptual model, In *J. Audio Eng. Soc.*, vol. 50, pp. 765-778, 2002, ISSN 1549-4950.

[8] ITU-T Rec. P.861 "Objective quality measurement of telephone-band (300-3400 Hz) speech codecs" International Telecommunication Union, Geneva (Switzerland), 1998.

[9] ITU-T Rec. P.862 "Perceptual Evaluation of Speech Quality "International Telecommunication Union, Geneva (Switzerland), 2001.

[10] ITU-T Rec. P.862.3 "Application guide for objective quality measurement based on Recommendations P.862, P.862.1 and P.862.2 "International Telecommunication Union, Geneva (Switzerland), 2005.

[11] ITU-T Rec. P.56 "Objective measurement of active speech level "International Telecommunication Union, Geneva (Switzerland), 1993.

[12] Počta, P., Mrvová, M., Kortiš, P., Palúch, P., Vaculík, M.: A systematic study of PESQ's behavior in simulated VoIP environment (from reference signals characteristics perspective), In *Proceedings of conference MESAQIN 2008*, Prague (Czech Republic), 2008, pp. 13-21, ISBN 978-80-01-04193-2.

[13] Hoene, Ch., Dulamsuren-Lalla, E.: Predicting Performance of PESQ in Case of Single Frame Losses, In *Proceedings of conference MESAQIN 2004*, Prague (Czech Republic), 2004, ISBN 80-01-03017-2.

[14] Sun, L., Ifeachor, E.C.: Subjective and Objective Speech Quality Evaluation under Bursty Losses, In *Proceedings of conference MESAQIN 2002*, Prague (Czech Republic), 2002, ISBN 80-01-02515-2.

[15] Varela, M., Marsh, I., Gronvall, B.: A systematic study of PESQ's behavior, In *Proceedings of conference MESAQIN 2006*, Prague (Czech Republic), 2006, ISBN 80-01-03503-4.

[16] ITU-T Rec. G.729 "Coding of speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Exited Linear prediction (CS-ACELP)" International Telecommunication Union, Geneva (Switzerland), 2007.

[17] ITU-T Rec. P.862.1 "Mapping function for transforming P.862 raw result scores to MOS-LQO "International Telecommunication Union, Geneva (Switzerland), 2003.

[18] Chochlík, M., Grondžák, K., Baboš, Š.: Windows Operating System Core Programming, (in Slovak), University of Žilina, Žilina (Slovakia), 2009, ISBN 978-80-8070-970-9.

[19] Yajnik, M., Moon, S., Kurose, J., Towsley, D.: Measurement and Modelling of the Temporal Dependence in Packet Loss, In *Proceedings of conference IEEE INFOCOM 99*, vol. 1, New York (USA), pp. 345–352, March 1999.

[20] Jiang W., Schulzrinne, H.: QoS Measurement of Internet Real-Time Multimedia Services, Technical Report, CUCS-015-99, Columbia University (USA), Dec. 1999.

[21] Sanneck H., Le, N. T. L.: Speech Property-Based FEC for Internet Telephony Applications, In *Proceedings of the SPIE/ACM SIGMM Multimedia Computing and Networking Conference*, San Jose, CA, (USA), pp. 38–51, Jan. 2000.

[22] ITU-T Rec. P.830 "Subjective Performance Assessment of digital telephone-band and wideband digital codecs" International Telecommunication Union, Geneva (Switzerland), 1996.

[23] Takahashi, A.: Objective quality evaluation based on ITU-T Recommendation P.862 by using long reference speech (NTT), COM12-D008,

International Telecommunication Union, Geneva (Switzerland), January 2005.

[24]ITU-T Rec. G.191 "Software tools for speech and audio coding standardization" International Telecommunication Union, Geneva (Switzerland), 2005.

7 Appendix

7.1 Independent losses

Table 2 provides the results of the analysis of variance (*ANOVA*) carried out on the objective independent losses test results (Dependent variable: *MOS-LQOn*) described in more detail in Section 3.1.

Table 2 Summary of ANOVA conducted on MOS-LQOn's in case of independent losses

Effect	SS	df	MS	F	p
Packet loss (1)	171.476	15	11.4318	456.38	0.0000
Active-Speech-Ratio (2)	7.708	2	3.8541	153.87	0.0000
(1)*(2)	3.577	30	0.1192	4.76	0.0000
Error	94.985	3792	0.025		
Total	277.746	3839			

7.2 Dependent losses

In Tables 3, 4 and 5, the results of the analysis of variance (*ANOVA*) for the dependent losses test results and all investigated *clp*'s (Dependent variable: *MOS-LQOn*) are shown. More details about this can be found in Section 3.2.

Table 3 Summary of ANOVA conducted on the MOS-LQOn's in case of dependent losses (clp = 50%)

Effect	SS	df	MS	F	p
ulp (1)	82.681	5	16.5362	524.1	0.0000
Active-Speech-Ratio (2)	1.994	2	0.9969	31.6	0.0000
(1)*(2)	1.882	10	0.1882	5.96	0.0000
Error	44.867	1422	0.0316		
Total	131.423	1439			

Table 4 Summary of ANOVA conducted on the MOS-LQOn's in case of dependent losses (clp = 30%)

Effect	SS	df	MS	F	p
ulp (1)	94.003	5	18.8005	635.6	0.0000
Active-Speech-Ratio (2)	0.879	2	0.4397	14.87	0.0000
(1)*(2)	1.61	10	0.161	5.44	0.0000
Error	42.062	1422	0.0296		
Total	138.554	1439			

Table 5 Summary of ANOVA conducted on MOS-LQOn's in case of dependent losses (clp = 15%)

Effect	SS	df	MS	F	p
ulp (1)	68.802	5	13.7604	384.59	0.0000
Active-Speech-Ratio (2)	0.952	2	0.4758	13.3	0.0000
(1)*(2)	1.639	10	0.1639	4.58	0.0000
Error	50.879	1422	0.0358		
Total	122.271	1439			