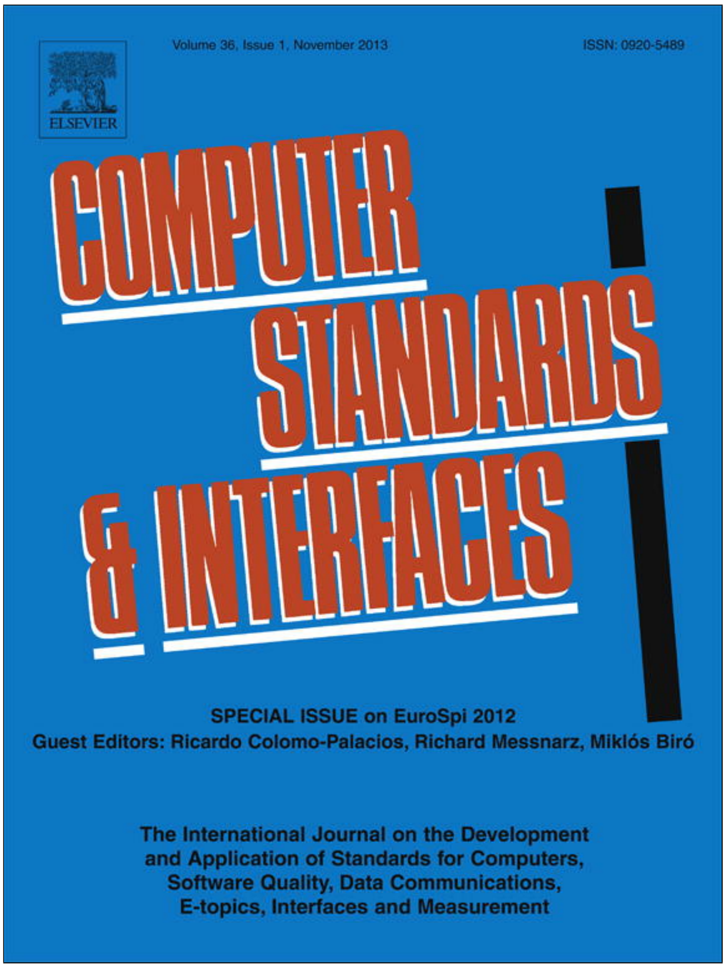

Volume 36, Issue 1, November 2013 ISSN: 0920-5489

COMPUTER STANDARDS & INTERFACES

SPECIAL ISSUE on EuroSpi 2012
Guest Editors: Ricardo Colomo-Palacios, Richard Messnarz, Miklós Biró

The International Journal on the Development
and Application of Standards for Computers,
Software Quality, Data Communications,
E-topics, Interfaces and Measurement

Computer Standards & Interfaces 36 (2013) 143–153

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

Effect of speech activity parameter on *PESQ*'s predictions in presence of independent and dependent losses

Peter Počta [a],*, Jan Holub [b]

[a] Dept. of Telecommunications and Multimedia, FEE, University of Žilina, Univerzitná 1, SK-01026 Žilina, Slovakia
[b] Dept. of Measurement K13138, FEE, CTU Prague, Technická 2, CZ-16627 Prague 6, Czech Republic

ARTICLE INFO

Article history:
Received 4 February 2011
Received in revised form 14 November 2011
Accepted 1 July 2013
Available online 16 July 2013

Keywords:
Speech quality assessment
Reference signal characteristic
Speech activity parameter
Packet loss
Perceptual Evaluation of Speech Quality (*PESQ*)

ABSTRACT

This paper deals with the investigation of *PESQ*'s (Perceptual Evaluation of Speech Quality; also known as *ITU-T* Recommendation *P.862*) behavior under independent and dependent loss conditions from a speech activity parameter perspective. The results show that an increase in amount of speech in the reference signal (expressed by the activity parameter) may result in an increase of the *PESQ* sensitivity to packet loss change as well as *PESQ*'s prediction accuracy improvement. On the other hand, it seems that human brain is a bit less sensitive to loss of some parts of words than *PESQ*. The reasons for those findings are particularly discussed.

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. In this case, speech quality is mainly impaired by packet loss, delay and jitter. Assessment of perceived speech quality in *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. Therefore, the most reliable approach for assessing speech quality is through subjective tests. The absolute category rating (ACR) test, defined by *ITU-T* Recommendation *P.800* [1], is one of widely accepted norms for speech quality assessment. In the test, listeners express their opinions on the quality of the speech materials in terms of five categories: excellent, good, fair, poor and bad with corresponding integer scores: 5, 4, 3, 2 and 1, respectively. The ratings are averaged and the result is usually known as mean opinion score (MOS). Subjective testing is expensive and time-consuming. It is the reason that subjective testing is impractical for the frequent testing such as routine network monitoring. An interested reader can find more details about subjective testing in [2]. Objective test methods have been developed in recent years. They are machine-executable and require a little human involvement. In principle, objective methods can be classified into two categories: signal-based methods and parameter-based methods. The former requires availability of speech signal to realize quality prediction process and can be according to [3] divided into two categories, intrusive or non-intrusive. Intrusive signal-based methods use two signals as the input to the measurement, 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. Several intrusive models have been developed during recent years, like Perceptual Speech Quality Measure (*PSQM*) [4], Measuring Normalizing System (*MNB*) [5,6], Perceptual Analysis Measurement System (*PAMS*) [7], and Perceptual Evaluation of Speech Quality (*PESQ*) [8,9]. In principle, the *PSQM* model is based on comparison of the power spectrum of the corresponding sections of reference and degraded signals. Naturally, auditory processing is involved in this process. The results of this model correlate more with the results of listening tests, in comparison with *MNB* [9]. It seems to be worth to mention that *MNB* uses a slightly different approach as *PSQM*, based on time and frequency measuring normalizing blocks involved in distance measure calculation part. More details about this complicated process can be found in [5]. In essence, *PAMS* model is based on similar approach as *PSQM*, however this model is equipped with better performing time-alignment module than *PSQM*. Let us remind you that good performing time-alignment module is crucial when speech quality of packet-based communications is evaluated. At the present, the *PESQ* model is frequently used. *PESQ* combines merits of *PAMS* and *PSQM99*

---

* Corresponding author.
E-mail addresses: pocta@fel.uniza.sk (P. Počta), holubjan@fel.cvut.cz (J. Holub).

0920-5489/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csi.2013.07.001

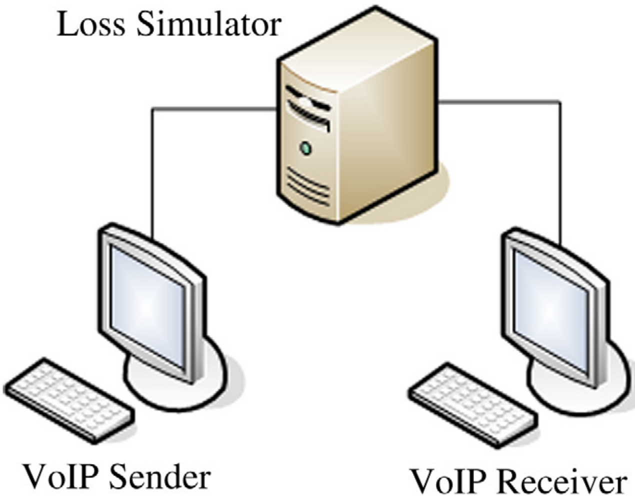

Fig. 1. Experimental scenario.

(an updated version of *PSQM*), and adds new methods for transfer function equalization and averaging distortions over time. *PESQ* also facilitates with very fine time-alignment. It can be used in wider range of network conditions, and gives higher correlation with subjective results than the other objective models [8–10]. Unlike the conversational model (for instance model proposed in [11]), *PESQ* is a listening-only model; the degraded sample is time-aligned with the reference sample during pre-processing. Predictions provided by *PESQ* do not reflect the effects of delay on speech quality. The disadvantages of this model include higher calculation load in comparison to *PSQM* and *MNB* which is caused by recursions in the model and inaccuracy when used in conjunction with some variables, like sidetone, loudness loss, effect of delay in conversational tests, talker echo and different listening levels as standard listening level of 79 dB SPL, extreme temporal clipping (for more details, see Table 2 in [10]). Among the models mentioned above, *PSQM* and *PESQ* were standardized by the *ITU-T* Recommendations such as *P.861* [12] and *P.862* [10] respectively. Moreover, *ITU-T SG12* has very recently standardized a new intrusive model called Perceptual Objective Listening Quality Assessment (*POLQA*) that will probably replace *PESQ* for some conditions mainly related to wideband and super-wideband speech transmission. Those conditions will be known when the characterization phase of this algorithm will be finished. In contrast to intrusive methods, the idea of the single-ended (non-intrusive) signal-based methods is to generate an artificial reference (i.e., an “ideal” undistorted signal) from degraded speech signal and to use this reference in a signal-comparison approach. Once a reference is available, a signal comparison similar to that of *PESQ* can be performed. The result of this comparison can further be modified by a parametric degradation analysis and integrated into an assessment of overall quality. The most widely used non-intrusive models include Auditory Non-Intrusive QUality Estimation (*ANIQUE*) [13] and internationally standardized *P.563* [14,15]. The former model utilizes the temporal envelope representation of speech and is based on the functional roles of human auditory systems and the characteristics of human articulation systems. The latter one is essentially based on models of voice production and perception and demonstrates better performance than *ANIQUE* model on more than 48 subjective experiments representing most distortions that occur on voice networks [16]. 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* [17] and is primarily used for transmission planning purposes. This 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.

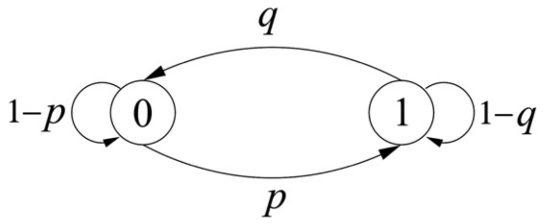

Fig. 2. Gilbert model.

Table 1
Activity parameters and numbers of active speech periods of the reference signals.

Reference signal	Activity par. of 42%	Activity par. of 62%	Activity par. 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 characteristics of reference signals for objective speech quality assessment provided by *PESQ* and *POLQA* are defined in *ITU-T* Recommendation *P.862.3* [18] and in brand new *ITU-T* Recommendation *P.863* [19], respectively. Those characteristics are crucial for speech quality assessment and can rapidly influence accuracy and reliability of results of the evaluation process, regardless of subjective or objective assessment. In particular, two reference signal characteristics are defined very broadly by both documents mentioned above from our point of view, namely the length of reference signal and speech activity parameter. The first one is clear and therefore we do not see any need to define this characteristic. The latter one is more complicated and is defined in the literature as the ratio in percentage points between the amount of active speech and the length of reference signal or sample. The *ITU-T* Recommendation *P.862.3* recommends using the reference signals in duration in the range from 8 s to 30 s for the purpose of *PESQ*'s measurements. The speech activity in the reference signals (speech activity parameter), which can be measured according to *ITU-T* Recommendation *P.56* [20], should be between 40% and 80% of their length. On the other hand, the new *ITU-T* Recommendation *P.863* provides the following information with regard to speech activity parameter: “The reference file has at least 40% activity and consists of at

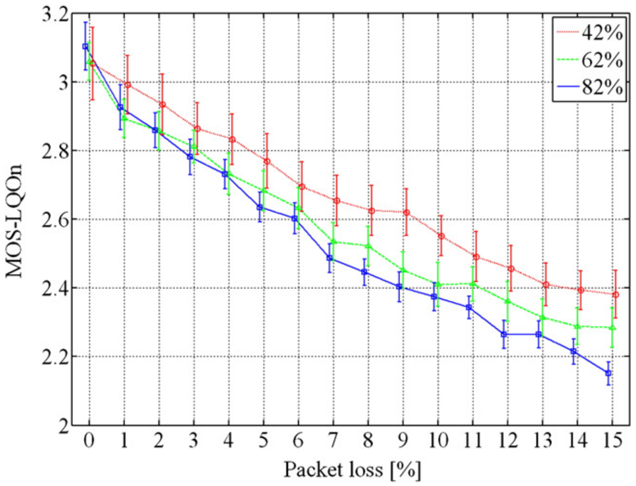

Fig. 3. *MOS-LQOn* as a function of packet loss for different values of activity parameter in case of independent losses. The vertical bars show 95% *CI* (derived from 80 measurements) for each loss. Table 8 (Appendix A.3.1) shows detailed *MOS-LQOn* values presented in this figure.

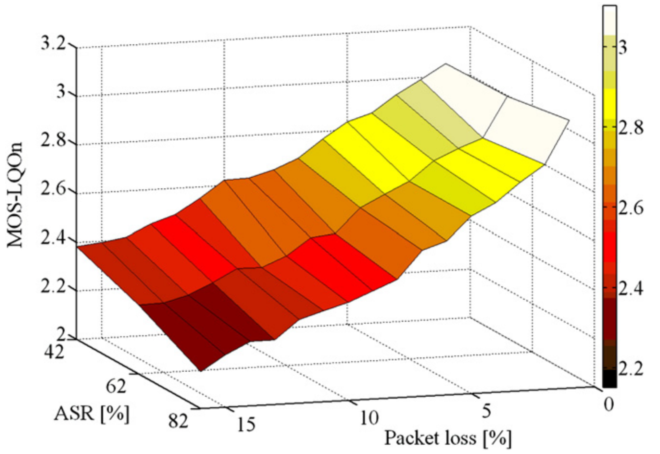

Fig. 4. *MOS-LQOn* versus packet loss and activity parameter for independent losses. Table 8 (Appendix A.3.1) shows detailed *MOS-LQOn* values presented in this figure.

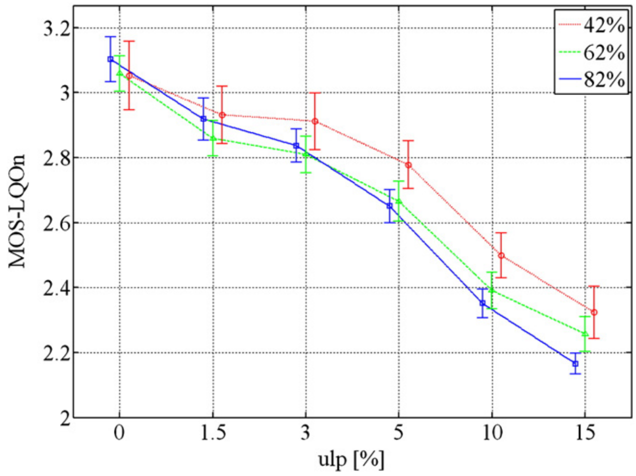

Fig. 6. *MOS-LQOn* as a function of unconditional loss probability for different values of activity parameter in case of dependent losses ($clp = 30\%$). Other detailed descriptions of Fig. 3 apply appropriately. Table 10 (Appendix A.3.2) shows detailed *MOS-LQOn* values presented in this figure.

least two sentences. The total amount of silence is split into at least two sections (typically three). Not more than 50% of the silence falls before the start or after the end of the file." Regarding the length of reference signal used for speech quality assessment provided by *POLQA*, the values defined for *PESQ* were not changed in this case. We suppose that those two characteristics (length of reference signal and speech activity parameter) can have an impact on final *PESQ*'s as well as *POLQA* predictions in presence of time-varying impairments. The investigation of both characteristics from *PESQ*'s prediction perspective has been done in [21]. As expected, both characteristics have an impact on speech quality assessed by test subjects (subjective test) as well as by *PESQ* model (objective model), when time-varying impairments, like packet loss are present. Some very important issues raised from [21] especially in the case of speech activity experiment. That is the reason for exhaustive investigation of the impact of activity parameter on speech quality predictions provided by *PESQ* from dependent and independent loss perspectives. Unfortunately, we cannot also investigate the behavior of *POLQA* with respect to this characteristic at this moment because *POLQA* model will be released for organizations participating in characterization phase at the beginning of this phase. We leave this point for further investigation, as can be also seen in Section 4 of this paper. One can argue why this analysis is going to be mainly done by *PESQ*, instead of using listening test results. We have several reasons to deploy *PESQ* as a useful tool for this analysis. Firstly, *PESQ* is recommended by *ITU-T* to assess speech quality in wide range of network conditions. Moreover, currently this algorithm is mainly used for such measurements around the globe and widely accepted by community for this purpose because of its very good accuracy in recommended conditions. Secondly, we know that *PESQ* behaves a little bit differently in case of such loss than human brain. Human brain tries to render the lost information on a basis of content heard. In other words, he attempts to predict the lost information (syllables, words) by the content of sentence processed. *PESQ* as well as the other objective tools are not able to do such prediction. On the other hand, when *PESQ* is supported by Packet Loss Concealment (*PLC*) algorithm (the lost information is concealed), its behavior is much closer to the behavior of human brain in such a case. Of course, there are some limitations, especially for extreme loss conditions when big parts of words or whole words get lost. In such conditions, *PLC* is not able to conceal the lost information properly and finally *PESQ* is not able to behave like human brain. Aforementioned extreme loss conditions will not be involved in this experiment. Thirdly, speech activity parameter is defined in [18] as one of the important input parameters, which mainly describe the characteristics of the used reference signals. As also mentioned at the beginning of this paragraph, this reference signal characteristic is defined very broadly by [18]. By this investigation, we would like to prove that this parameter have a big impact on final *MOS*-scores, provided by objective measures (like *PESQ*) as well as by listening tests, especially when the time-varying impairments, like packet loss are

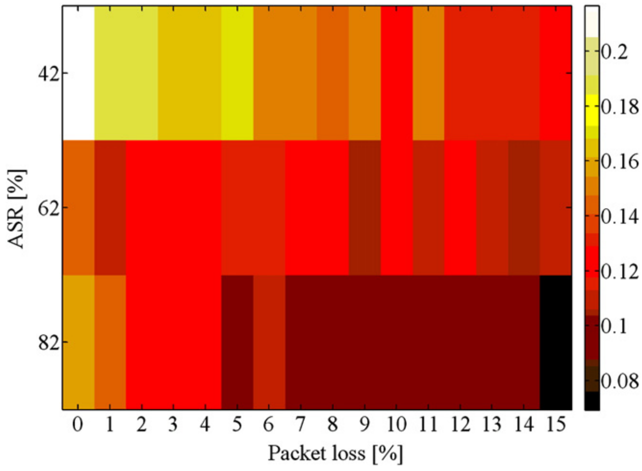

Fig. 5. *MAD* of *MOS-LQOn*'s at each point of loss space and activity parameter in case of independent losses. Table 9 (Appendix A.3.1) presents detailed *MAD* values presented in this figure.

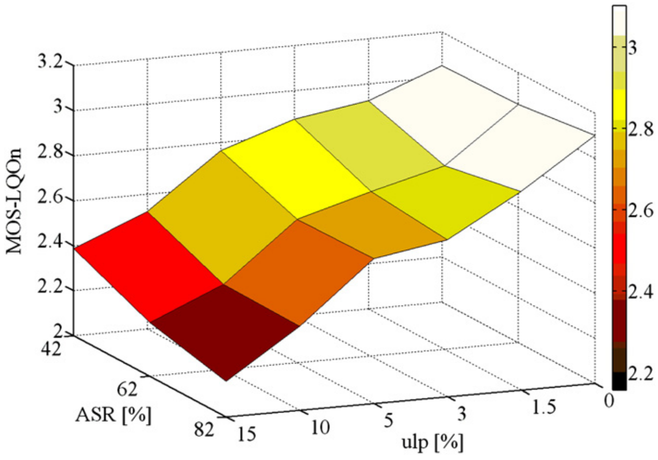

Fig. 7. *MOS-LQOn* versus unconditional loss probability and activity parameter for dependent losses ($clp = 50\%$). Table 11 (Appendix A.3.2) shows detailed *MOS-LQOn* values presented in this figure.

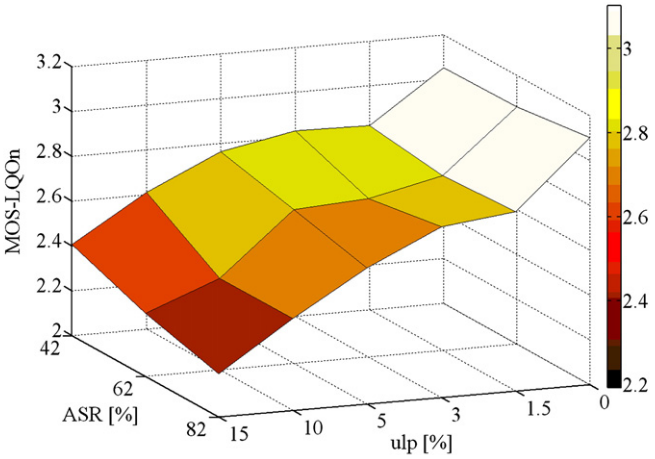

Fig. 8. *MOS-LQOn* versus unconditional loss probability and activity parameter for dependent losses (*clp* = 15%). Table 12 (Appendix A.3.2) shows detailed *MOS-LQOn* values presented in this figure.

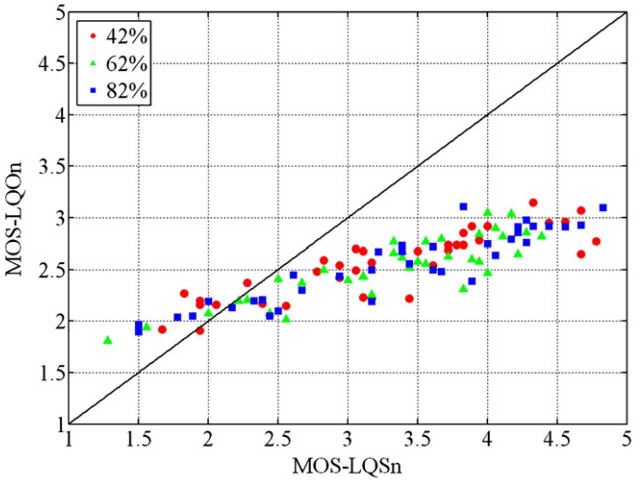

Fig. 10. Subjective results (*MOS-LQSn*) versus *MOS-LQOn* values (non-regressed). Tables 14–16 (Appendix A.4) show detailed values of *MOS-LQSn* and *MOS-LQOn* (non-regressed) for all tested conditions and all investigated activity parameters.

taken into account. Naturally, we will verify our objective results by listening test to prove that *PESQ* behaves correctly for different activity parameters as well as to evaluate an accuracy of its predictions. Fourthly, *PESQ* enables us to do very comprehensive study of the impact of this parameter on *MOS*-scores, which probably would not be feasible by doing only listening tests, because of known limitations of such tests (e.g. duration of test, number of samples presented without subjects fatigue). Finally, we believe that all aforementioned reasons (high acceptability and accuracy, comprehensive study) allow us to deploy this algorithm in this case.

Some works have been carried out on the study of *PESQ*'s behavior under single frame, uniform and dependent losses. In [22], 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 [23], an investigation on 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 [24], 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. Moreover, the behavior of *PESQ* algorithm with regard to different speech activity parameters in presence of receiver-side comfort noise has been examined in [25]. It has to be noted that environment including receiver-side comfort noise is very different from environment investigated in this paper and naturally also creates additional speech quality impairments in comparison to the packet loss occurred in this investigation (very different results obtained, for more details, see [25]). In particular, here we focus on an impact of activity parameter on speech quality predictions provided by *PESQ* in case of independent and dependent losses. The reference signals with activity parameters of 42, 62 and 82% are investigated in this study. In addition, we assess the variability of *PESQ*'s predictions with respect to speech activities and loss conditions as well as its accuracy, by comparing the predictions with subjective assessments.

The rest of the paper is organized as follows: Section 2 introduces experimental scenario and experiments carried out in this study. In

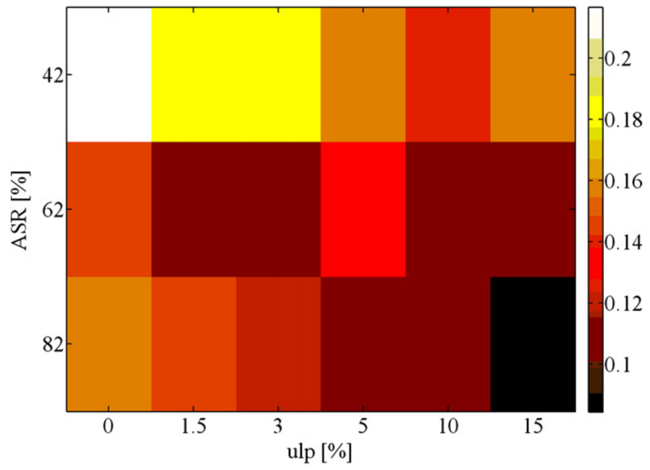

Fig. 9. *MAD* of *MOS-LQOn*'s at each point of loss space and activity parameter in case of dependent losses (*clp* = 30%). Table 13 (Appendix A.3.2) presents detailed *MAD* values presented in this figure.

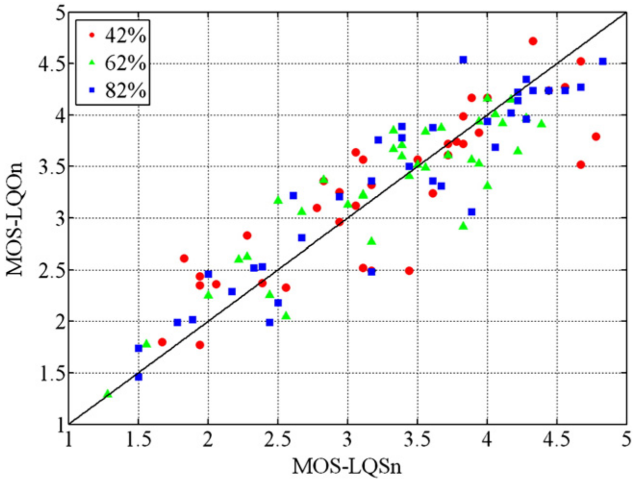

Fig. 11. Subjective results (*MOS-LQSn*) versus *MOS-LQOn* values (2nd order regression). Tables 14–16 (Appendix A.4) show detailed values of *MOS-LQSn* and *MOS-LQOn* (regressed) for all tested conditions and all investigated activity parameters.

Table 2
Pearson correlation coefficient and Root Mean Square Error between *MOS-LQSn* and *MOS-LQOn* before and after 2nd order regression.

Activity parameter	42%	62%	82%
ρ before regression	0.8607	0.8668	0.9214
ρ after regression	0.8608	0.8845	0.9285
δ before regression	0.9397	0.9537	1.0138
δ after regression	0.4462	0.3620	0.3530

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 (Fig. 1). In the case of loss simulator, two currently most widely used models were 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. For this experiment the *ITU-T G.729AB* encoding scheme [26] was chosen. This codec is one of the most popular narrowband voice compression algorithms used in *VoIP* since it operates with near toll quality using a very limited bandwidth—8 kbps. In the measurements, two frames (frame length = 10 ms) were encapsulated into a single packet; thus corresponding to a packet size of 20 ms. Adaptive jitter buffer, *G.729AB*'s native *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 the case of this experiment because of small constant jitter inserted by the loss simulator during the measurement. The Comfort Noise Generator (*CNG*) usage was disabled in the 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:

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

where *x* and *y* represent the raw *PESQ* score and the mapped *MOS-LQOn*, respectively. The equation mentioned is defined in *ITU-T* Recommendation *P.862.1* [27]. This mapping function was defined by *ITU-T*'s Study Group 12 to resolve the problem related to systematic deviation between the *MOS*-scores obtained in auditory listening tests (*MOS-LQS*: Listening Quality Subjective) and those derived from the signal-based measure (*PESQ*). Finally, the applying of such mapping function allows a linear comparison of objective results with subjective *MOS* scores, as described in [27]. In case of *PESQ* score calculation, we used some batch data processing techniques proposed in [28] for bulk data processing.

Table 3
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
Activity parameter (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			

Table 4
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
Activity parameter (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			

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 [29,30] have been proposed for modeling network losses, the currently most widely used of them will be briefly discussed in the following subsections.

2.2.1. Bernoulli model

The Bernoulli model is a model for a random process that consists of Bernoulli trails. A Bernoulli trial is an experiment whose result is random and can be either of two possible outcomes, loss (packet dropped) and no loss (packet received). In the Bernoulli model, each packet loss is independent (memoryless), regardless of whether the previous packet was lost or not. In other words, each experiment is independent of previous trails. In this case, there is only one parameter describing loss process, namely the average packet loss rate (P_{pl}), which can be mathematically described by the following formula:

$$P_{pl} = \frac{n_1}{n}100 \tag{2}$$

where n_1 is the number of lost packets and n is 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 [29–31]. In 2-state Gilbert model as shown in Fig. 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. When $p = 1-q$, the 2-state Gilbert model reduces to a Bernoulli model. $1-q$ is also referred to as the conditional loss probability (*clp*). The probability of being in State 1 is referred to as the unconditional loss probability (*ulp*). The *ulp* provides a measure of the average packet loss rate and is given by [32]:

$$ulp = \frac{p}{p+q}. \tag{3}$$

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

Sixteen independent loss and dependent loss conditions were chosen to cover all cases of interest. They consist of combinations of

Table 5
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
Activity parameter (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 6
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
Activity parameter (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			

packet loss rates (from 0% to 15%) in case of independent losses and unconditional loss probabilities (*ulp*, 0%, 1.5%, 3%, 5%, 10% and 15%), conditional loss probabilities (*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 signal selections should follow the criteria given by *ITU-T* Recommendations *P.830* [33] and *P.800* [1]. The reference signals should include active speech periods separated by silence periods, and are normally of 1–3 s 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 s with activity parameters of 42, 62 and 82% were applied. All reference signals used for this investigation were spoken by the same people (as defined in Table 1), also for different values of activity parameter. The differences between the used reference signals are only in case of number of active speech periods (sentences), resulting in different activity parameters. In case of higher values of activity parameter, the new sentences were added, as an extension. New sentences added also upgraded the distributions of speech and non-speech periods; in particular they changed the parts of the previous non-speech periods into speech periods.

The decision about using reference signals in length of 30 s came from our previous published work [21]. 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 speech activity parameter on speech quality predictions, assessed by *PESQ*. The long reference signal usage for the speech quality assessment by *PESQ* has been also investigated in [34]. 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 s. The results of this work have been included in *ITU-T* Recommendation *P.862.3*.

The activity parameters and numbers of active speech periods for each of the used reference signals are presented in Table 1. The activity parameter 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 [35], known as *sv56*.

Table 7
Summary of *ANOVA* conducted on the *MOS-LQSn*'s in case of dependent losses ($clp = 30\%$).

Effect	SS	df	MS	*F*	p^*
ulp (1)	1158.76	5	231.751	285.39	0.0000
Activity parameter (2)	0.7	2	0.349	6.80	0.0011
(1)∗(2)	49.19	10	4.919	6.06	0.0000
Error	1564.04	1926	0.812		
Total	2772.68	1943			

Table 8
Average *MOS-LQOn* values obtained for independent losses.

Packet loss [%]	Activity par. of 42%	Activity par. of 62%	Activity par. of 82%
0	3.05	3.06	3.10
1	2.99	2.89	2.93
2	2.93	2.86	2.86
3	2.86	2.81	2.78
4	2.83	2.73	2.73
5	2.77	2.68	2.63
6	2.69	2.63	2.60
7	2.65	2.53	2.49
8	2.63	2.52	2.45
9	2.62	2.45	2.40
10	2.55	2.41	2.37
11	2.49	2.41	2.34
12	2.46	2.36	2.27
13	2.41	2.31	2.27
14	2.39	2.29	2.21
15	2.38	2.28	2.15

2.4. Subjective assessment

The subjective listening tests were performed in accordance to *ITU-T* Recommendation *P.800* [1]. Always up to 8 listeners were seated in listening chamber with reverberation time less than 190 ms and background noise well below 20 dB SPL (A). All together, 18 listeners in the age of 19–30 years participated in the tests, the number of male and female listeners being balanced.

The samples were played out using high quality studio equipment in random order. Results in Opinion Score 1 to 5 were averaged to obtain *MOS*-Listening Quality Subjective narrowband (*MOS-LQSn*) values for each sample.

Because of the huge amount of objective measurement data, we had to make the decision which condition is the closest to real network conditions in order to limit the number of samples used in subjective tests. Finally, we decided on the basis of the available measurement results [29,30,32] that one of the dependent loss conditions is the best one for this purpose, namely $clp = 30\%$. The subjective tests were only done for this condition.

All together, 108 speech samples were selected for subjective testing. Always 6 samples represented one network testing condition (the combination of *ulp*'s and *clp*) and activity parameter. The 6 samples mentioned above were composed of 3 male and 3 female samples. In each sample collection, the best, average and worst cases were chosen from speech quality perspective. These were selected out of all recorded samples by expert listening.

Table 9
MAD of *MOS-LQOn* values obtained for independent losses.

Packet loss [%]	Activity par. of 42%	Activity par. of 62%	Activity par. of 82%
0	0.22	0.14	0.16
1	0.18	0.11	0.14
2	0.19	0.12	0.12
3	0.17	0.12	0.12
4	0.17	0.12	0.12
5	0.17	0.13	0.10
6	0.15	0.13	0.11
7	0.15	0.12	0.10
8	0.15	0.13	0.10
9	0.15	0.10	0.10
10	0.12	0.13	0.10
11	0.15	0.11	0.09
12	0.13	0.12	0.09
13	0.13	0.11	0.09
14	0.12	0.10	0.09
15	0.13	0.11	0.07

Table 10
Average *MOS-LQOn* values obtained for dependent losses ($clp = 30\%$).

ulp [%]	Activity par. of 42%	Activity par. of 62%	Activity par. of 82%
0	3.05	3.06	3.10
1.5	2.93	2.86	2.92
3	2.91	2.81	2.84
5	2.78	2.66	2.65
10	2.50	2.39	2.35
15	2.32	2.26	2.17

Table 12
Average *MOS-LQOn* values obtained for dependent losses ($clp = 15\%$).

ulp [%]	Activity par. of 42%	Activity par. of 62%	Activity par. of 82%
0	3.05	3.06	3.10
1.5	2.82	2.78	2.80
3	2.83	2.71	2.76
5	2.76	2.69	2.61
10	2.61	2.41	2.41
15	2.41	2.28	2.19

3. Experimental results

In this section, we describe and explain experimental results for objective assessment and comparison with subjective scores in more details, respectively.

3.1. Experimental results for objective assessment

The measurements were independently performed 80 times (20 different loss locations/patterns and 4 reference signals, see Section 2.2 and 2.3, respectively) under the same packet loss (independent losses), the same pair of *ulp* and *clp* (dependent losses) and the same value of activity parameter. The average *MOS-LQOn* score, 95% Confidence Interval (*CI*) and Mean Absolute Deviation (*MAD*) were calculated. The next subsections provide the detailed description of experimental results for the both examined types of losses.

3.1.1. 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 activity parameter. Figs. 3 and 4 depict differences between investigated activity parameters in speech quality evaluation, provided by *PESQ*. It can be seen from abovementioned figures that the difference in activity parameter has a significant impact on overall speech quality. This fact supports our preliminary assumption that an increasing amount of speech in reference signal (expressed by the activity parameter) has to result in increase of *PESQ* sensitivity to packet loss change. That may be explained by the increase/decrease of information (speech) loss probability at the same packet loss rate in the case of using higher/lower values of activity parameter. It is caused by a greater number of active speech periods in reference signals with higher activity parameter. 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. Naturally, the packet loss is concealed by *PLC* algorithm in all investigated cases but any concealment algorithm is not able to perfectly conceal the lost information which causes the differences between investigated and reference signal in speech quality domain. When higher value of activity parameter is deployed, the number of captured impairments as well as the aforementioned differences is getting higher. 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. The abovementioned effect is depicted in Fig. 3. In more detail, it can be seen from this figure that *MOS-LQOn* for higher value of activity parameter (82%) decreases faster in comparison with the remaining values of this parameter, namely 42% and 62%.

However, it can be seen from Fig. 3, that one bias has been obtained in case of 1% packet loss. It can be explained by a bit more captured losses at active speech periods (effective losses) in case of 62% activity parameter. For instance, in the mentioned case we captured approximately 15 loss events and only one odd effective loss can 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. 5 shows *MAD* of *MOS-LQOn*'s, which has been obtained for this experiment. It can be seen from Fig. 5 that the deviation of predictions is much smaller in case of higher values of activity parameter, especially when network condition degrades (packet loss increase). First fact is related to sensitivity effect mentioned above and second one can 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 value of activity parameter is used, this effect can be even more markedly achieved. This effect can be characterized as the sensitivity effect gain.

A two-way analysis of variance (*ANOVA*) was conducted on *MOS-LQOn*'s using packet loss and activity parameter as fixed factors (Appendix A.1.1, Table 3). We found clearly the highest *F*-ratio for the packet loss ($F = 456.38$, $p^* < 0.01$). Moreover, the activity parameter showed a little bit weaker effect on quality than packet loss, with $F = 153.87$, $p^* < 0.01$. The realized *ANOVA* revealed that speech activity parameter has affected the average *MOS-LQOn* values more weakly than packet loss.

3.1.2. 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 activity parameter. The experimental results for all investigated *clp*'s are depicted in Figs. 6–8. We can observe how speech quality drops, as expected, with both *clp* and *ulp*. Also, it is clear that the different values of activity parameter can seriously influence the quality in case of dependent losses. Obviously, we obtained the same effect as in first case (independent losses). It means that using higher values of activity parameter leads to increase of *PESQ* sensitivity to packet loss change, also in case of dependent losses. One more bias in comparison with first case has been achieved in case of 3% *ulp* (Fig. 6). The reason for that is same as mentioned above.

Table 11
Average *MOS-LQOn* values obtained for dependent losses ($clp = 50\%$).

ulp [%]	Activity par. of 42%	Activity par. of 62%	Activity par. of 82%
0	3.05	3.06	3.10
1.5	2.92	2.82	2.88
3	2.87	2.73	2.70
5	2.76	2.64	2.64
10	2.52	2.38	2.37
15	2.39	2.24	2.16

Table 13
MAD of *MOS-LQOn* values obtained for dependent losses ($clp = 30\%$).

ulp [%]	Activity par. of 42%	Activity par. of 62%	Activity par. of 82%
0	0.22	0.14	0.16
1.5	0.19	0.11	0.15
3	0.19	0.11	0.12
5	0.16	0.13	0.11
10	0.14	0.11	0.10
15	0.16	0.11	0.08

Table 14
MOS-LQSn, *MOS-LQOn* (non-regressed) and *MOS-LQOn* (regressed) values obtained for dependent losses ($clp = 30\%$) and activity parameter of 42%.

ulp [%]	Case	Sample	*MOS-LQSn*	*MOS-LQOn* (non-regressed)	*MOS-LQOn* (regressed)
0	Best	Male	4.33	3.15	4.72
		Female	4.67	3.07	4.52
	Average	Male	4.56	2.96	4.27
		Female	4.78	2.77	3.79
	Worst	Male	4.44	2.95	4.24
		Female	4.67	2.65	3.52
1.5	Best	Male	3.89	2.92	4.17
		Female	3.94	2.78	3.83
	Average	Male	3.72	2.74	3.72
		Female	3.83	2.74	3.72
	Worst	Male	3.06	2.70	3.64
		Female	3.61	2.54	3.24
3	Best	Male	4.00	2.92	4.17
		Female	3.78	2.74	3.74
	Average	Male	3.11	2.68	3.57
		Female	3.72	2.69	3.61
	Worst	Male	2.83	2.59	3.36
		Female	2.94	2.42	2.96
5	Best	Male	3.83	2.85	3.99
		Female	3.50	2.68	3.57
	Average	Male	3.17	2.57	3.32
		Female	2.94	2.54	3.25
	Worst	Male	3.06	2.49	3.12
		Female	3.17	2.22	2.49
10	Best	Male	2.78	2.48	3.10
		Female	3.11	2.23	2.52
	Average	Male	1.83	2.27	2.61
		Female	3.44	2.22	2.49
	Worst	Male	2.06	2.16	2.36
		Female	1.94	1.91	1.77
15	Best	Male	2.28	2.37	2.83
		Female	1.94	2.20	2.44
	Average	Male	2.39	2.17	2.37
		Female	2.56	2.15	2.33
	Worst	Male	1.94	2.16	2.35
		Female	1.67	1.92	1.80

Table 15
MOS-LQSn, *MOS-LQOn* (non-regressed) and *MOS-LQOn* (regressed) values obtained for dependent losses ($clp = 30\%$) and activity parameter of 62%.

ulp [%]	Case	Sample	*MOS-LQSn*	*MOS-LQOn* (non-regressed)	*MOS-LQOn* (regressed)
0	Best	Male	4.00	3.05	4.16
		Female	4.22	2.65	3.65
	Average	Male	3.94	2.84	3.94
		Female	3.94	2.58	3.53
	Worst	Male	3.89	2.60	3.57
		Female	3.83	2.31	2.92
1.5	Best	Male	4.17	3.04	4.15
		Female	3.67	2.80	3.88
	Average	Male	4.11	2.82	3.92
		Female	4.00	2.47	3.31
	Worst	Male	3.56	2.77	3.84
		Female	3.17	2.26	2.77
3	Best	Male	4.28	2.86	3.97
		Female	4.06	2.90	4.01
	Average	Male	3.39	2.68	3.71
		Female	3.72	2.63	3.61
	Worst	Male	3.44	2.52	3.41
		Female	3.39	2.62	3.60
5	Best	Male	3.33	2.77	3.85
		Female	4.39	2.82	3.91
	Average	Male	3.56	2.56	3.49
		Female	3.50	2.58	3.52
	Worst	Male	2.67	2.37	3.06
		Female	3.11	2.43	3.22
10	Best	Male	3.33	2.66	3.67
		Female	2.83	2.50	3.37
	Average	Male	3.00	2.40	3.13
		Female	3.11	2.44	3.23
	Worst	Male	2.28	2.21	2.63
		Female	2.56	2.02	2.05
15	Best	Male	2.50	2.41	3.17
		Female	2.44	2.08	2.26
	Average	Male	2.22	2.20	2.60
		Female	2.00	2.08	2.25
	Worst	Male	1.56	1.94	1.78
		Female	1.28	1.81	1.29

In Fig. 9, we can see the *MAD* of *MOS-LQOn*'s for 30% *clp*. Unsurprisingly, *PESQ*'s predictions deviation 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 can explain this phenomenon. Naturally, that is a point for a future investigation because exhaustive study is needed to validate, and interpret this phenomenon. On the basis of those results, we can pronounce that higher values of activity parameter usage may lead to *PESQ*'s prediction deviation decreasing, 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 activity parameter as fixed factors (Appendix A.1.2, Tables 4–6). In principle, we obtained similar results as in the case of independent losses. However, the smaller impact of activity parameter (expressed by *F*-ratio) was obtained for dependent losses (increased by higher burstiness (higher *clp* values)), see Tables 3 and 4–6. It means that impact of packet loss is much more dominant for dependent losses, probably due to poorer performance of packet loss concealment algorithm under bursty losses, as widely reported in scientific papers (for instance in [36]).

3.2. Comparison with subjective scores

As mentioned above, the subjective tests were only realized for dependent losses ($clp = 30\%$). The results obtained by means of subjective testing (*MOS-LQSn*) are compared with *MOS-LQOn* results in Figs. 10 and 11. Obviously, the sensitivity of PESQ to activity parameter modification is a bit weaker than that of human subjects (see Fig. 10). As the attempt to use 3rd order regression (as recommended in *ITU-T* Recommendation *P.862*) has lead to non-monotonic results, the 2nd order regression was used instead. Fig. 11 depicts the results after the 2nd order regression.

The *PESQ*'s performance from activity parameter perspective is characterized by the Pearson correlation coefficient ρ and Root Mean Square Error (*RMSE*) δ. The statistics for ρ and δ are summarized in Table 2. It is seen from this table that higher values of activity parameter usage may lead to higher correlation with subjective results as well as to *PESQ*'s prediction accuracy improvement (expressed by Pearson correlation coefficient and *RMSE* (after regression)). The prediction accuracy can be also increased by higher packet losses, as can be seen from Fig. 11. The results obtained in the case of smaller *MOS* values are mainly closer to diagonal line than higher *MOS* are. In general, high packet losses generate lower *MOS* scores.

The two-way *ANOVA* was also conducted on *MOS-LQSn*'s using *ulp* and activity parameter as fixed factors (Appendix A.2, Table 7). We got clearly the highest *F*-ratio for the *ulp* ($F = 285.39$, $p^* < 0.01$). Moreover, the activity parameter showed a little bit weaker effect on quality than *ulp*, with $F = 6.8$, $p^* < 0.01$. In comparison with the same test for objective data, a bit higher *F*-ratios were obtained for the *ulp* as well as for the activity parameter (Appendix A.1.2, Table 5). On the other hand, the same *F*-ratio was observed in case of an interaction of both parameters. Based on similar *F*-ratios, we can say that the results of this *ANOVA* test confirmed the sensitivity effect defined in Section 3.1 but a bit weaker than what were obtained in objective assessment. It means that human brain is a bit less critical to loss of some parts of words than *PESQ* supported by

Table 16
MOS-LQSn, *MOS-LQOn* (non-regressed) and *MOS-LQOn* (regressed) values obtained for dependent losses (*clp* = 30%) and activity parameter of 82%.

ulp [%]	Case	Sample	*MOS-LQSn*	*MOS-LQOn* (non-regressed)	*MOS-LQOn* (regressed)
0	Best	Male	3.83	3.11	4.54
		Female	4.83	3.10	4.52
	Average	Male	4.44	2.92	4.24
		Female	4.56	2.91	4.24
	Worst	Male	4.33	2.92	4.24
		Female	4.22	2.91	4.22
1.5	Best	Male	4.22	2.86	4.14
		Female	4.28	2.98	4.35
	Average	Male	3.39	2.73	3.89
		Female	4.28	2.76	3.96
	Worst	Male	4.06	2.64	3.69
		Female	4.00	2.75	3.94
3	Best	Male	4.17	2.79	4.02
		Female	4.67	2.93	4.27
	Average	Male	3.22	2.67	3.76
		Female	3.39	2.68	3.78
	Worst	Male	3.61	2.50	3.36
		Female	3.44	2.56	3.50
5	Best	Male	3.61	2.72	3.88
		Female	3.67	2.48	3.31
	Average	Male	3.17	2.50	3.36
		Female	3.89	2.39	3.06
	Worst	Male	2.61	2.45	3.22
		Female	3.17	2.19	2.48
10	Best	Male	2.50	2.10	2.18
		Female	2.67	2.30	2.81
	Average	Male	2.39	2.21	2.53
		Female	2.33	2.20	2.52
	Worst	Male	2.94	2.44	3.21
		Female	2.44	2.05	1.99
15	Best	Male	2.00	2.19	2.46
		Female	2.17	2.13	2.29
	Average	Male	1.89	2.05	2.02
		Female	1.78	2.04	1.99
	Worst	Male	1.50	1.97	1.74
		Female	1.50	1.90	1.46

PLC algorithm. It might be due to the cognitive load involved in rendering process realized by human brain to predict the lost parts of words. The subject concentration on the quality can partly be relaxed by this parallel task (prediction of lost information).

On the basis of this comparison, we can pronounce that the subjective tests confirm the objective experimental results (presented in Section 3.1) but the behavior of *PESQ* should be modified to better model the impact of the investigated reference signal characteristic on speech quality.

The experimental results show that the change of activity parameter has a significant impact on overall speech quality in presence of packet loss. This fact is our motivation for deriving activity parameters for typical conversation scenarios frequently occurring in current telecommunication networks. Naturally, an issue of activity parameter setup with regard to different conversation scenarios is also open for discussion. The exact values of activity parameter for typical conversation scenarios 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 outlook

This paper has investigated an impact of the activity parameter 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 the behavior of *PESQ*'s predictions under different activity parameters as well as to assess their accuracy by comparing the predictions with subjective assessments.

The results presented in the paper have approved our hypothesis that an increase in amount of speech in the reference signal (expressed by the activity parameter) may result in an increase of *PESQ* sensitivity to packet loss change as well as *PESQ*'s prediction accuracy improvement. Both effects previously mentioned can be explained by higher probability of losses obtained at active speech intervals (effective loss probability), which have not been properly concealed by *PLC* algorithm. In addition, prediction accuracy can be even increased by higher packet losses. This effect can be characterized as the sensitivity effect gain. More specifically, higher number of effective losses may lead to small variation in a quality score predicted by *PESQ* representing a certain network condition. Moreover, we can pronounce (on the basis of *ANOVA* results) that human brain is a bit less critical to loss of some parts of words than *PESQ* supported by *PLC* algorithm. It might be due to the cognitive load involved in rendering process realized by human brain to predict the lost parts of words. The subject concentration on the quality can partly be relaxed by this parallel task (prediction of lost information).

The results have some implications for the designers of *PESQ* model. As mentioned before, the activity parameter has an impact on speech quality in presence of packet loss. This suggests that this characteristic has to be defined more precisely, especially for time-varying impairments like packet loss. Moreover, it seems that the exact values of this parameter derived for frequently occurring conversation scenarios in current networks should be helpful to provide for more reliable speech quality assessment in such conditions (see outlook at the end of conclusion section). On the other hand, the results also show that human brain is a bit less sensitive to loss of some parts of words than *PESQ*. This implies that the behavior of cognitive model employed in *PESQ* should be altered to better model this effect. This can be done by retraining current cognitive model employed in *PESQ* on databases including impact of activity parameter on speech quality in presence of packet loss. It should be noted that this process is very complex and requires an involvement of more parties, due to availability of the databases for different languages mentioned above. Naturally, this process can have a positive impact on the accuracy of *PESQ*'s predictions. Finally, the presented findings can be also used for improving the performance of brand new *ITU-T* intrusive model for predicting speech quality, namely *POLQA* model.

A future work will focus towards the following issues. At first, we intend to investigate the performance of brand new *ITU-T* intrusive model for predicting speech quality, namely *POLQA* under conditions investigated here (as a part of characterization phase of this model). Secondly, we will attempt to derive the activity parameters for typical conversation scenarios frequently occurring in current telecommunication networks. Some preliminary results in this regard have been already published in [37]. Apparently, this point could be very interesting for other speech quality laboratories around the world. By this process, we might refine on the existing broadly recommended activity parameters, defined by *ITU-T* Recommendations *P.862.3* and *P.863* and provide for more reliable speech quality assessment, especially with regard to time-varying impairments. Thirdly, we plan to exhaustively study the highest *MAD* at 0% packet loss and find out the reason for that.

5. Acknowledgments

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.

Appendix A

A.1. ANOVA for objective assessment

In the next subsections, the detailed results of the analysis of variance (*ANOVA*) conducted on *MOS-LQOn* for independent and dependent losses can be found.

A.1.1. Independent losses

Table 3 provides the results of *ANOVA* carried out on the independent losses test results (dependent variable: *MOS-LQOn*) described in more details in Section 3.1.1.

A.1.2. Dependent losses

In Tables 4–6, the results of *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 3.1.2.

A.2. ANOVA for subjective assessment

Table 7 provides the results of *ANOVA* carried out on the dependent losses test results (dependent variable: *MOS-LQSn*) described in more details in Section 3.2.

A.3. Detailed results obtained for objective assessment

In the next subsections, detailed results displayed in figs. of Section 3.1 can be found.

A.3.1. Independent losses

Table 8 provides average *MOS-LQOn* values displayed in Figs. 3 and 4 (Section 3.1.1). Table 9 shows *MAD* of *MOS-LQOn* values displayed in Fig. 5 (Section 3.1.1).

A.3.2. Dependent losses

In Tables 10–12, average values of *MOS-LQOn* obtained for the dependent losses and all investigated *clp*'s are shown. The presented values are displayed in Figs. 6–8 of Section 3.1.2. In addition, Table 13 presents *MAD* of *MOS-LQOn* values displayed in Fig. 9 (Section 3.1.2).

A.4. Detailed results obtained for comparison with subjective scores

Tables 14–16 provide *MOS-LQSn*, *MOS-LQOn* (non-regressed) and *MOS-LQOn* (regressed) values for all tested conditions and all investigated activity parameters displayed in Figs. 10–11 (Section 3.2).

References

[1] ITU-T Rec., P.800: Methods for Subjective Determination of Transmission Quality, International Telecommunication Union, Geneva (Switzerland), 1996.

[2] Handbook of practical procedures for subjective testing, International Telecommunication Union, Geneva (Switzerland), 2011.

[3] S. Moeller, A. Raake, Telephone speech quality prediction: towards network planning and monitoring models for modern network scenarios, Speech Comm. (ISSN: 0167-6393) 38 (1–2) (2002) 47–75.

[4] J.G. Beerends, J.A. Stemerdink, A perceptual speech quality measure based on a psychoacoustic sound representation, J. Audio Eng. Soc. (ISSN: 1549-4950) 42 (1994) 115–123.

[5] S. Voran, Objective estimation of perceived speech quality — part I: development of the measuring normalizing block technique, IEEE Trans. Audio Speech Lang. Process, 7, 1999, pp. 371–382, (ISSN 1063-6676).

[6] S. Voran, Objective estimation of perceived speech quality — part II: evaluation of the measuring normalizing block technique, IEEE Trans. Audio Speech Lang. Process 7 (1999) 383–390, (ISSN 1063-6676).

[7] A.W. Rix, M.P. Hollier, The perceptual analysis measurement system for robust end-to-end speech quality assessment, Proceedings of IEEE ICASSP 2000, Istanbul (Turkey), 3, 2000, pp. 1515–1518.

[8] A.W. Rix, M.P. Hollier, A.P. Hekstra, J.G. Beerends, Perceptual evaluation of speech quality (PESQ) — the new ITU standard for objective measurement of perceived speech quality, part I — time-delay compensation, J. Audio Eng. Soc. (ISSN: 1549-4950) 50 (2002) 755–764.

[9] J.G. Beerends, A.P. Hekstra, A.W. Rix, M.P. Hollier, J.G. Beerends, A.P. Hekstra, A.W. Rix, M.P. Hollier, Perceptual evaluation of speech quality (PESQ) — the new ITU standard for objective measurement of perceived speech quality, part II — psychoacoustic model, J. Audio Eng. Soc. (ISSN: 1549-4950) 50 (2002) 765–778.

[10] ITU-T Rec., P.862: Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs, International Telecommunication Union, Geneva (Switzerland), 2001.

[11] M. Gueguin, R. Le Bouquin-Jeannes, G. Faucon, V. Gautier-Turbin, V. Barriac, A step further to objective modeling of conversational quality, Proceedings of 14th European Signal Processing Conference (EUSIPCO 2006), Florence (Italy), September 4-8 2006.

[12] ITU-T Rec., P.861: Objective Quality Measurement of Telephone-band (300–3400 Hz) Speech Codecs, International Telecommunication Union, Geneva (Switzerland), 1998.

[13] D.-S. Kim, ANIQUE: an auditory model for single-ended speech quality estimation, IEEE Trans. Audio Speech Lang. Process 13 (5) (2005) 821–831, (ISSN 1063-6676).

[14] L. Malfait, J. Berger, M. Kastner, P.563 — the ITU-T standard for single-ended speech quality assessment, IEEE Trans. Audio Speech Lang. Process (ISSN: 1558-7916) 14 (2006) 1924–1934.

[15] ITU-T Rec., P.563: Single-ended Method for Objective Speech Quality Assessment in Narrow-band Telephony Applications, International Telecommunication Union, Geneva (Switzerland), 2004.

[16] A.W. Rix, J.G. Beerends, D.-S. Kim, P. Kroon, O. Ghitza, Objective assessment of speech and audio quality. Technology and applications, IEEE Trans. Audio Speech Lang. Process (ISSN: 1558-7916) 14 (6) (2006) 1890–1901.

[17] ITU-T Rec., G.107: The E-model: A Computational Model for Use in Transmission Planning, International Telecommunication Union, Geneva (Switzerland), 2009.

[18] 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.

[19] ITU-T Rec., P.863: Perceptual Objective Listening Quality Assessment, International Telecommunication Union, Geneva (Switzerland), 2011.

[20] ITU-T Rec., P.56: Objective Measurement of Active Speech Level, International Telecommunication Union, Geneva (Switzerland), 1993.

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

[22] Ch. Hoene, E. Dulamsuren-Lalla, Predicting performance of PESQ in case of single frame losses, Proceedings of MESAQIN 2004 Conference, Prague (Czech Republic), ISBN: 80-01-03017-2, 2004.

[23] L. Sun, E.C. Ifeachor, Subjective and objective speech quality evaluation under bursty losses, Proceedings of MESAQIN 2002 Conference, Prague (Czech Republic), ISBN: 80-01-02515-2, 2002.

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

[25] P. Počta, H. Vlčková, Z. Polková, J. Holub, Impact of different active-speech-ratios on PESQ's predictions in case of independent and dependent losses (in presence of receiver-side comfort-noise), Radioengineering (ISSN: 1210-2512) 19 (1) (April 2010) 79–88, (Part 10).

[26] 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.

[27] ITU-T Rec., P.862.1: Mapping Function for Transforming P.862 Raw Result Scores to MOS-LQO, International Telecommunication Union, Geneva (Switzerland), 2003.

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

[29] M. Yajnik, S. Moon, J. Kurose, D. Towsley, Measurement and modelling of the temporal dependence in packet loss, Proceedings of Conference IEEE INFOCOM 99 Conference, New York (USA), 1, 1999, pp. 345–352.

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

[31] H. Sanneck, N.T.L. Le, Speech property-based FEC for Internet telephony applications, Proceedings of the SPIE/ACM SIGMM Multimedia Computing and Networking Conference, San Jose (USA), 2000, pp. 38–51.

[32] W. Jiang, H. Schulzrinne, Modelling of packet loss and delay and their effect on real-time multimedia service quality, Proceedings of 10th International Workshop Network and Operations System Support for Digital Audio and Video (NOSSDAV), Chapel Hill (USA), 2000.

[33] ITU-T Rec., P.830: Subjective Performance Assessment of Digital Telephone-band and Wideband Digital Codecs, International Telecommunication Union, Geneva (Switzerland), 1996.

[34] A. Takahashi, 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.

[35] ITU-T Rec., G.191: Software Tools for Speech and Audio Coding Standardization, International Telecommunication Union, Geneva (Switzerland), 2005.

[36] L. Sun, G. Wade, B.M. Lines, E.C. Ifeachor, Impact of packet loss location on perceived speech quality, Proceedings of Internet Telephony Workshop (IPtel 2001), New York (USA), 2001.

[37] P. Počta, M. Mrvová, Derivation of Speech Activity Parameter Values in the Context of Speech Quality Testing, Proceedings of MESAQIN 2011 conference, Prague (Czech Republic), ISBN: 978-80-01-04848-1, 2011, pp. 24–30.

Peter Počta was born in 1981, in Nové Zámky, Slovakia. He received his M.S. and Ph.D. degrees from University of Žilina, Faculty of Electrical Engineering, Slovakia in 2004 and 2007, respectively. He is currently an Assistant Professor in the Department of Telecommunications and Multimedia at University of Žilina and is involved with International Standardization through the ETSI TC STQ as well as ITU-T SG12. His research interests include speech, audio and video quality assessment, VoIP, VoWLAN, VoLTE and cross-layer optimization. He has published more than 20 journal and conference papers indexed in most widely accepted bibliographic databases.

Assoc. Prof. Ing. Jan Holub, Ph.D. graduated from the Czech Technical University, Faculty of Electrical Engineering. Since 1999 he works as an assistant professor there. He was awarded by A. Menchetti prize for the best contribution of younger scientist in ISDDMI 98 (10th International Symposium on Development in Digital Measuring Instrumentation, Naples, Faculty of Engineering). His research interests and activities cover AD and DA converters, digital signal processing, psychoacoustics and measurements in telecommunication networks. He was awarded several patents and published more than 30 articles in scientific magazines and 110 contributions for conferences (62% international).