SERBIAN JOURNAL OF ELECTRICAL ENGINEERING
Vol. 16, No. 3, October 2019, 405-417

UDC: 621.395.31/.38:621.316 **DOI: https://doi.org/10.2298/SJEE1903405L**

Model for Speech Signal Quality Estimation in Packet Network of Electricity Supply Industry

Aleksandar Lebl[1], Dragan Mitić[1], Vladimir Matić[1], Žarko Markov[1]

Abstract: In this paper we present the dilemma dealing with the choice of speech segment duration and the calculation of speech signal quality in the packet telephone network of Electricity supply industry (ESI). The analyzed signal is packetized by one packet. The characteristic of this network is that disturbances of long duration in power system produce burst packet loss. The longer speech segments increase the packet delay, but decrease the number of lost packets in one burst, and vice versa. Here we present why it is better to choose speech segments of short duration. Also, we suggest the corrected method for the calculation of speech signal quality in packet network under the influence of long duration disturbances.

Keywords: Delay effects, Electricity supply industry, Internet telephony, Power system transients, Quality of service.

1 Introduction

Some impacts, which were not important in classic telephone technic or were unknown, have the influence on the telephone signal quality in packet telephone networks. The speech signal delay and echo in classic telephone technic were only important in the connections with satellite transmission. The speech signal compression and the packet loss are the impacts unknown in classic telephone network. Their impact in one packet network can be planned and calculated using the so-called E-model [1, 2].

The implemented models in telephone services may be divided into subjective and objective models. Such models are implemented not only for the analysis of packetized speech signals [3], but also for the analysis of video signals [4]. E-model as the computational model for speech signal analysis is one of the objective methods for transmission quality estimation.

Various types of communications in Electricity supply industry (ESI) are analyzed in [5]. Among them voice communication using telephone network is fundamental. Telephone network of ESI can be designed using packet technics [6]. This network has some specific characteristics, as high availability and the

[1]IRITEL a.d., Beograd, Batajnički put 23, 11080 Beograd, Serbia;
E-mails: lebl@iritel.com; mita@iritel.com; vmatic@iritel.com; zmarkov@iritel.com

structure formed in one level. One of this network uniqueness is the great influence of overvoltage disturbances on telecommunication equipment [7]. These disturbances can impair speech signal more than it is usual [8]. In this paper it is presented how we determined the length of speech signal segment, which is packetized using one packet of ESI telephone network. Also, we try to solve the problem of calculating the speech signal quality in the case of random and burst packet loss for the equal speech segments. Some considerations about packet duration are presented in [9], Section 4. The shorter segments increase the probability of burst packet loss and thus the speech signal quality is decreased. Longer speech segments increase packet delay and the speech signal quality is also decreased. This paper is our attempt to resolve this dilemma.

2 Packetization Delay and Speech Signal Quality

Packet delay has great influence on the quality of packetized speech signal. According to E-model [1], speech signal quality dependence of packet delay is expressed as $R(T_d)$, meaning rating factor R as the function of packet delay T_d. The delay consists of more components, which refer to connection send side, network and connection receive side. On send side, which interests us, we distinguish packetization delay, compression delay, look ahead delay, serialization delay. These delay types are fixed and we cannot avoid the first one and the last one, even without compressors. Serialization delay usually has smaller values and the packetization delay is equal to the duration of speech signal segment (or segments) (t_p), which is (are) packetized in one packet. In modern packet telephony it is avoided to send signal by more speech segments in the same packet.

That's why packetization delay, t_p, can last from 10ms to 50ms, which is usually the duration of speech signal segment, placed in one packet. According to [10], **Table 1**, Note 1, it can be concluded that it is difficult to keep network delay in the class 0 bounds, where IP Packet Transfer Delay is IPTD $\leq$ 100ms. In [10], Appendix VII, it is seen that endpoint delay can be few tens of ms. We can conclude that delay from the user (talker) to the user (listener) can be greater than the bound T_d = 160 – 180 ms, where the function $R(T_d)$ changes its slope from the value $\Delta R/\Delta T_d \approx -0.02$/ms (area 1) to the value $\Delta R/\Delta T_d \approx -0.12$/ms (area 2), Fig. 1. (The graph of the function $R = f(T_d)$ presented in Fig. 1 is taken from [11], where the designation "one-way delay" is used for T_d, or from [12], where T_d is called "mouth-to-ear delay").

Example 1: Let us suppose that the sum of total network delay and dejitter buffer delay is 170ms and that one speech segment of 10ms duration is placed in one packet, i.e. let us suppose that packetization delay is 10ms. We can see that the delay $T_{d1} \approx$ 180ms causes speech signal quality degradation from the value R = 94 to the value $R_1 \approx 90$, Fig 1.

Example 2: Let us suppose that packetization delay in the same network is 50ms. Now the total delay is $T_{d2} \approx 220$ms and the speech signal quality drops down to $R_2 \approx 83$, Fig. 1. We can conclude that speech signal quality degradation, $\Delta R = R_1 - R_2$ is caused by the longer duration of speech segment in one packet, i.e., longer packetization delay. In this, the worst case, it can be seen that speech signal quality degradation $\Delta R \approx 7$ is caused by the increase of speech segment duration in one segment from 10ms to 50ms.

As the conclusion of this section we can say that it is very important to decrease packetization delay as much as possible, because so the total delay is decreased and the quality of transmitted packetized speech signal is increased. This conclusion is especially important in the networks where the delay is greater than about 150ms.

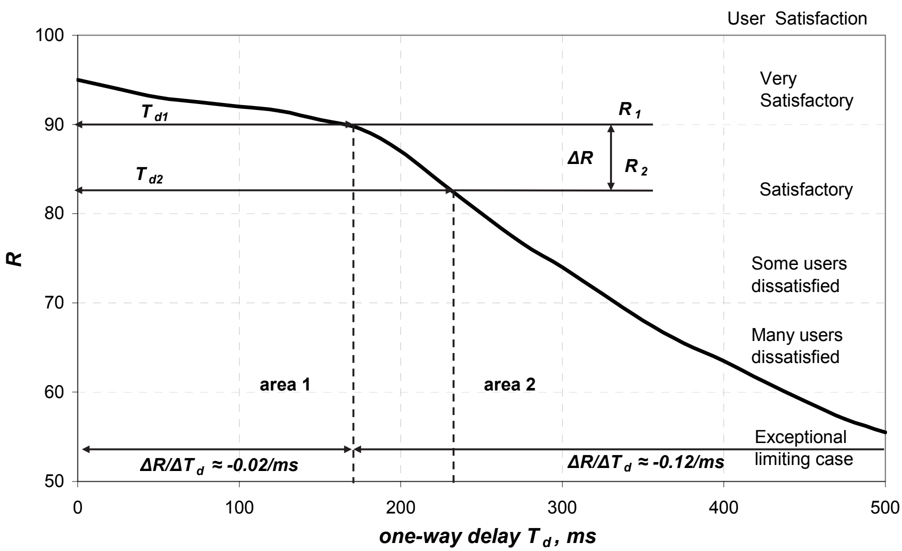

Fig. 1 – *Influence of packet delay on degradation of speech signal quality.*

3 Disturbances in Power System and Packet Loss

Telecommunication network of ESI is under the influence of disturbances in power system. It is presented in [7] that these disturbances can be tens and even hundreds of ms long. That's why packet loss in packet network of ESI appears as burst packet loss. As it is well-known from [1], burst packet loss has the greater influence on degradation of transmitted speech signal quality than the random packet loss when packet loss probability is equal. According to E-model, degradation of speech signal quality at the receive side, caused by the packet loss is expressed by the equation (7 – 29) from [1]:

$$I_{e\text{-}eff} = I_e + \frac{(95 - I_e) \times P_{pl}}{\frac{P_{pl}}{BurstR} + B_{pl}}, \tag{1}$$

where the meaning of the variables is:

– I_e: equipment Impairment Factor; it expresses impairment caused by the use of low bit-rate codec;

– $I_{e\text{-}eff}$: effective Equipment Impairment Factor. It is the value of I_e corrected by the influence of packet loss;

– P_{pl}: packet-loss Probability, expressed in percents;

– *BurstR*: Burst Ratio*,* the factor which expresses the burstiness of packet loss;

– B_{pl}: packet-loss Robustness Factor; it is stated in [1], subsection 3.5 that B_{pl} has codec specific value; some values of B_{pl} are suggested in [13].

For the simplified model where packet loss probability is small and bursty, i.e. if in one group of lost packets are *n* packets, it can be proved that it is $BurstR \approx n$. The influence of packet loss burstiness according to (1) can be seen on the simplified connection model in ESI telephone network as the one presented in [6], where there is no speech signal compression (I_e=*0*) and no packet loss concealment. That's why the equation (1) is simplified to:

$$I_{e\text{-}eff} = \frac{95 \times P_{pl}}{\frac{P_{pl}}{BurstR} + B_{pl}}. \tag{2}$$

Let us calculate degradation of speech signal quality for two models in packet network without compressors and without packet loss concealment.

The first model refers to random packet loss (*BurstR=1*), and the duration of speech signal segment is t_p = 50ms. From (2) we can calculate the following results which are obtained for 3 values of P_{pl}:

– case a1): $\Delta I_{e-eff} \approx 17.9$ for $P_{pl} = 1\%$;

– case a2): $\Delta I_{e-eff} \approx 24.5$ for $P_{pl} = 1.5\%$ and

– case a3): $\Delta I_{e-eff} \approx 30.1$ for $P_{pl} = 2\%$.

In the second model duration of speech segment is t_p = 10ms and the groups of lost packets contain 1, 2, 3, 4 and 5 grouped packets, Fig. 2. In Fig. 2 we present the calculated Effective Equipment Impairment Factor as the function of the number of lost packets in one burst. Calculations are made for the same values as in the first model (P_{pl} = 1%, 1.5% and 2%). The results in

Fig. 2 are obtained from equation (2) using the value B_{pl} = 4.3 from **Table 1.3** in [13], which is valid for coder G.711.

According to this calculation and from Fig. 2, it can be concluded that the degradation of speech signal quality, in the case that n = 5 packets are lost in each group of packets carrying speech signal of 10ms duration, is:

– case b1): $\Delta I_{e-eff} \approx 21.1$ for $P_{pl} = 1\%$;

– case b2): $\Delta I_{e-eff} \approx 30.9$ for $P_{pl} = 1.5\%$ and

– case b3): $\Delta I_{e-eff} \approx 40.4$ for $P_{pl} = 2\%$.

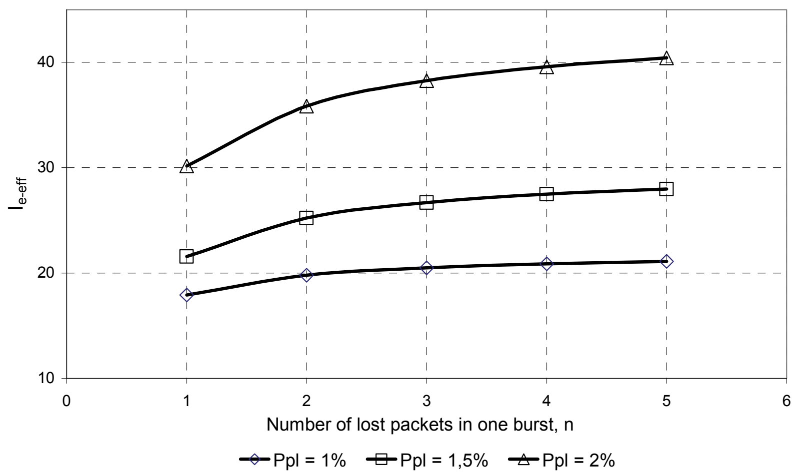

Fig. 2 – *Degradation of speech signal quality as the function of the number of lost packets in one burst.*

When comparing the results for two models (cases a) and b)), it is obtained that degradation of speech signal quality when bursts of n = 5 packets of 10ms duration are lost compared to the situation of random packet loss of 50ms packets is:

– cases a1) and b1): $\Delta I_{e-eff} \approx 3.2$ for $P_{pl} = 1\%$;

– cases a2) and b2): $\Delta I_{e-eff} \approx 6.4$ for $P_{pl} = 1.5\%$, and

– cases a3) and b3): $\Delta I_{e-eff} \approx 10.3$ for $P_{pl} = 2\%$.

The main conclusion of this section is that the calculation of speech signal quality is not good enough when considering random loss of long packets (the first model) and burst loss of short packets (the second model), which are used for the packetization of the same speech signal segment (50ms).

4 Determination of Speech Segment Duration

Our first problem is to determine the duration of the speech segment, which is placed in one packet in packet telephone network of ESI where we expect burst packet loss. The only criterion in this determination is the speech signal quality. The satisfactory bitrates are not the problem, because ESI network uses high bitrate optical links (OPtical Ground Wire, OPGW), [6]. When this segment duration is determined, we must consider two opposite requirements.

The first one is that speech segment has to be as short as possible, because packetization delay is then small. In this way total delay decreases and degradation of speech signal quality also decreases.

The second requirement is that we must decrease burst packet loss and the number of lost packets in one burst. It can be achieved using longer speech segments, packetized in one packet.

We choose the first requirement considering the following reason: the loss of one packet carrying the speech segment of duration $n \cdot t_p$ has the same consequence on the quality of speech signal as the loss of n consecutive packets, each carrying speech segments of duration t_p. But, as can be seen from previous numerical examples, the implementation of calculation leads to the different conclusion. How can we overcome this controversy? It can be concluded that recommendations [1, 13] are incomplete when they consider burst packet loss for small values B_{pl}, as it is the case for G.711 coder without concealment of lost packets. This incompleteness will be decreased considering not only the type of coder, but also speech segment duration in one packet. In this way it will be possible to express by approximate values the consequences of n consecutive packet loss, carrying speech segments of duration t_p each and the consequences of one packet loss, carrying speech segment of duration $n \cdot t_p$. In this calculation it will be obvious that it is necessary to choose as short as possible speech segment, because the gain will be in less total delay.

In [14] it is explained that values of B_{pl} must depend on the duration of speech segment which is packetized by one packet in order to equate the values of speech quality when it is packetized by speech segments of different duration. A few values of B_{pl} are suggested for one specific value of P_{pl} and different packetization intervals. In this paper we suggest that decrease of speech signal quality is calculated using the following equation (although this equation is only one possible solution how to present the approximate equation):

$$I_{e\text{-}eff} = \frac{95P_{pl}}{\dfrac{P_{pl}}{BurstR} + B_{pl}\left(1 + \dfrac{X}{1-X}\right)}, \tag{3}$$

where it is

$$X = \frac{\left(T_{pck} - 10\right)}{10}\left(\frac{0.12\left(P_{pl} - 1\right)}{P_{pl}} - \frac{0.075 P_{pl}}{BurstR}\right). \tag{4}$$

The value *BurstR* can be expressed as:

$$BurstR \approx n = \frac{T_{dis}}{T_{pck}}. \tag{5}$$

In these equations T_{pck} presents duration of speech segment, which is packetized by one packet and T_{dis} is duration of the disturbance, which causes speech packets loss. Both variables are expressed in ms.

The results for $T_{pck} = 10$ms obtained using (3) are exactly the same as the results obtained by (2). When the value of T_{pck} is other than 10ms, the results according to these two equations are different.

For the comparisons made in this paper the most important are the results presented in Fig. 3, which correspond to the disturbances of 50ms. The results are presented for the second model (small packet duration –10ms, burst loss) and for the first model (long packet duration –50ms, random loss). When considering the first model, the results are presented according to (2), which is taken from recommendation [1] and according to the new equations (3) and (4). It is obvious that the values of speech signal quality according to the new equation are much closer one to the other when these two models are applied for the disturbances of the same duration.

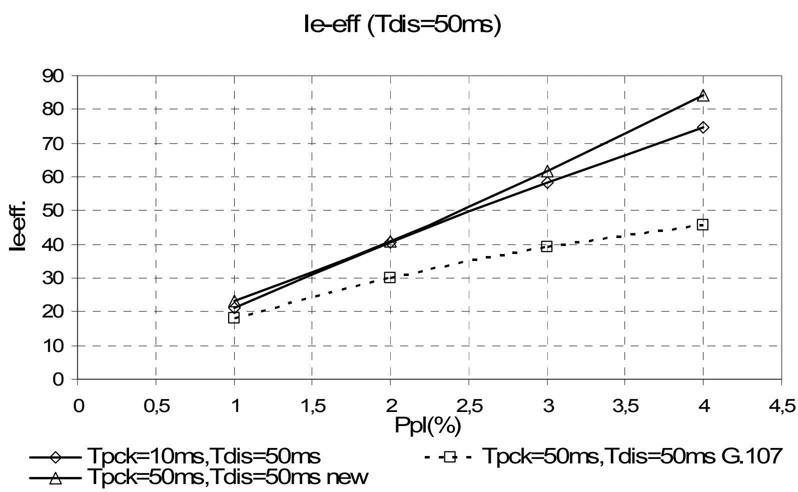

Fig. 3 – *Effective Equipment Impairment Factor* ($I_{e\text{-}eff}$) *for first and second model when disturbances are* T_{dis} = 50ms *long and packetization interval is* T_{pck} = 10ms *and* T_{pck} = 50ms.

Figs. 4 – 6 present the speech signal quality in the cases when disturbances are 30ms, 40ms and 60ms long, respectively. The results in Figs. 5 and 6 include two packet durations besides the duration of 10ms as the parameter, while the Figs. 3 and 4 are presented for only one packet duration besides 10ms.

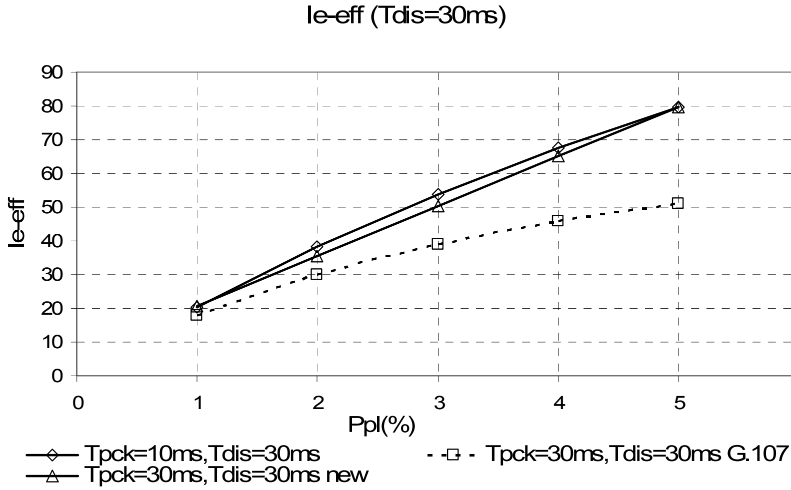

Fig. 4 – *Effective Equipment Impairment Factor* ($I_{e\text{-}eff}$) *for first and second model when disturbances are* T_{dis} = 30ms *long and packetization interval is* T_{pck} = 10ms *and* T_{pck} = 30ms.

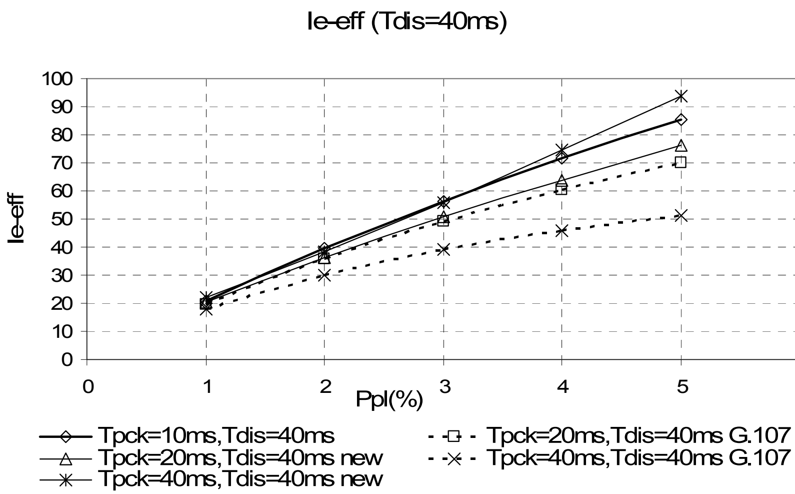

Fig. 5 – *Effective Equipment Impairment Factor* ($I_{e\text{-}eff}$) *for first and second model when disturbances are* T_{dis} = 40ms *long and packetization interval is* T_{pck} = 10ms, T_{pck} = 20ms *and* T_{pck} = 40ms.

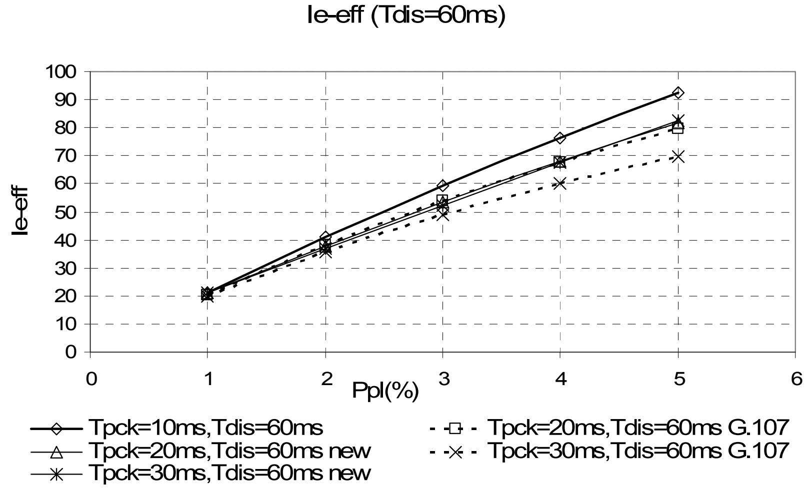

Fig. 6 – *Effective Equipment Impairment Factor* ($I_{e\text{-}eff}$) *for first and second model when disturbances are* T_{dis} = 60ms *long and packetization interval is* T_{pck} = 10ms, T_{pck} = 20ms *and* T_{pck} = 30ms.

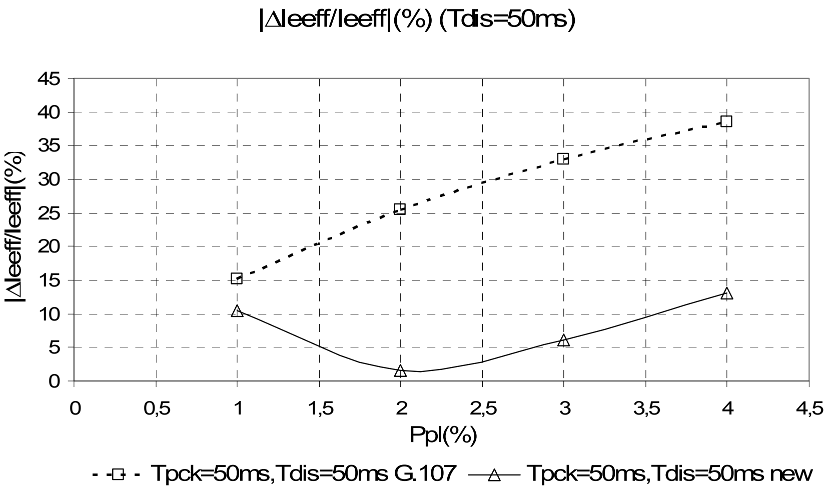

Fig. 7 – *Absolute value of relative error* ($|\Delta I_{e\text{-}eff}/I_{e\text{-}eff}|$) *in calculation of* $I_{e\text{-}eff}$ *when the results obtained by new equations* (3) *and* (4) *and by the equation from* [1] *are compared for first and second model, disturbances are* T_{dis} = 50ms *long and packetization interval is* T_{pck} = 50ms.

Figs. 7 – 10 present absolute value of relative error ($|\Delta I_{e\text{-}eff}/I_{e\text{-}eff}|$) of the results obtained by the calculation according to the new equations (3) and (4), compared to the results according to the equation from recommendation [1].

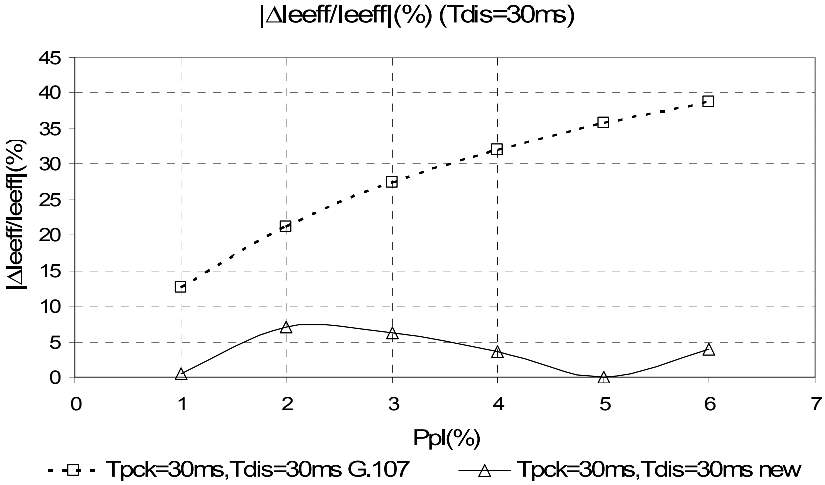

Fig. 8 – *Absolute value of relative error* ($|\Delta I_{e\text{-}eff}/I_{e\text{-}eff}|$) *in calculation of* $I_{e\text{-}eff}$ *when the results obtained by new equations* (3) *and* (4) *and by the equation from* [1] *are compared for first and second model, disturbances are* T_{dis} = 30ms *long and packetization interval is* T_{pck} = 30ms.

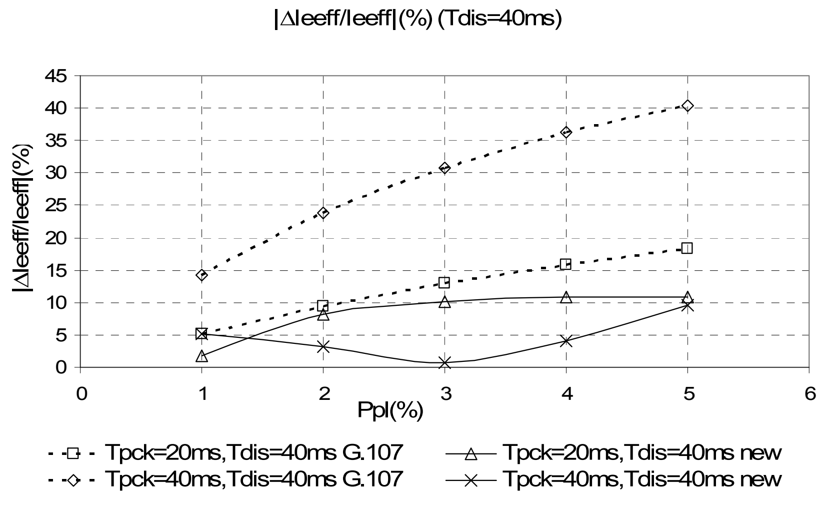

Fig. 9 – *Absolute value of relative error* ($|\Delta I_{e\text{-}eff}/I_{e\text{-}eff}|$) *in calculation of* $I_{e\text{-}eff}$ *when the results obtained by new equations* (3) *and* (4) *and by the equation from* [1] *are compared for first and second model, disturbances are* T_{dis} = 40ms *long and packetization interval is* T_{pck} = 20ms *and* T_{pck} = 40ms.

The Figs. 7 – 10 demonstrate the effectiveness of (3) and (4) – the difference between the estimations of speech signal quality when considering two models of calculation are much less when we implement the new equation instead of the equation from the recommendation. Exception is only the case when disturbance duration is 60ms, $T_{pck} = 20$ms and P_{pl} is 2% or 3%. But, in these cases relative error is less than 10%, so the deviation of the result is not important. Generally, relative error of the results according to the new equation in all presented cases is less than 13%, while the relative error according to equation from [1] reaches 40%.

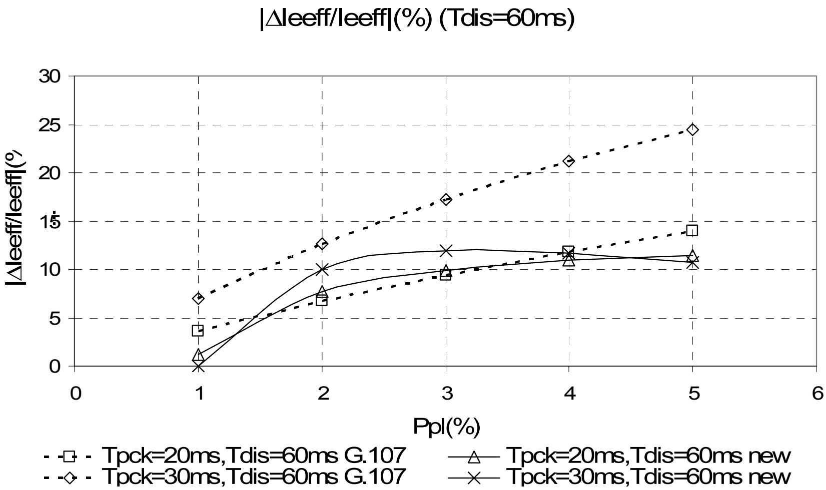

Fig. 10 – *Absolute value of relative error* ($|\Delta I_{e\text{-}eff}/I_{e\text{-}eff}|$) *in calculation of* $I_{e\text{-}eff}$ *when the results obtained by new equations* (3) *and* (4) *and by the equation from* [1] *are compared for first and second model, disturbances are* $T_{dis} = 60$ms *long and packetization interval is* $T_{pck} = 20$ms *and* $T_{pck} = 30$ms.

6 Conclusion

The contributions of the paper are twofold. The first one is evident from the (2) and the corresponding results. According to the results presented in this paper, in packet telephone network of ESI is desirable to send one speech signal segment, which is as short as possible per each packet. In this way total delay is shortened and the degradation of speech signal quality, caused by packetization delay, is decreased. The disturbances in packet telephone network of ESI are long-term and they cause burst packet loss. If we choose longer speech segments, we can decrease the number of consecutively lost packets, but we cannot shorten the part of lost speech signal.

The second, more important contribution follows from the new formulas (3) and (4), which are original development of the presented analysis. Suggested method of calculation shows that the difference in the obtained values of Effective Equipment Impairment Factor ($I_{e\text{-}eff}$), when the disturbances of the same duration are packetized by speech segments of different duration, can be decreased compared to the values according to ITU-T Recommendations. This is especially important for the estimations of speech quality in the networks where disturbances are relatively long. The correction of the equation from ITU-T Recommendation is realized by introducing the correction factor of B_{pl}. Value of B_{pl} according to (3)–(4) depends not only on the type of coder (compressor), but also on packetization interval, probability of packet loss and the number of lost packets in one burst. The parameters in our original equations are adjusted to obtain optimal change in the value of $I_{e\text{-}eff}$ when speech signal is packetized by segments in the range between 10ms and 60ms. This is a new, analytical and more comprehensive approach for B_{pl} determination comparing to the approach in previous literature [14], where B_{pl} values were only tabulated.

7 Acknowledgments

This work is supported by the Ministry of Education, Science and Technological Development, Republic of Serbia, within the Projects TR32007 and TR32051.

8 References

[1] Telecommunication Standardization Sector of ITU-T Rec. G.107: The E-Model: A Computational Model for Use in Transmission Planning, June 2015.

[2] A. Kovac, M. Halas, M. Orgon, M. Voznak: E-Model MOS Estimate Improvement through Jitter Buffer Packet Loss Modelling, Advances in Electrical and Electronic Engineering, Vol. 9, No. 5, 2011, pp. 233 – 242.

[3] F. De Rango, M. Tropea, P. Fazio, S. Marano: Overview on VoIP: Subjective and Objective Measurement Methods, International Journal of Computer Science and Network Security, Vol. 6, No. 1B, January 2006, pp. 140 – 153.

[4] D. Stanojević, B. Bondžulić, B. Pavlović, V. Petrović: The Impact of Quality of Service Parameters to the Subjective and Objective Video Quality Assessment, Serbian Journal of Electrical Engineering, Vol. 15, No.1, April 2018, pp. 97 – 114.

[5] H. Lehpamer: Introduction to Power Utility Communications, Artech House, Boston, London, 2016.

[6] N. Krajnović: The Design of a Highly Available Enterprise IP Telephony Network for the Power Utility of Serbia Company, IEEE Communications Magazine, Vol. 47, No. 4, April 2009, pp. 118 – 122.

[7] IEEE Stdandard 634-2004: IEEE Guide for Power – Line Carrier Applications, IEEE, New York, USA, June 2005.

[8] A. Lebl, D. Mitić, Ž. Markov: Influence of Connection Length on Speech Signal Quality in Packet Network of Electric Power Utility, Revue Roumaine des Sciences Techniques, Vol. 56, No. 3, September 2011, pp. 295 – 304.

[9] A. D. Potorac: Considerations on VoIP Throughput in 802.11 Networks, Advances in Electrical and Computer Engineering, Vol. 9, No. 3, October 2009, pp. 45 – 50.

[10] Telecommunication Standardization Sector of ITU-T Rec. Y.1541: Network Performance Objectives for IP-Based Services, December 2011.

[11] Telecommunications Industry Association TSB-116- Telecommunications IP Telephony Equipment Voice Quality Recommendations for IP Telephony, March 2006.

[12] Telecommunication Standardization Sector of ITU-T Rec. G.114: One-Way Transmission Time, May 2003.

[13] Telecommunication Standardization Sector of ITU-T Rec. G.113: Transmission Impairments due to Speech Processing, November 2007.

[14] M. Stanić, D. Mitić, A. Lebl: A Correction of E-Model in Quality Estimation of Packetized Speech Signal, AEU - International Journal of Electronics and Communications, Vol. 67, No. 9, September 2013, pp. 793 – 794.