COMPARISON OF TEST SEQUENCES FOR INTRUSIVE MEASUREMENT OF VTQOS WITH SPEECH SEQUENCES IN THE ENVIRONMENT OF IP NETWORKS BY MEANS OF PESQ ALGORITHM

Peter Počta, University of Žilina, Slovakia

Abstract This paper describes simulations of test and speech sequences transmission for intrusive measurement of Voice Transmission Quality of Service (*VTQoS*) in the environment of IP networks. The aim of simulations was detect environment influences on the quality of transmission sequences, particularly an influence of coding schemes, packets loss and the jitter on the transmission sequences. The evaluation of test and speech sequences simulations were based on the calculation of the *MoS* (Mean Opinion Score) values. The *PESQ* algorithm was used for calculation of *MoS* values. Reconsideration of a convenience of the given test sequences, which are composed from simple signals, on intrusive measurement of *VTQoS* in the environment of IP networks is the aim of this paper.

1. INTRODUCTION

VTQoS (Voice Transmission Quality of Service) is one of the important parts of *QoS* (Quality of Service). It is very important for granters as well as for users. When communication network includes in more and more transmission technologies, an increase in complication and the complexity of networks is seen. Measurement of the voice transmission quality becomes only platform that is available for simultaneous comparison of different transmission technologies and that is the most relevant to the view of the users. Of course, it is possible to measure and evaluate the transmission parameters of the networks. But only the evaluation of end-to-end quality provides optimal results because of the complexity of network technologies. Thus, it is the evaluation by the same way as users do. Since voice service is the most wide-spread service, in which a user uses filter and predicative abilities of human brain, it is crucial to optimally evaluate a quality of such service. Evaluation of a quality of the voice service may be performed using intrusive or non-intrusive methods, objectively or subjectively. Using non-intrusive method, we only monitor existing dialogue. The drawback of this method is that evaluation algorithm can not utilize an original sample of the primary signal. Thus, it is very difficult to detect some types of signal distortion that occure during transmission. In the intrusive methods, only a test voice sample is transmitted. These methods have been known since the beginning of the telecommunication technologies, when the special sequences of vowels (known as logathoms) were transmitted after the connection had been built-up. A receiver had to recognize these logathoms. This way of subjective evaluation is used to nowadays (e.g. method *MoS*).

Today's technical and software facilities provide an objectification of this measurement method by transmitting the sound sample defined beforehand, its receiving on the destination side, and a comparison of the transmission sample and the original sample using the suitable algorithm that imitates the way of perception and evaluation of the quality transmission opinion by an average listener. It is for example E-model defined in *ETR-250*, or algorithm *PSQM* (Perceptual Speech Quality Measurement) defined in *P.861 ITU-T* also *PESQ* (Perceptual Evaluation of Speech Quality) defined in *P.862 ITU-T*. The algorithm *PSQM* is based on comparison of the power spectrum of the corresponding sections of the original and the received signals. The results of this algorithm more correlate with the results of listening tests, in comparison with *E-model*. At present, this algorithm is no more used because of a raw time alignment. Instead of it the algorithm *PESQ* is used. The *PESQ* algorithm facilitates with very fine time alignment and one single interruption are also taken into account in the calculation of *MoS*.

A choice of the optimal test sequence is very important for all these methods. The test sequence would consist of non-speech-like (fully artificial) signals. These signals are closer defined in *P.501 ITU-T* and the recommendation divides them into deterministic and random signals. An advantage of these signals is simplicity and possibility of a

comparison of the results measured in different language areas. Thus, the test sequence composed from those signals enables the comparison of networks of individual countries within one corporation (e.g. Deutsche Telecom, Orange and Vodafone) from the point of the view *VTQoS*.
Here we focus to the influence of used coder, packet loss and jitter on the quality of transmission sequences.

2. DESCRIPTION OF THE FIRST TYPE OF TEST SEQUENCE (testseq1)

The length of the test sequence is set to 90 sec. This period equals to the length of a phone call of average user. The test sequence is composed from the following signals introduced and evaluated in [6]:

- Sinusoidal signal with frequencies 300, 800, 1000, 1700, 2400, 3000 Hz ,
- Square bipolar signal with frequencies 300, 400, 500, 600, 635, 670 Hz ,
- Gaussian white noise with μ = 0 and δ= 0,0001; 0,001; 0,01; 0,1; 0,5; 1.

The principle of the creation of the final test sequence is based on arrangements of the parts of the test sequence, which are shown in Figure 1 and Figure 2. The final test sequence consists of six sections. Each section consists of five parts. The arrangement shown in Figure 1 is used once and then the arrangement shown in Figure 2 is used four times to form the first section of final test sequence. The arrangement shown in Figure 2 is used five times to form the other sections of the final test sequence. The signals step-by-step have got the values defined above. That means, in the second section of the test sequence (from 15 sec. to 30 sec.), the signals have the following values: Square bipolar signal f = 400 Hz, Gaussian white noise δ = 0,001 and Sinusoidal signal f = 800 Hz. The values of the signals in the first section of the test sequence (from 0 sec. to 15 sec.) are the same as those in Figure 1 and Figure 2.

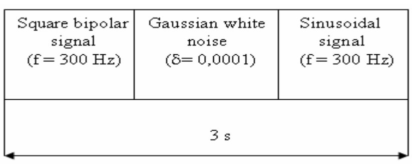

Fig.1 Initial part of the first type of test sequence (testseq1)

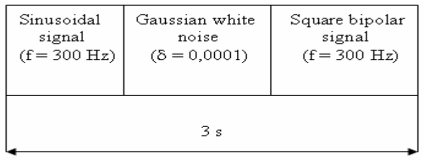

Fig.2 Second part of the first type of test sequence (testseq1)

The choice of first type of test sequence for intrusive measurement of *VTQoS* is published in [3]. The optimization of this type of the test sequence for coder *G.723.1 ITU-T* and *G.729 ITU-T* is published in [4].

3. DESCRIPTION OF THE SECOND TYPE OF TEST SEQUENCE (testseq2)

The length of the test sequence is set to 90 sec. This period equals with the length of a phone call of average user. The sequence was created of the signals, whose convenience as verified in [6]. The creation of test sequence was based on superposing Sinusoidal signal and Gaussian white noise on the square bipolar signal. We must hold on the condition of orthogonality. This rule only relates to periodic signals. These signals with competent parameters were used to create the sequence:

- Sinusoidal signal with frequencies 500, 1000, 1500, 2000, 2500, 3000 Hz ,
- Gaussian white noise with μ= 0 and δ= 0,0001; 0,005; 0,001; 0,05; 0,025; 0,01.

The square bipolar signal with frequency 500 Hz as used as a carrier signal.

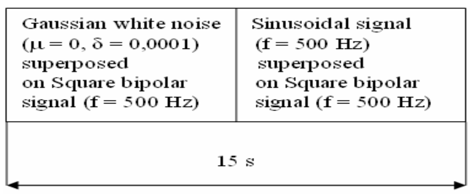

Fig. 3 Initial part of the second type of test sequence (testseq2)

The principle of the creation of the final test sequence is based on an arrangement of initial part of relevant test sequence, which is shown in Figure 3. The arrangement shown in Figure 3 is used six times to form the final test sequence. Thus, final test sequence consists of six parts. The signals step-by-step get the values defined above. Hence, in the second part of the test sequence (from 15 sec. to 30 sec.), the signals have the following values: Square bipolar signal f = 1000 Hz, Gaussian white noise δ= 0,005, the parameters of carrier signal are not changed. The values of the signals in the first part of the test sequence (from 0 sec. to 15 sec.) are the same as those in Figure 3.

The choice of second type of test sequence for intrusive measurement of *VTQoS* is published in [10]. The optimization of this type of test sequence for *G.729 ITU-T* codec can not realized, because the creation of this type of test sequence was based on superposing Sinusoidal signal and Gaussian white noise on the square bipolar signal. The degradation of square bipolar signal by means of *G.729 ITU-T* codec is described in [4]. We can not use this type of test sequence for *G.729 ITU-T* codec.

4. DESCRIPTION OF THE SPEECH SEQUENCES

The recommendation *P.830 ITU-T* [9] recommends to use minimum 2 female and 2 male voices for evaluation of speech quality in telecommunication network. The best choice is 8 male, 8 female and 8 infant voices. We decided to use 2 female and 2 male voices for the needs of our simulations. The length of the speech sequences is set to 90 sec. This period equals to the length of a phone call of average user. The speech sequences are composed from speech records. These speech records come from Slovak database.

5. SIMULATION DESCRIPTION

The transmission simulations were carried out on Gaoresearch's (freeware) online simulator [5]. The simulation model of transmission chain with coder *G.723.1 ITU-T* is depicted in Figure 4. The simulation model enables to change jitter rate and packet loss parameters in the range from 0 % to 10 %. The simulation model renders *VAD* (Voice Activity Detector) and *PLC* (Packet Loss Control) functions.

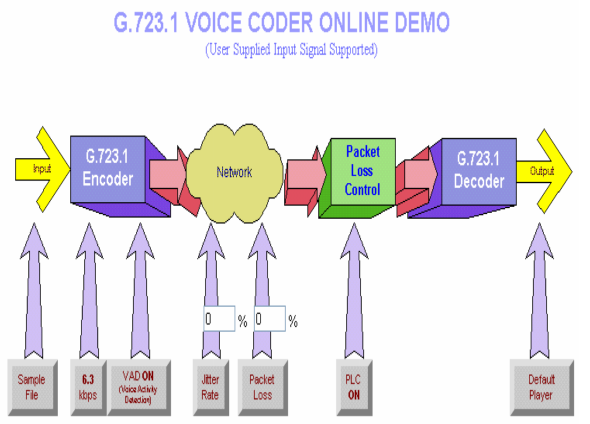

Fig.4 The simulation model of transmission chain with coder G.723.1 ITU-T

5.1 Principle of simulation

The source sequences described in chapter 2 and chapter 3 and 4 were used for the simulations. The simulations were realized for different setting of packet loss and jitter rate parameters and with using these 2 coders:

- *G.723.1 ITU-T (5.3 kbps, 6.3 kbps),*
- *G.729 ITU-T.*

The simulations of jitter influence were done for the values of jitter rate in the range from 0 % to 10 %. Jitter rate is defined as percentual number of packets, whose jitter value is above maximum tolerated jitter for given connection. Jitter is a measure of variation in latency over time. Jitter is caused by random variation of the momentary traffic load. This simulator tolerates the jitter below 90 ms. The packets delivered after this time are further not processed, they are also dropped out. The packets loss influence was investigated for the values of packet loss in the range from 0 % to 10 %. The packet loss parameter is defined as the percentual number of the packets that were lost during transmission. Packets may lost, due to high bit error rate of transmission channel and high traffic load. *VAD* and *PLC* functions were activated for all performed simulations.
The source and the destination sequences are compared after finishing the simulation. The comparison will be realized by means of *PESQ* (Perceptual Evaluation of Speech Quality) algorithm.
Study of an influence of the competent coders, packets loss and the jitter on sequences transmission was the aim of simulations.

5.1.1 *PESQ*

The *PESQ* algorithm is defined in recommendation P.862 ITU-T [8]. This algorithm was developed for the evaluation of speech quality. *PESQ* belong to the group of intrusive algorithms. It means, that we need a original speech sequence and too a destination speech sequence for the evaluation of the speech quality. These sequences will be compared by means of *PESQ* algorithm and then we obtain the *PESQMoS* value. The *PESQMoS* value represents the speech quality. The *PESQMoS* can have the values in the range from -0.5 to 4.5. The value -0.5 is the bad speech quality and the value 4.5 is the excellent speech quality.
This algorithm consists of two parts. In the first part of this algorithm is realized the time correlation and too amplitude correlation. In the second part of this algorithm is processed the speech by means of the perception model. The

perception model realizes the comparison of the equivalent parts of the original speech sequence and the destination speech sequence.

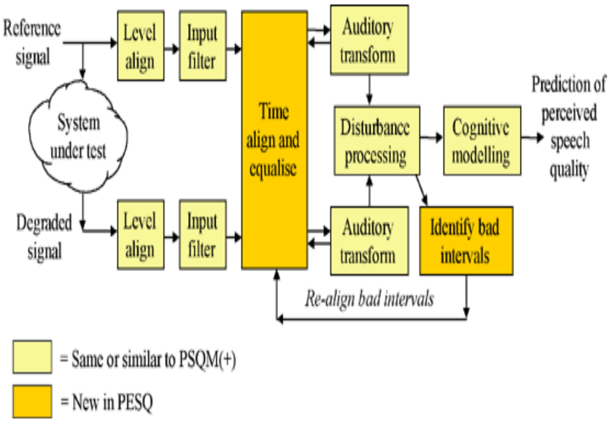

Fig. 5 The principal scheme of PESQ algorithm

It is possible to use *PESQ* in mobile networks as well as in networks based on packet transmission. The disadvantages include impossibility to use it for codec with data rate lower than 4 kbps and higher calculation load what is caused by recursions in the algorithm.

6. PRESENTATION OF RESULTS

6.1 The simulation results of jitter influence on the sequences transmission

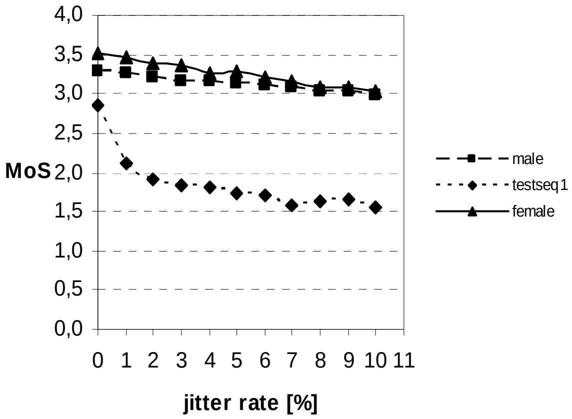

Fig.6 Graphical presentation of the simulation results of jitter influence for *G.729 ITU-T* codec

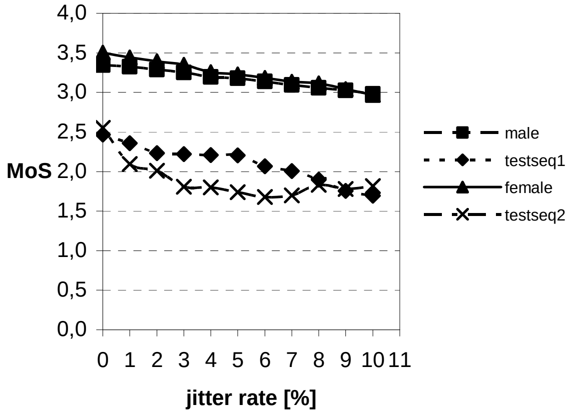

Fig.7 Graphical presentation of the simulation results of jitter influence for *G.723.1 (5.3 kbps) ITU-T* codec

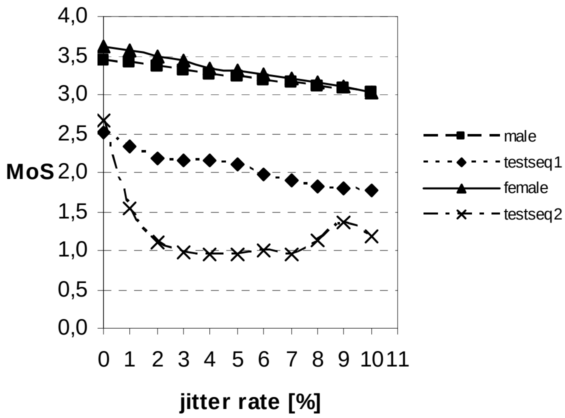

Fig.8 Graphical presentation of the simulation results of jitter influence for *G.723.1 (6.3 kbps) ITU-T* codec

The graphs represent the dependence of *MoS* values change of the percentual number of packets, whose jitter oversteped the time of 90 ms. Every packet whose jitter oversteped the time of 90 ms is dropped. Non-uniform distribution of the number of dropped packets during transmission causes a smooth undulation of characteristic. In the case of zero jitter rate, *MoS* value change is cause only by coder. The graphs only represent average values for female speech sequences and for male speech sequences.

6.2 The simulation results of packets loss influence on the sequences transmission

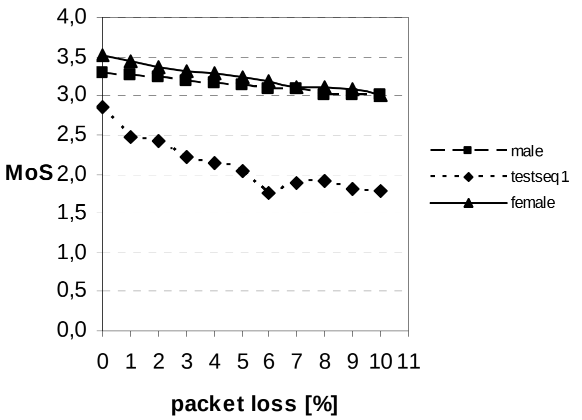

Fig.9 Graphical presentation of the simulation results of packets loss influence for *G.729 ITU-T* codec

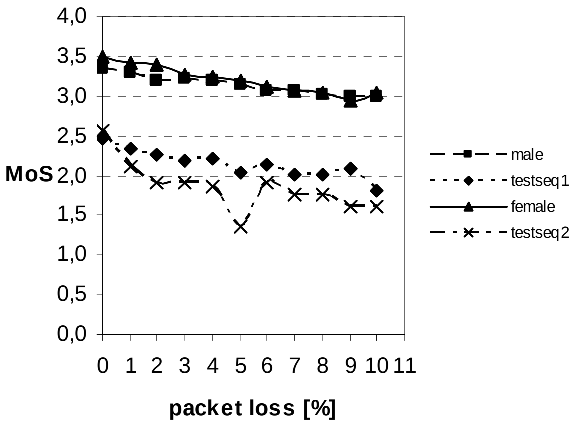

Fig.10 Graphical presentation of the simulation results of packets loss influence for *G.723.1 (5.3 kbps) ITU-T* codec

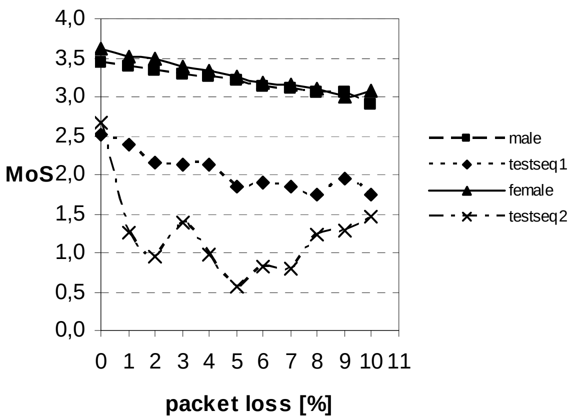

Fig.11 Graphical presentation of the simulation results of packets loss influence for *G.723.1 (6.3 kbps) ITU-T* codec

The graphs represent the dependence of *MoS* values change of packets loss, that means of percentual number of the packets, which were not delivered. The smooth undulation of the characteristic is caused by non-uniform distribution of the number of lost packets during given transmission. In the case of zero packet loss, *MoS* value change is cause only by coder. The graphs only represent average values for female speech sequences and for male speech sequences.

7. CONCLUSION

The results show, that the test sequences are more sensitive to the disturbing influences that origin from transmission in the environment of IP networks, because the test sequences are composed from simple signals. High sensitivity of the test sequences is suitable for intrusive measurement of *VTQoS* and enables more precise measurement of disturbing influences, which rise in IP networks. High sensitivity enables to predict qualitative changes in IP network. In the future, convenience of this test sequences for intrusive measurement of *VTQoS* will be verified practically by real measurements in convergent network of the University of Žilina.

8. BIBLIOGRAPHY AND AUTHOR

BIBLIOGRAPHY

1. POČTA, P., VACULÍK, M.: *Vergleich der Testsequenz für intrusive Messung VTQoS mit den Sprachproben in der Mobilumgebung*, In *Proceedings of 17. IWKM*, Mittweida (Germany), ISSN 1437-7624, 2005, no. 2, p. 51-54.

2. POČTA, P., VACULÍK, M. *Vergleich der Testsequenz für intrusive Messung VTQoS mit den Sprachproben in der Umgebung des Festnetzes*, In *Proceedings of conference* „4th Science Day at the University of Applied Sciences "Next Generation Networks” - Future of Networks" “, Leipzig (Germany).

3. POČTA, P., VACULÍK, M.: *Determination of optimal test sequence for automatic intrusive measurement VTQoS on environment of fixed telecommunication network*, In *Proceedings of conference RTT 2005*, Ostrava (Czech republic), 2005.

4. POČTA, P., VACULÍK, M.: *Researching of coders influence on basic measurement signals used in optimal test sequence*, This paper was sent to the *AEEE Journal* for evaluation, Žilina, Slovakia.

5. Gaoresearch online simulator http://www.gaoresearch.com/products/speechsoftware/speechsoftware.php

6. POČTA, P., VACULÍK, M.: *Method of choice of test signals for automatic intrusive measurement VTQoS*, In *Proceedings of conference MESAQIN 2005*, Prague (Czech republic), 2005, ISBN 80-01-03262-0.

7. KŘENEK, J., HOLUB, J.: *Hodnocení hlasových přenosů v telekomunikačních sítích, ST 6/2001*, 1996, pp.3-5.

8. P.862 ITU-T

9. P.830 ITU-T

10. POČTA, P.: *Determination of test sequence for intrusive measurement of VTQoS in mobile environment*, In *Proceedings of conference POSTER 2006, Prague* (Czech republic), 2006.

11. 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.

12. NARBUTT, M., DAVIS, M.: *Assessing the Quality of VoIP Transmission, Affected by Playout Buffer Scheme*, In *Proceedings of conference MESAQIN 2005*, Prague (Czech republic), 2005, ISBN 80-01-03262-0.

13. MELVIN, H., MURPHY, L.: *Exploring the Extent and Impact of Playout Adjustments within VoIP Applications on MOS Scale*, In *Proceedings of conference MESAQIN 2005*, Prague (Czech republic), 2005, ISBN 80-01-03262-0.

14. HOLUB, J., DROZDOVÁ, A.: *Proprietary Low Bit-rate Radio-communication Network – Objective and Subjective Speech Transmission Quality Assessment*, In *Proceedings of conference MESAQIN 2006*, Prague (Czech republic), 2006, ISBN 80-01-03503-4.

15. BEERENDS, J. G., LARSEN, E., IYER, N.: *Measurement of speech intelligibility based on the PESQ approach*, In *Proceedings of conference MESAQIN 2004*, Prague (Czech republic), 2004, ISBN 80-01-03017-2.

16. HOLUB, J., OCENASEK, J.: *Processing Power Optimisation for PESQ*, In *Proceedings of conference MESAQIN 2004*, Prague (Czech republic), 2004 ISBN 80-01-03017-2.

17. 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.

18. RIX, W. A.: *Comparison between subjective listening quality and P.862 PESQ score*, In *Proceedings of conference MESAQIN 2003*, Prague (Czech republic), 2003, ISBN 80-01-02822-4.

19. HOLUB, J., ŠMÍD, R., BACHTÍK, M.: *Child Listeners as the Test Subject – Comparison with Adults and P.862*, In *Proceedings of conference MESAQIN 2003*, Prague (Czech republic), 2003, ISBN 80-01-02822-4.

20. PENNOCK, S.: Accuracy of the Perceptual Evaluation of Speech Quality (PESQ) algorithm, In *Proceedings of conference MESAQIN 2002*, Prague (Czech republic), 2002, ISBN 80-01-02515-2.

Author:

MSc. Peter Počta
University of Žilina
Faculty of Electrical Engineering
Department of Telecommunications
Univerzitná 1
010 26 Žilina, Slovakia
tel. +421 41 513 2227
fax +421 41 513 1520
email: pocta@fel.utc.sk