Comparison of test sequence for intrusive measurement of VTQoS with speech sequences in the environment of IP networks

Peter Počta, Martin Vaculík

Dept. of Telecommunications, FEE, University of Žilina , Univerzitná 1, 010 26, Slovakia

pocta@fel.utc.sk , vaculik@fel.utc.sk

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 mean square measure. Reconsideration of a convenience of the given test sequence, which is composed from simple signals, on intrusive measurement of VTQoS in the environment of IP networks is the aim of this paper.

Keywords

VTQoS, test sequence, correlation coefficient, mean square measure, intrusive measurement

1 Introduction

VTQoS (Voice Transmission Quality of Service) is one of the important parts of QoS (Quality of Service). It is very important for providers as well as for users. When communication networks incorporate 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 in 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 the evaluation algorithm cannot 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 has been used till 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 algorithm PESQ facilitates with very fine time alignment and one single interruption are also taken into account in the calculation of MoS (Mean Opinion Score). 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.

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 using these signals is simplicity and possibility of a comparison of the results measured in different language areas. The test sequence composed from those signals enables the comparison of networks of individual countries within one corporation (e.g. Deutsche Telecom, Orange, Vodafone) from the point of the view VTQoS.

It was found that the test sequence composed from simple signals is more sensitive to the interferences in mobile environment and environment of fixed telecommunication network than a sequence composed from speech samples. Hence, such test sequence is more suitable for prediction of a qualitative change in the network. We compared these two types of sequences in [1], [2] .

Nowadays, the VTQoS intrusive measurements are performed by using samples of speech signal but the comparison is possible only within the single-language area in this case.

New phenomenona, which degrades quality of voice transmission e.g. impulse noise, short-term fading, clipping, or non linear distorsion (introduced in codecs with loss compression), appears in actual telecommunication systems.

Here we focus to the influence of used coder, packet loss and jitter on the quality of transmission sequences.

2 Description of the test sequence

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.

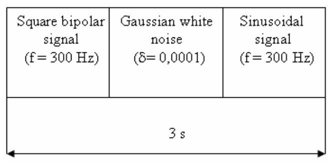

Fig.1 Initial part of the test sequence

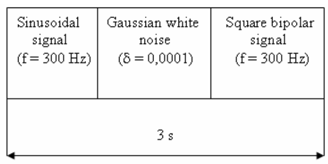

Fig.2 Second part of the test sequence

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

3 Description of the speech sequences

The recommendation P.830 ITU-T 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.

More detailed description of speech sequences:

- Speech sequence womanSK1 is recorded by a woman, who comes from West Slovakia and she is 57.
- Speech sequence womanSK2 is recorded by a woman, who comes from West Slovakia and she is 52.
- Speech sequence manSK1 is recorded by a man, who comes from West Slovakia and he is 38.
- Speech sequence manSK2 is recorded by a man, who comes from West Slovakia and he is 39.

4 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 3. 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.

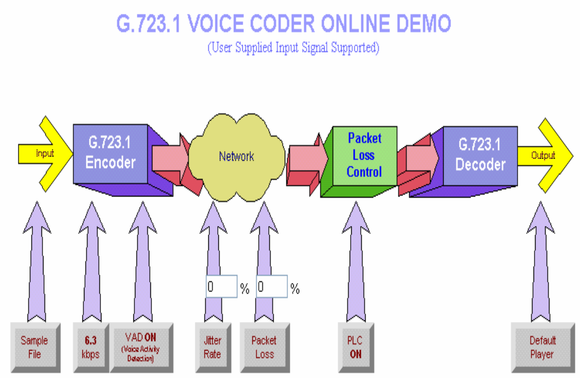

Fig.3 The simulation model of transmission chain with coder G.723.1 ITU-T

4.1 Principle of simulation

The source sequences described in chapter 2 and chapter 3 were used for the simulations. The simulations were realised 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 was 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 paramater 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 principle of the comparison is described by the following steps:

1. Reading in source and destination sequences.
2. Segmentation of the sequences into n intervals, each with 8000 samples. The following parameters are calculated in each interval:
    1. Coefficients FFT,
    2. Mean square measure d_i.
3. Calculation of the average values:

$$\overline{d} = \frac{\sum_{i=1}^{n} d_i}{n}, \qquad (1)$$

where n is the number of intervals, d_i is the mean square measure of the i-th interval.

The segmentation of the sequences into n intervals and calculation of relevant parameters in these intervals enables to obtain more precise results.

The mean square measure is based on the spectral comparison of the test microsegment with the reference microsegment. The most common norm is L_2-norm, which is defined as:

$$d_i(t,r) = \left[\sum_{j=1}^{N} \left(y_{tj} - y_{rj} \right)^2 \right]^{1/2}, \qquad (2)$$

where N is the number of FFT points in given microsegment, y_{tj} is the absolute value of the j-th FFT coefficient of the test microsegment, y_{rj} is the absolute value of the j-th FFT coefficient of the reference microsegment.

Study of an influence of the competent coders, packets loss and the jitter on sequences transmission was the aim of simulations.

5 Presentation of results

5.1 The simulation results of jitter influence on the sequences transmission

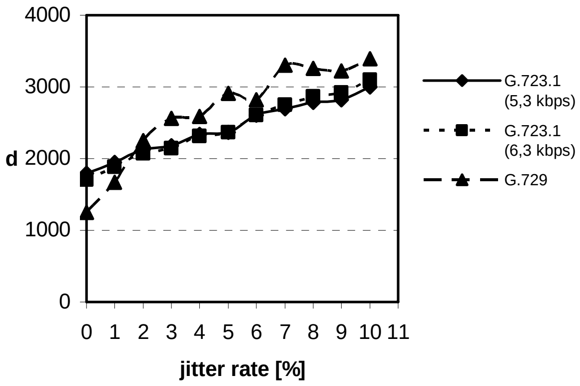

Fig.4 Graphical presentation of the simulation results of jitter influence on the test sequence

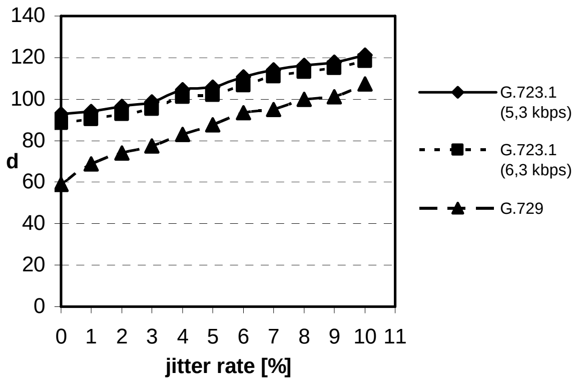

Fig.5 Graphical presentation of the simulation results of jitter influence on male speech sequences

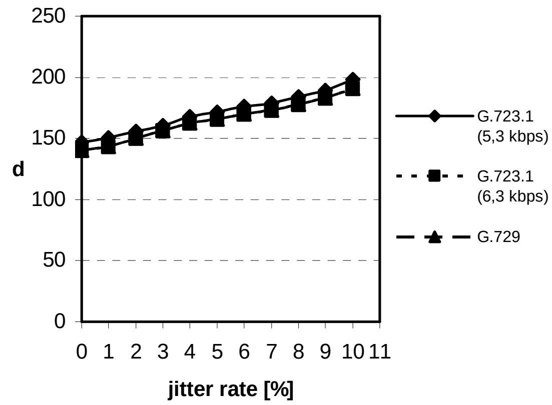

Fig.6 Graphical presentation of the simulation results of jitter influence on female speech sequences

The graphs represent the dependence of spectrum 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. Spectrum change is evaluated as the spectral distance defined by equation (2). Non-uniform distribution of the number of dropped packets during transmission causes a smooth undulation of characteristic. In the case of zero jitter rate, spectrum change is cause only by coder. The graphs only represent average values for female speech sequences and for male speech sequences.

5.2 The simulation results of packets loss influence on sequences transmission

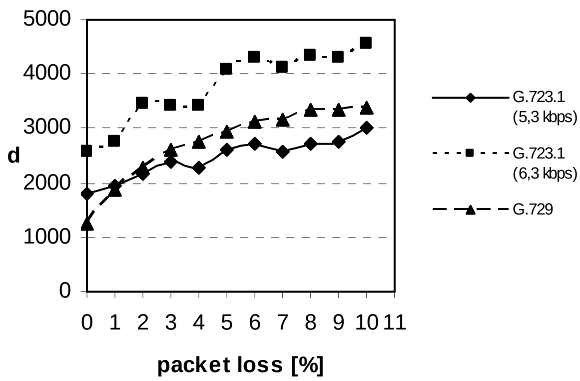

Fig.7 Graphical presentation of the simulation results of packets loss influence on the test sequence

test sequence for intrusive measurement of VTQoS will be verified practically by real measurements in convergent network of the University of Žilina.

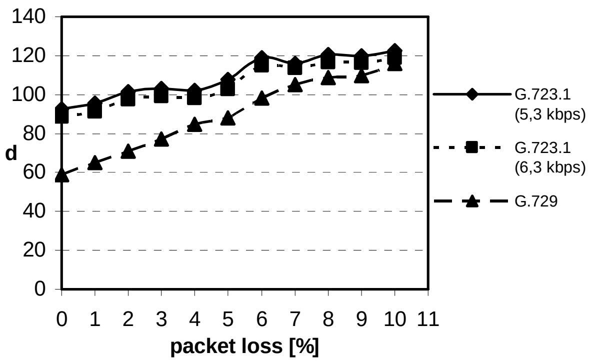

Fig.8 Graphical presentation of the simulation results of packets loss influence on male speech sequences

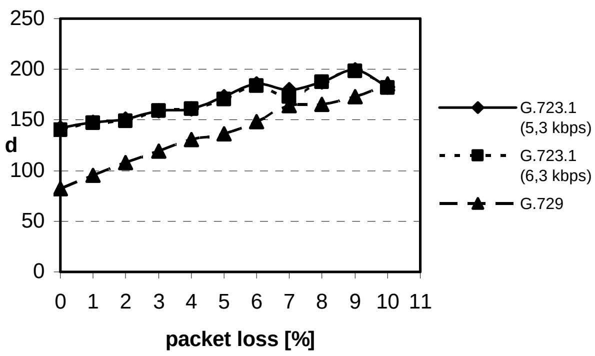

Fig.9 Graphical presentation of the simulation results of packets loss influence on female speech sequences

The graphs represent the dependence of spectrum change of packets loss, that means of percentual number of the packets, which were not delivered. Spectrum change is evaluated as the spectral distance defined by equation (2). 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, spectrum change is cause only by coder. The graphs only represent average values for female speech sequences and for male speech sequences.

6 Conclusion

The results show, that the test sequence is more sensitive to the disturbing influences that origin from transmission in the environment of IP networks, because the test sequence is composed from simple signals. High sensitivity of the test sequence 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 sequence for intrusive measurement of VTQoS will be verified practically by real measurements in convergent network of the University of Žilina.

7 Acknowledgements

The work is partially supported by the State Program of R&D within the project Intelligent Speech Communication Interface.

8 References

[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. Reaserching of coders influence on basic measurement signals used in optimal test sequence, This paper was sent to the *AEEE Journal* for evaluation.

[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] PSUTKA, J. Komunikace s počítačem mluvenou řečí, ACADEMIA, 1995, 287 p, ISBN 80-200-0203-0.

[9] FRANEKOVÁ, M. Modelovanie komunikačných systémov v prostredí Matlab, Simulink a Communications Toolbox, EDIS, 2003, 129 p, ISBN 80-8070-027-3.