UDC 681.3.06

VOICE QUALITY ASSESSMENT DURING VOICE TRANSMISSION BY VARIOUS NETWORK TECHNOLOGIES

K. Kleinová, J. Bača, F. Jakab, P. Feciľak

Computers Network Laboratory, Dpt. of Computers and Informatics, Technical University of Košice, Letná 9, 040 01 Košice, Slovakia, tel.+421 55 602 2573.
E-mail: katarina.kleinova@tuke.sk, jan.baca@tuke.sk, frantisek.jakab@tuke.sk, peter.fecilak@tuke.sk

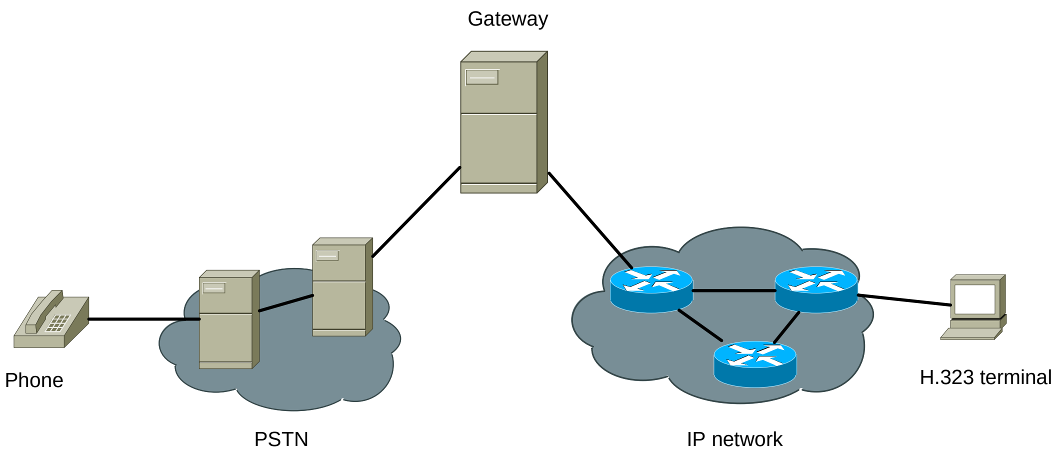
This paper deals with the voice quality system assessment including calculation of the noise floor during voice transmission by various network technologies. Voice over Internet Protocol (VoIP) development presupposes solution of the problems related to effective voice transmission with required transmission parameters (Quality of Services - QoS). One of the mechanisms that may be used for assurance of service quality in IP networks is the QoS model. The QoS model is used to evaluate the VoIP quality and to design the VoIP networks. The voice quality may be determined using the passive method – E-model – which respects the used codec, the packet loss and the delay or the jitter. Attention is also paid to the calculation of the noise floor emphasizing the possibilities of the signal-to-noise ratio estimating. This paper also presents results of experimental verification of the designed solution implementation functionality.

Описано систему оценки качества голоса включая определение порога шума в течении передачи голоса различными сетевыми технологиями. Развитие передачи голоса интернет протоколом (VoIP) предпочитает решение проблем относящихся к эффективной передачи голоса с требованными параметрами передачи (качество сервиса - QoS). Одним из механизмов, который может быть использован для обеспечения качества сервиса в IP сетях является QoS модель. QoS модели используются для оценки качества VoIP и для проектирования VoIP сетей. Качество голоса может быть определено с использованием пасивного метода – E-модели, которая отражает использован кодек, потерю пакетов, опаздывание и его колебание. Внимание уделяется определению порога шума подчеркивающее возможности оценки отношения сигнал-шум. Приведены результаты экспериментальной проверки функционирования имплементации предложенного решения.

1. Introduction

The issues of the voice information quality assessment are related to the interconnection of various voice network technologies. Apart from the voice transmission in the classical telecommunication network (PSTN) (VoIP technology) the voice can be transmitted also by on the basis of the IP protocol and other technologies, such as ISDN, ATM (VoATM – Voice-over-Asynchronous Transfer Mode), Frame Relay (VoFR – Voice over Frame Relay) or in wireless networks.

The convergence of the classical telecommunication and IP networks necessitates that the IP networks are expanded by the mechanism that secures the quality of service that is necessary for the voice transmission (Pic. 1).



Pic. 1. Example of a converged VoIP/PSTN network

In order to make sure that the voice transmission in the data network is not an issue of one technology only, the convergence of technologies is necessary such as the Frame Relay, ATM and IP in one communication service that would not be limited by the applied technology. Even though the fragmentation techniques in Frame Relay, ATM or IP are similar, the prioritization techniques, signalization protocols and algorithms for voice compression are not mutually compatible. In spite of specific standardization within each protocol cooperation of all this voice communication standards was impossible. Therefore the mutually cooperating solutions were processed.

It is important to realize that information on transmission parameters cannot be transmitted by all technologies of voice transmission. The information may be transmitted between various technologies as the data network provide independent transport interface. However in case of interconnection of two data networks by analog voice transmission technologies (e.g. PSTN) such information on transmission parameters cannot be transferred. 1

2. Voice information quality assessment on the basis of E-model

The resulting VoIP system quality is influenced by many technical attributes. These attributes include the selection of the algorithm for voice coding (vocoder), delay system (latency), connection reliability etc. The crucial role is played by the used codec, packet loss influence and the delay or jitter impact.

The E-model is a tool for assessment of the combined effects of variants of various transmission parameters impacting the speech quality 1. The E-model output is the resulting R-factor with the values from 0 to 100 with the acceptable value being from 50 to 100. The R-factor takes into account the influence of the noise, volume, quantizing distortion, coding method, echo, delay etc. It is specified for the whole transmission chain between the acoustic interfaces of the telephone network, i.e. not only the actual telephone canal is taken into account but also its terminal equipments and ambient (rooms) noise.

The E-model is based on the assumption of the additive interaction of individual distortion influences and is described by the equation for calculation of the parameter ***R***, that represents the overall transmission quality: 2

$$R = R_0 - I_S - I_d - I_{e-eff} + A \tag{1}$$

where the parameter R_0 representing the basic value of the parameter R derived from the signal/noise ratio

$$R_0 = 15 - 1{,}5\left(SLR + N_0\right) \tag{2}$$

where N_0 being the sum of powers of various noise sources (including the noise floor) and SLR as the measure of volume in the transmission direction.

The parameter I_S is the sum of all quality declines that may exist concurrently with the voice transmission. The parameter I_d represents the quality decline caused by a delay; it expresses the echo on the distant and close end as well as the effects of the delay.

The parameter I_{e-eff} includes the quality decline caused by the equipment (codec type, coding method) and the deterioration caused by the packet loss.

$$I_{e-eff} = I_e + (95 - I_e)\frac{Ppl}{Ppl + Bpl} \tag{3}$$

where I_e being the quality decline caused by the equipment, Bpl being the factor of robustness of the packet loss while Ppl is the probability of an accidental packet loss. 2

The factor *A (Advantage Factor)* is also described as the factor of expectation and takes into account the lower requirements of the participants in the voice service along with other advantages as compared with the traditional network (such as mobility). There is no relationship between the factor A and other transmission parameters. It can accrue the values specified in the Tablicia.

Tablicia. Factor A values

Type of the terminal	Value of the factor A
Fixed terminal	0
Mobile terminal inside the building	5
Mobile terminal in the moving vehicle	10
Terminal in localities with difficult access	20

In order to assess the voice quality the R-factor may be mapped for the estimated MOS (Mean Opinion Score) within the scale 1 – 5

For $R < 0$ $\quad MOS = 1$

For $0 < R > 100$ $\quad MOS = 1 + 0{,}035R + R(R-60)(100-R)7 \cdot 10^{-6}$

For $R > 100$ $\quad MOS = 4{,}5$

2.1. The methods of noise floor calculation

In spite of the effort to eliminate the noise (jamming) from the useful signal we can say that the noise is always present in the acoustic signal. Noise comes from various sources and it is always present in the signal in spite of the tools that can reduce it.

The noise floor is a combination of the background noise and the noise present in the microphone or in the sound card. It is defined as the signal size created by the sum of all noise sources and of undesirable signals in the system.

The standard criterion of the noise level in the acoustic signal is the distance of the signal from the noise SNR (Signal-to-Noise Ratio). 6 In order to achieve the best voice quality possible the maximum distance of the signal from the noise needs to be secured. Various noise types reduce the value of SNR. The noise sources are:

- internal (system) sources,
- external sources.

The internal noise sources include for example the noise of the used electric parts (telephone circuit noise). The external noise sources include the room noise – the background noise, speech as a noise (undesirable speakers), impulse noise. 4

The noise floor may be calculated on the basis of the following equation

$$N_{for} = -174 + NF + 10\log B \tag{4}$$

where N_{for} being the noise floor, NF the noise number (dB) specifying the degree of deterioration of the distance of the signal from the noise and B being the bandwidth (Hz). The calculated value determining the noise floor is expressed in decibels with minus sign. Currently the recommended value for the noise floor on the receiving side is -64 dBmp.

2.2 Estimate of speech signal SNR

When measuring the level of noise in the actual signal with nose background the performance of the speech and of the noise from one signal should be estimated. The higher value of the SNR estimate indicates that the signal contents are less damaged by the noise. On the other hand the low value of the SNR estimate indicates that the surrounding noise is dominant in the signal. 3

SNR can be calculated so that only the noise performance in the background can be estimated

$$SNR = 10\log\frac{\hat{\sigma}_s^2}{\hat{\sigma}_n^2} = 10\log\frac{\sigma_x^2 - \hat{\sigma}_n^2}{\hat{\sigma}_n^2}, \tag{5}$$

and the calculation of the local SNR is then based on the noise performance estimate in the actually processed segment.

The estimate of the segmental SNR (SSNR) is based on the estimate of the local SNR (SNRi) and is defined by the equation

$$SSNR = \frac{1}{K}\sum_{i=0}^{L=1} SNR_i \cdot VAD_i , \tag{6}$$

where VAD_i being the carrier of the information concerning the speech activity in the i-th segment (1 – speech, 0 – break), L being the total number of analyzed segments in the signal and K being the number of segments with the speech activity. The estimate precision is determined by two basic factors:

- Local SNR estimate precision,
- Speech activity detection precision that needs to be implemented from the noisy speech signal.

The algorithm principle for the estimation of the global SNR is the estimate of the noise performance from the breaks in the relevant speech and, vice-versa, the speech performance estimate only from the speech signal sections. The algorithm of the SNR estimate can then be summarized in the following equations:

$$SNR = 10\log\frac{\hat{\sigma}_s^2}{\hat{\sigma}_n^2}; \tag{7}$$

$$\hat{\sigma}_n^2 = \frac{1}{l_n}\sum_{n=0}^{l-1} x^2[n]\cdot\left|1 - vad[n]\right|; \tag{8}$$

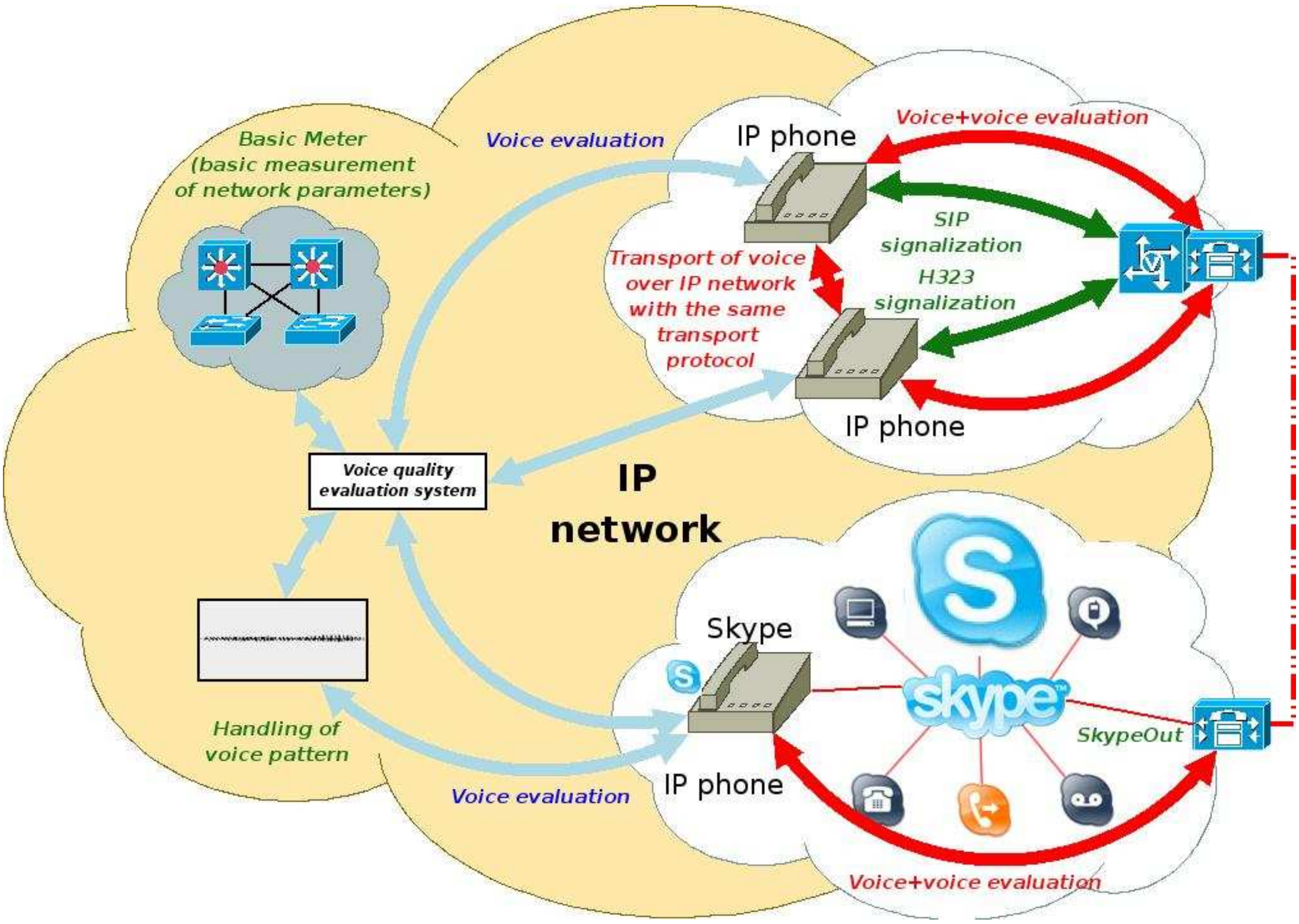
$$\hat{\sigma}_s^2 = \frac{1}{l_s}\sum_{n=0}^{l-1} x^2[n]\cdot vad[n] - \hat{\sigma}_n^2; \tag{9}$$

3. The voice quality assessment system

The voice quality assessment system may be classified in three functional blocks:

- The active measurement process (collection of the input information – voice sample, transmission parameters),
- The voice analysis (determination of SNR, SLR, RLR),
- The voice quality assessment and the exchange of assessments.

The role of the active measurement process is to identify the volume and time features of the transport medium that serve as the input parameters in the calculation of the R-factor on the basis of the E-model. These input features include the volume features such as the use of the bandwidth, the amount of transmitted packets, the loss-rate of the packets and the time features such as the delay fluctuation, one-way and two-way delay. (Pic. 2).

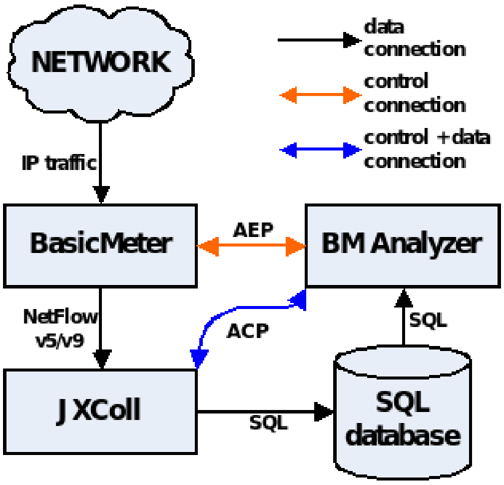


Pic. 2. Measurement of the time and volume features of the transport network

The basic measurements of the volume and time features of the voice transport medium are taken by the measurement tool BasicMeter with the following components of the architecture:

- BEEM
- BM analyzer
- JXCollector
- SQL database

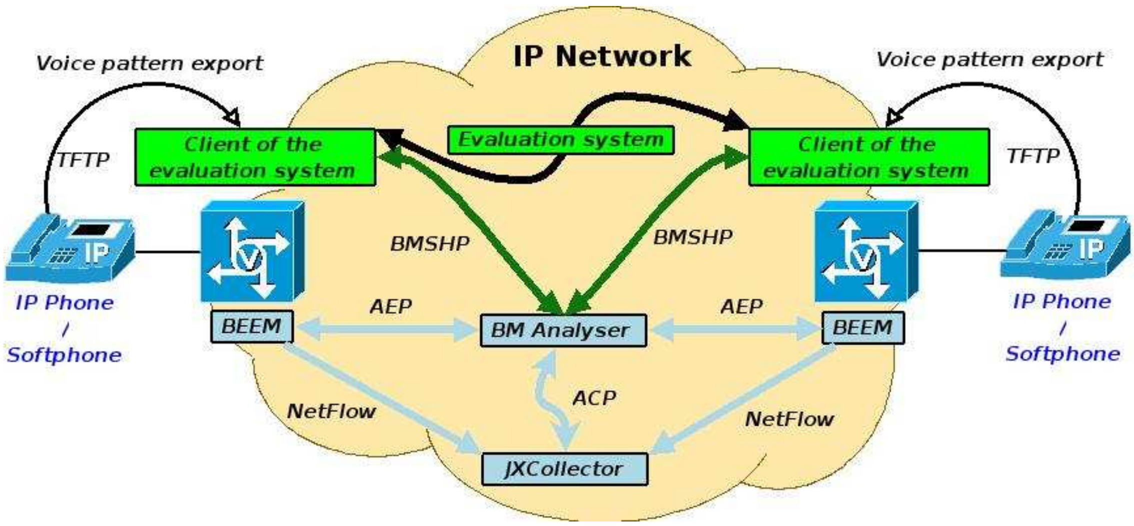
BEEM serves as the passive interceptor of the network communication using various sampling mechanisms and adaptive sampling using fuzzy regulator to export the information to BM analyzer using the AEP protocol (Analyzer Exporter Protocol). The role of BM analyzer is to assess the samples received by the lowest layer of the measuring tool (BEEM) and to determine the transmission and volume features of the transmitted packets. JXCollector serves to gather the information from several measuring points (in the actual implementation the measuring points are both sides transmitting the voice). The ACP protocol (Analyzer Collector Protocol) is implemented between the BM Analyzer and JXCollector to exchange information. The SQL database is used in order to archive the measured values for later processing.



Pic. 3. Measuring tool BasicMeter architecture

The BasicMeter tool includes the mechanisms of the measuring point time synchronization 5, adaptive sampling and the mechanisms using fuzzy regulator.

The active measurement process includes recording and analysis of the voice sample for the needs of the noise floor identification, signal/noise spacing, sample volume on the transmitting and receiving side that enter as variables in the E-model for calculation of the R-factor and its consecutive mapping for the MOS value (Mean Opinion Score). The Pic. 4 shows the combination of the Basic Meter measuring tools for measurement of the time and volume features of the transport medium with implementation of the system for exporting and analysis of the voice sample of the transmitted speech.



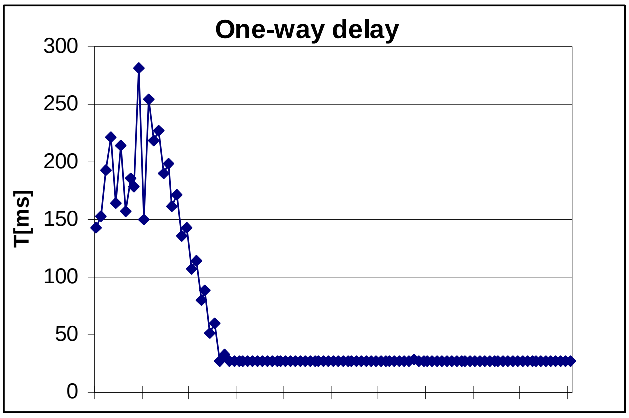
Pic. 4. Implementation of the voice sample export with the BasicMeter measuring tool

4. Experimental verification of the designed solution

The measurements in the model infrastructure were implemented for the purposes of the experimental verification of the proposed solution.

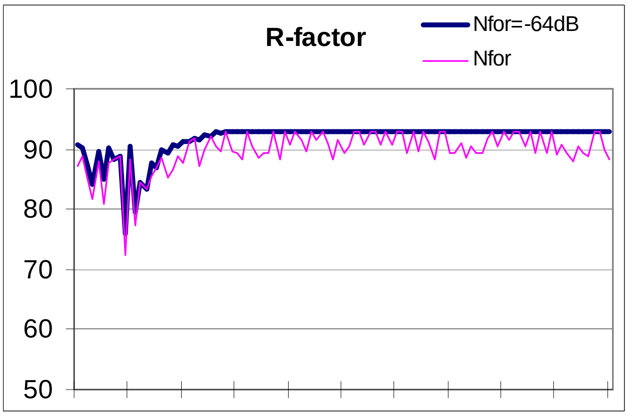
The active measurement process includes the measurement of the volume and time features of the transmission environment with the option to analyze the transmitted voice at present and the noise level in the transmitted voice on the transmitting and receiving side. Using the BasicMeter measurement tool it is possible to acquire the time features between two measurement points such as one-way and two-way delay or delay fluctuation. Based on this tool it is possible to gain also the volume feature of the packet loss rate that is necessary for calculation of the R-factor in accordance with the E-model.

Using codec G.711 one-way delay in the range from 29.06ms to 282.14ms was measured in a certain speech interval (Pic. 5).



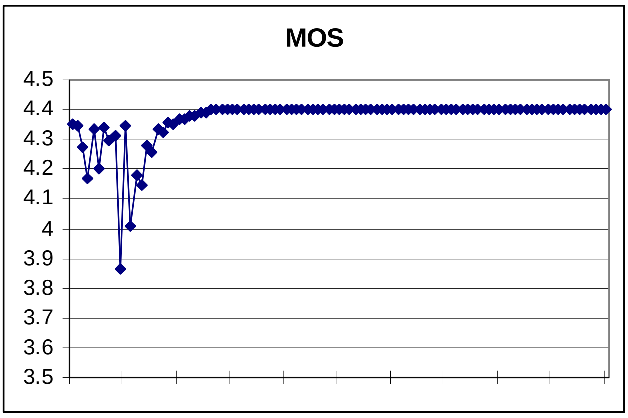
Pic. 5. One-way delay development

As in the development of the one-way delay the resulting R-factor should remain within the values from 92.8142 to 75.986. The chart of the R-factor development in the relevant speech time interval created by the system of the voice quality assessment is depicted in the Pic. 6. The voice quality assessment system included also the calculation of the noise floor; the resulting voice quality acquired by the assessment system was worse than in case of default noise floor (-64dB), but still adequate and corresponding to the relevant situation.



Pic. 6. R-factor development

As a result the R-factor development is depicted in the MOS assessment. The Pic. 7 shows the MOS assessment development provided by the assessment system for the relevant time interval with the default noise floor (-64dB).



Pic. 7. MOS assessment development

Conclusion

The experiments have shown that the assessment system is correctly designed and its results reflect the actual condition. Concurrently with the voice quality assessment system the speech was also subjectively assessed. The speech quality assessment by 10 participants resulted in MOS=4. Future activities in this field will be focusing on the selection of a suitable algorithm for the calculation of the signal-to-noise ratio and the consequent noise floor determination.

1. *Biran* G. Voice over Frame Relay, IP and ATM. The Case for Cooperative Networking. Accessible at: http://www.protocols.com/papers/voe.htm.
2. *ITU*-T Recommendation G.107. The E-Model, a computational model for use in transmission planning. 2003. Accessible at: http://www.itu.int/itudoc/itu-t/aap/sg12aap/history/g107/g107.html.
3. *Pollak P*. Estimation Methods of Speech SNR, Metody odhadu odstupu signálu od šumu v řečovém signálu – Methods of the estimation of the signal/noise spacing in the speech signal. (in the Czech language) In Akustické listy. – 2001. – Vol. 7. P. 10-12.
4. *Pollak P*. Efficient and Reliable Measurement and simulations of noisy speech background. In EUSPICO 2002. September 2002.
5. *Révés M.* Accuracy of Time Synchronization in Computer Networks using NTP, 7th PhD Student Conference and Scientific and Technical Competition of Students of the Faculty of Electrical Engineering and Informatics at the Technical University of Košice. Slovakia. 23 May 2007. – P. 161–162.
6. *Slezák J.* Určovaní poměru signál – šum, SNR kriteria – Determination of the signal/noise ratio, SNR criteria (in the Czech language). CTU. Prague 2001.
7. *Klimo M.* Calculation of Packet Loss Impact for E-model, EURESCOM workshop, Sophia Antipolis, France, February 2003, http:\\www.etsi.org
8. *Jakab F., Andoga R., Bača J., Giertl J.* Adaptive Sampling for Network Monitoring, ISAST Transactions on Communications and Networking: 2007. – N 1, 2007. – Vol. 1. – P. 52–61.