Investigation of Affects of Some Parameters of VoIP Networks on the Quality of Transmitted Speech

Bikes Agayev
Institute of Information Technology of ANAS, Baku, Azerbaijan
depart6@iit.ab.az

***Abstract*— The paper studies the assessment issues of the quality of transmitted speech in packet-switched networks. A hybrid method that simplifies the assessment process of the quality and reduces the result errors in comparison with existing methods is proposed.**

***Keywords*— speech signals; IP-telephony; methods of quality assessment; delayed delivery; packet loss**

I. Introduction

In recent years, packet technologies of voice transmission have been developed rapidly. In 2010 the volume of stored and transmitted digital information was already 986 exabytes, while this figure was 161 exabytes in 2006. More than half of this information is audio information [1].

Along with this, the demand for the quality of transmitted audio-messages through telecommunications networks is growing. Therefore, the development of effective methods for the quality assessment of speech signals in these networks is an important and urgent task. These questions are the most important not only for users but also for developers and creators of audio information networks and systems.

The issues of quality of speech, as well as its quality assessment, transmitted thourgh telecommunications networks and systems are in the focus of the international organizations, such as ITU-T (International Telecommunication Union), ETSI (European Telecommunications and Standard Institute) and STQ (Speech Transmission Quality). These organizations have developed some subjective (assessment is carried out by experts) and objective (technical - assessment is carried out by technical tools) methods. Currently, subjective expert-articulatory method of ITU-T REC. P.800 [2] and objective method E-model of ITU-T. Rec. G.107 [3] are adopted as a standardized methodology for quality assessing of the classical channel-switched telephone networks (PSTN, ISDN) and packet telephony (IP-telephony). In the first method the experts (speakers) of specially trained articulation teams vocalize the expressions of the standard syllable articulation tables at the input of the system. Other experts (auditors) accept them at the outlet of the system. The quality of accepted speech is assessed by psycho-physiological states of the auditors, i.e. by the comfort degree of acceptance of audio messages.

In [2] hundred-point system *R* (Quality Rating) and a five-point MOS (Mean Opinion Score) are the unit of speech quality measurement.

E-model as a technical method allows us to evaluate the speech quality by the calculation of the functions of some network parameters:

$$R = R_0 - I_s - I_d - I_e$$

where R_0 is R quality of the input signal, I_s, I_d, I_e are the degree of distortion caused by the quantization procedure and the echo signal, overall delay and quality of network hardware.

II. Problem Statement

Subjective methods developed by ITU-T based on Rec. P.800 have numerous serious drawbacks:

– attracting multiple articulating team, which requires significant financial expenditures and the procedure is very tedious;

– need for a specially equipped place with a low noise, etc.

Lack of E-model is that it is very difficult to calculate the value of R0, Id, Ie, particularly the latter one, as it depends on many factors, which are very difficult to take into account. Simplified versions of the model even have greater error. Both options do not allow to automate the process of quality assessment. Therefore, the development of the methods, that simplify the assessment and increase the accuracy of the results, is an urgent task.

III. Problem Solution

Recommendation of ITU-T Rec. P.800 (extra-ITU-T Rec. P.84 [4]) is a standardized methodology for subjective assessment of the quality of transmitted speech for the classical analogous, digital and packet networks telephony. It describes the conditions of tests performed by the articulation and expert team, the table of speech samples, the grading system and method for data analysis. Assessment is derived in rankings R units (R-quality) based on a hundred-point scale or in MOS units (MOS-quality) based on a five-point scale (Table 1).

TABLE I. CATEGORY OF SPEECH QUALITY BY ITU-T Rec. P.800

Psycho-physiological state of the auditor	Quality category	In MOS units	In R units
Complete absence of tension in conversation	best	4,34-4,50	90<R<100
One should be attentive during the conversation, but any significant efforts are not required	high	4,03-4,34	80<R<90
Constant, but not exclusively tension of attention is required during the conversation	medium	3,60-4,03	70<R<80
Conversation is possible, but has great difficulty	low	3,10-3,60	60<R<70
Conversation is impossible, or possible with great difficulty	poor	2,58-3,10	50<R<60

It is known that in the classical telephony the analogous audio signals (in particular, speech signals) of subscribers are submitted in the transmission network (communication) continuously and uniformly, i.e. in the rate of verbal exchange. Speech quality in these networks fundamentally depends on the amplitude-frequency characteristics of the channel, the noise and non-linear distortion. These parameters are quite stable, and a single method of modulation – PCM is used to encode the digitized speech.

The principles of packet switching of speech require applications of additional network equipments (different types of codec, vocal gateways, routers, etc.) and methods of speech signal processing (compression, packetization, routing, etc.). Consequently, there exist other factors that affect the quality of speech signals which are transmitted via IP networks. Therefore, criteria and methods that differ from those used for channel switching telephony networks are required to assess the quality of packet transmission of speech, in particular VoIP.

VoIP network, in a form of "phone-phone" connecting subscribers can be conditionally shown as composite network. The assessment of speech quality in the given composite network can be made by the subjective standard methodologies, according to [2, 4]. However, as noted above, these methods have some serious shortcomings. Therefore, the challenge is to develop such methods, where some procedural elements of the subjective methodology would be replaced by the elements of the objective methods.

As the Table 1. shows each value of R or MOS corresponds to a certain specific state of an expert, that tests the quality of the received speech. Such states of the expert can be emulated by the manipulation of certain parameters of the network of any type.

If we use the reasons causing the distortion of speech signals in IP-based networks as the affecting factors, the work of the experts could be replaced by measurement and computation procedures. The expediency of such a replacement arouses no doubt from a technological and organizational point of view, because measuring the network parameters is much easier than implementing expertise with the articulation team.

Obviously, the main causes of distortions of the speech signal in packet audio information networks are [5, 7]:

- delay in delivery of speech signals to the subscriber;
- packet loss in the network;
- packet jitter;
- codec type.

Consequently, according to the idea of the proposed method, the process of quality assessment is to determine the numerical values of these parameters, which cause physiological state of the experts listed in Table 1. Thus, using one-time service of the expert team, it is possible to determine the dependence of the quality of transmitted speech on these parameters separately. Obtained dependences, for example, in the form of graphs, can then be used to monitor the transactions of speech signals in IP-based networks.

As a result of the tests there were obtained graphic forms of dependencies between the quality of speech signals and the overall input-to-output delay, as well as packet loss in the corporate VoIP network (Fig.1, Fig.2). The experiments were conducted for two types of voice codec gateways: with standard G.711 (64 kbit/s) and low-speed G.729 (8.0 kbit/s).

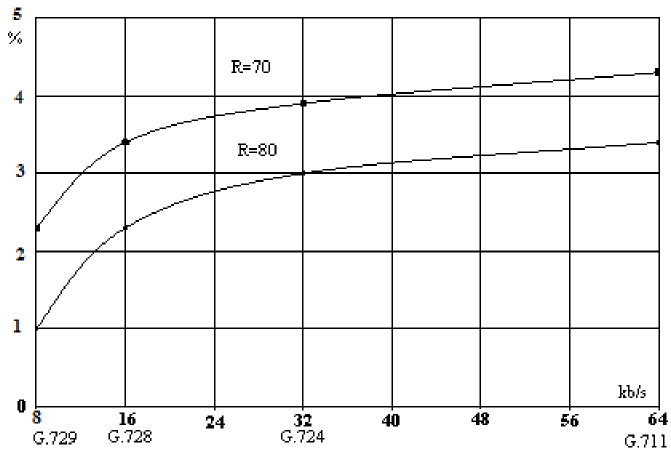

Figure 1. Dependence of the speech quality on the overall delay

Since, by definition, jitter is the maximum difference in delivery delay between the first and the last number of packages to the end IP network terminal (in this case, to the speech gateway), it can be considered as constituting the overall delay, so the dependence of the quality on the jitter is not analyzed separately. Impact of delays, the values of which are greater than 400 ms, is also not analyzed since, it is not recommended to create an IP network with the value of the permissible overall delay by Rec.G.114 and other sources [5, 6], which is greater than 400 ms.

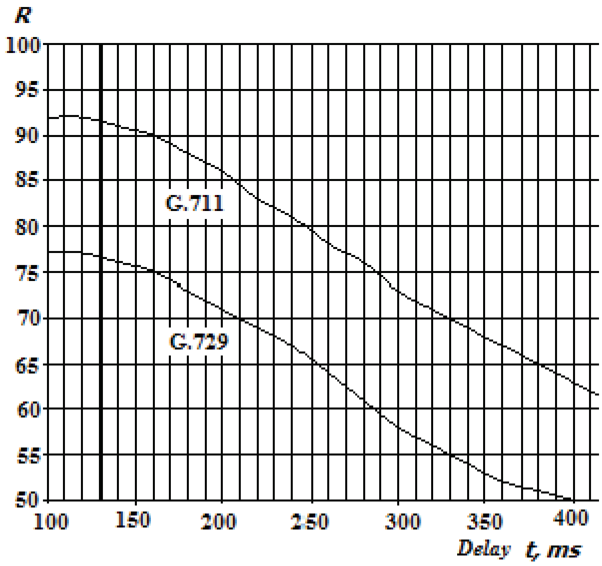

Figure 2. Dependence of the speech quality on the packet loss

The value of MOS quality is associated with R quality of a complex nonlinear dependence and requires more computation [3]. For practical work, a simple linear function can be used:

$$MOS = R/20 \tag{1}$$

However, an error of the computation in the nonlinear part is too large – about 20%. The piecewise linear function proposed by us can be applied for more accurate calculations.

Let's divide a priori, definitional domain of MOS = f(R) function into the N line segments. As a result, we obtain:

$$\left.\begin{array}{l} [R_0, R_1) \to [MOS_0, MOS_1) \\ : \\ [R_i, R_{i+1}) \to [MOS_i, MOS_{i+1}) \\ [R_{N-1}, R_N] \to [MOS_{N-1}, MOS_N] \end{array}\right\} \tag{2}$$

Each icon can be represented as a linear function:

$$f_i(R) = k_i R + b_i \tag{3}$$

where $R \in [R_{i-1}, R_i)$. k_i and b_i for each segment are determined by the method of least squares.

As a result, we obtain:

$$MOS = \begin{cases} k_1 R + b_1; R \in [R_0; R_1) \\ k_2 R + b_2; R \in [R_1; R_2) \\ : \\ k_N R + b_N; R \in [R_{N-1}; R_N] \end{cases} \tag{4}$$

There are several techniques to calculate the overall delay in the chain of IP networks. The simplest option is to calculate by technical indicators (passport data) of network equipments. This is especially practical in corporate IP networks, since the number of local (remote) segments corresponds to the topology of distributed organizations. In another version, the tests are carried out by software, for example, Psy VoIP of Psytechnics company, NetIQ Chariot of NetIQ company [7], though they are quite expensive. However, the convenient fact is that they can simulate sending traffic, entering different types of codec, defining test delays and the jitter length, as well as the packet loss values. For offline mode testing it is important to note that the more delayed components will be considered in a function of packet delivery period, the higher the accuracy of assessing the sound quality will be. For this purpose, for example, there can be used the calculation model of the overall delivery delay period proposed in [8].

IV. Conclusion

The paper proposes a hybrid (subjective and technological) method of the quality assessment of transmitted speech in telephony networks with packet switching, particularly in VoIP networks. The method makes it possible to realize an assessment of R and MOS speech quality by calculating the overall delay in delivery of audio messages to the end nodes and packet loss. This method simplifies the assessment process and increases the accuracy of the results, as an expert articulation team service is not used.

References

[1] Kharlamov A.A. Editorial paper. Speech technology. 2010, № 1, p. 111.
[2] ITU-T. Rec. P.800. Methods for subjective determination of transmission quality.
[3] ITU-T. Rec. G.107. The E-model a computational model for use in transmission planning.
[4] ITU-T. Rec. P.84. Subjective listening test method for evaluating digital circuit multiplication and packetized voice system.
[5] Tarasov B.Y., Drozdov K.I. Speech quality in IP networks. Communication Networks, 2007, № 12, pp. 39-45.
[6] Fomin A.F. Assessment of IP telephony network parameters. Information Technology, 2003, № 2, pp. 22-32.
[7] Kurilov O.S. An objective analysis of the speech quality in IP telephony. Technology and Communications, 2002, № 4, pp. 76-78.
[8] Agayev B.S., Pashayev F.H. Assessment and management of the speech quality in the distributed processing networks of packet switching audio information. News of ANAS. Series of Physical-Technical and Mathematical Sciences. Baku, 2010, XXX, № 6, pp. 145-150.