

Evaluation of VoIP Speech Quality Using Neural Network

Angel Garabitov and Aleksandar Tsenov

German Engineering and Industrial Management Education at Technical University of Sofia, Sofia 1756, Bulgaria

Abstract: The automatic quantification of the speech quality is an important and unsolved problem. The ability this task to be executed in real time is particularly interesting. Many parameters affect the voice quality, but only subjective assessments are decisive. The article makes a proposal to create cost effective and efficient solution for prediction the subjective speech quality assessment. A neural network is used as a “customer”. For training of the neural network, BYE Statistics is used. It gives the client’s view for the voice quality. Result of some practical experiments is given in article.

Key words: VoIP, quality of speech, neural network.

1. Introduction

The evaluation of the quality of VoIP telephony is usually a difficult task that is performed by the statistical processing of the satisfaction of a group of users. This paper proposes an idea for pseudo-subjective quality assessment of VoIP telephony. To evaluate the effects of various QoS-related parameters, we can use a neural network that can convert the multidimensional assessment criteria in simple relationships. For training the neural network are used various QoS-related parameters such as PSQM +, end-to-end delay, packet loss in VoIP system. Results confirmed that this method can effectively evaluate the overall QoS level QoS of several related factors.

The paper is organized as follows: Section 2 explains theoretical background; Section 3 describes the practical solution of the problem; Section 4 gives some results; Section 5 is the conclusion; Section 6 contains the future work.

2. Theoretical Background

There aren’t many methods for artificial subjective quality rating and analyze of perceived voice quality in VoIP. The determination of Subjective

measurement methods is a difficult task. Subjective measurements of QoS are carried out by use of a group of people [1-4]. A test phrase is recorded and then volunteers listen to it under various conditions. These tests are performed in special rooms, with background noise and other environmental factors, which are kept under control for test executions. Some examples are: conversation opinion test, hearing test review, an interview and survey test. There are also some disadvantages to this type of tests. These tests are very expensive and are not implemented in practice because of the large number of people is needed. The obtained results are truthful only when large numbers of people are engaged.

The existing listening tests allow the subjective assessment of speech. The goal of those tests is to evaluate the performance of the individual connections and algorithms under different conditions. Some of the known hearing tests are as follows: ACR (absolute category rating), DCR (degradation category rating) and CCR (comparison category rating).

Another possible method for quality analyzing is the use neural network with suitable topology and training data set. The education network in a specific order searches the training set. This order can be sequential or random. Some networks are trained without a teacher just once to search the sample. Other networks are trained with a teacher who browses some

Corresponding author: Angel Garabitov, M.Sc., research fields: communication systems and networks.

of the sample several times in the pass of learning. Even if they are successful at first glance, the behavior of the neural network is not always such that, what is required by the network. Assessment of quality of training of the neural network must be performed by examples that have not participated in its design. In addition, the number of test cases must be greater to achieve higher quality of the training of the network. It turns out that the testing of well trained neural networks is a very difficult task. The assessments of the quality related parameters are possible with the BYE message.

QoS measurement and speech quality prediction is an object of this research, which is a classical problem in VoIP networks. Upon examination of the problem, theoretically the most sense factors on speech quality are jitter, packets sent, lost packets, codec and quality.

All VoIP manufacturers [5] “Send In” the speech quality. This is shown in Fig. 1. As described, the input data of the neuron network are jitter (shortened to the BYE message with JI), packet sent (shortened to the BYE message with PS), packet loss (shortened to the BYE message PL) and codec type (shortened to the BYE message with EN).

The first variable is *Jitter*. In packet switched network, when a sequence of packets is send from

point A to point B, each of the packets will require a slightly different time to reach the destination. The various maturities are not a problem if a web page is downloaded, but it is when we want to transfer a stream of data in real time.

Formally the jitter is defined as the statistical variance of the RTP (real-time transport protocol) data packet inter-arrival time. In the RTP we measure jitter in timestamp units.

The first step to deal with jitter is to know its value. However, we do not need to calculate the exact value. In the RTP the receiving endpoint calculates an estimate with a simplified formula (a first order estimator).

The jitter estimate is sent to the other party using RTCP (real time control protocol).

The second and the third variables are *Packet loss* and *Packet sent*. Important and useful are the following measures:

(1) The average proportion of packets lost during the measurement period. This definition is used in VQmon, RTCP XR and RFC3550/RTCP.

(2) The average proportion of packets discarded due to late arrival at the receive jitter buffer. This definition is used in VQmon and RTCP XR. The average proportion of packets both lost and discarded

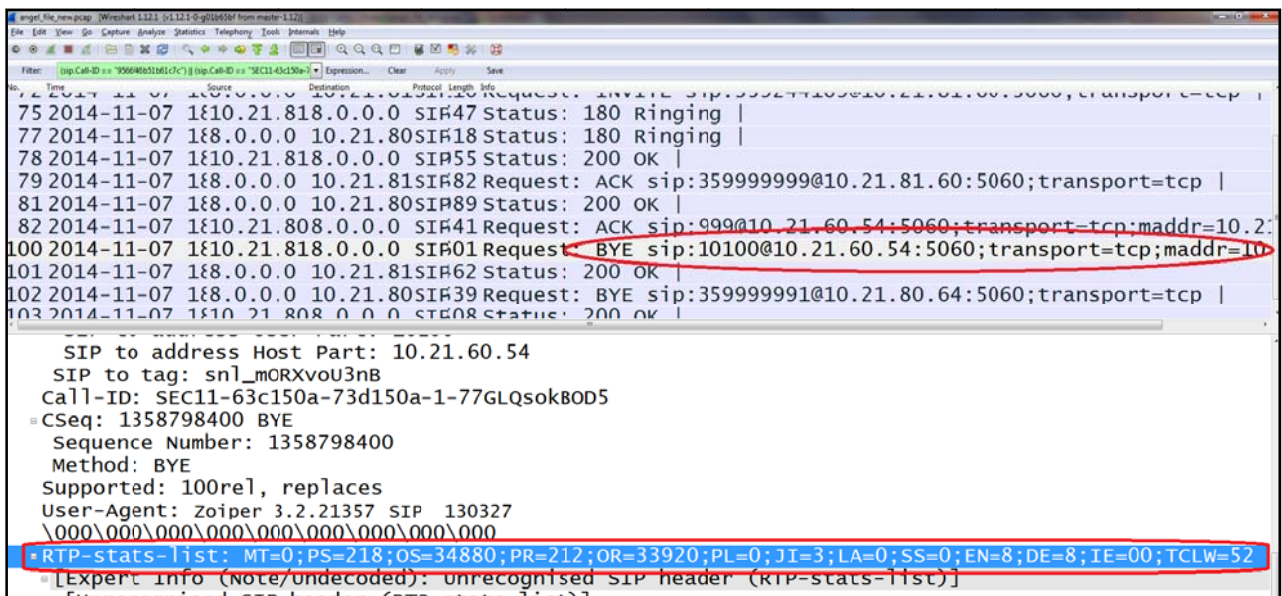


Fig. 1 BYE message.

during periods of low density loss, and the average length of these periods. A “gap” is often defined as a time period during which density loss of less than 5 percent and packet loss isolated or widely dispersed. This definition is used in VQmon and RTCP XR.

(3) VoIP receiving endpoints generally capture both packet loss and packet discard counter, so that you can determine for what type of problem you eliminate errors. Cisco phones include these statistics either in manual user interface on the phone set, or through their web interface. Avaya RTCP reports these statistics, the suddenly requires a separate RTCP receiver to capture the statistics. As VoIP control software is updated, you may find these statistics available within the VoIP controller.

The fourth variable is *Codec*. A codec converts the analog audio file into digital data packets and decodes

them back at the receiver. Most codecs use compression of data at the same time in order to reduce the bandwidth required for transmission. Mostly used codecs are G.711noPLC, G.711PLC and G.729A.

3. Practical Solution of the Problem

For construction of the neural networks, the application STATISTICA [6] is used. SANN (STATISTICA automated neural networks) is one of the most modern and efficient neural networks application on the market. It offers numerous unique advantages to neural network experts and to new users.

Experts have a wide range of network types and training algorithms. New users are directed through the necessary procedures for creating neural networks

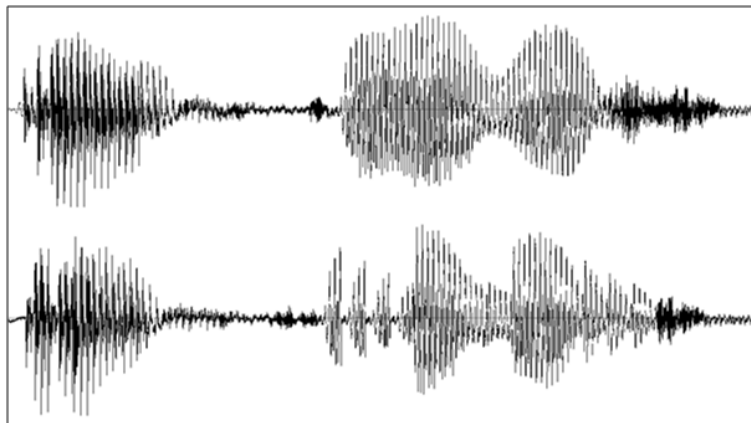


Fig. 2 Comparison between 5 and 20 ms Jitter.

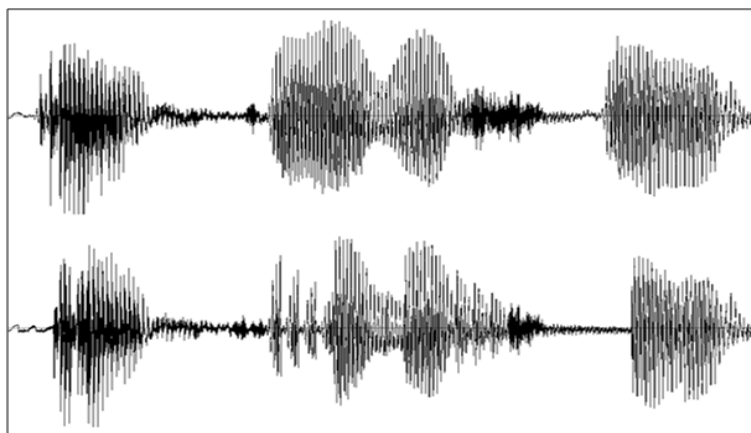


Fig. 3 Comparison between 2 % and 40% Packet loss.

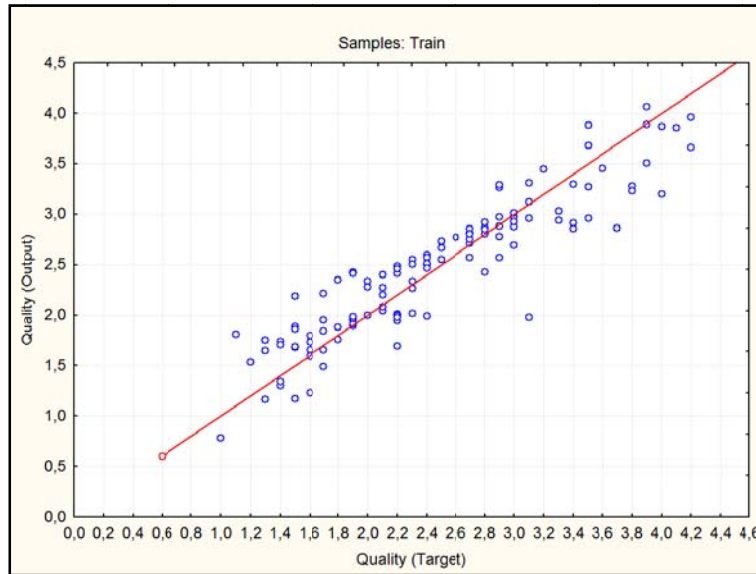


Fig. 4 Quality (output) as a function of quality (target).

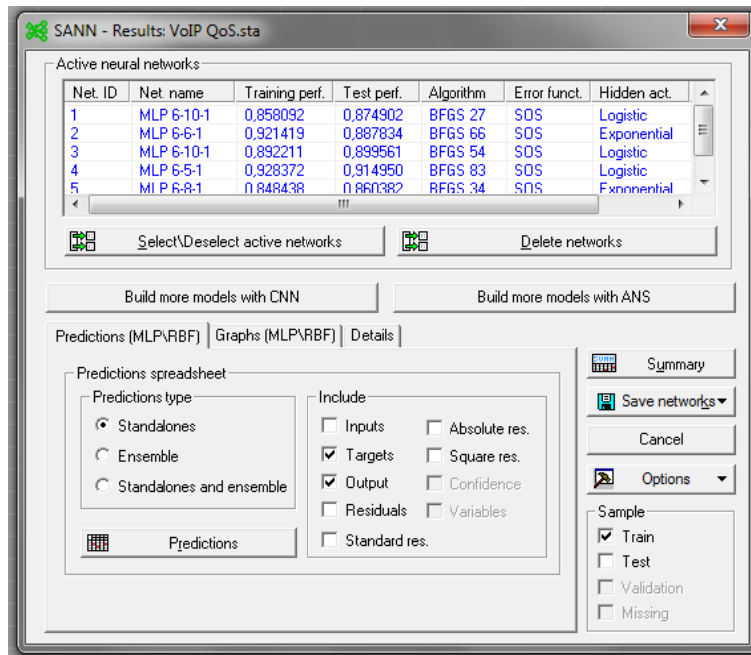


Fig 5 SANN results.

via the Automated Network Search tool. There is the possibility to save the trained network in C code.

These are the real samples of voice quality in relation to any parameter of the VoIP speech quality (jitter, packet sent, lost packet and codec). Special equipment generates the samples, which allows the perception of the real voice quality at various jitter or lost packets.

After the construction and training of the neural

network, all networks are tested. The best results are obtained by 6-10-1 MLP network with logistic activation function (Fig. 6).

To be sure, that this neural network solves the given task, it was tested by 150 results with known quality and input data (Figs. 4 and 5).

We think that this number of tests should be enough because of the constant average of the deviation of the voice quality error. The error can be decreased with

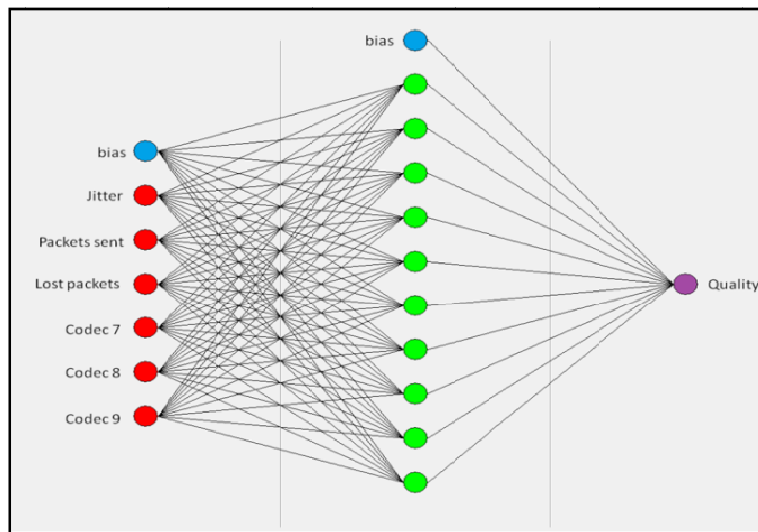


Fig. 6 The used neural network.

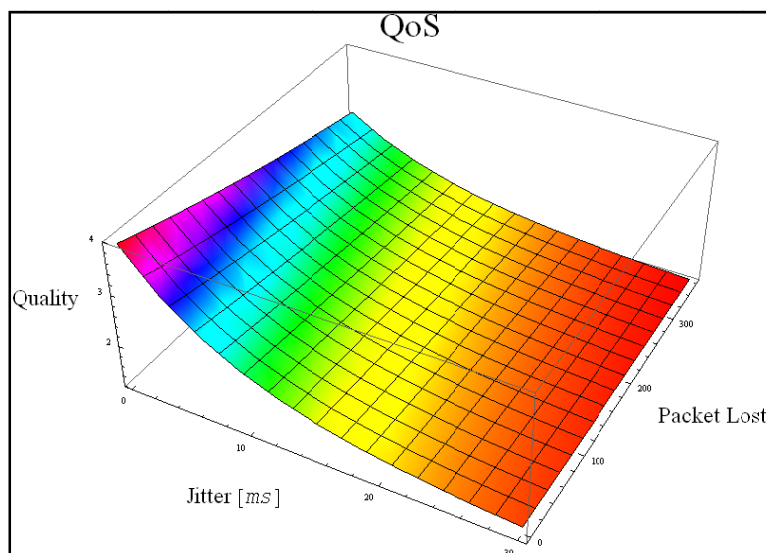


Fig. 7 ANN of 500 sent packets.

further train samples. Increasing the neuron number can also shrink the error.

The trained network can be exported in C source code. This code can be inserted into the server later. If needed, the network can be retrained with new data and the code recompiled.

4. Results

All graphics for the representation of the network will be presented when using the best codes (G.711PLC) and constant number of sent packages. Figs. 7 and 8 show the 3D graphics speech quality

depending on jitter and lost packets. On the x-axis is the jitter values are shown in milliseconds. The y-axis shows the variation of the lost packets. The third axis (z-axis) shows the speech quality in MOS LQ. It is easy to notice how with enlargement of jitter, the quality drops down. It is interesting that the lost packets do not affect the voice quality significantly.

The figure clearly shows that the parameters jitter and lost packets carry the vast information about call quality. Comparing the results obtained with the neural network, and those derived from the study of the subjective evaluation confirm this. Using BYE

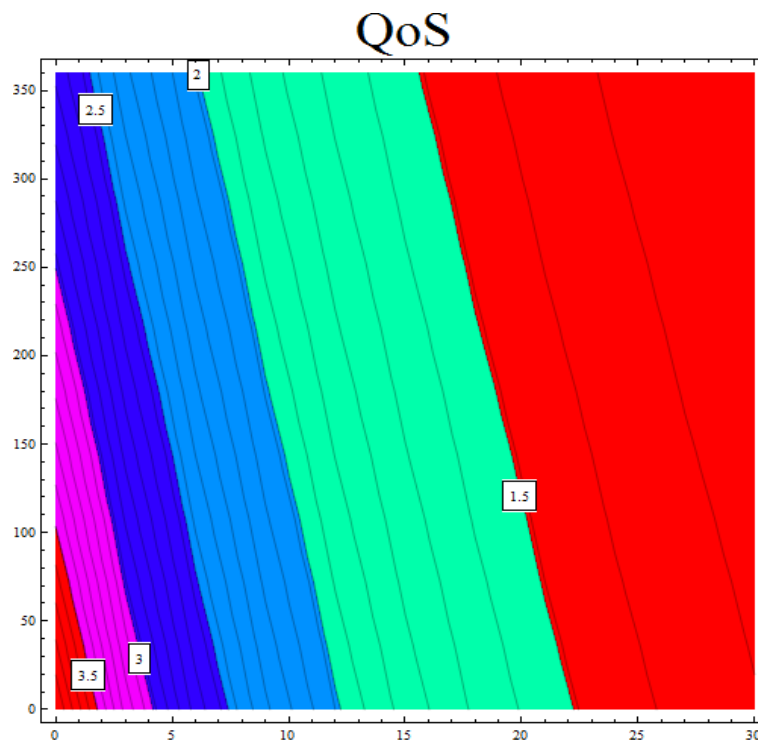


Fig. 8 ANN of 500 sent packets.

message receipt of digital expression of these two parameters is the least expensive way to obtain an objective assessment of quality.

Slightest deviation from the subjective results is comparable with the confidence interval of the study.

5. Conclusions

The proposed solution provides simple implementable QoS monitoring. As a result, this program is a cheap tool for QoS and VoIP network monitoring and for the VoIP network errors as well. This solution allows predicting the subjective user opinion of the voice quality only with the usage of statistics in the BYE-message and the corresponding OK message between the devices and the signaling server. The parameters considered by use of different codecs show, that the highest loss rate and the average loss burst size are provided by the packetization. In this study, a 0.874902 correlation coefficient between the prediction value and the result with the neural network as a subjective score is obtained.

6. Future Work

The use of a neural network, which assesses the voice quality subjectively, has a great advantage. The neural network module can easily improve the the quality of service by the use of the accumulated statistical data. In the updated module only the weights of the neurons are different from the original network. For update, only the weights in activating functions should be changed. After training the network, whose subjective evaluation is highly precious, corresponds to the customer evaluation.

References

- [1] Lakaniemi, A., and Rosti, J. 2001. "Subjective VoIP Speech Quality Evaluation Based on Network Measurements." Presented at the IEEE ICC.
- [2] Koumaras, H., Liberal, F., and Sun, L. 2008. "PQoS Assessment Methods for Multimedia Services." IGI Global publication, ISBN: 978-1-59904-820-8, July.
- [3] Lakaniemi, A., Rosti, J., and Raisanen, V. I. 2001. "Subjective VoIP Speech Quality Evaluation Based on Network Measurements." Presented at the IEEE ICC.
- [4] ITU-T Recommendation P.800. 1996 "Methods for

Subjective Determination of Transmission Quality.”
Revised in Geneva.

http://www.cisco.com/en/US/docs/ios-xml/ios/voice/cube_sip/configuration/15-2mt/voi-report-end-cal.html.

[5] CISCO. Accessed September 2012.

[6] <http://www.statsoft.com/>. 5.12.2014.