

# Evaluation of VoIP speech quality using neural network

Angel Garabito<sup>1</sup> and Aleksandar Tsenov<sup>2</sup>

**Abstract** – An important and unsolved problem is the automatic quantification of the speech quality. Particularly interesting is the ability of this task to be executed in real time. 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.

**Keywords** – VoIP, quality of speech, neural network

## I. INTRODUCTION

The evaluation of the quality of a VoIP telephony is usually a difficult task that is performed by the statistical processing of the satisfaction of a group of users. In this paper is developed the 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 been given 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.

## II. THEORETICAL BACKGROUND

### A. Using artificial subjective quality rating to analyze 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, 2, 3, 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, such as the experience, the humor and the culture of the subject. These tests are very expensive and are not implemented in practice because of the large number of people needed and running truthful results obtained.

<sup>1</sup>Angel Garabito is a former master student at the Faculty for German Engineering and Industrial Management Education at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: angel.garabito@fdiba.tu-sofia.bg.

<sup>2</sup>Aleksandar Tsenov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: akz@tu-sofia.bg

This type of test is used for unidirectional transmission, and it will have been carried out on transmission by examination of the recorded conversations or phrases. The goal of this test 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).

### B. Selecting an appropriate network topology

The education network in a specific order searches the training set. This order can be sequential or random. Some networks, that are trained without a teacher, just once to search the sample. Other networks that are trained with a teacher who browse some 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 is possible with the BYE message.

### C. Speech Quality

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 data packet inter-arrival time. In the Real Time Protocol 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.

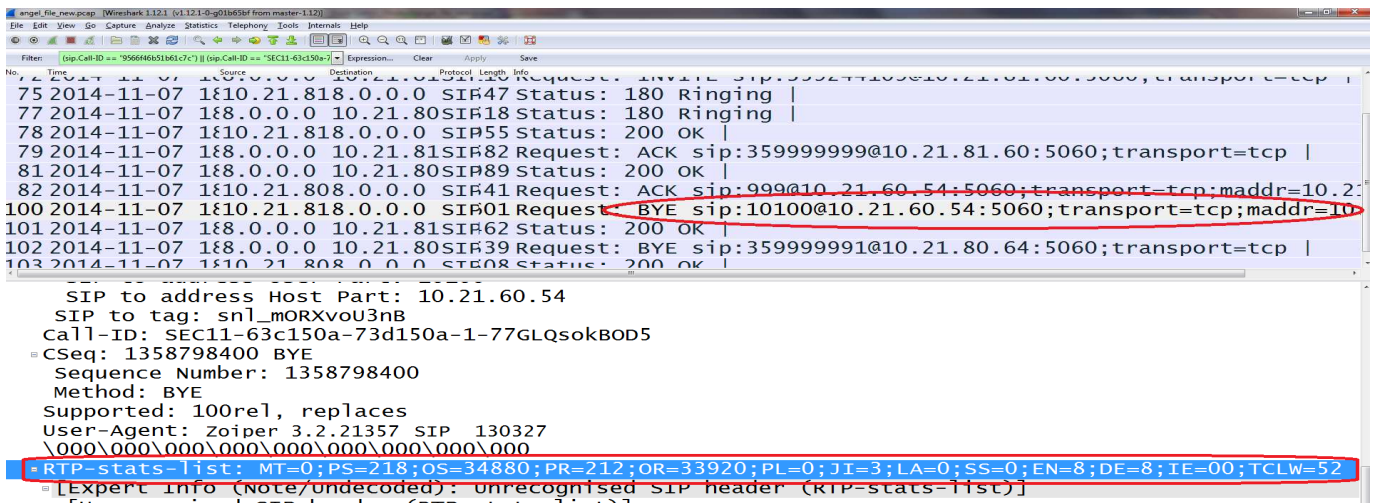


Fig. 1. BYE Message

Important and useful are the following measures:

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

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

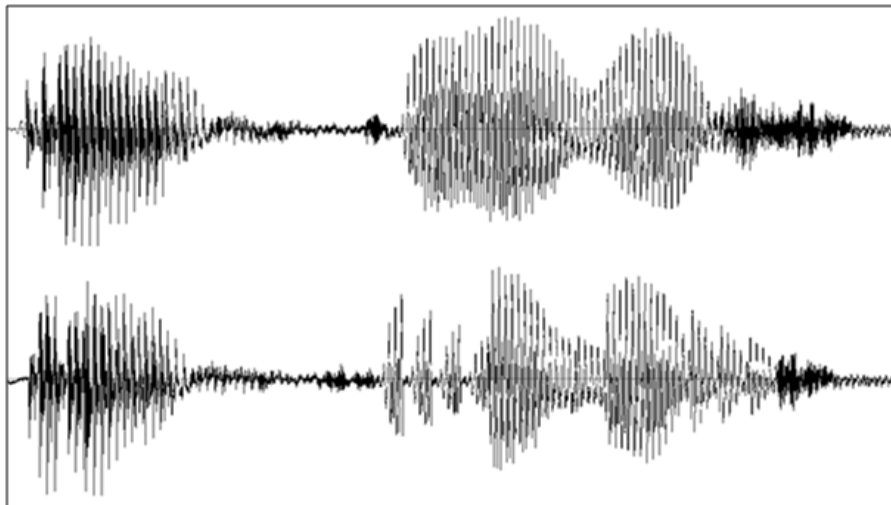


Fig. 2. Comparison between 5 and 20 ms Jitter

The average proportion of packets both lost and discarded 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.

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 report 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.

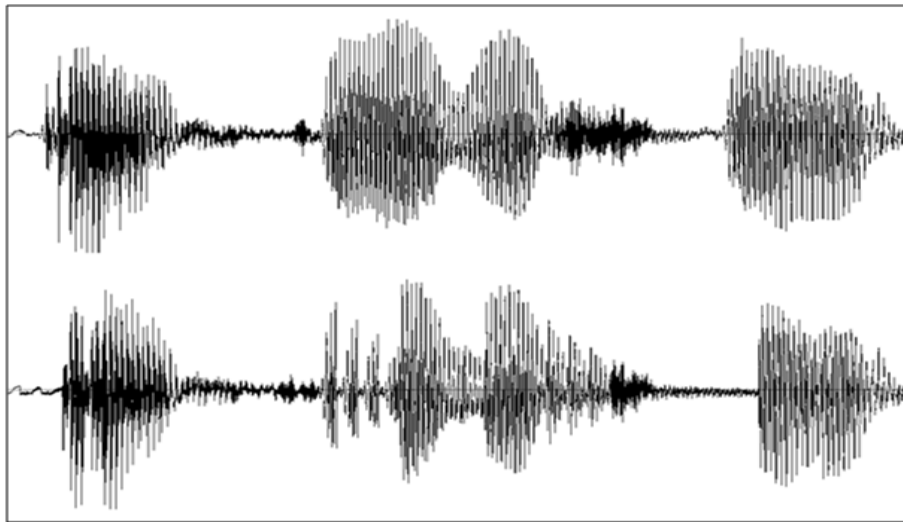


Fig. 3. Comparison between 2 % and 40% Packet loss

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.

### III. PRACTICAL SOLUTION OF THE PROBLEM

#### A. Building the neural network

For construction of the neural networks, the application STATISTICA [6] is used. STATISTICA Automated Neural Networks (SANN) 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 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). A 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. 4).

To be sure, that this neural network solves the given task, it was tested by 150 results with known quality and input data (Fig. 4). 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 further train samples. Increasing the neuron number can also shrink the error.

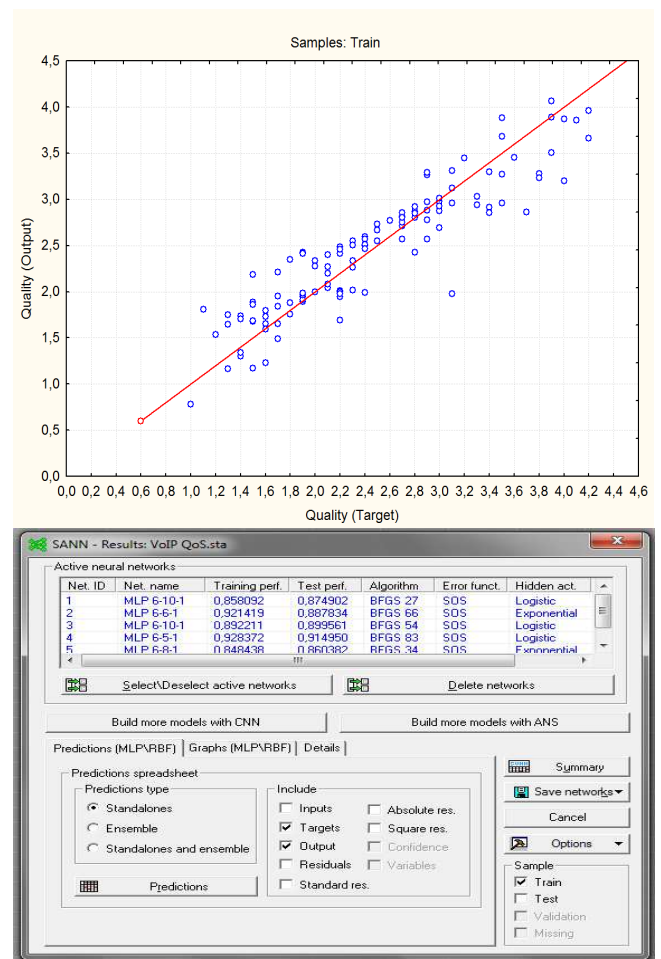


Fig 4. Quality (output) as a function of quality (target)

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.

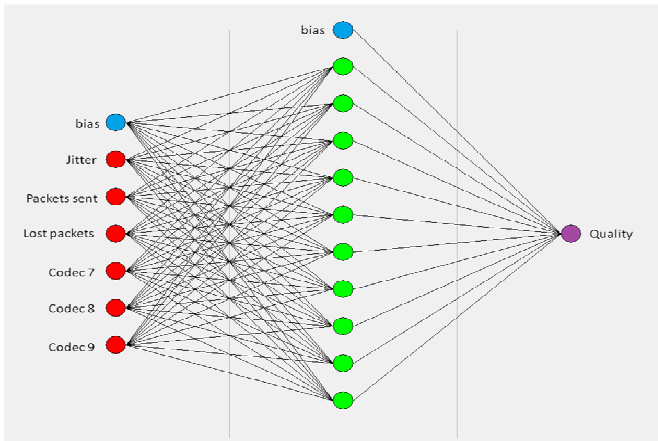


Fig 5. The used neural network

**B. 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. In Figure 6 is 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

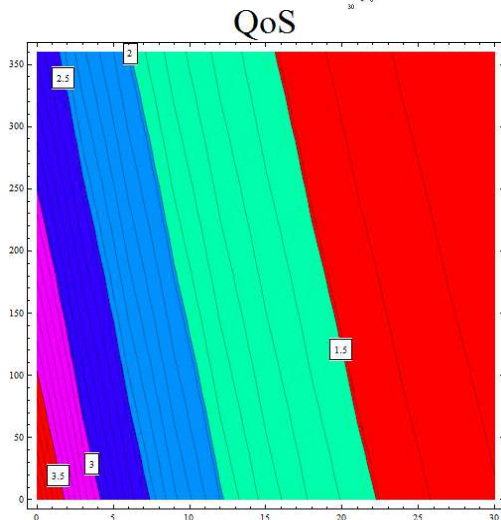
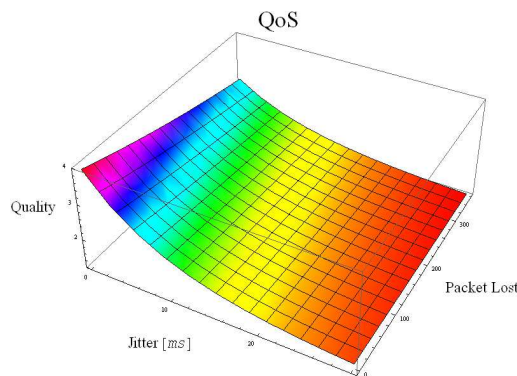


Fig 6. ANN of 500 sent Packets

with enlargement of jitter, the quality drops down. It is interesting that the lost packets does 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 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 are comparable with the confidence interval of the study.

**IV. CONCLUSION**

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 to predict 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,

**V. 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] A. Lakaniemi, J., "Subjective VoIP speech quality evaluation based on network measurements", Rosti, V. I. Raisanen IEEE ICC 2001;
- [2] H. Koumaras, F. Liberal, L.Sun, "PQoS Assessment Methods for Multimedia Services", IGI Global publication, ISBN: 978-1-59904-820-8, July 2008.
- [3] A. Lakaniemi, J., "Subjective VoIP speech quality evaluation based on network measurements", Rosti, V. I. Raisanen IEEE ICC 2001;
- [4] ITU-T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality", 1996;
- [5] [http://www.cisco.com/en/US/docs/ios-xml/ios/voice/cube\\_sip/configuration/15-2mt/voi-report-end-cal.html](http://www.cisco.com/en/US/docs/ios-xml/ios/voice/cube_sip/configuration/15-2mt/voi-report-end-cal.html), CISCO, 09.2012
- [6] <http://www.statsoft.com/>, 5.12.2014