

# Monitoring VoIP Call Quality Using Improved Simplified E-model

Haytham Assem, David Malone

Hamilton Institute, National University of Ireland,  
Maynooth  
Hitham.Salama.2012, David.Malone@nuim.ie

Jonathan Dunne, Pat O'Sullivan

Systems and Performance Engineering, IBM Dublin,  
Software Lab  
Jonathan\_dunne, patosullivan@ie.ibm.com

**Abstract**—ITU-T recommendation G.107 introduced the E-model, a repeatable way to assess if a network is prepared to carry a VoIP call or not. Various studies show that the E-model is complex with many factors to be used in monitoring purposes. Consequently, simplified versions of the E-model have been proposed to simplify the calculations and focus on the most important factors required for monitoring the call quality. In this paper, we propose simple correction to a simplified E-model; we show how to calculate the correction coefficients for 4 common codecs (G.711, G.723.1, G.726 and G.729A) and then we show that its predictions better match PESQ scores by implementing it in a monitoring application.

**Keywords**-VoIP; E-model; PESQ; Monitoring

## I. INTRODUCTION

The evaluation of data networks depends on several factors. Thus, it is argued that it is not appropriate to use a single metric to evaluate the quality of data networks. Yet in the telephony world, a single number is typically given to rate call quality. Such value is used as a basis of monitoring and tuning the network. Voice over Internet Protocol (VoIP) is an example of such data network application [1].

In previous years, VoIP has become an important application and is expected to carry more and more voice traffic over TCP/IP networks. In real-time voice applications, the speech quality is impaired by the packet loss, jitter, delay and bandwidth. Consequently, VoIP applications require low delay, low packet loss rates, low jitter and sufficient bandwidth in order not to affect the interaction between call participants.

VoIP is based on IP network; however IP networks frequently provide best effort services, and may not guarantee delay, packet loss, and jitter [2]. So, the prediction of voice quality in different environments and traffic loads may be as important part of network monitoring in order to measure voice quality and prevent critical problems before they occur.

As measuring voice quality is important to the service providers and end users, ITU-T provides two test methods subjective and objective testing. Subjective testing was considered the earliest attempts on this issue to evaluate the speech quality by giving Mean Opinion Scores (MOS). The MOS test is one of the widely known accepted tests that give a speech quality rating. ITU-T Rec. P.800 [3] presents the MOS test procedures as users can rate the speech quality from 1(Poor) to 5 (Excellent) scale. Of course, the numbers

of the listeners are considered an important factor in estimating accurate scores. Thus, subjective testing using MOS is time consuming, expensive and does not allow real time measurement. Consequently, in recent years new methods were developed for measuring MOS scores in an objective way (without human perception): PESQ [4], E-model [5] and several others.

PESQ, Perceptual Evaluation of Speech Quality, is considered an objective method for predicting the speech quality. It is an intrusive testing method which takes into account two signals; one is the reference signal while the other one is the actual degraded signal. Both signals are sent through the test that uses a PESQ algorithm and the result is a PESQ score. Consequently, this approach cannot be used to monitor real time calls.

Nowadays, a new objective method proposed by TU-T G.107 [5] defines the E-model, a mathematical model that combines all the impairment factors that affect the voice quality in a single metric called R value that is mapped to MOS scale. The E-model was designed to provide estimated network quality and has shown to be reasonably accurate for this purpose. It has not been accepted as a valid measurement tool for live networks. The ITU-T G.107 Recommendation [5] states at the beginning of the document that “*it is considered only estimates for the transmission planning purposes and not for actual customer opinion prediction*” unlike the PESQ [4] which is developed to model subjective tests commonly used in telecommunications to assess the voice quality by human beings.

Increasingly and against ITU recommendations, the E-model is being used nowadays by industry and research as a live voice quality measurement tool. Thus, simple versions of E-model [1, 6] have been proposed to simplify the complexity of the original E-model [5] and focus on most important part that affect the VoIP call quality.

The objective of our work is to provide a monitoring system using a simplified version of the E-model corrected for 4 common codecs to better predict PESQ MOS scores as PESQ is generally considered to provide more accurate predictions of user experience than the E-model.

This paper is organized as follows: Section 2 describes the proposed improved simplified E-model. In Section 3 we show how we derived the correction coefficients used in the improved simplified E-model. In Section 4 we propose our results using the derived model by implementing it in a monitoring application. Finally, we conclude and summarize the paper in Section 5.

## II. IMPROVED SIMPLIFIED E-MODEL

In this section, we will first give a brief description of the simplified E-model [6] and then we will describe our proposed improvements to the simplified E-model with the method of calculation of the various parameters used in the model in order to be applicable in monitoring purposes.

### A. Simplified E-Model

The original E-model is very complex [7] and involved with many factors. Moreover, the voice processing is not related significantly to the instantaneous judgment of QoS. Thus, a simplified version of the E-model [6] has been introduced to focus on the most important parts and afterwards it was used in a monitoring system [2]. This model takes in to account the codec and the present network conditions which are the main two factors that affect the voice quality. The simplified E-model is expressed by equation (1) by calculating the evaluation value R.

$$R = R_0 - I_{\text{codec}} - I_{\text{packetloss}} - I_{\text{delay}} \quad (1)$$

Where  $R_0$  represents the basic signal to noise ratio,  $I_{\text{delay}}$  represents the delays introduced from end to end,  $I_{\text{codec}}$  is the codec factor and the  $I_{\text{packetloss}}$  is the packet loss rate within a particular time. Finally the R value is mapped to MOS score.

### B. Improved simplified E-Model

The objective of this model is to determine the voice quality MOS rating by a simplified modified version of the previous E-model described above. The computational model consists of a mathematical function of parameters of the transmission system. The computation itself can be split into several elements and can be expressed by the following equation (2).

$$R = R_y - I_d + A \quad (2)$$

Where  $R_y$  is a second order function corrected using curves fitted to PESQ scores which is the standard objective method defined by ITU-T recommendation P.862 [4],  $I_d$  is the average delay time within specified period and A is the expectation factor due to the communication system. The description and method of calculating the previous parameters ( $R_y$ ,  $I_d$  and A) in (2) are as follows:

#### 1) $R_y$ :

$R_y$  as mentioned above is a second order function model corrected with PESQ scores to obtain more accurate results in our monitoring system.  $R_y$  can be expressed by the following equation (3).

$$R_y = aR_x^2 + bR_x + c \quad (3)$$

Where  $R_x$  is a part of the simplified E-model (1) which is corrected with PESQ scores,  $R_x$  can be obtained by the following expression (4) and a, b, c are codecs coefficients as shown in Table I and derived in section III.

$$R_x = R_0 - I_e - I_{p,l} \quad (4)$$

TABLE I. CODECS SPECIAL COEFFICIENTS

Codec	a	b	c
G.711	0.18	-27.90	1126.62
G.723.1 5.3k	0.039	-4.2	166.61
G.726 24k	0.046	-4.53	168.09
G.729A	0.063	-8.08	311.72

#### 1.1) $R_0$

$R_0$  is the basic signal to noise ratio, including noise sources such as circuit and room noise. However, currently it is really difficult to calculate  $R_0$  directly. Thus, ITU-T G.113 [8] provides the common value of  $R_0$ . Since, the inherent degradation that occurs when converting actual spoken conversation to a network signal and back reduces the theoretical maximum R-value (94.2) with no impairments to 93.2 [5]. So, we set the  $R_0$  value to 93.2.

#### 1.2) $I_e$

$I_e$  is the equipment impairment (codec quality) factors as defined in [8] and [9]. It represents the codec distortion which leads to voice distortion and impairments arising because of signal conversions. Nowadays, its value is determined by looking up the codec in the ITU-T Recommendation G.113 literature [8] as Table II is part of it.

TABLE II. SOME CODING INFORMATION

Encoder Type	References	BitRate (Kbit/s)	Ie value
PCM	G.711	64	0
ACELP	G.723.1	5.3	19
ADPCM	G.726	24	25
CS-ACELP	G.729A	8	11

#### 1.3) $I_{p,l}$ :

$I_{p,l}$  is the packet loss percentage within a particular period measured by certain number of packets. The percentage measured is the loss of packets occurred when the sender's packets is not received by the receiver. It can be expressed by the following formula (5).

$$I_{p,l} = \left(1 - \frac{N}{DS}\right) \times 100\% \quad (5)$$

Where DS is the difference between the largest and smallest sequence number of N packets. Statistics and calculation of the Real-time Transport Protocol (RTP) packets can be used to calculate this percentage by the following expression (6).

$$DS = LS - SS + 1 \quad (6)$$

Where LS and SS are the largest and smallest sequence numbers respectively. They are extracted from the RTP header of the sequence number field from the packets received.

2)  $I_d$ :

The delay components contributing to  $I_d$  provided in ITU-T G.107 [5] are  $T_a$ , the average absolute one way mouth to ear delay.  $T$ , the average one way delay from the receive side to the point in the end to end path where a signal coupling occurs as a source of echo.  $T_r$ , the average trip delay in the 4 wire loop. G.107 [5] gives a fully analytical expression for the function  $I_d$ , interms of  $T_a$ ,  $T$ ,  $T_r$  and parameters associated with a general reference connection describing various circuit switched and packet switch inter-working scenarios. Assuming perfect echo cancellation, all the factors in  $I_d$  can be collapsed in a single points as shown in (7) and  $I_d(d)$  is now function only of the one way delay  $d$ .  $I_d(d)$  can be calculated by a series of complex equations in ITU-T G.107 [5] as shown with the plotted curve of  $I_d$  vs one way delay in Fig.1 (labeled “G.107”).

$$d = T_a = T = T_r / 2 \quad (7)$$

The one way delay ( $d$ ) is the time it takes to get data across the network. The one-way delay measured from one end of the network to the other end is mainly composed of four components that can be expressed in equation (8).

$$d = t_0 + t_1 + t_2 + t_3 \quad (8)$$

Where  $t_0$  is the propagation delay,  $t_1$  is the transport delay,  $t_2$  is the packetization delay and  $t_3$  is the jitter buffer delay.

In this paper we approximate these four components by measuring the response time (round-trip delay) as in most modern devices  $t_1, t_2$  shall be small. Thus, ping should be reasonable.

In our model we used a simplified version of (9) as provided in [10]. This model shows accuracy for one way delay less than 400ms as shown in Figure 1 (labeled “AT&T simplified model”). We found this model reasonable as ITU-T recommend that one-way delay should not be more than 150 ms for good speech quality[11].

$$I_d = 0.024d + 0.11(d - 177.3)H(d - 177.3) \quad (9)$$

Where

$$H(x) = 0, \text{ if } x < 0$$

$$H(x) = 1, \text{ if } x \geq 0$$

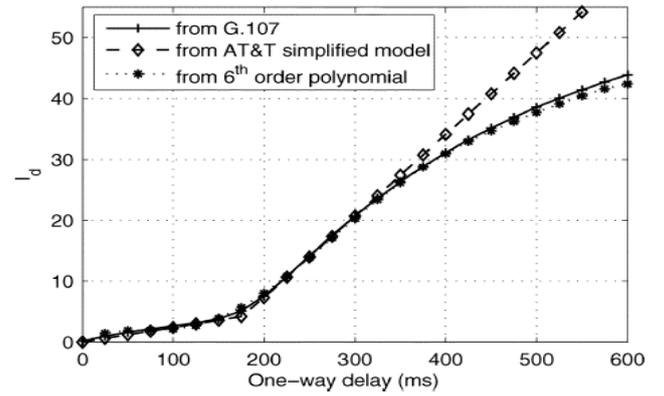


Figure 1.  $I_d$  versus one-way delay

3) A:

The advantage factor, A represents an “advantage of access”, introduced into transmission planning for the E-model (ITU-T G.107) [5]. This value can be used directly as an input parameter to the E-model. Provisional A values are listed in [5] as show in Table III. Assuming our communication system is conventional then we neglect A value.

TABLE III. EXAMPLES FOR THE ADVATAGE FACTOR A

Communication System	Maximum value of A
Conventional (wire bound)	0
Mobility by cellular networks in a building	5
Mobility in a geographical area or moving in a vehicle	10
Access to hard-to-reach locations, e.g., via multi-hop satellite connections	20

The R value of the E-model is finally transformed to MOS score that will reflect the user level of satisfaction as shown in Table IV, theoretical range of transmission performance rating factor R from 0 to 100. R=0 represents of the worst quality and R=100 represents the best quality. The R factor value for estimated average score of MOS can be expressed by equation (10).

For  $R < 0$ : MOS=1

For  $0 \leq R \leq 100$ :

$$MOS = 1 + 0.035 \times R + R(R - 60)(100 - R) \times 7 \times 10^{-6}$$

For  $R > 100$ : MOS=4.5 (10)

TABLE IV. RELATIONSHIP BETWEEN R-VALUE AND USER'S SATISFACTION

R-Value	Satisfaction Level	MOS
90-100	Very satisfied	4.3+
80-90	Satisfied	4.0-4.3
70-80	Some users dissatisfied	3.6-4.0
60-70	Many users dissatisfied	3.1-3.6
50-60	Nearly all users dissatisfied	2.6-3.1
0-50	Not recommended	1.0-2.6

### III. CORRECTION OF THE SIMPLIFIED E-MODEL

In this section we show how we derived the values of a, b and c (Table I) used in our improved simplified E-model described in the previous section.

In our experiment shown in Figure 2 we have developed a java program that stream RTP packets using 4 main audio codecs (G.711, G.726, G.723.1 and G.729A). We recorded the voice at both ends and measured the PESQ scores under different random packet loss rate ranges from 0-20%. For each packet loss rate, we repeated the experiment 10 times taking the average MOS PESQ score in order to increase the accuracy of the results as much as possible.

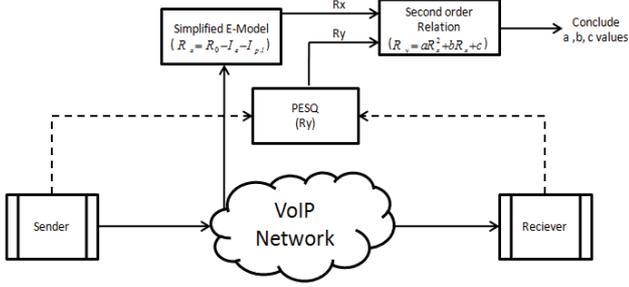


Figure 2. Deriving codecs coefficients a, b and c

The PESQ scores are converted from MOS to R value and this can be conducted by a complicated Candono's Formula as in [12] or by the simplified 3<sup>rd</sup>-order polynomial fitting [13] as shown in (11).

$$R = 3.026MOS^3 - 25.314MOS^2 + 87.060MOS - 57.336 \quad (11)$$

The converted PESQ scores from (11) will represent the  $R_y$  values shown in the below graphs (Fig.3-Fig.4) on the y axis. Since PESQ does not take the delay factor in its account, so we correct the rest part of the model which we name it  $R_x$  (see equation 3) represented on x axis.

We found that it is well matched second order relation function and then we derived the coefficients a, b and c as in Table I using least-squares fitting method. The graphs below (Figure 3-Figure 4) show the correlation between the converted values from PESQ and the R values from the simplified E-model for individual codecs in different loss ranges.

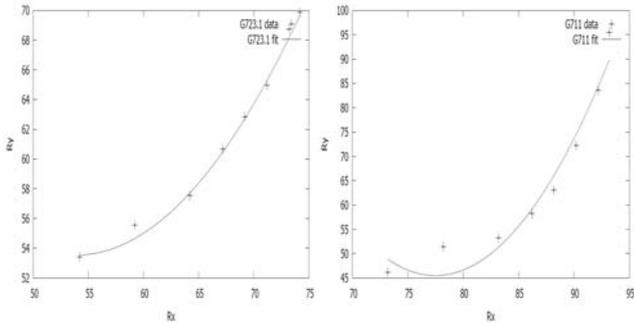


Figure 3. Relationship between  $R_x$  and  $R_y$  (G723.1 and G711)

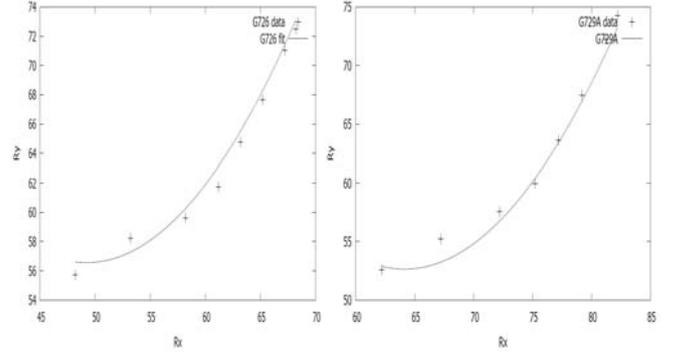


Figure 4. Relationship between  $R_x$  and  $R_y$  (G726 and G729A)

### IV. MONITORING SYSTEM DESIGN AND RESULTS

The monitoring system could target specific number of RTP packets to capture and perform an effective MOS value calculation. The system will use a coefficient database for the codec used in the call. This monitoring system is developed for monitoring VoIP quality at the network terminals, and the environment could be a personal or family network with voice quality monitoring.

The whole system works as follows: The system uses network capturing module to capture a certain number of packets passed to specific IP and port. Non RTP packets will be filtered. When this process completes the packet capture, the system will analyze the data, delay and packet loss rate as described previously in section II. The MOS score is calculated to assess voice call quality in this period of the call. We took our results online with introducing random packet loss rate in the network using Dummynet [14].

We compared our monitoring system using MOS scores based on the codec's coefficients (see Table I) derived for 4 main codecs with the simplified version of the E-model that is used in monitoring purposes [1, 6] and the PESQ scores. The graphs (Figure 5-Figure 8) show our results for the 4 codecs. It can be observed that the MOS scores of our improved simplified E-model based on the coefficient database (Table I) are very close to the PESQ scores unlike the simplified E-model which gives an advantage for the corrected model in monitoring purposes for the VoIP call quality.

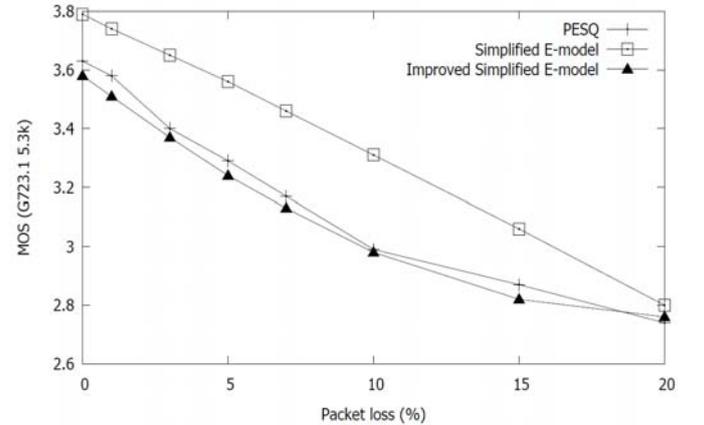


Figure 5. Comparative Analysis (G.723.1)

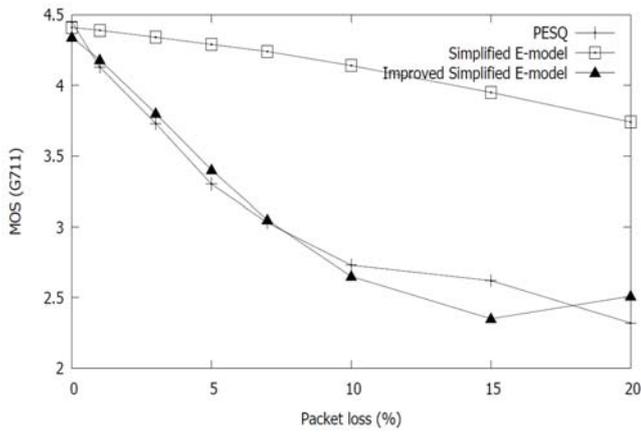


Figure 6. Comparative Analysis (G.711)

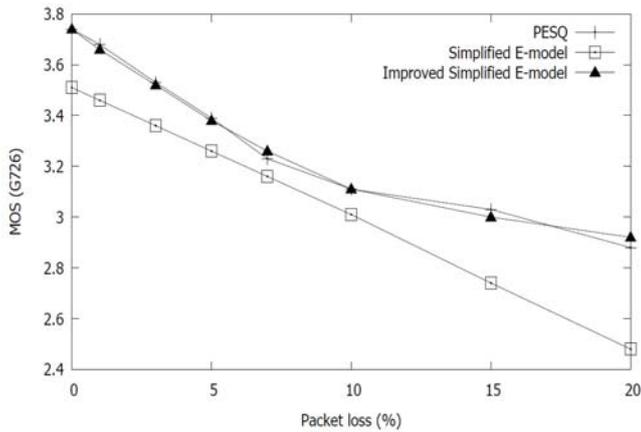


Figure 7. Comparative Analysis (G.726)

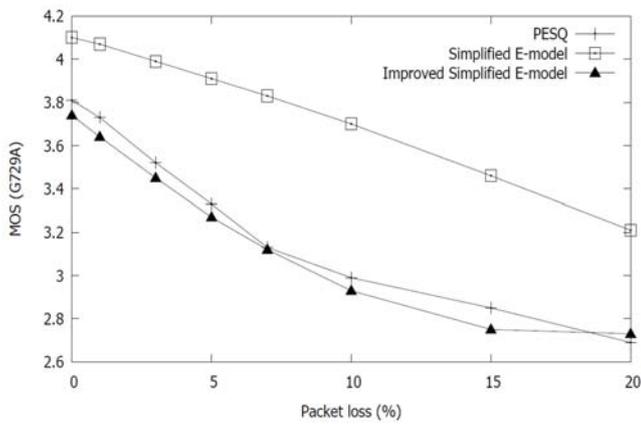


Figure 8. Comparative Analysis (G.729A)

## V. DISCUSSION AND CONCLUSION

The E-model brings a new approach to the computation of estimated voice quality. The main advantage of using E-model that it is classified as an objective non intrusive method that can be applied in real time. On contrary to the ITU-Recommendation, simplified versions of E-model have been introduced by researchers and industry to be used for monitoring purposes and predicting the VoIP call quality.

Consequently, we have proposed an improved simplified E-model and show how we derived the coefficients used in the model for 4 common codecs (G.711, G.723.1, G.726 and G.729A). We demonstrate its results by implementing it in a monitoring system; our system analyzes the impact of voice quality encoding factors under various network conditions and uses our simplified improved E-model to assess voice quality. The main advantage of our improved simplified version that, it is less complex than the original E-model model and it is more accurate than the simplified versions used.

We stress three benefits of our work. The first as confirmed by the experiment, the simplified version of E-model does not provide accurate results compared to PESQ scores. The second, the correction coefficients derived enhance the simplified E-model to monitor/predict the call quality. The third, proposing a complete design of monitoring system using our improved simplified E-model for 4 common codecs. Another output of our work is a java application that stream RTP packets using number of codecs.

## ACKNOWLEDGMENT

The authors were supported by Science Foundation Ireland (SFI) grants 07/SK/I1216a and 08/SRC/I1403.

## REFERENCES

- [1] John Q. Walker, "Assessing VoIP Call Quality Using the E-model", NetIQ Corporation.
- [2] Junsheng Zhang and Xiaohua Sun, "The VoIP phone QoS protection in the wide-area network", *Computer learning* .2006 No.6 .17-18.
- [3] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality", Geneva, 08/1996.
- [4] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality(PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", February 2001.
- [5] ITU-T Recommendation G.107, "The E-model: A computational model for use in transmission planning", Geneva, 04/2009.
- [6] Chunlei Jiang and Peng Huang, "Research of Monitoring VoIP Voice QoS", *International Conference on Internet Computing and Information Services*, 2011.
- [7] Pystechincs Limited, "The E-Model, R Factor and MOS", 23 Museum street Ipswich, Suffolk United Kingdom. December 2003. 10-41.
- [8] ITU-T Recommendation G.113, "Transmission impairments due to speech processing", 2001.
- [9] ITU-T Recommendation P.833, "Methodology for derivation of equipment impairment factors from subjective listening-only tests", 2001.
- [10] R.G.Cole and J.Rosenbluth, "Voice over IP performance monitoring", *ACM comput. Commun. Rev.*, vol. 31, no. 2,pp. 9-24, April 2001.
- [11] S.Pracht and D.Hardman: "Voice Quality in Converging Telephony and IP Networks, Agilent Technologies", White Paper, <http://literature.agilent.com/litweb/pdf/5980-0989E.pdf>
- [12] C. Hoene, H. Karl, and A. Wolisz, "A perceptual quality model for adaptive VoIP applications", *Int. Symp. Performance Evaluation of Computer and Telecommunication Systems(SPECTS'04)*, SanJose, CA.
- [13] L. Sun, "Speech Quality Prediction for voice Over Internet Protocol Networks", Ph.D dissertation, Univ. Plymouth, UK., Jan 2004.
- [14] Marta Carbona and Luigi Rizzo, "DummyNet Revisited", *ACM SIGCOMM Computer Communication Review* Volume 40 Issue 2, April 2010, 12-20.