

Voice Quality Evaluation of a Call Using Fuzzy Logic

Oyetade Durojaiye*, Elizabeth. N. Onwuka

Telecommunications Engineering Department, Federal University of Technology (FUT) Minna, Niger State, Nigeria

Abstract Voice will remain a fundamental communication media that cut across people of all walks of life. It is therefore important to make it very affordable. VoIP has been increasingly popular in recent times due to its affordability, however, poor reliability and voice quality remain important factors that limit the widespread adoption of VoIP systems. Good voice quality is a key factor for users transiting from the Public Switched Telephone Network (PSTN) to VoIP networks. It has been shown that line echo is one of the key factors that deteriorate voice quality in VoIP systems. Several non-real-time algorithms have been developed in literature to estimate various aspects of voice quality in VoIP systems. But there is no real-time algorithm that estimates the echo content of a VoIP conversation, which could enable the operator take some corrective actions to improve the quality while the call is in progress. In this paper, the authors propose a real-time fuzzy algorithm to estimate the strength of the line echo component of the voice quality in VoIP networks. The results obtained shows that the algorithm is able to track and estimate echo content of a live VoIP traffic in real-time. This algorithm could be embedded in VoIP systems to enable operators monitors calls in real-time.

Keywords VoIP, Quality of Service (QoS), Mean Opinion Score (MOS), Perceptual Evaluation of Speech Quality (PESQ), Perceptual Analysis/ Measurement System (PAMS), Perceptual Speech Quality Monitor (PSQM), ETSI- Computation Model (E-model)

1. Introduction

A modern real-life telephony network very often consists of local Public Switched Telephone Network (PSTN) that collects voice traffic from users; and is connected to a long-distance network optimized for data transport by means of the Internet Protocol (IP). The interconnection of the PSTN to the IP network is implemented through a Media Gateway (MG), which can convert the PSTN signals into the format required by the IP, and vice versa[1][9]. In order to ensure good voice quality, several non-real-time algorithms that estimate the voice quality such as the Perceptual Analysis Measurement System (PAMS), Perceptual Evaluation of Speech Quality (PESQ), Perceptual Speech Quality Monitor (PSQM) and the European Telecommunications Standards Institute (ETSI) Computation Model (E-model) have been developed.

A major hindrance in carrying voice traffic over data networks is the increased echo content, which is as a result of longer delays encountered in these networks. Echo detection and cancellation is therefore critical in achieving good quality voice signal in packet switched networks, which face longer delays due to the bursty nature of the IP network[10].

1.1. Theoretical Background

In most cases our everyday conversations take place in the presence of echoes. We hear echoes of our speech waves as they are reflected, for instance, from the floor and the walls. However, if the reflected waves arrive shortly after we spoke them, we do not perceive them as echo but as some reverberation. On the other hand if the reflected wave takes 20 or 30 milliseconds (ms) to come back to us, we will identify it as an annoying echo[5].

Echo is mainly dependent on the amount of delay present in the circuit or network. Most callers will hear echo of their own voice if the circuit contains as little as 30 milliseconds of round-trip delay. If the round-trip delay approaches 50 ms, virtually all callers will complain of echo if it is left uncontrolled[3][6]. Echo is closely related to other factors such as delay, jitter and packet loss that affect the voice quality of a VoIP call.

Delay is introduced into the telecommunications network primarily by transmission facilities and transmission equipment. The delay could be negligible or significant[12]. Depending on the network topology, and the type of transmission equipment used in the network, 30 ms of roundtrip delay can occur in connections that are across country or just across town[7].

The International Telecommunication Union (ITU) has a guide for the amount of delay introduced by specific transmission medium as shown in Table 1.

* Corresponding author:

tade@sainttade.com (Oyetade Durojaiye)

Published online at <http://journal.sapub.org/ijnnc>

Copyright © 2012 Scientific & Academic Publishing. All Rights Reserved

Table 1. Some typical transmission facility delays

Transmission Facility	Delay per 100 miles
T1 carrier over copper	1ms
Fiber optic cable	1ms
Microwave Radio	0.7ms

Jitter is another problem encountered by voice calls; it is caused by retransmission of lost packets, which is directly linked to delay in the IP network. Delay could be caused by different factors already mentioned above.

Packet Loss can occur for a variety of reasons; these include link failure; traffic congestions; and misrouted traffic. In an IP environment, the packets could be retransmitted or re-routed and this causes delay[2][3].

1.2. Related Works

Researchers have worked on evaluating the voice quality of a VoIP calls by improving on previous methods mainly PAMS, PESQ and the E-Model. Modified E-Model[2][8], an introduction of a Packet-based Echo Canceller and others have also been proposed, but none of these has been real-time. The main reason for the delayed success of the VoIP might be attributed to the fact that the Internet was designed to be a fault tolerant data exchange/transmission medium, and that traffic re-routing is the primary target in the case of Internet server additions and removals. The delivery time or transmission delay was never the primary concern in the design phase although the Internet Protocol (IP) has support for real-time transmissions[2]. According to Periakarruppan *et al.*:[3], the common problems that occur when a VoIP network is utilized are echo, delay, jitter, and loss of packets. All this problems are closely inter-linked.

As is described by Ditech Networks[14], there are three classes of objective voice quality evaluation metrics: (i) network-parameter based metrics, (ii) psycho-acoustic metrics, and (iii) elementary metrics.

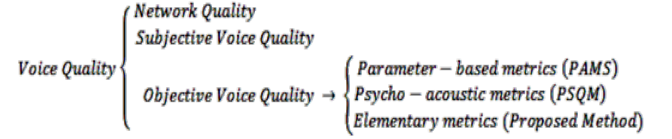
(i) Parameter-based metrics do not consider the actual voice signal. Instead, these metrics sum impairment factors that characterize the individual components of the communication system. For instance, in the E-model the packet loss and delay in a VoIP system are translated into impairment factors. Parameter-based metrics such as the E-model hold promise for predicting subjective voice quality but still requires extensive refinements and verifications.

(ii) Psycho-acoustic metrics transform voice signals to a reduced representation to retain only perceptually significant aspects. These metrics aim to predict the subjective quality over a wide range of voice signal distortions. One example of such metric is the PESQ algorithm

(iii) Elementary objective voice quality metrics rely on low-complexity signal processing parameters and techniques to predict subjective voice quality. Elementary metrics generally have smaller correlations with subjective voice quality than highly complex psycho-acoustic metrics and do not provide the perception modeling needed for psycho-acoustic coder algorithm development. However,

elementary metrics represent a good engineering tradeoff for communication and networking system researchers and developers in that they allow for fairly detailed conclusions about voice quality while having low computational complexity.

We summarize the various aspects of voice quality evaluation in Figure 1.

**Figure 1.** Classification of voice quality algorithms for VoIP systems

To mitigate the annoying effect of echo in voice traffic and improve on user perceived QoS, it is necessary to estimate the echo content of voice traffic in real-time. This could help an operator or the system take a corrective action such as switching a user to a less echo laden channel while a call is in progress. In this work we propose a simple, objective, low computationally complex algorithm using fuzzy inference system to evaluate the echo content of the voice traffic over VoIP network in real-time. The proposed algorithm is to address the non real-time nature of algorithms mentioned at the beginning of this section.

The remainder of this paper is organized as follows. Section II presents our proposed real-time echo detection solution. Section III presents a detailed performance evaluation of our proposed solution by employing extensive simulations using MATLAB. The work is concluded in Section IV.

2. Proposed Algorithm

In this section we develop an algorithmic tool for evaluating the echo content of voice traffic using a fuzzy inference system. We used the objective evaluation technique for the following reasons:

1. Contrary to the subjective quality method, an objective method can be automated since it does not require human intervention or feedback. If well designed, such a method can also estimate the quality in real-time.

2. Contrary to the network quality method, an objective method takes into consideration the user's perception of the call quality, and not only parameters that qualify the performance of the IP network.

The proposed fuzzy algorithm estimates the echo component of the voice quality factors and serves as a building block for an objective, passive, voice quality algorithm based on elementary metrics. As discussed earlier, the critical issues in delivering good voice quality over IP networks are: packet loss, delay, echo and jitter. These issues are all correlated, but there is a stronger correlation between jitter, delay and packet loss. Jitter in VoIP systems is normally compensated for by using a playout buffer at the receiving end, which introduces delay and additional packet loss. So

we can imagine a fuzzy inference system that evaluates the voice quality in a VoIP network described by Figure 2:

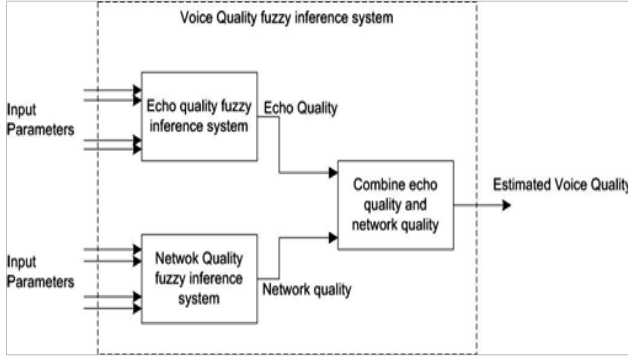


Figure 2. Fuzzy inference system to estimate the voice quality in a VoIP network

Different echo cancellers may estimate different sets of parameters. Therefore, we need extra computations to estimate the required parameters for the algorithm proposed here. The input parameters that we get from a standard echo canceller are:

1. Echo return loss (ERL) which is the amount of echo attenuation provided by the hybrid .
2. Combined loss (ACOM) which is the sum of the ERL, cancellation loss and nonlinear processing loss,
3. Receive speech power - an estimate of the speech power in the receive path.
4. Receive noise power - an estimate of the noise in the receive path.
5. Transmit speech power - an estimate of the speech power in the send path after the echo cancellation.
6. Transmit noise power - an estimate of the noise in the send path.

In order to obtain a real-time, low computationally complex algorithm, we chose to use as inputs to our fuzzy inference system four parameters that are already being computed or estimated by the echo canceller, such as estimates for the Echo return Loss (ERL) and the Combined Loss (ACOM), receive speech power estimations and transmit noise power estimations.

Table 2 shows the fuzzy sets associated with the input parameters proposed.

The three output membership functions which will give an estimate of the echo content of a voice traffic can be described by the following equations:

The membership function for the fuzzy set “bad echo (be)” is given as:

$$\mu_{be}(x) = \begin{cases} 1 - 2x, & 0 \leq x \leq 1/2 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The membership function for the fuzzy set “moderate echo (me)” is given as:

$$\mu_{me}(x) = \begin{cases} 2x, & 0 \leq x \leq 1/2 \\ 2(1 - x), & 1/2 < x \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The membership function for the fuzzy set “good echo

(ge)” is given as:

$$\mu_{ge}(x) = \begin{cases} 2x - 1, & 1/2 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Eqs (1) – (3) are represented graphically in Fig 3.

Table 2. Fuzzy sets associated to the input parameters

Fuzzy Set	Description
<i>Good ERL</i>	Represents values of ERL that will help the echo canceller to realize a good echo cancellation.
<i>Bad receive speech power</i>	The receive speech powers in this set are either too low or too high, making it difficult for the echo canceller to generate the signal that must be subtracted in the send path.
<i>Bad transmit noise power</i>	Represents values of the transmit noise that may disrupt the convergence of the adaptive filter.
<i>Bad ACOM</i>	With high probability, VoIP systems with ACOM values in this set will have echo problems and the voice quality will be bad.
<i>Moderate ACOM</i>	Represents values of ACOM that may indicate that the echo cancellation was not good enough and some echo may be leaked to the far end.
<i>Good ACOM</i>	VoIP systems with ACOM values in this set are able to cancel most of the echo in the calls.

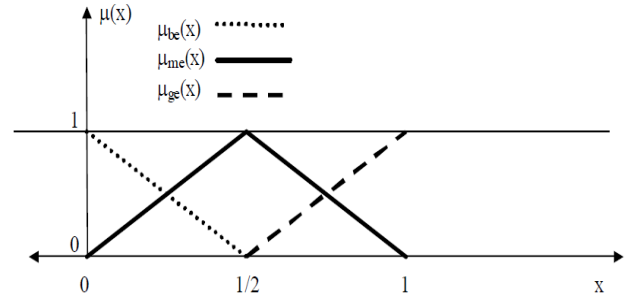


Figure 3. graphical representation of the output membership functions for Echo content

The approach we used is to define the input membership functions based on empirical reasoning, then we spent some time tuning those membership functions. But the tuning is done after the fuzzy inference system is designed because we need to use the output of the algorithm as a feedback for tuning. The next paragraph describes the proposed fuzzy inference system with the complete set of fuzzy rules, operations, and defuzzification.

The fuzzy rules together with the fuzzy membership functions are the main elements that reflect the empirical reasoning behind the proposed fuzzy inference system. Table 3 describes the fuzzy rules that we adopted for our proposed algorithm and the empirical reasoning behind each rule. The fuzzy implication operator that we chose for our algorithm is Larsen Operator. The defuzzification method that we chose is the center of mass (centroid) method (Ross, 2004). An advantage of using fuzzy logic is that we can first define the fuzzy input variables and elaborate the fuzzy rules and then we can tune the membership functions by running the algorithm for calls for which we know the MOS. That is exactly what we did in order to define the following membership functions for each one of the fuzzy inference input variables.

Table 3. Fuzzy rules to evaluate the echo component of the voice quality

Fuzzy Rule	Empirical reasoning
IF ACOM is bad THEN echo is bad.	The ACOM is a major parameter for estimating the quality of the echo signal. If the ACOM is bad, most probably the user is perceiving echo.
IF ACOM is good, THEN echo is good.	With a good ACOM, some echo is being cancelled successfully, independently of the other parameters.
IF ACOM is moderate AND ERL is good THEN echo is moderate.	If ACOM is moderate, there is some uncertainty about the quality of the echo signal. So we use the ERL to better estimate it.
IF receive speech power is bad AND transmit noise is bad THEN echo is bad	The signal levels for transmit and receive speech as well as for transmit noise are all contributing to a bad echo signal.

We used Matlab and its fuzzy logic toolbox to implement our proposed algorithm and ran a set of 16 calls. We implemented it in a way that we could feed Matlab 2008a version with different fuzzy inference systems at a time. The difference between the fuzzy systems was only in terms of the membership functions. All systems had the same input variables, fuzzy operations, and fuzzy rules, but different membership functions for the fuzzy variables. Then, after running all 16 calls for each one of the 5 fuzzy inference systems we could compare the results of the fuzzy algorithm to the expected MOS scores (after the echo cancellation).

We chose to use 5 different fuzzy inference systems in each tuning step because there are so many parameters that can be changed in a membership function that one easily gets lost if one tries to change several parameters at once. So, for instance, if we are tuning a specific triangular membership function that has a positive and a negative slope we would first tune the positive slope of the triangle and try it with say 3 to 5 different positive slopes. As a result of the tuning process described above we derived the following membership functions for the fuzzy sets shown in Table 2.

The membership function for the fuzzy set “good ERL (gerl)”

$$\mu_{gerl}(x) = \begin{cases} x - 20/10, 20 \leq x \leq 30 \\ 1, x > 30 \\ 0, otherwise \end{cases} \quad (4)$$

The membership function for the fuzzy set “bad ACOM (bacom)”

$$\mu_{bacom}(x) = \begin{cases} 23 - x/17, 6 \leq x \leq 23 \\ 1, x < 6 \\ 0, otherwise \end{cases} \quad (5)$$

The membership function for the fuzzy set “moderate ACOM (macom)”

$$\mu_{macom}(x) = \begin{cases} x - 12/11, 12 \leq x \leq 23 \\ 36 - x/13, 23 < x \leq 36 \\ 0, otherwise \end{cases} \quad (6)$$

The membership function for the fuzzy set “good ACOM (gacom)”

$$\mu_{gacom}(x) = \begin{cases} x - 23/17, 23 \leq x \leq 40 \\ 1, x > 40 \\ 0, otherwise \end{cases} \quad (7)$$

The membership function for the fuzzy set “bad receive speech power (brsp)”

$$\mu_{brsp}(x) = \begin{cases} -25 - x/5, -30 \leq x \leq -25 \\ 0, -25 < x < -15 \\ x + 15/10, -15 \leq x \leq -5 \\ 1, otherwise \end{cases} \quad (8)$$

The membership function for the fuzzy set “bad transmit noise power (btntp)”

$$\mu_{btntp}(x) = \begin{cases} x + 45/9, -45 \leq x \leq -36 \\ 1, x > -36 \\ 0, otherwise \end{cases} \quad (9)$$

3. Performance Evaluation

To evaluate the performance of our proposed real-time echo detection solution, extensive simulations were performed using MATLAB for a one-to-one VoIP voice call scenario. In this section, we present a detailed description of the evaluation process and results obtained.

The proposed algorithm was run for 16 calls with different levels of voice quality. These calls were all previously generated during a live VoIP call and recorded in PCM format. The calls were splitted into two groups. The first group contained 8 calls in which echo was effectively cancelled or no echo was present in those calls. The second group contained calls with various levels of echo signals and background noise. Then, we passed these calls through a Texas Instruments DSP platform with a line echo canceller. We then collected the measurements from this echo canceller, analyzed these measurements and finally used them as inputs to our fuzzy algorithm. As described in the previous section, the collected measurements from the echo canceller were the ERL, ACOM, transmit speech power, receive speech power, transmit noise power and receive noise power as shown in Table 4.

Table 4. Average reading of 16 VoIP Calls

Call No	ERL (dB)	ACOM (dB)	Recive Speech power (dBm)	Transmit Noise power (dBm)	Fuzzy Output	Call Quality
1	20.00	21.80	-46.50	-11.60	0.441	Bad
2	20.30	22.30	-45.50	-11.61	0.383	Bad
3	20.39	22.33	-50.24	-15.55	0.412	Bad
4	20.39	22.34	-49.34	-13.67	0.413	Bad
5	20.41	22.37	-49.50	-13.10	0.423	Bad
6	20.96	21.19	-47.70	-14.14	0.402	Bad
7	24.38	28.55	-46.06	-18.55	0.562	Bad
8	21.98	21.52	-43.53	-13.20	0.470	Bad
9	25.90	34.90	-47.50	-15.70	0.780	Good
10	27.33	35.47	-48.15	-19.41	0.810	Good
11	27.23	35.34	-45.61	-18.52	0.803	Good
12	26.80	35.09	-49.64	-18.67	0.790	Good
13	27.90	35.67	-48.75	-20.55	0.820	Good
14	27.67	35.43	-48.60	-20.86	0.808	Good
15	27.38	35.28	-49.34	-19.14	0.800	Good
16	26.92	35.11	-49.19	-17.89	0.791	Good

For each simulated VoIP call we had the near end signal, the far end signal and the measurements provided by the echo canceller. As was described, the proposed algorithm gives an instantaneous estimation of the echo signal for the call. During the simulations these estimations were computed every two seconds. The aim was to provide an average of the outputs of the fuzzy inference system throughout the call duration. The results obtained are shown in Figs 4,5, and 6.

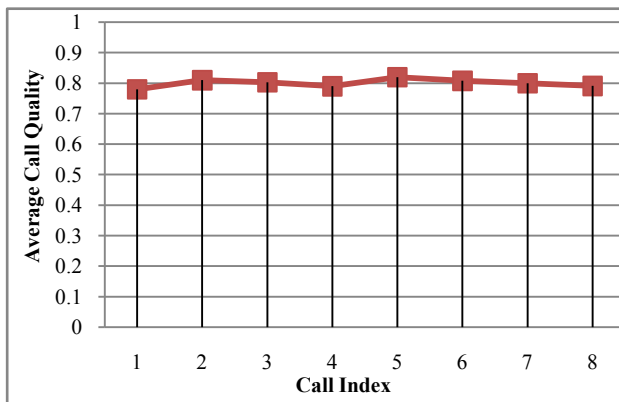


Figure 4. Average quality for calls with reduced echo component

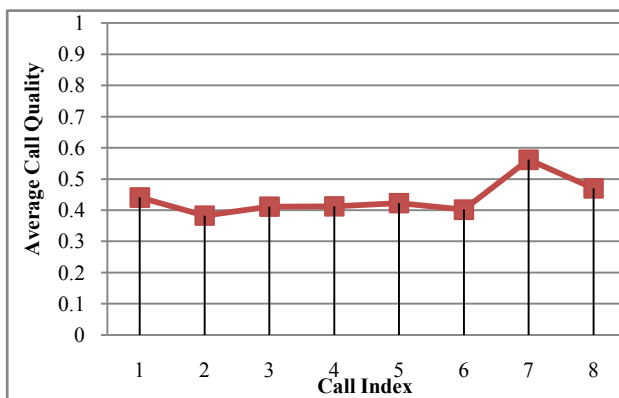


Figure 5. Average quality for calls with increased echo component

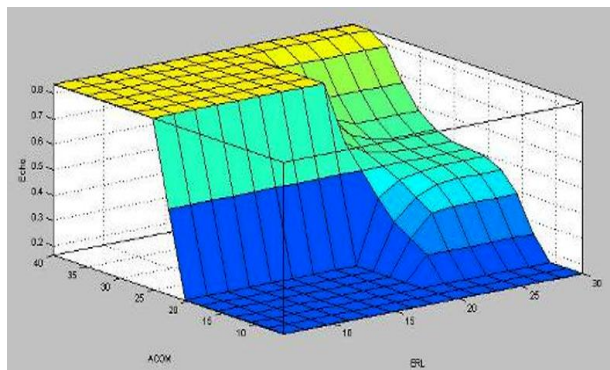


Figure 6. Matlab surface view for Echo Content, ERL and ACOM

Figure 4 shows the estimation of echo content of the first group of recorded calls, which are known to contain little or no echo. Note that the overall call quality is averagely good. This shows that our algorithm is able to track the echo content of the voice traffic. Again figure 5 shows that echo was not effectively cancelled or there was some high noise or

high difference level between the receive-path signal and the send-path signal. Note this is from the second group of recorded call which are known to have reasonable echo content. We see from this figure that the algorithm was able to track and detect several calls as having bad echo quality. However, some calls (like call number 7) obtained a much higher score than expected. This could be attributed to averaging error since average of the sample was used to obtain the final out. For instance, if we get an average score for the voice quality in a given call, we do not know if the call had a good quality for approximately half of the time and a bad quality in the remaining time or if the call quality was average during the whole call. This can be resolved by increasing the frequency of result samples taken for each call. Figure 6 shows a 3-D view of the echo content, echo and the combined loss and their relationship. One can conclude from these results that the proposed algorithm correctly tracks and estimates the echo content of a voice traffic in a VoIP system. We should note that due to our choice of defuzzification method and the output membership functions, even for calls with perfect quality the output wouldn't be 1.0. So the output shown in Figure 4 really reflects very good voice quality as we expected.

5. Conclusions

This work presented an algorithm which is a building block of an objective, passive, voice quality algorithm that can run in real-time and estimates the voice quality for live calls in a VoIP systems. The use of fuzzy logic was motivated by its ability to give reasonably good result with low computational complexity. The proposed algorithm can run and give results in real-time in the embedded system that processes the VoIP calls, thereby giving operators an almost instantaneous estimation of the quality of their network with respect to echo without the need for a reference signal. On the other hand, the algorithm has a tradeoff between precision and complexity, since it is not as precise in its estimates as the PESQ, MOS or the E-model methods but it is much less complex. Further work is needed to improve on the precision of the estimates while retaining the real-time feature.

REFERENCES

- [1] Manjunath, T. (2009) Limitations of perceptual evaluation of speech quality on VoIP systems, Proceedings of the Broadband Multimedia Systems and Broadcasting, 2009. BMSB '09. IEEE International Symposium on communications, vol. 13-15 pp.1-6.
- [2] Ren, Jiuchun; Zhang, ChongMing; Huang, WeiChao; Mao, Dilin, (2010) Enhancement to E-Model on standard deviation of packet delay, Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference, vol. 23-25, pp.256-259.

- [3] Periakarruppan, G.; Low, A.L.Y.; Azhar, H.; Rashid, A., (2006) Packet Based Echo Cancellation for Voice Over Internet Protocol Simulated with Variable Amount of Network Delay Time, TENCON 2006. 2006 IEEE Region 10 Conference, pp 1 – 4.
- [4] Adaptive Digital Technologies, (2010) The Echo "Phenomenon", Causes and Solutions, http://www.adaptivedigital.com/product/echo_cancel/echo_explain.htm.
- [5] Falk, T.H. and Wai-Yip Chan. (2008) Hybrid Signal-and-Link-Parametric Speech Quality Measurement for VoIP Communications, Audio, Speech, and Language Processing, vol.16, no.8, pp.1579-1589.
- [6] Lakaniemi, A.; Rosti, J.; Raisanen, V.I., (2001) Subjective VoIP speech quality evaluation based on network measurements, Communications, 2001. Proceedings of the ICC 2001. IEEE International Conference, vol.3, pp.748-752.
- [7] Ngamwongwattana, B. and Thompson, R. (2010) Sync & Sense: VoIP Measurement Methodology for Assessing One-Way Delay Without Clock Synchronization- Instrumentation and Measurement, vol. 59, pp 1318 – 1326.
- [8] Harjit Pal Singh Dr.Sarabjeet Singh Dr.Jasvir Singh, (2010) Computer Modelling & Performance Analysis of VoIP under Different Strategic Conditions, Computer Engineering and Applications (ICCEA), 2010 Second International Conference, vol.1, pp 611 – 615.
- [9] Justin Yackoski and Chien-Chung Shen, (2010) Managing End-to-End Delay for VoIP Calls in Multi-Hop Wireless Mesh Networks, IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings.
- [10] Lingfen Sun and Emmanuel C. Ifeakor, (2006) Perceived Speech Quality Prediction for Voice over IP-based Networks, Department of Communication and Electronic Engineering, University of Plymouth, Plymouth PL4 8AA, U.K.
- [11] Paglierani, P. and Petri, D. (2009) Uncertainty Evaluation of Objective Speech Quality Measurement in VoIP Systems, Instrumentation and Measurement, vol.58, no.1, pp.46-51.
- [12] Ross, T. (2004) Fuzzy Logic with Engineering Applications, Wiley, 2004.
- [13] Ditech Networks. Echo Basics Tutorial including echo cancellers and echo's effect on QoS" (2010) <http://www.ditchnetworks.com/learningCenter/echoBasics.html>.