International Journal of Applied Information Systems (IJAIS) ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 11 - No. 3, August 2016 - www.ijais.org

Predicting VoLTE Quality using Random Neural Network

Duy-Huy Nguyen
SAMOVAR, Télécom SudParis, CNRS,
Université Paris-Saclay
9 rue Charles Fourier - 91011 Evry Cedex

Hang Nguyen
SAMOVAR, Télécom SudParis, CNRS,
Université Paris-Saclay
9 rue Charles Fourier - 91011 Evry Cedex

Éric Renault
SAMOVAR, Télécom SudParis, CNRS,
Université Paris-Saclay
9 rue Charles Fourier - 91011 Evry Cedex

ABSTRACT

Long Term Evolution (LTE) was initially designed for a high data rates network. However, voice service is always a main service that drives huge profits benefit for mobile phone operators. Hence the deployment of Voice over LTE (VoLTE) is very essential. LTE network is a fully All-IP network, thus, the deployment of VoLTE is quite complex, specially for guaranteeing of Quality of Service (QoS) for meeting quality of experience of mobile users. The key purpose of this paper is to present an object, non-intrusive prediction model for VoLTE quality based on Random Neural Network (RNN). In order to simulate an experiment, a three-layer feed-forward RNN architecture with gradient descent training algorithm is applied. The inputs of this model are object network impairments such as Packet Loss Rate (PLR), Delay and Jitter. The VoLTE quality was predicted in term of the Mean Opinion Score (MOS). The simulation results show that this model offers MOS values which are quite close to well-known method is WB-PESQ (Wideband Perceptual Evaluation of Speech Quality) model. The results also show that the proposed model is very suitable for predicting voice quality over LTE network.

Keywords

Voice quality, VoLTE, MOS, WB-PESQ, RNN

1. INTRODUCTION

LTE network is developed by the Third Generation Partnership Project (3GPP) [1]. It is a mobile network which has high data rate, low delay and is fully packet-based. This network improves upon some of the capabilities of the legacy system by increasing data rates and extending QoS for various multimedia applications. Voice over LTE network (called VoLTE) is a main service of LTE network. Since LTE network is a full packet-switched, thus, the deployment of VoLTE service is very complicated. All voice traffics over LTE network are VoIP (Voice over IP) [6]. Also according to [6], there are two types of voice traffic over LTE network, those are VoLTE and VoIP. VoLTE is really a VoIP with QoS guaranteed [6]. However, in order to guarantee VoLTE quality is an extreme challenge.

In communications systems, the perceived voice quality is usually represented as the MOS. MOS can be attained by many methods. These methods are divided into two groups called subjective methods and objective methods. In subjective method, a group of human subjects listen to a live audio stream or a previously recorded audio file generated through the technology in question and rate the quality on a scale of 1 (poor) to 5 (excellent) [8].

These methods have some disadvantages such as too expensive, time consuming and are not suitable for a large network infrastructure. Otherwise, objective methods have more advantages, they eliminate the limitations of subjective methods. Objective methods are classified into two approaches: intrusive and non-intrusive ones. The intrusive methods (e.g. WB-PESQ [9]) are more exact and are widely utilized to predict aware voice quality. However they are not suitable for real-time services such as VoIP because they require original signals to reference. The non-intrusive methods (e.g. ITU-T Wideband E-model [10]) are computational models that are used for transmission planning. They are not accurate as the intrusive approaches and they have complex mathematical operations. The obtained results from objective methods do not always well relate to human perception. The main advantage of the non-intrusive methods are they predict voice quality without any reference to the original signals and they require less parameters than the intrusive methods. RNN models are used to predict quality multimedia flows such as voice or video over an IP network. They are non-intrusive methods and have many advantages compared to WB-PESQ model, specially they are very suitable for real-time services such as Voice or Video calls. However, they may be not as exact as the WB-PESQ model which involves nonlinear and time consuming operations. According to [2], RNN models are more advantageous than some other well-known models such as PESQ, E-model or INTERMON. RNN models are applied in many different fields where there are communication systems such as: (1) controlling the routing of packets in the Cognitive Packet Network (CPN), and (2) for automatic quantification of the quality of traffic associated with real-time multimedia applications such as video, speech and interactive voice [17].

According to our knowledge, there are no papers that predict VoLTE quality using RNN. Authors in [14], [15] and [13] proposed utilization of RNN to predict voice quality over Internet network. These models are not suitable for VoLTE service because they are used to predict voice quality for narrow band audio. In [5], authors proposed using RNN to predict video quality over LTE network.

The remainder of this paper is organized as follows: Overview of the system model is described in section 2. In section 3, we present the proposed simulation model. The simulation results and performance evaluation of the proposed model are analysed in section 4. The conclusion and future work are represented in section 5.

2. THE SYSTEM MODEL

2.1 The VoLTE traffic flow

2.1.1 The VoLTE architecture. When a VoLTE packet transmitted over LTE network, it is encapsulated sequentially with network protocols. For the downlink direction, the VoLTE packet uses Real-time Transport Protocol (RTP), User Datagram Protocol (UDP) and Internet Protocol (IP) at Application layer. It is then packeted with radio protocols such as Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC) and Medium Access Control (MAC). The VoLTE service utilizes both IPv4 and IPv6 protocols. The IPv4 header size is 40 bytes (IPv6 is 60 bytes), thus, in order to reduce the overhead, Robust Header Compression (RoHC) is deployed. The IP header is then compressed by RoHC down to only 1-3 bytes.

2.1.2 The VoLTE Codec. For the source encoder, VoLTE uses Adaptive Multi-Rate Wideband (AMR-WB). This vocoder utilizes a sampling rate of 16 kHz which covers a bandwidth in range of 50-7000 Hz. It also has the better voice quality when compared to Adaptive Multi-Rate Narrowband (AMR-NB). AMR-WB has nine codec modes where its bit rates are in range of 6.6-23.85 kbps. Since the audio bandwidth of AMR-WB is wider than AMR-NB, thus, it may offers clearly more natural sound and better voice quality in comparison with AMR-NB one.

2.2 The Random Neural Network model

2.2.1 The general model. RNN model was introduced by Gelenbe [4] at the end of eighties of the last century. It has been efficiently applied in many fields and applications such as pattern recognition, classification, image processing, combinatorial optimization and communication systems [17]. RNN is a mathematical model that combines Artificial Neural Network (ANN) and queuing model. So that, it is very suitable for predicting voice quality. According to [2], [5], RNN model has several advantages as follows:

—The results obtained well with human awareness;
—It doesn't request complex operations;
—It can be utilized in real-time applications;
—It can be easily extended when need to add input parameters;
—It has standard learning algorithm which has low complexity and strong generalization possibility;
—It can be implemented easily in both software and hardware because its neurons can be represented by simple counters.

RNN model consists of a set of interconnected neurons as shown on Figure 1. These neurons are arranged in different layers. They exchange signals from one neuron to another as well as the environment. The signals are transmitted immediately between neurons also with the environment. In RNN, each neuron is called a spike and is represented by an integer, specifically +1 for excitation spikes and -1 for inhibition spikes. The spikes can be departed from another neuron within the network or from outside. Positive neurons are permitted to send out spikes to one of two neuron types in the network. When a neuron sends out a spike, it loses one unit that is going from state q_i to state $q_i - 1$. Assume $p^+_{i,j}$, $p^-_{i,j}$ are the probabilities that the spike signal sent out by neuron i to neuron j as a positive one and as a negative one, respectively, d_i represents the signal leaving the network to the environment. If N is the total number of neurons in the network, for all $i = 1..N$, the following condition must be met [17]:

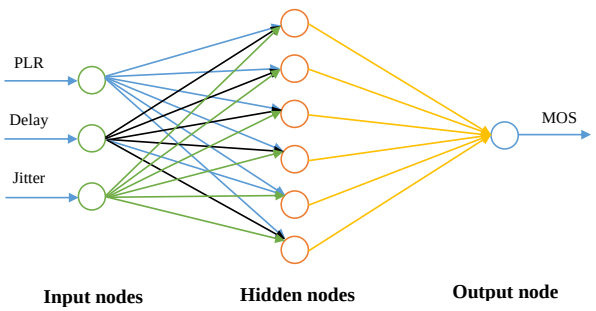

Fig. 1. Three layers Feed-forward architecture of RNN

$$d_i + \sum_{j=1}^{N} p^+_{i,j} + p^-_{i,j} = 1 \tag{1}$$

With: $p_{i,j} = p^+_{i,j} + p^-_{i,j}$
When signals leaving a neuron are not allowed to return directly back to the same neuron, it means $p_{i,j} = 0$. When a neuron receives a positive signal, its potential is increased by 1. It is reduced by 1 when it receives a negative one if it is strictly positive, and it doesn't change if its value is 0. Otherwise, when a neuron sends a positive or negative signal, its potential reduces by 1. It is strictly positive because only excited neurons sending signal. The neuron i receives signals from outside according to a Poisson process with rate λ^+_i for positive signal and λ^-_i for negative one. Excited neurons are denoted as $w^+_{i,j} = r_i \times p^+_{i,j}$ and $w^-_{i,j} = r_i \times p^-_{i,j}$ in the weight w is explained similarly to the one in ANNs. Nevertheless, they represent specifically the excited and inhibitive spike rates and they are not negative that means $w^+_{i,j} \geq 0$ and $w^-_{i,j} \geq 0$. Each neuron can fire only when its potential is strictly positive. Assume r_i is a Poison firing rate with exponential distributed interim pulse intervals. It is calculated as follows [5]:

$$r_i = \sum_j (w^+_{i,j} + w^-_{i,j}) \tag{2}$$

2.2.2 The network state and steady-state probability. The probability distribution of the network state and steady-state probability that neuron *i* excited is described as the following: At time t, the network state is denoted by vector of $k(t) = k_1(t), ..., k_N(t)$, the stationary probability distribution $p(k)$ is presented as follows [16]:

$$p(k) = \lim_{t \to \infty} prob[k(t) = k] \tag{3}$$

The quantity q_i is given by:

$$q_i = \frac{\lambda^+_i}{r_i + \lambda^-_i} \tag{4}$$

Where: λ^+_i, λ^-_i is described previously and total arrival rates $i = 1..N$ meet the system of non-linear concurrent formulas below:

$$\lambda^+_i = \wedge_i + \sum_{j=1}^{N} q_i r_i p^+_{ji} \tag{5}$$

$$\lambda^-_i = \wedge_i + \sum_{j=1}^{N} q_i r_i p^-_{ji} \tag{6}$$

In which: q_i must meet the following condition:

$$0 \leq q_i = \frac{\wedge_i + \sum_{j=1}^{N} q_i r_i p^+_{ji}}{r_i + \wedge_i + \sum_{j=1}^{N} q_i r_i p^-_{ji}} \leq 1 \tag{7}$$

With $\wedge_i$ are rates of the exogenous excitation. If there is a non-negative solution which exists λ^+_i, λ^-_i for the equations of (5) - (7), such that $q_i \leq 1$, then:

$$p(k) = \prod_{i=1}^{N} (1 - q_i) q_i^{k_i} \tag{8}$$

The network is stable which ensures the excitation grade of each neuron remains finite with probability 1. The average potential at a neuron *i* can be computed as $A_i = \frac{q_i}{1-q_i}$. If $q_i > 0$, this means neuron *i* is saturated or unstable and the neuron *i* is continuously firing in the steady-state.

3. THE PROPOSED SIMULATION MODEL

3.1 Simulation model

In order to create the database of the degraded audio signals, we have to build a simulation model for live voice streaming. This model should be suitable for VoLTE flow. The proposed simulation scenario is represented on Figure 2. In this diagram,

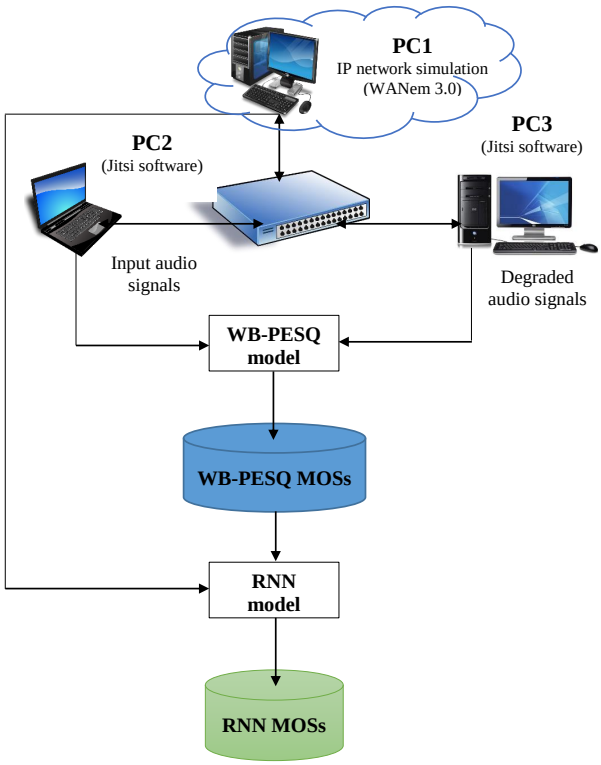

Fig. 2. The simulation scenario of the experiment

in order to make a voice call from PC2 to PC3, we propose to use Jitsi software. Jitsi is an audio/video Internet phone and instant messenger. It supports some of the most popular instant messaging and telephony protocols such as SIP, Jabber/XMPP [12]. It was formed from SIP communicator project and is now an open source multiplatform voice (VoIP), videoconferencing and instant messaging application [19]. Jitsi supports many different types of voice codec where there is AMR-WB which is mandatory for VoLTE application. In this experiment, we have configured Jitsi settings in such a way that it uses SIP protocol for signaling, AMR-WB codec for source codec and UDP protocol for delivering voice packet over an IP network for both PC2 and PC3. These characteristics are very suitable for VoLTE flow. When a voice packet is transmitted over LTE network (or IP network), its quality is affected by many factors, such as: Speech codec, latency, PLR, Jitter, Packet size, Silence suppression, Echo, Network parameters, Bandwidth, Equipment, Router, Phone frequencies, Weather conditions, Location of the hardware, and etc. Many of these are not supported by most service providers. Therefore, they can only improve voice quality through the network impairments. The following three factors have a direct and strong effect on VoLTE quality: packet loss rate (PLR), delay and jitter. These factors are features of the IP network. So that, when simulate VoLTE service, the simulation environment should be similar to real environment. There are many IP network simulation tools where some popular ones are Netdisturb [3], WANem [18], PacketExpert, IPLinkSim, PacketCheck [7], etc. All of these softwares allow adding network impairments to VoIP stream. One of the most featured softwares is WANem (Wide Area Network emulator) which provides a real simulation environment in Internet/WAN/LAN networks. It can be used to simulate characteristics of WAN network for live stream traffics. This means WANem allows applications such as voice, video or data, etc. to be tested in a realistic conditions of WAN network with an affordable cost [18]. It is an open source software and is widely used in research and academic communities. Therefore, in this study, we use WANem for simulating IP network. It is very suitable for the proposed simulation scenario. In fact, when a VoLTE packet transmitted from a sender to a receiver, it has to pass all layer protocols consisting of PDCP, RLC, MAC, and then Physical Layer (PHY) before it is delivered over the air interface. However, according to several authors, tasks of these layer protocols such as header addition, header compression, etc. which affect not significantly to voice quality. Hence, in this study, we do not mention the effects of them to VoLTE flow. The simulation process is described as follows:
The input audio signal is a prerecorded audio file. This is a .wav file that was taken from [11]. The audio file is in English and has a duration of 15 seconds. Firstly, we setup Jitsi software using SIP protocol, AMR-WB codec and UDP protocol for both PC2 and PC3. Secondly, we configure the voice flows that have to pass WANem software (run on PC1). We also configure parameters for WANem according to test conditions. Thirdly, from PC2, we make a voice call to PC3. The voice stream is transmitted from PC2 to PC3 through WANem. We have 64 test conditions thus we have to simulate 64 times. Then at the PC3, we recorded the received voice flows. These are the degraded audio signals after going through an IP network. Lastly, in order to measure voice quality, WB-PESQ [9] model is deployed resulting in MOS values (called WB-PESQ MOS values). And then RNN model is applied for predicting the new values of MOS. The inputs of RNN model are the network impairments such as PLR, delay, jitter and WB-PESQ MOS values. After two periods of training and testing will generate new MOS values (called RNN MOS values). The detailed description is represented in the following subsection.

3.2 Simulation parameters

In the simulation model as shown on Figure 2, the IP network simulation (WANem) plays a very important role. When a voice packet is transmitted over an IP network, it is affected by may factors where there are three very important factors such as PLR, delay and jitter. So that, during the simulation, these factors have to be changed in limited range. In this experiment, the threshold range for PLR, delay and jitter are 0-10 %, 5-100 ms and 0-15 ms, respectively.
PLR is percentage of packets which have not reached their destination. They could have been dropped over the network. Some packets not delivered to the destination will degrade the voice quality. There are many reasons that cause PLR such as link failure, congestion, buffer overflow or wrong route. The undelivered packets can be restored by Packet Loss Concealment (PLC) at the receiver. For the VoIP, PLR characteristics are bursty. In this study, we select four different values of PLR including 0, 4, 7 and 10 %.
Delay is the average time that a voice packet is transmitted from the source to destination. For the VoLTE, 3GPP recommended that this value should not exceed 100 milliseconds (ms). According to ITU-T recommendation, this value doesn't exceed 400 ms

for general networks. In order to compute one-way delay of voice packet, ITU-T Recommendation provides the E-model for narrow audio band and Wideband E-model for wide audio band. For the real-time services such as VoIP or Video calls, the end to end delay shouldn't exceed 150 ms for a better quality. In this experiment, we choose 4 different values for delay consisting of 5, 30, 60 and 100 ms. We do not setup delay equal to 0 ms because in this case, we can not configure different values for jitter. It means when delay equals 0 ms, value of jitter is always equal to 0.

Jitter is considered as variation of network delay. It is the variability of packet arrival time at the receiver. Jitter caused by different packets of the same conversation have different queue lengths or different routes when they are transmitted over the network. In order to avoid the reduction of voice quality, jitter has to be removed before replaying at the receiver. In order to reduce jitter, a de-jitter is used at the end user to delay the initiation of the replay process. In this study, we pick out four various values of jitter including 0, 5, 10 and 15 ms.

Based on triples of PLR, delay, and jitter, we have 64 values (test conditions). Each times make a voice call from PC2 to PC3, we change triple (PLR, delay, jitter). The voice sample is a prerecorded audio file. This file transmitted as a voice stream over IP simulation tool (WANem) that is affected by network impairments. At the PC3, the conversation is recorded into an audio file. This simulation process is repeated for each of 64 test conditions.

When predicting voice quality using RNN model, the WB-PESQ MOS values are used to validate the results of RNN model. In the RNN model, we use the architecture of three layers with the gradient descent algorithm. The number of nodes of hidden layer is 6. The required stop Mean Square Error (MSE) is used equals 2×10^{-5} and threshold for number of iteration is 1000. The proposed model is implemented through two periods including training and testing ones. The database of WB-PESQ MOS values is divided into two parts: first one for training phase and the remaining one for testing phase. After completing the prediction, we will have voice quality values called RNN MOS values.

4. PERFORMANCE EVALUATION

In order to assess the performance of the proposed model, we consider the effects of PLR, delay and jitter on perceived voice quality. We compare results obtained from two models of WB-PESQ and RNN. Figure 3 shows results between test conditions versus MOS values of WB-PESQ and RNN models. Such as

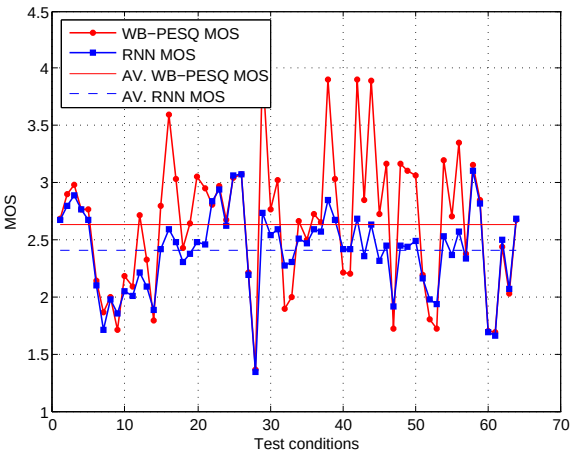

Fig. 3. Effects of test conditions on voice quality

shown on Figure 3, the MOS values of WB-PESQ model are actual MOS ones and they change according to a wide amplitude. For the MOS values predicted by RNN model modify according to a less wide amplitude. In order to assess the accuracy of WB-PESQ and RNN models through all test conditions, we compute the average MOS of each method. This value of WB-PESQ method is 2.63 while of RNN one is 2.40. It is clear that the accurateness of WB-PESQ model higher than that of RNN model isn't significant (a reduction of about 8.7%).

Figures 4, 5 and 6 show the effects of PLR, delay and jitter on voice quality, respectively. In order to represent these charts, we calculate the average MOS values of each method when the PLR, delay and jitter are fixed, respectively. Through these charts, it can be said that the effects of PLR, delay and jitter on voice quality is not much when they are in allowable limitations. In most cases, the voice quality reduces when PLR, delay and/or jitter augment. It can be conclusive that when guaranteeing PLR, delay and jitter in allowable limitations, the effects of them to voice quality is not significant.

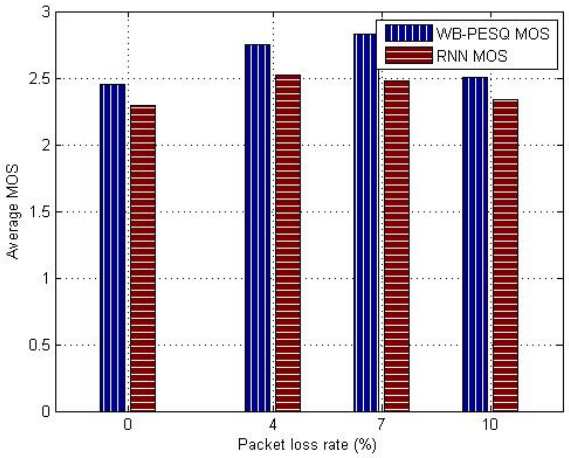

Fig. 4. Effects of PLR on voice quality

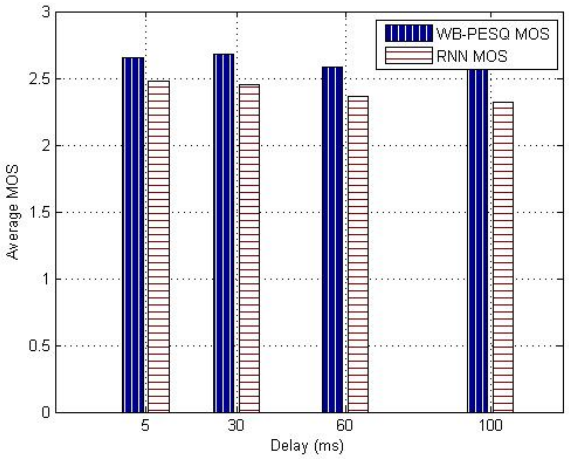

Fig. 5. Effects of delay on voice quality

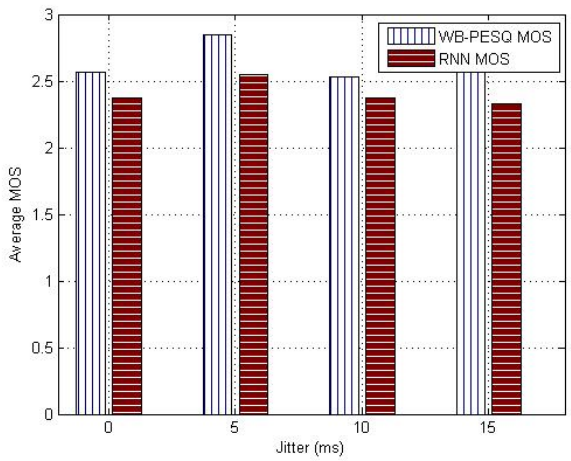

Fig. 6. Effects of jitter on voice quality

5. CONCLUSION

This paper proposes a new model for predicting voice quality over LTE network using RNN model with the architecture of three layers and the gradient descent algorithm. In order to evaluate the effects of network impairments on voice quality, three main factors are used including PLR, delay and jitter. WB-PESQ model is used to measure voice quality. The MOS values obtained from this model are used to validate RNN model results. Through obtained results from two methods, it is clear that the results attained from the proposed RNN model are quite accurate and are very close to WB-PESQ model ones. This means if network impairments such as PLR, delay and jitter are in thresholds then the effects of them on voice quality are not remarkable. The proposed model is suitable and can be applied to wideband audio flows such as VoLTE. It can be located at the receiver to measure immediately voice quality. In addition, the proposed model also can be used to predict video quality over LTE network. For the future work, we will complement more factors such as network impairments, languages, etc. to the proposed model.

6. REFERENCES

[1] 3GPP. http://www.3gpp.org.

[2] Charlie C Chen. An overview of Quality of service measurement and optimization for voice over Internet. 8(1):2053–2063, 2015.

[3] ZTI Communications. http://www.zti-communications.com/netdisturb/.

[4] Erol Gelenbe. Random neural networks with negative and positive signals and product form solution. *Neural computation*, 1(4):502–510, 1989.

[5] Tarik Ghalut and Hadi Larijani. Non-intrusive method for video quality prediction over lte using random neural networks (rnn). In *Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2014 9th International Symposium on*, pages 519–524. IEEE, 2014.

[6] Jonghwan Hyun, Jian Li, ChaeTae Im, Jae-Hyoung Yoo, and James Won-Ki Hong. A volte traffic classification method in lte network. In *Network Operations and Management Symposium (APNOMS), 2014 16th Asia-Pacific*, pages 1–6. IEEE, 2014.

[7] GL Communications Inc. http://www.gl.com/products.html.

[8] ITU-T. http://www.itu.int/rec/T-REC-P.800/en.

[9] ITU-T. https://www.itu.int/rec/T-REC-P.862.2/en.

[10] ITU-T. https://www.itu.int/rec/T-REC-G.107.1/en.

[11] ITU-T. http://www.itu.int/net/itu-t/sigdb/genaudio/AudioForm-g.aspx?val=1000050.

[12] Jitsi. https://jitsi.org.

[13] Kailash Chandra Mishra and Padma Charan Das. Measuring quality of service of voip based on artificial neural network approach. *International Journal*, 5(3), 2015.

[14] Samir Mohamed, Gerardo Rubino, and Martın Varela. Performance evaluation of real-time speech through a packet network: a random neural networks-based approach. *Performance evaluation*, 57(2):141–161, 2004.

[15] Kapilan Radhakrishnan and Hadi Larijani. A study on qos of voip networks: a random neural network (rnn) approach. In *Proceedings of the 2010 Spring Simulation Multiconference*, page 114. Society for Computer Simulation International, 2010.

[16] Kapilan Radhakrishnan and Hadi Larijani. Evaluating perceived voice quality on packet networks using different random neural network architectures. *Performance Evaluation*, 68(4):347–360, 2011.

[17] Stelios Timotheou. The random neural network: a survey. *The computer journal*, 53(3):251–267, 2010.

[18] WANem. http://wanem.sourceforge.net.

[19] Wikipedia. Jitsi — wikipedia, the free encyclopedia, 2015. [Online; accessed 21-October-2015].