

Real-Time, Non-intrusive Evaluation of VoIP

Adil Raja¹, R. Muhammad Atif Azad², Colin Flanagan¹, and Conor Ryan²

¹ Wireless Access Research Center

Department of Electronic and Computer Engineering

² Biocomputing and Developmental Systems Group

Department of Computer Science and Information Systems

University of Limerick, Limerick, Ireland

{adil.raja,atif.azad,colin.flanagan,conor.ryan}@ul.ie

Abstract. Speech quality, as perceived by the users of Voice over Internet Protocol (VoIP) telephony, is critically important to the uptake of this service. VoIP quality can be degraded by network layer problems (delay, jitter, packet loss). This paper presents a method for real-time, non-intrusive speech quality estimation for VoIP that emulates the *subjective* listening quality measures based on *Mean Opinion Scores* (MOS). MOS provide the numerical indication of perceived quality of speech. We employ a Genetic Programming based symbolic regression approach to derive a speech quality estimation model. Our results compare favorably with the International Telecommunications Union-Telecommunication Standardization (ITU-T) PESQ algorithm which is the most widely accepted standard for speech quality estimation. Moreover, our model is suitable for real-time speech quality estimation of VoIP while PESQ is not. The performance of the proposed model was also compared to the new ITU-T recommendation P.563 for non-intrusive speech quality estimation and an improved performance was observed.

Keywords: VoIP, Non-Intrusive, Speech Quality, GP, Symbolic Regression, MOS.

1 Introduction

Speech quality estimation for VoIP can be performed either *subjectively* or *objectively*. In the former case, speech quality is estimated by averaging the opinions of a set of suitably trained human subjects [1]. Each of the testers assigns an *Opinion Score* – on an integral scale from 1 (unacceptable) to 5 (excellent) – to the speech signal under test. The opinion scores of the testers are averaged into a MOS. Subjective MOS has been found to be, by far, the most reliable technique of speech quality estimation. However, it is expensive, time-consuming and laborious.

Recently, objective speech quality assessment has become a very active research area. This is an attempt to circumvent the limitations of subjective testing by simulating the opinions of human testers algorithmically. There are two distinct approaches to objective testing: intrusive and non-intrusive.

Intrusive speech quality estimation techniques compare the test (i.e., network distorted) speech signal, as reconstructed by the decoder, to the reference, input

speech, basing their estimation on the measured amount of distortion. ITU-T P.862 (PESQ) [2] is a popular example of intrusive estimation model.

On the other hand non-intrusive schemes assess the quality of the distorted signal in the absence of the reference signal. This approach is effective in environments where the reference speech signal is not accessible. P.563 is the new ITU-T Recommendation for non-intrusive evaluation speech quality in narrow-band telephony applications [3]. Intrusive models are more reliable than the non-intrusive ones as the former have access to a reference speech signal to compare the distorted speech signal with.

However, the afore-mentioned models are compute-intensive as they base their results on the time and/or frequency domain analysis of the speech signal under test. They also require the test call to be recorded for a considerable duration before it can be analysed. Hence, they are not suitable for real-time and continuous monitoring of speech quality. This makes non-intrusive models like ITU-T Recommendation G.107 (E-Model) [4] more attractive for real-time speech quality estimation as they base their results on networks traffic parameters. Despite the fact that E-model is a *transmission planning* tool, it has been deployed in various commercial applications. First of all a different version of it exists for various network conditions such as codec type and bursty or non-bursty network conditions. Moreover, it is restricted to a limited number of codecs and network conditions due to its reliance on subjective tests [5].

In this paper we have employed Genetic Programming (GP) based symbolic regression approach to estimate the speech quality as a function of impairments due to IP network and encoding algorithms. A main advantage of GP is that it can produce human-readable results in the form of analytical expressions. Moreover, GP deals with the significant input parameters and aids in the automatic pruning of the irrelevant ones. These features of GP make our results superior to the past research based on Artificial Neural Networks (ANNs) by Sun and Ifeakor [6], Mohamed et. al. [7] [8] and on lookup tables by Hoene et. al. [9]. We have used PESQ as a reference for evolutionary modeling. The results of proposed models show a high correlation with PESQ. Moreover, our models are suitable for real-time and non-intrusive estimation of VoIP quality.

The rest of the paper is organized as follows: section 2 talks about the VoIP architecture briefly. To gather the relevant data characterising the speech traffic we have employed a VoIP simulation as described in section 3. Section 4 elucidates how this data is used to evolve the speech quality estimation models. Section 5 presents the results and carries out an analysis of the current research. The paper concludes in section 6 outlining the major achievements and future ambitions.

2 VoIP

As opposed to traditional circuit switched telephony (PSTN), in VoIP the routing of voice conversations takes place over the Internet or an IP based network in the form of packets. The issues related to VoIP communication are governed by

various signaling and transport protocols. Once digitized, human speech is compressed by using a suitable encoding algorithm such as G.711, G.723.1, G.729 and AMR etc. Various speech codecs (encoder/decoder) differ from each other in terms of features such as encoding bit-rate (kbps), algorithmic delay (ms), complexity and, subsequently, speech quality (MOS). After compression and encoding into a suitable format the speech frames are packetized. RTP, UDP and IP packet headers are appended to the frames and the packets are sent to the receiver. During transmission some packets may be lost due to congestion and/or (wireless) transmission errors. The receiver processes the packets and presents them to the playout (depacketizing) buffer which is a temporary storage that aims to accumulate enough packets so that they can be played out to the listener as a steady stream as opposed to fragmented clips of voice. The playout buffer seeks to smooth out the variation of the inter-arrival delay (jitter) between the successive voice packets. If packets arrive too late to be played out on time, they are regarded as lost. Consequently, the losses as observed by the application are a superposition of losses due to late arrivals on the losses that occur elsewhere in the VoIP network. After the playout buffer the speech frames are decoded and in doing so any lost frames may be camouflaged by the decoder using a packet loss concealment (PLC) algorithm. Finally the decoded signal is translated in to its acoustic representation. Fig. 1 shows the steps required for mouth-to-ear transportation of voice over an IP network. Silence suppression or discontinuous transmission (DTX) is also supported by VoIP whereby the periods of a conversation when the speaker is silent are not coded or transmitted. DTX is aimed at bandwidth saving. A voice activity detector (VAD) is used to implement DTX.

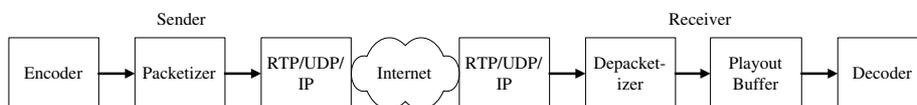


Fig. 1. VoIP system

3 VoIP Traffic Simulation

A simulation based approach was pursued for this research. Such an approach has been employed by various authors such as [10][11]. The main advantage of this approach is that various network distortion scenarios can be emulated precisely. Moreover, the tests are easily repeatable. This section describes the VoIP simulation methodology employed in this work. Before proceeding to the details of actual VoIP simulation environment it is pertinent to discuss the nature of VoIP packet loss. This is described in the following section along with a suitable packet loss model.

3.1 Packet Loss Model

VoIP packet loss is bursty in nature as it exhibits temporal dependency. In the current context the term burst is used to describe the event of a consecutive

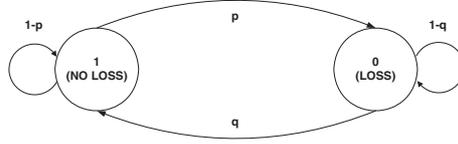


Fig. 2. The Gilbert Model

loss of a number of packets. So, if packet n is lost then normally there is a higher probability that packet $n + 1$ will also be lost. ITU-T Recommendation G.1050 [12], which presents a network model for evaluating multimedia transmission performance over IP, has proposed to use the Gilbert model to capture temporal dependency. Fig. 2 shows the state diagram of this 2-state Markov model.

In this stochastic automaton, p is the conditional probability that the packet numbered $n + 1$ is lost given that packet n is successfully received and q is the converse. $1 - q$ corresponds to the conditional loss probability (clp). Usually $p < 1 - q$. Moreover, the Gilbert model reduces to a Bernoulli model if $p + q = 1$. In (1) mlr corresponds to the mean loss rate and mbl corresponds to the mean burst length.

$$mlr = \frac{p}{p + q}, mbl = \frac{1}{q} \quad (1)$$

The values of p and q can be calculated using the loss length distribution statistics of a network traffic trace.

3.2 VoIP Simulation Environment

This section describes the network simulation environment and the testbed used in this study. A schematic of the simulation environment is shown in Fig. 3. The system includes a speech database, encoder(s)/decoder(s), a packet loss simulator, a speech quality estimation module (PESQ), a parameter extraction module for computing the values of different parameters and a GP based speech quality estimation model. Three popular codecs were chosen in the current research, namely;

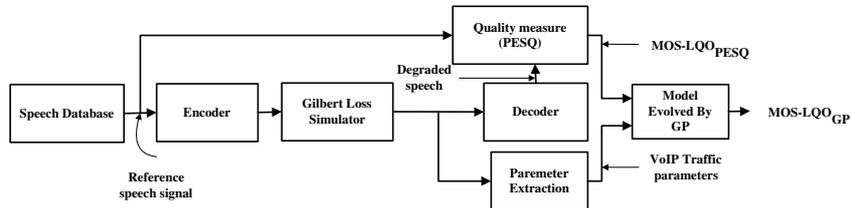


Fig. 3. Simulation system for speech quality estimation model

G.729 CS-ACELP (8 kbps) [13], AMR-NB [14] and G.723.1 MP-MLQ/ACELP (6.3/5.3 kbps) [15]. All of these are based on Linear Predictive Coding of speech. LPC is a scheme whereby the spectral envelop of human speech can be represented in a compressed form. AMR was used in its 7.4 and 12.2 kbps modes whereas G.723.1 was used in its 6.3 kbps mode only. All of these low-bitrate codecs aim at VoIP traffic bandwidth saving. These codecs also have built-in VAD and PLC mechanisms.

The choice of network simulation characteristics was driven by ITU-T Recommendation G.1050 [12] which describes a model for evaluating multimedia transmission performance over an IP network. Bursty packet loss was simulated using the Gilbert model (Fig. 2). It also models packets discarded by the playout buffer due to late arrivals. Packet loss was simulated for different values of mlr and clp . Twelve different values for mlr were chosen; 0, 2.5, 5, 7.5, 10, 12.5, 15, 20, 25, 30, 35 and 40%. The peak loss-rate (i.e. 40%) was kept an order of magnitude higher than that specified for an unmanaged network in ITU-T G.1050 (i.e. 20%) so as to gather more representative data for model derivation. For each value of mlr , clp was set to 10, 50, 60, 70, and 80%.

After subjecting the VoIP streams under test to various network impairments, they were evaluated using the PESQ algorithm. The PESQ algorithm compares a degraded speech signal with a reference (clean) speech signal and computes an objective MOS score ranging between -0.5 and 4.5, albeit for most cases the output will be between 1.0 and 4.5. It must be mentioned that PESQ simulates a listening test and is optimized to represent the average evaluation (MOS) of all listeners. It is statistically proven that the best possible result that can be obtained from a listening only test is never 5.0, hence it was set to 4.5¹. The PESQ algorithm is widely acclaimed for its high correlation with the results of formal subjective tests for a wide range of network distortion conditions. It is the current *de jure* standard for objective speech evaluation. In the current context, the MOS scores obtained by the PESQ algorithm and the MOS predicted by the GP based model are differentiated by the abbreviations $MOS-LQO_{PESQ}$ and $MOS-LQO_{GP}$ respectively. The term $MOS-LQO$ is an acronym for *Mean Opinion Score-Listening Quality Objective* and the various subscripts are used to identify the objective quality estimation models used. This terminology is based on [16].

Altogether, five VoIP traffic parameters have been chosen in the current analysis which form the input variables for evolutionary modeling. These parameters are: codec bit-rate (kbps), packetization interval (PI), frame duration (ms), mlr_{VAD} and mbl_{VAD} . A lower value of bit-rate corresponds to a higher compression of the speech signal, thus resulting in a lower bandwidth requirement at the expense of quality. The packetization interval, which specifies the acoustic information worth a certain duration of time to be contained in a VoIP packet, was varied between 10 to 60 ms. Considerable bandwidth saving can be achieved by encapsulating multiple speech frames in one VoIP packet thus reducing the need for RTP/UDP/IP headers that would have been required for encapsulation and

¹ <http://www.opticom.de/technology/pesq.html>

transportation of speech frames if they were to be sent individually. However, higher packetization intervals have certain associated drawbacks too. First, the end-to-end delay of the VoIP stream is increased as the sender has to buffer speech frames for a considerable duration before subsequent frames become available by the encoder. Second, for a large packetization interval, typically higher than 40 ms, loss of a single packet results in noticeable degradation of speech quality. Hence, packetization interval presents a trade-off between the speech quality and bandwidth saving. Frame duration has a similar effect on the quality as that of packetization interval. Higher frame durations may have other disadvantages, for instance, in LPC speech signal is assumed to be stationary (non-transient) for a given frame duration. However, for higher frame durations this assumption may considerably deviate from the reality. Thus, a codec with such a feature may obfuscate the final speech content. The parameter extraction module (Fig. 3) is used to obtain the values of the aforementioned parameters from the VoIP traffic stream under test. The corresponding $MOS-LQO_{PESQ}$ of the decoded VoIP stream under test subjected to these network conditions forms the target output value for training purposes. In actual VoIP applications this information would be gathered by parsing the RTP headers and bitstreams of the encoded frames. The information would then be used as an input for the GP based model to estimate $MOS-LQO_{GP}$ after processing.

Table 1. Common GP Parameters among all simulations

Parameter	Value
Initial Population Size	300
Initial Tree Depth	6
Selection	Lexicographic Parsimony Pressure Tournament
Tournament Size	2
Genetic Operators	Crossover and Subtree Mutation
Operators Probability Type	Adaptive
Initial Operator probabilities	0.5 each
Survival	Half Elitism
Generation Gap	1
Function Set	plus, minus, multiply, divide, sin, cos, \log_2 , \log_{10} , \log_e , sqrt, power,
Terminal Set	Random real-valued numbers between 0.0 and 1.0. Integers (2-10). $m_{lr_{VAD}}$, $m_{bl_{VAD}}$, PI , br , fd

4 Experimental Setup

As discussed earlier GP was the machine learning algorithm of choice for deriving a mapping between network traffic parameters and VoIP quality. GPLab was used as the preferred GP environment in this study. GPLab is a Matlab toolbox developed by Sara Silva ². A total of 4 GP simulations were conducted.

² <http://gplab.sourceforge.net/>

The common parameters of all the simulations are listed in Table 1. In all of the simulations the population size was set to 300. Each simulation was composed of 50 runs whereas each run spanned 50 generations. Adaptive genetic operator probabilities were used ³ [18]. Tournament selection with Lexicographic Parsimony Pressure (LPP) [19] was used in all of the simulations. Survival was based on elitism. The elitist criterion was such that half of the population of a new generation would be composed of best individuals from both parents and children. The other half of the population would be formed of remaining children on the basis of fitness. This elitism criteria is termed as *half elitism* in GPLab.

In simulation 1 mean squared error (mse) was used as the fitness function and tournament size was set to 2. For simulation 2 (and subsequent simulations) scaled mean squared error (MSE_s) was used as the fitness criterion and is given by equation (2).

$$MSE_s(y, t) = 1/n \sum_i^n (t_i - (a + by_i))^2 \quad (2)$$

where y is a function of the input parameters (a mathematical expression), y_i represents the value produced by a GP individual and t_i represents the target value which is produced by the PESQ algorithm. a and b adjust the slope and y-intercept of the evolved expression to minimize the squared error. They are computed according to equation (3).

$$a = \bar{t} - b\bar{y}, b = \frac{cov(t, y)}{var(y)} \quad (3)$$

where \bar{t} and \bar{y} represent the mean values of the corresponding entities whereas var and cov mean the variance and covariance respectively. This approach is known as *linear scaling* and is found to be very beneficial for the symbolic regression tasks with GP [20]. In simulation 2 (and subsequent simulations) *protected* functions were not used. Instead any inputs were admissible to all the functions. For the input values outside the domain of the functions *log*, *sqrt*, *division* and *pow*, NaN (undefined) values are generated. This results in the individual concerned being assigned the worst possible fitness.

The selection criterion in simulations 3 and 4 was based on the notion that population diversity can be enhanced if mating takes place between two, fitness-wise, dissimilar individuals, as suggested by Gustafson et. al. [21]. This selection scheme has been shown to perform better in the symbolic regression domain and, hence, it was employed in this research. This simple addition to the selection criterion only requires one to ensure that mating does not take place between individuals of equal fitness. In simulation 4 the maximum tree depth was changed from 17 to 7 to see if parsimonious individuals with performance comparable to those of earlier simulations can be obtained. Statistics pertaining to simulations and the results are presented in the next section.

³ Adaptive operator probabilities are discussed on page 31 of the GPLab manual.

5 Results and Analysis

Nortel Networks speech database containing high quality voice signals was used for analysis. The database contains 240 speech files corresponding to two male (m_1, m_2) and two female (f_1, f_2) speakers. Duration of speech signals in the files was between 10-12s. A total of 3360 speech files were prepared for various combinations of afore-mentioned values of network traffic parameters. The simulation parameters include frame duration, bit-rate, packetization interval, mlr and clp . 70% and 30% of the data of distorted speech files corresponding to speakers m_1 and f_1 were used for training and testing of the evolutionary models respectively. Distorted speech files corresponding to speakers m_2 and f_2 were used to validate the performance of the chosen model against speaker independent data. In other words network traffic parameters and corresponding $MOS-LQO_{PESQ}$ of 1177, 503 and 1680 speech files were used for training, testing and validation respectively.

Table 2(a) lists the statistics about the MSE of the training/testing data and of final tree size (in terms of number of nodes) of the 4 simulations under consideration. A Mann-Whitney-Wilcoxon test was also performed to decide if a significant difference exists between the simulations. Its results are tabulated in Table 2(b). At 5% significance level a '0' in the tableau indicates that no significant difference exists between the two simulations with respect to that *metric* (i.e. MSE_{tr} , MSE_{te} or *Size*). A '1' indicates the converse and an 'x' marks that the metric is not to be compared with itself.

A keen look at the tables 2(a) and 2(b) shows that simulation 2 (which used linear scaling) performed significantly better than simulation 1. When we compare it with simulation 3, we see that simulation 3 produces significantly smaller trees than simulation 2, albeit with marginally inferior fitness. Finally, simulation 4 exhibits similar traits, as its fitness is marginally worse again, although its trees are significantly smaller. The objective in the current research

Table 2. Statistical analysis of the GP simulations

(a) MSE Statistics for Best Individuals of 50 Runs for Simulations 1-4

Stats	Sim1			Sim2			Sim3			Sim4		
	MSE_{tr}	MSE_{te}	Size									
Mean	0.0980	0.1083	42.6	0.0414	0.0430	38.8	0.0434	0.2788	28.5	0.0436	0.0436	18.0
Std.												
Dev.	0.0409	0.0507	24.1	0.0040	0.0044	21.2	0.0042	1.0986	15.1	0.0037	0.0060	7.1
Max.	0.2135	0.2656	103	0.0543	0.0568	104	0.0519	6.8911	74	0.0520	0.0782	38
Min.	0.0449	0.0464	8	0.0368	0.0370	5	0.0378	0.0390	9	0.0370	0.0387	8

(b) Results of Mann-Whitney-Wilcoxon Significance Test

Stats	Sim1			Sim2			Sim3			Sim4		
	MSE_{tr}	MSE_{te}	Size									
Sim1	x	x	x	1	1	0	1	1	1	1	1	1
Sim2	1	1	0	x	x	x	1	0	1	1	1	1
Sim3	1	1	1	1	0	1	x	x	x	0	0	1
Sim4	1	1	1	1	1	1	0	0	1	x	x	x

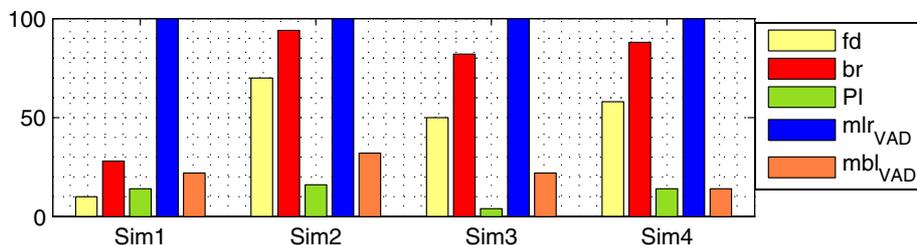


Fig. 4. Percentage of the best individuals employing various input parameters in the 50 runs of each of the four simulations

was to find fitter individuals with small sizes. Hence, simulation 4 was scavenged for plausible solutions.

Fig. 4 delineates the significance of various network traffic parameters in terms of the number of best individuals using them in each of the four GP simulations. It turns out that mlr_{VAD} had a 100% utility in all of the simulations. Codec bit-rate (br) and frame duration (fd) were the second and third most frequently availed parameters respectively. Whereas, both PI and mbl_{VAD} have shown advantage in least number of runs of all simulations.

Two of the models derived from this work are shown in this paper by equations (4) and (5). The MSE_s and Pearson's product moment correlation coefficients (σ) of equations (4) and (5) are compared with each other in Table 3. Equation (5) is a function of mlr_{VAD} solely. Whereas equation (4), which was the best model discovered, additionally has br and fd as independent variables. This was the best model of all the runs too. Fig. 5 shows the scatter plots of equation (4) for training and testing data. It is noticeable that equation (5) which is a function of mlr_{VAD} only, however, has comparable fitness to equation (4). Evaluating a single variable would be computationally cheap for a real time analysis. In the light of this and the earlier discussion on Fig. 4 mlr_{VAD} seems to be the most crucial parameter for VoIP quality estimation.

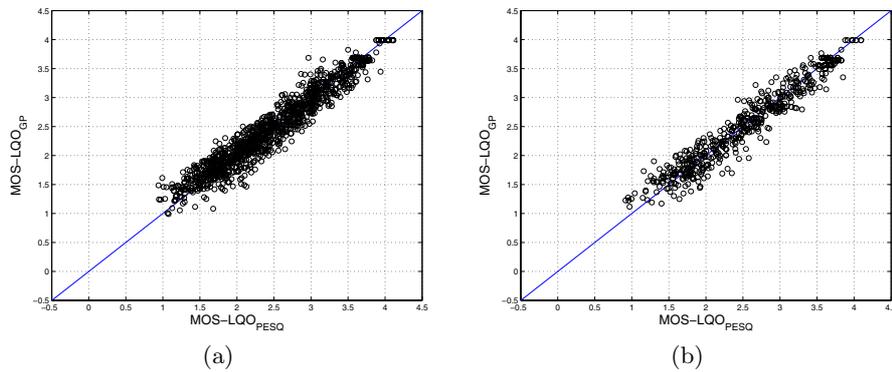
$$MOS - LQO_{GP} = -2.46 \times \log(\cos(\log(br)) + mlr_{VAD} \times (br + fd/10)) + 3.17 \quad (4)$$

$$MOS - LQO_{GP} = -2.99 \times \cos\left(0.91 \times \sqrt{\sin(mlr_{VAD})} + mlr_{VAD} + 8\right) + 4.20 \quad (5)$$

As stated earlier ITU-T P.563 is the new recommendation for non-intrusive speech quality estimation. A correlation analysis was done between $MOS-LQO_{PESQ}$ and the corresponding objective MOS values obtained by ITU-T P.563 ($MOS-LQO_{P.563}$). It turned out that the correlation coefficients (σ) varied between 0.65-0.82 under various network traffic conditions. This also highlights the significance of current research. It is reiterated to emphasize that ITU-T P.563 is a non-real-time process as it relies upon complex *digital signal processing* techniques to estimate the quality of the speech signal under test. The proposed models, on the other hand, are the functions of network traffic parameters that can be gathered efficiently by parsing VoIP packets.

Table 3. Performance Statistics of the Proposed Models

Data	Equation (4)		Equation(5)	
	MSE_s	σ	MSE_s	σ
Training	0.0370	0.9634	0.0520	0.9481
Testing	0.0387	0.9646	0.0541	0.9501
Validation	0.0382	0.9688	0.0541	0.9531

**Fig. 5.** MOS-LQO predicted by the proposed individual vs MOS-LQO measured by PESQ for (a) training data and (b) testing data for equation 4

6 Conclusions and Future Work

The problem of real-time quality estimation of VoIP is of significant interest. This paper has shown an approach for solving this problem by employing GP. One of the main objectives of this research was to estimate the effect of burstiness on speech quality. It turned out that burst length was least used by the best individuals of various runs. This is due to the fact that the PESQ algorithm does not model the effect of burstiness on speech quality [10] [11]. Hence, the effect of burstiness can be mapped only by conducting suitably designed formal subjective tests [22]. Despite this limitation, PESQ is the best objective quality estimation model and has been used to model the effect of packet loss by various studies. The proposed models are good approximations to PESQ and computationally more efficient. Hence, they are useful for real-time call quality evaluation. For the codecs considered in this study, we have also proposed a model (equation(5)) that is a function of mlr_{VAD} only with performance comparable to the other models. This is considerable since such a model can be deployed conveniently on a wide variety of platforms.

Our results are better than the past research both in terms of performance and nature of the proposed models. For instance, Sun and Ifeachor [6] and Mohamed et. al. [7] [8] proposed ANN based models for VoIP quality estimation

with the number of input parameters ranging between 4–5. However, a major limitation of ANNs is that the model interpretation remains an insurmountable proposition upon successful learning and, as a consequence, there is no direct method for estimating the significance of various input parameters. As stated earlier, evolutionary search prunes off the less significant input parameters leading to simpler models proposed in this paper. Similarly, in their award winning paper Hoene et. al. [9] present a look-up table based VoIP quality estimation model. The various MOS and corresponding parameter values would be stored in a lookup table. In the case the table does not contain a particular value of a parameter, linear interpolation is used to calculate MOS. Moreover, the model is not developed against a wider variety of input parameters. Although codec type is suggested as a network traffic variable in the abstract presentation of their VoIP quality estimation model, the number of codecs is actually restricted to 1 (i.e. AMR codec) in the model proposed therein. Our proposed models are free from such limitations. They can be used to assay the VoIP quality for any values of the input parameters which fall under the permissible range. Moreover, our models have been evolved against highly varying network conditions.

The focus of the current research has been on estimating the effect of those VoIP traffic parameters that affect the listening quality of a telephone call. A future objective would be to derive a model for conversational quality estimation of a call. Conversational quality suffers due to increase in the end-to-end delay of a call. Clearly, our next objective would be to estimate the combined effect of VoIP traffic parameters including the end-to-end delay on call quality.

References

1. ITU-T: Methods for subjective determination of transmission quality. International Telecommunications Union, Geneva, Switzerland. (1996) ITU-T Recommendation P.800.
2. ITU-T: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. International Telecommunications Union, Geneva, Switzerland. (2001) ITU-T Recommendation P.862.
3. ITU-T: Single-ended method for objective speech quality assessment in narrowband telephony applications. International Telecommunications Union, Geneva, Switzerland. (2005) ITU-T Recommendation P.563.
4. ITU-T: The E-Model, a computational model for use in transmission planning. International Telecommunications Union, Geneva, Switzerland. (1998) ITU-T Recommendation G.107.
5. ITU-T: Methodology for Derivation of Equipment Impairment Factors From Subjective Listening-Only Tests. International Telecommunications Union, Geneva, Switzerland. (2001) ITU-T Recommendation P.833.
6. Sun, L.F., Ifeachor, E.C.: perceived speech quality prediction for voice over ip-based networks. In: IEEE International Conference on Communications (ICC). Volume 4. (2002) 2573–2577

7. Mohamed, S., Cervantes-Perez, F., Affi, H.: Integrating networks measurements and speech quality subjective scores for control purposes. In: Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM). (2001) 641–649
8. Mohamed, S., Rubino, G., Varela, M.: A method for quantitative evaluation of audio quality over packet networks and its comparison with existing techniques. In: Measurement of Speech and Audio Quality in Networks (MESAQIN). (2004)
9. Hoene, C., Karl, H., Wolisz, A.: A perceptual quality model for adaptive voip applications. In: In Proc. of International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), San Jose, California, USA (2004)
10. Pennock, S.: Accuracy of the perceptual evaluation of speech quality (pesq) algorithm. In: Measurement of Speech and Audio Quality in Networks (MESAQIN). (2002)
11. Sun, L.F., Ifeachor, E.C.: Subjective and objective speech quality evaluation under bursty losses. In: Measurement of Speech and Audio Quality in Networks (MESAQIN). (2002)
12. ITU-T: Network model for evaluating multimedia transmission performance over internet protocol. International Telecommunications Union, Geneva, Switzerland. (2005) ITU-T Recommendation G.1050.
13. ITU-T: Coding of Speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP). International Telecommunications Union, Geneva, Switzerland. (1996) ITU-T Recommendation G.729.
14. ETSI EN 301 704 V7.2.1: (Digital cellular telecommunications system; Adaptive Multi-Rate (AMR) speech transcoding)
15. ITU-T: Dual rate speech coder for multimedia communication transmitting at 5.3 and 6.3 kbit/s. International Telecommunications Union, Geneva, Switzerland. (1996) ITU-T Recommendation G.723.1.
16. ITU-T: Mean opinion score (MOS) terminology. International Telecommunications Union, Geneva, Switzerland. (2003) ITU-T Recommendation P.800.1.
17. Chu, W.C.: Speech Coding Algorithms: Foundation and Evolution of Standardized Codecs. John Wiley and Sons Inc (2003)
18. Davis, L.: Adapting operator probabilities in genetic algorithms. In: Proceedings of the Third International Conference on Genetic Algorithms, San Mateo, CA (1989)
19. Luke, S., Panait, L.: Lexicographic parsimony pressure. In et. al., W.B.L., ed.: GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference, New York (2002) 829–836
20. Keijzer, M.: Scaled symbolic regression. Genetic Programming and Evolvable Machines 5(3) (2004) 259–269
21. Gustafson, S., Burke, E.K., Krasnogor, N.: On improving genetic programming for symbolic regression. In: Proceedings of the 2005 IEEE Congress on Evolutionary Computation. (2005)
22. ITU-T: Subjective performance assessment of telephone-band and wideband digital codecs. International Telecommunications Union, Geneva, Switzerland. (1996) ITU-T Recommendation P.830.