

Real-time Audio Quality Evaluation for Adaptive Multimedia Protocols

Lopamudra Roychoudhuri and Ehab S. Al-Shaer

School of Computer Science, Telecommunications and Information Systems,
DePaul University, 243 S. Wabash Ave., Chicago, IL-60604, U.S.A

Abstract. The quality of audio in IP telephony is significantly influenced by the impact of packet loss rate, burstiness and distribution on a specific audio compression technique. In this paper, we propose a novel statistical-based on-line audio quality assessment framework, Audio Genome, that can deduce the audio quality of an on-going Internet audio for many different codecs under any network loss condition at real-time. Our approach is superior to proposed learning-based techniques in terms of computational speed and ease of deployment. Our extensive evaluation experiments, that include large simulation scenarios, show that our approach is accurate and viable for adaptive real-time audio mechanisms. Finally, we show a deployment of Audio Genome as an integral part of an adaptive rate control mechanism.

1 Introduction

Audio codecs have a diverse range of compression degrees and underlying technologies. The main factors that significantly influence the evaluation of audio quality in IP telephony thus include codec type, loss rate, loss burst, inter-loss gap, delay, and recency [1] [2] [4]. ITU specifies E-Model [5] to deduce relative impairments to voice quality. However, it is a real challenge to establish a framework that derives audio quality on-line considering all of these factors. Such a framework can be highly beneficial for quality monitoring of an ongoing VoIP communication, and can be part of adaptive multi-codec audio control mechanisms that switch and mix codecs according to changing bandwidth and delay conditions to maintain optimal quality.

Audio quality of any speech processing system is generally described in terms of MOS (Mean Opinion Score) [10], the formal subjective measure of received speech quality, which is a real number between 1 and 5, where 1 is bad and 5 is excellent. In contrast to subjective testing, objective testing schemes, such as PESQ (Perceptual Evaluation of Speech Quality [12]) are automated and repeatable speech testing schemes that take into account the subjective nature of human perception. ITU-specified E-model [5][6][7][11] provides a computational model to derive relative impairments to voice quality and to estimate subjective MOS. But ITU provides no analytic methods that can directly measure the impairment due to *random loss* conditions of bursts and inter-loss gaps.

The objective of Audio Genome is to provide a statistical framework that first quantifies the effects of packet loss on various codecs by considering loss bursts, inter-loss gaps and various loss rates. We establish the audio quality as a set of functions that are derived from sufficient data generated from a large set of simulation experiments considering various codecs and a wide range of loss scenarios. Interpolation is used as the modelling technique to accurately characterize the curves representing the audio quality for codecs under any loss scenario. The resulting repository of quality information is used real-time that can assess the expected audio quality for an ongoing communication. Audio Genome, being a statistical approach, guarantees speed, accuracy and less overhead in terms of computation and data storage.

Many researchers have attempted to establish audio quality prediction models based on packet loss [2][3][13], but their work is not as comprehensive and complete as ours. Other researchers have used of neural networks [14][19] in order to provide an ongoing quality based on a set of codecs under various loss rates and distribution. Compared to these methods, our work is of lower complexity and computational delay, that can be directly applied to adaptive multimedia control mechanisms, and is also designed simple enough for easy deployment in hand-held devices. We provide a model of such an application, where Audio Genome is used to deduce the quality score for the current loss conditions in order for an adaptive multi-codec audio mechanism to take proper rate control actions.

Subsequent sections are organized as follows. Section 2 contains the related work. In section 3 we present the Audio Genome Approach. We describe the evaluation and experiment results in section 4, an application of Audio Genome in section 5, and conclusion and future work in section 6.

2 Related Work

ITU-specified E-model describes a computational model to derive relative impairments to voice quality and to estimate subjective MOS. ITU provides the measure of equipment impairment I_e for many codecs under no loss condition [11] and a limited number of codecs under very limited loss condition scenarios [7]. In addition, ITU framework does not directly consider random loss conditions of bursts and inter-loss gaps in measuring the impairment I_e . Ideally, we would like to be able to express I_e values for various codecs in fully analytic form as a function of packet loss and burstiness. However, at this point not enough subjective measurements and their specifics are available by ITU or in the literature.

Many authors have presented extensions to E-model. Cole and Rosenblath [3] described a method for monitoring VoIP applications based upon E-model, where they used curve fitting of ITU-published I_e values for selected codecs for various loss percentages. However, since they pointed out that ITU does not show a complete description of algorithms to generate loss data, they were unable to provide a complete framework of codec quality assessment, as in the case of our Audio Genome approach. In [13], the authors addressed the problem of predicting

the quality of telephone speech and classified quality prediction models based on E-Model. But they did not provide a comprehensive study of impairments due to packet loss. VQMon [2] is a non-intrusive passive monitoring system for VoIP using an extended E-model incorporating packet loss and recency effect. It derives I_e values dynamically by solving the Markov model using probabilities of loss from the observed loss. However, VQMon uses a limited "burstiness" model that, for example, does not distinguish between "burst" situation when 3 packets are lost consecutively vs. 3 packets are lost with a gap of 1 or more packets in between each loss pair, these two scenarios producing completely different quality results. A different approach has been taken by training a neural network with MOS for a set of codecs under various loss rates and distribution [14][19]. This approach is less attractive to us because of complexity in training and computational delay, which can be a problem in practical deployment of an online quality assessment mechanism.

Audio Genome attempts to bridge this gap by providing a comprehensive framework for on-line audio assessment that is easy to deploy and can be extended to many codecs. Watson and Sasse [20] have conducted extensive subjective evaluation of audio quality under packet loss compensation in multimedia conference systems. We use PESQ instead, since subjective testing is time-consuming, cumbersome, error-prone and non-repeatable.

3 The Audio Genome Approach

The Audio Genome approach can be described as follows.

- o *Generation of audio clips with packet loss scenarios:* We choose a periodic drop framework: we use a wide set of fixed inter-loss gaps (from 300 down to 2 packets) causing increasing degrees of packet loss, with a set of fixed loss burst lengths (1-4 packets) for a large number of short audio clips (9-12 sec) made by female and male voices. For a set of chosen codecs, we drop packets from audio clips using the periodic drop framework.

- o *MOS evaluation and observations:* Using PESQ, we compare 'pure' and 'poisoned' audio clips to deduce MOS scores under loss. We also observe the characteristics and behavior of each codec under various packet loss conditions.

- o *Validation of collected data:* We compare the results of representative clips with subjective testing and show that what we get analytically is close to subjective test MOS.

- o *Codec Quality Function Derivation:* We deduce codec quality functions for the collected data under loss using Interpolation.

- o *Online prediction of audio quality:* We use interpolation functions determined in the previous step to deduce the MOS for the ongoing transmission. We use weighted aggregation schemes that calculate MOS values for observed inter-loss gaps and burst sizes, and produce a combined MOS for the session so far.

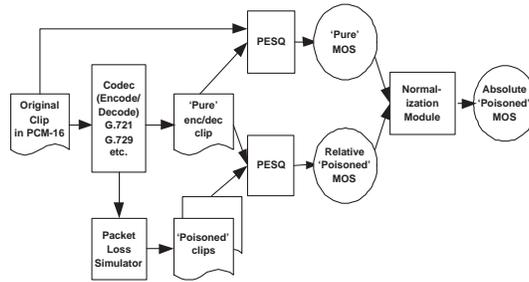


Fig. 1. Packet Loss Simulation Framework

Table 1. Audio Codecs used in Experiments

codec	Underlying technology	Bitrate kb/s	Measured MOS
G.711	waveform PCM 16 kHz	128	4.5
G.721	waveform ADPCM	32	3.04
G.729	ABS CS-ACELP	8	3.646
G.723.1	ABS MP-MLQ	5.3	3.485
GSM FR 06.10	ABS RPE-LTP	13	2.721
G.722.2			
mode (8)	ACELP	23.85	3.469
mode (6)	ACELP	19.85	3.392
mode (5)	ACELP	18.25	3.326
mode (3)	ACELP	14.25	3.1165
mode (1)	ACELP	8.85	2.718
mode (0)	ACELP	6.6	2.421

3.1 Audio Clip Generation with Loss Scenarios

Experiment Codec Set We choose G.711, as the standard audio compression technique or 'codec', a waveform PCM-16 (16 bit Pulse Code Modulation) coder of bitrate 128kbs [8] with the best quality. In addition, we choose multiple codecs of varied bitrates and underlying technology (Table 1) representing complex coding methods, such as Analysis By Synthesis (ABS) and Codebook Excited Linear Prediction (CELP). Apart from these codecs, G.722.2 (AMR-WB) is a fairly new Adaptive Multi-Rate Wideband codec with multiple bitrates [9] with not much testing results available. We choose 6 out of 9 bitrates of G.722.2 to evaluate how the different bitrates of the codec behave under degrees of packet loss in relation to each other. We measure the audio quality of each codec under no loss (Table 1 column 4) or the 'Pure' MOS, to create a referee for our measurements.

Experiment Methodology Fig. 1 depicts the packet loss simulation framework. Each original PCM-16 audio clip is encoded and decoded with every codec to create a 'pure' image with no loss, and is compared with the original clip to deduce the 'Pure' MOS. To create the 'Poisoned' clips, we drop packets during

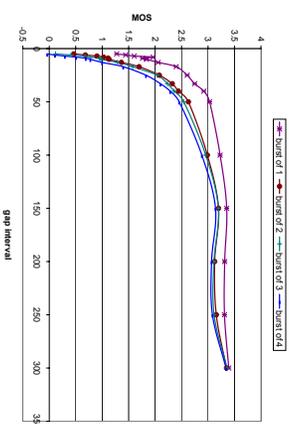
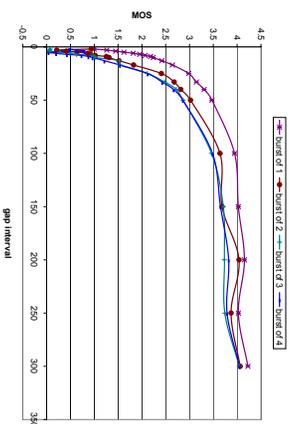


Fig. 2. Observed MOS (a) G.721, (b) G.723.1

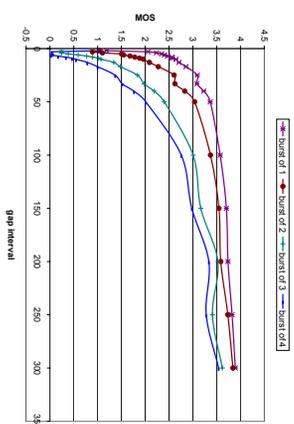
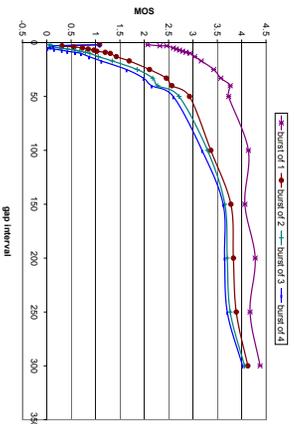


Fig. 3. Observed MOS (a) GSM, (b) G.729

encoding with each combination of gap and burst, and decode back to PCM-16. In the Packet Loss Simulator we use a wide combination of fixed inter-loss gaps from 300 down to 2 packets and a set of fixed loss burst lengths of 1, 2, 3 and 4 packets (the most occurring burst sizes as observed in the Internet [1]). We achieve a wide range of loss rates from 0.3% to 66.7% for the sake of completeness, though loss rates greater than 30% are too high for any meaningful result. We choose fixed inter-loss gaps and burst sizes in order to measure the effect of packet loss on each codec in a precise and controlled manner. We compare the 'Pure' and the 'Poisoned' images using PESQ to deduce the 'Poisoned' MOS. Since the 'Pure' image is an encoded-decoded clip, the comparison produces a quality score relative to the code score under no loss. The absolute quality is obtained by normalizing the relative quality with the ratio of measured codec 'Pure' MOS (Table 1 column 4) and 4.5, the PCM MOS under no loss, as a scaling factor in the Normalization process.

We choose a total of 24 short PCM-16 audio clips, created by sequentially truncating 4 larger clips, about 1 minute each, spoken by 2 males and 2 females, into 6 short segments of 9-12 seconds, as prescribed by PESQ [12]. The periodic drop experiment is run using every clip for each codec, and the MOS for each gap-burst combination is taken as the average of the MOS for all clips under the same degree and distribution of loss. For each codec, the Data Collection step produces four tables of data, one for each burst size, containing the MOS scores for all possible gaps of 2 to 300 packets. For example, a part of the table for

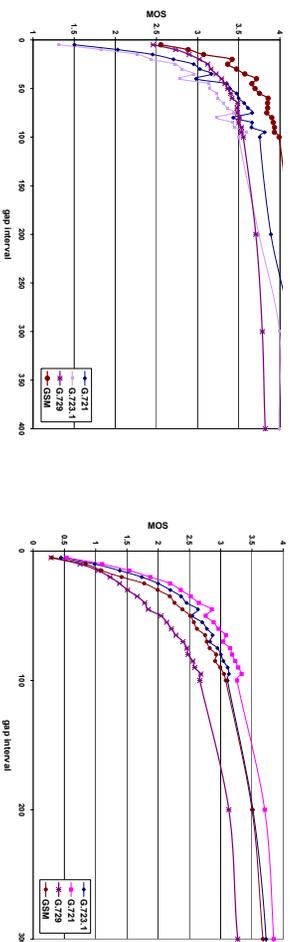


Fig. 4. Comparison of codecs (a) single loss, (b) loss bursts of 4

G.722.2 burst size 1 is $\{(x, y)\} = \{(13, 1.94), (17, 2.32), (25, 2.58), (33, 2.74)\dots\}$, where x is Inter-loss gap and y is the corresponding measured MOS.

3.2 MOS Evaluation and Observations

We observe that the quality patterns vary from codec to codec under similar loss distribution scenarios. The codecs exhibit some inherent differences in their quality degradation patterns. Fig. 2(a) and (b) and Fig. 3(a) show that the quality curves for bursts of 2, 3 and 4 are close to each other for G.721, G.723.1 and GSM, at a lower level than burst of 1. Thus for these codecs, bursts of 2, 3 and 4 have comparable detrimental effects on quality. This is counter-intuitive, since the common expectation may be that the quality degrades somewhat linearly with increasing burst sizes. Fig. 3(b) shows that G.729, on the other hand, has four distinct levels of quality degradation for 4 burst sizes. In Fig. 3(a) we see that GSM in particular, reacts to bursts comparatively more than others, the burst-of-1 curve being at a significantly higher level than other burst curves, but still manages to maintain consistently better quality than other codecs under all loss conditions. In Fig. 4(a), GSM performs as the best codec in high loss range. G.729 starts well in high loss conditions (2.45 at gap 5, i.e. 20% loss), but performs much worse than others under low loss. Fig. 4(b) shows that all codecs perform worse under high loss bursts than single loss, especially G.729 deteriorates much worse in comparison to others.

For G.722.2, surprisingly, all bitrates show comparable, almost identical quality degradation patterns under all loss scenarios. In Fig. 5(a) we take the average of all observations as the dataset for G.722.2 as a whole codec. This significantly reduces the data that need to be stored for this codec, even though it can be used with multiple bitrates.

3.3 Codec Quality Function Derivation using Interpolation

We use Newton's General Interpolation Formula with Divided Difference (Eqn. 1) [18] of degrees 1 and 2 to interpolate a MOS score from the scores for the preceding and the succeeding gaps, where x_0 is the preceding gap, y_0 is the

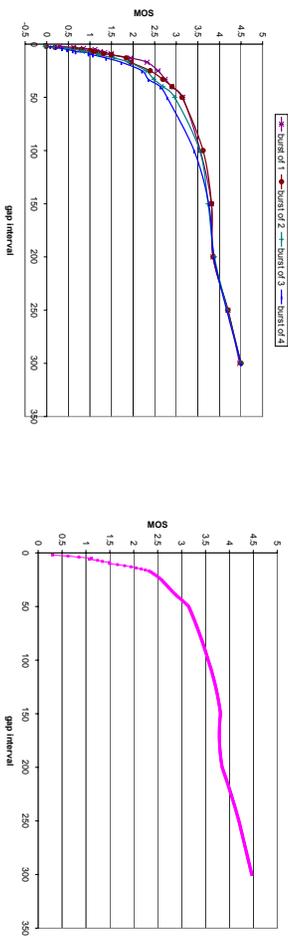


Fig. 5. G.722.2 (a) Observed MOS, (b) Interpolation of degree 2 for burst of 1

corresponding MOS, and x_1, \dots, x_n are the succeeding gaps.

$$\begin{aligned}
 y &= y_0 + (x - x_0) * \delta(x_0, x_1) + (x - x_0) * (x - x_1) * \delta(x_0, x_1, x_2) + \dots \\
 &+ (x - x_0) * (x - x_1) * \dots * (x - x_{n-1}) * \delta(x_0, x_1, x_2, \dots, x_n) \\
 &+ (x - x_0) * (x - x_1) * \dots * (x - x_n) * \delta(x, x_0, x_1, x_2, \dots, x_n)
 \end{aligned} \tag{1}$$

For example, the MOS for G.722.2 burst size 1 and inter-loss gap of 21 is estimated to be 2.45, interpolating between the values at gap length 17 and 25 using Eqn. 1 of degree 2. In Figure 5(b) the Interpolation data of degree 2, computed between each pair of data points from 2 to 300, fit the observed data for G.722.2 burst-of-1 (Fig. 5(a)) fairly well.

3.4 On-line Prediction of Audio Quality

We use the interpolation audio quality functions determined in the previous step to deduce the MOS for the ongoing transmission. The process can be described in the following steps:

- (i) *Loss Pattern Identification*: First, we monitor the loss for the session to keep track of the loss distribution and degree of burstiness observed so far, in terms of single, burst of 2, burst of 3, burst of ≥ 4 , and inter-loss gaps preceding loss bursts, in a series of (gap,burst) pairs. This will be input to determine the quality (MOS) so far.
- (ii) *Real-time deduction of MOS*: We use the packet loss pattern observed so far to derive the individual MOS values for gaps and bursts using interpolation functions, and combine them to deduce an aggregate MOS for evaluation of the ongoing session so far.

To find the MOS for each individual (gap, burst) pair as follows, we use interpolation for each observed (gap, burst) combination, where we select the stored codec table for the particular burst size, and interpolate the MOS values stored for gaps preceding and succeeding the observed gap to derive the intermediate MOS. We deduce the aggregate MOS score by combining the individual MOS scores using weighted aggregation schemes described next.

Weighted Aggregation Schemes - We first attempt to deduce the aggregate MOS by calculating the average of the individual (gap, burst) MOS values.

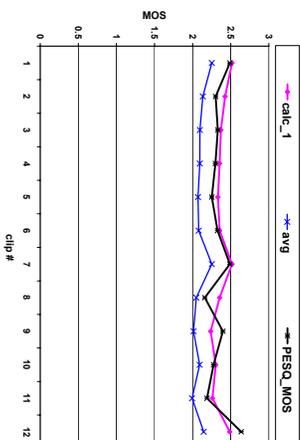


Fig. 6. Comparing Computed and PESQ MOS: G.729

However, PESQ dampens the effects of individual segment disturbances, after treating each occurrence of packet loss as a disturbance [12]. We mimic the dampening by accentuating the effects of larger inter-loss gaps, as they are analogous to lack of disturbance. We give higher importance to larger gaps to increase their effects on the overall calculated quality. We increase the accentuation in a weighted average scheme, where we give larger weights to better individual MOS values due to bigger gaps, by using a factor of the gap value as the weight. The weighted average schemes *avg* and *calc_1* are computed as follows:

$$avg = \left(\sum_{i=1}^P MOS_i \right) / P \qquad calc_1 = \sum_{i=1}^P (gap_i / 10) * MOS_i$$

where P is the number of (gap,burst) pairs observed so far.

In order to evaluate the aggregation schemes and to derive a unified MOS deduction scheme for each codec, we conducted random packet drop experiments ranging in a wide degree of packet loss ratio (2% to 30%) and burst degree (single burst to burst of 4). Fig. 6 is an example of a randomized test on G.729 for 12 clips with 9-12% loss per clip. It shows a comparison of MOS scores derived using the aggregation schemes, compared with PESQ MOS. Since the goal is to assess the quality degradation during transmission, our aim is to deduce the aggregation scheme for each codec that will predict MOS closest to the PESQ MOS under all loss circumstances. We see that PESQ MOS matches *calc_1* closer than *avg* in Fig. 6. We define the MOS deviation as the deviation between the PESQ MOS and the predicted MOS to determine the measure of the prediction accuracy, as follows:

$$MOSdev = \left(\sum_{i=1}^M (|MOS_PESQ_i - MOS_pred_i|) \right) / M \quad (2)$$

where M is the set of clips in an experiment set. We calculate the error percentage as $MOSdev/4.5$, since 4.5 is the best score possible.

Figure 7 depicts the similarities and the differences of the behavior of the aggregation schemes on different codecs. Figs. 7(a) and (b) show the overall

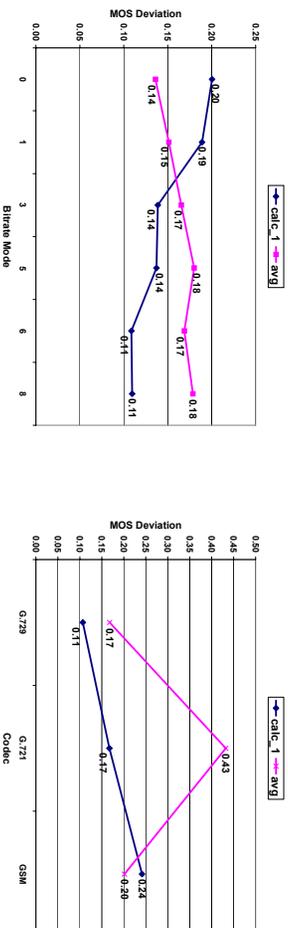


Fig. 7. Accuracy of Aggregate schemes (a) G.722.2 all modes, (b) G.729, G.721 & GSM

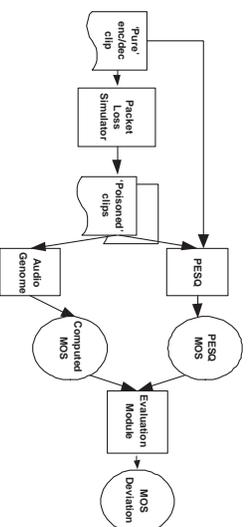


Fig. 8. Evaluation Experiment Framework

comparison of the aggregation schemes for all codecs. For most of the codecs *calc_1* stands out to be the best, except for GSM and G.722.2 modes 0 and 1, all 3 of which have *avg* as the preferred scheme. It is worth noting that the low bitrate modes 0 and 1 of G.722.2 exhibit a behavior different from the rest of the modes, even though all the modes were very close in quality under periodic loss scheme at the Data Collection phase. Table 2 column 2 depicts the aggregation schemes of highest accuracy for each codec under test. We use this set of schemes in the Audio Genome framework for evaluation in the next section.

4 Evaluation Results

The purpose of the evaluation experiments is to assess the accuracy of the Audio Genome framework under various ranges of packet loss scenarios. Fig. 8 depicts the experimental framework we use to evaluate the accuracy of Audio Genome. In the Packet Loss Simulator we conducted sets of random packet drop experiments ranging in a wide degree of packet loss ratio (2% to 40%) and burst degree and distribution. We appended small audio clips spoken by the same person in sequence to create incrementally larger 6 clips for each speaker and poisoned them randomly in order to consider the recency factor [4] of disturbance in PESQ scoring. For each of the 6 small clips as well as 6 extended clips for 4 speakers, 5 loss-degree groups, 3 testing schemes, i.e. a total of 720 clips per codec (a total

Table 2. Test Codec Set: Genome Accuracy

Codec	Overall Scheme	MeanMOSdev	Error %	Std Dev
G.729	<i>calc_1</i>	0.11	2.4%	0.07
G.721	<i>calc_1</i>	0.17	3.8%	0.09
GSM	<i>avg</i>	0.20	4.4%	0.11
G.722.2 0	<i>avg</i>	0.14	3.1%	0.12
G.722.2 1	<i>avg</i>	0.13	2.9%	0.12
G.722.2 3	<i>calc_1</i>	0.13	2.9%	0.12
G.722.2 5	<i>calc_1</i>	0.12	2.7%	0.10
G.722.2 6	<i>calc_1</i>	0.11	2.4%	0.12
G.722.2 8	<i>calc_1</i>	0.11	2.4%	0.09

of 6480 clips), we performed the following: (i) Extract the sequence of (gap, burst) pair data from each "poisoned" clip, (ii) Evaluate MOS_{PESQ} using PESQ, (iii) Deduce MOS_{pred} from Audio Genome, (iv) Evaluate the accuracy by computing the MOS deviation MOS_{dev} (eqn. 2) in Evaluation Module.

We calculated MOS_{dev} for all experiment sets and computed the accuracy of Genome under various loss degree and burstiness scenario. We observed that Audio Genome shows good accuracy for the appended clips in particular, which shows that it accommodates PESQ recency factor very well. We present the aggregate and an overall accuracy result of Genome for each codec in Table 2. Though the loss rate and degree is varied considerably over the experiments, the mean MOS deviation of Audio Genome for every codec is observed to be in a low range with low error percentage and standard deviation. Hence Audio Genome shows high accuracy under a wide range of loss scenarios.

5 An Application of Audio Genome: A Rate-Quality Optimization Mechanism

We present an example application, where Audio Genome is used to deduce the quality score for the current loss conditions in order for an adaptive multi-codec audio mechanism to take proper rate control actions. The objective of the Rate-Quality Optimization problem is to derive a codec combination set that will maximize the audio quality of the ongoing connection under the current constraints of available bandwidth, end-to-end delay and packet loss. The solution of the optimization problem is a combination ratio of codecs and/or bitrates that ensures the highest possible audio quality under current network conditions.

Problem Formulation. Maximize the audio quality under the constraint of available bandwidth and link delay.

$$\text{Maximize } z = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

subject to

$$b_1x_1 + b_2x_2 + \dots + b_nx_n \leq B$$

$$c_1x_1 + c_2x_2 + \dots + c_nx_n \leq 4.3$$

$$x_1 + x_2 + \dots + x_n = 1$$

$$d_1x_1 + d_2x_2 + \dots + d_nx_n \leq D$$

$$c_1x_1 + c_2x_2 + \dots + c_nx_n \geq 3.5$$

$$x_i \geq 0, i = 1 \dots n$$

where

x_1, x_2, \dots, x_n = percentage of each codec (type+bitrate) in the transmission mix

c_1, c_2, \dots, c_n = MOS score for each codec under current loss

b_1, b_2, \dots, b_n = bit rate of each codec

d_1, d_2, \dots, d_n = (packet size in bytes)*(encode/decode delay for 1 byte)

B = available bandwidth, $D = 400$ - link OWD

The rationale behind the constraints is as follows. *Constraint of Available Bandwidth*:The total bandwidth consumption by the codecs, expressed as the sum of the products of bitrate and percentage of each codec, should not exceed the available bandwidth. *Constraint of Delay*:The total codec delay, expressed as the sum of products of encode/decode delay and percentage of each codec, should not exceed the difference of the maximum allowable M2E delay (400ms) and the link OWD. *Constraint of Quality*:The quality sum cannot exceed the maximum quality value 4.3 (the MOS of G.711 under no loss), and should be greater than or equal to 3.5 (lower bound of acceptable speech quality). The objective function is the audio quality to be maximized, and is expressed in terms of the sum of the product of codec percentage and the codec quality score under current loss condition, as determined by Audio Genome.

6 Conclusion and Future Work

This paper presents a novel statistics-based real-time audio quality assessment framework, Audio Genome, that can deduce the audio quality of an on-going Internet audio for many different codecs under any network loss condition. We first provide an extensive experimental framework with 5 codecs G.721, G.729, G.723.1, GSM and 6 bitrate modes of G.722.2, where we quantify the effect of packet loss on the audio quality objectively by considering a wide range of loss bursts, inter-loss gaps and loss rates using interpolation as the modelling technique. For an ongoing communication, we evaluate 2 aggregation schemes to compare the predicted MOS with observed PESQ MOS. We derive a unified set of aggregation schemes for the codecs under test as the Audio Genome Model. We evaluate Audio Genome by conducting a set of extensive random loss experiments with loss degrees ranging from 2% to 40% and a wide range of packet burst distribution. For all codecs, under all loss scenarios, Audio Genome shows good accuracy: 96%-98% in average and never less than 91%. Audio Genome, being a statistical approach, guarantees speed, accuracy and less overhead in terms of computation and data storage. The framework being well-defined and repeatable, can be easily extended with any other codec. As an application, we provide a model of an adaptive multi-codec audio control mechanism that uses Audio Genome to perform rate control and maintain optimal quality.

As a future work, we plan to use multiple regression analysis for modelling the audio quality-packet loss relationships. We also plan to implement an adaptive multi-codec audio control mechanism with Audio Genome as an integral component. We intend to use Audio Genome to provide real-time feedback to

end-to-end Internet audio transport protocols in order to increase the reliability and quality of the audio session.

References

1. Bolot, J., Vega-Garcia, A.: Control Mechanisms for Packet Audio in the Internet. IEEE Infocom, San Francisco (1996) 232-239
2. Clark, A.D.: "Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality." IPTel April (2001)
3. Cole, R.G., Rosenbluth, J.H.: Voice over IP Performance Monitoring. ACM SIGCOMM (2001)
4. Cox, R., Perkins, R.: Results of a Subjective Listening Test for G.711 with Frame Erasure Concealment. Committee contribution T1A1.7/99-016 (1999)
5. ITU-T Recommendation G.107 (12/98). "The E-Model, a computational model for use in transmission planning."
6. ITU-T Recommendation G.108 (09/99). "Application of the E-model: A planning guide."
7. ITU-T Recommendation G.113 (02/96). "Transmission impairments."
8. ITU-T Recommendation G.711 (11/88). "Pulse code modulation (PCM) of voice frequencies."
9. ITU-T Recommendation G.722.2 (11/02). "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)."
10. ITU-T Recommendation P.800 (08/96). "Methods for subjective determination of transmission quality."
11. ITU-T Recommendation P.833 (02/01). "Methodology for derivation of equipment impairment factors from subjective listening-only tests."
12. ITU-T Recommendation P.862 (02/01). "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs."
13. Moeller, S., Raake, R.: Telephone speech quality prediction: towards network planning and monitoring models for modern network scenarios. Speech Communication, Vol. 38, Issue 1, pp. 47-75 (2002)
14. Mohamed, S., Cervantes-Perez, F., Afifi, H.: Integrating Network Measurements and Speech Quality Subjective Scores for Control Purposes. IEEE Infocom, Anchorage, Alaska, (2001)
15. Roychoudhuri, L., Al-Shaer, E., Hamed, H., Brewster, G.B.: Audio Transmission over the Internet: Experiments and Observations. IEEE International Conference on Communications (ICC), Anchorage, Alaska (2003)
16. Roychoudhuri, L., Al-Shaer, E., Hamed, H., Brewster, G.B.: Experiments on Audio Transmission over the Internet. Technical Report, April (2002). <http://www.mnlab.cs.depaul.edu/lroychou/techrep.pdf>
17. SAS Institute Inc. <http://www.sas.com/>
18. Scarborough, J.B.: Numerical Mathematical Analysis, 6th Ed., Johns Hopkins University Press (1966)
19. Sun L. and Ifeachor E. C.: Perceived Speech Quality Prediction for Voice over IP-based Networks. IEEE International Conference on Communications (IEEE ICC'02), New York, NY (2002) pp.2573-2577
20. Watson A. and Sasse M. A.: Evaluating Audio and Video Quality in Low-Cost Multimedia Conferencing Systems. Interacting with Computers Vol. 8 No. 3 (1996), pp. 255-275