Voice Quality Assessment for Mobile to SIP Call over Live 3G Network

G.Venkatakrishnan, I-H.Mkwawa and L.Sun

Signal Processing and Multimedia Communications,
University of Plymouth, Plymouth, United Kingdom
e-mail: L.Sun@plymouth.ac.uk

Abstract

The purpose of this paper is to assess the voice quality for mobile to SIP call on the live 3G network and to investigate the effects of codec and packet losses on the perceived speech quality. Asterisk based test platform is used with SIP client on one end and connected to 3G network on the other end to measure the speech quality on the live environment. More than 200 voice recordings are measured in the designed test bed on different codec combinations of GSM-GSM and G711-GSM (codec from SIP to asterisk – codec from asterisk to mobile phone) and results are analysed. Packet losses are introduced in the network to analyze the impact on the speech quality on the live network. The result shown that GSM-GSM codec had more impact on the speech quality with the MOS scores less than 3 whereas the G711-GSM had a fair quality with MOS scores above 3 for most cases. Packet loss is found to have major impact on voice quality and minor impact on call signalling on all the calls established with the duration of 180 seconds or lesser. A formula is derived to predict the MOS values on different packet loss conditions and validation tests on the proposed formula shown good accuracy with the prediction errors range between ±0.3 MOS for most cases. The work should help to better understand the voice quality for new services such as from 3G mobile to SIP call.

Keywords

Speech quality, codec, packet loss, PESQ, MOS, 3G

1 Introduction

Voice transmission is the most important service in mobile, telecommunication and VoIP networks that decide the Quality of Service (QoS) provided to the customer. To provide more effective services, the 3GPP (Third Generation Partnership Project) is producing a technical specifications which is globally applicable for 3G mobile system. The 3GPP group uses the IP technology end-to-end to deliver voice and other multimedia content to mobile handsets. The signalling function and the call control from terminal to network and in between network nodes are fulfilled by SIP (Session Initiation Protocol). This gives more offers to the customers to make calls between 3G phones and SIP phones apart from making calls only between 3G users. The evaluation of the perceived speech quality on such services in the live 3G network becomes an imperative task to the service providers to satisfy their customer's expectation. It is very important to investigate the speech quality and to provide information on the speech quality degradations due to different network impairments on these services. Asterisk based test platform is used with SIP client on

one end and connected to 3G network on the other end to measure the speech quality on the live environment. The different degradation factors that affect the voice quality in the live 3G network are voice codec used in the end-to-end network, network delays, delay variations and packet losses (Nortel, 2003).

In this research paper, we mainly focus on the effects of two different codec in combination in the end-to-end call. The effect of packet loss on the speech quality and call signalling from SIP client to the mobile handset through asterisk server is also analysed and the formula is proposed based on the MOS values obtained during different packet loss size. The rest of the paper is organized as follows: section 2 describes the experimental setup and different scenarios carried out using the test platform. Sections 3 present the experiment results and analysis made on the result. It also explains the proposed formula and model validation test results and prediction error range. Section 4 concludes the paper and suggests some future studies.

2 Experimental Setup and Scenarios

2.1 Test platform architecture:

A speech quality test platform is setup to objectively measure the perceived speech from the SIP phone to the mobile phone. Figure 1 shows the architecture of test bed used to provide the necessary network connectivity to evaluate the speech quality in a 3G mobile network. Four main components of the architecture are SIP client, asterisk server, network emulator and the mobile network connecting mobile phone. Asterisk is an open source hybrid TDM and full featured packet voice PBX system, used as a mediator between 3G mobile networks and the SIP phone IP network to establish live calls. The SIP client and mobile phone are used as two end users where calls are established and the voice quality measurements are evaluated. Network emulator or Netem emulates the properties of wide area network and provides the network emulation functionality for testing protocols. In this research, netem is used to introduce packet loss of different packet loss size inside the network.

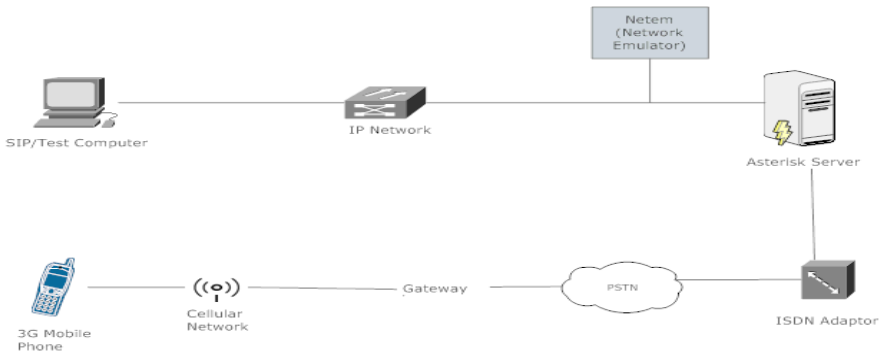

Figure 1: Test platform for speech quality evaluation

Once the test platform is set and the call is established, the speech signals are to be played in the sender end (SIP phone) and recorded in the receiver end (mobile

phone). To perform this play and record operations from SIP phone to the mobile handset, the mobile handset has to be connected to the test computer. This is done by using an electrical cable to replace the air interface such that the audio samples are played and recorded directly through the soundcard instead of hearing the sample from the ear piece and playing the sample from the microphone. It is also important to make sure that the sound card used in the process is of high quality and avoids general distortions caused when using general soundcards such as unwanted noise and gaps. Software used with the soundcard is also tested and found reliable. Loop test is also carried out on the cable that connects the mobile phone with the test computer to make sure that no distortions are introduced by the software or hardware when playing or recording the speech samples. The speech samples used in our experiments are British English reference samples and it satisfies the specifications mentioned in ITU-I P862.3

2.2 Experiment scenario to analyze the impact of codec:

2.2.1 GSM – GSM codec analysis

The calls are made from mobile phone to SIP phone. and the call is transferred from mobile phone to asterisk and asterisk to SIP phone. As it involves two different network, say IP network and the 3G network, two different codec negotiations takes place as shown in the diagram: one between the SIP phone and the asterisk server and the other between the asterisk server and the mobile phone.

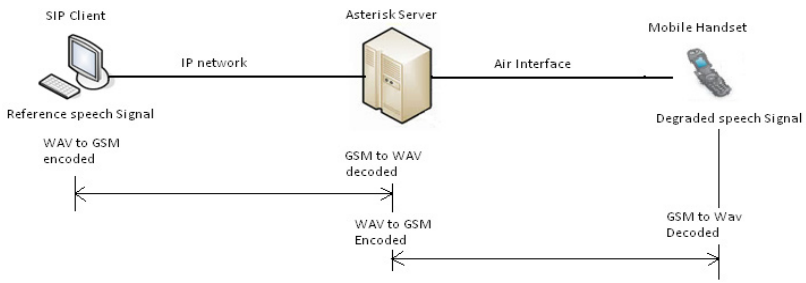

Figure 2: Experiment setup for GSM-GSM codec

In our first experiment, the testing is carried out with GSM codec as shown in the figure 2. However both the codec from SIP client to asterisk and asterisk to mobile handset are same, two different encoding and decoding happened in the process. Session initiation protocol assumes SIP client and asterisk as two user agents and negotiates the codec to compress on SIP client and decompressed on asterisk server. Asterisk server again encodes it using GSM and sends it in the 3G network and decoded in the mobile handset. The result showed that quality of speech signal degraded significantly after sending it through 3G networks. This quality degradation is mainly due to the facts; (i) the voice is encoded to GSM format and decoded to WAV format twice and (ii) the speech samples are carried through 3G network.

2.2.2 G711 – GSM codec analysis

This experiment also follows the same concept as the previous experiment and here we use G711a law codec to encode and decode from SIP client to asterisk instead of GSM codec. G711a also formally known as *Pulse code modulation (PCM) for voice frequencies* is a waveform codec using a sampling rate 8000 sample per second and operates in almost all telephony applications at 64kbps.

2.3 Experiment scenario to analyse the impact of packet loss

In order to investigate the effect of packet loss on the perceived speech quality, the 8 kHz sampled speech signal is processed by the encoder in test computer. Then the parameter based bit stream is sent to the decoder in the asterisk server. Here the asterisk server performs network emulation functionality providing a percentage of packet loss in the network downlink as well as the network uplink. After this loss simulation process, the bit streams are further processed by the 3G network and the degraded speech signal is recorded from the mobile phone. Netem, the network emulator is used to introduce the packet loss size of 5%, 10% and 20% in the IP network and different degraded MOS value measurements are taken.

Initially the speech quality MOS scores resulted in a normal quality of speech signal without having any impact on the emulated packet loss in the network. On further analysis, it is found that the packet loss created on the network using network emulator affected only the network downlink. In our experiment, the voice packets are transferred from SIP to the asterisk server, meaning the packet loss has to be introduced in the network uplink.

3 Experiment Results:

In this chapter, we present and explain the experimental results of the voice quality assessment over 3G networks.

3.1 Codec effects on live 3G network

To evaluate the quality degradation due to the codec combination, we used two different codec combination, GSM – GSM and G711 – GSM (codec from SIP phone to asterisk – Codec from asterisk to mobile phone)

From the experiments made on the GSM-GSM codec, we found that around 87.5% of the MOS scores measured on different speech samples resulted in between 2 and 3 and the remaining 12.5% went below 2 MOS score resulting in bad quality of the perceived voice sample. This is due the fact, that GSM codec has severe impact on the voice quality just by encoding and decoding and it is further reduced, when the encoded voice quality is transmitted through live 3G network.

3.1.1 Comparison of codec effects on live 3G network:

Previous research works on this field shown that, voice quality reduces to 3.55 average MOS score just by encoding and decoding and further reduces to 3.03 average MOS score when transferred through live network. The standard deviation is also found to be in the range of 0.23 and 0.26 (Goudarzi *et al.,* 2008).

	Only GSM encode and decode	**GSM on live network** (Asterisk - Mobile phone)	**GSM - GSM** (SIP - Asterisk - Mobile phone)	**G711 - GSM** (SIP - Asterisk - Mobile phone)
Avg MOS	3.555313	3.034167	2.748	2.959
STDDEV	0.235	0.262414	0.8851	0.8462

Table 1: Statistical summary on the comparison

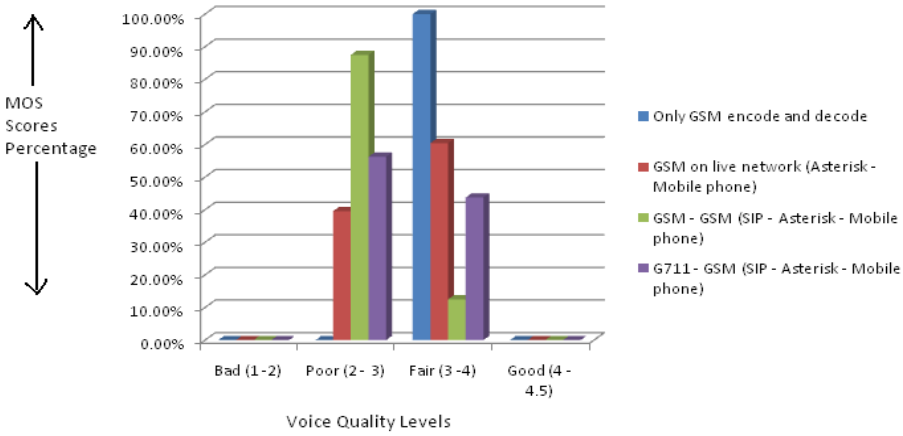

Figure 3: Comparison of PESQ – Only GSM, GSM, GSM-GSM, G711-GSM

The statistical summary in table 1 and the graph in figure 3 gives the comparison of current results with the previous work. In our first experiment case, we have encoded and decoded twice and hence the voice quality degraded more to 2.75. So, in live end-to-end call transfer, it is found that voice quality has severe effect when GSM-GSM codec is used. In the second case using G711-GSM codec, the voice quality is found to be far better than the voice quality of the previous case, having the MOS score of 3 after encoding and decoding twice. The usage of G711 codec instead of GSM codec from SIP client to asterisk increased the average MOS value to 3 from 2.75 (which is the average value of GSM – GSM codec). Another important factor to be noted here is that standard deviation of the MOS scores on codec combination is comparatively high (in the range of 0.85 – 0.89) than the standard deviation of MOS score on the single codec experiments (in the range of 0.23 – 0.26)

3.2 Packet loss effects on live 3G network

3.2.1 Impact of the packet loss on voice quality

Packet losses are introduced with different packet loss sizes, 5%, 10% and 20% and experimented. On each packet loss size, each sample is measure thrice and averaged to get reliable MOS scores. The below mentioned tabular column, table 3 gives the averaged MOS scores on voice quality on different packet loss sizes. It is clear that the MOS value decreases as the packet loss size increases from 5% to 20%

	PESQ MOS Score			
	Loss 0%	**Loss 5%**	**Loss 10%**	**Loss 20%**
Female	2.8	2.3	2	1.6
Male	3.1	2.6	2.3	1.9

Table 3: PESQ MOS scores on different packet loss size on the network

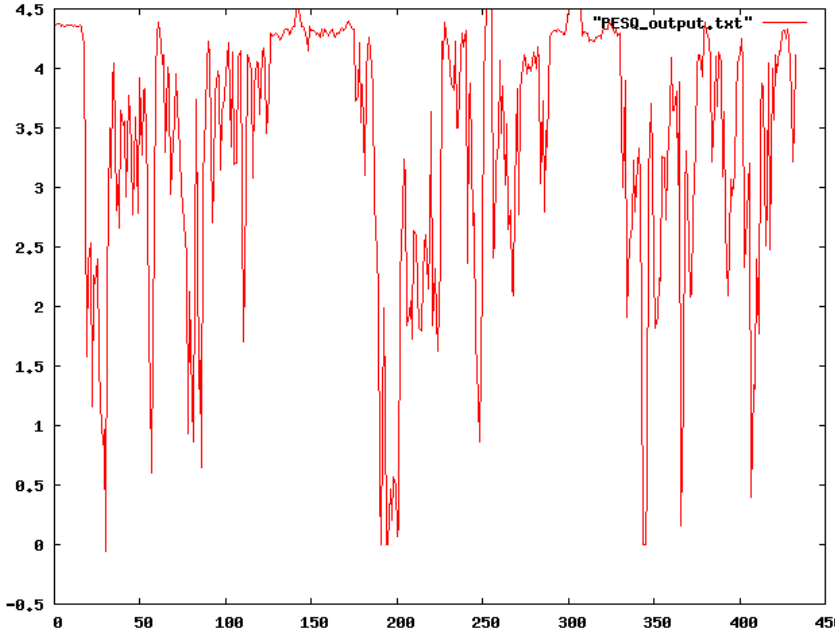

Figure 4: MOS score over number of frames for B_eng_m7.wav

Impact of talk spurt and silence on speech sample: As already mentioned in the previous sections, the speech sample used as a reference signal had three talk spurts separated by silence, is used for quality measurement. The loss at the silence segment had less impact on the perceived speech quality on all codec types and different network condition than the voiced speech segment. Moreover, it is found that the beginning of the voiced segment had more impact than the continuous speech on the perceived speech quality. Figure 5.7 shows the variation of objective MOS value over the number of frames for a speech sample B_eng_m7.wav under

packet loss size of 5%. As it can be seen, two silences in between 3 voice segment in the sample are taken place in the frames 125 -175 and 275-325 and here the PESQ scores are in the maximum of 4.5. Alternatively, the voice quality is decrease to Zero at around 25th frame, after 175th frame and 325th frame (approx) indicating the beginning of talk spurts. The impact is such that the PESQ scores are completely reduced to zero in these parts. Later parts in the talk spurts had more improved scores than the beginning indicating the quality recovery from the beginning. Due to this, the overall voice quality reduced and varied in between the MOS scores of 2.75 and 3. By this analysis, it can be concluded, that the speech samples with more talk spurts will have more impact than the speech sample having less talk spurts.

3.2.2 Impact of the packet loss on signalling part

In the previous section, the voice quality is analyzed and in this section the effect of signalling to establish the call in packet loss conditions are analyzed. To do this, numerous calls are made from mobile phone to SIP phone without any packet losses and then with packet losses. Wireshark, a packet analyzer tool is used to record the signalling of each call and to check the total time taken to establish the call. In our experiment, the call signalling happens between the asterisk server and the sip client. Once the call is initiated from mobile phone, the call is forwarded to the exchange, which in our case is the asterisk server. The asterisk server acts as a user agent and begins the message exchange by sending SIP INVITE message. The INVITE message contains the details of the type of codec supported by the called party. In our case asterisk supports G711a and GSM 06.10 codecs and this information is included in INVITE message. Then the 180 RINGING message sent in response to INVITE from SIP client to asterisk server alerting the ringing in the asterisk server. When the call is accepted in the SIP client, 200 OK response is sent to the asterisk server wherein the SIP client sends the codec information supported by it. In our case SIP is configured to support G711a codec. Asterisk server, in turn, acknowledges with the OK reply to use G711a codec there by establishing a call between SIP client and asterisk server.

In normal case without any packet loss condition, when the call is made between two clients, the calling client will take approximately 101ms to reach the calling client. In this research work, we tried to introduce packet loss of 20% and analysed. Since the packet loss introduced in the network is a random packet lost, some times the calls are established without any packet loss impact in the signalling part. But in many cases, the time taken is increased to 400ms and 1000ms. Signalling part however had the impact of increasing the signalling time to 4-10 times higher than the usual signalling time; the impact is less visible in the overall voice call connection since the time variations are in milliseconds. On all the experiments it is noted that it does not affect the call establishment between two clients. In other words, the call is established and stay connected in all the calls that has carried out for 180 seconds or lesser. From this analysis, it can be concluded that the random packet loss have less effect on signalling part of the call and the effects are visible only on the voice packets that are transferred through the network.

3.2.3 Voice Quality Prediction formula

Totally 16 speech samples are used and more than 200 sample recordings are taken under network condition with different packet loss size, 0%, 5%, 10% and 20%. Three sets of recordings each having sixteen speech samples is taken on all packet loss sizes and the average is calculated from three sets on each packet loss sizes. Figure 5.11 shows the graph plotted on the MOS scores for one sample file with packet loss size in the X-axis and MOS scores on Y axis. On the curve, a second order polynomial line is plotted fitting the original MOS score values and the polynomial equation is derived.

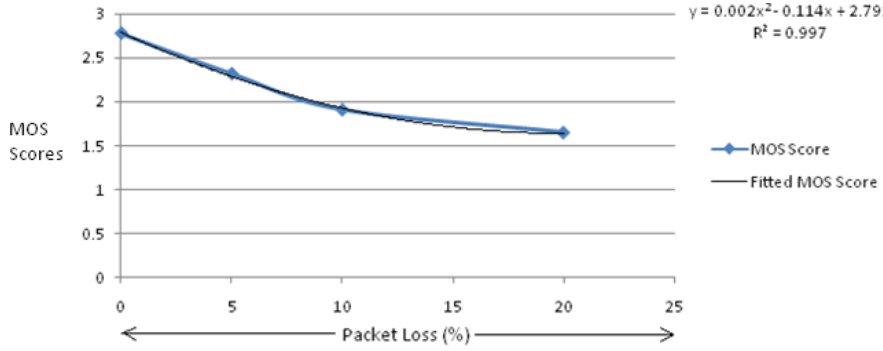

Figure 5 – MOS Score and Fitted MOS score value

The polynomial equation is found to be, $y = 0.002x^2 - 0.114x + 2.791$. Where coefficient a=0.002, b= -0.114 and c = 2.791 ≈ 2.78 which is the actual MOS value. This is the equation derived for one audio sample, *B_eng_f1.wav*.Similarly the co-efficient are calculated for other 15 sample files and the averaged co-efficient are made into an equation to give the prediction formula. The MOS score of any voice quality under x percentage of packet loss is given by the following equation,

$$y = 0.0018x^2 - 0.107x + C$$

Where C is the actual MOS score in the network without any packet loss. To determine the accuracy of the proposed formula in MOS prediction, a set of speech samples are measured for PESQ MOS score and compared with the calculated MOS score. It is found to have good accuracy on the proposed formula with the prediction errors range between ±0.3 MOS for most cases. It can be concluded that MOS can be directly predicted from the formula for the given network condition if the packet loss size and the MOS score without any packet loss is known.

4 Conclusion and Future Works:

This paper assessed the voice quality on live 3G network and effect of codec and the effect of packet losses are analyzed. GSM codec is found to have higher impact on the voice quality just by encoding and decoding the voice sample. When used on

combination, G711-GSM codec is found to have less impact than GSM-GSM codec. On different packet loss size experiments, it can be concluded that the signalling part of the call had less impact on the calls connected for 180 seconds (or lesser) and the call is established between the users all the time and the voice quality had more impacts due to packet loss. After doing more than 200 degraded sample recording, a formula is proposed to measure the MOS scores on different packet loss size, provided the percentage of packet loss on the network and original MOS score of the voice quality in the network.

In our project, both IP network and 3G network are involved and real SIP to 3G call scenario is more complicated. To name a few complexities, packet loss may be bursty instead of random packet losses, transcoding may happen in the call path instead of having two encoding and decoding process in the end-to-end call path, some techniques such as Voice Activity Detection (VAD) may be used. Future works may concentrate on these issues to evaluate more accurate MOS scores on the perceived speech quality.

5 References

Barrett, P.A. and Rix, A.W. (2002) *Applications of speech quality measurement for 3G*. Rix, A.W. (Ed). 3G Mobile Communication Technologies, 2002. Third International Conference on (Conf. Publ. No. 489). pp 250-255.

Goudarzi, Mohammad., Sun, L. and Ifeachor, E. (2008) PESQ and 3SQM measurement of voice quality over live 3G networks, 8th International MESAQIN (Measurement of Speech, Audio and Video Quality in Network) Conference, Prague, June 11, 2009

Nortel Networks. (2003) *Voice over packet An assessment of voice performance on packet networks*. Available online at: www.nortel.com/products/library/collateral/74007.25-09-01.pdf(Accessed: 10/07/09)