Time Scale Modification as a Packet loss Concealment in VoIP Applications

Hassan Yeganeh[1], Farid Mousavipour[2], Amir Hassan Darvishan[1]
[1]Iranian Telecommunication Research Center
Tehran, Iran
[2]Shahab Danesh Institution Technology of Qom
Qom, Iran
{yeganeh, darvishan}@itrc.ac.ir, Mousavipour@shahabdanesh.ac.ir

***ABSTRACT*: *In this work, Attempts focused on improving the voice quality in VoIP toward packet loss effect. The main aim was to find an efficient time scale modification method for compensating lost voice samples. This work started by introducing typical voice codecs; then, some of the well-known voice codecs were applied to the low packet loss condition and their intrinsic resistance was evaluated. Some time scale modification methods are applied for compensating the lost packet referred as packet loss concealment (PLC) method. Codec characteristics were surveyed to determine the best method for improving the quality. Evaluation of quality assessment was gained by ITU P.862.1 (narrowband PESQ) score in narrowband codecs. The results show that the quality of degraded voice significantly improved.*

Keywords: VoIP, Packet Loss Concealment, Time Scale Modification, Quality Improvement

Received: 11 December 2011, Revised 19 January 2012, Accepted 24 Janurary 2012

1. Introduction

In the past decade, a fast growth of Internet Protocol (IP) based network has occurred and Voice over IP (VoIP) services have been developed. VoIP is simply a way to make phone calls through the broadband networks such as the Internet. In a general sense, VoIP transmits voice packets via an IP based broadband network.

Due to the advance in technology over the years, traditional voice communication over the PSTN is has been characterized by high quality, often referred to as toll quality. Some factors impaired PSTN voice quality such as circuit noise and group delay distortion. In a VoIP network such impairments do not exist although some other problems are challenging VoIP networks.

The most important issue in the VoIP network is voice quality. Some factors impaired voice quality more than PSTN in VoIP network .Attempts is focused on reducing these impairments.

In the VoIP networks, voice samples are packetized that their schemes are depend on codec type. For example, in G.711 each sample represents a packet but in G.729 each 10ms frame represented a packet. In VoIP, voice quality is affected by two main factors. First, network quality of service (QoS) and second, voice codec. The main QoS impairments in VoIP network are delay, jitter and packet loss [1].

Voice intelligibility is heavily impaired by long end-to-end delay. In a VoIP network, delay may increase when two participants use different voice codecs. In [2,3] ITU-T recommends some delay limits for VoIP connections.

Jitter is the result of network congestion and improper queuing. At the sending side, voice packets are transmitted at a constant rate, while at the other end, packets may be received at an uneven rate. Jitter values between 30 and 75 ms are acceptable. High end-to-end delay and jitter can lead to the packet loss.

Packet loss is a punctual metric. An IP packet can be lost because of the congestion in the broadband network or a packet may also be dropped at the destination when arrives too late.

Some recent studies have focused on estimation of lost data as packet loss concealment methods [4]. In some early PLC methods, lost packets are replaced by random signal or copies of last received packets [5]. In other PLC methods, time-scale modification algorithms have been used to minimize the effects of lost packets. For example, in [6, 7, 8] use waveform similarity overlap-and-add technique (WSOLA) to compensate packet loss effects.

In this work, some substantial time scale modification schemes are employed for packet loss concealment. As a frame is lost in the received audio signal, the frames before and after the lost one can be employed by time scale modification algorithms to cover the gap of the lost frame. VSOLA and Hybrid techniques are three important time scale modification algorithms in speech processing [17, 18, 19]. It is expected that using this techniques results significant improvement in PLC.

Quality assessment measure is ITU-T P.862.1 which is referred as narrowband PESQ [9].

This paper is organized as follows. A brief summary of existing voice codecs is provided in Section 2. In section 3, the relationship between the quality of different codecs and packet loss in low loss condition is demonstrated. In section 4 some time scale modification methods is described and section 5 provides an evaluation of the PLC algorithms through simulation. Section 7 concludes the paper.

2. Voice Codecs

Voice codecs are the algorithms that compress the digital voice signal and encode into a specified format. Codecs are different in complexity, bandwidth and quality.

Most domestic PSTN systems operate with sample rate 8 kHz and 8-bit nonlinear quantization scheme according to [10], which encodes at 64 kb/s. The well-known voice codecs, classified to two main categories: narrowband codecs and wideband codecs. Narrowband codecs operate on voice signals filtered to a frequency range from 300 to 3400 Hz and sampled at 8 kHz, wideband codecs sampled at 16 kHz and operate on voice signals filtered to a frequency range from 50 to 7000 Hz or higher. Some parameters are mentioned for each codec such as compressing scheme, voice quality, complexity and codec delay. Codec delay is the total algorithmic delay plus processing time interval.

Narrowband codecs are classified into two categories, traditional codecs and modern codecs. Some old G.7xx series codecs that developed by ITU-T are traditional codecs. Modern codecs are recent developed codecs that announced by commercial projects such as AMR that introduced by 3rd Generation Partnership Project (3GPP). They generally have high complexity than traditional codecs.

2.1 Traditional Codecs

2.1.1 G.711

G.711 use Pulse Code Modulation (PCM) scheme to compressing and generating one 8-bit value due to 8 kHz sample rate. This codec is the oldest codec that operates in PSTN systems and has two forms, A-law that is used in Europe and Middle East and µ-Law that is used in North American and Japan. G.711 is the standard codec that used in H.323 in ISDN networks. Each frame of G.711 contains one sample of digital signal and makes the best quality than other traditional codecs as respects to low complexity. Its codec delay is 0.25 ms.

2.1.2 G.726

This codec is developed after G.723. It works at 4 bitrates, i.e., 16, 24, 32, 40 kb/s. In this work, 40 kb/s bitrate is used to gain fine quality. G.726 is a waveform speech coder which uses Adaptive Differential Pulse Code Modulation (ADPCM) and has very low complexity less than 1 millions of instructions per second (MIPS). It's codec delay is equal to G.711. It also is the standard codec used in DECT wireless phone systems.

2.1.3 G.729

This codec has lower bitrate than other traditional codecs and developed for suffering of more calls in to limited bandwidth [11]. Original G.729 runs at 8 kb/s and has good quality. Its compression scheme is CS-ACELP that works with 10 ms frames. Complexity of this codec is high (more than 20 MIPS) and codec delay is high than previous codecs which is near 25 ms. For this reason ITU-T introduced Annex A (G.729A), that has medium complexity with slightly lower quality. Speech that is encoded with G.729 can be decoded by G.729A decoder and contrariwise. G.729 Annex B comprises a VAD module and DTX module additionally.

2.2 Modern Codecs

2.2.1 AMR

This codec is an Adaptive Multi-Rate (AMR) codec. AMR was designed for speech compression in the third generation mobile telephony. AMR works at several bitrates such as, 12.2, 7.95, 5.9, and 5.15. In this work 12.2 kb/s and 5.9 kb/s bitrates are used to indicating variation of their qualities.

This codec is introduced by 3GPP project as a narrowband codec, but some modification of this codec developed for wideband codecs. AMR uses ACELP compression scheme for all bitrates. AMR can be considered to be most popular codec in mobile communication. AMR is the most widely deployed codec in the world today.

2.2.2 iLBC

Internet Low Bitrate Codec (iLBC) is a free codec, developed by Global IP Sound [12]. This codec is used in some important projects such as Skype and Google Talk. iLBC uses a block-independent linear-predictive coding (LPC) algorithm and has support for two basic frame lengths: 20 ms at 15.2 kb/s and 30 ms at 13.33 kb/s. each block samples is independent from previous block and make this codec able to endure a certain degree of packet loss. Codec delay for 20 ms and 30 ms frame modes is 40 and 60 ms respectively. In this paper 20 ms frame length is used to evaluating the quality of transmitted voice.

2.2.3 Speex

Speex codec is designed to be very flexible [13]. Speex is not designed for mobile phones but rather for packet networks application. Speex is beneficial to handle VoIP and audio books. Speex is based on CELP scheme uses 20 ms frames and design to compress voice at variable bitrate ranging from 2.2 to 44 kb/s. This codec can change bitrate dynamically. In speex, phonemes like vowels require higher bitrate to achieve good quality. Unvoiced phonemes (such as 'b' sound) can be coded with fewer bits. Speex also employs packet loss concealment for wireless environments.

Figures 1 and 2 show comparison between the quality scores of traditional and modern codecs.

The quality scores are obtained by PESQ algorithm for each voice sample then averaging scores of all speech samples (16 samples, consist of 8 males and 8 females, totally 200 seconds). The results of this measure are given on a scale from 1 to 4.5.

Figure 1 shows quality of traditional codecs and Modern codecs quality evaluation have been shown in Figure 2. In this work, the reference speech database was taken from the ITU-T dataset [14]. These sample rates are 8 kHz. Quality evaluation was performed in no packet loss condition or clean channel. Objective evaluation method is PESQ for narrowband speech signals.

For the former speech codecs, female and male results were compared to evaluate the performance of algorithm. The results show that no clear difference exists between them. Figures 1 and 2 show that the male and female quality scores are close together.

Among traditional narrowband codecs, G.711 (upper 4.1) has the best quality which G.726 has the worst quality (near 3.6). Speex and AMR (12.2) have the best quality among modern codecs and low bitrate AMR (5.15) has the worst quality (lower 3.4).

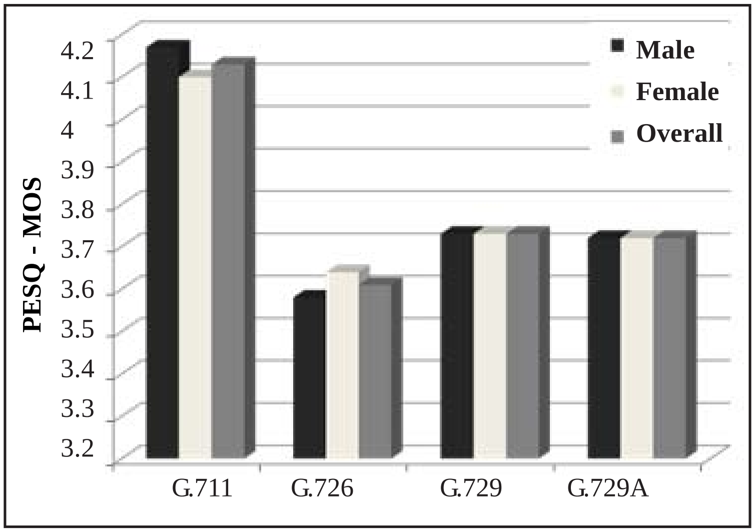

Figure 1. Traditional codec PESQ scores in no loss condition

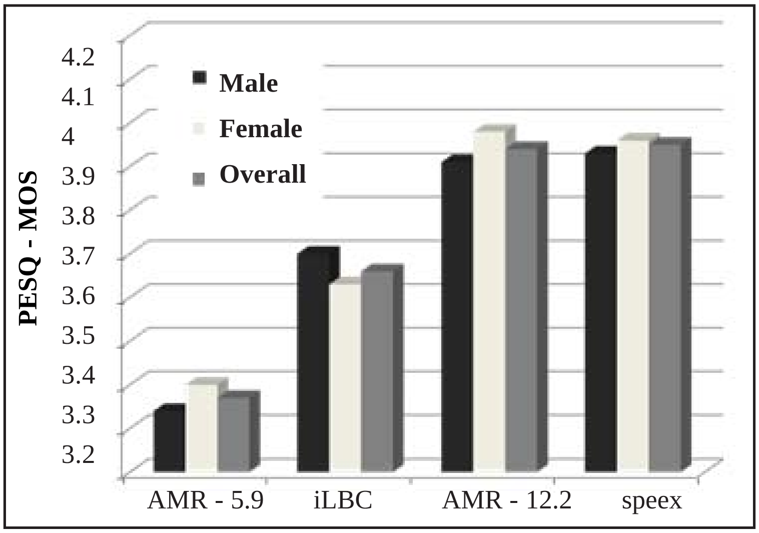

Figure 2. Modern codec PESQ scores in no loss condition

3. Packet Loss Effect

In this section the packet loss effect on voice quality is demonstrated. IP broadband networks (e.g., the Internet) are inherently best-effort networks with variable delay and loss. Voice traffic can tolerate some packet loss, where lost packets are replaced by zeros. If the packet loss rate is greater than 5%, it is considered harmful to the voice quality [11] and a good concealment technique is required for replacing instead of the lost packets. The maximum packet loss rate and required concealment algorithms are depend on the nature of the encoding algorithm.

In the VoIP networks, the original voice signal is encoded and packetized in encoder and then transmitting over the network. After IP packets pass encoder, packet loss can be occurred. On the receiver side, packet loss leads to impairment the quality of decoded voice packets. Some codecs employ their own packet loss concealment such as: packet replication.

In this work, a packet loss rate was generated from 0% to 5%, in an incremental step of 1% with a narrowband PESQ score is also evaluated quality of output voice. Figures 3 and 4 show PESQ score of traditional and modern codecs under low packet loss condition.

In Figure 3 the G.711 quality has more affected from other traditional codecs. G.711 uses PCM scheme and this feature leads to this effect. G.729 and G.729.A have similar reaction to packet loss and have close score. G.726 has the worst result but in higher packet loss rate G.711 quality decreased more than G.726. Among traditional codecs G.729 has the best resistance opposite packet loss effect but G.729.A has the best efficient codec that It performance is too close to G.729.

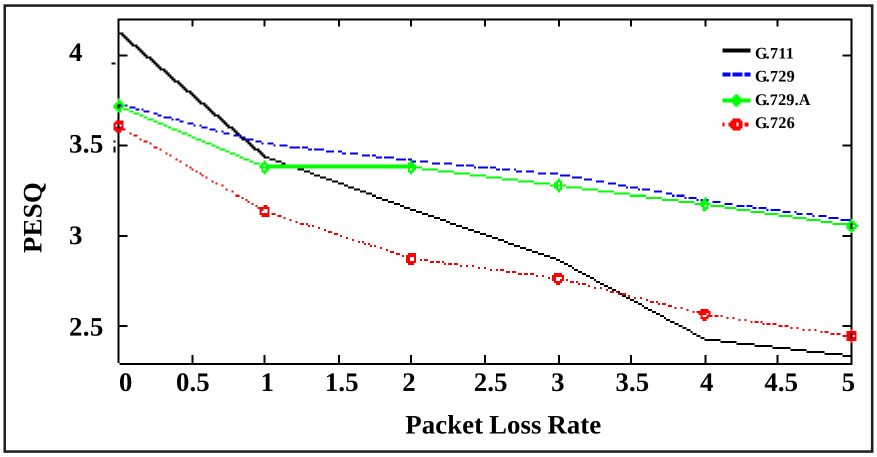

Figure 3. Traditional codecs quality versus packet loss rate

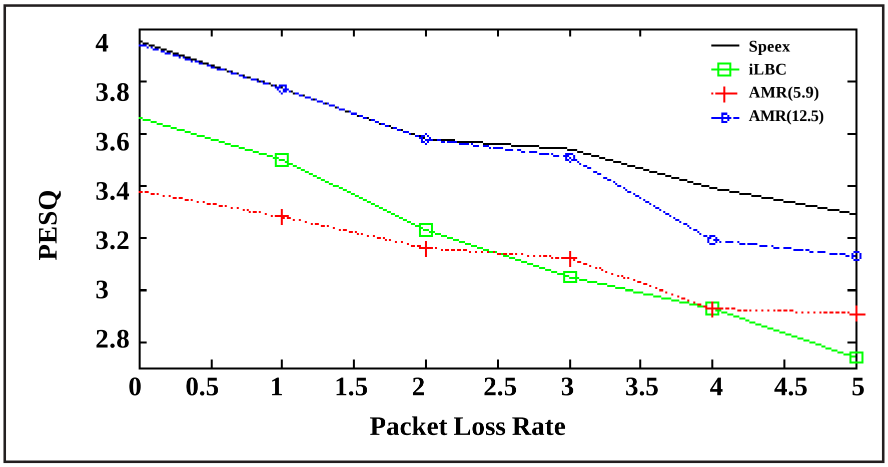

Figure 4. Modern codecs quality versus packet loss rate

In modern codecs that are shown in Figure 4 while there are speex and AMR (12.5) have the lowest sensitivity from packet loss effect and maintain their performance. In higher packet loss rate (upper 3%) speex is resisted over impairment but AMR (12.5) is impaired clearly. In this section packet loss concealment is not used for any codecs. AMR (5.9) in no packet loss condition has the lowest PESQ score near 3.4 and in higher packet loss condition its performance is closed to iLBC. Among Modern codecs, speex has the best performance.

4. Packet Loss Concealment

Many Methods have been proposed to conceal the effect of packet loss in VoIP networks. This section has been introduced some of the techniques in brief to get an idea about various approaches taken for minimizing the effect of packet loss in the VoIP networks. Two main techniques are replaced with the lost packets and time scale modification approaches. Some other PLC techniques such as: adaptive playout scheduling [16] needs jitter buffer. Its function is to buffer packets and then play them out at a steady rate.

4.1 Replacing Lost Packets

This technique consists of some methods that they replace lost packets with random signal or last samples.

Repetition is the first idea which comes to mind. With this method the lost packets are replaced by copies of last received packets. This method has not complexity and does not need high memory usage. This procedure can be enhanced by interpolating next packet and previous packet of lost packet. Interpolation action done by averaging the packets after and before the lost packet. Other method is noise replacement. This method also is applied in PSTN network. In PSTN communication systems during periods of transmit silence, when no voice signals are sent. On the received side, a comfort random signal generator (CNG) unit generates a local noise signal that it presents to the listener during silent periods. White Gaussian noise is substituted with lost voice samples.

4.2 Time Scale Modification

In this approach some time scale modification algorithms have been used to minimize the effect of packet loss. Time scale modification technique changes the duration of a voice signal while holding the signals local frequency content. This procedure leads to speeding up or slowing down the playback rate of an audio signal without changing the pitch period of original signal.

As illustrated in Figure 5, when one voice frame of the input signal is lost, the time scale modification (tsm) technique is applied to extend the time duration of several frames before the lost frame, in order to make them span across the gap of the lost frame. In this procedure, the output voice signal will appear as if there is no frame lost, and the quality of the restored signal will not decrease much. It is because the one who is listening to the output signal will further restore the waveform according to the current context consciously, and slight differences between the original signal and output signal could be properly eliminated in this procedure.

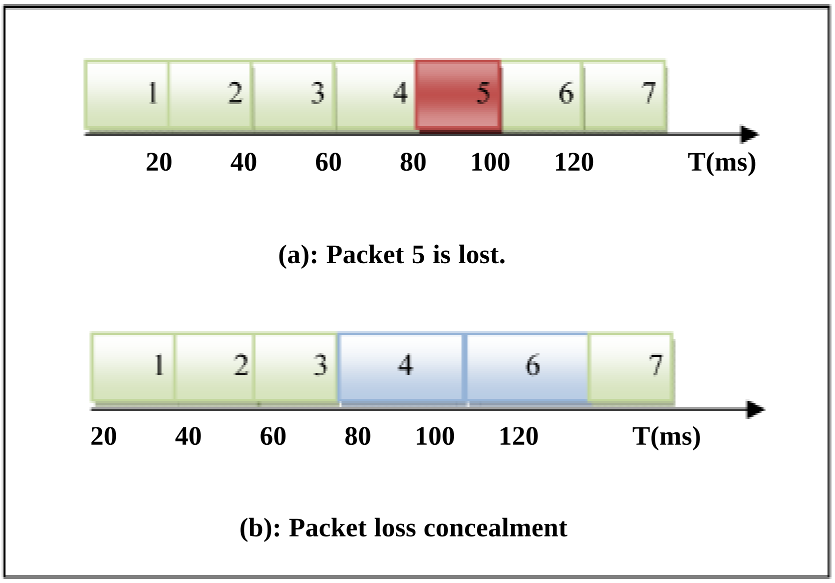

Figure 5. Packet loss concealment by time scale modification algorithm

Details of the time-scaling of the tsm algorithm are illustrated in Figure 5, and the procedure of restoring a voice frame lost in the transmission by the tsm algorithm is shown. When frame loss is occurred, the tsm algorithm extracts several voice frames of L samples from the frames before the lost one. Then the voice segments extracted is overlapped and added together with a constant step length.

The main attentions in choosing a time scale algorithm are the performance quality of output voice and the efficiency of algorithm. Time scale modification methods are classified to time domain and frequency domain. Some variations of time scale modification applied in this work such as WSOLA, hybrid and VSOLA.

4.2.1 SOLA

SOLA algorithm [13] estimates a signal from its modified short-time Fourier transform (STFT). This algorithm minimizes the mean squared error between the STFT of the estimated signal and the modified STFT. This algorithm can use an iterative procedure to estimate a signal from its modified STFT magnitude. SOLA especially attempts to minimize a similar measure though manipulation of the input signals STFT. This algorithm requires numerous iterations (about 50 iterations) of the STFT analysis for each synthesis cycle. This makes that high computationally demanding for contemporary real-time application such as VoIP.

4.2.2 WSOLA

WSOLA specifies the timing tolerance of the input voice frames during overlap-and-add procedure. The WSOLA is operated based on minimizing the distance function between the short-time Fourier transforms (STFTs) of the original signal and the time scaled signal [7,8]. But this kind of solution is likely to destroy the original phase relationships, and reduce the quality of the original voice signal, such as the pitch frequency. Fortunately, the WSOLA algorithm does all the work in the time domain, and could time-scale the input signal without the side effect of modifying the pitch frequency.

4.2.3 PSOLA

Phase vocoder is a well known technique for time scaling and pitch shifting voice signal via modification of their STFT [14]. Phase vocoder is a frequency domain technique that employs a fixed overlap-add approach. Synchronization between overlapping frames is obtained by changing phases in signal's STFT. The phase vocoder algorithm introduces specific artifact such as "*phasiness*" and "*transient smearing*". Transient smearing occurs even with modification factors that are close to 1. Phasiness or loss of presence has been occurred even with near unity modification factors. [14] Is presented new phase calculating techniques that significantly reduce the artifacts. In addition, new PSOLA method reduces the computational cost of the phase vocoder by more than a factor of two.

4.2.4 VSOLA

Variable parameter synchronized overlap-add (VSOLA) [15] is an efficient audio time scale modification algorithm that takes advantage over SOLA algorithm.

Standard SOLA parameters are generally fixed. For example length of overlapping frame N is typically 20-30 ms and analysis step size Sa is N/2.in VSOLA an adaptive and efficient SOLA parameter set is proposed that Sa and N describe in equation 1 and 2.

$$\text{Sa} = \frac{L_{stat} - SR}{|1 - \alpha|} \qquad (1)$$

$$\text{N} = \text{SR} + \alpha \, \text{S}_a \qquad (2)$$

Where L_{stat} is the stationary length of signal (20-30 ms) and SR is the search range that means the longest likely pitch period in speech samples. á also is time scale modification factor. VSOLA use algorithm that operate within a sub band framework.

4.2.5 Hybrid

The hybrid algorithm uses the appropriate features of both time domain and frequency domain of time scale modification and reduces the presence of the phasiness artifact associated with frequency domain techniques, without the need to find pitch period accurately. Hybrid algorithm uses the same principles to provide a good initial set of phase estimate.

The algorithm is both robust and efficient and has high performance. These attributes make it particularly appropriate for the time scale modification of general speech signal where no prior knowledge of the original speech signal exists.

5. Simulation Results

This section shows the main results have obtained by simulation for packet loss concealment techniques that was for compensate lost packets.

In order to evaluate the PLC methods two category of methods are implemented. Firstly replacing lost packets category that include random signal generating, repeating and repeating-interpolation methods. Second category is time scale modification techniques that are include: WSOLA, VSOLA, and Hybrid methods.

These PLC techniques are applied to G.711 and speex codecs. Each voice sample is segmented to equal frame length then given number is assigned to each frame as frame number. Lost rate is work based on the generation of random number in range of the frame number. Packet loss rate is from 0 to 30 % in an incremental step of 2%. In Figure 6 and 7 performance evaluation of G.711 demonstrated. WSOLA is applicable method that demonstrated in various papers such as [7, 8].

VSOLA, WSOLA, and Hybrid simulated for comparison among tsm methods.

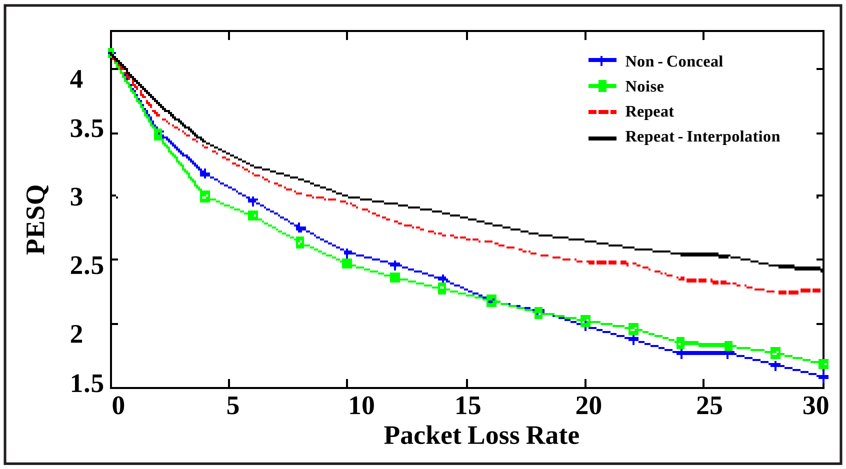

Figure 6. Replacing PLC techniques for G.711

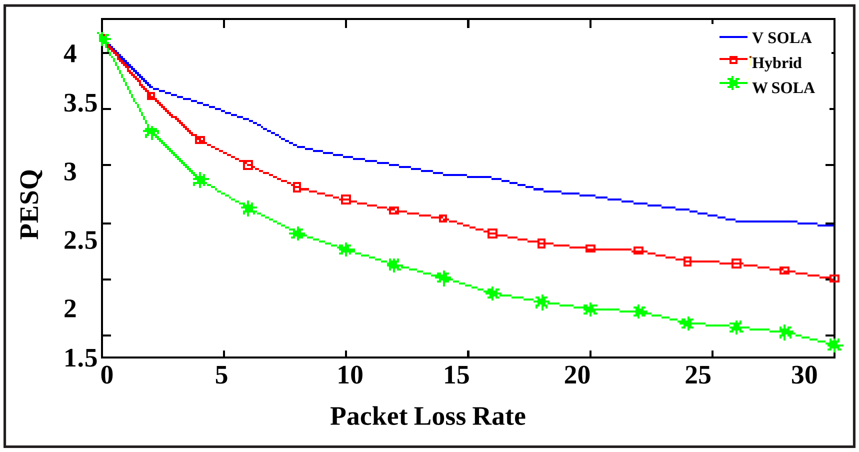

Figure 7. Time scale modification PLC techniques for G.711

In Figure 6, noise is a white Gaussian noise that is replaced with lost samples. In this Figure, Noise mode is similar non-concealment mode that means this technique is not proper unless for high loss rate (upper18%).Repeat-interpolation technique is very effective and compensate lost packet properly. In high loss rate (near 30%) PESQ score is about 2.5 and shows fair quality of speech signal.

Figure 7 shows in G.711 codec, VSOLA has the best performance in low and high loss conditions. WSOLA method has the worst performances. Hybrid method also has the acceptable performance with respect to high cost computation among time scale modification methods.

Figure 8 and Figure 9 show performance of PLC methods in Speex codec. Figure 8 is similar Figure 6 that means repeat-

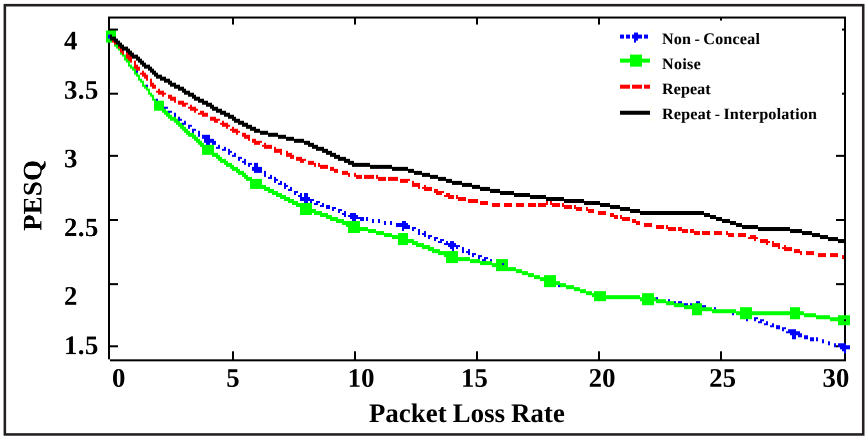

Figure 8. Replacing PLC techniques for speex

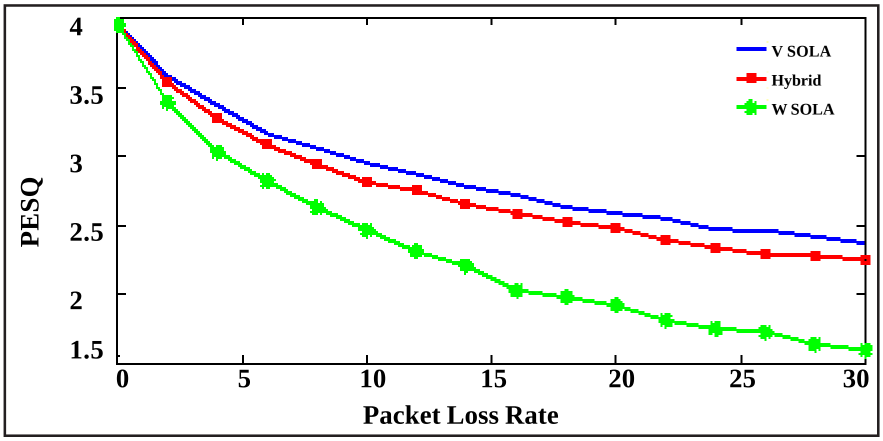

Figure 9. Time scale modification PLC methods for speex

interpolation technique still has the best performance.

In time scale modification techniques still VSOLA has the best performance with lower cost computation than WSOLA and Hybrid.

6. Conclusion

In a broadband IP network that is designed for VoIP application, end-users connect via ADSL modem to the network with a significant application being VoIP. The quality of transmitted voice is an important issue in VoIP network. This work attempts to improve the quality of speech signal in the VoIP. The PLC technique has been used to compensate lost speech data and improve voice intelligibility respectively.

A number of traditional and modern voice codecs were selected for evaluating the voice quality under non packet loss condition. Speex and AMR (12.5) among modern codecs and G.711 among traditional codecs were found to provide the best quality with no clear difference between males and females. The AMR (5.2) among modern codecs and G.726 among traditional codecs had the worst quality. Our evaluation measure score is based on narrowband PESQ.

In the next step the packet loss condition has been applied to the both Modern and traditional codecs. In this condition, all the codecs impaired the signal but the performance of speex and G.729 remain at an acceptable level.

Two major techniques deal with PLC techniques were implemented. Random packet loss condition is simulated for evaluating performance of techniques. VSOLA method among time scale modification category has the best performance and repeat-interpolation also was the best choice in replacement PLC category. The speex codec in both low packet loss and high packet loss conditions exhibits fine performance. The main disadvantage of this codec is its high complexity computation, and increased latency in encoder and decoder operation (30 ms for each frame).

Finally, this work demonstrated speex as the best codec for VoIP application in broadband networks that expect to have high quality and low packet loss ratio.

7. Acknowledgment

The work is supported in part of an ITRC grant under the Monitoring and Governance of Design and Implementation Research Pilot for Broadband Network in the Technology, Network and User Area Project.

References

[1] Karapantazis, S., Stylianos, F. P, (2009). VoIP: A comprehensive survey on a promising technology. Computer Networks. 53, 2050-2090.

[2] ITU-R Recommendation G.114. (2003). SERIES G: transmition systems and media, digital systems and networks: One way transmision time.

[3] ITU Center of Excellence for Arab Region; Quality of Services for Wireless/Fixed Communication Systems, (2010).

[4] Qiao, Z. (2008). Enhancement of Perceived Quality of Service for Voice Over Internet Protocol Systems, PhD thesis, University of Plymouth, UK.

[5] ITU-T Recommendation G.711 Appendix I, Ahigh quality low complexity algorithm for packet loss concealment with G.711, (1999).

[6] Moon-Keun Lee, Sung-Kyo Jung, Hong-Goo Kang, Young-Cheol Park, Dae-Hee YounA. (2003). Packet loss concealment algorithm based on time-scale modification for CELP-type speech coders, ICASSP, 1 (4) I-116 - I-119.

[7] Li Mojia, Wu Muqing, Wu Dapeng, Wang Lizhong, Xu Chunxiu. (2009). Packet Loss Concealment Using Enhanced Waveform Similarity OverLap-and-Add Technique with Management of Gains, WiCom '09, p. 1– 4.

[8] Lizhong Wang, Muqing Wu, Mojia Li, Lulu Wei. (2009). A packet loss concealment method base on GWSOLA algorithm and signification transient detected, IC-NIDC, p.745–749.

[9] ITU-T Recommendation P.862 Corrigendum 1, (2007).

[10] ITU-T Recommendation G.711.0, (2009). Lossless compression of G.711 pulse code modulation.

[11] ITU-T Recommendation G.729, (2008). Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP).

[12] iLBC Official Website. <http://www.ilbcfreeware.org>.

[13] Speex Official Website. <http://www.speex.org>.

[14] ITU-T Recommendation P. 50 Appendix 1, Test signals, (1998).

[15] Merazka, F. (2008). Repetition-based packet lost concealment method for CELP-based coders in packet networks, Systems, Signals and Devices,. IEEE SSD. p. 1–5.

[16] Liang, Y. J., Farber, N., Girod, B. (2003). Adaptive Playout scheduling and Loss Concealment for voice communication over IP Networks", Multimedia, IEEE on, 5 (4) 532–543.

[17] Griffen, D, W., Lim, J. S. (1984). Signal Estimation from modified short-time Fourier Transform, IEEE trans on acoustics, speech and signal processing, 32 (2) 236-243.

[18] Laroche, J., Dolson, M. (1999). Improved Phase Vocoder, Speech and Audio Processing, IEEE Transactions on speech and audio processing, 7 (3) 323 –332.

[19] Dorran, D., Lawlor, R. (2003). An efficient time-scale modification algorithm for use within a subband implementation, International Conference on Digital Audio Effects, p. 339-343.