Comparison of Common PLC Methods Used in VoIP Networks

Z. Becvar, J. Zelenka, M. Brada and L. Novak
Department of Telecommunication Engineering
Czech Technical University
Technicka 2, Prague, 166 27, Czech Republic
Phone: (+420) 2-2435 5994 Fax: (+420) 2-3333 9810 E-mail: becvaz1@fel.cvut.cz

Keywords: Speech quality assessment, PLC, VoIP, Subjective test, Objective test, MOS

Abstract - VoIP is very emerging technology in last years and telecommunication operators seek to profit from it. The main advantage of this technology is usage of existing infrastructure in the form of wide coverage of Internet connection. Unfortunately, this advantage brings some weak points that are expectable because of the quality of Internet connection. The quality of a signal on a receiving side is affected by many disturbances that have to be suppressed. This paper deals with some methods which are used for suppressing of consequences of these effects. There were compared several methods implemented only on the receiving side of telecommunication chain. The efficiency of these methods was assessed by both subjective and objective tests.

1. INTRODUCTION

We can recognize several problems connected with VoIP (Voice over Internet Protocol) technology such as the packet loss, packet delay and packet delay variation (jitter). The most of these problems almost lead to the loss of data flow continuity and furthermore to the loss of a signal information element. This loss is represented to a human perception system like a dropout.

Generally, the methods for avoiding these dropouts can be distinguished to a couple of kinds. First, the methods those operate both on a transmitting side and on a receiving side of a telecommunication chain. They are commonly called transmitter–receiver based PLC (Packet Loss Concealment) methods. As a logical reason, there is a need of cooperation between transmitting and receiving side, but this cooperation can be hardly performable, especially if we consider a cross-country connection over many telecommunication operators. These transmitter–receiver PLC methods have to be standardized to ensure the right function procedure. The second group of PLC methods is formed by receiver based PLC methods that operate only on a receiving side of the telecommunication chain. These methods benefit from no need of cooperation with a transmitting side and they can be applied everywhere without knowledge of the origin of the signal. In this paper we will consider only these one sided methods.

The solution of quality improvement is based on PLC methods which are constructed on very different approaches like LP (Linear Prediction) analysis, waveform modeling, HMM (Hidden Markov Models) models etc. Some PLC methods are implicitly contained in audio codecs used in VoIP or in GSM network and they are defined by a number of recommendations. Other PLC methods are strictly classified and their source codes and principals are unknown such in proprietary systems as Skype is.

The rest of the paper is organized as following. Next chapter describes some well known receiver based PLC methods. The third chapter deals with the tests and the fourth chapter brings some results from the tests. The fifth section presents our conclusions.

2. OVERVIEW OF TESTED PLC METHODS

There were created following receiver based PLC methods in MATLAB: repetition, silence substitution, G.711 Appendix 1 (G.711), LP, OLA (OverLap and Add) and WSOLA (Waveform Similarity OverLap and Add).

2.1 Repetition

This method [1] replaces lost packet with the last packet that was received correctly. The advantage of this principle is simple implementation. This method can be improved by the gradual changes in amplitude.

2.2 Silence substitution

Silence substitution [1] inserts a zero level signal on missing signal place. This method is also very easy to implement. The listening quality can be increased by inserting background noise instead of zero level signal [2].

2.3 ITU-T G.711 Appendix 1

This method [3] is intended as a supplement of ITU-T G.711 PCM codec. Its objective is to generate synthetic speech with characteristics not diverging the original signal.

During the normal operation this method copies the received packets into circular history buffer. Pitch period is estimated by finding the peak of cross-correlation function between the history buffer and last received speech signal.

When the dropout is detected the method tries to reconstruct the first 10 ms of speech signal at first. It takes the most recent 1.25 pitch periods of last received signal for the substitution of 10 ms of missing signal. If the pitch period is shorter, it is repeated multiple times to fill the whole 10 ms segment.

If the erasure time is longer than 10 ms further action is required. The count of repeated pitch periods increases to two because of using only the one, some unnatural artifacts would be introduced into to the speech signal. After the first good frame is detected, the reconstruction continues for a while and resulting signal is seamlessly mixed with original speech signal.

2.4 Linear Prediction

LP method [4] is based on the vocal tract parameters estimation and synthesis of the missing signal using LP model. During the short time intervals the speech can be considered stationary and the coefficients of the filter are static. We can estimate them and split the speech signal into a set of coefficients and the residual error signal.

Coefficient estimation is done by autocorrelation method and Levinson-Durbin algorithm.

The most of the work has to be done when the first packet is lost. The LP coefficients are computed and signal is filtered. If the signal energy is too low, the filter parameters are considered to be zero. Then the coefficients are saved for later use and residual signal processed for pitch detection.

Pitch period is estimated from autocorrelation of the residual signal by searching for peak locations. Then the pitch period and residual signal are used to generate excitation signal for the inverse LP filter.

When the erase is longer, then no pitch period and LP coefficients are estimated for the second and other lost packets.

Before the reconstructed signal is played, it is scaled by factor K (K=1 at the beginning) which decrease constantly with the time.

2.5 OLA

This method [5] is based on repeating the last received packet and smoothing the transition between the original and repeated packet.

When considering quasi-stationary signal as the speech signal, we can use the short-time Fourier transform (STFT) to express the relation between time and frequency domain.

$$X(w,m) = \sum_{n=-\infty}^{\infty} x(n+m)w(n)e^{-j\omega n}$$

The OLA synthesis reconstructs the original signal which STFT is maximally close to X(w,m) in the least square sense. The reconstructed signal is defined by (1).

$$y(n) = \frac{\sum_k w(n-k)\hat{y}_w(n-k,k)}{\sum_k w^2(n-k)} \quad (1)$$

2.6 WSOLA

The WSOLA method [5] is an improvement of the OLA method. It produces high quality signal and it is very robust to the background noises.

It assumes that time-scaled waveform sounding similarly to the original if it maintains similarity in all neighbors of related samples indices $m=\tau^{-1}(n)$.

We require to meet the condition (2).

$$y(n+m)w(n) \Leftrightarrow x(n+\tau^{-1}(m)+\Delta_m)w(n) \quad (2)$$

Where ⇔ means "maximum similarity" and *w(n)* is the windowing function.

The reconstructed signal is defined by (3).

$$y(n) = \sum_k v(n-kS) \times x(n+\tau^{-1}(kS)-kS+\Delta_k) \quad (3)$$

Many implementations of WSOLA algorithms are based on this idea, typically, one uses 20 ms length Hanning window with 50% overlap.

3. TESTS COMPOSITION

There were realized subjective listening test and objective test. PESQ [6] was used for objective testing. All tests were realized with Czech utterances.

Input samples used in both tests were distorted by removing of audio segment with specific length and the percentage of removed segments varied in those samples. Used segments duration was selected as a typical packet lengths used in VoIP communication. The length was 10 ms. Percent occurrences of removed segments were 2, 5, 7, 10 and 20%. Further, the removed segments were substituted by a new packet created by one of PLC method.

The parameters of these speech samples were selected to be in accordance with recommendations ITU-T P.862, P.800 and P.830.

3.1 Subjective test realization

The subjective test was composed according to ITU-T P.800 [7] and P.830 [8] recommendations. These standards define procedures of selecting speakers for source speeches, number of speakers, methods for recording and preparing of input samples (signal level, sampling rate, frequency characteristics...), required sample quantities and formats, parameters of testing environment, the amount of listeners and methods for selection and guidance of test attendants. Used test parameters and test environment setting were following:

1) Speaker selection: There were used a professional speakers from radio Akropolis and the actors. The speakers were selected regarding to their good articulation and fluent talk. There were selected 5 male and 5 female speakers and there were created about 100 different utterances.

2) Speeches preparation: Tested utterances were original studio records acquired from radio Akropolis. All acquired signals were recorded in maximal quality and then digitally processed to required form.

The utterance length was approximately 8 seconds and it was composed from burst with length between 1 and 3 seconds separated by silence. The speech was active between 40% and 80% according to the ITU-T P.862. Tested speech was 16-bit linear PCM (Pulse Code Modulation) sampled with 8 kHz sample rate.

3) Testing environment: The testing environment is the place where the listening process is in progress. There was used a quiet laboratory at Czech Technical University and listening was realized according to recommendations. The test was done with high-quality headphones in this testing place and prevented from any interruption or distortion.

4) Listeners: Applicable set of listeners was acquired from local university students and employees. The final number of test attendants was 20 persons.

5) Test phases and classification: The test was composed from two parts: Training part and Testing part. Training part consists of 3 utterances and was instrumental to identification of listeners with test classification. This part was not added into the results.

The second part of test was consisting of 70 utterances – 7 methods x 5 packet loss ratios x 2 speech per loss ratio. The utterances were played in random order, but same for all listeners. Whole listening test was designed to take between 25 to 30 minutes.

MOS scale from 5 to 1 (Excellent = 5; Good = 4; Fair = 3; Poor = 2; Bad = 1) was used for classification.

3.2 Objective test realization

Objective test was applied on the same utterances as the subjective test. Objective PESQ results and the converted PESQ results (according to recommendation ITU-T P862.1 [9]) were compared with subjective test results.

The signal is rated with MOS (Mean Opinion Score) [7]. PESQ MOS value range is defined between -0.5 and 4.5 according to ITU-T P.862 and between 1 and 5 according to ITU-T P.862.1.

The recommendation ITU-T P.862.1 [9] was created for better conformity with subjective test results. This recommendation allows conversion of objective results obtained by ITU-T P.862 to subjective scale using (4).

$$y = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.4945*x+4.6607}} \quad (4)$$

4. RESULTS

Overall performance of PLC methods in subjective test is illustrated in Fig. 1 and the average MOS scores for all percentage of packet losses are summarized in Table 1.

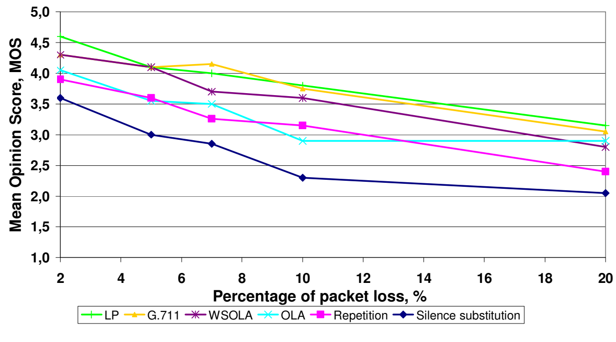

Fig. 1. Subjective MOS score of PLC methods.

PLC Method	Average MOS Score
LP	3,93
G.711	3,87
WSOLA	3,70
OLA	3,38
Repetition	3,26
Silence substitution	2,76

Table 1. Average MOS score of PLC methods obtained from subjective test

The best results in subjective test were achieved by the usage of PLC method based on linear predictive coding LP. The second most efficient PLC method was G.711, followed by methods based on overlap-and-add technique WSOLA and OLA. Two left PLC methods repetition and silence substitution did not enhance the subjective speech quality as other tested PLC methods.

The results of PLC methods in objective tests according to ITU-TP.862 are illustrated in Fig. 2 and the average MOS scores for all percentage of packet losses are summarized in Table 2.

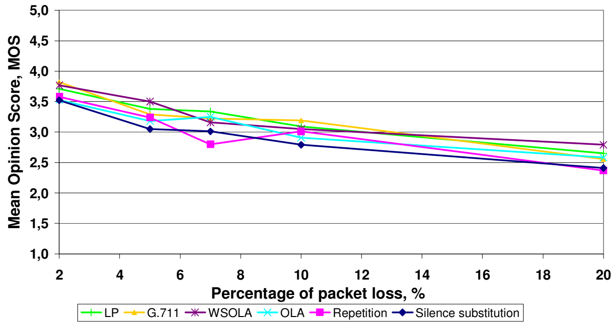

Fig. 2. MOS score of PLC methods obtained by P.862.

PLC Method	Average MOS Score
LP	3,23
G.711	3,22
WSOLA	3,25
OLA	3,09
Repetition	3,00
Silence substitution	2,96

Table 2. Average MOS score of PLC methods obtained by P.862

The recalculated results of PLC methods in objective tests according to ITU-TP.862.1 are illustrated in Fig. 3 and the average MOS scores for all percentage of packet losses are summarized in Table 3.

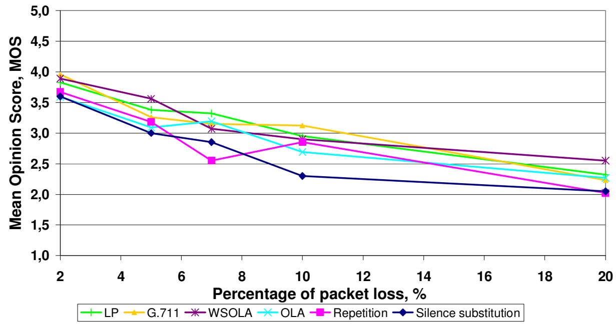

Fig. 3. MOS score of PLC methods obtained by P.862.1.

PLC Method	Average MOS Score
LP	3,16
G.711	3,14
WSOLA	3,19
OLA	2,96
Repetition	2,82
Silence substitution	2,76

Table 3. Average MOS score of PLC methods obtained by P.862.1

The best results in subjective test were achieved by LP, G.711 and WSOLA. The results of these three methods are very close to decide about the best methods. OLA method obtains the results little bit worse results. Two left PLC methods – repetition and silence substitution achieve the worst results similarly as in subjective tests.

In comparison of subjective and objective methods can be observed that subjective test provides overestimation of all values. An order of individual methods is nearly same in subjective and objective testing. The difference between P.862 and P.862.1 evaluation consist only in recounting the numbers; the character of curves is similar.

5. CONCLUSIONS

Generally, the best result of tested PLC methods has algorithm based on LP analysis that is the most powerful apparatus in connection with human speech processing.

Less satisfaction results have methods based on waveform modeling. These methods can not calculate with frequency and phase characteristics of speech and therefore the usage of these methods evoke some dissonant disturbances in speech.

The simplest methods silence substitution and repetition offer relatively good results. This is achieved because the disruptions of a signal are very short and spectral characteristic is barely intact. Practically, these methods are not used, they were considered for a comparison.

The subjective test provides overestimation of all values in comparison with objective test. The difference between P.862 and P.862.1 results is very small.

PLC methods based on LP analysis can be improved by usage of double-sided reconstruction algorithm that operates with a packet before and after the loss.

ACKNOWLEDGEMENT

This work has been supported by the grant “Specification of a qualitative criterion and resources optimization for high speed access networks” No. NPV 1ET300750402 funded by Academy of Sciences of the Czech Republic.

REFERENCES

[1] C. Perkins, O. Hodson, V. Hardman, "A Survey of Packet-Loss Recovery Techniques for Streaming Audio", IEEE Network, vol.12, No. 5, p.40-48, 1998.

[2] R. M. Warren, *Auditory Perception*, Pergamon Press, 1982.

[3] ITU-T Recommendation G.711 Appendix 1*: A high quality low-complexity algorithm for packet loss concealment with G.711*, 1999.

[4] ANSI T1.521a - 2000 (Annex B): *American National Standard for Packet Loss Concealment for Use with ITU-T Recommendation G.711*, 2000.

[5] E. Mahfuz, "Packet Loss Concealment for Voice Transmission over IP Networks", Diploma thesis, Montreal, Canada, 2001.

[6] ITU-T Recommendation P.862, *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*, 2001.

[7] ITU-T Recommendation P.800.1, *Mean Opinion Score (MOS) terminology*, 2003.

[8] ITU-T Recommendation P.830, *Subjective Performance Assessment of Telephone-Band and Wideband Digital Codecs*, 1996.

[9] ITU-T Recommendation P.862.1, *Mapping function for transforming P.862 raw result scores to MOS-LQO*, 2003.