A Packet Loss Concealment Algorithm Robust to Burst Packet Loss for CELP-type Speech Coders

Choong Sang Cho[1], Nam In Park[2], and Hong Kook Kim[2]
[1]SoC Research Center, Korea Electronics Technology Institute (KETI)
68 Yatap-dong, Bundang-gu, Seongnam-si, Gyeonggi-do 463-816, Korea
[2]Department of Information and Communications, Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro (Oryong-dong), Buk-gu, Gwangju 500-712, Korea
E-mail: [1]idealfisher@keti.re.kr, [2]{naminpark, hongkook}@gist.ac.kr

Abstract: In this paper, a packet loss concealment (PLC) algorithm for CELP-type speech coders is proposed which improves the quality of decoded speech under burst packet loss conditions. The proposed PLC algorithm is based on the reconstruction of excitation by combining voiced excitation and random excitation, where the voice excitation is obtained from the adaptive codebook excitation scaled by a voicing probability and the random excitation is generated by permutating the previous decoded excitation. The voicing probability is estimated from the correlation using the decoded excitation and pitch of the previous frames. In addition, a linear regression-based gain amplitude is estimated and applied to the reconstructed excitation for the compensatation of the undesirable amplitude change under a burst packet loss condition. The proposed algorithm is implemented as a PLC algorithm for G.729 and its performance is compared with PLC employed in G.729 by means of perceptual evaluation of speech quality (PESQ), a waveform comparison, and an A-B preference test under random and burst packet loss rates of 3% and 5%. It is shown that the proposed algorithm provides significanlty better speech quality than the PLC of G.729, especially under burst packet losses.

1. Introduction

With the increasingly popular use of the Internet, IP telephony devices such as voice over IP (VoIP) phones and voice over WiFi (VoWiFi) phones have attracted wide attention for speech communications. In order to realize an IP phone service, speech packets are transmitted using a real-time transport protocol/user datagram protocol (RTP/UDP), but the RTP/UDP does not check it out whether or not the transmitted packets are correctly received [1]. Due to the nature of this type of transmission, the packet loss rate would be higher as the network becomes congested. In addition, depending on the network resources, the possibility of burst packet loss occurring also increases, potentially resulting in severe quality degradation of the received speech [2].

In this paper, a packet loss concealment (PLC) algorithm for CELP-type speech coders is proposed as a means of improving the quality of decoded speech under burst packet loss conditions. The proposed PLC algorithm is based on the reconstruction of excitation by combining voiced excitation and random excitation, where the voice excitation is obtained from the adaptive codebook excitation scaled by a voicing probability and the random excitation is generated by permutating the previous decoded excitation. In addition, a linear regression-based gain amplitude is estimated and applied to the reconstructed excitation for the compensatation of the undesirable amplitude change under a burst packet loss condition. In order to measure the performance of the proposed PLC algorithm, we apply the proposed algorithm to the G.729 standard and compare its performance with a conventional algorithm by measuring the perceptual evaluation of speech quality (PESQ).

2. Conventional PLC Algorithm

The PLC algorithm employed in the G.729 standard reconstructs the speech signal for the current frame based on previously-received speech information. In other words, the PLC algorithm replaces the missing excitation with an equivalent characteristic of a previously received frame, though the excitation energy gradually decays. In addition, it uses a voicing classifier based on a long-term prediction gain. During the error concealment process, a 10 ms frame is declared as a voice signal if at least a 5 ms sub-frame of the 10 ms frame has a long-term prediction gain of more than 3 dB. Otherwise, the frame is declared an unvoiced signal. An erased frame inherits its class from the previous speech frame. The synthesis filter in a lost frame uses the linear prediction (LP) filter parameters of the last good frame; LP filter parameters are computed from the repeated line spectrum frequency (LSF) parameter. Finally, the gains of the adaptive and fixed codebooks are attenuated by a constant factor. The pitch period of the lost frame uses the integer part of the pitch in the previous frame. To avoid repetition of the same periodicity, the pitch period is increased by one for each subsequent subframe.

3. Proposed Algorithm

Fig. 1 shows an overview of the proposed PLC algorithm. Contrary to the conventional PLC algorithm, the proposed PLC algorithm consists of five blocks: voicing probability estimation, periodic/random excitation generation, linear predictive coding (LPC) smoothing, new excitation generation, and signal amplitude control. If frame erasure occurs, the number of lost frames is counted until next good frame. In the case of single packet loss, LPC coefficients of the previous good packet are first scaled to smooth the spectral envelope. Next, a new excitation signal is estimated by combining the periodic excitation obtained from the estimated voicing probability and the random excitation obtained via a permutation of the previous decoded excitation signal. If the next frame is also erased, i.e., consecutive frames losses occur, the signal amplitude estimation for the

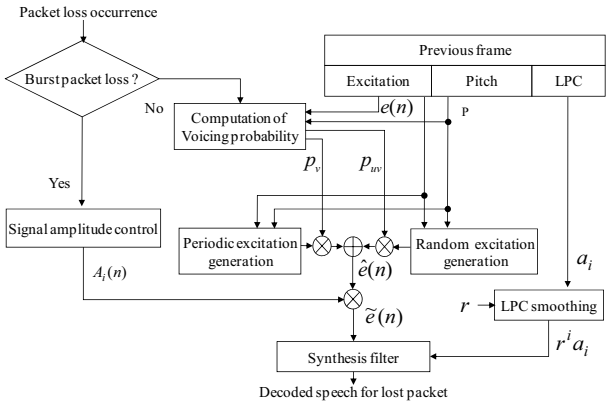

Figure 1. Overview of the proposed PLC algorithm.

erased frame is first performed prior to the excitation estimation described above. Finally, decoded speech corresponding to the erased packet is obtained by filtering the estimated new excitation with the smoothed LPC coefficient.

3.1 Generation of periodic and random excitation

The main feature of the proposed algorithm is generation of the excitation of lost frame represented by the probabilistic sum of the voiced/unvoiced excitation using the pitch and the excitation of the previous frame as shown in Fig 1. First of all, voiced excitation is generated from an adaptive codebook by repeating excitation of the previous frame during the pitch period, which is referred to as periodic excitation in this paper. That is, the periodic excitation, $e_p(n)$ is given by

$$e_p(n) = e(n-P) \tag{1}$$

where e(n) is the excitation of the previous frame and P is the estimate of the current frame pitch period.

Next, to generate unvoiced excitation which is referred to as random excitation as a fixed codebook for the PLC, temporal excitation is produced based on a random permutation of the previous excitation as following

$$x(n) = p_\pi(e(n)) \tag{2}$$

where $x(n)$ is the temporal excitation, p_π is permutation matrix and n is generated by a random sequence in the range of P. A sample is selected randomly from within a selection range having the same length of pitch period. For the selection of the next sample, the position of the pitch period, P is slightly modified by adding one to the value of the previous frame to prevent the same sample from being selected.

In addition, based on the fact that the fixed codebook contributes to periodicity of the speech signal as an adaptive codebook [3], we can compute the maximum cross correlation between periodic excitation and temporal excitation using

$$m^* = \arg\max_{0\le m\le 72} \frac{\left(\sum_{i=0}^{79} e_p(n)x(n-m)\right)^2}{\sum_{i=0}^{79} x^2(n-m)}. \tag{3}$$

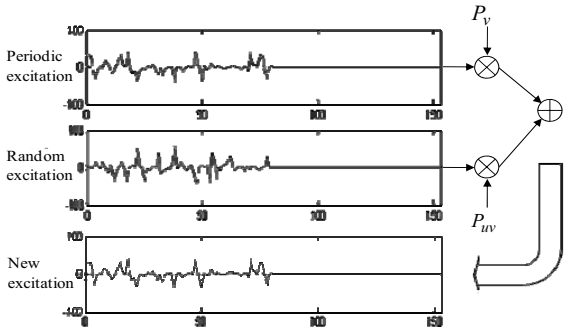

Figure 2. Structure of the new excitation generation for the proposed PLC algorithm.

By using m^*, we obtain the best random excitation which contributes to periodicity of speech signal, defined as

$$e_r(n) = x(n-m^*) \tag{4}$$

where $e_r(n)$ is the random excitation. As shown in Fig. 2, to recover the erased frame, we can obtain the reconstructed excitation using the periodic and random excitation, defined as

$$\hat{e}(n) = p_v e_p(n) + p_{uv} e_r(n) \tag{5}$$

where $\hat{e}(n)$ is the reconstructed excitation, p_v is the voicing probability, and p_{uv} is the unvoicing probability.

3.2 Calculation of voicing and unvoicing probability

The voiced and unvoiced speech signals can be classified by a correlation coefficient [4]. The voiced speech signal is highly correlated with adjacent speech signals, whereas the unvoiced speech signal has a low correlation coefficient with adjacent speech. Here, a correlation coefficient close to 1 infers that the speech signal has characteristics of voiced speech. In order to estimate the voiced and unvoiced characteristics, the maximum correlation coefficient of previous good frame is calculated by the pitch and excitation of the last received good packet, based on

$$r = \frac{\left|\sum_{n=0}^{k} e(n)e(n-P)\right|}{\sqrt{\sum_{i=0}^{k} e^2(n)}\sqrt{\sum_{i=0}^{k} e^2(n-P)}} \tag{6}$$

where r is the maximum correlation coefficient, k is the number of sample in a frame and P is the pitch period of the last received good packet.

The voicing and unvoicing probabilities p_v and p_{uv} can then be calculated as (7) and (8), respectively,

$$p_v = \begin{cases} 1, & if \quad r > 0.49 \\ \dfrac{r-0.09}{0.4}, & if \ 0.09 \le r \le 0.49 \\ 0, & if \quad otherwise \end{cases} \tag{7}$$

$$p_{uv} = 1 - p_v . \tag{8}$$

As a result, we obtain a reconstructed excitation as described in Eq. (5).

3.3 Speech amplitude control using linear regression

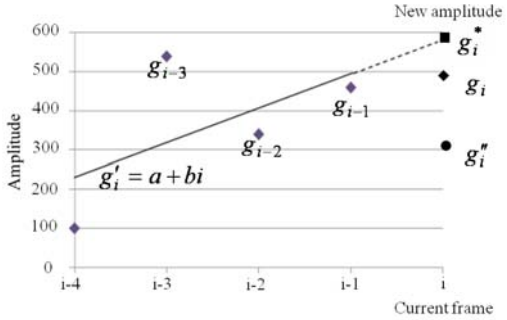

Figure 3. Amplitude prediction using linear regression.

The speech signal reconstructed using the voicing probability and periodic/random excitation has the undesirable speech amplitude due to abrupt speech amplitude change. The amplitude of the original speech rapidly decreases, though the reconstructed speech signal maintains the same amplitude. Conversely, when the amplitude of speech is rapidly increased, the amplitude of the reconstructed speech decreases based on the number of consecutive lost frames. To overcome this problem, modeling of the previous data is applied to amplitude control.

Fig. 3 shows the signal amplitude prediction using linear regression. Assuming that i is the current frame and the g_i is the original speech amplitude, the PLC employed in G.729 estimates the amplitude, g_i'' by attenuation of the codebook gain, whereas the proposed PLC estimates the amplitude, g_i^* by the linear regression. As shown in figure, amplitude obtained by linear regression estimates more than one by attenuation of the codebook gain. Then, the linear regression is based on the linear model as

$$g_i' = a + bi \tag{9}$$

where g_i' is the newly predicted current amplitude. a and b are coefficients for the first order linear function, and i is the frame [5]. Assuming that measurement errors are normally distributed and the past four amplitude values are used, we find a and b so that each difference the original speech amplitude and estimated speech amplitude from the figure is minimized. In other words, the minimum error for determining a and b is

$$\varepsilon = \sum_{j=i-4}^{i-1} \left(g_j - g_j' \right)^2 \tag{10}$$

where ε is the squared error, g_j is original past j-th amplitude. To minimize (10) with respect to a and b, the derivative regarding a and b should be equal to zero, that is, a^* and b^* are the optimized parameters such that ε is mini mized. By using these parameters, an estimate of each g_i^* is denoted by

$$g_i^* = a^* + b^* i \quad . \tag{11}$$

To obtain the amplitude of a lost packet, the ratio between the current amplitude and the amplitude of the last previous frame is first defined as

$$\sigma_i = \frac{g_i^*}{g_{i-1}} \tag{12}$$

where i is the current frame and σ_i is the i-th ratio.

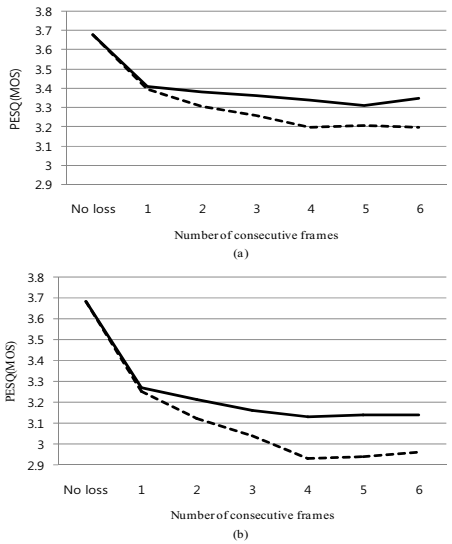

Figure 4. Comparison of PESQ scores using the proposed PLC (solid line) and G.729 PLC (dotted line) under a packet loss rate of (a) 3% and (b) 5%.

Moreover, the number of consecutive lost frames is taken into consideration by observing that if consecutive packet losses occur, the speech amplitude also decreases. We define the scale factor as

$$S_i = \begin{cases} 1.0, & if \;\; l(i) = 1,2 \\ 0.9, & if \;\; l(i) = 3,4 \\ 0.8, & if \;\; l(i) = 5,6 \\ 0, & otherwise \end{cases} \tag{13}$$

where S_i is the scale factor and $l(i)$ is the number of consecutive lost frames of the i-th frame. Then, the estimated amplitude A_i' can be determined using

$$A_i' = \begin{cases} S_i \times \sigma_i \;\; , k \neq 0 \\ 1 \qquad\quad , k = 0 \end{cases} \tag{14}$$

To prevent the discrete attenuation of the estimated amplitude, A_i' is smoothed by

$$A_i(n) = -\frac{A_{i-1}' - A_i'}{N} \cdot n + A_{i-1}' \,, \quad n = 0,...,N-1 \tag{15}$$

where $A_i'(n)$ is the smoothed amplitude of time n and N is the frame size.

Finally, we multiply $A_i(n)$ to the excitation $\hat{e}(n)$ obtained in Eq. (5) such that the amplitude ajusted excitation, $\tilde{e}(n)$, is applied to the synthesis filter.

4. Performance Evaluation

To evaluate the performance of the proposed PLC algorithm, we replaced the PLC algorithm already employed in G.729 [6] with the proposed PLC algorithm, then compared the performances by measuring their respective PESQ [7] scores. The PESQ value estimates a mean opinion score (MOS) ranging from 1.0 to 4.5; a MOS close to 4.5 infers high speech quality. For the PESQ test, 92 speech sentences (48 males and 48 females from the NTT-AT speech data-

Table 1. A-B preference test results.

Burstiness/ packet loss rate		Preference Score (%) PLC of G.729	No difference	Proposed PLC
r=0.0	3%	8.33	44.75	47.92
	5%	4.17	41.67	54.17
r=0.99	3%	2.08	54.17	43.75
	5%	4.20	37.50	58.30
Average		4.69	44.27	51.04

base [8]) were processed by G.729 using both the proposed PLC algorithm and the original PLC algorithm under packet loss rates of 3% and 5%.

Fig. 4 shows the MOS result at packet loss rates of 3% and 5%. To see how the proposed PLC algorithm works under burst packet loss conditions, the number of consecutive lost packets was changed from 1 to 6. From the Figure, it can be seen that the proposed algorithm had higher PESQ scores than the original G.729 algorithm for all packet loss rates and number of consecutive missing frames.

In order to evaluate the subjective performance, we performed an AB preference listening test, where ten speech sentences (five males and five female) were processed by both the algorithm and the proposed one under the packet loss rate 3% and 5%. The result in Table 1 shows significantly the relative preference of the proposed PLC algorithm to the PLC employed in G.729. Finally, to visually compare the PLC of the G.729 standard with the proposed PLC, we then compared the waveforms of each algorithm, as shown in Fig. 5. Fig. 5(b) shows the decoded speech waveform without loss of the original signal. After applying the packet error pattern shown in Fig 5(c), the speech reconstructed by the proposed PLC is significantly more similar to the original than what could be obtained by the PLC of the G.729 standard, as evidenced by the dotted circles in Fig 5(d) and 5(e).

5. Conclusion

In this paper, we proposed a packet loss concealment algorithm for a CELP-type speech coder that reduces the degradation of speech quality due to packet loss. The proposed PLC consists of a voicing probability, a periodic/random excitation generation, and speech amplitude control. We evaluated the performance of the proposed algorithm on G.729 under random and burst packet loss rates of 3% and 5% and compare it with PLC employed in G.729. From the PESQ measure, an A-B preference tests and a waveform comparison, it was shown that the proposed PLC algorithm provided better speech quality than the PLC employed in G.279 under packet loss rates of 3% and 5%.

Acknowledgement

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (R01-2008-000-10243-0) and by the GIST Technology Initiative (GTI).

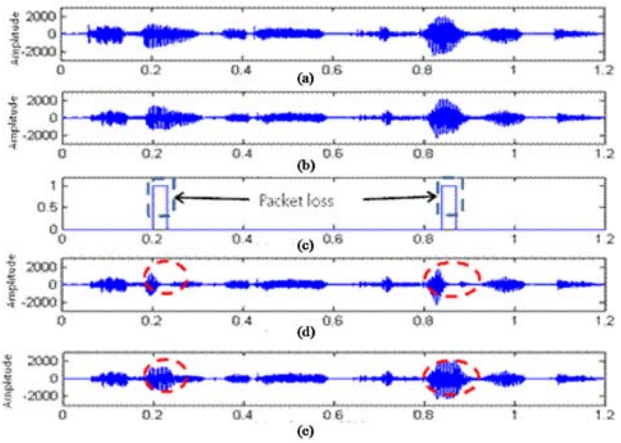

Figure 5. Waveform comparison of a decoded signal using PLC; (a) original wave, (b) decoded speech signal without packet loss, (c) packet loss pattern, (d) reconstructed speech signal using the PLC of the G.729 standard, and (e) repaired speech signal using the proposed PLC.

Reference

[1] B. Goode, "voice over internet protocol (VoIP)," *Proceedings of the IEEE*, vol. 90, no.9, pp. 1495-1517.month, 2002.

[2] W. Jian and H. Schulzrinne, "Comparision and optimization of packet loss repair methods on VoIP perceived quality under bursty loss," in *Proc. NOSSDAV,* pp. 73-81, May 2002.

[3] H. K. Kim and M. S. Lee, "A 4 kbps adaptive fixed code excited linear prediction speech coder," in *Proc. ICASSP*, vol. 4, pp. 2303-2306, Mar. 1999.

[4] J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*, IEEE Press, 2000.

[5] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes The Art of Scientific Computing,(3rd Ed.)*, Cambridge University Press, 2007.

[6] ITU-T Recommendation G.729, *Coding of speech at 8 kbit/s using conjugate-structure code-excited linear prediction (CS-ACELP)*, Feb. 1996.

[7] ITU-T Recommendation P.862, *Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech coders*, Feb. 2001.

[8] NTT-AT, *Multi-lingual speech database for telephonometry,* 1994.