MDCT-Domain Packet Loss Concealment for Scalable Wideband Speech Coding

Nam In Park and Hong Kook Kim

School of Information and Communications
Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea
{naminpark,hongkook}@gist.ac.kr

Abstract. In this paper, we propose a modified discrete cosine transform (MDCT) based packet loss concealment (PLC) algorithm in order to improve the quality of decoded speech when a packet loss occurs in scalable wideband speech coders using MDCT as spectral parameters. The proposed PLC algorithm is realized by smoothing MDCT coefficients between the low and high bands for scalable wideband speech coders. In G.729.1, a typical scalable wideband speech coder standardized by ITU-T, two different PLC algorithms are applied to low band and high band in time and frequency domain, respectively. Thus, the MDCT coefficients around the boundary between the low and high band can be mismatched. The proposed PLC algorithm is replaced with the PLC algorithm applied to the high band, and it compensates for the mismatch in the MDCT domain at the boundary. Finally, we compare the performance of the proposed PLC algorithm with that of the PLC algorithm employed in G.729.1 by means of perceptual evaluation of speech quality (PESQ), an A-B preference test, and a waveform comparison under different random and burst packet loss conditions. It is shown from the experiments that the proposed PLC algorithm provides significantly better speech quality than the PLC of G.729.1.

Keywords: Packet loss concealment (PLC), wideband speech coding, modified discrete cosine transform (MDCT), G.729.1.

1 Introduction

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

Most speech coders in use today are based on telephone-bandwidth narrowband speech, nominally limited to about 200-3400 Hz and sampled at a rate of 8 kHz. On

T.-h. Kim et al. (Eds.): UCMA 2011, Part II, CCIS 151, pp. 11–18, 2011.

the contrary, the wideband speech coders have been developed for the purpose of smoothly migrating from narrowband to wideband quality (50-7,000 Hz) at a rate of 16 kHz in order to improve speech quality in services. That is, voice services using wideband speech not only increase the intelligibility and naturalness of speech, but also add feeling of transparent communications. Especially, G.729.1, a scalable wideband speech coder, improves the quality of speech by encoding the frequency bands left out by the narrowband speech coder, G.729. Therefore, encoding wideband speech using G.729.1 is performed in two different approaches which are applied to low band and high band in time and frequency domain, respectively. In particular, when a frame loss occurs, the low band and the high band PLC algorithm work separately. In other words, the low band PLC algorithm reconstructs excitation and spectral parameters of the lost frame from the last good frame. Also, the high band PLC algorithm reconstructs spectral parameters, e.g., typically modified discrete cosine transform (MDCT) coefficients, of the lost frame from the last good frame. Of course, the excitation of the low band could be used for reconstructing the excitation of the high band if bandwidth extension from the low band to the high band is employed in wideband speech decoding [3][4].

In this paper, we propose a modified discrete cosine transform (MDCT) based packet loss concealment (PLC) algorithm in order to improve the quality of decoded speech when a packet loss occurs in a scalable wideband speech coder using MDCT as spectral parameters. Here, we select ITU-T Recommendation G.729.1, which is a typical MDCT-based scalable wideband speech coder employing G.729 as a narrowband speech coder. Since two different PLC algorithms in G.729.1 are applied to low band and high band in time and frequency domain, respectively, the MDCT coefficients around the boundary between the low and high band can be mismatched. The proposed PLC algorithm is replaced with the PLC algorithm applied to the high band, and it compensates for the mismatch in the MDCT domain at the boundary.

The remainder of this paper is organized as follows. Following this introduction, Section 2 describes a conventional PLC algorithm currently employed in the G.729.1 decoder [5]. After that, Section 3 proposes an MDCT-based PLC algorithm that is implemented by replacing the conventional PLC algorithm in G.729.1. Section 4 then demonstrates the performance of the proposed PLC algorithm, and this paper is concluded in Section 5.

2 Conventional PLC Algorithm

The PLC algorithm employed in the G.729.1 standard reconstructs speech signals of the current frame based on previously received speech parameters such as excitation in the low band and the MDCT coefficients in high band, which is shown in Fig. 1 [5]. In other words, the PLC algorithm replaces the missing excitation and MDCT coefficients with an equivalent characteristic from a previously received frame, while the excitation energy gradually decays. In addition, for frame error correction (FEC) it uses a voicing classifier based on the parameter C, which is a signal classification such as voice, unvoice, transition, and silence. During the frame error concealment process for the low band, the gain or the energy parameter, E, is adjusted according to C. Next, the synthesis filter for the lost frame uses the linear predictive coding (LPC) coefficients of the last good frame. Similarly, the pitch period of the lost frame uses the integer part of the pitch period from the previous frame. To avoid becoming desynchronized, the phase parameter, P_{resync}, is used for the recovery after the lost voice onset period.

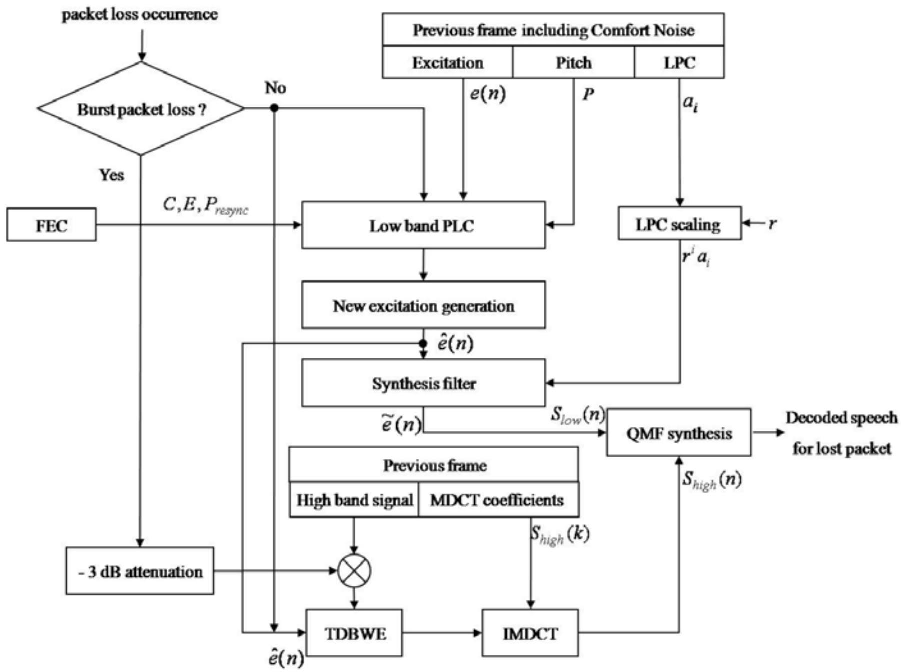

Fig. 1. Overview of the G.729.1 PLC algorithm

During the frame error concealment process for the high band, the high-band signals of the previous good frame are applied to the time domain bandwidth extension (TDBWE) by using the excitation generated by the low-band PLC. Next, the MDCT coefficients of previous good frame are used to generate the high-band signals. Finally, the decoded speech for the lost frame is generated through a quadrature mirror filter (QMF) synthesis, which consists of the 64-tap filter coefficients.

3 Proposed PLC Algorithm

As shown in Fig. 1, according to a 4-kHz boundary frequency, the PLC algorithm for the G.729.1 speech coder consists of low-band and high-band PLCs. Note that the PLC algorithm for regenerating the low-band signals is executed in the time domain, whereas the PLC algorithm for the high-band signals is executed in the frequency domain. Since different PLC algorithms are applied for the low- and high-bands, a frequency mismatch [7] could occur.

Contrary to this conventional PLC algorithm, the proposed algorithm is designed as shown in Fig. 2. In the figure, the PLC of the low-band signal is equivalent to the conventional PLC algorithm. On the other hand, in the high-band PLC, the synthesized low-band signal is transformed into the frequency domain by using MDCT to smooth the boundary frequency between the low- and high-bands. In this case, the weighting filter [8] for smoothing the frequency is given by

$$S'_{high}(k) = 0.6 \cdot S_{high}(k) + 0.4 \cdot S_{avg}, \qquad k = 0,1,\cdots,39 \tag{1}$$

where $S_{low}(k)$, $S_{high}(k)$, and $S'_{high}(k)$ are the low-band signal, high-band signal, and the smoothed high-band signal in the MDCT frequency-domain, respectively, and k is the MDCT frequency bin with a range of 0 to 159. S_{avg} is the frequency average at the boundary frequency (4 kHz) and denoted as $S_{avg} = \frac{1}{2}\left(S_{low}(159) + S_{high}(0)\right)$. After passing through the weighting filter, the smoothed signal is transformed into the time domain by applying an inverse modified discrete cosine transform (IMDCT), as shown in the figure. Finally, the decoded speech for the lost frame is recovered through QMF synthesis, which consists of the 64-tap filter coefficients.

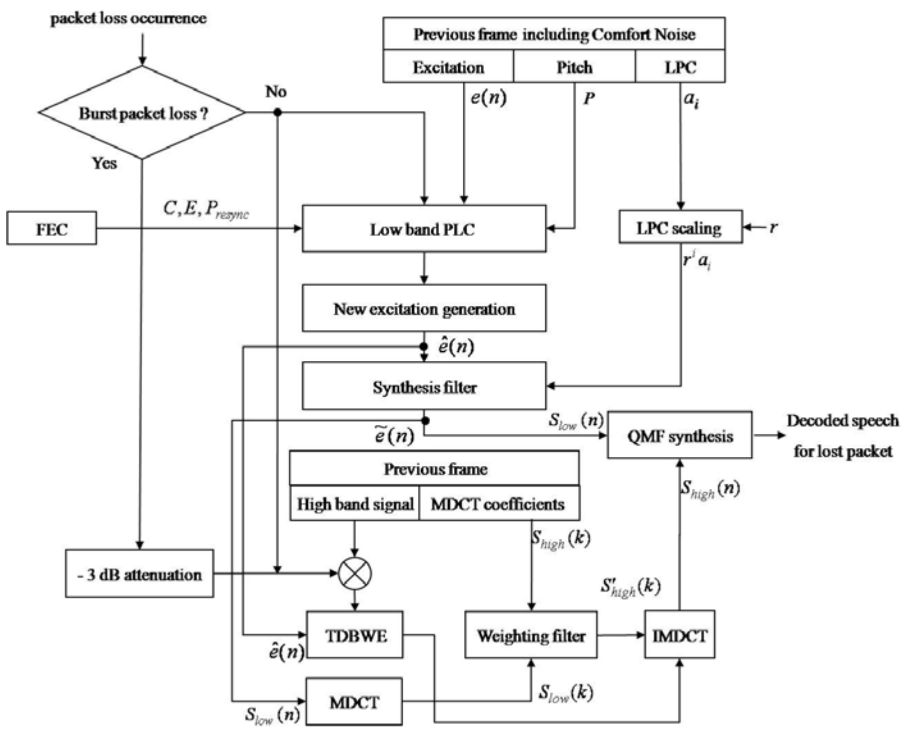

Fig. 2. Overview of the proposed PLC algorithm

4 Performance Evaluation

To evaluate the performance of the proposed PLC algorithm, we replaced the PLC algorithm currently employed in G.729.1 [5] with the proposed PLC algorithm, and then obtained perceptual evaluation of the speech quality (PESQ) scores according to ITU-T Recommendation P.862 [8]. For the PESQ test, 76 audio files were taken from the SQAM audio database [9] and processed by G.729.1 using the proposed PLC algorithm under different packet loss conditions. The performance was also compared to that using the PLC algorithm employed in G.729.1, referred to here as G.729.1-PLC. In this paper, we simulated two different packet loss conditions, including random and burst packet losses. During these simulations, packet loss rates of 3, 5, and 8% were generated by the Gilbert-Elliot model defined in ITU-T Recommendation G.191 [10]. Under the burst packet loss condition, the burstiness of the packet losses was set to 0.66. Thus, the mean and maximum consecutive packet losses were measured at 1.5 and 3.7 frames, respectively.

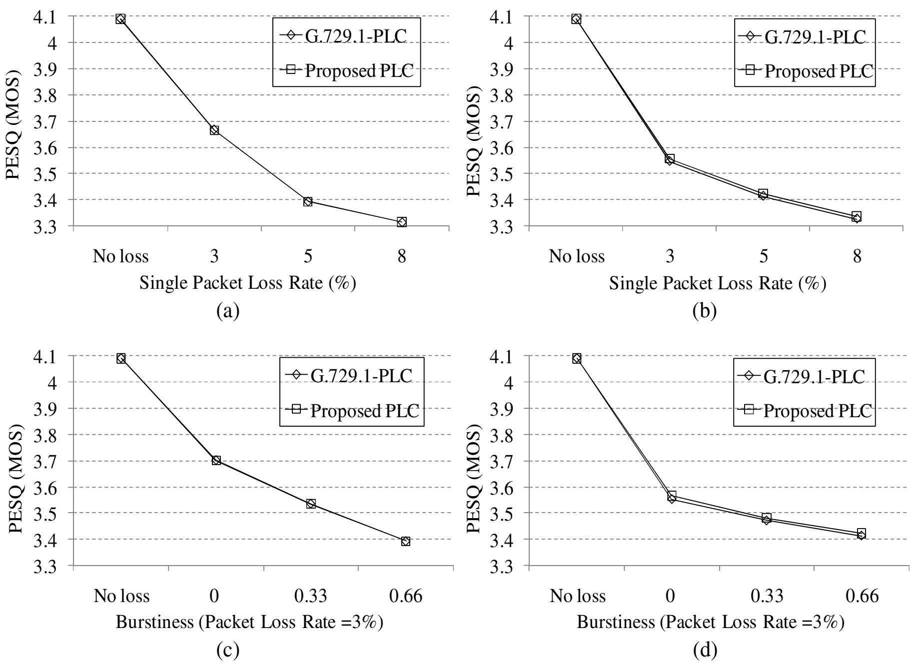

Fig. 3. Comparison of PESQ scores of the proposed PLC and G.729.1-PLC for (a) speech data and (b) music data under single packet loss conditions, and for (c) speech data and (d) music data under burst packet loss conditions

Fig. 3 compares the PESQ scores when the proposed PLC and G.729.1-PLC were employed in G.729.1 under single packet loss conditions and burst packet loss conditions at a packet loss rate of 3%, respectively. It was shown from the figure that the proposed PLC algorithm had PESQ scores comparable to the G.729.1-PLC algorithm for all conditions in the case of the speech data. However, the effectiveness of the proposed PLC algorithm was investigated when packet losses occurred in audio data such as music.

In order to evaluate the subjective performance, we performed an A-B preference listening test, in which 6 speech files (3 males and 3 females) and 2 music files were processed by both G.729.1-PLC and the proposed PLC under random and burst packet loss conditions. Table 1 shows the A-B preference test results. As shown in the table, the proposed PLC was significantly preferred than G.729.1-PLC.

Table 1. A-B preference test results

		Preference Score (%)		
Burstiness/Packet loss rate		G.729.1-PLC	No difference	Proposed PLC
$\gamma = 0.0$ (random)	3%	15.62	46.88	37.50
	5%	12.08	56.67	31.25
	8%	21.88	45.31	32.81
$\gamma = 0.66$	3%	18.75	51.56	29.69
	5%	14.06	54.69	31.25
	8%	15.63	57.81	26.56
Average		16.34	52.15	31.51

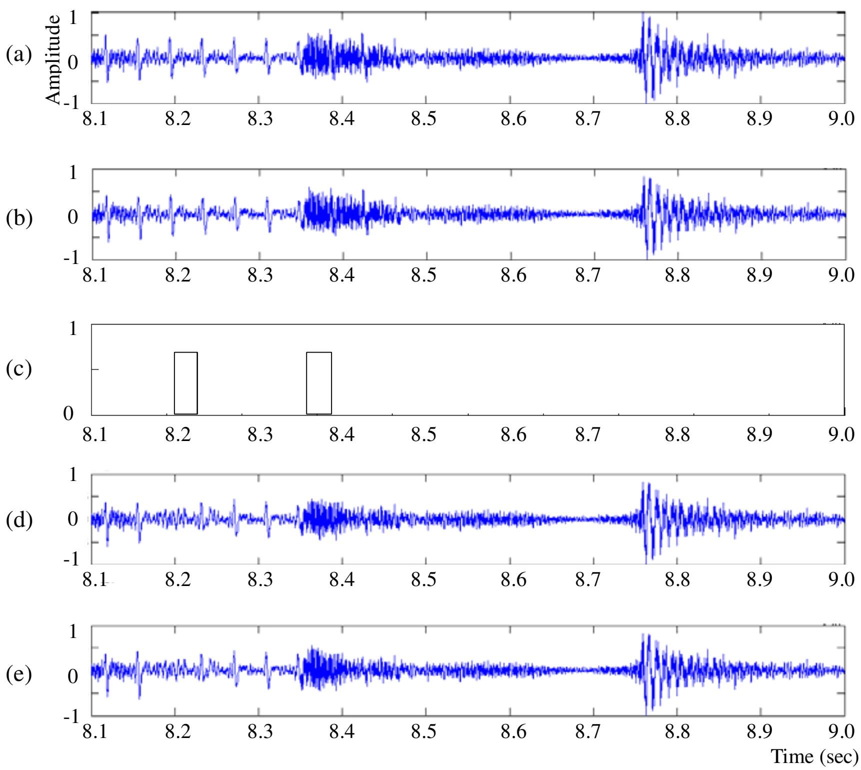

Fig. 4. Waveform comparison: (a) original waveform, (b) decoded speech signal with no packet loss, and reconstructed speech signals using (c) packet error patterns, (d) G.729.1-PLC, and (e) proposed PLC

Finally, Fig. 4 compares the waveform comparison of speech reconstructed by different PLC algorithms. Figs. 4(a) and 4(b) show the original speech waveform and the decoded speech waveform with no loss of the original signal, respectively. After applying the packet error pattern (expressed as a solid box in Fig. 4(c)), the proposed PLC (Fig. 4(e)) reconstructed the speech signals better than in G.729.1-PLC (Fig. 4(d)).

5 Conclusion

In this paper, we proposed a packet loss concealment algorithm for the G.729.1 speech coder to improve the performance of speech quality when frame erasures or packet losses occurred. To this end, we proposed an MDCT-based approach, where MDCT coefficients between the low and high bands were smoothed in order to improve the quality of decoded speech. Next, we evaluated the performance of the proposed PLC algorithm on G.729.1 under random and burst packet loss rates of 3, 5, and 8%, and then compared it with that of the PLC algorithm already employed in G.729.1 (G.729.1-PLC). It was shown from the comparison of PESQ scores, A-B preference, and waveforms that the proposed PLC algorithm provided similar or better speech quality than G.729.1-PLC for all the simulated conditions.

Acknowledgments. This work was supported in part by the "Fusion-Tech Developments for THz Information & Communications" Program of the Gwangju Institute of Science and Technology (GIST) in 2011, by the Mid-Career Researcher Program through an NRF grant funded by MEST, Korea (No. 2010-0000135), and by the Ministry of Knowledge Economy (MKE), Korea, under the Information Technology Research Center (ITRC) support program supervised by the National IT Industry Promotion Agency (NIPA) (NIPA-2010-C1090-1021-0007).

References

1. Goode, B.: Voice over internet protocol (VoIP). Proceedings of the IEEE 90(9), 1495–1517 (2002)
2. Jian, W., Schulzrinne, H.: Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss. In: Proceedings of NOSSDAV, pp. 73–81 (2002)
3. Gournay, P., Rousseau, F., Lefebvre, R.: Improved packet loss recovery using late frames for prediction-based speech coders. In: Proceedings of ICASSP, pp. 108–111 (2003)
4. Tommy, V., Milan, J., Redwan, S., Roch, L.: Efficient frame erasure concealment in predictive speech codecs using glottal pulse resynchronisation. In: Proceedings of ICASSP, pp. 1113–1116 (2007)
5. Rogot, S., Kovesi, B., Trilling, R., Virette, D., Duc, N., Massaloux, D., Proust, S., Geiser, B., Gartner, M., Schandl, S., Taddei, H., Yang, G., Shlomot, E., Ehara, H., Yoshida, K., Vaillancourt, T., Salami, R., Lee, M.S., Kim, D.Y.: ITU-T G.729.1: an 8-32 kbit/s scalable coder interoperable with G.729 for wideband Telephony and voice over IP. In: Proceedings of ICASSP, pp. 529–532 (2007)

6. Taleb, A., Sandgren, P., Johansson, I., Enstrom, D., Bruhn, S.: Partial spectral loss concealment in transform coders. In: Proceedings of ICASSP, pp. 185–188 (2005)
7. ETSI ES 202 050, v1.1.3.: Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Advanced Front-End Feature Extraction Algorithm; Compression Algorithm (2003)
8. 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 (2001)
9. EBU Tech Document 3253: Sound Quality Assessment Material, SQAM (1998)
10. ITU-T Recommendation G.191: Software Tools for Speech and Audio Coding Standardization (2000)