Artificial Bandwidth Extension Based Packet Loss Concealment for CELP-Type Speech Coders

Nam In Park, Jin Ah Kang, and Hong Kook Kim

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

Abstract. In this paper, a packet loss concealment (PLC) algorithm is proposed to improve the quality of decoded speech when packet losses occur in a code-excited linear prediction (CELP)-type speech coder. The proposed PLC algorithm is based on artificial bandwidth extension (ABE) from narrowband to wideband, consisting of packet loss concealment in the narrow-band, ABE in the modified discrete cosine transform (MDCT) domain, and smoothing of wideband MDCT coefficients with those of the last good frame. The effectiveness of the proposed PLC algorithm is demonstrated by an informal listening test.

Keywords: Packet loss concealment (PLC), Artificial bandwidth extension (ABE), Wideband speech coding

1 Introduction

Most speech coders in use today are based on telephone-bandwidth narrowband speech, nominally limited to about 300–3,400 Hz at a sampling rate of 8 kHz. In contrast, wideband speech coders have been developed for the purpose of smoothly migrating from narrowband to wideband quality (50–7,000 Hz) at a sampling rate of 16 kHz. In particular, the ITU-T Recommendation G.729.1 [1], a scalable wideband speech coder, improves the quality of speech by encoding the frequency bands ignored by the narrowband speech coder, ITU-T Recommendation G.729. Therefore, encoding wideband speech using ITU-T G.729.1 is performed via two different approaches that are applied to the low-band and high-band in the time and frequency domains, respectively.

When a frame loss occurs, the low-band and high-band packet loss concealment (PLC) algorithms work separately. In other words, the low-band PLC algorithm reconstructs the excitation and spectral parameters of the lost frame from the last good frame, while the high-band PLC algorithm reconstructs the spectral parameters such as typically modified discrete cosine transform (MDCT) coefficients of the lost frame from the last good frame [2]. However, it is easier to reconstruct the low-band signal than the high-band signal under packet loss conditions. Thus, the quality of reconstructed speech can be improved if the reconstructed signal of the low band is used for a high-band PLC algorithm.

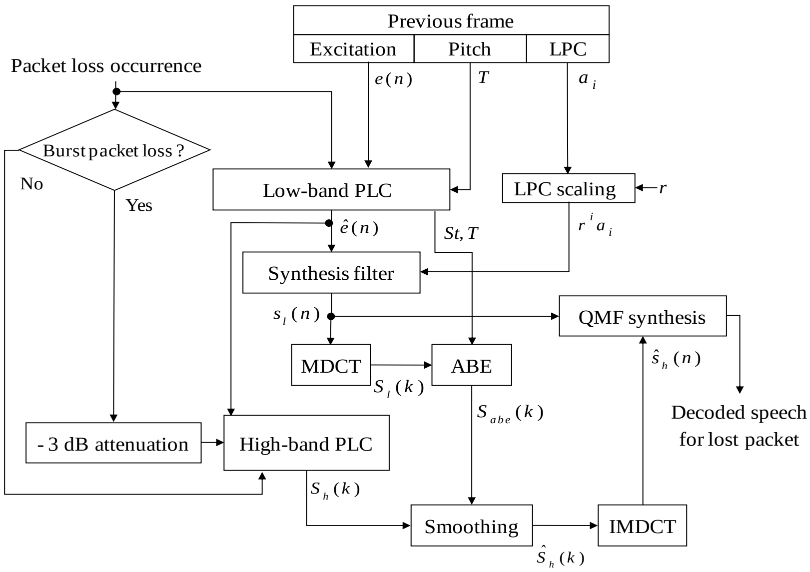

Fig. 1. Block diagram of the proposed ABE-based PLC algorithm.

Therefore, this paper proposes an artificial bandwidth extension (ABE)-based PLC algorithm for high-band signal reconstruction in order to improve the quality of decoded speech under packet loss conditions. The proposed PLC algorithm is mainly composed of three functions including PLC in the narrowband, ABE in the MDCT domain, and smoothing of the wideband MDCT coefficients using those of the last good frame. In particular, the ABE algorithm is performed using two different approaches that are applied to the 4–4.6 kHz and 4.6–7 kHz bands, respectively. That is, the ABE algorithm reconstructs the MDCT coefficients of the 4–4.6 kHz band from the harmonic spectral band replication and correlation-based replication approaches. On the other hand, the MDCT coefficients for the 4.6–7 kHz band are obtained by spectral folding [3]. The performance of the proposed PLC algorithm is subsequently evaluated after being implemented in the ITU-T G.729.1 decoder, and it is compared with that of the PLC algorithm currently used in ITU-T G.729.1.

2 Proposed ABE-Based PLC Algorithm

Fig. 1 shows a block diagram of the proposed ABE-based PLC algorithm. When a frame loss occurs, the proposed PLC algorithm first reconstructs the low-band speech signal of the lost frame, $s_l(n)$, via the low-band PLC module. Simultaneously, the high-band signal is reconstructed by extending the glottal pulse using the high-band spectral envelope of the last good frame, which is denoted as $S_h(k)$. Next, the high-band signal is obtained by applying an ABE algorithm to extend the low-band signal

Table 1. Comparison of A-B preference scores (%) for speech files between the proposed PLC and G.729.1-PLC under random and burst packet loss conditions with different packet loss rates.

Burstiness (r)	Packet Loss Rate (%)	G.729.1-PLC	No Difference	Proposed PLC
0	10	21.43	45.24	33.33
	20	28.57	35.71	35.72
	30	19.05	54.76	26.19
0.99	10	14.29	52.38	33.33
	20	26.19	40.48	33.33
	30	16.67	47.62	35.71
Average		21.03	46.03	32.94

in the MDCT domain, resulting in $S_{abe}(k)$. Subsequently, $\hat{S}_h(k)$ is obtained by smoothing $S_{abe}(k)$ with $S_h(k)$. By applying an inverse MDCT (IMDCT) to $\hat{S}_h(k)$, a time-domain high-band signal, $\hat{s}_h(n)$, is obtained. Finally, $s_l(n)$ and $\hat{s}_h(n)$ are concatenated by using quadrature mirror filterbank (QMF) synthesis.

3 Performance Evaluation

The effectiveness of the proposed ABE-based PLC algorithm is demonstrated by comparing its performance with that of the PLC algorithm currently employed in the ITU-T G.729.1 decoder, which is referred to as G.729.1-PLC. For the comparison, 6 audio files (3 male and 3 female voice files) were prepared from the sound quality assessment material (SQAM) database [4]. Since the files were originally recorded in stereo at a sampling rate of 44.1 kHz, the right channel signal of each file was down-sampled from 44.1 kHz to 16 kHz. In addition, two different packet loss conditions such as random and burst packet losses were simulated. In particular, packet loss rates of 10%, 20%, and 30% were generated by the Gilbert-Elliot model defined in ITU-T Recommendation G.191 [5]. To simulate burst packet loss conditions, the burstiness of the packet losses was set to 0.99, where the mean and maximum consecutive packet losses were measured at 1.9 and 5.6 frames, respectively.

An A-B preference listening test was performed to evaluate the subjective quality. To this end, 6 speech files were processed by both G.729.1-PLC and the proposed PLC algorithm under random and burst packet loss conditions. In this paper, seven people with no auditory disease participated. Table 1 compares the A-B preference test results for the speech data. It was shown in the table that the speech signals decoded by the proposed PLC algorithm were preferred to those by the G.729.1-PLC algorithm.

4 Conclusion

In this paper, a packet loss concealment (PLC) algorithm has been proposed to improve the performance of decoded signal quality when frame erasures or packet losses occurred. The proposed PLC algorithm was based on artificial bandwidth extension from the low-band to the high-band in the MDCT domain. The performance of the proposed PLC algorithm was evaluated by replacing the PLC algorithm currently employed in the ITU-T Recommendation G.729.1 decoder, G.729.1-PLC, It was also compared with that of G.729.1-PLC under random and burst packet loss rates of 10, 20, and 30% by an A-B preference test. It was shown from the test that the proposed PLC algorithm provided better quality of decoded speech signals than G.729.1-PLC for all the simulated packet loss conditions.

Acknowledgments. This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No.2012–0006777), 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-2013-H0301-13-2005).

References

1. Ragot, S., Koves, 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. and 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. Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, Honolulu, HI, 15-20 (2007)
2. Tommy, V., Milan, J., Redwan, S. and Roch, L.: Efficient frame erasure concealment in predictive speech codecs using glottal pulse resynchronisation. Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, Honolulu, HI, 15-20 (2007)
3. Pulakka, H., Laaksonen, L., Vainio, M., Pohjalainen, J. and Alku, P.: Evaluation of an artificial speech bandwidth extension method in three languages. IEEE Transactions on Audio, Speech and Language Processing, 16, 1124-1137 (2008)
4. EBU, Sound Quality Assessment Material Recoding for Subjective Tests, (1998)
5. ITU-T Recommendation G.191, Software Tools for Speech and Audio Coding Standardization, (2000)