An effective hybrid low delay packet loss concealment algorithm for MDCT-based audio codec

Zhibin Lin[1], Jing Lu[1] and Xiaojun Qiu[2]

[1] Key Laboratory of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China

[2] Center for Audio, Acoustics and Vibration, Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia

Abstract: This paper proposes a hybrid packet loss concealment (PLC) algorithm for the MDCT-based audio codec with different PLC strategies on tone dominant source signals and noise like signals respectively. It is meaningful to find that the phase angle of the MDCT-MDST coefficients decreases linearly with the increase of the frame index but the amplitude keeps unchanged for the stationary source signal with dominant tonal components. Therefore an efficient frame interpolation method is designed to accurately estimate the phase angle and the magnitude of the MDCT-MDST coefficients of the lost frame. For the noise-like signals without overwhelming tonal components, a modified shaped-noise insertion is proposed to improve the audio perception. Both objective and subjective test results show that the proposed algorithm provides better performance than the existing ones for both music and voiced speech signals.

Keywords: Packet Loss Concealment; Audio Coding; Discrete Cosine and Sine Transforms; Signal Classification.

1. Introduction

Delivery of broadcasting multimedia such as real-time audio streaming over the internet has become more and more popular in recent years. However, data packets sometimes are delayed or discarded in the network transmissions due to the unwarranted quality of service of the internet delivery. Missing packets create a gap in the streamed audio, which significantly degrades the subjective perception of the audio quality. Thus, speech codecs must implement a Packet Loss Concealment (PLC) technique which conceals these frames losses and reduces the degradation in the synthesized audio signal.

Most PLC algorithms were developed for speech codecs with time-domain predictive coding [1-4]. They concentrate attentions on digital speech transmission and work well for speech audio signals but yield poor results for audio signals with music. In the bibliography, these techniques can be classified into repetition methods, interpolation/extrapolation methods and more sophisticated regeneration methods based on a speech model [5-8]. There are also some PLC algorithms proposed for modified discrete cosine transform (MDCT) based codecs [9-14]. The simplest one of them only uses packet repetition, as suggested in ITU standard G.719 [10]. Its complexity is low, but the performance is poor, especially for tone dominant source signals. A PLC approach based on the gapped-data amplitude and phase estimation (GAPES) was proposed, which provides better concealment quality by converting the MDCT coefficients to DSTFT (Discrete Short-Time Fourier-Transform) or MDCT-MDST (Modified Discrete Sine Transform) coefficients at the cost of higher computational complexity, larger memory requirement, and extra delay [11, 12]. In the Advance Audio Codec (AAC) decoder for the 3rd generation partnership project (3GPP), a shaped-noise insertion method is used to provide moderate audio concealment quality [13]. Although this method is computationally efficient and works well for noise-like signals, it suffers from audible artifacts or reconstruction seams for tone dominant source signals as well as extra delay.

This paper proposes a hybrid PLC algorithm for MDCT-based audio codecs. It is found that the phase angle of the MDCT-MDST coefficients decreases linearly with the increase of the frame index but the amplitude keeps unchanged for the stationary source signal with dominant tonal components. Based on this feature, the phase and magnitude of the MDCT-MDST coefficient of the lost frame can be estimated by taking the real part of the interpolated MDCT-MDST coefficient for tone-dominant spectral bins. For noise-like spectral bins, a shaped-noise insertion is implemented without additional delay. A real-time MDCT-based signal classification method is also proposed so that different strategies can be implemented efficiently on tone-dominant signals and noise-like signal respectively. Compare to the previous work, our technique has three merits: low computational complexity, low delay and better performance for music signals and voiced speech, and satisfactory performance for unvoiced speech signals. Further work is on-going to improve the performance of the unvoiced speech signals.

This paper is organized as follows. In section 2, an analysis of the hybrid PLC algorithm is presented, also

a real time MDCT-based signal classifier is proposed. The framework of the proposed algorithm is outlined at the end of this section. Then in Section 3, the objective and subjective test results are described. The paper is concluded in the last section.

2. The proposed hybrid PLC algorithm

2.1. MDCT-MDST analysis for a tone dominant source signal

The k-th MDCT coefficient of a sine windowed signal sequence in the m-th frame $x_m(n)$ of $2N$ samples is defined as [15]

$$
\begin{aligned}
C_m(k) &= \frac{2}{N}\sum_{n=0}^{2N-1} x_m(n)\sin\left[\left(n+\frac{1}{2}\right)\frac{\pi}{2N}\right]\times\cos\left[\frac{\pi}{N}\left(n+\frac{N+1}{2}\right)\left(k+\frac{1}{2}\right)\right] \\
&= \frac{1}{N}\sum_{n=0}^{2N-1} x_m(n)\left\{\sin\left[(k+1)\frac{n\pi}{N}+\frac{\pi}{2N}(k+1)+\frac{k\pi}{2}+\frac{\pi}{4}\right]-\sin\left[\frac{nk\pi}{N}+\frac{k\pi}{2}+\frac{\pi}{4}+\frac{k\pi}{2N}\right]\right\}
\end{aligned}
\tag{1}
$$

where N is the number of the MDCT bins. For a single stationary sinusoid

$$
x_m(n) = A\sin(n\omega+\varphi_m),
\tag{2}
$$

where A is the amplitude, ω is the frequency and

$$
\varphi_m = \varphi_{m-1}+\omega N,
\tag{3}
$$

is the phase. Substituting Eq. (2) into Eq. (1) yields

$$
\begin{aligned}
C_m(k) = \frac{A}{2N}\Bigg\{ &\sum_{n=0}^{2N-1}\cos\left[n\left(\omega-(k+1)\frac{\pi}{N}\right)+\varphi_m-\left(\frac{\pi}{2N}(k+1)+\frac{k\pi}{2}+\frac{\pi}{4}\right)\right] \\
&-\sum_{n=0}^{2N-1}\cos\left[n\left(\omega+(k+1)\frac{\pi}{N}\right)+\varphi_m+\left(\frac{\pi}{2N}(k+1)+\frac{k\pi}{2}+\frac{\pi}{4}\right)\right] \\
&-\sum_{n=0}^{2N-1}\cos\left[n\left(\omega-k\frac{\pi}{N}\right)+\varphi_m-\left(\frac{k\pi}{2}+\frac{\pi}{4}+\frac{k\pi}{2N}\right)\right] \\
&+\sum_{n=0}^{2N-1}\cos\left[n\left(\omega+k\frac{\pi}{N}\right)+\varphi_m+\left(\frac{k\pi}{2}+\frac{\pi}{4}+\frac{k\pi}{2N}\right)\right]\Bigg\}
\end{aligned}
\tag{4}
$$

By using

$$
\sum_{n=0}^{2N-1}\cos(n\omega+\varphi) = \mathrm{Re}\left\{\frac{e^{j\varphi}(1-e^{j2N\omega})}{1-e^{j\omega}}\right\},
\tag{5}
$$

Equation (4) can be rewritten as

$$C_m(k) = \frac{A}{2N} \text{Re} \left\{ \left(1 - e^{j2N\omega}\right) \left[\frac{e^{j\varphi_m} e^{-j\varphi_1}}{1 - e^{j\left[\omega - (k+1)\frac{\pi}{N}\right]}} - \frac{e^{j\varphi_m} e^{j\varphi_1}}{1 - e^{j\left[\omega + (k+1)\frac{\pi}{N}\right]}} - \frac{e^{j\varphi_m} e^{-j\varphi_2}}{1 - e^{j\left(\omega - \frac{k\pi}{N}\right)}} + \frac{e^{j\varphi_m} e^{j\varphi_2}}{1 - e^{j\left(\omega + \frac{k\pi}{N}\right)}} \right] \right\} , \tag{6}$$

where Re{.} denotes the real part of the complex number, and

$$\varphi_1 = \frac{\pi}{2N}(k+1) + \frac{k\pi}{2} + \frac{\pi}{4}, \tag{7}$$

$$\varphi_2 = \frac{k\pi}{2} + \frac{\pi}{4} + \frac{k\pi}{2N}. \tag{8}$$

It can be seen from Eq. (6) that, for a single stationary sinusoid signal with frequency ω, the most energy concentrates in a central region of the main-lobe of the analysis window [15]

$$I_0 = \left\{ k \left| \min \left\{ \left| \omega - (k+1)\frac{\pi}{N} \right|, \left| \omega - k\frac{\pi}{N} \right| \right\} < W \right. \right\}, \tag{9}$$

where W is the main-lobe width of the analysis window. For simplicity, only the frequency bin $k \in I_0$ is considered, so the second term and the fourth term of Eq. (6) are much smaller than the first and third terms. Eq. (6) becomes

$$C_m(k) \approx \frac{A}{2N} \text{Re} \left\{ \frac{e^{j\varphi_m} e^{-j\varphi_1} \left(1 - e^{j2N\omega}\right)}{1 - e^{j\left[\omega - (k+1)\frac{\pi}{N}\right]}} - \frac{e^{j\varphi_m} e^{-j\varphi_2} \left(1 - e^{j2N\omega}\right)}{1 - e^{j\left(\omega - \frac{k\pi}{N}\right)}} \right\}. \tag{10}$$

In the same way, the k-th MDST coefficient in the m-th frame can be approximately expressed as

$$S_m(k) \approx -\frac{A}{2N} \text{Im} \left\{ \frac{e^{j\varphi_m} e^{-j\varphi_1} \left(1 - e^{j2N\omega}\right)}{1 - e^{j\left[\omega - (k+1)\frac{\pi}{N}\right]}} - \frac{e^{j\varphi_m} e^{-j\varphi_2} \left(1 - e^{j2N\omega}\right)}{1 - e^{j\left(\omega - \frac{k\pi}{N}\right)}} \right\}, \tag{11}$$

where Im{.} denotes the imaginary part of the complex number. The complex MDCT-MDST coefficient can be formed by the MDCT coefficient as the real component and the MDST coefficient as the imaginary component

$$D_m(k) = C_m(k) + jS_m(k) \approx \left[e^{j\varphi_m} A(k) \right]^*, \tag{12}$$

where

$$A(k) = \frac{A\left(1 - e^{j2N\omega}\right)}{2N} \left\{ \frac{e^{-j\varphi_1}}{1 - e^{j\left[\omega - (k+1)\frac{\pi}{N}\right]}} - \frac{e^{-j\varphi_2}}{1 - e^{j\left(\omega - \frac{k\pi}{N}\right)}} \right\}. \tag{13}$$

It should be noted here that $A(k)$ is independent of the frame index m. Using Eqs. (3) and (12), the phase and the amplitude of the MDCT-MDST coefficient can be obtained as

$$\angle D_m(k) = -\varphi_m - \angle A(k) = -\varphi_{m-1} - \angle A(k) - \omega N = \angle D_{m-1}(k) - \omega N ,$$ (14)

$$D_m(k) = |A(k)| .$$ (15)

Equations (14) and (15) show that for the tone dominant signal, the phase of the MDCT-MDST coefficient decreases linearly with the increase of the frame index m while the amplitude keeps unchanged. For example, for the source signal $x(n) = \sin(16.3\pi n / N) + 0.5\sin(30.7\pi n / N)$ and $N = 640$, the phase and the amplitude of MDCT-MDST coefficient at tone-dominant bins are shown in Figs. 1 and 2, respectively. This feature is meaningful and is the basis for the PLC algorithm design as will be shown in Sec. 2.4.
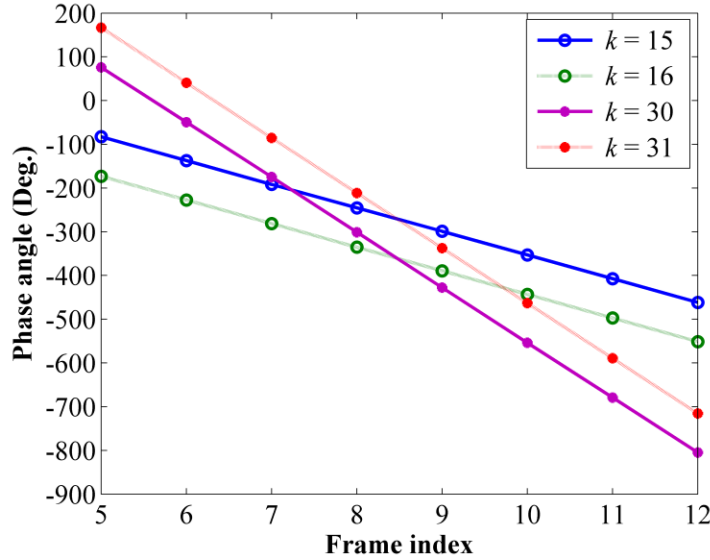
Fig. 1. The phase angle of MDCT-MDST coefficient at k = 15, 16 and 30, 31.
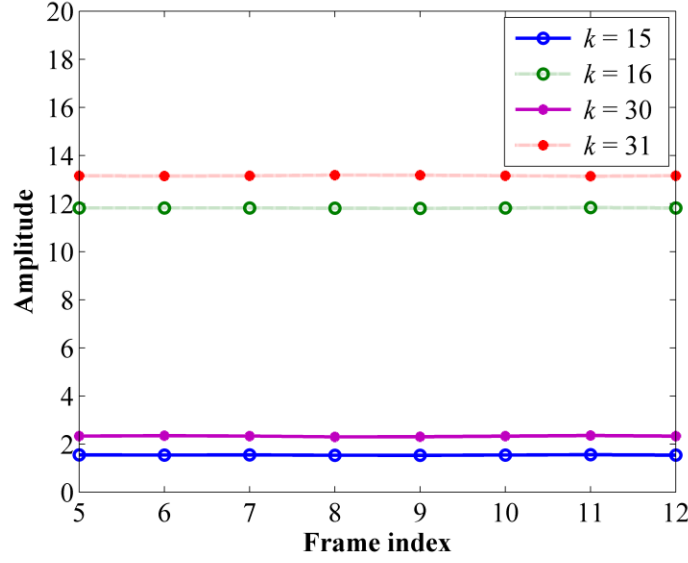
Fig. 2. The amplitude of MDCT-MDST coefficient at $k = 15, 16$ and $30, 31$.

2.2. Modified shaped noise insertion

The packet loss concealment module in the core AAC decoder estimates the MDCT coefficient of the lost frame via shaped-noise insertion method by assuming that the lost frame and its neighboring frames can be well modeled as a noise-like signal [13]. Although this low complexity technique appears to work effectively for noise-like signals, it brings additional delay due to the estimation of the energy scale factor. Furthermore, its performance decreases significantly for continuous packet losses [13, 14]. Since the noise modeling and the psychoacoustic principle that noise-like signals can be parameterized with short time spectral energy envelope, the model parameters of the previous two frames are extracted by computing the spectral energy of the MDCT coefficients in each scale factor band. Because the special energy distribution within the scale factor band is not as psycho-acoustically important as energy distribution across scale factor bands. Then, the spectral energies of the lost frame can be estimated via the band wise energy extrapolation from the previous two frames. The aforementioned modification procedure is proposed in this paper to improve the performance as follows

$$\hat{C}_m(k) = C_{m-2}(k)\tilde{a}(k)\tilde{s}(k) , \qquad (16)$$

where $\hat{C}_m(k)$ is the k-th estimated MDCT coefficient vector of the loss m-th frame, $\tilde{s}(k)$ is a signed random variable whose value can be chosen as negative or positive, and $\tilde{a}(k)$ is the energy scale factor for a scale subband estimated by

$$\tilde{a}(k) = \frac{\sum_{k \in B_r} |C_{m-1}(k)|^2}{\sum_{k \in B_r} |C_{m-2}(k)|^2} ,\qquad (17)$$

where B_r is the set of MDCT bins in the r-th scale factor band containing the k-th bin of interest. In this modified shaped-noise insertion approach, the band-wise energy extrapolation is carried out by taking the root mean square of the band energies of the previous frames, and then $\tilde{a}(k)$ is determined so that the band energy of the MDCT estimated in the loss frame is consistent with the band energy continuity. Unlike that in the original AAC, at least one frame delay is introduced in order to estimate the MDCT coefficient of the loss frame, there is no delay introduced in this modified shaped-noise insertion technique. This improves the performance of the algorithm significantly under continuous packet loss conditions.

2.3. Signal classification

To implement the above proposed algorithm, a classification approach is proposed to classify each frame to either tone dominant or noise like signal. The classifier is based on 5 features of the signal: (1) the zero-crossing rates, (2) the modified harmonic structure stability, (3) the subband energy gradient, (4) the subband energy, and (5) the spectral centroid. The zero-crossing rate is defined as [16]

$$Z_r = \sum_{n=1}^{N-1} \{|sign[x(n)] - sign[x(n-1)]|\} / (N-1) .\qquad (18)$$

The harmonic structure stability is defined by Zhang *et al* [17] and modified in the MDCT domain as follows

$$H_{SS} = \frac{1}{N_H} \sum_{r=1}^{N_H} \sum_{m=1}^{M} (H_r^m - \overline{H}_r)^2 ,\qquad (19)$$

where N_H is the maximal harmonic number, $\overline{H}_r$ is the average harmonic structure of harmonic number r and the harmonic structure coefficient H_r^m of frame index m is defined as

$$H_r^m = \log(\rho^m A_r^m) / \log(\rho^m A_1^m), r = 1, ..., N_H ,\qquad (20)$$

where A_r^m is the amplitude of harmonic number r of frame index m and $\rho^m = const / A_1^m$ (*const* is an arbitrary constant) is a scale factor of frame index m used to eliminate the influence of the amplitude [17].

The subband energy gradient is proposed in this paper as an auxiliary feature to determine the differences of the subband energy over time,

$$\nabla_{SE} = \sum_{m=1}^{M-1} [\log E(m, H_r) - \log E(m-1, H_r)] , \qquad (21)$$

where $E(m, H_r) = \sum_{k \in B_r} |C_{m-1}(k)|^2$ is the root mean square of subband B_r in frame index m, M is the total frame index used in the calculation of the subband energy gradient defined as Eq. (6).

The spectral centroid is the barycenter of the spectrum [18]. It is calculated as the weighted mean of the frequencies presented in the signal, determined using a Fourier transform, with their magnitudes as the weights.
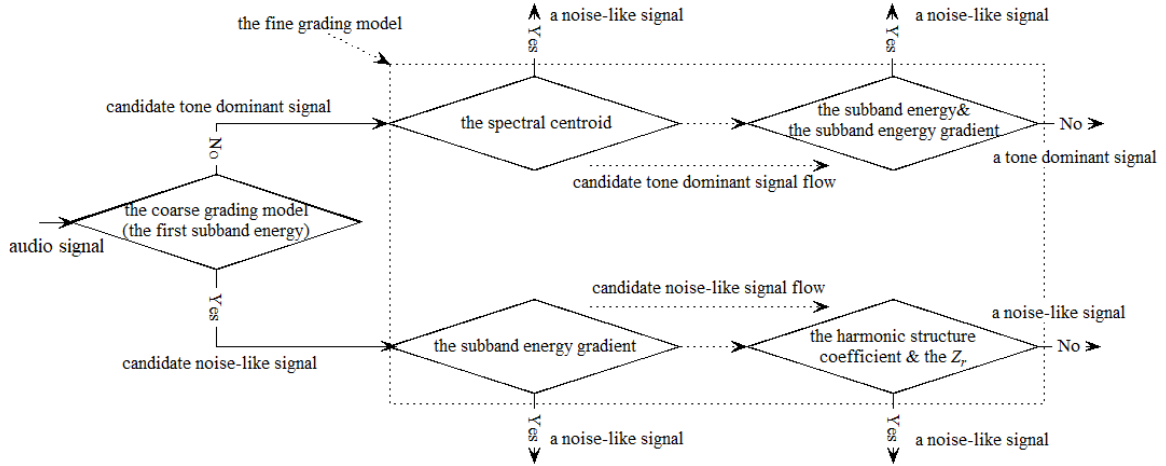

Fig. 3. The schematic diagram of the classification approach.

All 5 features are fed to a frame-level classifier, which is based on a coarse and fine grading model, which is shown in Fig. 3. In the model, the signal is classified initially as either a candidate tone dominant source signal or a candidate noise-like signal roughly according to the feature of the first subband energy. Then, for the candidate tone dominant source signals, the signal type is further classified as either a noise-like

signal or a candidate tone dominant source signal by using the threshold of spectral centroid. On the other hand, for the candidate noise-like signals obtained in the first step, the signal type is classified as either a noise-like signal or a candidate noise-like signal by using the threshold of the subband energy gradient. In the mean time, more features are implemented to make a decision either candidate or certain signal types for the two threads of the uncertain decision in the first step. The features are different for different threads respectively. The proceedings will be converged after five steps for the candidate noise-like signal and six steps for the candidate tone dominant signal. It should be note here that a database containing 100 sentences with tone dominant signals and noise like signals were used as the training database to obtain the thresholds of the aforementioned features.

2.4. The whole proposed PLC method

The block diagram of the proposed PLC method is shown in Fig. 4. With a previous signal type received in the decoder, the coefficients of the loss frames are estimated via the modified shaped-noise insertion module for noise like signals as described in Section 2.2, and then the following procedure is implemented for the tone dominant source signal.
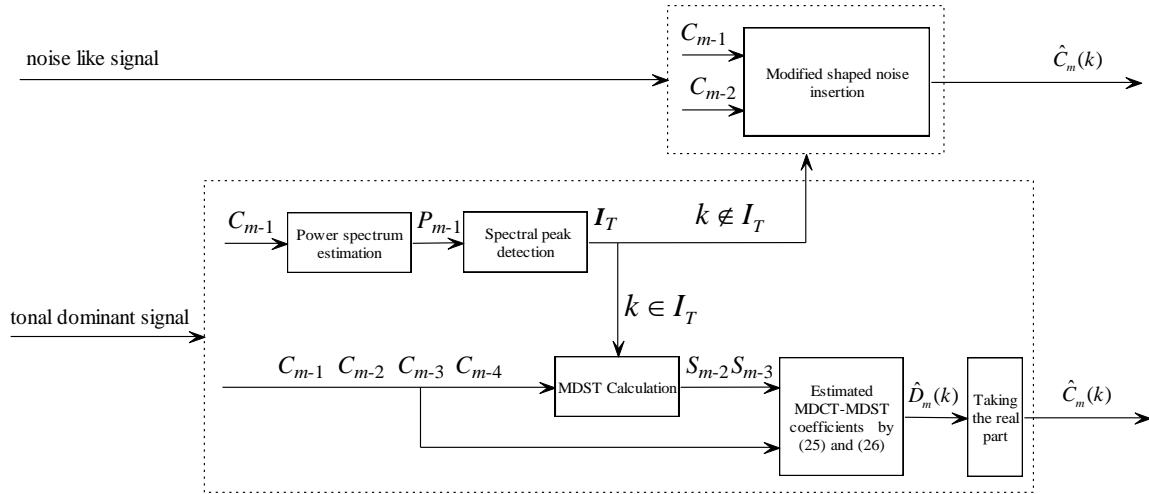
Fig. 4. The proposed hybrid PLC algorithm based on the interpolation technique in the complex spectral domain and the modified shaped-noise insertion.

Suppose that the source signal is modeled with L sinusoids with frequencies ω_l, the integral spectral peak at $k_l = \text{integer}(\omega_l N / \pi)$ can be detected from its power spectrum [19]

$$\hat{P}_{m-1}(k) = C_{m-1}^2(k) + \left[C_{m-1}(k+1) - C_{m-1}(k-1) \right]^2 , \qquad (22)$$

where $\hat{P}_{m-1}(k)$ is the pseudo-spectrum, which is used to regularize the MDCT coefficient. It has been reported that the pseudo-spectrum can be used to smooth the MDCT spectrum and such an approximation is sufficient for peak detection purposes [19].

With all the local maximum peak positions, the index set of the tone dominant source MDCT bins can be written as:

$$I_T = \bigcup_1^L I_l , \qquad (23)$$

where

$$I_l = \left\{ k \, \big| \, |k - k_l| < W \right\} . \qquad (24)$$

Here W is the main-lobe width of the analysis window. For $k \notin I_T$, the MDCT coefficients of the loss frame are estimated by the shaped-noise insertion method. For $k \in I_T$, the following steps are used to estimate the MDCT coefficients of the lost frame,

1) calculate the MDST coefficients $S_{m-3}(k)$ and $S_{m-2}(k)$ from the MDCT coefficients by the method proposed by Ryu *et al* [14],

2) estimate the phase of the MDCT-MDST coefficients by using the interpolation technique,

$$\angle \hat{D}_m(k) = 2 \left[\angle D_{m-2}(k) - \angle D_{m-3}(k) \right] + \angle D_{m-2}(k) = 3 \angle D_{m-2}(k) - 2 \angle D_{m-3}(k) , \qquad (25)$$

3) estimate the amplitude of the MDCT-MDST coefficients

$$\left| \hat{D}_m(k) \right| = \left| D_{m-2}(k) \right| , \qquad (26)$$

4) convert the MDCT-MDST coefficients back to the original MDCT domain

$$\hat{C}_m(k) = \text{Re} \left\{ \left| \hat{D}_m(k) \right| e^{j \angle \hat{D}_m(k)} \right\} . \qquad (27)$$

3. Experiments and discussions

To evaluate the performance of the signal classification, a database containing 100 sentences with tone dominant signals and noise like signals were used as the training database to obtain the thresholds of the aforementioned features in Section 2.3. The classification accuracy achieved for tone dominant signal and noise like signal classification are shown in Table 1. The six tone dominant source signal samples include some vowels and music signals such as horn and piano, and the six noise-like signal samples consisted of consonants including male speech and female speech. It can be seen from the table that the accuracy of the proposed method is larger than 95%.

Table 1 The accuracy rates of the classifier for tone dominant signal and noise like signal.

Sample group	1	2	3	4	5	6
Tone dominant signal (%)	97.6	96.2	96.7	95.8	94.9	94.3
Noise like signal (%)	97.1	97.6	95.0	96.5	96.0	98.4

Subsequently, the performance of the proposed PLC technique is evaluated by comparing with the 3GPP PLC technique (shaped-noise insertion approach) and the packet repetition method used in ITU-T standard G.719 [10, 13]. Both objective measurements and subjective listening tests were carried out to verify the performance of the proposed PLC techniques. The objective assessment tool used to evaluate the result of the concealment techniques is the perceptual estimation of audio quality (PEAQ) standard BS. 1387 developed by the ITU-R [20]. The standard uses a number of psycho-acoustical measures to give the measurement of the quality difference between two signals. The ODG (Objective Difference Grade) score is given in the range from − 4 to 0, which is similar to the standard mean opinion score (MOS) scale. A random loss test was performed at loss rates of 3% and 10%. Four different audio files of different types were used in the tests: piano, horn, voiced speech and unvoiced speech. All are mono-channel signals, 10 seconds long, sampled at 32 kHz and coded at a bit rate of 48 kb/s per channel. Figs. 5 and 6 summarize the average results, where it can be seen that the proposed algorithm performance is superior to both existing methods, especially for tone dominant signal.
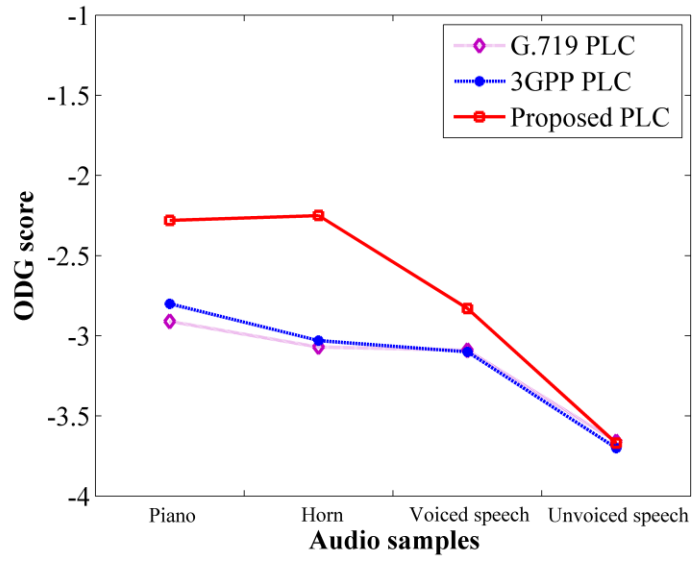
Fig. 5. Average PEAQ ODG scores for 3% random packet loss.
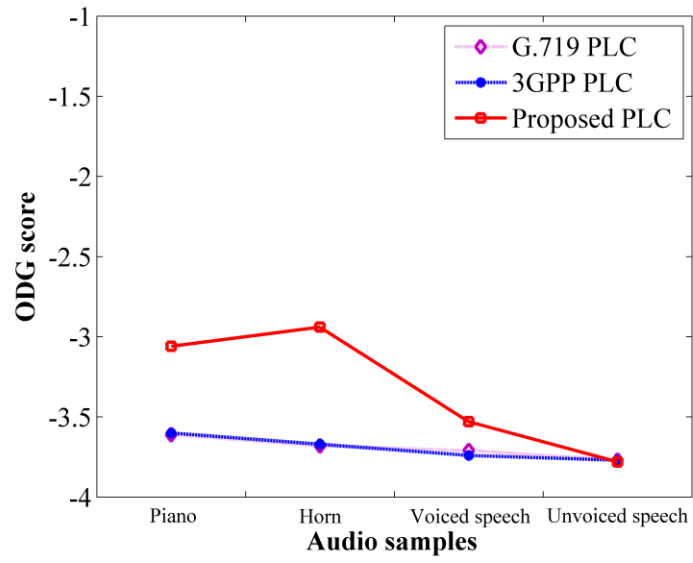
Fig. 6. Average PEAQ ODG scores for 10% random packet loss.

Table 2 Comparative test results of the proposed technique versus 3GPP. The numbers indicate how many

listeners voted in favor of each method.

Sample	Piano		Horn		Voiced speech		Unvoiced speech	
Loss rate	Proposed	3GPP	Proposed	3GPP	Proposed	3GPP	Proposed	3GPP
3%	15	1	16	0	14	2	8	8
10%	15	1	16	0	13	3	7	9

Table 3 Comparative test results of the proposed technique versus G. 719. The numbers indicate how many

listeners voted in favor of each method.

Sample	Piano		Horn		Voiced speech		Unvoiced speech	
Loss rate	Proposed	G. 719	Proposed	G. 719	Proposed	G. 719	Proposed	G. 719
3%	16	0	16	0	13	3	8	8
10%	16	0	16	0	12	4	8	8

Sixteen inexperienced young people with good hearing ability took part in the subjective listening tests. Every listener was asked to distinguish between pairs of audio files, where the packet losses in each file were concealed by a different algorithm. Table 2 shows the votes for the proposed PLC and the 3GPP PLC. On average, for the music signals (Piano, Horn), about 94% of the listeners voted in favor of the proposed PLC algorithm. For speech signals, about 66% of the listeners prefer the proposed algorithm. While compared with the G.719 PLC, the voting results presented in Table 3 confirm that the proposed method is more suitable for the packet loss concealment of the tone dominant signals such as the music signals and voice speech signal. Overall, 93% of the listeners voted for the proposed algorithm, so it can be concluded that the proposed method works effectively for music signals and voiced speech, and obtains satisfactory performance for unvoiced speech signals. As a result, it can be seen from Tables. 2 and 3 that the proposed method is more suitable for stationary signals whereas its performance degrades for noise like signals. On the other hand, for a MDCT-based audio codec, the improvement of the PLC performance of unvoiced speech signal is more complicated, more carefully designed approach might be needed to improve the

performance of the PLC method, such as the LP (Linear Prediction) model for speech signal. More research is on-going to improve the proposed PLC method.

4. Conclusion

In this paper, a hybrid packet loss concealment (PLC) method is proposed for an MDCT-based audio codec by combining different strategies on tone dominant source signal and noise-like signal. The tone dominant source signal PLC method uses the frame interpolation to estimate the phase and magnitude of the MDCT-MDST coefficient of the lost frame based on the observations that the phase of the MDCT-MDST coefficients of tone dominant source signals decreases linearly with the increase of the frame index and the amplitude keeps unchanged. The noise-like signal PLC method implements a modified shaped-noise insertion without introducing any additional delay. Compared with the packet repetition technique of G.719 and the shaped-noise insertion technique of 3GPP, the proposed algorithm works effectively for the tone dominant signals such as the music signals and voiced speech, and obtains satisfactory performance for unvoiced speech signals.

References

[1] D. Lopez-Oller, N. Benamirouche, et al., Speech excitation signal recovering based on a novel error mitigation scheme under erasure channel conditions, Speech Communication, vol. 97, 2018, pp. 73 -80.

[2] B. K. Lee, J. H. Chang, Packet Loss Concealment Based on Deep Neural Networks for Digital Speech Transmission, IEEE-ACM Trans. Audio, Speech, Lang. Process., vol. 24, No. 2, 2016, pp. 378-387.

[3] D. Goodman, G. Lockhart, O. Wasem, Waveform substitution techniques for recovering missing speech segments in packet voice communications, IEEE Trans. Acoust., Speech, Signal Process., vol. 34, 1986, pp. 1440-1448.

[4] M. Elsabrouty, M. Bouchard, T. Aboulnasr, Receiver-based packet loss concealment for pulse code modulation (PCM G.711) coder, Signal Processing, vol. 84, 2004, pp. 663-667.

[5] T. Gueham, F. Merazka, An enhanced insertion packet loss concealment method for voice over IP network services, 40th International Conference on Telecommunications and Signal Processing, 2017, pp. 377–382.

[6] M. Toyoshima, T. Shinamura, Packet loss concealment for VoIP based on pitch waveform replication and linear predictive coding, IEEE Asia Pacific Conference on Circuits and Systems, 2014, pp. 89–92.

[7] J. Chen, Packet loss concealment based on extrapolation of speech waveform, IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, 2009, pp. 4129–4132.

[8] H. Gokhan Ilk, S. Guler, Signal transformation and interpolation based on modified DCT synthesis, Digital Signal Processing, vol. 21, 2011, pp. 756-763.

[9] J. Lindblom, A sinusoidal voice over packet coder tailored for the frame erasure channel, IEEE Trans. Speech, Audio Process., vol. 13, 2005, pp. 787-798.

[10] I. Telecommunication Union, IUT-T Recommendation G.719, Low-complexity full-band audio coding for high-quality conversational applications, International Telecommunication Union, 2008.

[11] H. Ofir, D. Malah, Packet Loss Concealment for Audio Streaming Based on the GAPES Algorithm, in: Proc. 118th AES Conv., Barcelona, Spain, 2005, preprint 6334.

[12] H. Ofir, D. Malah, I. Cohen, Audio packet loss concealment in a combined MDCT-MDST domain, IEEE Signal Processing Letters, vol. 14, 2007, pp. 1032-1035.

[13] 3GPP TS 26.404, Enhanced aacPlus encoder SBR part, 3GPP, June, 2004.

[14] S-U. Ryu, K. Rose, A Frame Loss Concealment Technique for MPEG-AAC, in: Proc. 120th AES Convention, Paris, France, 2006, preprint 6662.

[15] J. P. Princen, A. B. Bradley, Analysis Synthesis Filter Band Design Based on Time Domain Aliasing Cancellation, IEEE Trans. Acoust., Speech, Signal Process., vol. 34, 1986, pp. 1153-1161.

[16] T. F. Quatieri, Discrete Time Speech Signal Processing: Principles and Practice, Prentice hall PTR, 2002.

[17] Y. G. Zhang, C. Zhang, Separation of Voice and Music by Harmonic Structure Stability analysis, Multimedia and Expo, ICME 2005, pp. 562- 565.

[18] K. K. Paliwal, Spectral Subband Centroid Features for Speech Recognition, in: Proc. IEEE ICASSP, vol. 2, 1998, pp. 617-620.

[19] L. Daudet, M. Sandler, MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction, IEEE Trans. Speech, Audio Process., vol. 12, 2004, pp. 302-312.

[20] I. Telecommunication Union, ITU-R recommendation BS. 1387, Method for objective measurements of

perceived audio quality, International Telecommunication Union, 1998.