Audio Error Concealment Based on Wavelet Decomposition and Reconstruction

Fei Xiao, Hexin Chen and Yan Zhao
School of Communication Engineering, Jilin University, Changchun, China
Email: xiaofei200411@163.com, {chx, zhao_y}@jlu.edu.cn

***Abstract*—The algorithm of wavelet decomposition and reconstruction are applied in the audio error concealment method in this paper. When there is lost frame of the audio signal, the correct frames which lie before and after the lost frame are wavelet decomposed firstly. Then the two sets of the wavelet coefficients obtained from the wavelet decomposition are utilized to get the wavelet coefficients for the lost frame. Finally the concealment of the lost frame is completed by the wavelet coefficients reconstruction. Comparing to the performance of the traditional audio concealment algorithm, the proposed method is better for audio frame reconstruction.**

***Index Terms*—error concealment, audio, wavelet, MALLAT, CELP**

I. INTRODUCTION

People living in the environment of a variety of sounds, language conveys information in social communication activities and the music expresses the feelings of the people. So the sound has a dual nature, one is objective reality and the other is subjective feeling of reflection. When the compressed audio signal is lost during the process of storage or transmission, the use of audio error concealment (EC) to deal with the loss of audio signal is necessary. The audio error concealment technique makes use of the short-term stable characteristics of the audio signal, combining with characteristics of human auditory system to cover up audio compression data which has some decoding errors caused by damaged storage media and transmission channel errors in order to improve the audio playback quality.

So far, researchers have put forward many techniques for dealing with the lost audio, such as waveform substitution, which refers to reconstruction of missing packets by substitution of past waveform segments and concludes pattern matching and pitch detection [1][2]. Another type of audio error concealment algorithm constructs audio packets at the receiver which need parameters from the encoder based on CELP to complete the process of concealment [3]. Because this method derives the encoder state from packets surrounding the loss and generate a replacement for the lost packet from that, so this process is complex to implement but can give good results. An error concealment algorithm focus on the reconstruction of the linear predictive coding (LPC) coefficients which represents the short-term spectral information of speech within a frame and preserving them plays a major role in the quality of the reconstructed speech [4].

Based on previous studies, when the short-term stability of the audio signal is invalid or the waveform of the audio signal is non-repetitive, the results of some methods are not satisfactory to reconstruct the loss of the audio signal. Therefore in this paper, an audio error concealment algorithm based on wavelet decomposition and reconstruction is proposed to improve the concealed audio quality.

Corresponding author: Yan Zhao;

II. BASIC THEORY OF WAVELET DECOMPOSITION AND RECONSTRUCTION

MALLAT algorithm which is based on the compactly supported wavelet of DAUBECHIES is used in the proposed method to decompose and reconstruct the audio signal.

A. MALLAT Algorithm

Let $\{V_j\}$ be a given multi-resolution analysis scale space, $f \in V_{J_1}$ (J_1 is a definite integer) is an arbitrary signal, which has the following wavelet decomposition [5].

$$f(t) = A_{J_1} f(t) = A_{J_1+1} f(t) + D_{J_1+1} f(t) \tag{1}$$

where:

$$A_{J_1+1} f(t) = \sum_{m=-\infty}^{\infty} C_{J_1+1,m} \varphi_{J_1+1,m} \tag{2}$$

$$D_{J_1+1} f(t) = \sum_{m=-\infty}^{\infty} d_{J_1+1,m} \psi_{J_1+1,m} \tag{3}$$

Define:

$$H = (H_{m,k}) = (h_{k-2m}), G = (G_{m,k}) = (g_{k-2m}) \tag{4}$$

Then:

$$C_{J_1+1} = HC_{J_1}, \quad d_{J_1+1} = GC_{J_1} \tag{5}$$

doi:10.4304/jsw.7.12.2742-2748

Similarly, there is:

$$f(t) = A_{J_2} f(t) + \sum_{j=J_1+1}^{J_2} D_j f(t) \qquad (6)$$

where

$$C_{j+1} = HC_j, d_{j+1} = GC_j, j = J_1, J_1+1, ..., J_2 \qquad (7)$$

This is the MALLAT pyramid decomposition algorithm, where $A_j f$ is called the continuous approximation of f under the resolution 2^j [5]. The discrete approximation signal C_{j-1} passes through the filter H and obtains the discrete approximation signal C_j under the resolution 2^j, and the discrete detail signal d_j can be obtained by C_{j-1} passing through the filter G under the resolution 2^j. MALLAT wavelet decomposition is shown in Fig. 1. Obviously, the inverse process of the wavelet decomposition is valid. MALLAT reconstruction algorithm is [5]:

$$C_j = H^* C_{j+1} + G^* d_{j+1}, j = J_2 - 1, J_2 - 2, ..., J_1 \qquad (8)$$

where H^* and G^* are the dual operators of H and G respectively [5].

B. Wavelet Decomposition of the Audio Signal

The wavelet decomposition method is applied to the audio signal. According to (7) of MALLAT decomposition, the smooth version (averages) of the audio signal can be got by the role of H and the detail version (details) of the audio signal can be got by the role of G. Then the smooth version is further wavelet transformed to get the smoother version and a more detailed version. Wavelet decomposition process of the audio signal is shown in Fig.2 [6].

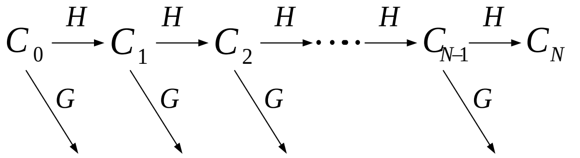

Figure 1. MALLAT decomposition

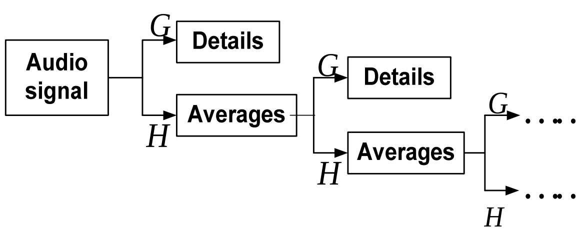

Figure 2. Flow chart for wavelet decomposition of the audio signal

The compactly supported wavelet which is called DAUBECHIES is used to decompose and reconstruct the audio signal. In the finite length FIR filter, DAUBECHIES wavelet function has the greatest regularity, so the waveform of the wavelet is relatively smooth and its time-frequency localization characteristic is better [7]. Let $H(\omega)$ be the Fourier transform of $h(n)$, that is:

$$H(\omega) = \sum_n h(n) e^{-j\omega n} \qquad (9)$$

After determining $H(\omega)$, the scaling function h_n and wavelet coefficients g_n can be obtained, which are defined as [6]:

$$N\varphi(t) = \sqrt{2} \sum_{\mathrm{n=0}}^{\mathrm{2N-1}} h_n \varphi(2t - n) \qquad (10)$$

$$N\psi(t) = \sqrt{2} \sum_{\mathrm{n=0}}^{\mathrm{2N-1}} g_n \varphi(2t - n) \qquad (11)$$

$$g_n = (-1)^n h_{2N-n-1}, n = 0, 1, 2, ..., 2N-1 \qquad (12)$$

Through the above operations, decomposed layers for high-frequency coefficients and the last layer of the low frequency coefficients can be obtained [6].

C. Wavelet Reconstruction of the Audio Signal

The reverse process of wavelet decomposition is the wavelet reconstruction of the audio signal. The decomposed high-frequency coefficients and low-frequency coefficients can be used to reconstruct the audio signal.

III. THE PROPOSED METHOD

A. Overview

The core idea of the proposed algorithm is the using of wavelet function and MALLAT algorithm to perform wavelet decomposition of the audio signals near the lost frame. Then the wavelet coefficients of the audio signals near the lost frame are used to estimate the wavelet coefficients of the lost frame. Finally, the estimated wavelet coefficients are used for wavelet reconstruction to replace the lost signal to complete the audio signal recovery.

Fig.3 shows the overall block diagram of the proposed method.

The audio signal is assumed to have L frames in total. Two cases are considered in our experiments which are unilateral concealment and bilateral concealment. In the unilateral concealment, only old frames before the lost frame are used for concealing the lost frame. In the bilateral concealment, both old and future frames of the lost frame are used for error concealment.

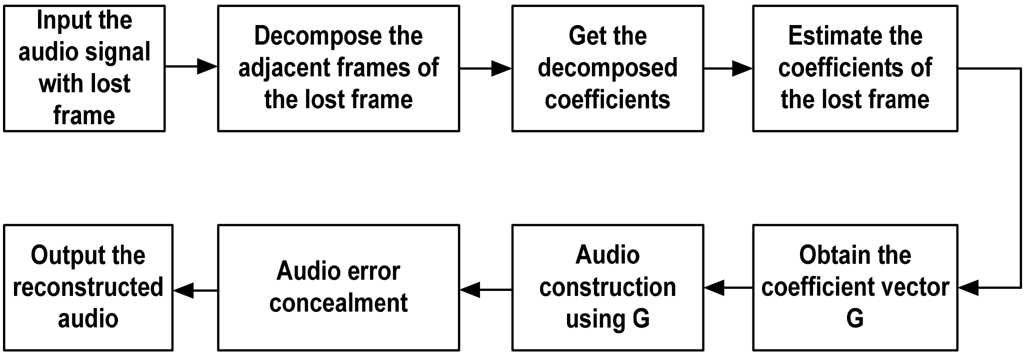

Figure 3. Overall block diagram

B. Unilateral Concealment of Wavelet Coefficients

Assume the number of the lost frame is l, then the $l-1$ frame is decomposed by n layer wavelet decomposition and the corresponding wavelet coefficients, $\vec{g}_{l-1} = (a_{(l-1)n}, d_{(l-1)n}, d_{(l-1)(n-1)}, ..., d_{(l-1)1})$ can be obtained, where $a_{(l-1)n}$ is the smooth coefficient and $d_{(l-1)n}, d_{(l-1)(n-1)}, ..., d_{(l-1)1}$ are the detail coefficients. Similarly, the frames with number from 1 to $l-2$ of the audio signal are decomposed respectively. Then the coefficients of every frame can be got and stored in the matrix C:

$$C = \begin{bmatrix} a_{1n} & d_{1n} & d_{1(n-1)} & \cdots & d_{11} \\ a_{2n} & d_{2n} & d_{2(n-1)} & \cdots & d_{21} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{(l-2)n} & d_{(l-2)n} & d_{(l-2)(n-1)} & \cdots & d_{(l-2)1} \end{bmatrix}_{(l-2)\times(n+1)} .$$

The cross-correlation value $a_i, i = 1, 2, ..., l-2$ between $a_{(l-1)n}$ in $\vec{g}_{l-1}$ and $a_{in}, i = 1, 2, ..., l-2$ obtained in C is calculated, which is further used to constitute a row vector of $\vec{A} = (a_1, a_2, ..., a_{l-3}, a_{l-2})_{1\times(l-2)}$. To find the maximum value in $\vec{A}$ and get the frame number m_1 corresponding to the maximum, $a_{(m_1+1)n}$ is used as the wavelet coefficients of the corresponding lost frame. Same as above, the maximum values of all the detail coefficients $d_{(l-1)n}, d_{(l-1)(n-1)}, ..., d_{(l-1)1}$ are calculated, and in turn the frame numbers $m_2, m_3, ..., m_n, m_{n+1}$ are obtained, then $d_{(m_2+1)n}, d_{(m_3+1)(n-1)}, ..., d_{(m_n+1)2}, d_{(m_{n+1}+1)1}$ are used as the substitute wavelet coefficients of the corresponding lost frame. At last the resulting vector $\vec{g}_l = (a_{(m_1+1)n}, d_{(m_2+1)n}, d_{(m_3+1)(n-1)}, ..., d_{(m_{n+1}+1)1})$ is got. Using this vector as the wavelet coefficients corresponding to the lost frame of the audio signal, we can finally reconstruct the lost signal by wavelet reconstruction, thus complete the error concealment process.

C. Bilateral Concealment of Wavelet Coefficients

In this case, both old and future frames of the lost frame are used conceal the lost frame. For the old frames, the unilateral concealment method is used as above. While for the future frames, the following process is done. The $l+1$ frame are decomposed by n layer wavelet decomposition and the wavelet coefficients $\vec{g}_{l+1} = (a_{(l+1)n}, d_{(l+1)n}, d_{(l+1)(n-1)}, ..., d_{(l+1)1})$ can be obtained, and the frames with number from $l+2$ to L of the audio signal are decomposed respectively, and then the coefficients of every frame can be got and stored in the matrix D:

$$D = \begin{bmatrix} a_{(l+2)n} & d_{(l+2)n} & d_{(l+2)(n-1)} & \cdots d_{(l+2)1} \\ a_{(l+3)n} & d_{(l+3)n} & d_{(l+3)(n-1)} & \cdots d_{(l+3)1} \\ \cdots & \cdots & \cdots & \cdots \\ a_{Ln} & d_{Ln} & d_{L(n-1)} & \cdots\ d_{L1} \end{bmatrix}_{(L-l-1)\times(n+1)} .$$

The cross-correlation value $b_j, j = l+2, l+3, ..., L$ between $a_{(l+1)n}$ in $\vec{g}_{l+1}$ and $a_{jn}, j = l+2, l+3, ..., L$ obtained in D is computed, which is used to constitute a row vector of $\vec{B} = (b_{l+2}, b_{l+3}, ..., b_L)_{1\times(L-l-1)}$. The maximum value in $\vec{B}$ and its corresponding frame number p_1 can be determined. Then $a_{(p_1-1)n}$ is used as the wavelet coefficient of the corresponding lost frame. Similarly, the maximum values of all the detail coefficients $d_{(l+1)n}, d_{(l+1)(n-1)}, ..., d_{(l+1)1}$ are calculated, and in turn the frame numbers $p_2, p_3, ..., p_n, p_{n+1}$ can be obtained. Then $d_{(p_2-1)n}, d_{(p_3-1)(n-1)}, ..., d_{(p_n-1)2}, d_{(p_{n+1}-1)1}$ are used as the wavelet coefficients of the corresponding lost frame. At last the resulting vector $\vec{G}_l = (a_{(p_1-1)n}, d_{(p_2-1)n}, d_{(p_3-1)(n-1)}, ..., d_{(p_{n+1}-1)1})$ can be got, and then $\vec{G} = (2 * \vec{g}_l + \vec{G}_l)/3$ is calculated. Using vector $\vec{G}$ as the wavelet coefficients corresponding to the lost frame of the audio signal, and the lost signal is finally reconstructed by the process of wavelet reconstruction, thus completing the error concealment process.

IV. Experimental Results

In order to evaluate the performance of the proposed method, we test the proposed method in both the waveform coding system and the parameter coding system. Test audio sequences with different features are used in our experiments. The type of wavelet function used in our experiments is DAUBECHIES wavelet.

A. Results in the Waveform Coding

The first part gives the results of the proposed method comparing with the original audio error concealment based on pattern matching [2]. Without loss of generality, we assume a data packet is constituted of a frame signal in our experiments, so the packet loss rate is the frame error rate (FER). The recovery effect of the audio signal is measured by SNR (signal-to-noise ratio) [8].

Firstly, the pure background music signal is used as the test sequence, which sampling frequency is 8kHz. Each frame has 160 sampling values and the length of each frame is 20ms. The method based on pattern matching uses 80 samples for the template. The wavelet decomposition level is 3.

Table 1 gives the SNR results of the concealed audio signal under different FER (Frame Error Rate). From Table 1, we can see that the proposed method has much higher SNR than that of the pattern matching method in [2] for both bilateral concealment and unilateral concealment under all five different FERs. The results in Table 1 also show that the recovery quality of the background music using bilateral concealment is better than that of using unilateral concealment because more adjacent information has been utilized in the bilateral concealment.

Fig. 4 shows the comparison curves of using different methods. It is obvious to see that the concealed audio qualities are all improved under different FER comparing to the pattern matching method in [2].

Secondly, the music is used as the test sequence, which sampling frequency is 16kHz. Each frame has 640 sampling values and the length of each frame is 40ms.

TABLE I.
SNR(dB) for the Pure Background Music

FER (%)	SNR(dB)		
	Bilateral concealment	Unilateral concealment	D.Goodman's Method in [2]
2	24.199	22.439	19.775
5	19.449	18.879	10.611
8	15.838	14.733	9.3641
10	14.478	12.809	9.3382
15	13.238	11.533	6.1407

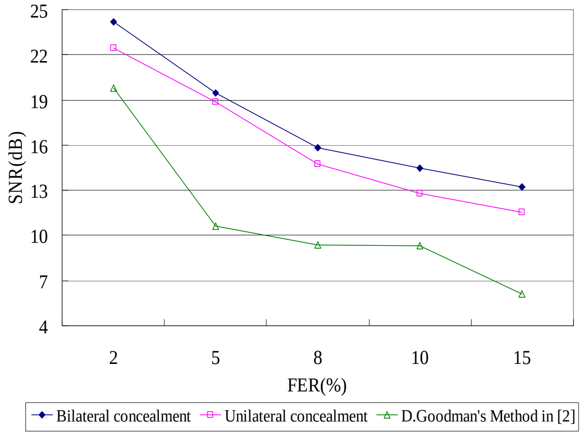

Figure 4. Comparison of different methods in SNR for pure background music

The number of samples in the template is 160. The wavelet decomposition level is 3. The concealment results are shown in Table 2 and Fig.5. From Table 2 and Fig. 5, show that the recovery effect of using the proposed method is also better than that of using the pattern matching method for music under different FERs. Comparing the data in Table 2 with the data in Table 1, we can see also that the recovery results for pure background music are better than those for music due to more correlation existing in the pure background music.

TABLE II
SNR(dB) for the Music

FER (%)	SNR(dB)		
	Bilateral concealment	Unilateral concealment	D.Goodman's Method in [2]
1	17.243	17.183	15.854
2	15.691	15.362	13.942
3	15.086	14.605	13.290
5	12.072	11.746	11.142
10	7.9959	7.4146	7.9177

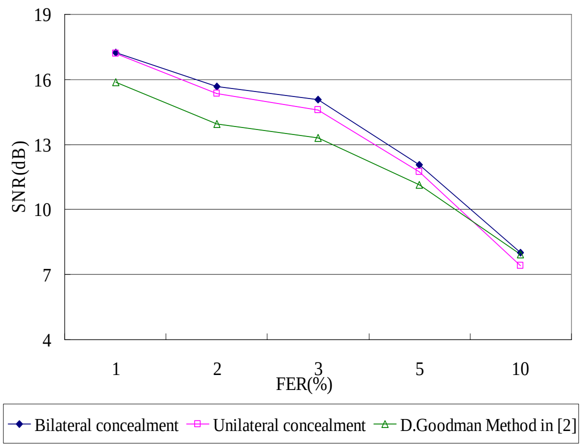

Figure 5. Comparison of different methods in SNR for music

Finally, the speech signal with sampling frequency 16kHz is used as the test sequence. Each frame has 320 sampling values. The length of each frame is 20ms. The number of samples in the template is 160. The wavelet decomposition level is 3. The concealment results are shown in Table 3 and Fig.6.

From Table 3 and Fig.6, we can see that the method in this paper improves the quality of the audio signal on the value of SNR comparing with the method based on pattern matching [2]. And the bilateral concealment is better than the unilateral concealment in most cases.

It is obvious to see in Fig. 6 that the recovery quality of using the proposed method is much improved comparing to the pattern matching method when the FER increases. The test sequence here is the speech signal, therefore there may exist pause when communicating between two persons. The method of pattern matching is to find the matching samples before the lost frame, and the searching range is limited. If the amplitude of the speech signal is weak in the searching range, the matching results will be very poor. While the proposed method makes use of two adjacent frames, which may have strong correlation with the lost frame, to conceal the lost frame to get improved recovery results.

B. Results in the Parameter Coding

The second part of our experiments present the results of the proposed method comparing with the CELP-based audio error concealment technique [3]. The CELP-based audio error concealment techniques use a number of characteristic parameters of the audio signal to recover the lost audio, so restoration of the audio waveform is biased and the recovery results measured by SNR value are not accurate. In this case, PESQ (Perceptual Estimation of Speech Quality) is often used to measure the recovery effect of the audio [9].

Using the background sound signal for the test sequence, which sampling frequency is 16kHz and the length of each frame is 20ms. Table 4 and Fig.7 give the comparison results of the proposed method and M. Chibani's method [3] in PESQ. From Table 4 and Fig.7, we can see that the proposed method has better concealment results than M. Chibani's method under different FER and the bilateral concealment is better than the unilateral concealment in most cases. For the special case, when FER=1% the unilateral concealment is better than the bilateral concealment, because the lost signal is more similar to the previous samples.

TABLE III
SNR(DB) FOR THE SPEECH AS FER CHANGED

FER (%)	SNR(dB)		
	Bilateral concealment	Unilateral concealment	D.Goodman's Method in [2]
1	15.365	15.365	15.251
2	15.153	15.155	14.815
3	15.058	15.066	14.730
5	11.545	11.535	10.070
10	10.989	10.981	9.1442

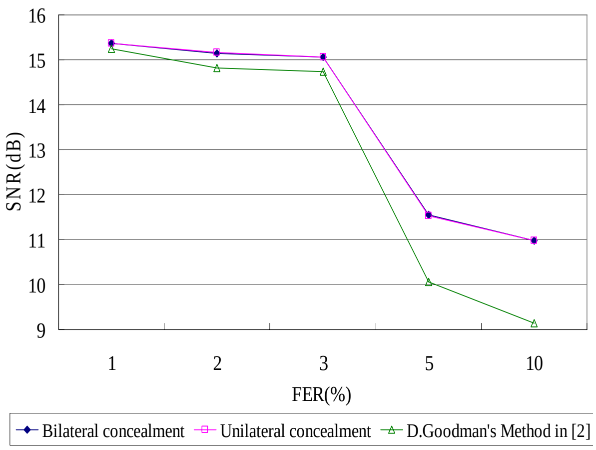

Figure 6. Comparison of different methods in SNR for speech signal

TABLE IV
PESQ UNDER DIFFERENT FER

FER (%)	PESQ		
	Bilateral concealment	Unilateral concealment	M.Chibani's Method in [3]
1	4.1581	4.1807	3.6821
5	4.1513	3.2195	2.9083
10	3.2072	3.1100	2.8060
15	3.0433	2.9181	2.7699
20	3.0323	2.9120	2.7664

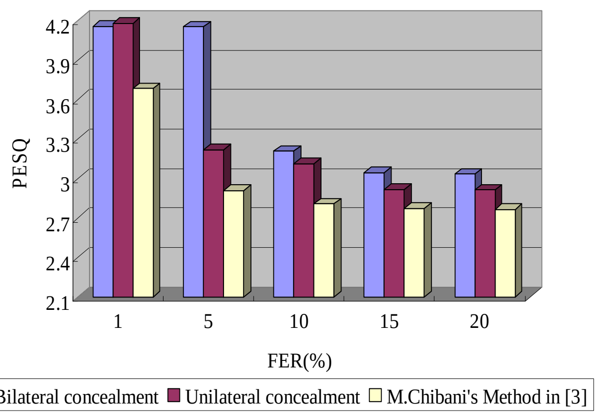

Figure 7. Comparison of different methods in PESQ

C. Results of Using Different Decomposition Level

The above two parts give the experimental results of audio error concealment algorithm based on wavelet decomposition and reconstruction comparing with the existing audio error concealment methods. The experiments are done with fixed wavelet decomposition level. Wavelet decomposition level, as a factor, is going to affect the recovery of the audio quality. Therefore selecting a suitable Wavelet decomposition level is important. We choose the Wavelet decomposition level n according to experimental results.

Take the background sound signal for the test sequence, which sampling frequency is 16kHz and the length of each frame is 20ms. When the FER is 5%, SNR values vary for different n. The specific results are shown in Table 5 and Fig.8. Experimental data in Table 5 and the curves in Fig.8 illustrate that the maximum SNR value is got when the level n is 7. However, the larger the level is, the higher the computational complexity will be. Furthermore the SNR varies very little for different decomposition level n. Therefore, the best choice of the decomposition level is 3 considering both reconstruction quality and complexity.

V. CONCLUSIONS

In this paper, we propose an audio error concealment algorithm based on the wavelet decomposition and reconstruction. This algorithm includes two cases which are unilateral concealment and bilateral concealment respectively.

The proposed algorithm uses the information from the old frames or both the old and future frames of the lost frame to conceal the lost frame. Comparing with the traditional audio error concealment algorithms, it can get more related information about the lost frame. In addition, the proposed algorithm can be used in both the waveform coding system and the parameter coding system. Therefore it can be widely used.

The experimental results show that the proposed algorithm works better to recover the lost audio signal than the traditional methods. The value of SNR and PESQ of the reconstructed audio signal are improved. Experimental results also show that the bilateral treatment in the wavelet coefficients to restore the lost audio signal is better than the unilateral treatment in most cases. However, the bilateral concealment method may bring a little delay which is not suitable for strict real-time application, such as interactive audio.

In general the algorithm proposed in this paper has reached the expected recovery effect of the audio signal.

TABLE V
SNR(dB) WITH DIFFERENT n (FER=5%)

n	SNR(dB)	
	Bilateral concealment	Unilateral concealment
3	19.449	18.879
5	19.359	18.709
7	19.491	18.887
9	19.335	18.497
11	19.339	18.497

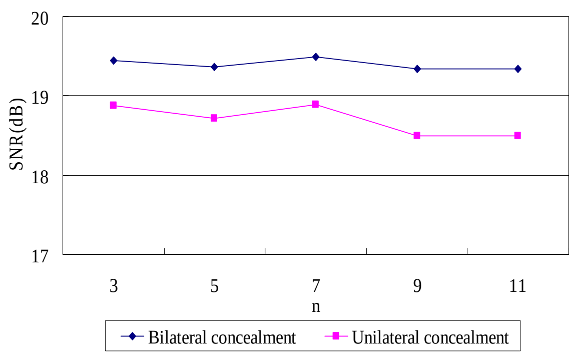

Figure 8. Comparison of using different n

ACKNOWLEDGMENT

This work was supported by the project of National Natural Science Foundation of China under Grant 60832002, 61171078 and in part by the Research Fund for Doctorial Program of Higher Education of China under Grant 20110061110084 and the Outstanding Youth Foundation of Jilin University under Grant 200905018.

REFERENCES

[1] O.J.Wasem, D.J.Goodman, C.A.Dvorak and H. G. Page, "The effect of waveform substitution on the quality of PCM packet communications". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.36, no. 3, pp. 342-348, Mar. 1988.

[2] D. Goodman, G.Lockhart and W.C.WONG, "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.34, pp.1440, Dec. 1986.

[3] Mohanmd Chibani, Roch Lefebver, Philippe Gourrnay, "Fast Recovery for a CELP-Like Speech Codec After a Frame Erasure", *IEEE Transactions on Audio, Speech, and Language Processing*, vol.15, no.8, pp.2485-2495, 2007

[4] Farshad, Lahouti, K.Khandani, "Soft Reconstruction of Speech in the Presence of Noise and Packet Loss", IEEE Transactions on Audio, Speech and Language Processing, vol.15, pp.44, Jan. 2007.

[5] Dongmei He, Wen Gao, "Complexity Scalability Audio Coding Algorithm Based on Wavelet Packet Decomposition", *IEEE International Conference on Signal Processings*, pp.659, 2000.

[6] Pramila Srinivasan, Leah H.Jamieson, "High-Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling", *IEEE Transaction On Signal Processing*, vol.46, pp.1085, Apr.1998.

[7] Fang Chen, Wei Li, Xiaoqiang Li., "Audio quality-based authentication using wavelet packet decomposition and best tree selection", *International Conference on*

Intelligent Information Hiding and Multimedia Signal Processing, pp.1265, August.2008.
[8] Nurgun Erdol, Claude Castelluccia and Ali Zilouchian, “Recovery of Missing Speech Packets Using the Short-Time Energy and Zero-Crossing Measurements”, *IEEE Transactions on Speech and Audio Processing*, vol.1, pp.295, Jul.1993.
[9] Antony W. Rix, John G.beerends, Michanel P. Hollier, et al., “Perceptual evaluation of speech quality(PESQ)-a new method for speech quality assessment of telephone networks and codecs”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.749-752, May 2001.

Fei Xiao was born in Ha’erbin, China, in 1985. She received the B.S. degree from Jilin University in 2008. She is now pursuing her M.S degree in School of Communication Engineering, Jilin University, China.

Her research interests include audio signal processing, error concealment and digital video processing.

Hexin Chen was born in Jilin, China, in 1949. He received the M.S. and Ph.D. degrees in communication and electronic in 1982 and 1990 from Jilin University of Technology, respectively.

He has been a visiting scholar in the University of Alberta from 1987 to 1988. From Feb.1993 to Aug.1993, he was a visiting professor in Tampere University of Technology in Finland. He currently is a professor of communication engineering. His research interests include image and video coding, multidimensional signal processing, image and video retrieval and audio and video synchronization.

Yan Zhao was born in Jilin, China, in 1971. She received the B.S. degree in communication engineering in 1993 from Changchun Institute of Posts and Telecommunications, the M.S. degree in communication and electronic in 1999 from Jilin University of Technology, and the Ph.D. degree in communication and information system in 2003 from Jilin University.

She has been a postdoc researcher in the Digital Media Institute of Tampere University of Technology in Finland from Mar.2003 to Dec.2003. From Mar. 2008 to Aug.2008, she was a visiting professor in the Institute of Communications and Radio-Frequency Engineering in the Vienna University of Technology. She currently is an associate professor of communication engineering. Her research interests include image and video coding, multimedia signal processing and err-or concealment for audio and video transmitted over unreliable networks.

Dr. Zhao is a member of IEEE.