Bilateral Waveform Similarity Overlap-and-Add Based Packet Loss Concealment for Voice over IP

J.F. Yeh*[1], P.C. Lin[2], M.D. Kuo[1,3], Z.H. Hsu[1]

[1] Department of Computer Science and Information Engineering,
National Chiayi University,
Taiwan (R.O.C.)
*ralph@mail.ncyu.edu.tw
[2] Department of Computer Science and Information Engineering,
[3] Department of Digital Design and Management,
Far East University,
Taiwan (R.O.C.)

ABSTRACT
This paper invested a bilateral waveform similarity overlap-and-add algorithm for voice packet lost. Since Packet lost will cause the semantic misunderstanding, it has become one of the most essential problems in speech communication. This investment is based on waveform similarity measure using overlap-and-Add algorithm and provides the bilateral information to enhance the speech signal reconstruction. Traditionally, it has been improved that waveform similarity overlap-and-add (WSOLA) technique is an effective algorithm to deal with packet loss concealment (PLC) for real-time time communication. WSOLA algorithm is widely applied to deal with the length adaptation and packet loss concealment of speech signal. Time scale modification of audio signal is one of the most essential research topics in data communication, especially in voice of IP (VoIP). Herein, the proposed the bilateral WSOLA (BWSOLA) that is derived from WSOLA. Instead of only exploitation one direction speech data, the proposed method will reconstruct the lost voice data according to the preceding and cascading data. The related algorithms have been developed to achieve the optimal reconstructing estimation. The experimental results show that the quality of the reconstructed speech signal of the bilateral WSOLA is much better compared to the standard WSOLA and GWSOLA on different packet loss rate and length using the metrics PESQ and MOS. The significant improvement is obtained by bilateral information and proposed method. The proposed bilateral waveform similarity overlap-and-add (BWSOLA) outperforms the traditional approaches especially in the long duration data loss.

Keywords: Packet loss concealment, waveform similarity overlap-and-add, VoIP, speech communication.

1. Introduction

Since the mobile computing has become one of most frequently used communications, distance transmission is increasing essential for human social activities by voice over IP. Compared to other animals, speech communication has become one of most natural and fluent for human. Instead of the traditional voice communication, internet telephony, also known as VoIP refers to transmitting voice data in data networks, is increasing complied with popularization of 3G communications of hand-held devices especially in smart phone. However, there are some situations such as roaming, moving and communication in wireless vehicular environment will decrease the reliability of data transmissions, Mantilla-Caeiros et al. proposed a pattern recognition based esophageal speech enhancement system to deal with the problem resulted from unclear speech communication [1].Considering of binaural speech intelligibility cross-correlation, Padilla-Ortíz, and Orduña-Bustamante proposed the noise reduction method [2]. Voice packets are lost frequently in unreliable communication. Therefore, packet loss will reduce service quality of VoIP applications. Due to the unreliability of the networks, voice data is possible to be lost during transmissions, resulting in a quality decreasing of VOIP application especially in using the handheld devices in wireless networking based on User datagram protocol (UDP). Furthermore, speech packet loss is more frequent during inter-vehicle Communication (IVC) resulted from the vehicle's mobility [3]. According to analysis provided by Angkititrakul et al., the speech packet loss rate is

from 13.78 to 20.70 percentages when the speech data length is during 10 to 30 ms [4]. To relieve the effects of packet loss, packet loss concealments (PLCs) is proposed in the latest decades. It makes the information lost and uncomfortable for the receiver, it may lead to misunderstanding caused by packet lost from the view point of speech recognition.

To relieve the effects of packet loss, the approaches for packet loss concealment (PLC) are proposed. Since 1998, C. Perkins et al. had much ink spent in survey of the methods about packet loss concealment [5]. The PLC systems can be divided into two categories: sender-based and receiver-based approaches. Sender-based methods focus on reducing the packet transmission failure and improved the distribution of packet loss to split the larger block into several smaller blocks. By this, the difficulty of reconstruction will reduced significantly. Receiver-based techniques concentrate on improving the quality of reconstructed speech. Receiver-based PLC methods recover the speech signal content of the lost packet from its adjacent packets or surrounding information. There are two useful methods: one repeated the frames with the same pitch period to reconstruct the voice data for lost packet as illustrated in [6]. Another one extended the frames before the lost one to cover the instantaneous discontinuities caused by the missed packet in the received speech [4]. These methods usually based on time scale modification (TSM) of audio signals, which is the process of modifying the duration of the signal not transforming other qualities such as the timbre and pitch frequency [7]. The reason to use the time domain methods mainly is the real time applications such VoIP. Recently, time domain pitch synchronous overlap-add (TD-PSOLA) and waveform similarity overlap-and-add (WSOLA) technique are used as effective algorithms in time scale modification. TD-PSOLA is widely used in speech synthesis to adapt the length and quality of speech signal. Pitch related information is essential to the packet loss concealment. However, pitch tracking is very hard in various environments and speakers especially in spontaneous speech [8]. Yeh and Yan proposed a method for word fragment detection by prosody features for spontaneous speech [9]. WSOLA algorithm is applied to the packet loss concealment by reconstructing the frames to cover the lost packet as illustrated in Figure 1. Wu et al. (2009) proposed an enhanced gain control approach based on WSOLA to solve the problem of standard WSOLA, lack of efficient amplitude controls [10]. Ilk and Güler (2006) invested a time scale modification of speech for graceful degrading voice quality by adaptive WSOLA [11]. Duration modification is one of the most important factors in speech communication [12]]. Ito and Nagano proposed an packet loss concealment to solve the voice transmission by G.729 [13]. Jagla et al. presented an algorithm for the real time synthesis of internal combustion engine noise to deal with the problems resulted from packet loss [14]. The foundation about the WSOLA is shown in Figure 1. The packet is lost and the WSOLA algorithm is applied to extend speech signal content of the received packets before the lost packet to concealment the gap of the lost packet by duplicating the signal of previous speech packet to that of lost packet. However, the WSOLA algorithm sometimes may be lead to the amplitude of reconstruction voice signal is not consistent with the original voice signal. To solve this problem, Li et al. proposed the GWSOLA, which introduced the gain control into the WSOLA algorithm [10][15-17]. Sun et al used a modified weighted overlap and add-based by spectral subtraction [18]. However, most of the distortion in the packet loss concealment (PLC) output signal is due to misalignment between reconstruction waveform and the decoded waveform [19]. It is requires not only speech signal content before lost packet but also future speech signal content to obtain the consistent and gain control reconstruction. Kondo and Nakagawa used the linear prediction method to speech packet loss concealment and achieved good results [20]. Linenberg et al. treated the packet loss problem by adaptive lattice modelling [21]. Wang et al. applied the speech synthesis prediction skills to speech packet loss recovery and achieve significant improvement [22]. Yeh and Hsu invested an approach based on spectral block clustering transformation functions to measure the similarity of speech segments [23].

In this paper, we proposed the bilateral WSOLA (BWSOLA), which based on WSOLA and it can use both the forward and afterward speech signal waveform to reconstruction voice waveform of lost packet. Because the bilateral WSOLA using both the forward and afterward speech information, the misalignment between reconstruction waveform and the original waveform can be solve. Here, we further abbreviate bilateral WSOLA as BWSOLA. Therefore this method can maintain consistency of amplitude, frequency and phase between

reconstructed voice and adjacent voice, thus resulting in noticeable audio quality improvement. Mantilla-Caeiros et al. used the codebook with formant, pitch and zero crossing rate information to speech enhancement system [24].

The remainder of this paper is organized as follows. Section 2 describes the framework and basic concepts about bilateral WSOLA. The proposed method is illustrated in section 3, we also describes the reconstruction techniques of the bilateral WSOLA in detail. In section 4, the proposed method is evaluated according to Mean Opinion Score (MOS) measure and signal quality according to the experimental results. Some findings and future works are illustrated in the conclusions in section 5.

2. Framework of bilateral WSOLA

According to the related works described in the previous section, to enhance the WSOLA for improving the packet loss concealment is with potential and necessary. Herein, Figure 2 illustrates a framework of the proposed bilateral WSOLA. The received original data is first fed in lost packet detection module. By checking the serial number of data gram, we can find the speech packets are lost or not. The length of loss data is also obtained by the serial number of received data gram. When packet lost is detected, both the forward and afterward speech signals contents are stored in data buffer; they are used to reconstruct the lost speech data. Since there are different strategies for voice speech data and unvoiced speech data, the forward and afterward speech signal contents are determined as voiced or unvoiced in voiced detection module [25].

Herein, we utilize autocorrelation function (ACF) to decision the speech signal is voiced or not by pitch tracking. Since there are different strategies for voiced speech segment and unvoiced speech segment, we must decide the speech signal is voiced or not. According to the results of voiced detection of the forward and afterward speech signals, the reconstruction strategies of lost speech data can be categorized into four categories: 1) BV where the both forward and afterward speech signals are voiced, 2) PV where the forward speech signals is voiced and the afterward speech signals is unvoiced, 3) NV where the forward speech signals is unvoiced and the afterward speech signals is voiced, 4) BU where both the forward and afterward speech signals are unvoiced. Final, according to result of voiced detection, there are different strategies to reconstruction speech signal of lost packet using the BWSOLA algorithm.

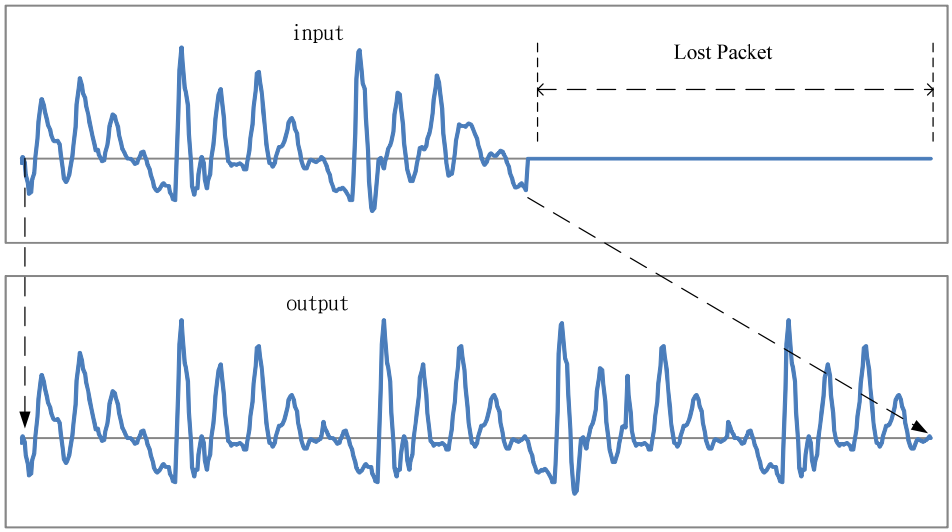

Figure 1. Packet loss concealment using waveform similarity overlap-and-add (WSOLA).

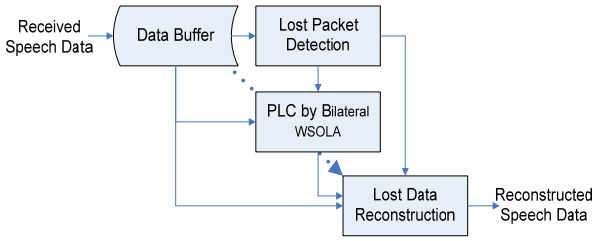

Figure 2. Framework of the proposed bilateral WSOLA.

3. Reconstruction of lost frame using WSOLA

Since unbiased signal reconstruction is very essential for proposed approach in this paper, we must first decide which strategy used. Similar to the method proposed by Guo and Chen in [26]. Figure 3 shows the detail processing for reconstruction for lost frame using the proposed method, bilateral WSOLA (BWSOLA). The input speech signals are extended according to the around speech signals, previous and cascading signals are both included, by the waveform similarity overlap-and-add (WSOLA). Herein, the around speech signals means forward and afterward speech signals is stored in the data buffers: L_WSOLA and R_WSOLA respectively such as shown in Figure 4. According to results of voiced detection, it further chooses one of BV, PV, NV and BU algorithms to reconstruction speech signal. Here, we describe the bilateral WSOLA (BWSOLA) as follows.

3.1 Bilateral Waveform similarity overlap-and-add

As illustrated in Figure 4, when the packets of the input speech signal are lost, the WSOLA is applied to extend the previous two packets speech signal to three packets speech signal for the lost packet that is should be numbered as three. In this way, the extended speech signal can be covered the gap of the lost packet, and the quality of the reconstruction speech signal won't decrease much for human hearing.

In extending process, the bilateral WSOLA (BWSOLA) extracts some voice segments form the packets before the lost one. The length of voice segment is given by following equation.

$$L = \frac{ml}{n} \tag{1}$$

where L denotes length of voice segment, l is length of one packet, m is number of packets after extended, n is number of packets before extended. At the example of Figure 4, the length of one packet l is 256 samples, the m is three and the n is two, so L equals to 384. Then the voice segments extracted is overlapped and added with Δy, in which $\Delta y = L/2$. We search the voice segment which should be extracted through maximum cross-correlation result. The cross-correlation can be representing the distance between the original signal and the time-scaled signal. The procedure is given by following equation.

$$C = \sum_{j=0}^{L-1} X(j+R)X(j+ola) \tag{2}$$

where C is cross-correlation value, X is the original signal, *ola* is start point of overlap-and-add voice segment, R is search region of the voice segment. The each point in search region is candidate for start point of voice segment. Assume each start point of voice segments extracted is end point of search region, the end point of final voice segments extracted is equal to the a point before the lost packet. The end point of search region is given by following equation.

$$nl = (K-1)r + L \tag{3}$$

where K is search step, r is end point of search region. The K is given by following equation.

$$K = 2(m-1) - 1 \tag{4}$$

The search region is defined as the eq. (5)

$$R_k = rk, rk-1, ..., rk-Ls \tag{5}$$

where L_S means the length of search region. The best start point of voice segment r_k' is given by following equation.

$$r_k' = \arg\max_R \sum_{j=0}^{L-1} X(j+R_k)X(j+ola_k) \tag{6}$$

The voice segment S_k is defined as the eq. (7).

$$S_k(n) = X(r_k' + n) \quad n = 0,\ 1,...,\ L-1 \tag{7}$$

The speech signals $Y(n)$ after extend by WSOLA is given by following equation.

$$Y(n) = \sum_{k=0}^{K-1} S_k(n + ola_k) \tag{8}$$

where *Y* denotes speech signal extended. It is derived from the received signal.

3.2 The methodology for BV

Both the forward and afterward speech signal is voiced, the gap caused by missing packet can be seen as process of the pitch period transforms. If only use speech signal of one direction to reconstruction waveform of lost may be resulting in misalignment between reconstruction waveform and the original waveform. So we use both the forward and afterward speech signal. In process of reconstructing, we extend the forward and afterward speech signal by BWSOLA, and then making some adjustments for gap of the lost packet. In order to avoid problem of misalignment, the reconstruction waveform based on afterward speech waveform. We use cross-correlation function to search the start point of waveform, which the most similar to the forward speech waveform. The processing is following.

$$p' = \arg\max_{p} \sum_{j=0}^{n-p} Y_L(j) Y_R(j+p) \tag{9}$$

where Y_L and Y_R are L_WSOLA and R_WSOLA respectively, p' denotes the start point of the most similar waveform. p is the index for search region, its duration is from zero to L_p, that is to say, $p = 0,\ 1,...,\ Lp$. The L_p is length of search region, In this research, we assume the L_p is quarter of packet length. The reconstruction waveform of lost packet is

$$\sum_{j=p'}^{l} Y_R(j) \tag{10}$$

Because the length of reconstruction waveform is less than the length of the packet, we need to extend reconstruction waveform. The extend method is the average expansion of the overall reconstruction waveform; see Figure 5. The blue line is original waveform with packet loss from frame 197 to 253 before extend. The red line means the speech waveform that is extended. The Figure 6 shows the reconstructed waveform by BV method, the blue waveform is original speech signals; the red and green waveforms are reconstruction speech signals by WSOLA and GWSOLA respectively; the purple waveform is reconstruction speech signals by proposed method in this paper.

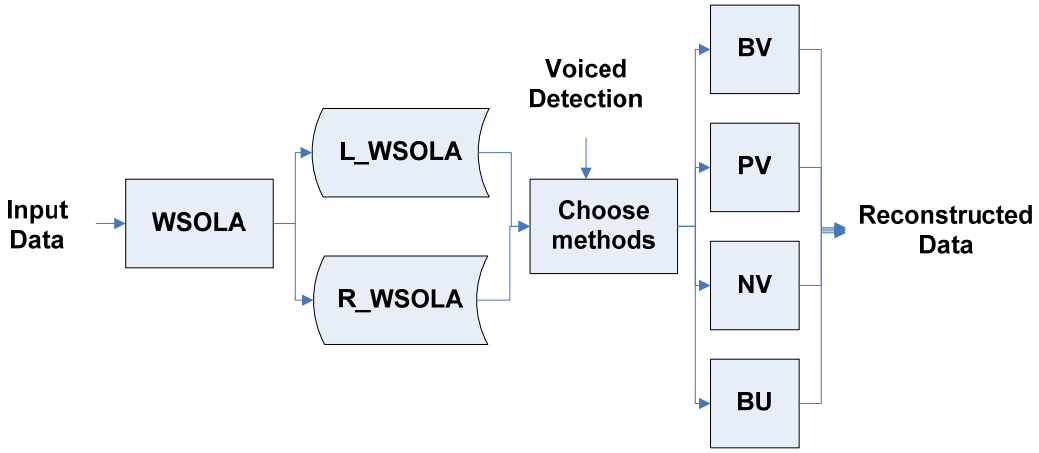

Figure 3. Flow-chart of reconstruction of lost speech frame by Bilateral WSOLA.

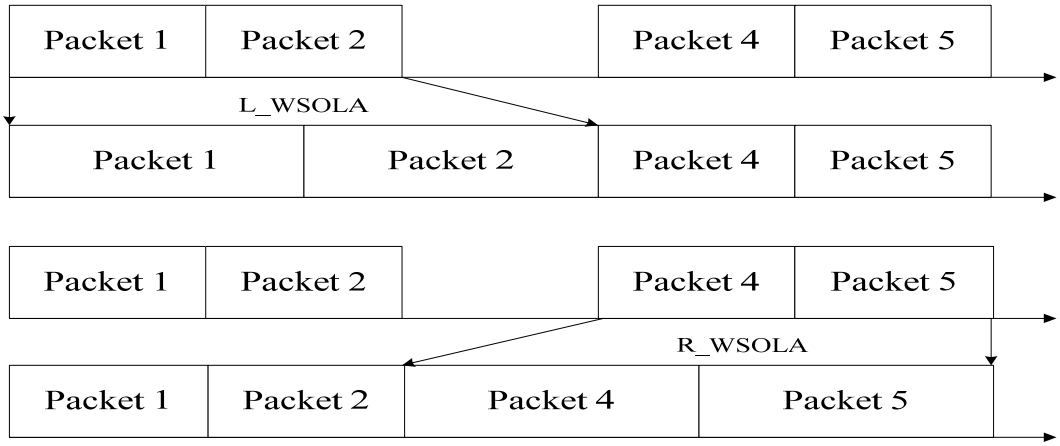

Figure 4. The Illustrations about traditional L_WSOLA and R_WSOLA.

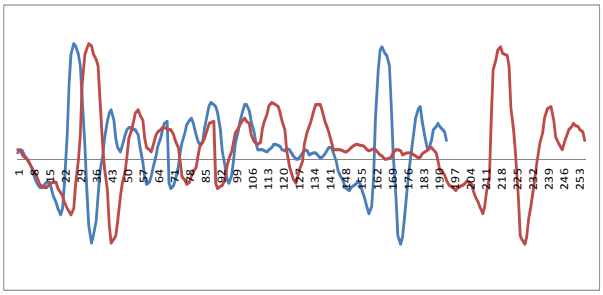

Figure 5. The reconstruction waveform after extend by WSOLA.

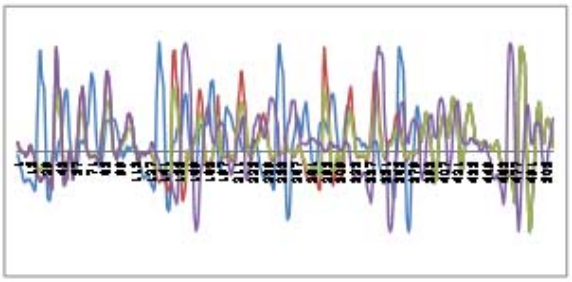

Figure 6. The reconstruction waveform of BV by proposed BWSOLA.

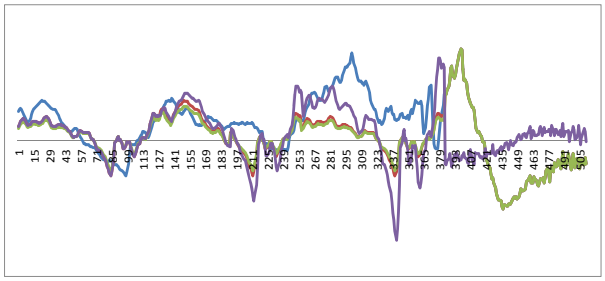

Figure 7. The reconstruction waveform of PV by proposed BWSOLA.

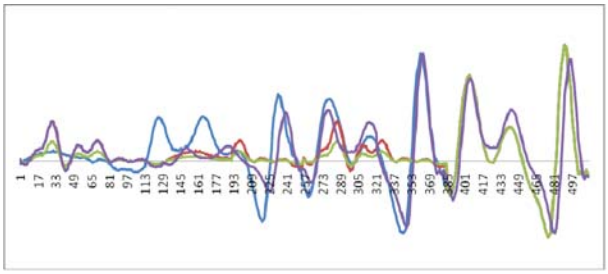

Figure 8. The reconstruction waveform of NV by proposed BWSOLA.

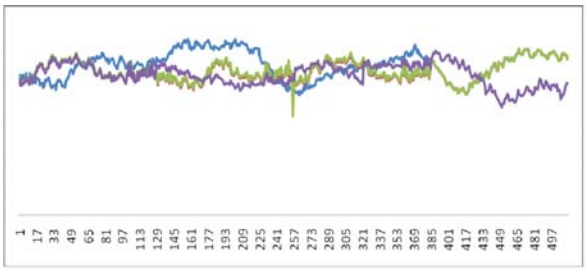

Figure 9. The reconstruction waveform of BU by proposed BWSOLA.

3.3 The methodology for PV or NV

When one direction of speech signal is voiced, another is unvoiced, we extend voiced speech signal and adjust amplitude. The procedure is following.

$$Y(n) = \sum_{j=0}^{l-1} S(j)A(j) \tag{11}$$

where $S(\cdot)$ is the voiced speech signal, $A(\cdot)$ is the adjust factor. The adjust factor is given by following equation.

$$A(n) = 1 - n\frac{E_V - E_{UV}}{E_V(l-1)} \quad n = 0,\ 1,...,\ l-1 \tag{12}$$

where E_V is the energy of the voiced speech, E_{UV} is the energy of unvoiced speech. The reconstruction waveform shows on Figure 7 and Figure 8. The notations of color lines are the same as those of Figure 6.

3.4 The methodology for BU

If both the forward and the afterward speech signals are unvoiced, then we can suppose the speech signal of missing packet is noise. In this case, we can use the fade-in and fade-out method to reconstruct waveform. The reconstruction waveform shows on Figure 9. The notations of color lines are the same as those of Figure 6.

4. Experimental results and discussion

Two experiments results are described as follows. The first experiment was designed to assess the speech quality in variant packet loss rate; the other one measurement is aimed at speech quality according to different lost packet length. The probability of packet loss was independent from the other packets in experimental setting. In our experiments, two metrics about speech quality were used to evaluations the performance of the proposed method: perceptual evaluation of speech quality (PESQ, ITU-T P.862-2) and the Mean Opinion Score (MOS). The Table 1 and the Table 2 showed the PESQ scores and the MOS scores in experiment of fixed packet loss rate respectively. The Table 3 and the Table 4 showed another experiment results. We can see that the waveform restored by proposed method in this paper usually gets a higher PESQ and MOS value. For comparisons, the conventional WSOLA and GWSOLA were implemented here.

4.1 Performance assess on packet loss rates

The reconstructed speech quality is very sensitive to the ratio of lost data. In this experiment, four packet loss rates were used to evaluate proposed packet loss concealment, DSESOLA: 10, 20, 30 and 40 percentages. According to the results shown in Table 1, we can find that the proposed BWSOLA can achieve higher PESQ values compared to conventional WSOLA and GWSOLA. This means that the proposed method outperforms WSOLA and GWSOLA especially in the higher packet loss rate. The reason is the bi-directional inference can achieve the better reconstruction for lost data. There were above 0.3 in PESQ value difference when the packet loss rates are 30% and 40%. The MOS values on different packet loss rates were illustrated in Table 2. As expected, the higher packet loss rate, the lower MOS value. According to the observations, we can find that the most improvement obtained in BV and NV by proposal BWSOLA, the contribution came from considering of the afterward waveform.

	10%	20%	30%	40%
WSOLA	3.177	2.226	1.728	1.395
GWSOLA	3.201	2.298	1.749	1.566
BWSOLA	3.324	2.335	2.104	1.810

Table 1. The PESQ values of the methods on different packet loss rates.

	10%	20%	30%	40%
WSOLA	3.13	2.53	2.23	2.06
GWSOLA	3.17	2.89	2.51	2.12
BWSOLA	3.56	3.32	2.84	2.53

Table 2. The MOS values of the methods on different packet loss rates.

4.2 Performance assess on packet loss rates

Since the continuous data lost and long packet lost affect the speech quality critically. Herein, we measured speech quality according to different lost packet length. We can found the longer data loss, the lower values of PESQ and MOS. The proposed BWSOLA can achieve significant improvement compared to WSOLA and GWSOLA due to the reconstruction according to the forward and afterward data gram. The improvements achieved by BV, PV, NV, and BU are illustrated here.

	1	2	3	4
WSOLA	2.6	2.3	1.3	1.3
GWSOLA	2.7	2.1	1.1	0.9
BWSOLA	3.0	2.5	1.6	1.3

Table 3. The PESQ values on different lost packet length.

	1	2	3	4
WSOLA	3.5	3.3	2.5	2.0
GWSOLA	3.7	3.2	2.4	2.0
BWSOLA	3.9	3.5	2.9	2.2

Table 4. The MOS values of the different methods on different lost packet length.

5. Conclusions

This paper presents the bilateral waveform similarity overlap-and-add algorithm (BWSOLA) for speech packet loss concealment. The proposed method is more suitable for packet loss concealment than others methods such as traditional WSOLA and GWSOLA because of the dual side information. The proposed BWSOLA reconstructed the gap caused by lost packet by extending the forward and the afterward waveforms to solve the gain control problem. Because the proposed method references the afterward waveform to solve misalignment problem and keep the continuity of reconstructed speech data by considering of consistency of amplitude, frequency and phase between reconstructed voice and adjacent voice. Both PESQ scores subjective test and MOS value objective test shows BWSOLA algorithm outperforms WSOLA and GWSOLA significantly. As a packet loss concealment method, the time domain approaches can obtain the real time performance for practice speech communication. Since the data processing capability is increasing higher recently, the method in frequency domain should be considered in the near future. Due to speech can be interpreted as harmonic and noise components, the approaches in frequency domain should obtain another contribution in reconstructed speech in quality and understanding [27].

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive feedback and helpful suggestions.

References

[1] A. Mantilla-Caeiros et al., “Pattern Recognition Based Esophageal Speech Enhancement System,” Journal of Applied Research and Technology, vol. 8, no. 1, pp. 56-71, 2010.

[2] A. L. Padilla-Ortíz, and F. Orduña-Bustamante, “Binaural Speech Intelligibility and Interaural Cross-Correlation Under Disturbing Noise and Reverberation,” Journal of Applied Research and Technology, vol. 10, no. 3, pp. 347-360, 2012.

[3] J.-F. Yeh et al., "Speech Enabling Services for Wireless Access in Vehicular Environments" ICIC Express Letters, Part B: Applications- An International Journal of Research and Surveys, Vol. 2, No. 3, 2011, pp.705-710.

[4] H. Sanneck H. er al, “A New Technique for Audio Packet Loss Concealment,” Global Telecommunications Conference, 1996, pp. 48-52, 1996.

[5] C. Perkins et al., "A Survey of Paket Loss Recovery techniques for Streaming Audio," IEEE Network, Vol. 12, No. 5, pp. 40-48, 1998.

[6] H. Svensson et al., “Implementation Aspects of a Novel Speech Packet Loss Concealment Method,” IEEE International Symposium on Circuits and Systems, Kobe, Japan ,Vol. 3, pp. 2867-2870, 2005.

[7] S. Grofit and Y. Lavner, “Time-Scale Modification of Audio Signals Using Enhanced WSOLA With Management of Transients,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no.1, 2008, pp. 106-115, 2008.

[8] H.-P. Shen et al., “Speaker Clustering Using Decision Tree-based Phone Cluster Models with Multi-Space Probability Distributions,” IEEE Trans. Audio, Speech, and Language Processing, Vol. 19, Issue 5, pp.1289-1300, 2011.

[9] J.-F. Yeh and M.-C. Yen, “Speech Recognition with Word Fragment Detection Using Prosody Features for Spontaneous Speech,” International Journal of Applied Mathematics & Information Sciences, V 6 No. 2S pp. 669S-675S, 2012.

[10] M. Li et al., “Packet Loss Concealment Using Enhanced Waveform Similarity OverLap-and-Add Technique with Management of Gains,” International Conference on Wireless Communications, Networking and Mobile Computing 2009 (WiCom '09), Beijing, China, 2009.

[11] H. G. Ilk and S. Güler, “Adaptive Time Scale Modification of Speech for Graceful Degrading Voice Quality in Congested Networks for VoIP Applications,” Signal Process, 86(1), pp. 127–139, 2006.

[12] K. Sreenivasa Rao and B. Yegnanarayana, “Duration Modification using Glottal Closure Instants and Vowel Onset Points,” Speech communication, Vol. 51, issue 21, pp.1263-1269, 2009.

[13] A. Ito and T. Nagano, "Packet Loss Concealment of VoIP under severe loss conditions," 15th International Symposium on Wireless Personal Multimedia Communications (WPMC), pp.489 - 490, 2012.

[14] Jagla et al. "Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-and-add method," J. Acoust. Soc. Am., vol. 132, no. 5, pp. 3098-3108, 2012

[15] L. Wang et al, “Waveform Similarity Over-and-add Technique with Gain Control,” IEEE International Conference on Broadband Network & Multimedia Technology, Beijing, China, pp. 735-739, 2009.

[16] L. Wang et al., “A Packet Loss Concealment Method Base on GWSOLA algorithm and Signification Transient Detect,” IEEE International Conference on Network Infrastructure and Digital Content 2009, Beijing, China, pp.745-749, 2009.

[17] W. Verhelst and M. Roelands, "An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modifications of speech," in Proc. of International Conference on Acoustics Speech and Signal Processing, Minneapolis, MN, 2, pp. 554-557, 1993.

[18] Y. Sun et al., "A modified weighted overlap and add-based spectral subtraction method," J. Acoust. Soc. Am. Volume 131, Issue 4, pp. 3444-3444, 2012.

[19] J.-H. Chen, “Packet Loss Concealment Based on Extrapolation of Speech Waveform,” IEEE International Conference on Acoustics, Speech and Signal Processing 2009, Taipei, Taiwan, R.O.C., pp.4129-4132, 2009.

[20] K. Kondo and K. Nakagawa, “A Speech Packet Loss Concealment Method Using Linear Prediction,” IEICE Transaction on Information & Systems, Vol. E89-D, No. 2, pp.806-813, 2006.

[21] N. Linenberg et al., “Packet Loss Concealment Using Adaptive Lattice Modeling,” 15th IEEE Mediterranean Electrotechnical Conference, pp.378-382, 2010.

[22] J.-F.Wang et al, “A Voicing Driven Packet Loss Recovery Algorithm for Analysis-by-synthesis Predictive Speech Coders Over Internet,” IEEE Transactions on Multimedia, vol. 3, pp. 98–107, 2001.

[23] J.-F. Yeh and C.-H. Hsu, “Sub-Syllable Segment-based Voice Conversion Using Spectral Block Clustering Transformation Functions,” Journal of the Chinese Institute of Engineers. Vol. 33, No. 7, pp. 1059-1067, 2010.

[24] A. Mantilla□Caeiros et al., “A Pattern Recognition based Esophageal Speech Enhancement System,” vol.8, no. 1, pp. 56-71, 2010.

[25] M.-Y. Zhu et al., “An Accurate Low Complexity Algorithm for Frequency Estimation in MDCT Domain,” IEEE Transactions on Consumer Electronics, Vol. 54, No.3, pp. 1022-1028, 2008.

[26] H.-Y. Gu and Z.-S. Chen., "A Packet Loss Concealment Method for Voice over IP," 18th Conference on Computational Linguistics and Speech Processing, 2006.

[27] W.-T. Liao et al., “Adaptive Recovery Techniques for Real-time Audio Streams,” Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM), vol. 2, , Anchorage, USA, pp.815-823, 2001.