Bilateral Waveform Similarity Overlap Add Approach based on Time Scale Modification Principle for Packet Loss Concealment of Speech Signals

Miss. Rohini D. Patil
Research Student,
Department of Electronics,
DKTE's Textile and Engineering Institute,
Ichalkaranji, Maharashtra

Prof (Mrs.) D. Y. Loni
Assistant Professor,
Department of Electronics,
DKTE's Textile and Engineering Institute,
Ichalkaranji, Maharashtra

***Abstract*— Packet arrival delays at receiver and packet loss in the network greatly affect the quality of received signals in Voice over IP (VoIP) networks. To reduce this quality degradation caused by packet loss, Packet Loss Concealment (PLC) techniques are used. In this paper we presented Bilateral Waveform Similarity Overlap Add approach (BWSOLA) using Time Scale Modification (TSM). BWSOLA is receiver based PLC algorithm, which uses both previous and afterward packets for concealment of lost packets. BWSOLA maintains consistency in amplitude, frequency and phase between concealed and adjacent speech signals. The paper includes detailed comparison of Waveform Similarity Overlap Add (WSOLA), Gain controlled Waveform Similarity Overlap Add (GWSOLA) and BWSOLA techniques. Results obtained clearly indicate the quality improvement in BWSOLA as compared to WSOLA and GWSOLA techniques used for concealment of lost packet. The parameters used for comparison are Euclidean, Manhattan and Chebychev distance.**

Keywords— ***VoIP; PLC; WSOLA; GWSOLA; BWSOLA; Euclidean distance; Manhattan distance; Chebychev distance.***

I. INTRODUCTION

Voice over IP (VoIP) permits the transport of voice over public internet. Internet gives various value added services by bringing together voice and data. Unfortunately, VoIP does not yet provide best quality of voice data over the VoIP network. So the quality of delivered data is an important issue while using VoIP for real time communication. For non-real time applications such as Telnet, FTP (File Transfer Protocol), email etc., VoIP can be used in an efficient way because TCP (Transmission Control Protocol) ensures reliable delivery of data. TCP is connection oriented protocol that uses best effort service. It integrates various retransmission policies based on feedback and timeout mechanism for reliable data transfer. But for real-time applications, TCP does not promise the required quality of data demanded by real-time applications [1].

IP networks are essentially best-effort networks with variable delay and loss. Voice traffic in the network can tolerate some packet loss. But, if the packet loss rate is greater than 5%, it is considered harmful to the voice quality and a good concealment technique is required for reconstruction of the lost packets [2]. Phoneme is the unit of sound that distinguishes one word from other. The length of a phoneme is typically between 80 to 100 ms. In voice transmission, when the duration of the packet loss is greater than the length of a phoneme it can change the meaning of a word [3]. VoIP applications use UDP (User Datagram Protocol) as the transport layer protocol. The Real Time Protocol (RTP) and Real Time Control Protocol (RTCP) are used to provide additional functionality such as adding sequence numbers and timestamps [4].

Packet loss on network is a major cause of audio signal loss. It is caused by the transmission impairments such as the excess of the transmission capacity and congestion and also may be caused in wireless channel due to noise, co-channel interference and fading [5]. Sound quality of the signal with lost packets can be improved by PLC. Packet Loss Concealment approaches are of two types, one is sender based and another is receiver based approach. Sender based approaches reduces the packet transmission failure while receiver based approaches focus on improving the quality of reconstructed speech signals [6].

Time Scale Modification (TSM) can be efficiently used for PLC of audio signals. Time scaling is a process of changing the speed or duration of an audio without affecting its pitch. Time Scale modification of speech signals can be used for supporting hearing impaired school children. Comparisons of results obtained from Overlap and Add (OLA) and Synchronous overlap and add (SOLA) shows that SOLA give improved results than OLA [7].

For non-uniform real-time speech stretching the algorithm uses combination of Synchronous overlap and add (SOLA) with vowels, consonants, and silence detector and the obtained results showed that the algorithm given high quality of stretched speech [8]. For high quality time-scale modifications of speech WSOLA can be used. The time scaling algorithm should produce an output that maintains maximal local similarity to original signal. SOLA algorithm does not maintain maximum local similarity, WSOLA technique was introduced which overcome the drawbacks of SOLA. WSOLA technique ensures sufficient signal continuity at segment joins that occurred in original signal [9]. Time scaling of correctly received packets over heterogeneous packet switched networks based on time-scale modification using WSOLA, reduces the error due to lost packets. Additional delay and complexity required for concealment was reduced by new parameter selection scheme. Using subjective listening tests, the results compared shows that overall sound quality is higher for proposed technique [10]. A new technique for improving quality of speech in voice over IP using Time Scale Modification using Global Local Search Time Scale Modification (GLS-TSM) can provide flexible delay cutoffs to late arrived packets and reduce additional buffering delay. TSM was performed on the packets already present in the playout buffer [11].

A sinusoidal model of successfully received speech signal is used to recover lost frames using both extrapolation and interpolation techniques. Especially for music signals, use of interpolation improves results [12]. The standard WSOLA algorithm only considers phase information and lacks in amplitude control, which leads to significant mismatch between the original and reconstructed speech signal. This problem can be overcome, by introducing gain in standard WSOLA algorithm. Gain controlled Waveform Similarity Overlap and Add (GWSOLA) adjusts the level of audio signal to maintain audio signal level consistent. GWSOLA performs better than WSOLA [13]. In the proposed algorithm, gain is introduced into the standard WSOLA technique which helped to adjust the amplitude of the voice. Simulation results show that introduction of a gain into the standard WSOLA technique, improves the quality of the restored voice signal than that standard WSOLA technique [14].

Incremental subspace learning along with non-negative matrix factorization can be used for recovery of lost speech segments [15]. Packet loss concealment problem can be solved using G.729. In the proposed work 50% loss is considered for concealment [16]. The receiver based PLC algorithm used adaptive thresholding for classification of packets and linear-prediction based packet loss concealment is performed. The algorithm enhances the quality of signal with low complexity and gives higher quality of voice [17]. Bilateral Waveform Similarity Overlap Add (BWSOLA) algorithm based on WSOLA uses the bilateral information for reconstruction of lost packet. As algorithm uses both sided information, the quality of the reconstructed speech signal of the BWSOLA is much better than WSOLA and GWSOLA. BWSOLA outperforms the traditional approaches especially in the long duration data loss [18].

II. FRAMEWORK OF BILATERAL WAVEFORM SIMILARITY OVERLAP ADD (BWSOLA)

TSM can be used for packet loss concealment. The BWSOLA algorithm starts working after the packet loss is detected at the receiver end. When packet lost is detected, both the forward and afterward speech signals contents are extracted they are used to reconstruct the lost speech data. The GWSOLA algorithm is applied to both sided packets and the results are stored in L_GWSOLA and R_GWSOLA respectively. As BWSOLA adopts different approaches for voiced and unvoiced speech signal, both sided signals needs to be assessed for determining its state (voiced or unvoiced). Autocorrelation function (ACF) along with periodogram technique is used to decide the state of speech signal by computing its fundamental frequency. Fig 1 shows the block diagram of Bilateral Waveform Similarity Overlap Add approach. According to the state of both sided packets they are categorized into four different categories as: 1) BV (Both Voiced) where the both previous and afterward speech signals are voiced, 2) PV (Previous Voiced) where the previous signals is voiced and the afterward signals is unvoiced, 3) AV (Afterward Voiced) where the previous speech signals is unvoiced and the afterward signal is voiced 4) BU (Both Unvoiced) where both the previous and afterward signals are unvoiced. Finally depending upon the voicing state, different reconstruction strategies are used to reconstruct the lost speech signal using BWSOLA.

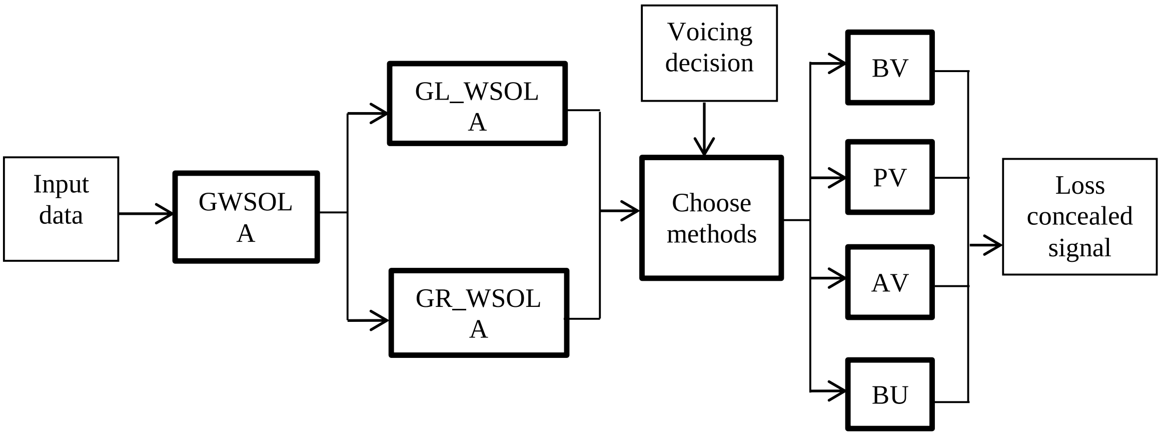

Figure 1 Block Diagram of Bilateral Waveform Similarity Overlap Add (BWSOLA)

III. Implementation of Waveform Similarity Overlap Add (WSOLA) approach

Standard WSOLA algorithm is suitable for packet loss concealment in real time voice communications. When the receiver detects a packet loss in the received audio signal, the packets before the lost packet are extended using TSM, to cover the gap of the missing packet such that TSM must preserve the pitch frequency and timbre of speech signal [10].WSOLA algorithm operates in two steps as:

1. Analysis Step 2. Synthesis Step

Fig 2. (a) shows the operation of analysis step in which input signal X[n] have 2 packets before the lost packet. In analysis step, 3 frames Sk's are extracted from the input signal, at time instant xk's according to the maximal cross correlated point and then in analysis step frames are added with less overlap in output at position yk's as shown in Fig 2. (b), and we got the output signal Y[n].

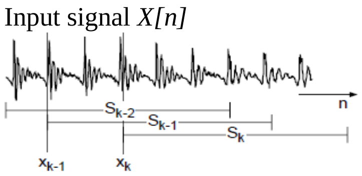

Figure 2. (a) Analysis step WSOLA algorithm: extract S_k

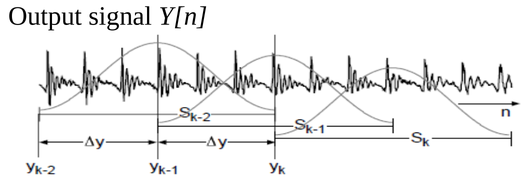

Figure 2. (b) Synthesis step WSOLA algorithm: overlap and add

For implementation we have considered 8 kHz sampling frequency of the input signal. Length of the packet is 20 ms i. e. 160 samples per packet. In the analysis step, n=2 number of packets before lost packet are extracted and m=3 number of analysis frames Sk's are formed, length of each analysis frame is L and is calculated using Equation (1)

$$L = \frac{ml}{n} \qquad (1)$$

We extract the first analysis frame of length L=240, then computed analysis time shift Δx using Equation (2)

$$\Delta x = \frac{(m-1)l - L}{2(n-1)} \qquad (2)$$

The synthesis time shift Δy is Calculated using Δy = L/2. We got Δx=40 and Δy=120 samples. The search region should be centered on Δx i.e. on 40th sample from the start of analysis frame. Length of search region is LS, it should not be less than one pitch period and should not be greater than a phoneme. Location of Search region for analysis frame is shown in Fig 3.

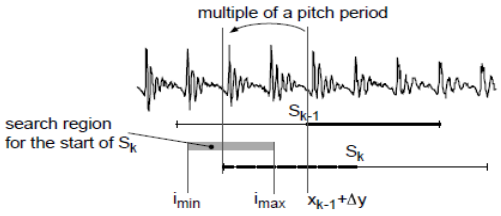

Figure 3. Decision of search region

In Fig 3, i_{min} and i_{max} are start point and end point of search region respectively. After deciding the search region, cross-correlation is computed according to Equation (3)

$$C_i = \sum_{j=0}^{\Delta y-1} X(i+j)X(x_{k-1}+\Delta y + j) \qquad (3)$$

Correlation value C_i gives the best index at which the next frame should start i. e. x_{k-1}, where *i* fall within the search region. Then according to the value of C_i extract the next analysis frame. The procedure is repeated to get the next analysis frames. At the end of analysis step we get the overlapping analysis frames. In synthesis step, the second and third analysis frames are shifted with Δ*y* and *2**Δ*y* respectively, retaining first analysis frame. Then the shifted synthesis frames are overlap added using Equation (4)

$$Y[n] = \sum_{k=0}^{L-1} w[n-y_k]\cdot X[n-y_k+x_k] \qquad (4)$$

w[n] = flattopwin window

IV. Implementation of Gain controlled Waveform Similarity Overlap Add (GWSOLA) approach

The operation of analysis step of GWSOLA is same as WSOLA approach. But in synthesis step, gain of each frame is introduced along with respective frames and then overlap add operation is performed. Gain of each frame is calculated using Equation (5)

$$g_k = \frac{\sum_{n=0}^{L-1} X(y_k+n)S_k(n)}{\sum_{n=0}^{L-1} S_k(n)^2} \qquad (5)$$

g_k = gain of k^{th} frame

Then the overlap add is performed using Equation (6)

$$Y(n) = \sum_{k=0}^{m} g_k S_K(n+y_k) \qquad (6)$$

Y[n] = output of GWSOLA

V. Implementation of Bilateral Waveform Similarity Overlap Add (BWSOLA) Approach

For implementation of BWSOLA algorithm, the input speech signals i. e. previous and afterward signals are extended using GWSOLA, and stored in L_GWSOLA and R_GWSOLA respectively.

As shown in Fig 4 packet 1 and 2 are extended using GWSOLA and stored in L_GWSOLA. Packet 4 and 5 are extended using GWSOLA and stored in R_GWSOLA. As different methods are used for voiced and unvoiced signals, initially we have to decide the state of speech signal. Depending upon voicing state further different reconstruction strategies are chosen from BV, PV or AV, and BU.

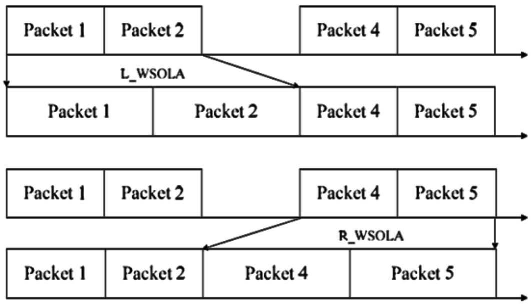

Figure 4. Evaluation of L_WSOLA and R_WSOLA

A. *Methodology for BV*

If both the previous and afterward speech signals are voiced, only previous or afterward speech signal is used to reconstruct lost signal it may cause misalignment between reconstruction waveform and the original waveform. To avoid this problem both the previous and afterward signal are used. In process of reconstruction, the previous and afterward speech signals are extended using GWSOLA. In order to avoid problem of misalignment, the reconstruction waveform based on afterward speech waveform. Cross-correlation function is used on L_GWSOLA and R_GWSOLA, to find the start point of reconstructed waveform, which the most similar to the previous speech waveform. We found the maximum correlation point over the length of search region and got the point of most similarity p', which denotes the start point of the most similar waveform. The reconstruction waveform of lost packet is extracted using equation 7

$$\sum_{j=p'}^{l} Y_{GR}(j) \qquad (7)$$

YGR is R_GWSOLA. The extracted reconstructed waveform is less than the packet length. To get the final reconstructed signal WSOLA is applied to the available reconstruction signal.

B. *Methodology for PV or AV*

When one sided speech signal is voiced, another is unvoiced, we extend voiced speech signal and adjust amplitude. The adjust factor is calculated using equation (8)

$$A(n) = 1 - n\frac{E_V - E_{UV}}{E_V(l-1)} \qquad (8)$$

Where EV is the energy of the voiced speech, EUV is the energy of unvoiced speech. And n=0, 1 …l-1

The adjust factor is applied on the voiced signal using equation (9)

$$Y(n) = \sum_{j=0}^{l-1} S(j)A(j) \qquad (9)$$

Where S (j) is the voiced speech signal, A (j) is the adjust factor.

C. *Methodology for BU*

If both previous and afterward packets are unvoiced then l/2 i.e. 80 samples are extracted from the end of the previous packet and l/2+10 i.e. 90 samples are extracted from start of afterward packet. Overlap add is performed on last 10 samples of previous and first 10 samples of afterward signal to get the output signal.

VI. Experimental Results and Discussion

To implement and validate the algorithm, we have initially created input signal with lost packet. Input signal used is “Speak up whatever you want”, and loss of 54th packet was created. Then using the WSOLA, GWSOLA, BWSOLA approaches of TSM the lost packet is concealed and experimental results of these approaches are discussed in a, b, and c respectively. As the packet loss is manually generated, we used original lost packet to compare how close the reconstructed packet is with respect to the known lost packet.

A. *Experimental results and Discussion of WSOLA*

Fig 5 shows the results of WSOLA using only previous packets (L_WSOLA).From fig 5, we can see that the major deviation occures in the second half of the signal, as this approach uses only previous packets for TSM.

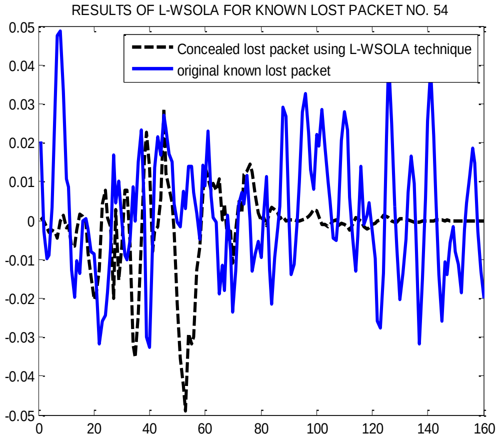

Figure 5. Comparison of the known lost packet and result of WSOLA approach using only previous packets (L_WSOLA)

Fig 6 shows the results of WSOLA using only afterward packets (R_WSOLA). From fig 6, it can be seen that major deviation takes place in the first half of the signal, as this approach uses only afterward packets for TSM.

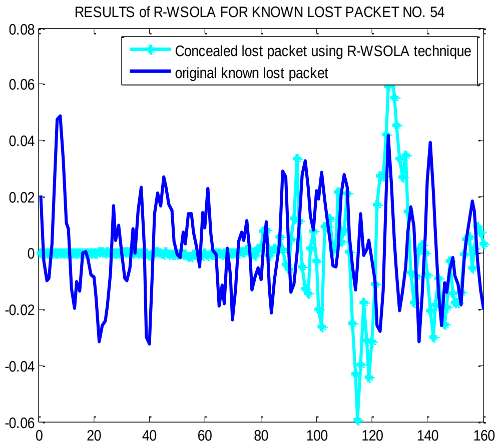

Figure 6. Comparison of the known lost packet and result of WSOLA approach using only afterward packets (R_WSOLA)

From fig 5 and 6, it can be said that WSOLA algorithm the concealed packet deviates widely at the initial and tail parts of the packet.

B. Experimental results and Discussion of GWSOLA

The comparison of known lost packet and the result of PLC by using GWSOLA method using only previous packets are shown in Fig 7. The results are improved than WSOLA but deviation still exists.

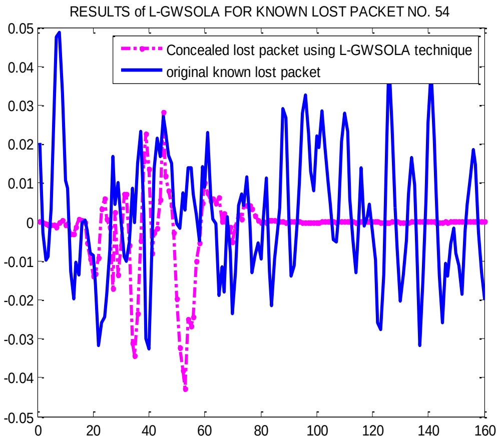

Figure 7. Comparison of the known lost packet and result of GWSOLA approach using only previous packets (L_GWSOLA)

Results of GWSOLA algorithm using only afterward packets are shown in fig 8. The comparison of known lost packet and the result of PLC shows that results are improved as compared to WSOLA but deviation still exist.

Comparing fig 7 and 8, it can be observed that GWSOLA algorithm improves the results but deviation remains same in terms of amplitude levels in the known lost packet and result of GWSOLA.

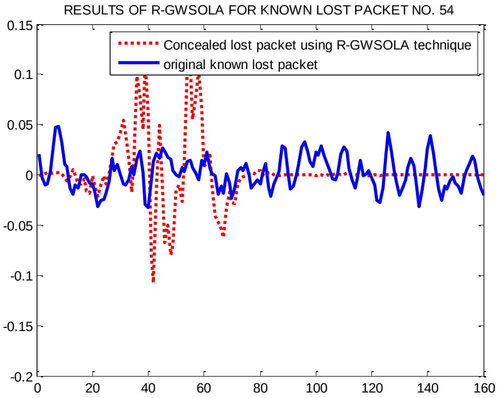

Figure 8. Comparison of the known lost packet and result of GWSOLA approach using only afterward packets (R_GWSOLA)

C. Experimental results and Discussion of BWSOLA

Fig 9, shows the comparison of known lost packet and the result of PLC by using BWSOLA algorithm. From Fig 9, we can see that deviation is reduced much more than WSOLA and GWSOLA. The concealed packet obtained using BWSOLA approach closely approximates the known lost packet compared to other approaches.

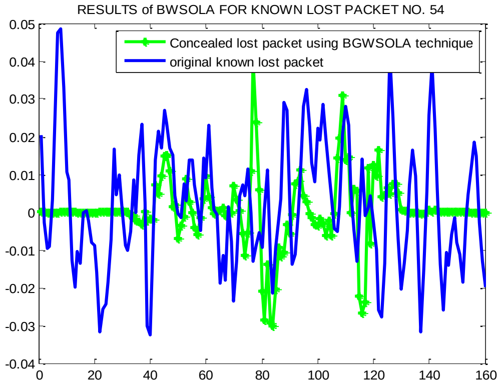

Figure 9. Comparison of the known lost packet and result of BWSOLA approach (BWSOLA)

Comparison of results is also performed by computing distance between the known lost packet and concealed packet using metrics like Euclidean distance, Manhattan distance and Chebychev distance on three test signals. Three test signals used as

1. Test signal-1 is “You Have Mail”.
2. Test signal-2 is “Good Bye”.
3. Test signal-3 is “speak up whatever you want”

Test signal-1 has 27 packets; out of 27 packets we got 2 results better for LWSOLA i.e. 7.4%. Test signal-2 have 28 packets out of 28 we got 4 results better for LWSOLA i.e. 14.28%. Test signal-3 have 82 packets out of 82 we got 6 results better for LWSOLA i.e. 7.31%. Accordingly percentage is calculated for three test signal with all algorithms.

Table I. Comparison WSOLA, GWSOLA and BWSOLA algorithms for 3 test signals using Euclidean Distance metric

	Test signal-1	Test signal-2	Test signal-3
LWSOLA	7.4	14.28	7.31
GLWSOLA	29.62	3.57	15.85
RWSOLA	11.11	10.71	14.63
GRWSOLA	7.4	28.57	23.17
BWOLA	44.44	42.85	39.02

By observing table I, we can see BWSOLA algorithm outperforms for all three test signals than WSOLA and GWSOLA algorithm.

Table II. Comparison WSOLA, GWSOLA and BWSOLA algorithms using Manhattan Distance metric for 3 test signals

	Test signal-1	Test signal-2	Test signal-3
LWSOLA	7.4	10.71	8.53
GLWSOLA	14.81	10.71	15.29
RWSOLA	25.92	3.57	14.63
GRWSOLA	14.81	32.14	17.07
BWOLA	37.03	42.85	42.5

From the comparison using Manhattan Distance in table II we can see that for all three test signals BWSOLA algorithm gives better results.

Table III. Comparison WSOLA, GWSOLA and BWSOLA algorithms using Chebychev Distance metric for 3 test signals

	Test signal-1	Test signal-2	Test signal-3
LWSOLA	22.22	7.14	7.13
GLWSOLA	14.81	10.71	21.95
RWSOLA	7.4	25	13.41
GRWSOLA	18.51	25	23.71
BWOLA	37.03	32.14	32.92

Table III gives comparison using Chebychev Distance after observing the results in table III, is can be said that WSOLA and GWSOLA gives very poor results as compared to BWSOLA.

VII. CONCLUSION

The standard WSOLA algorithm is appropriate for packet loss concealment in real time voice communications. It extends the previous packets of the lost packet, such that the gap caused by the lost packet can be concealed by the stretched version of the previous packets, while the pitch frequency and timbre of the voice transmitted are maintained unchanged. But WSOLA lacks in efficient amplitude controls, which lead to significant mismatch between the original signal and the concealed signal. As in GWSOLA algorithm gain is introduced, the amplitude mismatch problem is effectively solved. But both the algorithms WSOLA and GWSOLA considers only previous and afterward packets i.e. only one side packets for concealment of lost packets. In BWSOLA algorithm both sided packets are used for concealment and also GWSOLA algorithm is used so amplitude and phase will also be matched as in original signal. From the results, we can conclude that BWSOLA algorithm gives better results than WSOLA and GWSOLA algorithms.

REFERENCES

[1]. www.wekipedia.com

[2]. N. Jayant and S. W. Christensen, “Effect of packet losses in waveform coded speech and improvement due to an add-even sample-interpolation procedure,” *IEEE Trans. Communications*, vol. 29, pp. 101–109, Feb. 1981.

[3]. A. Watson and M. Sasse, “Measuring perceived quality of speech and video in multimedia conferencing applications,” *Proc. of ACM Multimedia*, pp. 55–60, Sept. 1998.

[4]. H. Shulzrinne, S. Casner, R. Frederick, and V. Jacobsen, “A Transport Protocol for Real-time applications.” *Network Working Group RFC-1889, Internet Engineering Task Force*, Jan. 1996.

[5]. Moon-Keun Lee; Sung-Kyo Jung; Hong-Goo Kang; Young-Cheol Park; Dae-Hee Youn, “A packet loss concealment algorithm based on time-scale modification for CELP-type speech coders.” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 116-119, 2003.

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

[7]. Adam Kupryjanow, Andrzej Czyżewski, **“**Time Scale modification of speech signals for supporting hearing impaired schoolchildren”, *IEEE International conference on signal processing algorithms, architectures, arrangements & applications*, pp. 159- 162, 2009.

[8]. Adam Kupryjanow, Andrzej Czyżewski., “a non-uniform real-time scale stretching method”, *IEEE International conference on signal processing and multimedia applications (SIGMAP)*,pp. 1-7, 2011

[9]. 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.

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

[11]. Agnihotri S., Aravindhan K., Jamadagni H. S., Pawate B. I., “A New Technique for Improving Quality of Speech in Voice Over IP Using Time Scale Modification”, *IEEE international conference on Acoustics Speech, and signal processing,(ICASSP’02)*, Vol. 2, pp. 2085-2088, 2002.

[12]. Jonas Lindblom and Per Hedelin,**“**Packet Loss Concealment Based On Sinusoidal Modeling”, speech coding, *IEEE Workshop Proceedings*, pp. 65-67, 2002

[13]. 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.

[14]. 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.

[15]. Jianjun Huang, Xiongwei Zhang, and Yafei Zhang, “Recovery of Lost Speech Segments Using Incremental Subspace Learning,” *ETRI Journal*, Vol. 34, pp. 645-648, August 2012.

[16]. 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

[17]. Byeong hoon kim, Hyoung-Gook kim, Jichai Jeong, and Jin Young Kim, “VoIP Receiver Based Adaptive Playout Scheduling And Packet Loss Concealment Technique,” IEEE Transaction on Consumer Electronics, Vol. 59, no. 1, 2013

[18]. J. F. Yeh, P. C. Lin, M. D. Kuo, Z. H. Hsu, “Bilateral Waveform Similarity Overlap-and-Add Based Packet Loss Concealment for Voice over IP”, *Journal of Applied Research and Technology*, Vol. 11, pp. 559-567,August 2013