PAPER

# A Speech Packet Loss Concealment Method Using Linear Prediction*

**Kazuhiro KONDO**†a), *Member and* **Kiyoshi NAKAGAWA**†, *Fellow*

**SUMMARY** We proposed and evaluated a speech packet loss concealment method which predicts lost segments from speech included in packets either before, or both before and after the lost packet. The lost segments are predicted recursively by using linear prediction both in the forward direction from the packet preceding the loss, and in the backward direction from the packet succeeding the lost segment. Predicted samples in each direction are smoothed by averaging using linear weights to obtain the final interpolated signal. The adjacent segments are also smoothed extensively to significantly reduce the speech quality discontinuity between the interpolated signal and the received speech signal. Subjective quality comparisons between the proposed method and the the packet loss concealment algorithm described in the ITU standard G.711 Appendix I showed similar scores up to about 10% packet loss. However, the proposed method showed higher scores above this loss rate, with Mean Opinion Score rating exceeding 2.4, even at an extremely high packet loss rate of 30%. Packet loss concealment of speech degraded with G.729 coding, and babble noise mixed speech showed similar trends, with the proposed method showing higher qualities at high loss rates. We plan to further improve the performance by using adaptive LPC prediction order depending on the estimated pitch, and adaptive LPC bandwidth expansion depending on the consecutive number of repetitive prediction, among many other improvements. We also plan to investigate complexity reduction using gradient LPC coefficient updates, and processing delay reduction using adaptive forward/bidirectional prediction modes depending on the measured packet loss ratio.

***key words:*** *speech packets, packet loss concealment, linear prediction, segment smoothing, subjective speech quality evaluations*

## 1. Introduction

In many of the real time speech communication systems that transmit signals using packets, *e.g.* the emerging VOIP systems, there exists a finite probability that some packets will be lost. The two major sources of packet loss are congestion in the intermediate nodes, and discarded packets at the receiving end due to packets arriving too late to be decoded and played out. These packet losses may in some cases lead to intervals with significant loss rate, requiring a robust concealment solution.

Various methods have been proposed to try to generate an estimate of the lost speech segments included in the lost packets [1],[2],[3]. Some require modifications to the data included in the packets, *e.g.* the integration of overhead information to help the accurate estimation of the lost segments, while others attempt this only from the speech signal in the previously received packets. We will only deal with the latter in this paper.

The simplest substitute for the lost signal is either silence or gain-adjusted pseudo-random noise. However, this has been known to degrade the perceived speech quality significantly. The next best solution is to simply copy the last packet, and substitute a replica of the speech signal included in the copied packet. This provides somewhat less degradation, but in many cases is still not sufficient. A somewhat better solution is to interleave the speech samples and send them in separate packets [4],[5]. If one of the interleaved packets are lost, they can be recovered by interpolating them from received interleaved samples. This provides a much better estimate of the lost segment, but at the cost of the added delay necessary to interleave. Time scale modification techniques have been used to stretch received packets to cover the lost packets, and have shown some success [6]. However, this technique also will degrade the quality significantly as packets are lost consecutively. Recently, statistical n-gram estimate of the lost speech vector from previous $n-1$ received speech vectors have been attempted, and have shown some promise [7]. Concrete results still needs to be seen.

ITU has also standardised a packet loss concealment (PLC) algorithm to be used with the G.711 PCM standard [8]. This algorithm uses pitch detection to estimate the best matching pitch period immediately after the last received data, and repeats the pitch period data to fill the lost segment. This algorithm provides reasonably good estimate of the lost segment at fairly low complexity. An enhanced algorithm, which performs pitch period repetition in the LPC residual domain, has been proposed and standardised by ANSI [9],[10]. The proposed enhancement showed modest improvement over the method described in [8].

In this paper, we propose a PLC algorithm that predicts lost speech segment from speech included in either packets before the lost packet, or in both packets before and after the lost packet. Linear prediction is employed repetitively both in the forward direction, *i.e.* from the packet preceding the loss, and in the

Manuscript received January 1, 2003.
Manuscript revised January 1, 2003.
Final manuscript received January 1, 2003.

†The authors are with the Faculty of Engineering, Yamagata University, Yonezawa, Yamagata, JAPAN
a) E-mail: kkondo@yz.yamagata-u.ac.jp

*This paper was presented in part at the Interspeech-ICSLP 2004.

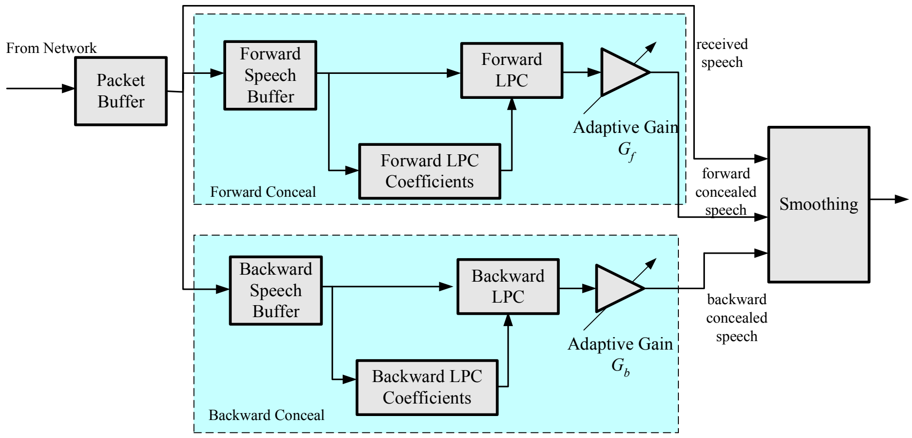


**Fig. 1** Block diagram of the proposed PLC algorithm.

backward direction, *i.e.* from the packet succeeding the lost segment. The predicted sample in each direction is smoothed to obtain the final interpolated signal. Smoothing is also applied to the adjacent segments to significantly reduce the speech quality discontinuity between the interpolated signals. We compared the subjective quality of our algorithm to the G.711 Appendix I PLC algorithm. The results show that the two algorithms show similar quality up to a packet loss rate of about 10%, but the proposed algorithm generally shows better quality above this packet loss rate [11].

In the next section, we will describe the proposed algorithm. In section 3, the speech quality evaluation tests along with its results are described. Section 4 compares the advantages as well as the disadvantage of the proposed algorithm with the G.711 Appendix I PLC algorithm. Finally, the conclusion is given in section 5.

## 2. Packet loss concealment algorithm using linear prediction

### 2.1 Linear prediction of speech segments in lost packets

Figure 1 shows the configuration of the proposed algorithm. All processing is done at the receiving end only. We will use linear prediction recursively to try to estimate the lost speech segment from speech samples in the neighbouring received packets. We can write the forward prediction of one sample from the preceding received packet using the standard linear prediction equation [12]:

$$\hat{x}_{f,n} = -1 * \sum_{i=1}^{N} a_i x_{n-i} \tag{1}$$

where $\hat{x}_{f,n}$ is the predicted sample, $x_{n-i}$ is the previously received speech sample, $N$ is the LPC prediction order, and $a_i$ is the LPC coefficient calculated from M samples preceding the lost packet, $x_{n-j}, j = 1, 2, \ldots, M$ where $M$ is the analysis window length. The LPC coefficients are calculated using the Levinson-Durbin algorithm [12] using the samples in the analysis window. There may be cases where some of the samples in the analysis window is not available due to packet loss. If this is the case, concealed samples will simply be used instead of actual received samples for LPC coefficient calculation.

The following samples, $\hat{x}_{f,n+i}$, are predicted recursively using both predicted samples and previously received speech samples. For example, both

$$\hat{x}_{f,n}$$

and

$$x_i, i = n-1, n-2, \ldots, n-N+1$$

are used to predict sample $\hat{x}_{f,n+1}$. In this case, the LPC coefficients $a_i$ are not updated; they remain fixed at values estimated from $x_{n-i}, i = 1 \ldots M$. This sample prediction is repeated for the whole lost segment, *i.e.* $\hat{x}_{f,n+i}, i = 0, 1, \ldots, L-1$, where $L$ is the lost segment length. $\hat{x}_{f,n+i}$ can generally be expressed as

$$\hat{x}_{f,n+i} = -1 * \sum_{j=1}^{N} a_j \hat{x}_{f,n+i-j} \tag{2}$$

and is depicted in Fig. 2. Each delay elements are initialised to $x_i, i = n-1, n-2, \ldots, n-N$, and from then on, $\hat{x}_{f,n+i}, i = 1, 2, \ldots$ can be predicted recursively.

As the prediction is repeated, the gain of the predicted speech was shown to gradually decrease. Thus, we introduced a linearly increasing gain $G_f$ starting from 1.0 at the start of the lost segment, and saturating at $G_{max}$ at the end of the segment. This gain is applied to the predicted speech samples. We used an empirical value of 1.8 for $G_{max}$.

The above prediction can be applied in the backward direction as well using the packet received after

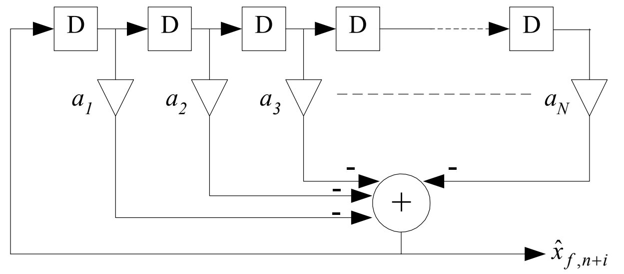


**Fig. 2** Block diagram of the LPC sample estimation algorithm.

the lost segment. Following the above notation, this can be written as:

$$\hat{x}_{b,n} = -1 * \sum_{i=1}^{N} b_i x_{n-i+N+1} \tag{3}$$

where $\hat{x}_{b,n}$ is the backward predicted sample, and $b_i$ is the backward LPC coefficient calculated from $K$ samples succeeding the lost packet, $x_{n+j}, j = 1, 2, \ldots, K$ where $K$ is the analysis window length. The LPC coefficients are also calculated using the Levinson-Durbin algorithm. Since the backward LPC requires packets after the lost packet, it is necessary to wait for the required packets to be received, *i.e.*, processing delay is required. Thus, we restricted the backward LPC analysis window to be exactly two packets, typically 160 samples at 8 kHz sampling, 10 msec packets. Moreover, there may be cases where one of these packets in the analysis window will not be available due to packet loss. If this is the case, we simply fall back into using only the forward prediction for concealment.

This prediction is repeated to obtain the whole lost segment in time-reversed order. Adaptive gain $G_b$ is also applied in a similar manner, this time starting from 1.0 at the end of the lost segment, to $G_{max}$ at the start of the segment in the backward direction.

### 2.2 Concealment of a single lost packet

When a single packet is lost, *i.e.* when a packet is lost, but both its preceding and succeeding packets are received intact, we can obtain both a forward LPC predicted estimate from samples in the preceding packet, and a backward LPC predicted estimate from samples in the succeeding packet. Thus,we have two estimates for the lost segment, *i.e.* $\hat{x}_{f,n+i}$ and $\hat{x}_{b,n+i}$. It can reasonably be assumed that the former ($\hat{x}_{f,n+i}$) is a better estimate of the earlier portions of the lost segment since the recursive repetition of the LPC estimation is smaller, while the latter ($\hat{x}_{b,n+i}$) is a better estimate of the later portions. Thus, we can combine the two estimates with a linear weight to obtain a single sample estimate:

$$\hat{x}_{n+i} = \omega_2 \hat{x}_{f,n+i} + \omega_3 \hat{x}_{b,n+i}, \tag{4}$$
$$\omega_2 = 1 - \alpha,$$
$$\omega_3 = \alpha$$

where $\alpha$ is a linearly increasing weight from 0 for $i = 0$, to 1 for $i = L - 1$.

### 2.3 Consecutive lost packets

When consecutive packets are lost, the processing will depend on the combination of normal reception or loss of both the preceding and succeeding packet.

(1) **Preceding packet received, succeeding packet lost:** In this case, only forward prediction will be used. LPC coefficients are calculated from the samples in the preceding packet. The lost segment is recursively calculated using the above LPC coefficients and samples in the preceding packet. The compensation gain $G_f$ is initially set to 1.0 and gradually increased to reach $G_{max}$ at the end of the segment.

(2) **Preceding packet lost, succeeding packet lost:** In this case, only forward prediction will be used. LPC coefficients are kept fixed at values calculated from speech samples in the last received packet. $G_f$ remains fixed at $G_{max}$.

(3) **Preceding packet lost, succeeding packet received:** In this case, only backward prediction will be used. Backward LPC coefficients are calculated from samples in the succeeding packet. The lost segment is calculated using the backward LPC coefficients and the samples in the succeeding packet recursively in the backward direction. The compensation gain $G_b$ is gradually decreased from $G_{max}$ at the beginning of the segment to 1.0 at the end.

### 2.4 Smoothing of adjacent received packets

In most cases, the predicted lost segments show some degree of perceived speech quality discontinuity with the adjacent received samples. This sudden change in quality seems to be a major source of perceived quality degradation. Thus, transition smoothing from received speech to the predicted samples is essential to reduce this degradation. An example of smoothing for two consecutive lost packet case is shown in Fig 3. There are a number of smoothing modes depending on the combination of whether the preceding and/or succeeding packet is lost, and whether the current packet is lost.

(1) **Current packet received, succeeding packet lost** (*e.g.* packet 1 in Fig 3): In order to smooth transition from received samples to the predicted samples, a forward-predicted sample is prepared for the received packet preceding the lost packet (packet 1). The LPC coefficients used here are calculated from the samples in the same packet,

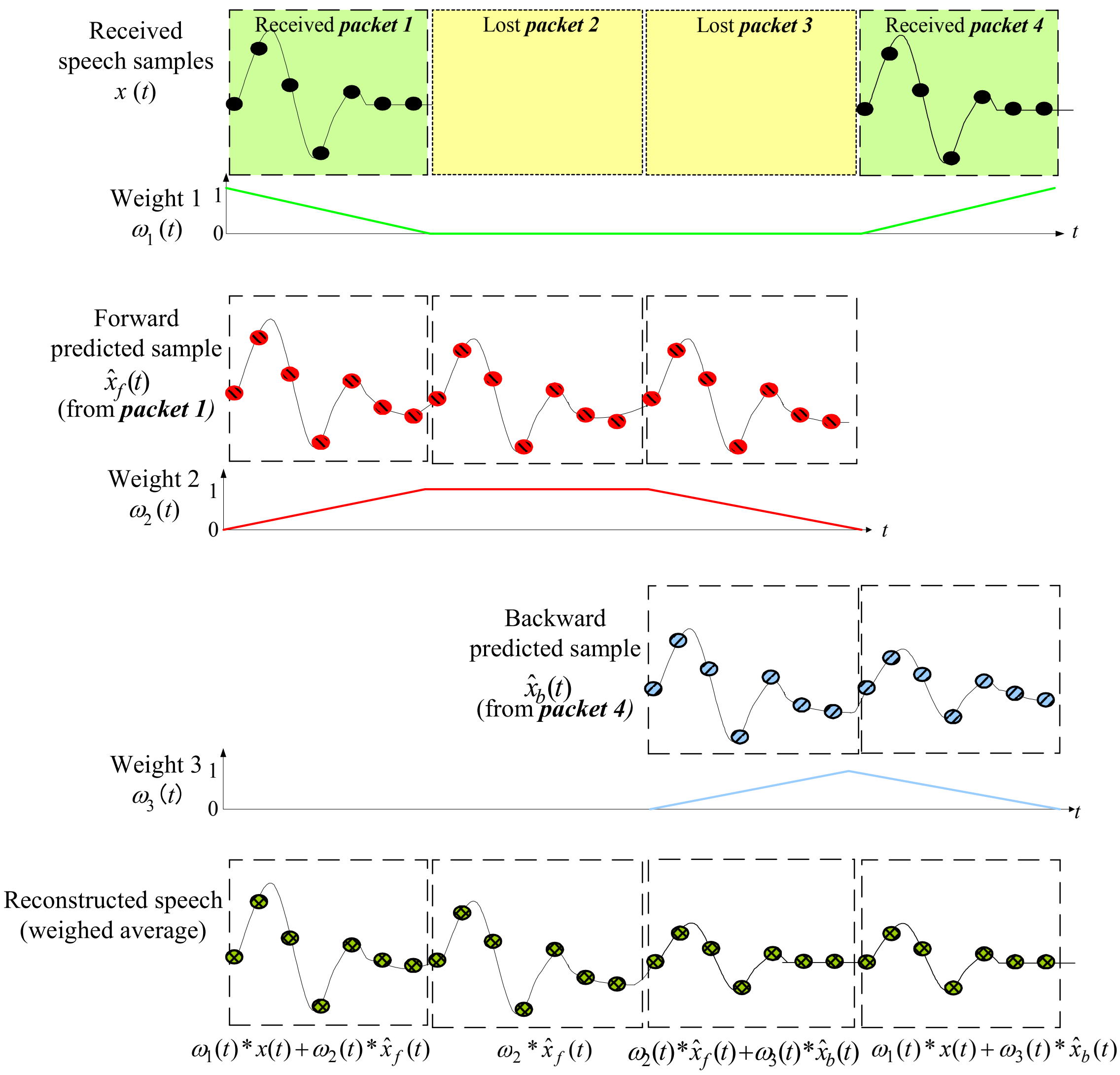


**Fig. 3** Smoothing between predicted and received samples.

*i.e.* packet 1. However, the samples included in the packet before packet 1 are used to predict these samples. This predicted segment is again linearly weighed and added to the actual received segment in packet 1; larger weight is given to the predicted segment $\hat{x}_{f,i}$ at samples in the smoothed segment just before the lost segment, while larger weight is given to the received samples $x_i$ at the beginning of the smoothed segment.

$$\begin{aligned} \hat{x}_i &= \omega_1 x_i + \omega_2 \hat{x}_{f,i}, \\ \omega_1 &= 1-\alpha, \\ \omega_2 &= \alpha \end{aligned} \tag{5}$$

where $\alpha$ is a linearly increasing weight from 0 for $i = 0$, to 1 for $i = L-1$.

(2) **Current packet received, preceding packet lost** (*e.g.* packet 4): The same smoothing can be applied to the samples after the lost segment, *i.e.* packet 4. Backward-predicted samples for the segment in packet 4 $\hat{x}_{b,i}$ are weighed and averaged with received samples $x_i$. This obviously requires more processing delay since two packets after the loss needs to be received in this case. Thus, in the listening tests in section 4, we decided not to use the backward-prediction smoothing described here; we simply use the received speech samples in packet 4. However, use of the above smoothing method does seem to improve the perceived quality to some degree at the cost of additional delay.

(3) **Current and preceding packet lost, succeeding packet received** (*e.g.* packet 3): In this case, both forward and backward predicted samples are available. The overlapped forward ($\hat{x}_{f,i}$) and backward predicted samples ($\hat{x}_{b,i}$) are averaged to smooth the discontinuity between the forward-predicted samples and backward-predicted samples. This is essentially the same processing that

is applied to the single lost packet case. Thus, this can be expressed as

$$\hat{x}_i = \omega_2 \hat{x}_{f,i} + \omega_3 \hat{x}_{b,i}, \quad (6)$$
$$\omega_2 = 1 - \alpha,$$
$$\omega_3 = \alpha$$

(4) **Current and succeeding packet lost, preceding packet received** (*e.g.* packet 2), **also, current, preceding and succeeding packet lost** (not shown in Fig): This is the only case where no smoothing is applied. The lost segment is simply replaced by forward predicted samples $\hat{x}_{f,i}$.

Extensive smoothing described here has shown to improve the subjective quality significantly in informal listening tests.

## 3. Quality comparison tests

We conducted subjective quality evaluation tests (mean opinion score tests) for the proposed algorithm. Twenty listeners with normal hearing evaluated all speech samples. The listeners were asked to rate each speech sample into one of the five standard categories which were assigned numerical scores shown below:

5: very good, 4: good, 3: normal, 2: bad, 1: very bad

The scores for each sample were then averaged to give the final opinion score.

We used samples in the ASJ continuous speech corpus [13], [14]. Two male and two female speakers, two samples each, were used. The sample consisted of read phonetically balanced Japanese sentences. Each was approximately 3 seconds long. The original sampling rate was 16k Hz, but was down-sampled to 8 kHz. All samples were in 16 bit linear PCM.

Packet length was assumed to be 10 [msec], or 80 samples at 8 kHz. Packets were randomly discarded. No statistical packet transmission model was assumed. The tested packet loss concealment schemes were as follows:

(1) Simple silence insertion (denoted **silence** in the results).
(2) ITU-T G.711 Appendix I (denoted **g711a1**). All parameters were set as described in the ITU standard. The ANSI C code accompanying the standard was used to simulate this algorithm.
(3) The proposed algorithm. We tested both the bidirectional LPC (denoted **bidirectional**) based prediction, and prediction in the forward direction only (denoted **forward**). As stated in section 2.4, we did not use the backward prediction smoothing, but all other smoothing modes described in section 2.4 were applied.

The choice of LPC analysis order was based on

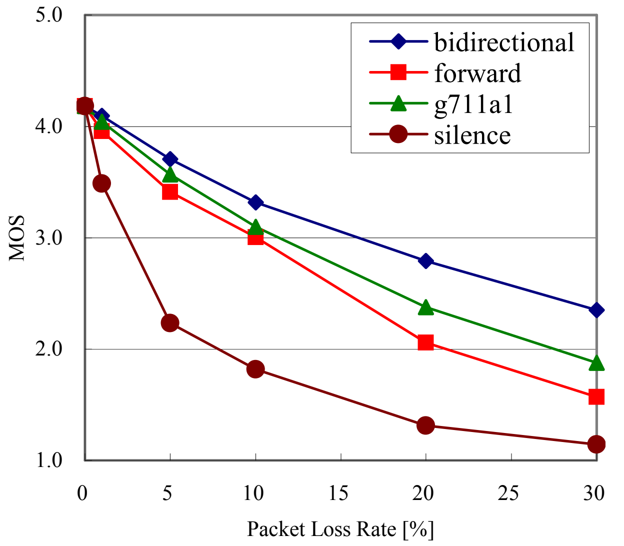


**Fig. 4** Mean opinion score vs. packet loss rate for clean speech.

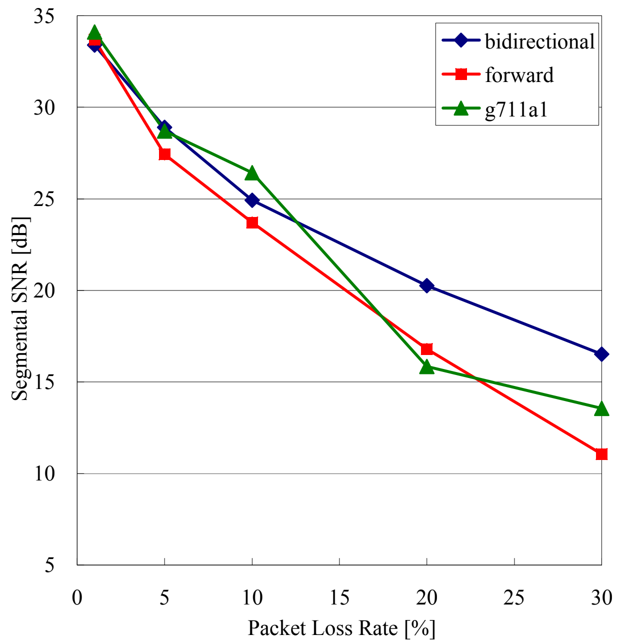


**Fig. 5** Segmental SNR vs. packet loss rate for clean speech.

the SNR estimate using the proposed concealment algorithm. Table 1 shows the SNR estimate for concealment of random packet loss rate of 10 % using forward only prediction. One-sided Hamming window was applied in all LPC analysis. Since the SNR is dependent on the lost packet position, the estimates shown are average of five different loss sequences with the same loss rate and the same speech utterance.

As can be seen in the table, the optimum LPC order seems to be speaker dependent. In the following experiments, we empirically chose an LPC order of 128 with analysis block length of 256 since this combination worked best on the male speech. However, careful

**Table 1** SNR for forward only loss compensation (10 % packet loss).

| LPC order | Analysis block length | SNR [dB] | |
|---|---|---|---|
| | | male speech | female speech |
| 32 | 64 | 9.20 | 11.22 |
| 64 | 128 | 9.60 | 12.40 |
| 96 | 192 | 9.44 | 11.52 |
| 128 | 256 | 9.74 | 11.20 |
| 256 | 512 | 9.46 | 10.52 |

reconsideration may be necessary.

Figure 4 shows the Mean Opinion Score (MOS) for packet loss rates between 0 and 30% with clean speech. As expected, **silence** insertion shows significant degradation with loss. The other methods show similar trends up to about 10% loss. However, above this rate, **bidirectional** LPC outperforms the G.711 Appendix I (**g711a1**) and the **forward** LPC. The **g711a1** starts to show synthetic quality due to long pitch repetition for high loss rates. On the other hand, the **forward** LPC shows some "background ringing" perhaps due to over-enhancement of the formant due to the long recursive LPC. We experimented with bandwidth expansion [15] to reduce this effect, but found the improvement to be marginal. Further optimisation is necessary for improvement.

Figure 5 shows the segmental SNR for one of the speech utterances used in the subjective quality tests. Interestingly, the objective measure roughly agrees with the subjective quality test results in this case. The **g711a1**, **bidirectional** and **forward** LPC all show similar segmental SNR up to 10 % loss, and above this loss, **bidirectional** LPC shows higher segmental SNR than the other two concealment algorithms. Thus, segmental SNR may provide a fair estimate of the subjective quality of the concealed speech.

We also conducted subjective quality tests with degraded speech inputs. The same speech samples as described above were used in these tests as well. Figure 6 shows the MOS vs. packet loss rate for speech coded and decoded using the ITU-T G.729 codec [16]. This kind of setup may occur in networks where bandwidth-restricted wireless networks and/or cellular networks and VOIP links are connected in tandem, where speech may first go through low bit rate speech coding and decoding processes in order to reduce the transmission bit rate requirement, and then goes through a packet transmission network. This test was intended to find out the effect of synthetic artifacts introduced by the codec on the concealment algorithm. The G.729 codec was simulated using the ANSI C code accompanying the ITU standard. The results seem similar to Fig. 4, but the difference between **bidirectional** LPC and **g711a1** is smaller. This may be due to the fact that the G.729 codec shows some low-pass characteristics, which consequently serves to make the perceived degradation due to pitch repetition less apparent.

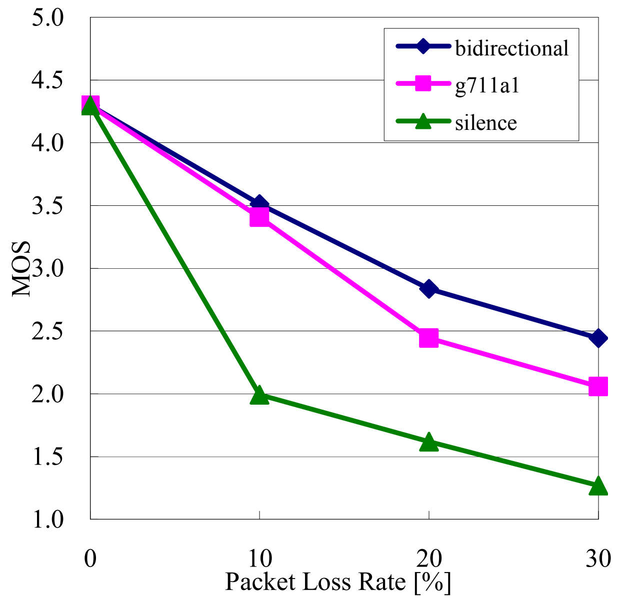


**Fig. 6** Mean opinion score vs. packet loss rate for G.729 transcoded speech.

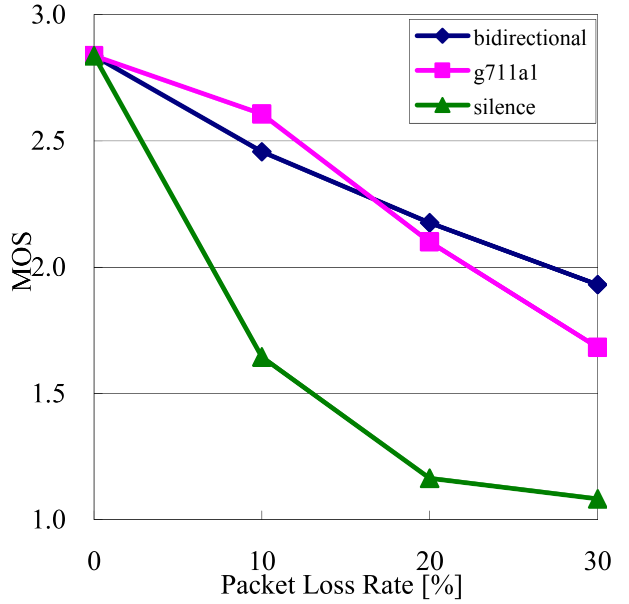


**Fig. 7** Mean opinion score vs. packet loss rate for speech mixed with babble.

Figure 7 shows MOS vs. packet loss rate for speech mixed with babble (multi-speaker noise). This test was intended to find out the effect of surrounding speech noise on the concealment algorithm. Babble was added to the clean speech sample at 10 dB SNR. Other SNR was unofficially tested, but the difference in the overall effect of babble noise seemed to be negligible. Again, the **bidirectional** LPC seems to maintain better quality than the **g711a1** at higher loss rates, although at 10% loss, **g711a1** shows slightly higher quality. The repetition of pitch period with babble seems to show synthetic quality with **g711a1**, which seems to degrade the perceived quality. We also start to notice the effect of pitch extraction errors on **g711a1** caused by the low SNR input. The pitch errors become quite annoying at higher loss rates, which degrades the perceived quality significantly.

## 4. Discussions

So far, we have compared our proposed algorithm based on LPC prediction to the ITU standard G.711 Appendix I based on pitch repetition. The comparisons can be summarised as follows:

(1) Generally, the G.711 algorithm requires a buffer with more than two pitch periods (maximum pitch) for pitch extraction - the reference pitch interval, and the adjacent pitch interval to calculate the correlation. The G.711 Appendix I standard recommend a buffer of 390 samples. On the other hand, the proposed algorithm needs only a buffer of about the pitch interval for prediction. Even if only a shorter interval is available, the prediction accuracy degrades gracefully, as we have seen in Table 1.

(2) For reasons stated in (1) above, the proposed algorithm achieves backward prediction by restricting the associated delay through shorter buffer length of two packets, or 160 samples, at the price of prediction accuracy. If this is still too much delay, it is possible to further reduce the buffer length at the cost of still less prediction accuracy. With the G.711 algorithm, the associated delay would need to be larger and fixed, the full 390 samples, to achieve backward concealment.

(3) A single parameter, the pitch period, plays a critical role in the performance of the G.711 algorithm. If the pitch accuracy is low, the concealed speech quality will degrade significantly. With the proposed LPC algorithm, since the concealment relies on added contribution of all LPC coefficients, as the LPC accuracy becomes lower, the degradation is generally graceful.

(4) Since the G.711 algorithm simply repeats the pitch interval, as the repetition gets longer, the concealed speech quality tends to become synthetic, and thus quite annoying. With the proposed algorithm, as the prediction is repeated, the concealed speech generally will tend to include more white-like noise, but will tend not to be as synthetic.

## 5. Conclusion

In this paper, we proposed a speech packet loss concealment algorithm, which uses LPC to estimate the lost segment. When both forward LPC from the packet preceding the lost packet and backward LPC from the packet succeeding is used, the subjective quality was shown to outperform the packet loss concealment in ITU-T G.711 Appendix I [8]. The proposed algorithm showed superior quality especially at loss rates above 10%.

The algorithm requires a high LPC order to be effective. Thus, the computational complexity is fairly high. However, use of block update or gradient update may still be effective while considerably lowering the complexity of the LPC coefficient update.

Also, the LPC order was shown to be speaker dependent. It is reasonable to assume that the optimum order is correlated with the pitch period. Thus, adaptive LPC order according to the estimated pitch period may improve the LPC gain, which in turn may improve the concealed speech quality.

When consecutive packets are lost, LPC is applied repeatedly to estimate the lost segment, resulting in "synthetic" quality. As stated in section 3, bandwidth expansion may improve this quality. However, it seems that the expansion factor may need to be adjusted according to the amount of LPC repetitions applied. That is, bandwidth expansion can be applied to estimated LPC coefficients as follows:

$$\tilde{a}_i = a_i \rho^i$$

where $\rho$ ($\rho \leq 1$) is the expansion factor, and $\tilde{a}_i$ is the bandwidth expanded LPC coefficients. When $\rho = 1$, bandwidth expansion does not take effect, and for smaller $\rho$, the expansion bandwidth becomes significant. Thus, $\rho$ can be kept close to 1 while the repetition is still small, and can be gradually decreased as the repetition is continued, thereby gradually increasing the expansion.

The bidirectional LPC shows improved loss concealment performance, but will require additional processing delay since it requires the speech signal in the packet succeeding the lost packet. However, it was shown that for loss rates up to 10%, the subjective quality is not significantly different from forward only LPC, which does not require the added delay. Thus, it may be possible to limit the overall delay by using a bimodal prediction scheme based on the observed packet loss rate; when the loss rate is below 10%, the forward LPC is used, while for loss rates above 10%, the LPC is switched to bidirectional. This should provide a balanced solution between the speech quality and processing delay. The switch between modes can be accomplished during inter-word pauses fairly transparently. Expanding or shrinking the pause slightly can also accomplish this delay adjustment. There have been reports that the human ear is immune to such alterations if the percentage of the alterations is limited to a small proportion [17].

## Acknowledgement

The authors would like to thank The Telecommunications Advancement Foundation for their support. They also thank the members of the Nakagawa Laboratory of the Electrical Engineering Department at the Yamagata University for their support and participation in the experiments.

### References

[1] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," IEEE Network, vol.12, no.5, pp.40–48, Sept./Oct. 1998.
[2] V. Hardman, M. Sasse, M. Handley, and A. Watson, "Reliable audio for use over the internet," Proceedings of the International Networking Conference, June 1995.
[3] B. Wah, X. Su, and D. Lin, "A survey of error-concealment schemes for real-time audio and video transmissions over the internet," Proceedings of the IEEE International Symposium on Multimedia Software Engineering, Dec. 2000.
[4] D.J. Goodman, G.B. Lockhart, O.J. Wasem, and W.C. Wong, "Waveform substitution techniques for recovering missing speech segments in packet voice communications," IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-34, no.6, pp.1440–1448, Dec. 1986.
[5] 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 Trans. Acoust., Speech, Signal Processing, vol.ASSP-36, no.3, pp.342–348, March 1988.
[6] H. Sanneck, A. Stenger, K. Younes, and B. Girod, "A new technique for audio packet loss concealment," Proceedings of the IEEE Global Internet, 1996.
[7] M. Lee, I. Zitouni, and Q. Zhou, "An n-gram model approach for packet loss concealment," Proceedings of the International Conference on Spoken Language Processing (Interspeech-ICSLP), Oct. 2004.
[8] ITU-T Recommendation G.711 Appendix I, "A high quality low complexity algorithm for packet loss concealment with G.711," 1999.
[9] ANSI Recommendation T1.521a-2000 (Annex B), "Packet loss concealment for use with itu-t recommendation G.711," 2000.
[10] E. Gündüzhan and K. Momtahan, "A linear prediction based packet loss concealment algorithm for pcm coded speech," IEEE Transactions on Speech and Audio Processing, vol.9, no.8, pp.778–785, Nov. 2001.
[11] K. Kondo and K. Nakagawa, "A packet loss concealment method using recursive linear prediction," Proceedings of the International Conference on Spoken Language Processing (Interspeech-ICSLP), Oct. 2004.
[12] S. Haykin, Adaptive Filter Theory, Prentice-Hall, Upper Saddle River, 1996.
[13] Japan Information Processing Development Corporation, "ASJ continuous speech corpus for research," 1991.
[14] S. Itahashi, "On recent speech corpora activities in japan," Journal of the Acoustical Society of Japan (E), vol.20, no.3, pp.163–170, May 1999.
[15] Y. Tohkura, F. Itakura, and S. Hashimoto, "Spectral smoothing technique in parcor speech analysis-synthesis," IEEE Transactions on Acoustics, Speech and Signal Processing, vol.ASSP-26, no.6, pp.587–596, Dec. 1978.
[16] ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)," Mar. 1996.
[17] W. Montgomery, "Techniques for packet voice synchronization," IEEE Journal on Selected Areas in Communications, vol.SAC-1, no.6, pp.1022–1028, Dec. 1983.

**Kazuhiro Kondo** received the B.E., the M.E., and the Ph.D. degrees from Waseda University in 1982, 1984, and 1998, respectively. From 1984 to 1992, he was with the Central Research Laboratory, Hitachi Limited, Kukubunji, Tokyo, Japan. During this time, he was engaged in research on speech and video coding systems. From 1992 to 1995, he was with the Texas Instruments Tsukuba R & D Center Limited, Tsukuba, Ibaraki Japan. From 1995 to 1998, he was with the DSP R & D Center, Texas Instruments Inc., Dallas, Texas, USA. During this time, he worked on speech recognition systems and multimedia signal processing. In 1999, he joined the Faculty of Engineering at Yamagata University, Yonezawa, Yamagata, Japan. His current interests include broad aspects of speech and audio signal processing, multimedia signal processing, and speech and audio quality evaluation methods. Dr. Kondo is a member of the IEEE, the International Speech Communication Association (ISCA), and the Acoustical Society of Japan.

**Kiyoshi Nakagawa** received his B.E., the M.E., and the Ph.D. degrees from Osaka University in 1968, 1970, and 1980, respectively. He joined the Electrotechnical Laboratory of the NTT Corp., Musashino, Tokyo, Japan in 1970, and has worked on digital optical communication networks among numerous digital communication technology. He joined the Faculty of Engineering at Yamagata University, Yonezawa, Yamagata, Japan in 1996 as Professor in information communication network technology, and has been engaged in broad aspects of optical communication technology. He has authored numerous technical publications, including several books. He has received the IEE Oliver Lodge Award, the IEE Electronics Letters Award, among other awards he has received from the IEICE. Professor Nakagawa is a member of the IEEE, the Optical Society of America (OSA), and the Laser Society of Japan.