# Packet Loss Concealment Estimating Residual Errors of Forward-Backward Linear Prediction for Bone-Conducted Speech

Ohidujjaman[1], Nozomiko Yasui[2], Yosuke Sugiura[3], Tetsuya Shimamura[4], Hisanori Makinae[5]
Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan[1,3,4]
Computer Science and Engineering, Daffodil International University, Savar, Dhaka 1216, Bangladesh[1]
National Institute of Technology, Kisarazu College, Chiba 292-0041, Japan[2]
Third Information Science Section, National Research Institute of Police Science, Chiba 277-0882, Japan[5]

***Abstract*—This study proposes a suitable model for packet loss concealment (PLC) by estimating the residual error of the linear prediction (LP) method for bone-conducted (BC) speech. Instead of conventional LP-based PLC techniques where the residual error is ignored, we employ forward-backward linear prediction (FBLP), known as the modified covariance (MC) method, by incorporating the residual error estimates. The MC method provides precise LP estimation for a short data length, reduces the numerical difficulties, and produces a stable model, whereas the conventional autocorrelation (ACR) method of LP suffers from numerical problems. The MC method has the effect of compressing the spectral dynamic range of the BC speech, which improves the numerical difficulties. Simulation results reveal that the proposed method provides excellent outcomes from some objective evaluation scores in contrast to conventional PLC techniques.**

***Keywords*—*Autocorrelation method; bone-conducted speech; modified covariance method; packet loss concealment; residual error***

## I. Introduction

Recently, much attention has been paid to the use of BC speech in the field of speech signal processing. BC speech travels through the vibration of skull bone, skin, and soft tissue as its pathway. A vibration sensor is used for BC microphone decoration. BC microphone captures the vibrations and converts them into an electric signal. Thus, BC speech is rarely affected by ambient noise. Therefore, BC speech delivers advantages over air-conducted (AC) speech in some cases. In [1], BC speech was synthesized based on the least squares (LS) method. Pitch detection for BC speech is discussed in [2]. In [3], BC speech was utilized with AC speech for speaker recognition. Speaker verification is also described in [4]. To improve the quality of speech, several types of algorithms have been derived for BC speech [5]-[7]. Rahman *et al*. [8] considered a noisy environment for BC speech and derived a noise-robust LP method. However, there are few works on PLC for BC speech. In [9], the MC method was derived for the PLC technique in noisy cases, in which how the transmitted BC speech can be reconstructed with higher accuracy by reducing the spectral dynamic range was investigated.

In recent years, voice over internet protocol (VoIP) has considerably increased due to the frequent use of internet telephony. VoIP utilizes a packet-switch network instead of circuit switching. Speech voice is degraded in packet-switched networks because of the delayed packet, additive noise, network congestion, packet arrival jittering, and bit errors [10]. The more bit errors finally cause the packet loss. Also, the high-speed network cannot conciliate a packet loss issue, resulting in degraded quality of speech [11]. Therefore, lost packets must be recovered to eschew speech quality degradation, which is known as PLC. Several methods have been recommended for PLC such as 2-sided LP, recovery sub-codec (RSC), deep neural network (DNN), hidden Markov model (HMM), adaptive recurrent neural network (A-RNN)-based PLC, and so on [11]-[17]. One representative of the classical PLC methods may be an LP-based PLC extensively utilized in global systems for mobile (GSM) technology [18]. Audio jitter buffer, network equalizer (NetEQ), is a web real-time communications (WebRTC's) default PLC whose performance is degraded in higher packet loss cases. A state-of-the-art PLC approach is WaveNetEQ (generative model) PLC built on DeepMind's wave recurrent neural network (WaveRNN) technology, which is deployed instead of the NetEQ PLC technique for accurate speech reconstruction [19]. Currently, the WaveNetEQ PLC approach is implemented in Google Due [20]. Among the above approaches, the DNN, HMM, WaveNetEQ, and A-RNN PLC techniques are state-of-the-art, but the large volume of speech data is required to be employed to train them, which is an inevitable hands-on problem in some circumstances, and they also face higher computational complexity. Contrariwise, the LP-based PLC technique provides better performance with a notably lower computational complexity. Few works are known for the LP-based PLC technique of BC speech in the literature. Therefore, there is sufficient scope to enhance the performance of the LP-based PLC approach.

In [21], a PLC method was proposed through sinusoidal extrapolation with pulse code modulation (PCM) coded at the recipient side. This PCM-coded PLC technique extrapolates filter coefficients and LP residuals from the last packet correctly received while speech packets are lost. In [22], the ACR method of LP was proposed for the PLC technique where the forward direction and bidirectional estimation processes were employed for 10% and more than 10% packet losses, respectively. The Levinson-Durbin (LD) algorithm was deployed to estimate the LP coefficients, and a better PLC was obtained by setting a large LP order. In [23], the PCM-coded PLC technique was proposed, where the predictive error signal of the prior packet and the attained pitch period are employed to stimulate an excitation for packet loss. The forthcoming packet was omitted for backward estimation, resulting in signal attenuation. In [11], a recovery sub-codec (RSC) PLC technique was proposed that employed low delay code-excited LP (LD-CELP) where the ACR method of LP was deployed. BC speech possesses an expanded spectral dynamic range, which leads to a large eigenvalues expansion of the ACR matrix in the ACR method employed in the RSC-based PLC technique. To evade the eigenvalues expansion, we utilize the forward-backward LP (FBLP), which is often called the MC method [9], instead of the ACR method.

In the above conventional PLC techniques, LP-based methods are usually implemented. According to the existing PLC approaches [9][11][21]-[23], none of them incorporated the residual error estimates. Commonly, the conventional PLC method utilizes the ACR method where the L-D algorithm is employed. The ACR method provides degraded performance in an expanded spectral dynamic range of an input signal. On the other hand, the MC method reduces the eigenvalue expansion of BC speech, resulting in an improved ill-condition. From this point of view, we consider the MC method by including the residual error estimates. This study emphasizes the residual error estimation in both forward and backward packets to the lost packet. The predictive residual errors are added in each lost packet exclusively during the packet loss estimation. For the simulations, some objective evaluations are done, and it is demonstrated that the proposed PLC technique performs better than the conventional PLC approaches.

The following is the organizational structure of this paper. Section I includes the introduction. Section II provides an overview of the conventional method, while Section III explains the proposed approach. Section IV outlines the experiments. Results are discussed in Section V. Section VI concludes the paper.

## II. Conventional PLC Method

In [9], the MC method was derived for the PLC technique in noisy cases, in which how the transmitted BC speech can be reconstructed with higher accuracy by reducing the spectral dynamic range was investigated. The AC speech was replaced by the BC speech to mitigate the additive noise problem, and the MC method was employed instead of the AC method to improve the ill-conditioned difficulties. The MC method of LP was derived from a least-squares (LS) method for estimating the LP coefficients by the concurrent minimization of the FBLP squared errors. In [9], without residual error estimates, the basic form of LP was defined as

$$s(n) = -\sum_{k=1}^{p} \alpha(k) s(n-k) \tag{1}$$

where $\alpha(k)$ and $p$ correspond to the LP coefficients and the LP order, respectively, and $s(n-k)$ denotes the prior data samples. The total squared errors were expressed as

$$\mathcal{E} = \mathcal{E}_f + \mathcal{E}_b \tag{2}$$

where $\mathcal{E}_f$ and $\mathcal{E}_b$ denote the forward and backward squared errors, respectively, and they are defined as

$$\mathcal{E}_f = \sum_n (s(n) + \sum_{k=1}^{p} \alpha(k) s(n-k))^2 \tag{3}$$

$$\mathcal{E}_b = \sum_n (s(n) + \sum_{k=1}^{p} \alpha(k) s(n+k))^2, \tag{4}$$

respectively. In (3) and (4), the speech signal sample at time $n$, $s(n)$, is predicted by forward and backward LP filters with the order of $p$. When the LP coefficients $\alpha(1)$, $\alpha(2)$, …, $\alpha(p)$ are represented in a vector form as

$$\boldsymbol{\alpha} = [\alpha(1), \alpha(2), \ldots, \alpha(p)]^T \tag{5}$$

where $T$ denotes transpose. The MC method was derived to obtain the LP coefficients as follows. For the unknown vector $\boldsymbol{\alpha}$, the residual error vector, $\boldsymbol{\epsilon}$, is defined as

$$\boldsymbol{\epsilon} = \mathbf{R}\boldsymbol{\alpha} - \mathbf{r} \tag{6}$$

where $\mathbf{R}$ and $\mathbf{r}$ correspond to the observation matrix and measurement vector, respectively. We can define the least squares criterion as

$$L = \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} \tag{7}$$

$L$ is expanded as follows:

$$L = (\mathbf{R}\boldsymbol{\alpha} - \mathbf{r})^T (\mathbf{R}\boldsymbol{\alpha} - \mathbf{r})$$

$$= (\mathbf{R}\boldsymbol{\alpha})^T (\mathbf{R}\boldsymbol{\alpha}) + \mathbf{r}^T\mathbf{r} - (\mathbf{R}\boldsymbol{\alpha})^T \mathbf{r} - \mathbf{r}^T (\mathbf{R}\boldsymbol{\alpha}) \tag{8}$$

where $\mathbf{r}^T(\mathbf{R}\boldsymbol{\alpha})$ is represented by $(\mathbf{R}\boldsymbol{\alpha})^T\mathbf{r}$. Thus, equation (8) is rewritten as

$$L = \mathbf{R}^T \boldsymbol{\alpha}^T (\mathbf{R}\boldsymbol{\alpha}) + \mathbf{r}^T\mathbf{r} - 2\boldsymbol{\alpha}^T \mathbf{R}^T \mathbf{r} \tag{9}$$

Since (9) is a quadratic form of $\boldsymbol{\alpha}$, by differentiating (9) with respect to $\boldsymbol{\alpha}$ and setting it to zero, we obtain the following forms as

$$2\mathbf{R}^T\mathbf{R}\boldsymbol{\alpha} - 2\mathbf{R}^T\mathbf{r} = 0 \tag{10}$$

$$\boldsymbol{\alpha} = (\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T\mathbf{r} \tag{11}$$

$$\boldsymbol{\alpha} = \mathbf{C}^{-1}\mathbf{c} \tag{12}$$

where $\mathbf{C} = \mathbf{R}^T\mathbf{R}$ and $\mathbf{c} = \mathbf{R}^T\mathbf{r}$ correspond to the MC matrix and MC vector, respectively. In this PLC technique, the MC method reduced the eigenvalue expansion, which improved the ill-condition. This is because there is a centrosymmetric characteristic in the MC matrix, $\mathbf{C}$, resulting in better speech reconstruction in the PLC technique. However, in this PLC technique, the predictive residual error is ignored, which is expected to be estimated and incorporated in the case of optimal speech reconstruction for the PLC.

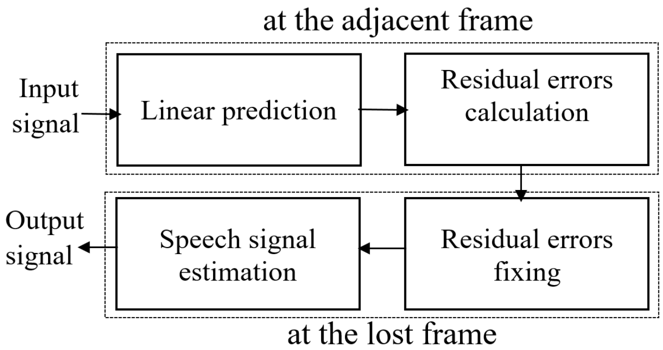


Fig. 1. Block diagram of the proposed method.

## III. Proposed Method

This paper proposes a progressive PLC technique focusing on residual error signal which is an effective factor to provide an improvement of the LP method. The PLC technique in [9] avoided residual error estimates in the basic LP Eq. (1), in which residual error must be incorporated for accurate speech reconstruction in the PLC technique. Therefore, we modify Eq. (1) by incorporating the residual error estimates as

$$s(n) = -\sum_{k=1}^{p} \alpha(k) s(n-k) + \epsilon(n) \tag{13}$$

where $\epsilon(n)$ denotes the predictive residual error. The basic block diagram of the proposed method is shown in Fig. 1.

### A. Residual Error Calculation

In the conventional LP-based PLC method, the long packet loss causes amplitude attenuation in the estimated speech signal, resulting in degraded speech quality. Therefore, the proposed method focuses on the estimation model of the speech signal in the packet loss and considers causality with the attenuation. The conventional LP form (1) is expanded, and the transition is performed then (1) becomes (14) as follows:

$$s(n) = -\alpha(1)s(n-1) - \alpha(2)s(n-2) - \cdots - \alpha(p)s(n-p)$$
$$\Longrightarrow \; s(n) + \alpha(1)s(n-1) + \cdots + \alpha(p)s(n-p) = 0 \tag{14}$$

Eq. (14) is a linear first-order combination of sample points, held in the LP method. By comparing Eq. (14) with Eq. (13), it is noted that the conventional estimation model Eq. (14) assumes the residual error, $\epsilon(n)$, to be 0. Thus, the estimated speech waveform is attenuated since the residual error is ignored in the speech estimation. Therefore, the residual error must be estimated that originally occurs in the lost segment from the adjacent packets, and needs to be incorporated into the speech estimation model to enhance the performance of the conventional LP-based PLC technique. The residual error is the difference between the true speech signal and the estimated speech signal. The LP error is calculated in the adjacent packets for the forward and backward directions. Through the LP errors $\epsilon(n)$ at the adjacent packets, the estimated LP residual error, $\hat{\epsilon}(n)$, in the lost packet is calculated as

$$\hat{\epsilon}(n) = s(n) - \hat{s}(n); (n = 0, 1, \ldots, L-1) \tag{15}$$

where $L$ denotes the length of a packet loss. The error estimation takes advantage of the fact that the LP error has the same periodicity as the original speech signal. Since the periodicity is confirmed, the LP error of the packet loss is calculated through the concept of the PWR method. The conventional PWR method processes the speech waveform, but in this case, the processing is performed for the LP error, hence the input to the MC function changes to an LP error as

$$R(m) = -\sum_{k=0}^{L-1} \epsilon(k)\ \epsilon(k-m) \tag{16}$$

**Algorithm 1** : Residual error calculation

**fun** $error \leftarrow$ ***residualError*** $(Tmp, PW, AR_{coef}, M)$
    ***for*** $i \leftarrow M+1 : PW$
        $xEsti \leftarrow -AR_{coef}^{T} \times Tmp(i-1\ to\ i-M)$
    ***end***
    ***if*** $size\ (xEsti, 1) \neq size(Tmp, 1)$
        $xEsti \leftarrow xEsti^{T}$
    ***end***
    $error \leftarrow Tmp - xEsti$
***end***

Algorithm 1 shows the residual error estimation process. In Algorithm 1, $Tmp$ and $PW$ correspond to samples for error estimate and window length for pitch extraction, respectively, and $AR_{coef}$ denotes autoregressive (AR) coefficients of LP order $M$. Finally, the estimated LP error $\hat{\epsilon}(n)$ is incorporated into the conventional estimation model to conceal the lost packet as follows:

$$\hat{s}(n) = -\sum_{k=1}^{p} \alpha(k)\ s(n-k) + \hat{\epsilon}(n) \tag{17}$$

where $\hat{s}(n)$ and $s(n-k)$ correspond to predicted samples and previous samples, respectively.

### B. Forward-Backward Packet Estimation and Addition

Eq. (17) with error estimate is concisely expressed as

$$\hat{s}(n) = -\boldsymbol{\alpha}\mathbf{s}^{T} + \hat{\epsilon}(n) \tag{18}$$

where $\boldsymbol{\alpha} = \big(\alpha(1) \;\; \alpha(2) \;\; \ldots \;\; \alpha(p-1) \;\; \alpha(p)\big)$,
$\mathbf{s} = \big(s(n-1) \;\; s(n-2) \;\; \ldots \;\; s(n-p)\big)$,
$\hat{\epsilon}(n) = \epsilon^{f}(n) + \epsilon^{b}(n)$, and
$\mathbf{s}$ and $\boldsymbol{\alpha}$ correspond to the input data vector and LP parameter vector, respectively. The LP coefficients in $\boldsymbol{\alpha}$ are obtained from the preceding and succeeding packets to the packet loss. Eq. (18) is more specifically expressed as

$$\hat{s}(n) = -\boldsymbol{\alpha}[1:p]\ \mathbf{s}^{T}[(n-1):(n-p)] + \hat{\epsilon}(n) \tag{19}$$

Eq. (19) is performed $L$ times for the forward-backward predictions to retrieve lost samples in the packet loss. For instance, lost samples are estimated in the forward prediction as follows:

$$s^{f}(1) = -\boldsymbol{\alpha}[1:p]\ \mathbf{s}^{T}[n:(n-p+1)] + \hat{\epsilon}(n)$$
$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$
$$s^{f}(L) = -\boldsymbol{\alpha}[1:p]\ \mathbf{s}^{T}[n+L-1:n+L-p] + \hat{\epsilon}(n+L-1) \tag{20}$$

**Algorithm 2** : Forward-backward packets addition

**fun** $s \leftarrow$ ***addition*** $(FEF, BEF)$
  $L \leftarrow length(FEF)$
  $s \leftarrow zeros([L \;\; 1])$
  ***for*** $i \leftarrow 1$ *to* $L$
    $\omega \leftarrow (i-1)/(L-1)$
    $s(n+i) \leftarrow FEF(i) \times (1-\omega) + BEF(i) \times \omega$
  ***end***
***end***

In the backward prediction, lost samples are estimated as,

$$s^b(1) = -\boldsymbol{\alpha}[1:p]\ \mathbf{s}^T[n:(n+p-1)] + \hat{\epsilon}(n)$$
$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$
$$s^b(L) = -\boldsymbol{\alpha}[1:p]\ \mathbf{s}^T[n-L+1:n-L+p] + \hat{\epsilon}(n-L+1) \quad (21)$$

In Eq. (20) and Eq. (21), the estimated errors are incorporated, which pays an advantage to the proposed method over the conventional PLC methods. The estimated packets (forward and backward predictions) are stored in the buffer until they are added. We add these packets from the buffer to restore packet loss. Algorithm 2 shows the estimated packet addition process. In Algorithm 2, $FEF$ and $BEF$ indicate the forward and backward estimated frames, respectively, and $L$ indicates the total number of samples in the lost packet. The linear weighting, $\omega$, is defined as $0 \leq \omega \leq 1$. Fig. 2 represents a diagram of the 2-sided MC method where estimated residual errors are incorporated in both forward and backward predictions. The forward prediction is done from the previous packet of the packet loss, and the backward prediction is from the future packet of the packet loss as shown in Fig. 2(c). Then, the speech signal is reconstructed by adding forward and backward estimated packets as shown in Fig. 2(d). The proposed method using the forward and backward adjacent packets of the lost frame inherits the advantage of the LP method that the distortion does not occur in the estimated speech, and it overcomes the drawback where the amplitude of the estimated speech is attenuated.

## IV. Experimental Conditions

We employed the National Research Institute of Police Science (NRIPS) database [24], from where the BC speech is implemented in this paper. In this database, the BC microphone TEMCO EML-1-A was used to measure BC speech, and a digital recorder EDIROL R-4 recorded the BC speech. We used 250 male and 250 female speech utterances of 50 different sentences pronounced by five males and five females for BC speech in the simulation. The evaluated result may differ for diverse speakers though the packet loss rate is identical. Therefore, the estimated score of each experiment of 250 times was averaged. As the objective evaluation, we utilized perceptual evaluation of speech quality (PESQ), log-spectrum distortion (LSD), and the log area ratio (LAR). The simulation details are shown in Table I.

## V. Results Discussion

In this section, we conducted several comparisons between the conventional and proposed methods.

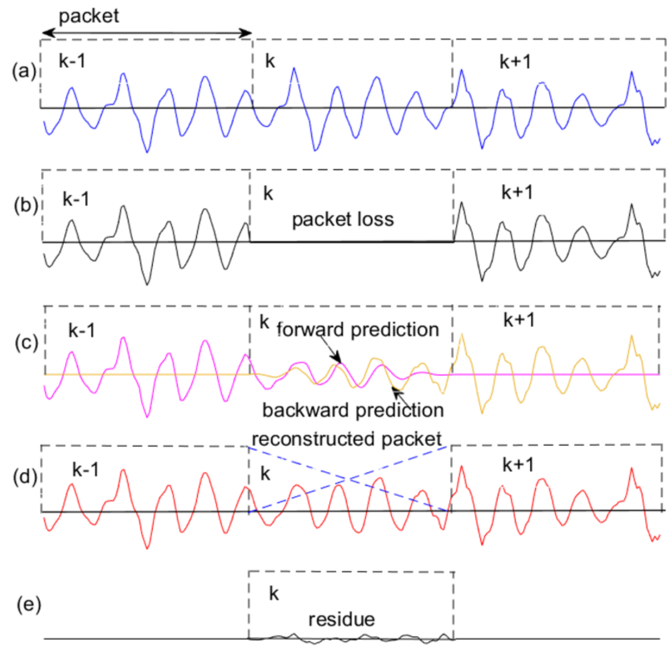


Fig. 2. Schematic diagram of 2-sided LP: (a) Original speech (b) Degraded speech (c) Forward and backward predictions (d) Reconstructed speech (e) Residue.

TABLE I. Simulation Specifications

| Items | Specifications |
|---|---|
| Accent | Japanese pronunciations |
| Speaker | Male and female |
| Sampling frequency | 8 kHz |
| Quantization | G.711 $\mu$-law |
| Speech signal length | 4-5 sec |
| Length of packet | 10 ms |
| Loss rate | 10% and 30% |
| LP order | 12 |
| LP analyzing window | 20 ms |
| Window type | Rectangular |

### A. Estimated Waveform

The proposed PLC method avoids the use of linear gain to the restored speech signal, whereas the conventional PLC methods apply a linear gain of 1.1 to 1.8 [9][11][21]-[23]. The reconstructed speech waveform for the MC method without error estimate (conventional method) and the MC method with error estimate (proposed method) in the PLC technique are shown in Fig. 3, where the estimated speech waveform obtained by the proposed method is well-matched to the original transmitted speech waveform. This is because the incorporated residual error signal influences the quality of the reconstructed speech signal.

### B. PESQ Score

The perceptual evaluation of speech quality (PESQ) was utilized as the objective evaluation to measure the excellence of speech for VoIP. The PESQ provides the quality of speech

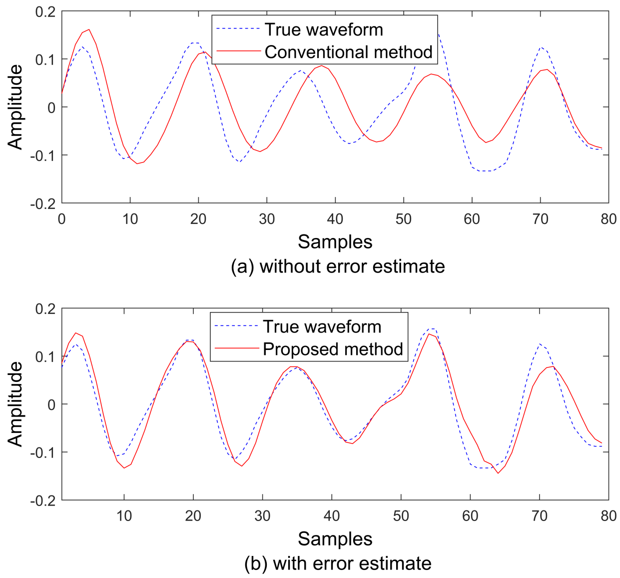


Fig. 3. Estimated waveform.

TABLE II. AVERAGE PESQ SCORES OF DIFFERENT METHODS

| Methods | 10% packet loss | | 30% packet loss | |
|---|---|---|---|---|
| | male | female | male | female |
| ACR method | 3.57 | 3.61 | 3.12 | 3.19 |
| RSC method | 3.66 | 3.69 | 3.23 | 3.27 |
| MC method | 3.78 | 3.81 | 3.39 | 3.41 |
| Proposed method | 4.03 | 4.07 | 3.63 | 3.66 |

signal in the extent from -0.5 to 4.5, in which the higher quantity represents that the quality of speech signal is better [25]. Table II shows the PESQ scores of the ACR, RSC, MC, and proposed methods. The MC method has shown better results up to now [9], but the proposed method provides the best result over the conventional methods. In the case of 10% packet loss, the conventional MC method provides PESQ scores of 3.78 for males and 3.81 for females, respectively. On the other hand, the proposed MC method provides PESQ scores of 4.03 for males and 4.07 for females, respectively. Similarly, for 30% packet loss, the proposed method provides acceptable PESQ scores in both male and female cases than the conventional MC method. Thus, the PESQ score is higher for compensated speech signals wherein the residual error is incorporated. We utilized the same database [24] for all considering methods in this paper for a fair comparison.

### C. LSD Score

Furthermore, the LSD score is observed for the restored BC speech for different methods in the PLC technique. We assumed a packet length of 20 ms and a frame-shifting of 10 ms for obtaining the LSD scores. The LSD scores of 24 frames from each restored speech are averaged through 10 ms frame

TABLE III. AVERAGE LSD SCORES OF DIFFERENT METHODS

| Methods | Speakers | |
|---|---|---|
| | male | female |
| ACR method | 12.08 | 11.89 |
| RSC method | 11.25 | 11.01 |
| MC method | 10.38 | 10.10 |
| Proposed method | 8.41 | 8.13 |

overlapping. The LSD is calculated as

$$LSD = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left|20\,\log_{10}\left(\frac{P(\omega_b)}{\hat{P}(\omega_b)}\right)\right|^2} \tag{22}$$

where $P(\omega_b)$ and $\hat{P}(\omega_b)$ correspond to the true and estimated power spectra, respectively, and $B$ denotes the upper-frequency bin number. The proposed method provides lower LSD scores for reconstructed BC speeches as shown in Table III. The lower LSD value indicates a better reconstruction of the transmitted speech signal [7].

### D. LAR Distances

LP coefficients are difficult to interpolate and sensitive to quantization errors. LP coefficients are normally converted into other parameters that are equivalent to LP coefficients. Also, the converted parameters are easy to handle before transmission. One such parameter is the log area ratio (LAR) which is used to represent LP coefficients for transmission over a channel [26]. The proposed method generates the LAR distance line nearer to the baseline as shown in Fig. 4. The LAR distance line nearer to the baseline discloses that the reconstructed speech signal is closer to the transmitted speech signal.

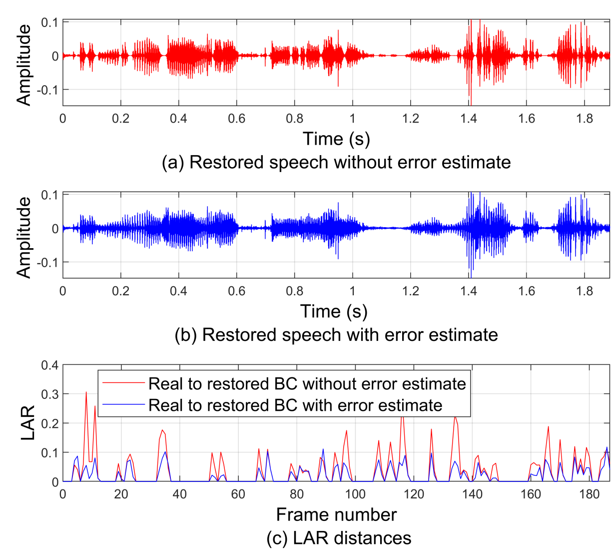


Fig. 4. LAR distances for estimated speech.

## VI. Conclusion

This paper suggested a PLC technique for the MC method where the estimated residual error was incorporated in the forward and backward directions for packet loss compensation. The MC method was considered for BC speech to confirm residual error minimization and to improve the ill-condition for obtaining the best set of the LP parameters. The MC method compressed an expanded spectral dynamic range of input signals since the centrosymmetric characteristic was involved in the MC matrix C. It was also noted that the conventional LP-based PLC methods use a linear gain of 1.1 to 1.8. On the other hand, the proposed PLC method did not use a linear gain at all. Especially the PESQ and LSD scores showed the best results obtained by the proposed method. In the case of 10% packet loss, the proposed method provided PESQ scores of 4.03 for males and 4.07 for females, respectively. Through the LSD evaluation, the proposed method obtained the best LSD scores such as 8.41 for males and 8.13 for females, respectively. Simulation results in all objective evaluations showed that the proposed PLC method improves the conventional ones.

In this proposed method, residual error incorporation indicates one kind of input signal addition. In the future, we would like to use the combination of LP and polynomial prediction approaches in the PLC technique for better speech reconstruction without error signal incorporation.

## References


[1] S. Zhang, Y. Sugiura, and T. Shimamura, "Bone-conducted speech synthesis based on least squares method," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 17, no. 3, pp. 425-435, 2022.

[2] M.S. Rahman and T. Shimamura, "Pitch determination from bone conducted speech," *IEICE Trans. Inf. & Syst.*, vol. E99–D, no. 1, pp. 283-287, Jan. 2016.

[3] S. Tsuge and S. Kuroiwa, "Bone- and air-conduction speech combination method for speaker recognition, *International Journal of Biometrics*, vol. 11, no. 1, pp. 35–49, 2019.

[4] S. Iijima and T. Shimamura, "Bone-conducted speech for speaker verification," *Proc. Int. Workshop Nonlinear Circuits and Signal Processing*, 2008, pp. 172-175.

[5] H. Q. Nguyen and M. Unoki, "Improvement in bone-conducted speech restoration using linear prediction and long short-term memory model," *Journal of Signal Processing*, vol. 24, no. 4, pp. 175-178, July 2020.

[6] P. Singh, M. K. Mukul, and R. Prasad, "Bone conducted speech signal enhancement using LPC and MFCC," *Int. Conf. on Intelligent Human Computer Interaction (IHCI), Springer*, vol. 11278, 2018, pp. 148-158.

[7] T. T. Vu, M. Unoki, and M. Akagi, "A blind restoration model for bone-conducted speech based on a linear prediction scheme," *Int. Symp. on Nonlinear Theory and its Applications*, 2007, pp. 449-452.

[8] M. A. Rahman, T. Shimamura, and H. Makinae, "LP-based quality improvement of noisy bone conducted speech," *IEEJ Transactions on Electronics, Information and Systems*, vol. 137, no. 1, pp. 197-198, 2017.

[9] Ohidujjaman, N. Yasui, Y. Sugiura, T. Shimamura, and H. Makinae, "Packet loss compensation for VoIP through bone-conducted speech using modified linear prediction," *IEEJ Trans. Electrical and Electronic Engineering, Wiley*, vol. 18, no. 11, pp. 1781-1790, Aug. 2023.

[10] J. Lecomte and T. Backstrom, " Packet loss and concealment," in *Speech Coding with Code-Excited Linear Prediction, Springer*, $1^{st}$ ed., Mar. 2017, pp. 161-184.

[11] T. Morinaga, S. Sasak, K Manol, and T. Kanekol, "Robust speech coding under packet-loss conditions using recovery sub-codec for broadband IP network," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 03, 2002, pp. 2713-2716.

[12] A. Bakri, A. Barouche, and M. Abbas, "An improved packet loss concealment technique for speech transmission in VOIP," $2^{nd}$ *Int. Conf. on Natural Language and Speech Processing (ICNLSP)*, April 2018, pp. 1-5.

[13] B.J. Borgstrom, P.H. Borgstrom, and A. Alwan, "Efficient HMM-based estimation of missing features, with applications to packet loss concealment," $11^{th}$ *Annual Conf. of the Int. Speech Communication Association, Chiba, Japan*, Sep. 26-30, 2010.

[14] M.K. Lee, S.K. Jung, H.K. Kang, Y.C. Park, and D.H. Youn, "A packet loss concealment algorithm based on time-scale modification for CELP-type speech coders," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2003, pp. 116–119.

[15] C.A. Rodbro, M.N. Murthi, S.V. Andersen, and S.H. Jensen, "Hidden Markov model-based packet loss concealment for voice over IP," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, 2006, pp. 1609–1623.

[16] B.K. Lee and J. H. Chang, "Packet loss concealment based on deep neural networks for digital speech transmission," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 2, Feb. 2016, pp. 378-387.

[17] R. Lotfidereshgi and P. Gournay, "Speech prediction using an adaptive recurrent neural network with application to packet loss concealment," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 5394–5398.

[18] M.M Mohamed, M.A Nessiem and B.W. Schuller, "On deep speech packet loss concealment," *A Mini- Survey*, arXiv preprint arXiv:2005.07794 [eess.AS], May 2020.

[19] S. Davy, N. Belton, J. Tobin, O.B. Zuber, L. Dong, and Y. Xuewen, "A causal convolutional approach for packet loss concealment in low powered devices," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2023.

[20] "Improving audio quality in Duo with WaveNetEQ," http://ai.googleblog.com/2020/04/improving-audio-quality-in-duo-with.html
Archived: https://web.archive.org/web/20220309043235/http://ai.googleblog.com/2020/04/improving-audio-quality-in-duo-with.html

[21] J. Lindbrom and P. Hedelin, "Packet loss concealment based on sinusoidal extrapolation," IEEE ICASSP, Vol. 1, 2002, pp. 173-176.

[22] K. Kondo and K. Nakagawa, "A speech packet loss concealment method using linear prediction," *IEICE Trans. Inf. & Syst.*, vol. E89-D, no. 2, 2006, pp. 806-813.

[23] E. Gunduzhan 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, Nov. 2001.

[24] K. Amino, T. Osanai, T. Kamada, H. Makinae H and T. Arai, "Effects of the phonological contents and transmission channels on forensic speaker recognition," *Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism (A. Neustein, H. Patil eds.), Springer-Verlag Berlin Heidelberg*, 2011, pp. 275-308.

[25] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," 2001.

[26] D. Chow and W.H. Abdulla, "Speaker identification based on log area ratio and Gaussian mixture models in narrow-band speech," *PRICAI 2004: Trends in Artificial Intelligence, Lecture Notes in Computer Science, Springer Berlin / Heidelberg*, vol. 3157, 2004, pp. 901-908.