Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 1732151, 7 pages
https://doi.org/10.1155/2018/1732151

Research Article

Fixed-Location Pulse Linear Prediction Coding Vocoder Model for Low Bit Rate Speech Coding

Zhen Ma

School of Information Engineering, Binzhou University, China

Correspondence should be addressed to Zhen Ma; 13954380080@139.com

Received 26 September 2018; Revised 4 December 2018; Accepted 13 December 2018; Published 31 December 2018

Academic Editor: Paolo Crippa

An arbitrary-location pulse determination algorithm based on multipulse linear prediction coding (MP-LPC) is presented. This algorithm can determine all the amplitudes of the pulses at a time according to given pulse locations without the use of analysis-by-synthesis. This ensures that the pulses are optimal in a least-square sense, providing the theoretical foundation to improve the quality of synthesized speech. A fixed-location pulse linear prediction coding (FLP-LPC) method is proposed based on the arbitrary-location pulse determination algorithm. Simulation of the algorithm in MATLAB showed the superior quality of the speech synthesized using pulses in different locations and processed using the arbitrary-location pulse determination algorithm. The algorithm improved speech quality without affecting coding time, which was approximately 1.5% of the coding time for MP-LPC. Pulse locations in FLP-LPC are fixed and do not need to be transmitted, with only LSF, gain, and 16 pulse amplitudes requiring coding and transmission. FLP-LPC allows the generation of synthesized speech similar to G.729 coded speech at a rate of 2.5 kbps.

1. Introduction

As one of the most common ways to exchange ideas, speech coding has been the topic of extensive research for many years [1–4]. In mobile communication systems, along with the explosive growth in the number of users, a dramatic increase in telephone traffic has led to limited bandwidth allotted to each speech channel.

The quality of synthesized speech and the coding rate are contradictory factors in communication systems. Parameter coding focuses on and extracts the parameters of vocal tract model and excitation, which can synthesize speech at bit rates below approximately 4 kbps. The synthesized speech has intelligibility, although it sounds unnatural. Parameter coding includes duality excitation linear prediction, mixed excitation linear prediction (MELP) [5] (McCree and Barnwell, 1995), waveform interpolation (WI) [6], sinusoidal transform coding (STC) [7], and multiband excitation (MBE) [8]. Hybrid coding can achieve good synthesized speech quality at coding rates ranging from 4 to 16 kbps, which is used in different areas. Hybrid coding includes multipulse linear prediction coding (MP-LPC) [9] and code-excited linear prediction [4, 10].

MP-LPC is a typical analysis-by-synthesis linear predictive coding (ABS-LPC) method in which dozens of pulses are selected as excitation signals [11]. MP-LPC can achieve good synthesized speech quality at low coding rates; however, its pulse determination algorithm is complex. MP-LPC was improved by simplifying the amplitudes and locations of excitation pulses. In regular-pulse excitation linear prediction coding (RPE-LPC) [12], equispaced pulses are used as the excitation signal. Therefore, only the amplitude of pulses and the first pulse location need to be determined and transmitted. In multipulse maximum likelihood quantization (MP-MLQ) [13], the locations of pulses can be either all odd or all even, and the amplitudes of pulses can be the signs (±1). In RPE-LPC and MP-MLQ, the locations or locations and amplitudes of pulses become more regular, and less information about them needs to be transmitted. In the present study, we present an arbitrary-location pulse determination algorithm based on conventional MP-LPC. In this algorithm, pulse locations can be assigned arbitrarily

without searching through the analysis-by-synthesis procedure. These pulses with arbitrarily assigned locations and optimal amplitudes can produce speech approaching original speech in a least-square sense. Based on the arbitrary-location pulse determination algorithm, a fixed-location pulse linear prediction coding (FLP-LPC) method is presented. The arbitrary-location pulse determination algorithm can synthesize speech with better quality within a shorter coding time than the sequential method. FLP-LPC not only yields synthesized speech of high quality and short coding time, but can also reduce the coding rate.

2. MP-LPC

For a speech frame of length N, M pulses are used as the excitation signal denoted as $p(n) = \sum_{k=1}^{M} g_k * \delta(n-n_k)$, where g_k and n_k are the amplitude and location of the kth pulse, respectively. The key of MP-LPC is to determine g_k and n_k to minimize the perceptual weighted error between original and synthesized speech.

The corresponding synthesized speech signal is

$$\hat{s}(n) = \hat{s}_0(n) + \sum_{k=1}^{M} g_k * h(n-n_k), \tag{1}$$

where $\hat{s}_0(n)$ is the zero-input response of the linear prediction (LP) filter that represents the effect of the previous frame, and $h(n)$ is the unit impulse response of the LP filter.

The error between $\hat{s}(n)$ and $s(n)$ is

$$e_s(n) = s(n) - \hat{s}(n) = \overline{e}_n(n) - \sum_{k=1}^{M} g_k * h(n-n_k), \tag{2}$$

where $\overline{e}_n(n) = s(n) - \hat{s}_0(n)$ is the equivalent speech without the effect of the previous frame, that is, the excitation signal filtered by the LP filter. $e_s(n)$ filtered by the perceptually weighted filter is

$$\begin{aligned} e(n) &= \left[\overline{e}_n(n) - \sum_{k=1}^{M} g_k * h(n-n_k)\right] * w(n) \\ &= \overline{e}_w(n) - \sum_{k=1}^{M} g_k * h_w(n-n_k), \end{aligned} \tag{3}$$

where $h_w(n)$ is the unit impulse response of the perceptually weighted filter. The perceptually weighted mean square error E_m is

$$E_m = \sum_{n=1}^{N} e^2(n) = \sum_{n=1}^{N}\left[\overline{e}_w(n) - \sum_{k=1}^{M} g_k * h_w(n-n_k)\right]^2. \tag{4}$$

The main idea of MP-LPC is to minimize E_m by selecting the appropriate g_k and n_k in the excitation signal. By setting the partial derivative of E_m to zero, M linear equations and M nonlinear equations can be obtained. Solving these $2M$ equations at a time is complicated. Therefore, in MP-LPC, a sequential method is used to determine the amplitude and location of one pulse in each iteration. After M iterations, the amplitudes and locations of M pulses can be determined. The n_j of the jth pulse is the location to maximize the following formula:

$$\frac{R_{eh}^2(n_j)}{R_{hh}(n_j, n_j)}. \tag{5}$$

Next g_j is determined according to the n_j determined above as follows:

$$g_j = \frac{R_{eh}(n_j)}{R_{hh}(n_j, n_j)}, \tag{6}$$

where

$$R_{eh}(n_j) = \sum_{n=1}^{N} \overline{e}_w(n) * h_w(n-n_j), \tag{7}$$

$$R_{hh}(n_k, n_j) = \sum_{n=1}^{N} h_w(n-n_k) * h_w(n-n_j). \tag{8}$$

3. Arbitrary-Location Pulse Determination Algorithm

From the above, the main idea of MP-LPC is to determine g_k and n_k for $\|e(n)\| = \|\overline{e}_w(n) - \sum_{k=1}^{M} g_k * h_w(n-n_k)\| = 0$, that is

$$\sum_{k=1}^{M} g_k * h_w(n-n_k) = \overline{e}_w(n). \tag{9}$$

The above M equations can be written compactly as

$$H\beta = E, \tag{10}$$

where

$$\begin{aligned} &H(n_1, n_2, \ldots, n_M) \\ &= \begin{bmatrix} h_w(1-n_1) & \cdots & h_w(1-n_M) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ h_w(N-n_1) & \cdots & h_w(N-n_M) \end{bmatrix}_{N\times M}, \end{aligned} \tag{11}$$

$$\beta = [g_1, g_2, \ldots, g_M]^T, \tag{12}$$

$$E = [\overline{e}_w(1), \overline{e}_w(2), \ldots, \overline{e}_w(N)]^T. \tag{13}$$

It can be inferred that H will remain unchanged once $n_1, n_2, \ldots, n_M$ are determined (even though assigned arbitrarily) in the beginning of the search. Therefore, formula (10) can be regarded as a linear system. If $N = M$ and H is nonsingular, then linear system (10) has a unique solution. However, in most cases $N > M$ and the system is overdetermined and may not be consistent; therefore, it does

not always have a unique solution. To determine g_k is simply equivalent to finding a least-squares solution $\widehat{\beta}$ of (10):

$$\left\| H\widehat{\beta} - E \right\| = \min_{\beta} \left\| H\beta - E \right\|. \tag{14}$$

The smallest norm least-squares solution of the above linear system can be determined by H^+, the Moore–Penrose generalized inverse of H [14]

$$\widehat{\beta} = H^+ E. \tag{15}$$

Several methods can be used to calculate the Moore–Penrose generalized inverse, including orthogonal projection, the orthogonalization method, the iterative method, and singular value decomposition (SVD) [14].

The locations of pulses in the excitation signal can be assigned arbitrarily, which has no effect on the existence of the smallest norm least-squares solution of (10). The process of the arbitrary-location pulse determination algorithm can be summarized as follows.

Step 1. Assign pulse locations n_i, $i = 1, \ldots, M$ arbitrarily.

Step 2. Calculate the unit impulse response matrix H.

Step 3. Calculate the pulse amplitude vector β: $\beta = H^+ E$.

The location of pulses in the excitation signal has no effect on the existence of the smallest norm least-squares solution of (10); namely, the synthesized speech can always approach the original speech in a least-square sense regardless of the pulse locations. Moreover, the transmission of pulse locations will increase the coding rate, leading to the waste of bandwidth. As discussed above, for different speech frames, the pulses in fixed locations but with different amplitudes can be used as the excitation signal. We therefore developed a method in which only the pulse amplitudes and LPC parameters are coded and the pulse locations do not need to be coded. This method is called fixed-location pulse linear prediction coding.

4. Results

4.1. Performance Evaluation of the Arbitrary-Location Pulse Determination Algorithm. To test the effect of a combination of different pulse locations on the quality of synthesized speech processed by the proposed arbitrary-location pulse determination algorithm, five sections of speech from five different speakers with equal content were analyzed. The sampling frequency was 8000 Hz. There were 813 frames in the five sections of speech, and each frame included 160 samples. The speech frames were analyzed in 50 trials, and the locations of pulses were arbitrarily selected at each trial; the amplitudes of all pulses were then calculated using the proposed arbitrary-location pulse determination algorithm.

The pulses at different locations and the synthesized speech generated by each for the same frame of speech are shown in Figure 1. The residual signal and original speech are shown in Figures 1(a) and 1(b), respectively. For direct comparison with the sequential method, the same locations of pulses were used in the arbitrary-location pulse determination algorithm and in the sequential method. The excitation signals obtained by the sequential method and arbitrary-location pulse determination algorithm (in the same locations but with different amplitudes) are shown in Figures 1(c) and 1(e), and the corresponding synthesized speech is shown in Figures 1(d) and 1(f). These two methods were capable of generating synthesized speech close to the original speech with signal to noise ratios (SNR) of 17.1471 and 19.9547, indicating that the proposed method is superior to the sequential method. Here SNR is defined as

$$SNR = 10 \log \left[\frac{\sum_{k=0}^{N-1} S_o^2(k)}{\sum_{k=0}^{N-1} \left(S_o(k) - S_r(k) \right)^2} \right], \tag{16}$$

where $S_o(k)$ and $S_r(k)$ are the original and synthesized speech signals, respectively. Figures 1(g)–1(p) show the speech synthesized with five strings of pulses in different locations, illustrating the superior SNRs obtained with the proposed method compared with those of the sequential method. These results indicate that high-quality synthesized speech can be obtained with pulses in different locations, and the amplitudes were calculated using the proposed method. The mean SNR for the 40650 trials performed was 19.2937. The mean coding time of the sequential method was 0.1224, and that of the proposed method was 0.0019, only 1.55% of the coding time for sequential method.

For the above speeches, 50 trials were conducted for all the speech frames. At each trial, 8, 10, 12, 14, 16, 18, 20, 22, 24, 28, and 32 pulses with arbitrarily selected locations were extracted, and the amplitudes were calculated using the proposed algorithm. The results are shown in Figure 2, which shows that the box heights decrease and the medians increase (Figure 2(a)). The increase in the mean SNR at all trials and the decrease in standard deviation with increasing pulse number indicate that the greater number of pulses leads to a more steady solving process and a more modest effect of the assigned pulse locations. As mentioned by Ma et al. [15], for a certain frame of speech, when the pulse number reaches a certain value, a continuing increase of pulses will not improve the quality of synthesized speech. Regarding the proposed algorithm, when the pulse number reaches a certain value, the rank of H does not increase subsequently and does not contribute significantly to solving for amplitudes. The average coding times for different pulse numbers are shown in Figure 2(b). An increase in the pulse number results in a gradual increase in the average coding time. The average coding time for 32 pulses was 0.0028 s, which was lower than that of the sequential method at 0.1224 s. The excitation signals obtained with different numbers of pulses and the synthesized speeches for the same frame of speech are shown in Figure 3. The results show that these synthesized speeches were close to the original speech and had good SNRs. Figures 2 and 3 indicate that when the pulse number is greater than 16, the quality of synthesized speech is improved in a nonobvious manner.

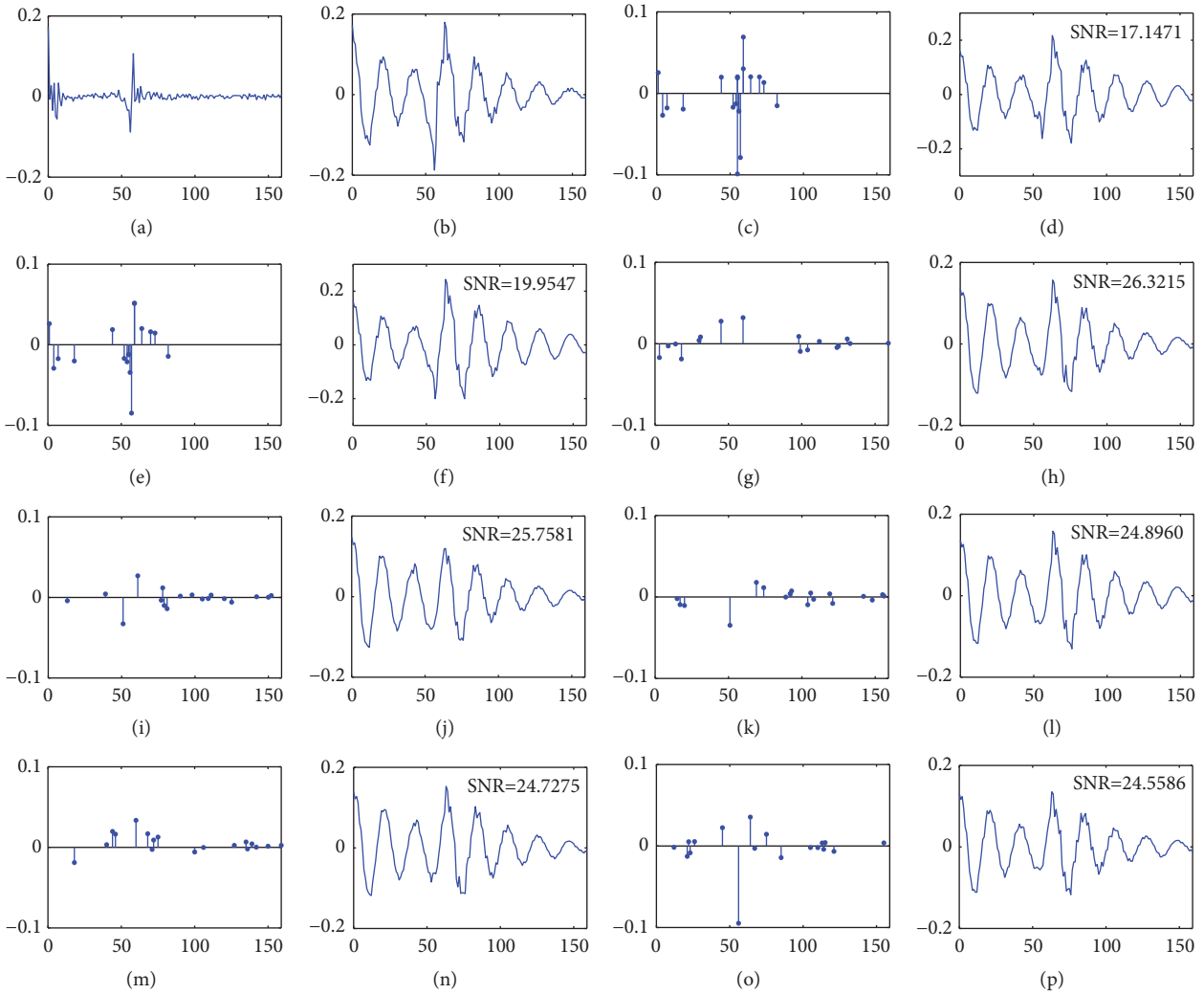

Figure 1: (a) Residual signal; (b) original speech; (c) multipulse excitation signal obtained by the sequential method; (d) speech synthesized using the excitation signal in (c); (e) multipulse excitation signal obtained by the presented method with the same locations as the signal in (c); (f) speech synthesized using the excitation signal in (e); (g–p) the pulses with the locations arbitrarily assigned and the speeches synthesized.

4.2. Performance Evaluation of FLP-LPC

4.2.1. Performance Evaluation of Unquantized FLP-LPC. In the proposed method, the locations of pulses can be assigned arbitrarily and do not need to be calculated using an algorithm. Therefore, the pulses at fixed locations but with different amplitudes can be used as the excitation signal for every speech frame. The proposed method and sequential method [11] were used to process the same speech spoken by a female and male from Chinese Central Television news broadcasting (2534.3950 s). The PESQ_MOS specified in the ITU standard P.862 and SNR were used to evaluate speech quality. Average SNR and PESQ_MOS as a function of pulse number are shown in Figure 4.

The speech obtained with FLP-LPC was more natural and intelligible than that obtained with MP-LPC. Average SNR and PESQ_MOS for MP-LPC and FLP-LPC increase with pulse number. But they increase insignificantly, when the pulse numbers are more than 18.

4.2.2. Coding Scheme of FLP-LPC and Performance Evaluation. The present results indicate that, for a speech frame of 20 ms, 16 pulses are sufficient to produce good-quality synthesized speech. Therefore, in the coding scheme of FLP-LPC, 16 evenly distributed pulses are used as the excitation signal. First, the amplitudes of pulses are normalized, and then the gain and normalized amplitudes are coded. The coded parameters are LSF, gain, and pulse amplitude. LSF and normalized amplitudes are multistage vector quantized and the gain is quantized using 4 bits. The specific bit allocation of FLP-LPC is shown in Table 1.

The test speeches included 20 sets of samples from two men and two women with a sample frequency of 8000 Hz. There were background noises in the recording, such as a creaking door and automobile noises. The test speeches included a database composed of 1560 sentences from 83 men and 83 women, the content of which was selected from People's Daily. These speeches were coded using FLP-LPC, G.723.1 and G.729, and the PESQ_MOS values were

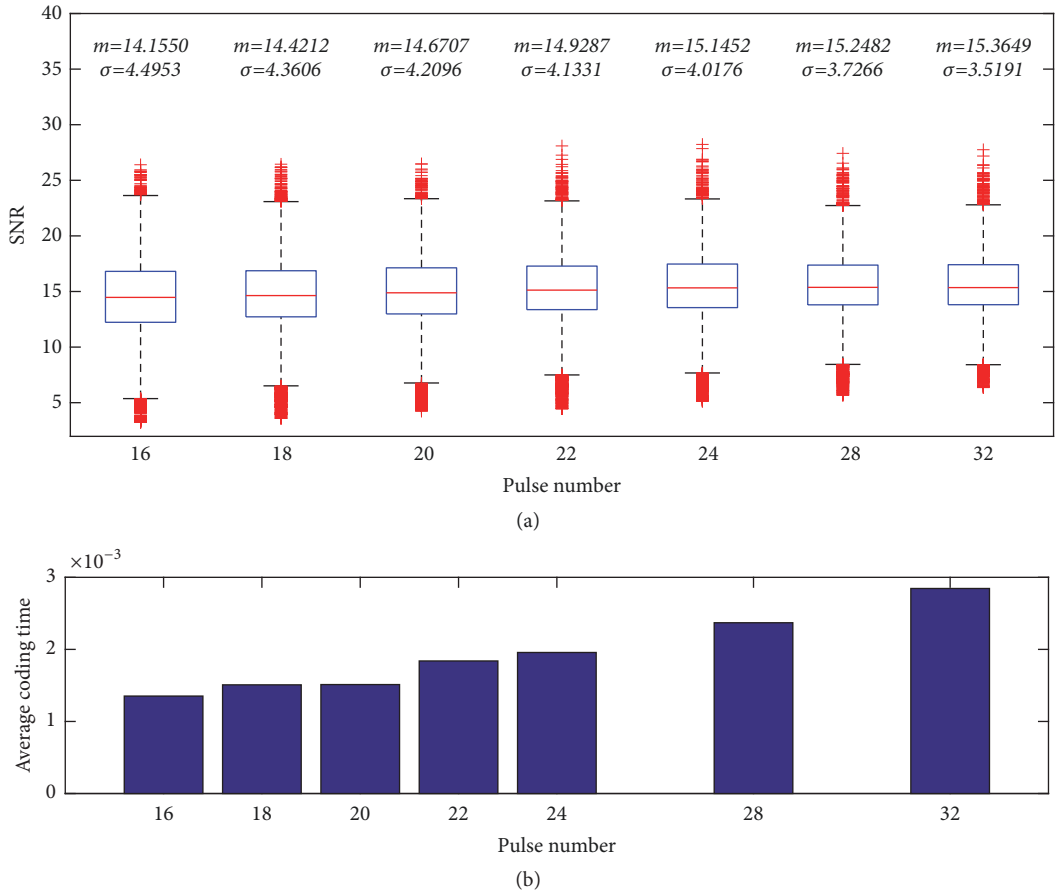

Figure 2: (a) SNR distribution with different pulse numbers; (b) average coding time with different pulse numbers.

Table 1: FLP-LPC bit allocation.

Parameter	Bits
LSF	18
Gain	4
Pulse amplitudes	28
total	50

Table 2: Performance comparison of FLP-LPC with G.723.1 and G.729.

Coding method/standard	Coding rate (kbps)	PESQ_MOS
FLP-LPC	2.5	3.731
G.723.1	5.3	3.497
G.729	8	3.765

calculated, as shown in Table 2. FLP-LPC synthesized speech with similar quality to that generated with G.729 and superior to that of G.723.1 at a coding rate of 2.5 kbps.

5. Conclusion

To solve the problems associated with MP-LPC, an arbitrary-location pulse determination algorithm was developed in the current study. In this algorithm, the pulse amplitudes are determined by solving a linear system of equations under the premise of arbitrarily assigned locations of pulses. The existence of a smallest norm least-squares solution of this linear system is not affected by the locations of pulses in the excitation signal. Tests performed in different speech frames showed that pulse combinations with different locations can be used as the excitation signal to synthesize high-quality speech, yielding better results than those obtained with the conventional sequential method. The sequential method determines one pulse at a time, which ensures that the added pulse is optimal at every iteration, whereas it does not guarantee that the combination of pulses after all iterations is optimal. In the proposed algorithm, the combination of pulses determined at a time is optimal in a least-square sense, which provides the theoretical basis for ensuring the quality of synthesized speech. The proposed algorithm does not increase coding time to improve synthesized speech quality, which is 1.5% of the coding time in MP-LPC. To study the effect of pulse number on the quality of synthesized speech, the excitation signals with different numbers of pulses were

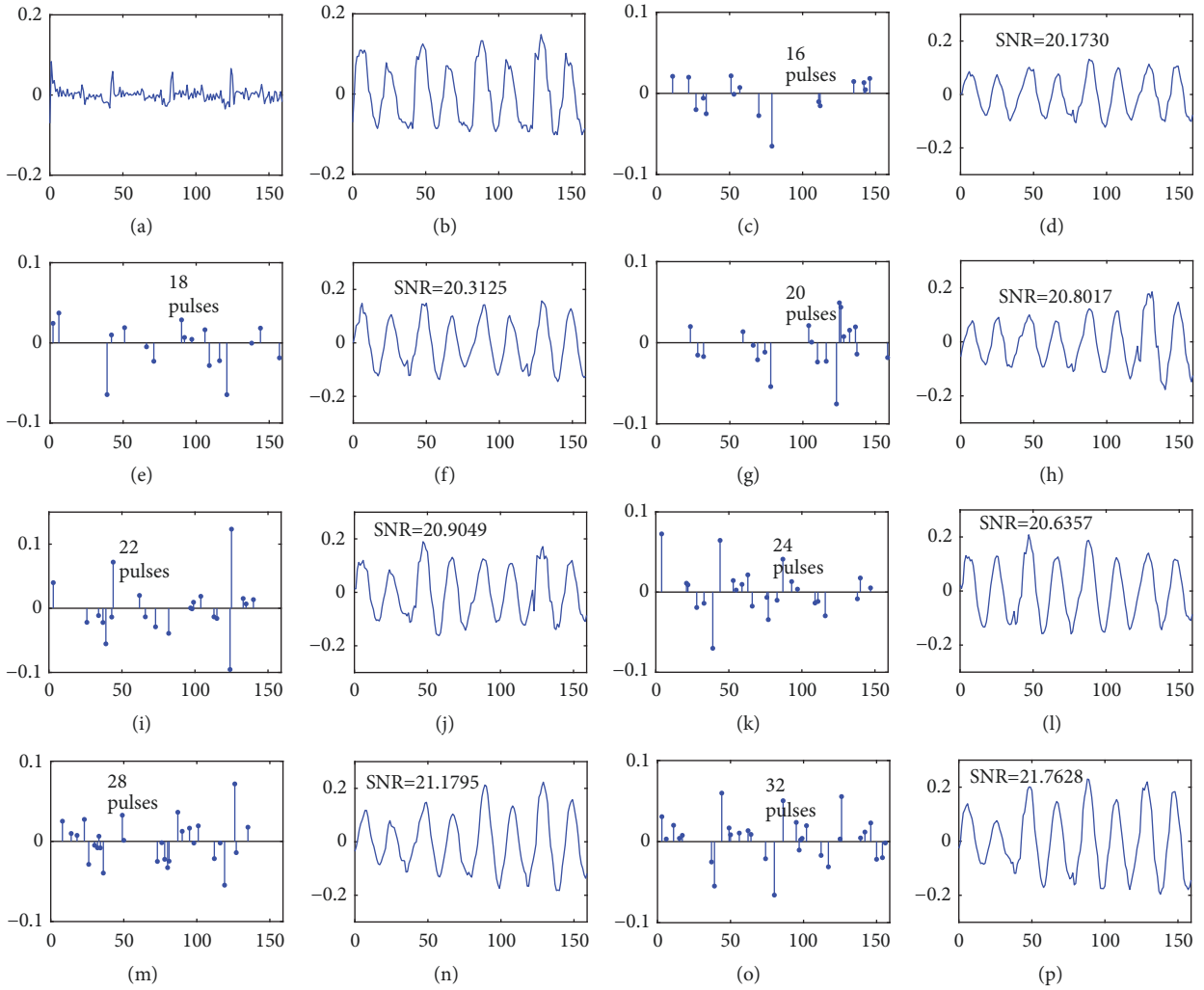

Figure 3: (a) residual signal; (b) original speech; (c–p) excitation signals with 16, 18, 20, 22, 24, 28, and 32 pulses and the speeches synthesized.

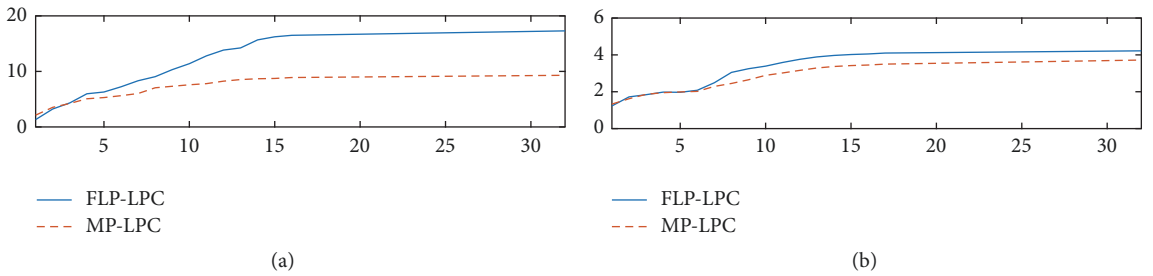

Figure 4: (a) average SNR and (b) PESQ_MOS as a function of pulse number for MP-LPC and FLP-LPC.

calculated for the same frames of speech. The results showed that 16 pulses were sufficient to generate 20 ms long speech.

In improved MP-LPC methods, such as RPE-LPC and MP-MLQ, the locations or amplitudes or both of pulses become more regular to reduce the information on the excitation signals to be transmitted. FLP-LPC was developed to further reduce the coding rate. The method is based on the premise that the pulse vector in a least-square sense is independent of the locations of pulses in excitation signal. In FLP-LPC, the locations of pulses are fixed, and only the amplitudes of pulses need to be determined through the arbitrary-location pulse determination algorithm. The pulse locations do not need coding or transmitting, which reduces the coding rate without affecting the quality of synthesized speech. Our results show that the SNR and the PESQ_MOS of the speech synthesized with FLP-LPC were higher than those of speech generated by MP-LPC. Moreover, we developed an FLP-LPC coding scheme in which the pulses are evenly

distributed. FLP-LPC can synthesize speech with quality similar to that of G.729 and superior to that generated by G.723.1 at a coding rate of 2.5 kbps. In conclusion, FLP-LPC can synthesize speech of high quality with a short coding time and can reduce the coding rate; however, it has the drawback of a bigger memory requirement for calculating the Moore–Penrose generalized inverse.

Data Availability

The speech data used to support the findings of this study were supplied by Chinese Linguistic Data Consortium under license and so cannot be made freely available. Requests for access to these data should be made to Mengyi Sun, service@chineseldc.org.

Conflicts of Interest

The author declares that they have no competing interests.

Acknowledgments

Project supported by Ministry of Education of China Humanities and Social Sciences Research Project (18YJCZH129); Natural Science Foundation of Shandong (ZR2014FL005); Binzhou University Scientific Research Fund Project (2016Y29).

References

[1] F. Lahouti, A. R. Fazel, A. H. Safavi-Naeini, and A. K. Khandani, "Single and double frame coding of speech LPC parameters using a lattice-based quantization scheme," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1624–1632, 2006.

[2] A. Mouchtaris, K. Karadimou, and P. Tsakalides, "Multiresolution Source/Filter Model for Low Bitrate Coding of Spot Microphone Signals," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2008, pp. 1–16, 2008.

[3] M. Deriche and D. Ning, "A novel audio coding scheme using warped linear prediction model and the discrete wavelet transform," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 6, pp. 2039–2048, 2006.

[4] N. Ku, C. Yeh, and S. Hwang, "An efficient algebraic codebook search for ACELP speech coder," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2014, no. 1, 2014.

[5] A. V. McCree and T. P. Barnwell, "A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 3, no. 4, pp. 242–250, 1995.

[6] W. B. Kleijn, "Encoding Speech Using Prototype Waveforms," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 1, no. 4, pp. 386–399, 1993.

[7] R. J. Mcauly and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[8] D. W. Griffin and J. S. Lim, "Multiband Excitation Vocoder," *IEEE Transactions on Signal Processing*, vol. 36, no. 8, pp. 1223–1235, 1988.

[9] BS. Jr Atal, "Remde, A new model of LPC excitation for producing natural-sounding speech at low bit rates," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 614–617, Paris, France, 1982.

[10] M. Schroeder and B. Atal, "Code-excited linear prediction(CELP): High-quality speech at very low bit rates," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 937–940, Tampa, FL, USA.

[11] S. Singhal and B. S. Atal, "Amplitude Optimization and Pitch Prediction in Multipulse Coders," *IEEE Transactions on Signal Processing*, vol. 37, no. 3, pp. 317–327, 1989.

[12] P. Kroon, E. F. Deprettere, and R. J. Sluyter, "Regular-Pulse Excitation—A Novel Approach to Effective and Efficient Multipulse Coding of Speech," *IEEE Transactions on Signal Processing*, vol. 34, no. 5, pp. 1054–1063, 1986.

[13] S.-W. Yoon, H.-G. Kang, Y.-C. Park, and D.-H. Youn, "An efficient transcoding algorithm for G.723.1 and G.729A speech coders: Interoperability between mobile and IP network," *Speech Communication*, vol. 43, no. 1-2, pp. 17–31, 2004.

[14] D. Serre, *Matrices: Theory and Applications*, vol. 216 of *Graduate Texts in Mathematics*, Springer, New York, NY, USA, 2002.

[15] Z. Ma, Y. Cao, and J. Zang, "Research on MPLPC excited-pulse abstract algorithm," in *Proceedings of the 2009 International Symposium on Computational Intelligence and Design, ISCID 2009*, pp. 489–492, China, December 2009.