

A Low Bit Rate Speech Coder Based on the Inflection Point Detection

Byeong-Gwan Iem

Department of Electronic Engineering, Gangneung-Wonju National University, Gangneung, Korea



Abstract

A low bit rate speech coder based on the non-uniform sampling technique is proposed. The non-uniform sampling technique is based on the detection of inflection points (IP). A speech block is processed by the IP detector, and the detected IP pattern is compared with entries of the IP database. The address of the closest member of the database is transmitted with the energy of the speech block. In the receiver, the decoder reconstructs the speech block using the received address and the energy information of the block. As results, the coder shows fixed data rate contrary to the existing speech coders based on the non-uniform sampling. Through computer simulation, the usefulness of the proposed technique is shown. The SNR performance of the proposed method is approximately 5.27 dB with the data rate of 1.5 kbps.

Keywords: Non-uniform sampling, Inflection point detection, Low rate speech coder

1. Introduction

Speech coding has been an important research area for both wired and wireless communications. Major techniques are waveform coding and vocoding [1, 2]. The waveform coding encodes speech samples directly by quantizing them. Prominent examples are the pulse code modulation (PCM), the delta modulation, and the adaptive differential PCM (ADPCM) [1]. The vocoding technique transmits the parameters which characterize a block of speech samples. The parameters include the voiced/unvoiced decision, the linear predictive coefficients (LPC), the pitch period, the energy, and so on [1, 2]. These parameters can be also used for the speech recognition [3, 4]. The examples of the vocoding methods are the code excited LPC, the multi-pulse LPC, the residual error excited LPC, and so on [2]. These speech coding techniques are performed after the uniform speech sampling.

The non-uniform sampling is a research area studied as an alternative to the conventional uniform sampling technique [5-12]. Due to less frequent sampling, the non-uniform sampling based speech coding shows lower bit rate than the PCM coding based on the uniform sampling. The non-uniform sampling is achieved by detecting local maxima and minima of a signal [5, 6] or inflection points of a signal [11, 12]. Since the number of maxima, minima and inflection points is variable depending on a signal, the bit rate of the non-uniform sampling based speech coder is not fixed. Thus, it is not an attractive candidate as a speech coder in communication where a planned and predetermined band is assigned for a communication channel.

In this paper, a new fixed and low bit rate speech coder is proposed. The coder is based on the inflection point detection method. To achieve a fixed bit rate, the coder compares the

Received: Dec. 9, 2015
Revised : Dec. 24, 2015
Accepted: Dec. 25, 2015

Correspondence to: Byeong-Gwan Iem
(ibg@gwnu.ac.kr)
©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

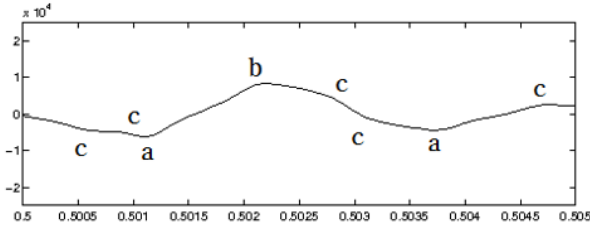


Figure 1. Enlarged plot of a speech signal with various inflection points.

detected inflection point (IP) pattern of a speech block with candidate IP patterns in a database (DB). And, the address of the closest IP pattern of the database is transmitted through the channel. At the receiver, the decoder fetches the IP pattern from the DB using the received address and estimates the speech signal through interpolation. As results, the data rate is not variable, but fixed. The structure of the paper is as follows. In next section, the inflection point detection (IPD) algorithm is explained in detail. The effect of the threshold value for the IPD is also considered. In Section 3, the structure of the encoder and the decoder is presented. And then, simulation results and conclusions are provided.

2. Inflection Point Detection

A segment of speech signal can be considered as a piecewise linear graph between inflection points. Figure 1 shows an enlarged plot of a signal. As shown in the figure, there are several types of inflection points: local maxima (point b), local minima (point a), and points of simple slope changes (point c). Non-uniform sampling techniques was proposed to detect these local maxima and minima to reduce the number of samples [5, 6]. Later, a method to detect the points showing slope changes (points a, b, c in Figure 1) was presented for speech coding [9].

In this paper, the inflection point detection (IPD) technique is more refined by considering the structure of inflection points. Figure 2 shows the typical inflection points when signal increases monotonically between t and $t+T$. For example, Figure 2(a) shows an inflection point of local maxima, where the next 2 samples show monotonically decreasing pattern between $t+T$ and $t+2T$. Figure 2(b) demonstrates an inflection point of slope change where the next 2 samples are constant pattern between $t+T$ and $t+2T$. Figure 2(c) shows an inflection point of another slope change where the next 2 samples are also monotonically increasing between $t+T$ and $t+2T$. The same IP patterns can be

considered for monotonically decreasing signal.

To detect these inflection points, the following IPD algorithm can be used. Let the consecutive differences of samples are expressed as

$$d_{21} = x_2 - x_1$$

$$d_{32} = x_3 - x_2$$

For the detection of local maxima and minima such as Figure 2(a), the product of consecutive differences is checked if it is less than 0, i.e.

$$d_{21} \cdot d_{32} < 0 \tag{1}$$

If it is less than 0, then the sample x_2 is a local maximum or a local minimum. To check slope change, the following identifier is defined and used:

$$\text{identifier (ID)} = \frac{|d_{21} - d_{32}|}{|d_{21}| + |d_{32}|} \tag{2}$$

The range of identifier value is $0 < \text{ID} \leq 1$ if there is a slope change. That is, for the slope change such as Figure 2(c), the ID value is $0 < \text{ID} < 1$. For a little change of slope, the ID value approaches to 0. If there is no slope change, the value for ID is 0. And for the big change of slope such as Figure 2(b), the ID value is 1. The same consideration can be applied for the monotonically decreasing signal pattern. Therefore, the IPD algorithm shown in Figure 3 can be used. That is, if the condition in (1) is satisfied, the sample is classified as an inflection point of local maxima or minima. Otherwise, the ID value in (2) is calculated and compare to a predetermined threshold. If the value is greater than the threshold, the sample is classified as an inflection point of slope change. By setting a threshold value for the ID, the IPD can adjust the amount of inflection points detected. That is, the smaller threshold value means the more detected inflection points.

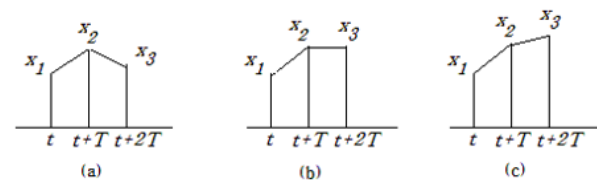


Figure 2. Types of inflection points in increasing signal at the early stage.

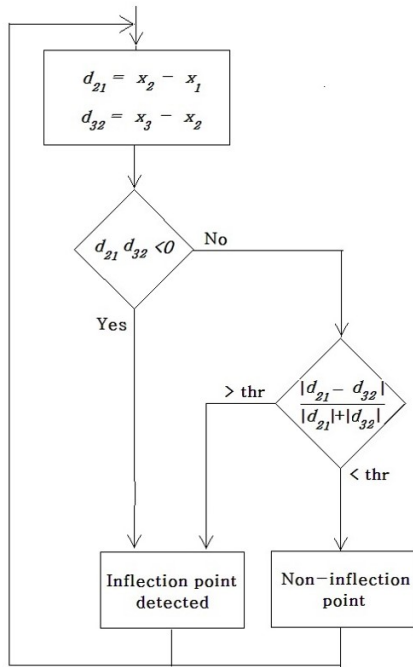


Figure 3. Inflection point detection algorithm.

3. Structure of the Speech Coder

The speech coder based on the non-uniform sampling technique shows variable data rate which is not suitable for communication application [5, 6, 11, 12]. In this paper, a new fixed and low bit rate speech coder is proposed based on the IPD. The structure of the speech coder is shown in Figure 4.

A block of speech signal is processed by the IPD algorithm, and the resulting IP pattern is normalized by its energy. The normalized IP pattern is compared with the elements of the IP pattern database. The address of the closest member of the database and the energy of the detected IP pattern are sent through communication channel.

At the receiver, using the received address and the energy information, the decoder reconstructs the speech signal. The decoder fetches the IP pattern from the database using the received address, and multiplies the obtained element of the DB with the received energy. Then, the decoder performs interpolation to get a speech estimates. Thus, the bit stream transmitted consists of the bits for the address and the energy for each speech block. For example, if a speech segment is taken as 20 ms with the sampling frequency of 10 kHz, the block has 200 samples, and there are 50 blocks per a second. The number of bits for an address is determined by the size of

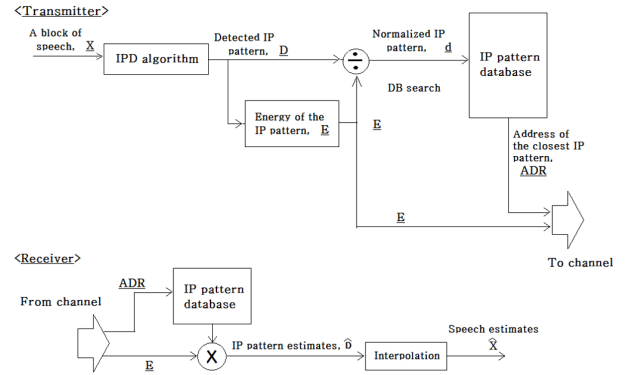


Figure 4. Structure of the speech coder. IPD, inflection point detection; IP, inflection point.

the IP pattern database. If the size of the database is N, the number of address bits is $\log_2 N$. Therefore, the data rate is $(\log_2 N + M)$ bits/block * 50 blocks/second, where M is the bits for the maximum energy of a detected IPD pattern.

4. Simulation Results

The computer simulation is performed to show the usefulness of the proposed speech coding technique. A speech is sampled at the rate of 10 kHz, and the speech is segmented as 20 ms blocks with 50% overlapping. The window function used in segmentation is the Hanning window. And the IP pattern database has 8519 entries, so the number of bits for the address is $\log_2 8519 = 14$, where x is the nearest integer greater than x. And the number of bits for the energy is 16 bits. As results, the data rate is 1500 bits/second. Figure 5 shows the processed signal results. Figure 5(a) is the original signal, Figure 5(b) the detected inflection point signal superimposed on the original signal of (a), and Figure 5(c) the reconstructed signal. Clearly, from Figure 5(a) and (c), the usefulness of the proposed speech coder can be confirmed. The SNR performance of the proposed speech coder is calculated as follows:

$$SNR = 10 \log_{10} \left[\frac{\text{signal power}}{\text{noise power}} \right]$$

where the noise is the difference between the original signal and the reconstructed signal. The SNR is 5.27 dB for the speech signal in Figure 5. The SNR value is comparable with that of uniform sampling based PCM coder [1]. The SNR performance of the uniform sampling PCM coder is theoretically given as

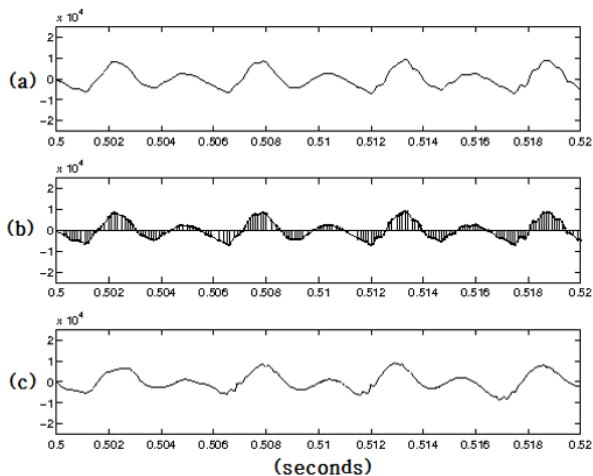


Figure 5. Processing results of the inflection point detection (IPD) based coding: (a) original speech, (b) detected inflection point superimposed on the original in (a), and (c) reconstructed speech at the receiver.

[1]

$$SNR \text{ (dB)} = 6B + 4.77 - 20\log_{10} \left[\frac{X_{max}}{\sigma_x} \right]$$

where B is the number of bits per sample, and X_{max} and σ_x are the maximum value and the standard deviation of a speech signal. For example, when $B = 3$ and $X_{max}/\sigma_x \cong 7.8$, theoretically, $SNR \text{ (dB)} \approx 4.91$ dB. In computer simulation, if the uniformly sampled speech is linearly quantized with 3 bits, the SNR is calculated as 4.18 dB. If the sampling rate is 10 kHz, the data rate is 30 kbps. Thus, the proposed IP based coding method shows similar SNR performance with much lower data rate comparing to uniform sampling PCM.

5. Conclusion

A new non-uniform sampling based speech coding technique has been proposed. Unlike existing non-uniform sampling based coding methods, the proposed coder shows a fixed data rate. The inflection points of a speech block are detected and compared with entries of IP pattern database. The address of the closest entry of the DB and the energy of the IP pattern are transmitted. At the receiver, the decoder fetches the DB entry and reconstructs the speech through interpolation. The computer simulation has shown the usefulness of the proposed speech coding technique. The SNR performance of the non-uniform sampling based coding has been compared with that of the uniform sampling based PCM coding. With relatively much

lower bit rate of 1.5 kbps, the IP based speech coder shows similar SNR of 5.27 dB to the uniform sampling PCM coder of 30 kbps.

As future research topics, the search algorithm of the IP pattern DB should be further studied for efficient implementation of the proposed coding method. And, the IP pattern DB itself can be also refined for high quality speech reconstruction.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgements

This work was supported by Gangneung-Wonju National University in 2015.

References

- [1] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [2] A. M. Kondozi, *Digital Speech: Coding For Low Bit Rate Communication Systems*. Chichester: John Wiley & Sons, 1994.
- [3] G. Lee and W. G. Kim, "Emotion recognition using pitch parameters of speech," *Journal of Korean Institute of Intelligent Systems*, vol. 25, no. 3, pp. 272-278, 2015. <http://dx.doi.org/10.5391/jkiis.2015.25.3.272>
- [4] W. G. Kim, "Robust speech recognition parameters for emotional variation," *Journal of Korean Institute of Intelligent Systems*, vol. 15, no. 6, pp. 655-660, 2005. <http://dx.doi.org/10.5391/jkiis.2005.15.6.655>
- [5] M. Bae, W. Lee, and D. Kim, "On a new vocoder technique by the non-uniform sampling," in *Proceeding of IEEE Military Communications Conference (MIL-COM'96)*, McLean, VA, 1996, pp. 649-652. <http://dx.doi.org/10.1109/milcom.1996.569428>
- [6] M. Budaes and L. Goras, "On speech signal reconstruction from local extreme values," in *Proceeding of International Symposium on Signals, Circuits and Systems (ISSCS)*, Lasi, Romania, 2005, pp. 315-318. <http://dx.doi.org/10.1109/isscs.2005.1509917>

- [7] L. D. Davisson, "Data compression using straight line interpolation," *IEEE Transactions on Information Theory*, vol. 14, no.3, pp. 390-394, 1968. <http://dx.doi.org/10.1109/tit.1968.1054160>
- [8] T. Fjallbrant, "Method of data reduction of sampled speech signals by using non-uniform sampling and a time-variable digital filter," *Electronics Letters*, vol. 13, no. 11, pp. 334-335, 1977. <http://dx.doi.org/10.1049/el:19770243>
- [9] P. K. Ghosh and T. V. Sreenivas, "Dynamic programming based optimum non-uniform samples for speech reconstruction and coding," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, 2006, pp. 1221-1224. <http://dx.doi.org/10.1109/icassp.2006.1660247>
- [10] J. W. Mark and T. D. Todd, "A non-uniform sampling approach to data compression," *IEEE Transactions on Communications*, vol. 29, no.1, pp. 24-32, 1981. <http://dx.doi.org/10.1109/TCOM.1981.1094872>
- [11] B. G. Iem, "A non-uniform sampling technique based on inflection point detection and its application to speech coding," *Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 903-909, 2014. <http://dx.doi.org/10.1121/1.4884882>
- [12] B. G. Iem, "A non-uniform sampling technique and its application to speech coding," *Journal of Korean Institute of Intelligent Systems*, vol. 24, no. 1, pp. 28-32, 2014. <http://dx.doi.org/10.5391/jkiis.2014.24.1.028>



Byeong-Gwan Iem received his B.S. and M.S. from Yonsei University, Seoul, Korea, in 1988 and 1990, respectively. He received his Ph.D. from the University of Rhode Island, RI, USA in 1998. He is a professor at Gangneung-Wonju National University,

Gangneung, Korea. His areas of study interests are DSP and its applications.

Tel: +82-33-640-2426

Fax: +82-33-646-0740

E-mail: ibg@gwnu.ac.kr