Original Article
International Journal of Fuzzy Logic and Intelligent Systems
Vol. 16, No. 4, December 2016, pp. 276-280
http://dx.doi.org/10.5391/IJFIS.2016.16.4.276

ISSN(Print) 1598-2645
ISSN(Online) 2093-744X

A Fixed Rate Speech Coder Based on the Filter Bank Method and the Inflection Point Detection

Byeong-Gwan Iem
Department of Electronic Engineering, Gangneung-Wonju National University, Gangneung, Korea

ljfis

Abstract

A fixed rate speech coder based on the filter bank and the non-uniform sampling technique is proposed. The non-uniform sampling is achieved by the detection of inflection points (IPs). A speech block is band passed by the filter bank, and the subband signals are processed by the IP detector, and the detected IP patterns are compared with entries of the IP database. For each subband signal, the address of the closest member of the database and the energy of the IP pattern are transmitted through channel. In the receiver, the decoder recovers the subband signals using the received addresses and the energy information, and reconstructs the speech via the filter bank summation. 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 confirmed. The signal-to-noise ratio (SNR) performance of the proposed method is comparable to that of the uniform sampled pulse code modulation (PCM) below 20 kbps data rate.

Keywords: Non-uniform sampling, Filter bank method, Inflection point detection

1. Introduction

Most of speech processing is based on the uniform sampling of speech signal [1–4]. It is easy and simple to implement to get discrete-time signal samples from a speech signal by band limiting and sampling according to Shannon's sampling theorem [1, 2]. However, there is large correlation between neighboring consecutive samples, and various speech coding techniques have been developed to reduce such correlation and redundant information [1, 2]. Differential coding and linear predictive coding (LPC) are some of prominent examples of speech coding [1, 2]. In differential coding scheme, the difference between a sample and its estimates is coded as in the delta modulation [1]. In LPC, a speech is analyzed based on the speech production model using linear predictive analysis, and obtained parameters are coded and transmitted [2]. These speech coding techniques require a lot of computation.

Non-uniform sampling technique is an alternative to overcome the information redundancy due to the high correlation between neighboring speech samples [5–11]. In non-uniform sampling, a speech is sampled not periodically. Typical examples of the non-uniform sampling method are the local maxima/minima detection method [6, 7] and the inflection point (IP) detection method [9–11]. The non-uniform sampling methods remove redundancies between samples by extracting irregularly, but the resulting signal shows variable code rate which is not

Received: Dec. 8, 2016
Revised : Dec. 12, 2016
Accepted: Dec. 13, 2016

Correspondence to: Byeong-Gwan Iem
(ibg@gwnu.ac.kr)

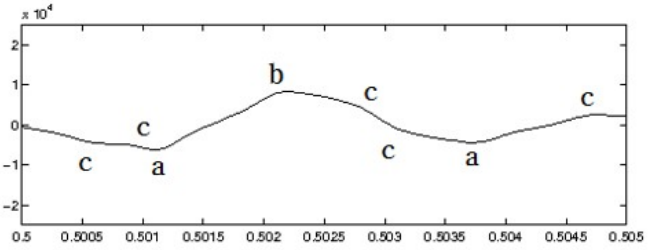

Figure 1. Enlarged plot of a speech signal with various inflection points [11].

desirable for communication channel.

In this paper, a fixed rate speech coder based on non-uniform sampling method is proposed. A speech signal is band passed via filter banks. Each subband signal is non-uniformly sampled using the IP detection scheme, and the obtained IP pattern is normalized by its energy. The coder compares the normalized inflection point signal with candidate patterns in a database, and the addresses of the selected patterns are transmitted with amplitude information. In the receiver, the decoder obtains the candidate IP patterns using the received addresses and amplitudes. The IP patterns are interpolated to obtain the subband signal estimates. The paper is written as follows. In the next section, the IP detection scheme is briefly summarized. In Section 3, the structure of the speech encoder/decoder is explained in detail. In this section, the design of IP pattern database is also considered. Simulation results and conclusions are followed.

2. Inflection Point Detection

A speech signal can be considered as a piecewise linear signal in a short period of time. By sampling non-uniformly at inflection points, the redundant information can be removed, and a smaller number of samples can be obtained. As shown in Figure 1, the inflection points can be taken at sample points where local minima (point a), local maxima (point b), or points of simple slope change (point c) happens.

The inflection point detection (IPD) technique is as follows. For three consecutive samples in uniform sampling, the consecutive differences of samples are defined as

$$d_{21} = x_2 - x_1,$$
$$d_{32} = x_3 - x_2.$$

The sample x_2 is determined as a local maximum or minimum point if the product of the consecutive differences is less than 0,

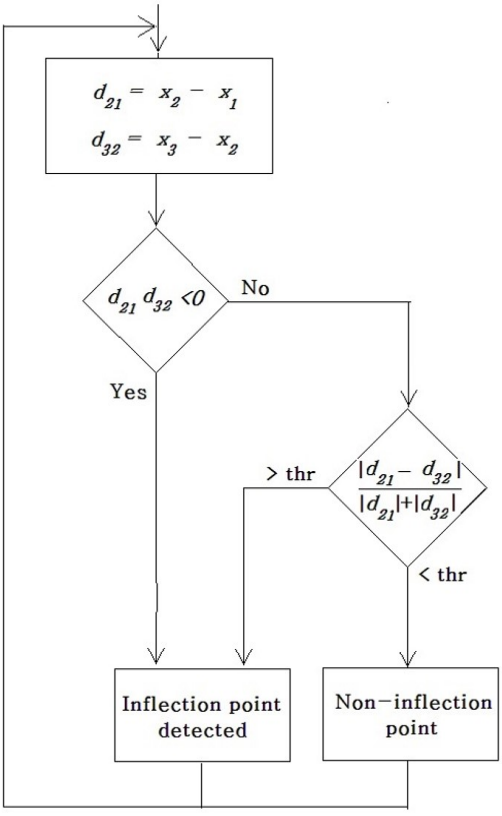

Figure 2. Inflection point detection algorithm [11].

i.e.,

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

The sample point with mere slope change can be obtained by checking following measure [11]:

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

The range of identifier value is $0 < \text{ID} \leq 1$ if there is a slope change. The larger the slope change is, the bigger the ID value is. Thus, by checking if the ID value is larger than a predetermined threshold, the sample point with slope change can be detected. Therefore, the inflection point detection algorithm shown in Figure 2 can be used. That is, if the condition in (1) is satisfied, the sample is determined as a local maximum or minimum. Otherwise, the ID value in (2) is compared to a predetermined threshold. If the value is greater than the threshold, the sample is classified as an inflection point of slope change [11].

3. The Speech Coder

3.1 The 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 [5–10]. In this paper, a new fixed bitrate speech coder is proposed based on the inflection point detection and filter bank method. The structure of the speech coder is shown in Figure 3. A block of speech signal is preprocessed by the filter bank of bandpass filters. And each band pass filtered speech is processed by the IPD algorithm, and the resulting IP pattern is normalized by its energy, and 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. The IP pattern database is furnished with IP patterns of band pass filtered white noise. At the receiver, using the received addresses and the energy information, the decoder reconstructs the speech signal using the same IP pattern database. The decoder fetches the IP patterns from the database using the received addresses, and multiplies the obtained element of the database with the received energy. Then, the decoder performs interpolation and synthesis to get a speech estimates. Thus, the bit stream transmitted over channel consists of the bits for the address and the energy for each subband speech block.

3.2 The IPD Pattern Database

The IP pattern database consists of IP patterns of band pass filtered white noise. The 25,600 white noise samples are taken from a zero mean unit variance Gaussian noise. Every 100 samples are grouped into a block. Each block is processed by M band pass filters. Then, the band-pass filtered signal goes through the IPD procedure. As results, 256 IP patterns are obtained for each subband. A block of target speech is processed by the same filter banks and the IPD algorithm. The band pass filter for the k-th band is defined as

$$h\left[n\right]=\frac{\sin\left(\frac{\pi n}{M}\right)}{\pi n}\cos\left(\pi\left(k-1\right)\frac{n}{M}\right)w[n], \tag{3}$$

where M is the number of subbands and $w[n]$ is a window function.

3.3 Frame Structure

The information transmitted includes the address of the IP pattern database and the energy of the IP block for each band pass filtered speech block. Therefore, the frame over communication channel can be as shown in Figure 4. Each 10 milliseconds speech has M subbands, and each subband should be encoded by IP pattern address and amplitude value. For example, if a speech segment is taken as 10 milliseconds with the sampling frequency of 10 kHz, the block has 100 samples, and there are 100 blocks per second. The number of bits for an address is determined by the size of 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 per subband is $(\log_2 N + L)$ bits/subband * M subbands/block*100 blocks/second where L is the bits for the maximum energy of a detected IPD pattern.

4. Simulation Results

The computer simulation result is provided to show the performance comparison under various situations and to show the usefulness of the proposed speech coding technique. The sampling frequency of a speech is 10 kHz, and the speech is segmented as 10-millisecond blocks with 50% overlapping. And the IP pattern database has 250 entries for each band, so the number of bits for the address is $\lfloor \log_2 250 \rfloor = 8$, where $\lfloor x \rfloor$ is the nearest integer greater than x. And the number of bits for the energy is 8 bits. As results, the data rate is $(3200 * M)$ bits/second when M band pass filters are used in the filter bank. Figure 5 shows the processed signal results when $M = 10$. Figure 5(a) is the original signal, and Figure 5(b) is the reconstructed signal at the receiver. From the figure, the usefulness of the proposed speech coder can be seen. In Table 1, the signal-to-noise ratio (SNR) performance is compared under various conditions for the number of bands in the filter bank method. The SNR of the proposed speech coder is calculated as follows:

$$SNR = 10\log_{10}\left[\frac{signal\ power}{noise\ power}\right], \tag{4}$$

where the noise is the difference between the original and the reconstructed signal. The SNR value is comparable with that of uniform sampling based pulse code modulation (PCM) coder [1]. The SNR performance of the uniform sampling PCM coder is defined as the ratio of signal power to quantization error power, and theoretically given as [1]

$$SNR(dB) = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) = 10\log_{10}\left[\frac{3\cdot 2^{2B}}{\left(X_{\max}/\sigma_i\right)^2}\right], \tag{5}$$

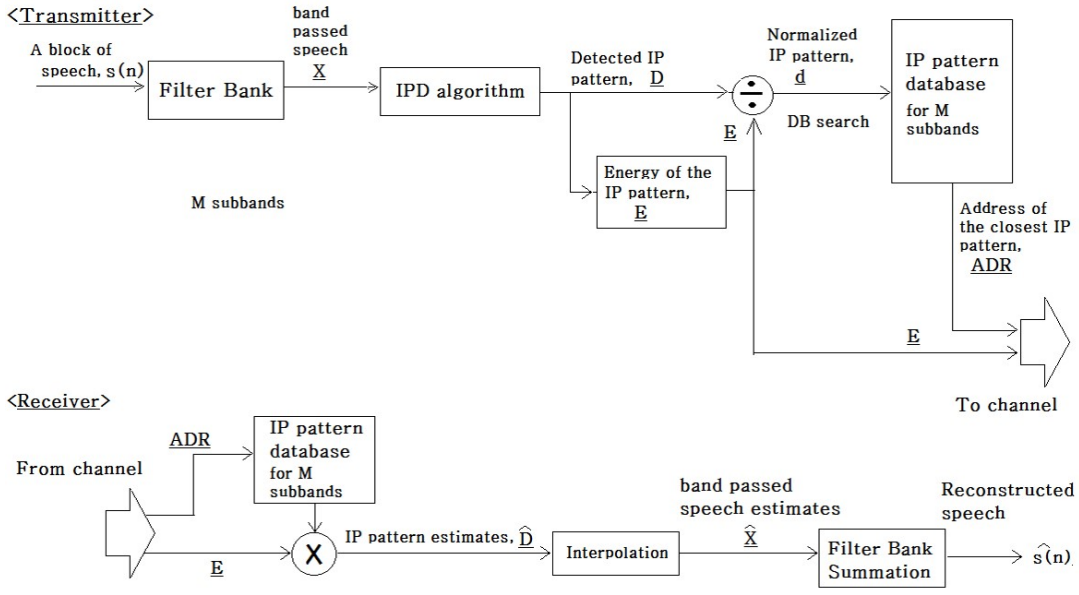

Figure 3. Structure of the speech coder.

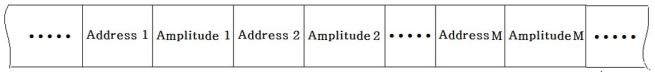

Figure 4. Frame structure over a channel when M subbands are used for a speech block.

where the quantization error power is calculated as $\sigma_e^2 = X_{\max}^2 /(3 \cdot 2^{2B})$ [1]. Here, 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. After some calculation, the SNR is [1]

$$SNR\,(\mathrm{dB}) = 10\log_{10}\frac{\sigma_x}{\sigma_e 2}6B + 4.77 - 20\log_{10}\left[\frac{X_{max}}{\sigma_x}\right]. \tag{6}$$

For example, when $B = 3$ and $X_{\max}/\sigma_x \cong 7.4$, theoretically, $SNR(\mathrm{dB}) \approx 5.4$ dB. If the sampling rate is 10 kHz, the data rate is 30 kbps. And, when $B = 2$ and $X_{\max}/\sigma_x \cong 7.4$, theoretically, $SNR(\mathrm{dB}) \approx -0.6$ dB and the data rate is 20 kbps. In Table 1, with fewer than 7 band pass filters, the proposed IP based coding method shows similar or better SNR performance with much lower data rate comparing to uniform sampling PCM.

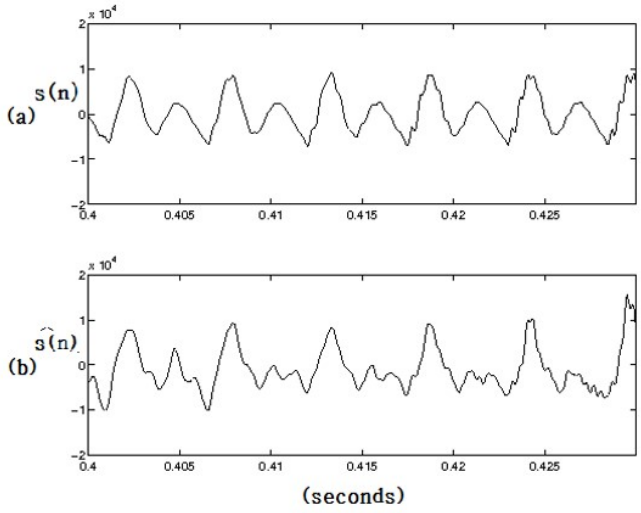

Figure 5. Processing results of the IPD based coding (a) original speech $s(n)$ and (b) reconstructed speech $\hat{s}(n)$.

5. Conclusion

A new speech coding technique based on the non-uniform sampling and filter bank method has been proposed. Unlike existing non-uniform sampling based coding methods, the proposed coder shows a fixed data rate. The inflection points of a band pass filtered speech block are detected and compared with entries of inflection point pattern database. For each subband of a speech block, the address of the closest entry of the database and the energy of the IP pattern are transmitted through channel. At the receiver, the decoder collects the database entries of each subband and reconstructs the speech through interpolation and filter bank summation. The computer simulation has shown

Table 1. SNR performance under various numbers of bands in the filter bank method

Number of bands M	Bit rate (kbps)	SNR (dB)
4	12.8	0.96
5	16	1.22
6	19.2	1.60
7	22.4	1.61
8	25.6	1.50
9	28.8	1.52
10	32	1.52

the usefulness of the proposed speech coding technique. The SNR performance of the non-uniform sampling and filter bank method based coding has been compared with that of the uniform sampling based PCM coding. With relatively much lower bit rate below 20 kbps, the IP based speech coder shows similar SNR to the uniform sampling PCM coder.

Conflict of Interest

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

References

[1] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals.* Englewood Cliffs, NJ: Prentice-Hall, 1978.

[2] T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall, 2002.

[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. J. Bae, W. C. Lee, and D. S. Kim, "On a new vocoder technique by the nonuniform sampling," in *Proceedings of Military Communications Conference (MILCOM'96)*, Mclean, VA, 1996, pp. 649-652. http://dx.doi.org/10.1109/milcom.1996.569428

[6] M. Budaes and L. Goras, "On speech signals reconstruction from local extreme values," in *Proceedings of International Symposium on Signals, Circuits and Systems*, Iasi, Romania, 2005, pp. 315-318. http://dx.doi.org/10.1109/ISSCS.2005.1509917

[7] L. 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] J. Mark and T. Todd, "A nonuniform 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

[9] B. G. Iem, "A nonuniform 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

[10] B. G. Iem, "A nonuniform 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

[11] B. G. Iem, "A low bit rate speech coder based on the inflection point detection," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 15, no. 4, pp. 300-304, 2015. http://dx.doi.org/10.5391/ijfis.2015.15.4.300

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