A Fixed Rate Speech Coder with Improved Database Search

Byeong-Gwan Iem

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

Abstract

A fixed rate speech coder having improved database search performance is proposed. In the speech coder, the 10 ms block of speech is band-passed using the filter bank, and the inflection points (IP) for each band are detected and analyzed to get the short-time statistics like the energy and IP rate. The obtained IP statistics are used to decide if the speech block is voiced or unvoiced. Then the encoder searches for the closest IP pattern either at the voiced part or at the unvoiced part of the IP database. For further reduction of the search time, the entries of the database parts corresponding to the neighboring subbands are checked for the obtained IP pattern. As results, about 94.5% decrease in search time is achieved compared to the exhaustive search. With 10 subbands and 240 entries for each band, the reconstructed signal shows about 5.2 dB SNR at 32 kbps data rate.

Keywords: Non-uniform sampling, Inflection point detection, Unvoiced/voiced decision

1. Introduction

Speech coding techniques exploiting the non-uniform sampling have been studied widely [1–7]. Most of non-uniform sampling based speech coders show variable code rates which are not suitable for a fixed communication channel [1–6]. Recently, a fixed rate speech coder based on the non-uniform sampling technique has been proposed in [7]. The speech coder uses the filter bank method and the inflection point (IP) detection technique. The obtained IP pattern for each subband is compared with entries in IP pattern database, and the address of the closest pattern is transmitted through communication channel. In the receiver, the decoder fetches the IP pattern from the same database with the received addresses, and reconstructs the speech with interpolation and filter bank summation [7]. The obtained result shows acceptable quality of speech with the relatively low bit rate. However, the search for the closest IP entry of the database at the encoder requires a lot of calculation and time, which is an obstacle to the implementation of the speech coder. Thus, an efficient search method exploiting the IP characteristics is desirable for the realization of the speech coder based on the non-uniform sampling.

The IP shows different patterns depending on the blocks processed. That is, if a speech block belongs to the unvoiced speech, the obtained IP samples shows small and alternating amplitude values. On the other hand, if the block is from the voiced, then the IP pattern has large amplitude. By analyzing the obtained IP patterns, the IP pattern database can be searched more efficiently. In this paper, several important speech parameters such as the

Received: Dec. 6, 2017
 Revised : Dec. 17, 2017
 Accepted: Dec. 20, 2017

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 [7].

voiced/unvoiced decision are obtained using the IP pattern analysis. The paper is written as follows: In the next section, the IP detection scheme is briefly reviewed. In Section 3, the characteristics of the inflection points are explained in detail. In this section, the design of IP pattern DB is also considered. Simulation results and conclusions are followed.

2. Inflection Point Detection

A speech signal can be approximated as a piecewise linear signal in a short period of time. By sampling irregularly at inflection points, a smaller number of samples can be obtained compared to the conventional uniform sampling. As shown in Figure 1, the inflection points can be one of three types: local minima (point *a*), local maxima (point *b*), or points of simple slope change (point *c*) [7].

For three consecutive samples x_1 , x_2 , and x_3 in uniform sampling, the differences of neighboring samples are defined as

$$\begin{aligned} d_{21} &= x_2 - x_1, \\ d_{32} &= x_3 - x_2. \end{aligned}$$

A local maxima or minima point x_2 is selected as inflection points if the product of the consecutive differences is less than 0, i.e.,

$$d_{21} \cdot d_{32} < 0. \quad (1)$$

Another kind of inflection point is the sample point with slope change which can be obtained by following measure [7]:

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

The larger the slope change is, the bigger the ID value is. Thus, by comparing the ID value in (2) with predetermined threshold ranged between 0 and 1, the inflection points can be detected [7]

Figure 2. Inflection point detection algorithm [7].

(Figure 2).

3. Characteristics of the Inflection Points

3.1 Short-Time Energy of the IP

Since the inflection points are samples from a speech signal, their amplitudes reflect where they are from. That is, if the speech is from the voiced block, the IP points show relatively high values. On the other hand, if the IP are from the unvoiced, they have small values. As results, the short time energy of the IP is similar to that of ordinary speech defined as [8]

$$E(n) = \frac{1}{L} \sum_{m=-\infty}^{\infty} s^2(n)w(n-m), \quad (3)$$

where $w(n)$ is the rectangular window of length L . Figure 3 shows simulation results of the short-time energy of uniformly sampled speech defined as in (3) and that of the IP signal. The short time energy of the IP resembles that of the original speech. Thus, the short time energy of the IP can be used to decide if a speech block is voiced or unvoiced. This information can be exploited for an efficient search of a proper IP pattern if

Figure 3. Comparison of the short-time energy. (a) Original speech signal, (b) short-time energy of a speech signal and (c) short-time energy of the inflection points.

a database is well organized reflecting the voiced/unvoiced information.

3.2 Comparison of the Rate of IP and the Zero Crossing Rate

The zero crossing rate (ZCR) is one of the well-known and widely used speech characteristics. It is closely related to the instantaneous frequency of a narrowband signal [9]. The ZCR counts the number of zero crossings of a signal and takes their average as defined in (4) [8]

$$\begin{aligned} \text{ZCR}(n) &= \frac{1}{2L} \sum_{m=-\infty}^{\infty} |\text{sign}(s(n)) - \text{sign}(s(n-1))| w(n-m). \end{aligned} \quad (4)$$

Here, $\text{sign}(\cdot)$ gives 1 or -1 depending on the sign of the value $s(n)$. The ZCR of a voiced speech block shows relatively low values whereas that of an unvoiced speech produces high value. Some of the inflection points may be exploited to get information similar to the ZCR. Local maxima and minima classified as inflection points of speech also reflect where the speech block is from. That is, the inflection points of the voiced speech are less frequent whereas those of the unvoiced are frequent. Thus, the rate of the inflection points in (5) can be used as a voiced/unvoiced detector like the ZCR. Figure 4 compares the ZCR and the IP rate of a speech to show their

Figure 4. Comparison of the zero crossing rate and the IP rate. (a) The original speech signal (b) the zeros crossing rate, and (c) the IP rate.

similarity. As can be seen in Figure 4, the shapes of the ZCR and the IP rate show the opposite pattern of the short time energy of the IP. Thus, the short time energy and the IP rate can be used to decide the unvoiced/voiced together.

$$\text{IPR}(n) = \frac{1}{L} \sum_{m=-\infty}^{\infty} |\text{sign}(s(n))| w(n-m). \quad (5)$$

3.3 Structure of the Speech Coder

At the transmitting side, the encoder detects the inflection points of a band-passed speech block. And the obtained inflection point patterns of subbands are compared with entries of IP pattern database. Before comparison, the encoder decides if the speech block is voiced or unvoiced using the short time statistics of the obtained IP pattern. Thus, the database search is performed not for whole database but only for part of the database. In this way, the search time can be reduced dramatically. At the same time, the database is aligned according to the order of subbands to reduce the scope of search into neighboring subbands. In this paper, each band has 240 entries of IP patterns for the unvoiced or the voiced. After the voiced/unvoiced decision, the database search process is performed for these 240 entries. The encoder/decoder structure is shown in Figure 5. The transmitted information includes the address of the IP pattern database and the energy of the IP block for each band pass filtered speech block.

Figure 5. Structure of the speech coder.

4. Simulation Results

The computer simulation has been performed to show the improvement in the database search and the usefulness of the proposed speech coding technique. The sampling frequency of a speech is 10 kHz, and the speech is segmented as 10 ms blocks with 50% overlapping. The length of the speech is 2.12 seconds. The IP pattern database has 240 entries for each band, so the number of bits for the address is $\lceil \log_2 240 \rceil = 8$ where $\lceil x \rceil$ 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. The search speed has been compared for two different search scenarios. One is the exhaustive search of the whole database and the other is the proposed search using the short time analysis of the IP pattern. The energy and rate of the obtained IP are calculated, and the speech block is classified as the voiced or the unvoiced. And then parts of the database are checked to reduce the search time. Matlab command 'cputime' has been used to compare two search methods. The result is 94.5% reduction of CPU time when the IP analysis and neighboring band search was used. Figure 6 shows the processed signal results when $M = 10$. Figure 6(a) is the original signal, and Figure 6(b) is the reconstructed signal at the receiver. From the figure, the usefulness of the proposed speech coder can be seen. The signal to noise ratio (SNR) for this reconstructed signal was calculated

to be 5.2 dB using

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

where the noise is the difference between the original and the reconstructed signal. The theoretical SNR performance of the conventional uniformly sampled PCM is around 5.4 dB at 30 kbps, and previously, the non-uniform sampling method in [7] produces 1.52 dB at around 30 kbps data-rate. The performance of the proposed method approaches that of the uniformly sampled PCM, and is better than the previous method in [7].

5. Conclusion

A new fixed rate speech coding technique with reduced database search has been proposed. The speech coder uses the non-uniform sampling at inflection points and filter bank method. The conventional speech statistics like the short time energy and zero crossing rate have been applied to the obtained inflection points to reduce the database search time. Unlike existing fixed rate non-uniform sampling coding method, the proposed coder searches the IP database partially. By using the short time statistics like the energy and IP rate, the encoder searches either the voiced part or the unvoiced part of the database. Furthermore, by exploiting subband information, the encoder compares IP entries at the neighboring bands. As results, the

Figure 6. Processing results of the IPD based coding: (a) original speech $s(n)$, and (b) reconstructed speech $\hat{s}(n)$.

database search time decreases about 94%. The reconstructed speech shows SNR of 5.2 dB with 32 kbps data rate when 10 subbands are used. The optimal database to get improved SNR should be considered as future study topic.

Conflict of Interest

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

References

- [1] M. Bae, W. Lee, and S. Kim, "On a new vocoder technique by the nonuniform sampling," in *Proceedings of Military Communications Conference (MILCOM)*, Mclean, VA, 1996, pp. 649-652. <http://dx.doi.org/10.1109/milcom.1996.569428>
- [2] M. Budaes, and L. Goras, "On speech signal reconstruction from local extreme values," in *Proceedings of International Symposium on Signals, Circuits and Systems (ISSCS)*, Iasi, Romania, 2005, pp. 315-318. <http://dx.doi.org/10.1109/isscs.2005.1509917>
- [3] J. Mark, and T. Todd, "A nonuniform sampling approach to data compression," *IEEE Transactions on Communica-*
- tions*, vol. 29, no.1, pp. 24-32, 1981. <http://dx.doi.org/10.1109/TCOM.1981.1094872>
- [4] B. G. Iem, "A nonuniform sampling technique based on inflection point detection and its application to speech coding," *Journals of the Acoustical Society of America*, vol. 136, no. 2, pp. 903-909, 2014. <http://dx.doi.org/10.1121/14884882>
- [5] 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>
- [6] 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>
- [7] B. G. Iem, "A fixed rate speech coder based on the filter bank method and the inflection point detection," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 16, no. 4, pp. 276-280, 2016. <http://dx.doi.org/10.5391/ijfis.2016.16.4.276>
- [8] L. Rabiner, and R. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [9] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. Part 2: Algorithms and applications," *Proceedings of IEEE*, vol. 80, no. 4, pp. 540-568, 1992. <http://dx.doi.org/10.1109/5.135378>

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