대역분리-비균일표본화 방법을 이용한 새로운 음성신호의 파형부호화 연구

A New Speech Waveform Coding Based on the Nonuniform Sampling Method with Separated to High-Low Band

(배 명 진*, 이 주 헌*, 임 성 빈*, 이 원 철*)
MyungJin Bae*, JooHun Lee*, SungBin Im*, WonCheol Lee*

Abstract

To reduce the redundancy within samples that resulted from uniform sampling method, nonuniform sampling or nonredundant-sample coding methods can be considered. However, it is well known that when conventional nonuniform sampling methods are applied directly to speech signal, the required amount of data is comparable to or more than that by uniform sampling method like PCM. To overcome this problem, a new nonuniform sampling method is proposed, in which nonuniform sampling is applied to the low-pass filtered speech signal and higher band is compensated by 8 colored Gaussian random noise with various noise levels. By this method, speech signal waveform can be encoded by 1.8 times larger compression ratio than the conventional nonuniform sampling method.

요 약

균일표본화에서 나타나는 샘플간의 잉여정보를 더욱 줄임으로써, 요구되는 데이타량을 크게 줄일 수 있는 방법으로 비균일표본화 방법이 고려된다. 그러나, 음성신호의 경우 이러한 비균일표본화 방법을 바로 적용하면, 필요한 데이타량이 균일표본화에 견주어 크게 줄어들지 않게 된다. 특히, 잡음환경하에서는 오히려 균일표본화의 경우보다도 데이타량이 커질 수 있다. 이러한 단점을 보완하기 위해서, 먼저 음성신호를 적당히 저대역 필터링을 한 후 비균일표본화를 적용하고, 고대역성분에서의 오차는 잡음신호로 보완하는 방법을 제안한다. 제안된 방법은 기존의 비균일표본화 방법보다 약 1.8배의 데이타압축효과를 얻을 수 있었다.

I. Introduction

The major objectives of speech coding are how much the transmission rate and/or data storage requirement can be reduced, how high quality of the decoded speech signal can be obtained and how fast the coding/decoding can be processed. In general, coding methods are classified into the following three categories : waveform coding, source coding and hybrid coding. Among them, from the viewpoints of intelligibility and naturalness, waveform coding is preferable to maintain high quality by preserving the shape of the waveform itself. This method is based on the sampling technique which consequently removes the inherent redundancy of waveform. PCM, ADM, DPCM, and ADPCM have been searched as one of the waveform coding methods.

However, since the inherent redundancy of waveform is not completely removed by uniform sampling, waveform coding method still has the major drawback to require large amount of data[5]. It means that there is still the redundancy of waveform left in the uniformly sampled data. These unnecessary

*Dept. of Telecommunication Engineering, Soongsil University, Seoul 156-743, Korea
숭실대학교 정보통신공학과
접수일자 : 1995년 7월 28일

redundant samples come from the relatively high correlation between the neighboring samples obtained by uniform sampling method. However, Licklider and Pollack experimental results show that the intelligibility test of differentiated and clipped waveform scores 97%, comparable to that of the original waveform, 99%. This means that the most significant information in the sense of intelligibility is the interval between maxima and minima since the uniquely remained information of the differentiated and clipped signal is the time index when the peaks occur. Therefore, according to the intelligibility test by Licklider and Pollack, those samples between the peak and the valley points may be considered as redundancy and ignored without loss of intelligibility.

To remove the redundant samples in the uniform sampling method, nonuniform sampling or nonredundant-sample coding method has been considered [4] and such researches as polynomial predictor [8] and interpolator using pan-algorithm [9] were proposed for nonuniform sampling technique. However, since these algorithms use the differences of the magnitudes or slopes between the neighboring samples, they still need large amount of data. Therefore, it is well known that those algorithms are improper to speech signal since the waveform varies rapidly and has nonstationary characteristics. Moreover, especially in noisy environment, the required amounts of data of those algorithms are comparable to or more than that of PCM.

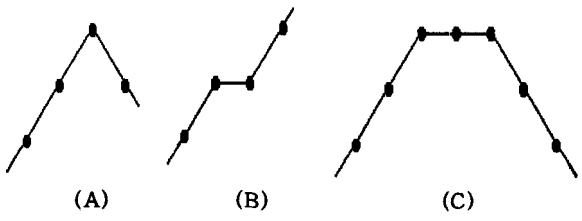


Fig. 1 Consideration on maxima and minima of PCM signal;
(a) normal signal,
(b) unresolved maxima and minima by quantization error,
(c) unresolved maxima by quantization error.

Ⅱ. The Conventional Nonuniform Sampling Method (CNSM)

In the sense of human perception, the information only related to the peak and the valley samples is enough to reconstruct the original speech signal. Therefore, the rest samples except the peak and the valley points in the uniform sampling are considered as redundancy in speech coding. To remove those redundant samples, nonuniform sampling technique can be considered. The peak and the valley points in nonuniform sampling are determined by examining the sign of the multiplication result of the consecutive 2 slopes obtained from the adjacent 3 samples. If the sign is plus, those samples are considered on the increasing or decreasing segment and therefore, neither peak nor valley exists in that segment. On the contrary, if the sign is minus, the sample in the middle of that segment may be the peak of the valley. More careful consideration is necessary when the sign is zero. In that case, 2 kinds of unresolved maxima and/or minima can be considered. Fig. 1-(b) and (c) show the examples of them, respectively. In case of fig. 1-(b), it is expected that there are one maxima and one minima in the original waveform. On the contrary, such case as fig. 1-(c) may happen when the excessive waveform is clipped and one maxima or minima is expected to be in the original waveform. Usually, the middle sample is considered as a peak or a valley sample. According to the above procedure, the peak and the valley points are determined and their magnitudes and the intervals are stored in buffer for transmission and reconstruction.

Waveform reconstruction is performed by using cosine interpolation method based on such parameters as the magnitudes and the intervals of the peak and the valley samples[8]. The reconstructed waveform, $y_k(n)$, obtained by the cosine interpolation method is represented as follows;

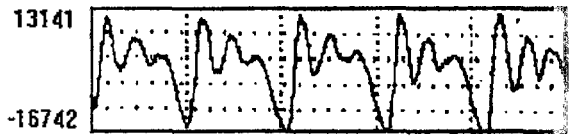
$$y_k(n) = [\frac{Mag(k-1)-Mag(k)}{2} \cos(\frac{\pi n}{Inter(k)}) + \frac{Mag(k-1)+Mag(k)}{2}]_L, \; I \le n \le Inter(k) \qquad (1)$$

where, Mag(·) is the magnitude of nonuniformly sampled data and Inter(·) is the interval of them. An example of reconstruction waveform using cosine interpolation is shown in Fig. 2. Fig. 2-(a) is the original waveform and interpolation and Fig. 2-(b) is the reconstructed waveform using cosine interpolation.

In noisy environment, however the required amount of nonuniformly sampled data may be comparable to that of uniformly sampled data because of

its higher frequency feature. To reduce the data rate without losing the merit of nonuniform sampling, a new nonuniform sampling method using separated high-low band for speech signal is proposed. In this method, speech signal is low-pass filtered by 2.67 kHz and then, nonuniform sampling is applied to this filtered signal to determine the magnitudes and their intervals of the peak and the valley points as coding parameters. To compensate the high frequency band, Gaussian random noise is added to the signal reconstructed by those parameters at the decoding part. Level and selection of eight Gaussian random noise is obtained from the difference signal between the original signal and the reconstructed signal at encoding part.

(A) original waveform



(b) cosine interpolation using peaks and valley

Fig. 2 Signal reconstruction by cosine interpolation ;
(a) Original waveform,
(b) Reconstructed waveform by cosine interpolation.

Ⅲ. The Proposed Nonuniform Sampling Method with Separated High-Low Band

According to the speech production mechanism, since higher frequency band is related to the sound produced from the constriction structure rather than from the resonance structure, the 3rd and the 4th formants have larger bandwidths. Moreover, from the viewpoint of speech perception, the higher frequency band components are not significant while the 1st and the 2nd formants are indispensible to reconstruct the high-intelligible speech. Therefore, the samples related to the frequency band higher than the 2nd formant are considered as redundant information in the speech perception. As shown in table 1, the 1st and the 2nd formant frequencies of most phonemes are less than 2.5 kHz. Also, the formants higher than this cut-off frequency have quite broad bandwidths. Therefore, nonuniform sampling can be applied only to the signal component of the original waveform less than 2. 5 kHz without significant degradation of intelligibility. Since the low-pass filtered signal is smoother than the original one, a relatively smaller number of the peak and the valley samples is obtained when nonuniform sampling is performed on it. For this reason, it is possible to achieve high compression ratio.

Table 1. Formant frequencies and their bandwidths of vowels.

[Hz]	/a/	/i/	/o/	/u/	/e/
F1	692	332	377	359	485
F2	1178	2231	611	836	1970
F3	2564	3049	2753	2492	2663
F4	3445	3562	3525	2914	3229
B1	84	55	46	36	80
B2	79	88	84	53	126
B3	107	232	118	841	167

To compensate the naturalness of speech, random Gaussian noise is added to the waveform reconstructed roughly with those filtered parameters. Generally, since the characteristic of the error signal between the original and the reconstructed low-band waveform is rather colored than white, we can roughly approximate the error signal as one of eight colored white Gaussian noise. By using this procedure, higher compression ratio can be obtained without serious loss of intelligibility and naturalness.

Fig. 3 shows the block diagram of the proposed method. In this diagram, s(t) is an analog signal to be coded and s(n) is its digitized signal by A/D converter and sampler. Then, s'(n) is the low-pass filtered signal at 2.67 kHz.

$$s'(n)=\frac{1}{M}\sum_{i=1}^{M-1}s(n-i), \tag{2}$$

Where M is the window size of LPF.

The conventional nonuniform sampling is applied to this low-pass filtered signal and such parameters as the magnitudes and the intervals of the peak and the valley points are stored into buffer to be transmitted. At the same time, with those parameters, s"(n) is reconstructed by the cosine interpolation technique to be compared with the original waveform. Then, e(n), the error signal between the original signal

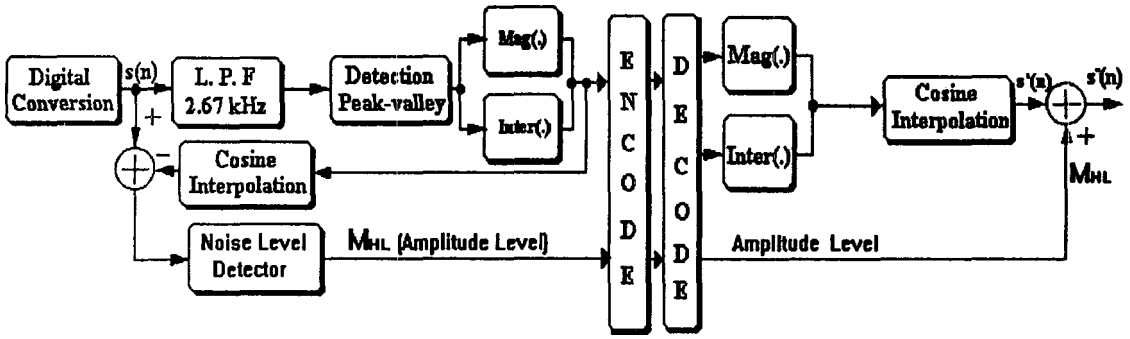


Fig. 3 Block diagram of the proposed method

and the reconstructed low-pass filtered signal is obtained. Noise level of this error signal, M_{HL} is parameterized by calculating its average for the analysis frame. Eq. (3) and Eq. (4) represent those e(n) and M_{HL} respectively.

$$e(n)=s(n)-s'(n), \tag{3}$$

$$M_{HL}=\frac{1}{N}\sum_{n=0}^{N-1}|e(n)|, \tag{4}$$

where N is the frame size.

Then, the buffered parameters and noise level, M_{HL} are quantized and transmitted to the decoder. This procedure can much reduce the data rate to achieve higher compression ratio than the conventional nonuniform sampling even in the noisy environment.

Ⅳ. Experimental Results

To compare the performances between the conventional nonuniform sampling method and the proposed method, ten phoneme-balanced Korean sentences were used. Each sentence was pronounced five times by 1 female and 2 male speakers. For simulation test, the speech signal was sampled at 8 kHz, low-pass filtered at 4 kHz and digitized with a 16 bits A/D converter. The simulation was performed by using personal computer (IBM-PC/pentium 75㎒).

Fig. 4 shows some examples of the proposed method. In this figure, (a) is the original waveform and (b), (c) are the reconstructed waveforms by using the conventional nonuniform sampling method and the proposed method, respectively. As shown in the examples, the proposed method reconstructs the original waveform with much lower data rate, that is, higher compression ratio.

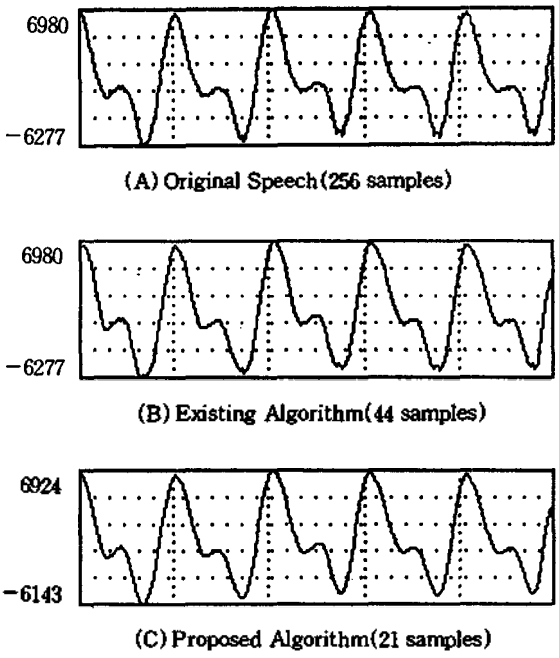


Fig. 4 Examples of the proposed method ;
(a) Original waveform,
(b) Reconstructed waveform by the conventional nonuniform sampling method,
(c) Reconstructed waveform by the proposed method.

Table 2 shows the comparison results of the conventional nonuniform sampling and the proposed method. From the table, the average compression ratio compared to 64 kbps μ-law PCM of the conventional nonuniform sampling method is 2.79 and that of the proposed method is 5.12. Therefore, the proposed method achieved 1.8 times higher compression ratio than the conventional nonuniform sampling method with little degradation of segmental SNR.

Table 2. Comparison results of SEGSNR and compression ratio between the conventional NSM and the proposed method.

	conventional NSM		proposed method	
	SEGSNR (dB)	Compression ratio	SEGSNR (dB)	Compression ratio
sent.1	15.04	2.59	14.86	5.23
sent.2	14.79	2.83	13.35	5.18
sent.3	14.93	2.59	14.21	4.95
avg.	14.92	2.79	13.84	5.12

V. Conclusion

The main objective of speech coding includes (1) high compression ratio for storage of transmission, (2) high synthesized speech quality in terms of the intelligibility and the naturalness and (3) fast processing speed. Generally, waveform coding is preferable to obtain high speech quality. However, the major drawback to that coding technique is that it requires large amounts of data. This large required data rate results from the inherent redundancy of uniform sampling. Higher the correlation between the neighboring samples of signal waveform increases, larger this redundancy does. To remove the redundancy, nonuniform sampling method was proposed. However, the conventional nonuniform sampling method is improper to a speech signal since speech is nonstationary especially when a speech signal changes rapidly. In that case, the data rate of the conventional nonuniform sampling method becomes comparable to that of the uniform sampling method such as PCM.

To overcome this problem a new nonuniform sampling method using separated high-low band. In the proposed method, the conventional nonuniform sampling technique is applied to the low-pass filtered speech signal to reduce the data rate without losing the 1st and the 2nd formants information, and higher band component is compensated by adding and selecting one of eight Gaussian noise to the reconstructed to form the final decoded speech. Experimental results with phoneme balanced Korean sentences show that the proposed method can achieve higher compression ratio with little degradation of segmental SNR compared with the conventional nonuniform sampling method. The average compression ratio compared to 64 kbps μ-law PCM of the proposed method is 5.12 while that of the conventional nonuniform sampling method is 2.79. This means that, with the proposed method, speech signal can be compressed 1.8 times higher than the conventional nonuniform sampling method without serious deterioration of the intelligibility and the naturalness.

References

1. M. J. Bae, D. S. Kim, H. Y. Jeon and S. G. Ann, "On a new predictor for the waveform coding of speech signal by using the dual autocorrelation and the sigma-delta technique," IEEE Proc. of ISCAS'94, Vol. 6, No. 3, pp. 261-264, June 1994.
2. T. J. Lynch, Data compression : Techniques and Applications, Lifetime Learning Pub., 1985.
3. Panos E. Papmichalis, Practical approaches to speech coding, Prentice-Hall, 1984.
4. J. W. Mark and T. D. Todd, "A nonuniform sampling approach to data compression," IEEE Trans. on Com., Vol. COM-29, No. 1, pp. 24-32, Jan. 1981.
5. N. S. Jayant and P. Noll, Digital Coding of Waveforms-Principles and Applicants to Speech and Video, Prentice-Hall, 1978.
6. L. R. Rabiner and R. W. Schafer, Digital processing of speech signals, Prentice-Hall, 1978.
7. T. J. Lynch, "The probability of a straight-line sequence from a uniform independent sample source," IEEE Trans. on Info. Theory, Vol. IT-14, No. 5, pp. 773-774, Sept. 1968.
8. L. D. Davission, "Data compression using straight line interpolation," IEEE Trans. on Info. Theory, Vol. IT-14, No. 3, pp. 390-394, May 1968.
9. L. Ehrman, "Analysis of some redundancy removal bandwidth compression technique," Proc. IEEE, Vol. 55, No. 3, Mar. 1967.

▲배 명 진(MyungJin Bae)
현재 : 숭실대학교 정보통신공학과 교수

▲이 주 헌(JooHun Lee)
현재 : 숭실대학교 정보통신공학과 연구원

▲이 원 철(WonCheol Lee)
현재 : 숭실대학교 정보통신공학과 교수

▲임 성 빈(SungBin Im)
현재 : 숭실대학교 정보통신공학과 교수