On a Multiband Nonuniform Sampling Technique with a Gaussian Noise Codebook for Speech Coding

가우시안 코드북을 갖는 다중대역 비균일 음성 표본화법

HyungGoue Chung*, MyungJin Bae*
정 형 교*, 배 명 진*

※이 연구는 1996년도 숭실대학교 교내연구비 지원금에 의해 이루어졌습니다.

ABSTRACT

When applying the nonuniform sampling to noisy speech signal, the required data rate increases to be comparable to or more than that by uniform sampling such as PCM. To solve this problem, we have proposed the waveform coding method, multiband nonuniform waveform coding(MNWC), applying the nonuniform sampling to band-separated speech signal[7]. However, the speech quality is deteriorated when it is compared to the uniform sampling method, since the high band is simply modeled as a Gaussian noise with average level. In this paper, as a good method to overcome this drawback, the high band is modeled as one of 16 codewords having different center frequencies. By doing this, with maintaining high speech quality as MOS score of average 3.16, the proposed method achieves 1.5 times higher compression ratio than that of the conventional nonuniform sampling method (CNSM).

요 약

잡음 음성신호에 비균일 표본화 부호화법을 적용하면, PCM 균일표본화의 전송율 정도로 데이타 전송율이 높아진다. 이러한 문제점을 해결하기 위해 비균일 표본화법을 성분분리된 음성신호에 적용하는 방법으로서 다중대역 비균일 파형 부호화(MNWC)법을 제안하였었다[7]. 그렇지만, 고대역의 성분에 대해 가우시안 잡음의 평균레벨로 단순하게 모델링 하였기 때문에, 비균일 표본화법에 비해 음질의 열화가 초래되었었다. 따라서 본 논문에서는 이러한 단점을 극복하기 위해 고대역의 성분을 중심주파수가 서로 다른 16가지의 가우시안 잡음으로 모델링하였다. 이렇게 하였을 때, 제안된 방법은 MOS평가가 평균 3.16 정도로 고음질을 유지하면서도 기존의 비균일 표본화법에 비해 1.5배 정도의 압축 율을 얻을 수 있었다.

Ⅰ. Introduction

The major objectives of speech coding include high compression ratio for limited bandwidth of transmission, high synthesized speech quality in terms of the intelligibility and the naturalness and fast processing speed. In general, coding methods are classified into the following three categories: the waveform coding, the source coding and the hybrid coding. Among them, from the viewpoints of intelligibility and naturalness, the 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[2].

However, since the inherent redundancy of waveform is not completely removed by uniform sampling, the waveform coding method still has the major drawback to require large amount of data[1][6]. It means that there is still the redundancy of waveform in the uniformly sampled data. These redundant samples come from the relatively high correlation between the neighboring samples obtained by uniform sampling method.

To remove the redundant samples in the uniform sampling

*숭실대학교 전기공학과
접수일자: 1997년 5월 9일

method, nonuniform sampling or nonredundant-sample coding method has been considered[1]. Such researches as polynomial predictor[4] and interpolator using pan-algorithm[5] 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[5].

To reduce the data rate without losing the merit of nonuniform sampling, we have proposed the nonuniform sampling method, multi-band nonuniform waveform coding (MNWC), using the separated high-low band and the nonuniform sampling method for speech signal[7]. In the method, speech signal is separated into two bands, the low and the high band. At the first, 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 the intervals of the peak and the valley points as coding parameters. Then, the high frequency band is modeled as a Gaussian random noise with an average noise level.

The noise level is obtained from the difference signal between the original and the temporary signal reconstructed with only the low band parameters at the encoding part. Therefore, the parameters of the MNWC are the nonuniform sampling parameters of the low-pass filtered speech signal and the noise level of Gaussian noise which will be added to the reconstructed signal. For several Korean sentences, this conventional MNWC achieved higher compression ratio, 5.12, when compared with the conventional uniform sampling method. However, in spite of the higher compression ratio, the synthesized speech quality of MNWC is worse than that of the uniform sampling method.

To compensate for quality degradation of MNWC, we proposed the modified MNWC, which the high band is modeled by one of 16 Gaussian random noise with different center frequencies as well as different noise levels.

Ⅱ. The Conventional Multiband Nonuniform Waveform Coding

In the sense of 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 these redundant samples, a 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, both peak and valley don't exist in that segment. On the contrary, if the sign is minus, the sample in the middle of that segment may be the peak or the valley. More careful consideration is necessary when the sign is zero[1][7].

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 points. The reconstructed waveform, $y_k(n)$, obtained by cosine interpolation method is represented as follows:

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

where, Mag(·) is the magnitude of nonuniformly sampled data and Inter(·) is the interval between them. However, in noisy environment, the required amount of information in nonuniformly sampled data may be comparable to that of uniformly sampled data because of its higher frequency feature. Therefore, a modified nonuniform sampling method is necessary to cope with that kind of problem.

According to the speech production mechanism, since higher frequency band is related to the sound produced from the constriction structure than from the resonance structure, the 3rd and the 4th formants in that band have broad bandwidths. Moreover, from the viewpoint of speech perception, 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 and can be ignored in the sampling procedure.

Since the 1st and the 2nd formant frequencies of most phonemes are less than 2.5 kHz and the formants higher than this cut-off frequency have quite broad bandwidths, nonuniform sampling can be applied only to the signal component of the original waveform less than 2.5 kHz without terrible loss of intelligibility. Since the low-pass filtered signal is smoother than the original one, the small number of the peak and the valley samples are obtained when nonuniform sampling is performed on it. This makes it possible to achieve high compression ratio. To compensate the naturalness random Gaussian noise is added to the waveform which is temporarily reconstructed with those filtered parameters. By doing this, the higher compression ratio with good speech quality can be obtained. However, the characteristic of the error signal between the original and the reconstructed low-band waveform is rather colored than white. the rough approximation of high band occurs a quality degradation compared to the conventional nonuniform sampling method(CNSM). Therefore, the conventional MNWC is modified as following section.

Ⅲ. The Multiband Nonuniform Waveform Coding with Noise Codebook

Fig. 1 shows the block diagram of a method proposed in this paper, multiband nonuniform waveform coding with a noise codebook. As shown in this figure, the basic structure of the proposed method is almost same as that of the conventional MNWC, but major difference is in the noise codebook. This codebook consists of 16 Gaussian random noises having different center frequencies.

In this block diagram, s(n) is speech signal digitized by A/D conversion and is performed by 2'nd order low-pass filter with stopband at 2.67 kHz. The filtered signal $s_L(n)$ is given by

$$s'(n) = \frac{1}{N}\sum_{i=0}^{N-1} a(i)s(n-i), \tag{2}$$

where N is set to 5 as the windowing size of LPF and a(.) as filter coefficients with a(0) = a(4) = 0.1, a(1) = a(3) = 0.2, and a(2) = 0.4.

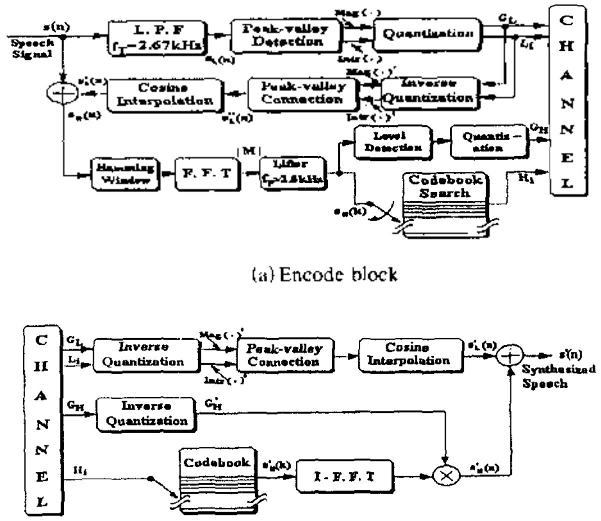
Then, the conventional *nonuniform sampling* is applied to this low-pass filtered signal, $s_L(n)$, the magnitudes of the peak and the valley points and their intervals are parameterized as *Mag*(·) and Inter(·), and then quantized as G_L and L_i, respectively. The signal waveform, $s_L{}'(n)$, of the low frequency is reconstructed by using these *Mag*(·) and Inter(·) and by using the cosine interpolation technique, and then subtracted from the original waveform. Therefore, the residual signal, $s_H(n)$, is obtained. A level of the residual signal is estimated by calculating the average of $s_H(n)$ for every analysis frame and is parameterized as G_H.

The noise codebook proposed in this method consists 16 colored Gaussian noises with different center frequencies. Their center frequencies are uniformly arranged with frequency interval of 75 Hz and bandwidth of 40 Hz in range between 2.75 kHz and 3.95 kHz. Unlike the codebook searching in CELP vocoder, the optimum index, H_i, of the codebook is found by the spectral matching technique between the magnitude spectrum of $s_H(n)$ and one of these 16 codewords.

In the decoding part, speech signal, s'(n) is synthesized by using these parameters as follows,

$$\begin{aligned} s'(n) &= s'_L(n) + s'_H(n) \\ &= Y|Mag(.), Intr(.)| + G_H{}' F^{-1}|S_H{}'(H_i)| \end{aligned} \tag{3}$$

where Y|.| and F^{-1}|.| are function of the cosine interpolation and the inverse Fourier transform, respectively.



(a) Encode block

(b) Decode block

Fig 1. Block diagram of the proposed multiband nonuniform waveform coding with noise codebook.

Ⅳ. Experimental Results

To compare the performances between the conventional nonuniform sampling method and the proposed method, it was used three phoneme-balanced Korean sentences. Each sentence was pronounced five times by 1 female and 2 male speakers. For a simulation test, speech signal was sampled at 16 kHz, low-pass filtered at 3.4 kHz and digitized with a 16 bits A/D converter. Th simulation was performed by using personal comput (IBM-PC/586-166MHz).

Fig. 2 shows some examples of the proposed method. In this figure, (a) is the original waveform and (b), (c) are the reconstructed waveform by using the cosine interpolation and the proposed method respectively. As shown in the example, the proposed method reconstructs the original waveform with much reduced data rate, that is, higher compression ratio.

Table I shows the comparison results of the conventional nonuniform sampling, the conventional MNWC and the proposed method. From the table 1, the average compression ratios compared to 64 kbps -Law PCM of the conventional nonuniform sampling method and the conventional MNWC are 2.79 and 5.12 respectively. Their MOS scores are 3.9 and 3.5 respectively. On the other hand, in case of the proposed method, the average compression ratio is 5.06 and the MOS score is 3.16. Consequently, it can be said that the proposed method improves the MOS score much while maintaining high compression ratio, compared to the conventional MNWC.

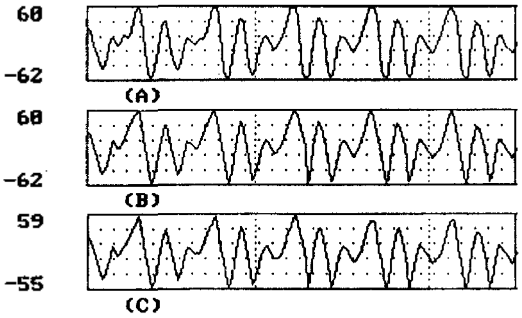


Fig 2. A waveform example of the proposed method :
(a) speech waveform for /uh/,
(b) LPFed waveform synthesized by the nonuniform sampling method,
(c) waveform synthesized by the proposed method.

Table 1. Comparison results of MOS score and compression ratio between the conventional NSM and the proposed method

sentence	conventional NSM		conventioanl MNWC		proposed method	
	Compression ratio	MOS score	Compression ratio	MOS score	Compression ratio	MOS score
sent.1	2.59	3.7	5.23	3.4	5.18	3.5
sent.2	2.83	3.9	5.18	3.5	5.11	3.8
sent.3	2.95	4.1	4.95	3.6	4.90	4.0
ave.	2.79	3.9	5.12	3.5	5.06	3.8

Ⅴ. Conclusion

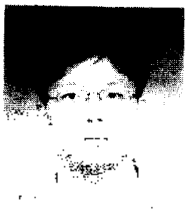
To improve the speech quality while maintaining high compression ratio in the waveform coding, the conventional MNWC is improved by introducing a Gaussian noise codebook. On the contrary to that, in the conventional MNWC, the high band of the coded speech signal is modeled as a white Gaussian noise with average level, the proposed method models it as one of colored Gaussian noises with 16-different center frequencies of average energy level. To this, the proposed method introduces the noise codebook into the encoding part, which consists of 16 colored Gaussian noises with different center frequencies.

After choosing the optimum codeword by spectral matching, the parameters of the conventional MNWC and this code index are transmitted to the decoding part. Experimental results show that, compared to the conventional MNWC, the compression ratio is deteriorated very little from 5.12 to 5.06, but average MOS score is much improved from 3.5 to 3.16. Consequently, by the proposed method, the speech quality is much improved while maintaining the high compression ratio, compared to the conventional MNWC.

References

1. 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.
2. N. S. Jayant and P. Noll, Digital Coding of Waveforms-Principles and Applicants to Speech and Video, Prentice-Hall, 1978.
3. 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.
4. L. D. Davission, "Data compression using straight line interpolation," IEEE Trans. on Info. Theory, Vol. IT-14, No. 3, pp 390-394, May 1968
5. L. Ehrman, "Analysis of some redundancy removal bandwidth compression technique," Proc. IEEE, Vol. 55, No. 3, Mar. 1967.
6. 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.
7. M. J. Bae, J. H. Lee, S. B. Im and W. C. Lee, "A New Speech Waveform Coding Based on the Nonuniform Sampling Method with Separated to High-Low Band," ASK, J., Acoust., Society, Korea, Vol. 14, No. 5, pp. 89-93, Oct. 1995.

▲정 형 교(HyungGoue Chung)　1971년 5월 12일생



1996년 2월 : 숭실대학교 정보통신공학과 졸업
1996년 8월～현재 : 숭실대학교 전기공학과 석사과정
※주관심분야 : 디지탈 음성신호처리, 통신 시스템

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