A Hybrid Coding with Adaptive Filters Using Non-uniform Technique in Speech Signals

Seong-Geon Bae [#1], Myung-Jin Bae [*2]
[#1] Information and Telecommunication of Dapartment, Soongsil University
369, Sangdo-ro Dongjak-gu, Seoul
[1] sgbae123@empal.com
[2] mjbae@ssu.ac.kr

Abstract—This paper proposes a new hybrid method in non-uniform sampling coding that uses adaptive filters as like zero-crossing and correlation. In speech signals, the modified zero-crossing rate method refers to the number of times in which the signal crosses zero in frame, the rate is high in noisy parts but low in voiced signal or unvoiced signal. And the correlation is high in voiced signals, low in noise. Also the non-uniform sampling uses the peaks and valleys, we consider that peaks and valleys is high in correlation. The method proposed in this paper uses the number of peaks and valleys in determining the threshold values of voiced, unvoiced signal and noise. Therefore, by decreasing the number of samples from sections with noise and voiceless sounds which have a high number of peaks and valleys, and by creating a signal that has little effect on the recognized signals, we may get higher compression rates.

Keyword- Waveform coding, non-uniform Sampling, Peak and Valley, Adaptive filter, Zero-crossing

I. INTRODUCTION

Researches in speech signal processing have been progressing in various fields such as coding, recognition, synthesis and analysis. In the field of speech coding, the characteristics of the sound is analyzed to separate important information out of the signals while passing through the source and the filter in order to increase the compression efficiency rate. The analysis methods can be classified according to the domain in which the signals are analyzed; the time domain, the frequency domain, and the hybrid domain which has merits of both time and frequency domain. The time domain based methods such as PCM, DPCM and ADPCM, utilize the quasi-periodic nature of signals. The frequency domain based methods such as LPC and LSP separate the components of signals by performing a domain transform. The time domain based method produces a sound which is natural and clear by its low compression efficiency. On the other hand, the frequency domain based method has relatively high compression efficiency rate but it uses intensive calculations and thus time consuming. In order to overcome these problems, the hybrid domain based method is often used because it is able to retain the naturalness and clarity of the sound similar to that of the time domain based method, while increasing the compression efficiency rate through accomplishing exact component separation of the frequency based method. The weakness of the time domain based method is its large data size. On the other hand, the frequency based method has a weakness in using a large number of calculations. This paper introduces an algorithm in which the compression rate is improved by removing the unnecessary components in the time domain and by producing a precise analysis.

In chapter 2, the pros and cons of existing time and frequency domain based speech coding methods are examined. In chapter 3, non-uniform sampling is reviewed. In chapter 4, the proposed hybrid domain based analysis method is explained by examining its characteristics and relevant processes. In chapter 5, the performance of the algorithm is evaluated through experiment and its results are discussed. In chapter 6, the conclusion of this paper is provided along with suggestions for future work.

II. NON-UNIFORM SAMPLING

The main focus of speech signal processing in terms of the information transmitted in the speech signal is the transmission of the data, the compression rate, the playback sound quality, and the processing speed. Typically it is known that the unnecessary residual components are due to the high correlation between the sample and the sampling interval. Therefore, in terms of sound quality, the unnecessary data is in the noise and voiceless sounds which have many peaks and valleys and not the voice sounds which has relatively larger amplitude. In order to improve the sound by removing the samples that do not affect its recognition, the recognition characteristics of the sound must be identified. Since only the information pertaining the peaks and valleys remain after taking the derivative and clipping the speech signal, it can be seen that the samples in between the peaks and valleys are unnecessary in terms of recognition [4]. Generally the peaks and valleys are utilized when using the non-uniform sampling method. This method detects the peaks and valleys that exists within the continuous speech period and transmits their amplitude and intervals. The transmitted data is used to restore the interval between the peaks and valleys of the sample using the cosine function. The sounds signals restored via this method has a high sound quality with a natural and clear sound. However this method limited in a sense that it requires far too many samples in the noise and voiceless sounds unlike voice sounds. Figure 1 shows typical forms of peaks and valleys.

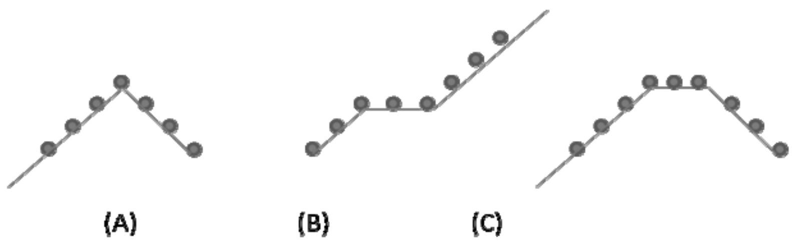

Fig. 1. General quantization of signals; Typical peak (b) Peak and valley simultaneously existing (c) Overload

In figure 1, (a) shows a typical peak, (b) shows a peak and a valley existing simultaneously, and (c) shows continuous peak when the sample is overloaded. Therefore these peak characteristics must to be considered during their restoration. Due to the fact that the non-uniform sampling uses the time domain characteristics of the sound, it has both naturalness and clarity but consequently requires a large number of samples. The peaks and valleys need to be classified into three types when searching for them using the existing method. The strong quasi-periodic characteristics of the voice sounds make them have relatively fewer peaks and valleys. The voiceless sounds have a large number of peaks and valleys and have energies similar to that of voice sounds; and these characteristics must not be neglected. The noisy section has a high zero crossing rate and thus creates a large number of valleys and peaks. Due to the fact that the existing method uses a single cosine function to code the voiceless and noisy sounds, they require a large number of samples relative to voice sounds reducing the compression efficiency.

This research implements a pre-filter that can separate the low and high frequency signals of the recognition signal. This can greatly affect the compression rate and the sound quality in using the non-uniform sampling, which is applied indiscriminately over the entire signal, by filtering it beforehand. Licklider and Pollack showed that that in a perception test that a sound signal that has been derived and clipped have a recognition level of 97 % which is similar to the recognition level of 99% of the original sounds, meaning that there is no loss in recognition due to the derivation and clipping. Therefore the recognition characteristics of the speech signal are affected by the effect on the peaks and valleys caused by the sampling and quantization. The restored signal made from the non-uniform sampling method of the voice sound section, which affects the recognition signal of the speech signal the most, can be represented as peaks and valleys. In the case of singletons the number of zero crossing can be used to find the frequency in which to design an appropriate low pass filter that can separate the high frequency component and thus the noise making a more accurate sound coding possible.

III.PAGE STYLE

In this paper, before coding using the peaks and valleys of the non-uniform sampling method, the signal is pre-filtered by frame. This minimizes the distortion of recognition information as well as increase the compression efficiency by applying the filter differently for voice sounds, voiceless sounds, and noise. Generally non-uniform sampling using peaks and valleys is represented as equation 1. The peak and valleys of the inputted frame is searched and their amplitude and interval is saved. The unnecessary samples in between the peaks and valleys are removed. This way, the compression rate is increased.

$$s_k(n) = [\frac{Amp(k-1) - Amp(k)}{2}\cos(\frac{\pi n}{Int(k)}) + \frac{Amp(k-1) + Amp(k)}{2}], \qquad 1 \le n \le I(k) \qquad (1)$$

where, *Amp(.)* is the magnitude of non-uniformly sampled data and *Int(.)* is the interval of them.

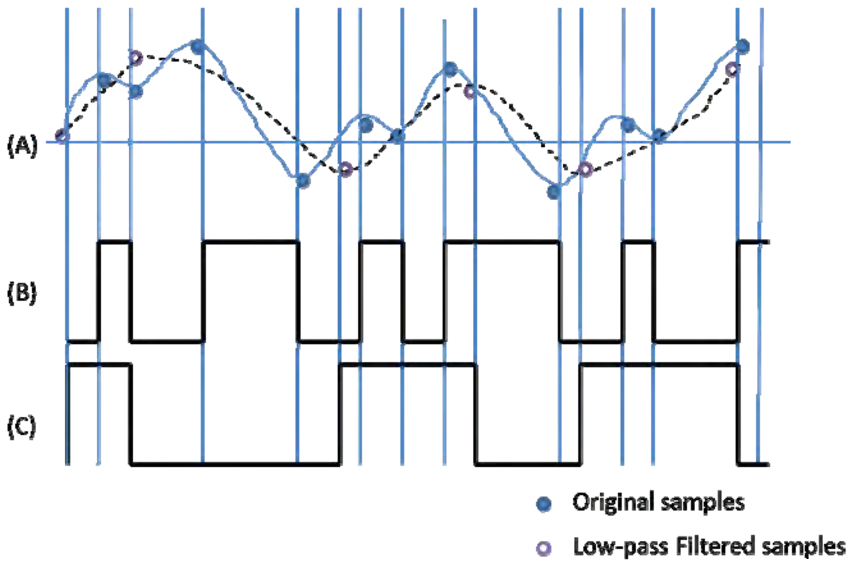

Fig. 2. Examples of non-uniform coding (A) peaks and valleys (B) original non-uniform coding (C) non-uniform coding using low pass filter signal with 2.75 kHz cutoff.

In figure 2, shows the typical examples of non-uniform coding; (a) shows how the peaks and valleys are searched and transmitted and restore at the restoring end by frame, (b) shows peaks and valleys searched after using a filter with a cut off frequency of 2.75 kHz. One can see that the number of peaks and valleys is decreased. If this is applied differently for the voice sounds, voiceless sounds and noise, using aforementioned characteristics one can increase the compression efficiency further.

First, the average crossing frequency is found using the modified zero-crossing rate within in the inputted signal frame. This average crossing frequency is high in case it is noise and low in case it is a voice sound. Using this characteristic feature, if the average zero-crossing rate is higher than 60, it should be recognized as noise and passes through a linear low pass filter using the coefficient of the average zero-crossing rate as the cutoff point. The average zero-crossing rate is used as the number of filter. If the number of the zero-crossing rate is lower than 60, the signal is recognized either as a voice or voiceless sound and thus passes through a 2.75 kHz filter to remove the fourth formant. Filtering the signals according to their characteristic properties can linearly smooth the signals by reducing the amplitude of peaks and valleys. This will, in turn, prevent the problem of reduced compression rate in noise and unvoiced signals due to the large number of peaks and valley as found in the existing methods. The filter used in this study averages the samples as shown in equation 2.

$$s_L(n) = \frac{1}{N}\sum_{k=0}^{N-1} s(n+k) \tag{2}$$

Here, k represents the cutoff frequency

In order to reduce the number peaks and valleys the voice sound was filtered at 2.75 kHz and the voiceless sound and noise was filtered using the crossing rate in the signal frame.

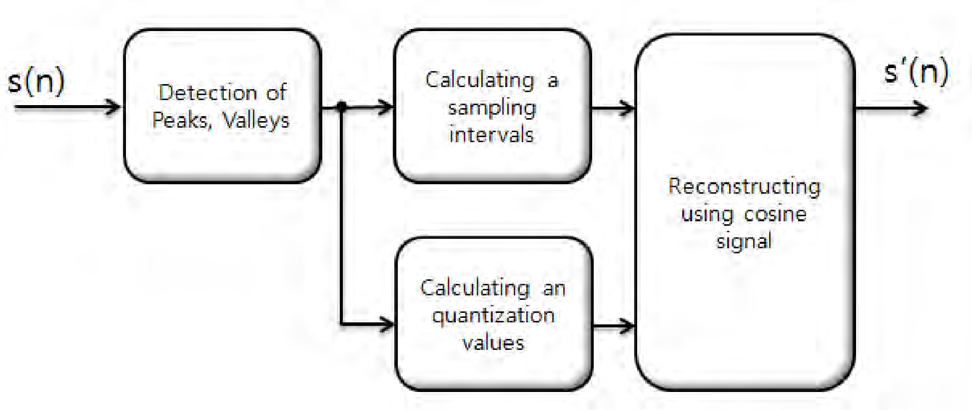

Fig. 3. Speech coding scheme using post filter

Figure 3 shows the speech coding scheme proposed in this paper. The analog signal is converted into a digital signal and divided into frames. And the zero-crossing rate is detected here. The detected coefficient is used as a filter coefficient in case the signal frame is noise or a voiceless sound. In case that it is a voice sound, it is filtered through a 2.75 kHz filter since the important information is within that range. In order to compensate the filtered high frequency component, the difference between the original signal and the filtered signal is found. This filtered signal is composed of high frequency components.

$$d\ (n) = s(n) - s_L(n) \tag{3}$$

Here, s(n) is the original signal, and $s_L(n)$ is the filtered signal.

$$h_H = \frac{1}{N}\sum_{k=0}^{N-1} d(n+k) \tag{4}$$

Here, k is the number of frames. The amplitude of the difference signal is found using this equation. During decoding, the filtered signal is added using the Gaussian random signal and the h_H value. Here, the filtered signal saves the amplitudes and intervals of the detected peaks and valleys. The peaks and valleys is later used in the restoring end to restore the signal using the non-uniform sampling method. The most important part of this research is the detection of peaks and valleys using the separation of low frequency signal and the separation of the high frequency signal using the difference signal. In order to compare the proposed and existing method they were implemented using C and MATLAB respectively. The standard of the sound quality was chosen to be 8-bit per sample using a u-law PCM. The computer used for the simulation use for the performance evaluation was an IBM-PC, interfaced with an AD/DA converter at a sampling rate of 8 kHz and coded at 16 bits per sample. In order to evaluate the performance, five representative sentence were spoken 3 times by men and women in their 20's, 30's and 40's were used as speech samples. The sample with a balanced melody was used for the performance evaluation.

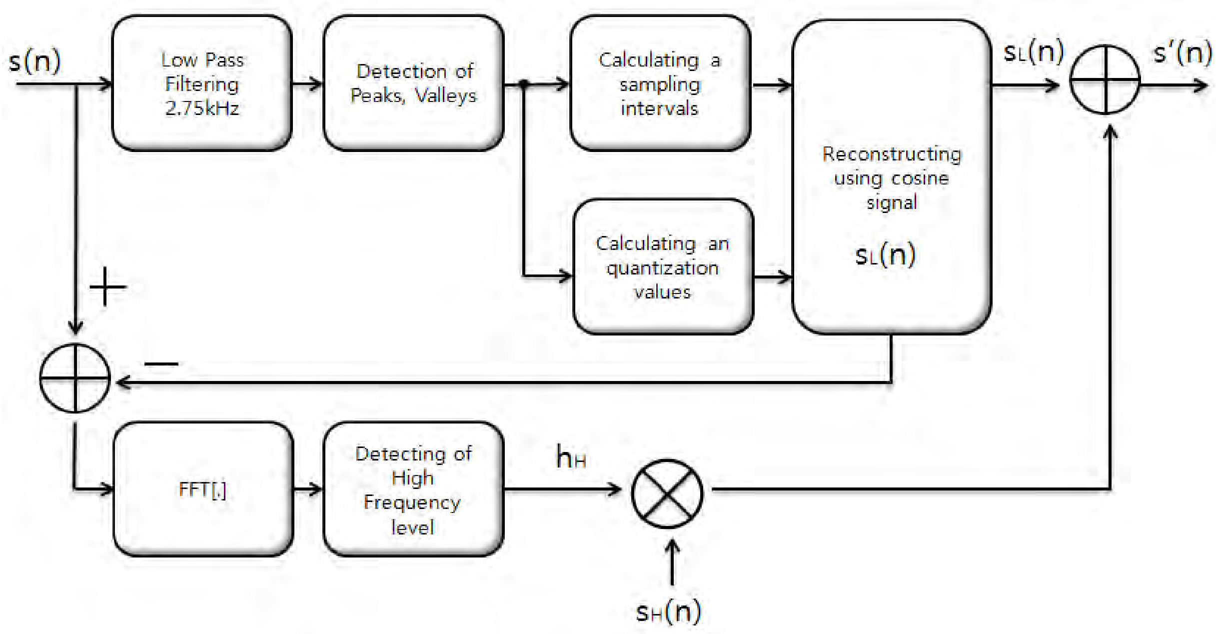

Fig. 4. A New speech coding scheme proposed for voiced signal

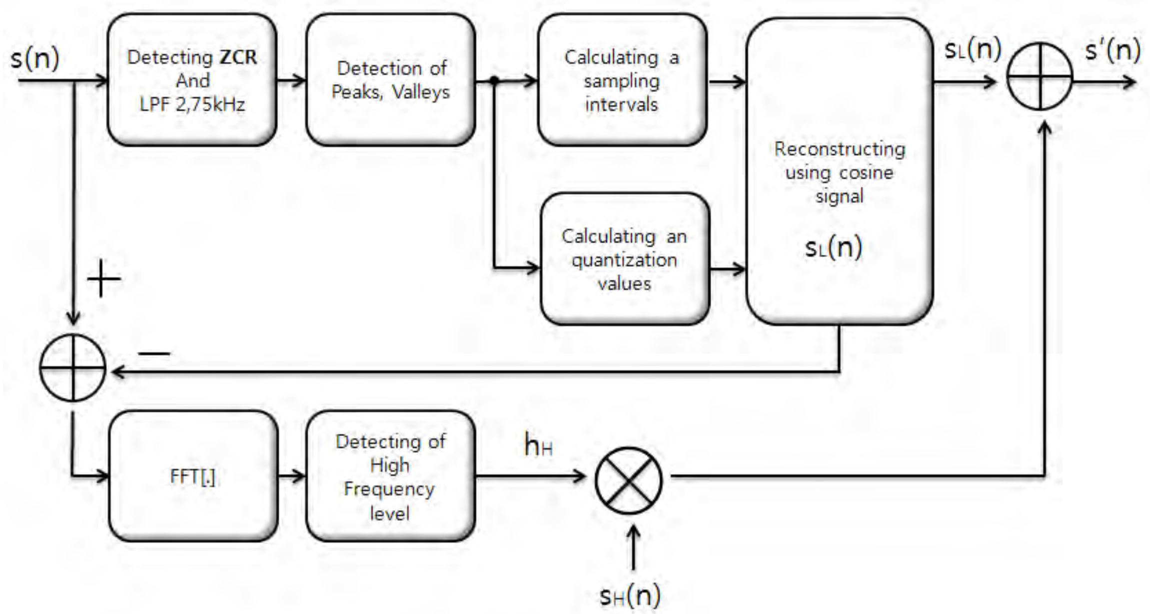

Fig. 5. A New speech coding scheme proposed for unvoiced signal

IV. CONCLUSION

It is difficult to separate the naturalness and intelligibility according to the characteristics of the inputted speech signals. Although there have been many methods proposed in order to separate the frequency component of a signal, these processes affect the recognition of signal. Therefore, in order to retain the sound's naturalness and clarity, the source coding method has been used. The short comings of this method are its long calculation process which is required to separate the source. This method has been usually used on the systems that are capable of conducting such a complex computation, and thus suggestions to improve the method have been put forward. Among the methods suggested so far, the hybrid coding which takes the advantages both of the domain based method and the frequency based method, has been considered performing well and thus studied in many areas. And the source coding was used for its high component separation performance as well as its high compression efficiency. Furthermore, a hybrid coding method using non-uniform sampling which has both advantages of the aforementioned methods was proposed. In the proposed method, the compression efficiency rate of the non-uniform sampling coding method and thus its naturalness and clarity of the sound were improved by using a filter to reduce the required number of peaks and valleys. This research shows that if we can come up with a new pre-filter more suitable for a higher sound quality, such as audio, the proposed method would become a better performing high sound quality hybrid coding method.

REFERENCES

[1] S.G. Bae, M.J. Bae, “A New Speech Coding using Harmonics Emphasis Filter,” ISAAC 2013, AACL Vol. 1, pp. 43-44, November 2013.

[2] S.G. Bae, M.S. Kim, and M.J. Bae, “A New Hybrid Non-Uniform Coding with Low Bit Rates for Sound Signal in Near Field Communication,” 2013, IEEE ICTC, pp. 1039-1040, ICTC 2013.

[3] S.G. Bae, M.S. Kim, and M.J. Bae, “On Enhancement Signal Using Non-uniform Sampling in Clipped Signals for LTE Smart Phones,” 2013, IEEE ICCE-berlin, pp.125-126, ICCE-berlin 2013.

[4] S.G. Bae, M.J. Bae, “On A New Hybrid Speech Encoder using Variable LPF,” ASTL Vol. 21, pp.275-278, April 2013.

[5] S.G. Bae, H.W. Park and M.J. Bae, “On a New Enhancement of speech Signal using Non-uniform Sampling and Post Filter,” ICHIT2012, LNCS 7425, pp.723-729, August 2012.

[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] 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. September (1968)

[8] L.D. Davission, “Data Compression Using Straight Line Interpolation,” IEEE Trans. On Infor. Theory, Vol. IT-14, No.3, pp. 390-394, May (1968)

[9] L. Ehrman, “Analysis of Some Redundancy Removal Bandwidth Compression Techniques,” Proc. IEEE, Vol. 55, No. 3, March (1967)

AUTHOR PROFILE

Seong-Geon Bae He received the M.S. degree in Electronics Engineering from Konkuk University. He is currently the Ph.D. degree at Soongsil University in 2014. He is currently the Professor of the Dept. of Broadcasting Sound & Visual at Daelim University.

Myung-Jin Bae He received the Ph.D. degree in Electronic Engineering from Seoul National University in 1987. He is currently the Professor of the Dept. of Information & Telecommunication at Soongsil University. He has authored and coauthored more than 300 journal articles and conference papers. His research interests are Speech Signal Processing, Audio Processing and Speech Communications