International Journal of Software Engineering and Its Applications
Vol.7, No.6 (2013), pp.391-398
http://dx.doi.org/10.14257/ijseia.2013.7.6.33

A New Non-uniform Sampling & Quantization by using a Modified Correlation

Seong-geon Bae[1] and Myungjin Bae[2]*

[1]Department Electronic Communications, Daelim University College
29, Imgok-rd, Dongan-gu, Anyang-Si, Gyeonggi-Do, 431-715, Korea
[2]Information and Telecommunication Department,
Soongsil University, 1-1 Sangdo 5 dong DongJak-Ku,
Seoul, 156-743, Republic of Korea
[1]sgbae123@empal.com, [2]mjbae@ssu.ac.kr

Abstract

We proposed a new non-uniform sampling and quantization using variables low-pass filter with correlation. It focuses on the naturalness and intelligibility of speech synthesis applications and the compression and signal-to-noise ratio of speech transmission applications. However, it is well known that when conventional sampling methods are applied directly to speech signal, the required amount of data is comparable to or more than that of uniform sampling method. To overcome this problem, a new non-uniform methods is proposed, in which time domain coding is applied to two low-pass filters in lower bandwidth and the remain signals are compensated by the Gaussian white signal, which is used to get high quality speech by correlation of signal .

Keywords: *Sampling, Quantization, Peak and valley, Low-passed filter, cross-correlation*

1. Introduction

In speech production, as well as in many human-engineered electronic communication systems, the information to be transmitted is encoded in the form of a continuously varying (analog) waveform that can be transmitted, recorded, manipulated, and ultimately decoded by a human listener. In the case of speech, the fundamental analog form of the message is an acoustic waveform, which we call the speech signal. In many speech communication settings, the presence of background interference causes the quality or intelligibility of speech to degrade. When a speaker and listener communicate in a quiet environment, information exchange is easy and accurate. However, a noisy environment reduces the listener's ability to understand what is said. In addition to interpersonal communication, speech can also be transmitted across telephone channel, loudspeakers, or headphones. The quality of speech, therefore, can also be influenced in data conversion, transmission, re-production. The purpose of many enhancement algorithms is to reduce background noise, improve speech quality, or suppress channel or speaker interference. In this paper, we discuss the general problem of speech enhancement with particular focus on algorithms designed to remove additive background noise for improving speech quality. In our paper, background noise will refer to any additive broadband noise component as white Gaussian noise, aircraft cockpit noise, or machine noise in a factory environment. Other speech processing areas that are sometimes included in a speech enhancement include suppression of distortion from voice coding

* Corresponding author : Myung Jin Bae(mjbae@ssu.ac.kr)

ISSN: 1738-9984 IJSEIA

algorithms, suppression of a competing speaker in a multi-speaker setting, enhancing speech as a result of a deficient speech production system, or enhancing speech for hearing-impaired listeners. The problem of enhancing speech degraded by additive background noise has received considerable attention in the past two decades. Many approaches have been taken, each attempting to capitalize on specific characteristics or constraints, all with varying degrees or success. The success of an enhancement algorithm depends on the goals and assumptions used in deriving the approach. Depending on the specific application, a system may be directed at one or more objectives, such as improving overall quality, increasing intelligibility, or reducing listener fatigue. The objective of achieving higher quality and/or intelligibility of noisy speech may also contribute to improved performance in other speech application, such as speech compression, speech recognition, or speaker verification.

Main point of speech coding is to process it by considering, especially, transmission and compression rate of data, signal quality of playback, and processing velocity among the information transmitted by the encoded data. In addition, there is the source coding method which is the parameter method to have improvement by detecting characteristic section of signal, and this has the disadvantage that the complexity and the calculation is too much to be applied to various signals. It is the hybrid coding method which is made by taking advantages of coding methods, and the improvements have been made for this owing to the development of computing. In this paper, in order to use this hybrid coding method, the method of coding with non-uniform quantization and sampling of recognized signal obtained from the time domain and with analyzing frequencies in frequency domain will be used.

At Chapter 2, the conventional non-uniform sampling method which is to get recognizable characteristics of signal, and at Chapter 3, the proposed method to use the new non-uniform sampling and quantization will be described, and at Chapter 4, experimental and results, and at Chapter 5, conclusion and study direction in the future will be presented.

2. General Non-uniform Sampling

A major objective in speech coding is to compress the signal, that is, to employ as few bits as possible in the digital representation of the speech signal. The efficient digital representation of the speech signal makes it possible to achieve bandwidth efficiency in the transmission of the signal over a variety of communication channels, or to store it efficiently on a variety of magnetic and optical media. Since the digitized speech is ultimately converted back to analog form for the user, an important consideration in speech coding is the level of signal distortion introduced by the digital conversion process. The coding method to use peak and valley of signal by analyzing waveform characteristic has been presented, and this has been used by considering the recognizable aspect. Therefore, this recognizable aspect of speech signal will be affected by the peak and valley in the sampling and quantization. In this characteristic of important information for recognition of speech signal, the roles of peak and valley point become remarkably important. By utilizing this characteristic, numerous applications for synthesis or coding section are possible, and especially for noisy environment, great support have been conducted for searching important factors of recognition signal by considering characteristics of peak and valley. These two sampling methods for linear sampling and non-uniform one. The interpolation method of non-uniform sampling in speech reconstruction is used as cosine interpolation. The waveform reconstruction is performed by using a cosine interpolation method based on such parameters as the magnitudes and the intervals of the peak and the valley.

3. A New Proposed Method

According to the speech production mechanism, the positive and negative signals have very high correlations. Moreover, from the viewpoint of speech recognition, higher frequency band components have low correlation, while the positive and the negative signals are separated to reconstruct the high-intelligible speech. Therefore, the samples related to the frequency band higher than the voiced signals are considered as redundant information in the speech perception.

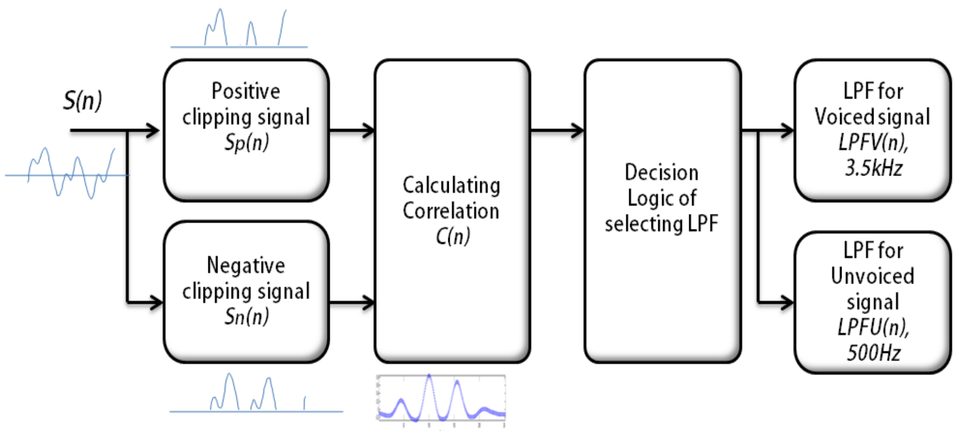
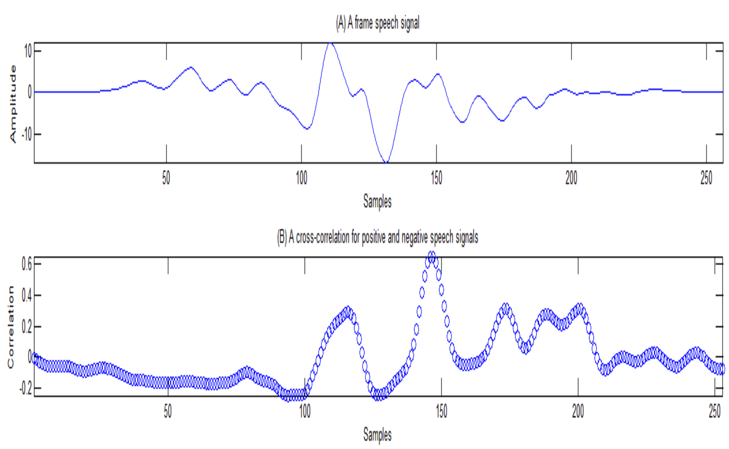
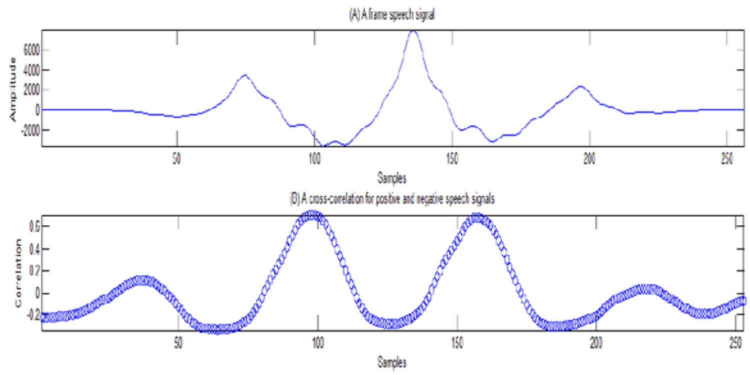


Figure 2. Examples of New Non-uniform Sampling and Quantization using LPFV and LPFU

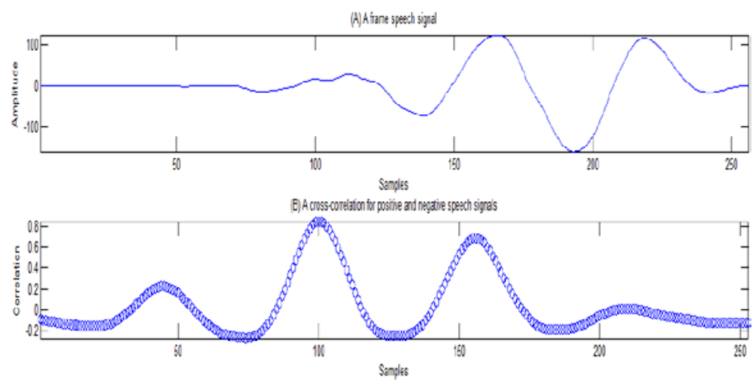
The correlation of positive and the negative samples in voiced signal are more than unvoiced signal. Also, the higher correlations have wide broad bandwidths. Therefore, non-uniform sampling can be only applied to the correlation of signal components in the original waveform to reduce loss of intelligibility. Since the lower correlation signal is smoother than the higher one, fewer number of the peak and the valley sample are obtained when non-uniform sampling is performed on it. Also if the signal is voiced, the low-pass filtered signal is used by 3.2kHz, the unvoiced by 500Hz. So fewer number of the peak and the valley sample in voiced signal. But when unvoiced, the results is smaller it. So we consider it as a decision logic to detect the voiced or unvoiced signal. Then this makes it possible to achieve high quality and high compression ratio.



(A) Correlations for Silence Signal



(B) Correlations for Unvoiced Signal



(C) Correlations for Voiced Signal

Figure 4. Correlations between the Positive and the Negative Signal

Figure 3 shows the encode block diagram of the method proposed in this paper. In the encode block diagram, S(n) is speech signal digitized uniformly by A/D converter, and $S_{L35}(n)$ is the law-pass filtered signal by 3.5 kHz as cutoff frequency, and $S_{L5}(n)$, the law-pass filtered signal by 500Hz. Then the conventional non-uniform sampling is applied to this two low-pass filtered signals and such parameters as the magnitudes, M(.) and the intervals, I(.) of the peak and the valley are quantized and sampled.

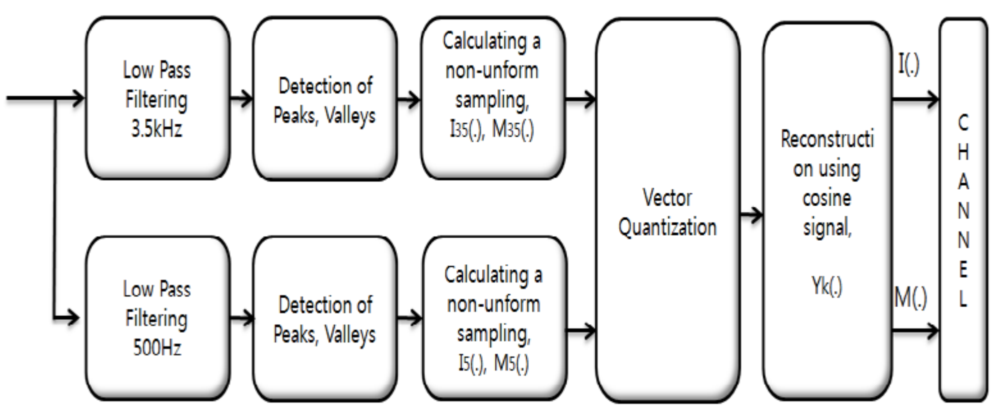


Figure 3 A New Speech Coding Scheme

At the same time, Re-constructed speech signal, $S_L'(n)$, is reconstructed by the cosine interpolation as the non-uniform sampling.

$$y_k(n) = [\frac{M(i-1)-M(i)}{2}\cos(\frac{\pi n}{I(i)}) + \frac{M(i-1)+M(i)}{2}], \qquad 1 \le n \le I(i) \tag{3-1}$$

Major components of the residual signal consist of the white signals. Generally, the magnitude spectrums of the higher components have wide bandwidth than that of lower signals and these are assigned to a little important information to signal intelligibility. To preserve the naturalness of signal for higher components, the Gaussian signal is added to the reconstructed roughly with Gaussian level from encoder. Since the characteristic of the residual signal between the original and the reconstructed signal is rather a pseudo colored than a white noise, we can roughly approximate the residual signal with Gaussian signal in decoder.

$$S_{LR}(k) = [\frac{M(i-1)-M(i)}{2}\cos(\frac{\pi k}{I(i)}) + \frac{M(i-1)+M(i)}{2}], \qquad 1 \le k \le I(i) \tag{3-2}$$

where, *M(.)* is the magnitude of non-uniformly sampled data, *I(.)* is the interval of them and *i* are samples in the interval.

Then, the level, the Gaussian random signal is reconstructed by $S_{LR}(n)$. And finally we get *S'(n)*, reconstructed signals. This procedure can much reduce the data rate to achieve higher compression ratio than that of the conventional non-uniform sampling even in the noisy environment and also get a good quality of speech by reconstructed harmonics signal.

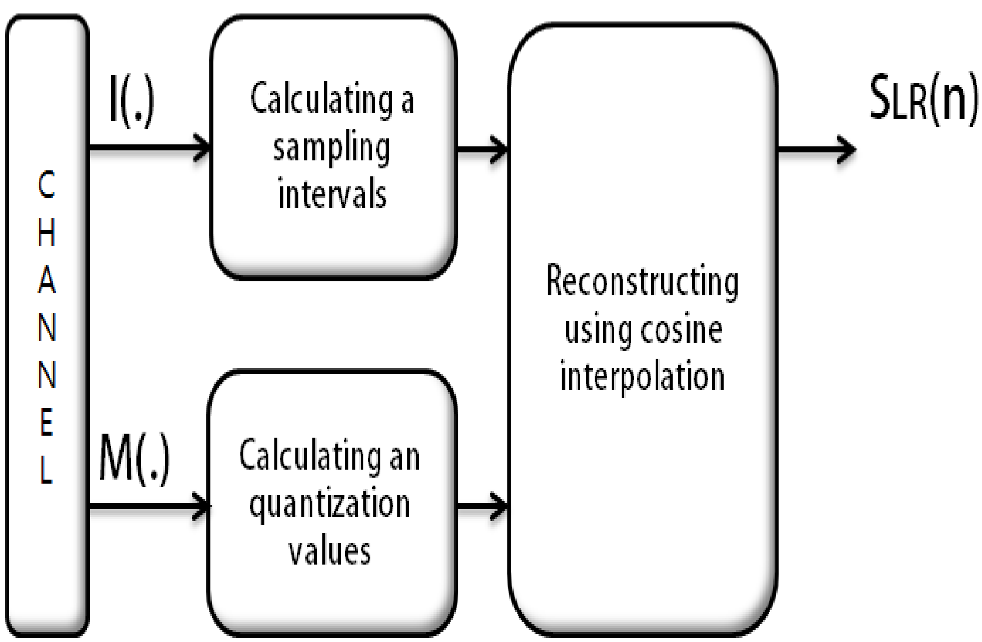


Figure 4. A New Speech Decoding Scheme

4. Experimental and Results

To compare the performances between the existing and the proposed method, three phoneme-balanced Korean and English sentences were used. Each sentence was pronounced by 32 female and 38 male speakers three times. For simulation test, speech signal was sampled at 8 kHz to 44.1kHz, filtered by anti-aliasing 10th order LPF with 3.6 kHz to 21kHz cutoff frequency and digitized with 16bit A/D converter. In Table 1, 2, That is experiment data those are the SNR of the signal by varying. We make the data of that the good SNR. And we enhance the signals by Gaussian random signal to high frequency in band. We use white Gaussian signal for compensation higher frequencies. This results show the higher compression rate and good quality in noisy or silence. Also we get that the compression was proportional to the sampling rate.

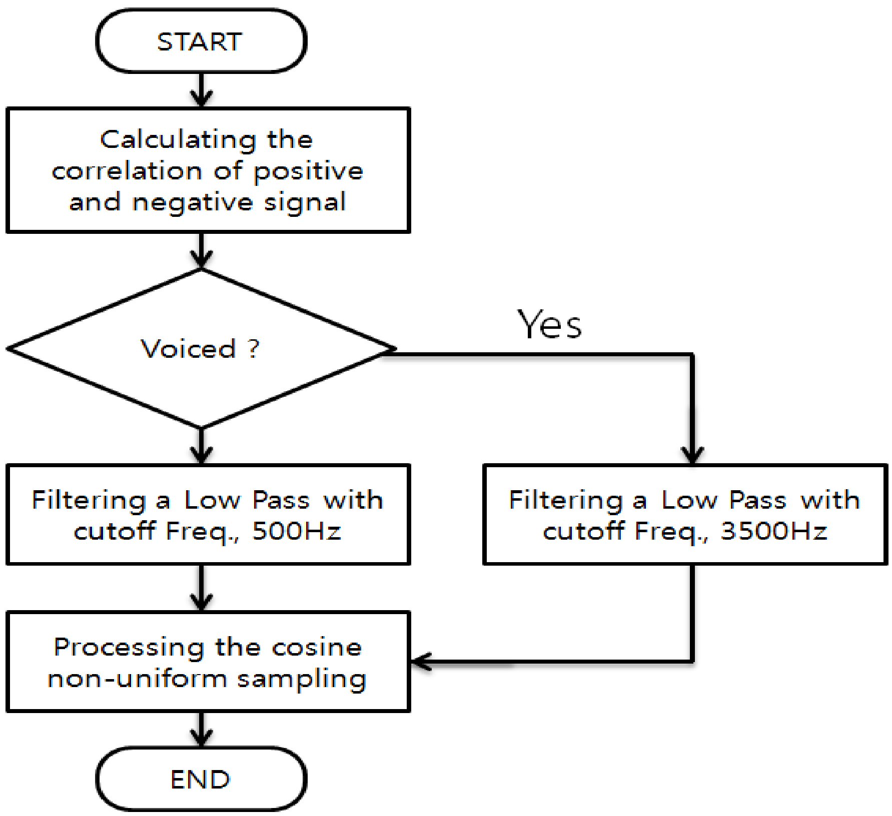


Figure 5. A Flow Chart of Proposed Algorithm

As show in Figure 5 is a flow chart of the proposed method by using the correlation method and the non-uniform sampling, respectively. We get to higher compression rate and good quality by sampling rates. We processed lower correlation in the unvoiced signal, higher correlation in the voiced signal. To get higher compression rate and good quality, we compare positive sample signals to negative ones by cross correlation values. Generally the correlation values have higher values in voiced and lower ones in unvoiced. And we use the correlation method for detection voice and unvoiced which is a positive and negative level clipping'.

Table 1. Results of Segmental SNR and Compression Rate

	Existing Method		Proposed Method	
	SNR(dB)	Compression Rate	SNR(dB)	Compression Rate
Sample 1	16.1	2.7	15.4	12.4
Sample 2	16.5	2.6	15.2	12.7
Sample 3	15.4	2.2	14.3	11.1
Sample 4	15.8	2.9	14.6	12.8
Sample 5	16.2	2.5	15.5	12.3
Avg.	16.0	2.5	15.0	12.26

Table 2. Results of Variables Data to Sampling

	Existing Method	Proposed Method
8 kHz	2.5	8.5
10 kHz	3.1	12.4
22.01 kHz	5.7	18.2
44.1kHz	12.4	34.2

5. Conclusion

We proposed a new non-uniform sampling and quantization using variables low-pass filter with correlation. It focuses on the naturalness and intelligibility of speech synthesis applications and the compression and signal-to-noise ratio of speech transmission applications. In this paper, to get high quality and compression rates, we proposed new non-uniform sampling and quantization method using variable LPF. The proposed technique is applied to two low-pass filters to speech signal to reduce the data by correlation information.

In conclusion, Experimental results with phoneme balanced Korean and English sentences show that the proposed method can achieve higher compression ratio with good of segmental SNR compared with the existing method.

References

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

[2] S.G. Bae, H.W. Park and M.J. Bae, “On a New Enhancement of speech Signal using Nonuniform Sampling and Post Filter,” ICHIT2012, LNCS 7425, pp723-729, August (2012)

[3] S.G. Bae, H.W. Park and M.J. Bae, “ A Study on Enhancement of Speech using Non-uniform Sampling,” IJHIT, Vol.5, No.2, pp237-242, April (2012)

[4] S. Gazor and W. Zhang, “A soft voice activity detector based on a Laplacian-Gaussian model," IEEE Trans. Speech Audio Processing, Vol. 11, No. 5, pp. 498-505, Sep. 2003

[5] A. J. Accardi and R. V. Cox, “A modular approach to speech enhancement with an application to speech coding,” J.Acout. Soc. Am, Vol. 10, No. 3, p.1245, Sep. 2001.

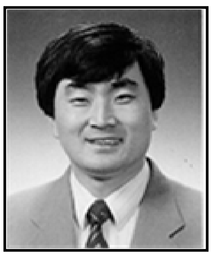
[6] T.J. Lynch, “The Probability of a Straight-line Sequence from a Uniform Independent Sample Source,” IEEE Trans. On Info. Theory, Vol. IT-14, No. 5, pp 773-774. September (1968)

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

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

Authors

Seong-Geon Bae, he received the M.S. degree in Electronics Engineering from Konkuk University in 1995. He is currently the under Ph.D. degree at Soongsil University. He is currently the Professor of the Dept. of Electronic Communications at Daelim University Collage.
E-mail : sgbae123@empal.com



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.
E-mail: mjbae@ssu.ac.kr

