A Study on Enhancement of Speech using Non-uniform Sampling

Seong-Geon Bae, Hyung-Woo Park and Myung-Jin Bae[1]

Information and Telecommunication Department,Soongsil University,
1-1 Sangdo 5 dong DongJak-Ku,Seoul,
156-743, Republic of Korea
sgbae123@empal.com, park.hyungwoo@gmail.com, mjbae@ssu.ac.kr

Abstract

There are various methods to improve signal under the clipped environment of voiced sound in the voice signal processing. In the case of the majority of clipping signal restoration, it will be concentrated in sound quality deterioration of low frequency as they are converged into voiced sound from the characteristic viewpoint of sound quality deterioration. Therefore, in the case of clipping of voiced sound section, great efforts have been contributed for sound quality improvement and restoration. From this study, the method to improve clipped voice signal will be proposed. Voiced sound section is created by utilizing the fact that differentiation and clipped signal in voice signal are supposed to give small influence for the evaluation of recognition level. In here, for the signal which is obtained by stressing amplitude at voiced sound section where most of them are distributed, the method of emphasizing synthesis with original signal was used.

Keywords: *Clipping, Non-uniform sampling, Peak and valley*

1. Introduction

Speech encoding method for storing or transmitting in signal can largely be classified into the waveform encoding method, the source coding method, the hybrid encoding method, etc., and in order to maintain clarity and natural characteristic, the wave form encoding method is mainly used. This coding process is the method of storing and synthesizing after eliminating repeated, unnecessary remaining ingredient, and it is consisted of PCM, ADPCM, ADM, etc. However, it has the disadvantage that large amount of memory is required due to enormous quantity of data. Main point of speech coding is to process it by considering, especially, transmission and compression rate of data, sound quality of playback, and processing velocity among the information transmitted by the speech data. In general, unnecessary remaining ingredient which is existed in the speech signal is known to be derived from relatively high interrelation existed between samples. Therefore, in order to reduce data quantity or transmission rate for storing or transmission of data, the remaining ingredient which is existed in the uniform standardized sample, i.e., the sample which has high interrelation between samples and gives lesser influence from the viewpoint of recognition shall be removed. In general, for voiced signal, this problem can be generated for the recognition when voiced signal is clipped in accordance with the characteristic of input signal for speech signal. Therefore, in order to maintain clarity and natural characteristic of sound, work to improve clipped sound shall be carried out. Due to the improvement of calculation time of computing, various calculation methods have been presented. One of these methods is the waveform

[1] Corresponding author : Myung Jin Bae(mjbae@ssu.ac.kr)

synthesizing method in which improvement work is performed at time zone, and it is known to be excellent in terms of clarity and natural characteristic. In addition, there is the frequency synthesizing method which is the parameter method to have improvement by detecting characteristic section of sound, and this has the disadvantage that calculation quantity is too much to be applied to various sounds. It is the mixed synthesizing method which is made by taking advantages of synthesizing method, and numerous improvements have been made for this owing to the development of computing.

In this study, in order to use this mixed synthesizing method, the method of synthesizing with non-uniform standardization and sampling of recognized signal obtained from the time zone will be used. Especially, in the case of clipping signal, as it is concentrated in the voiced sound with large amplitude, the improvement method which is searching and using peak and valley which are characteristics of voiced sound for synthesizing in order to emphasize linearity is to be presented. At chapter II, existing method which is to seek recognizable characteristic of voice signal, and at chapter III, signal method to use the proposed non-uniform standardized method will be described, and at chapter IV, conclusion and study direction in the future will be presented.

2. Existing Method

According to the study which considers recognizable characteristic of voice signal, it can be understood that in the case of the signal made by differentiating and clipping original voice wave from the recognition test of Licklider and Pollack, almost no deterioration is generated when comparing to 99% of recognition level of original signal as its value of recognition level is 97%. The coding method to use peak and valley of sound by utilizing this characteristic has been presented, and this has been used by considering the recognizable aspect. Therefore, this recognizable aspect of voice signal will be affected by the peak and valley which are appeared in the standardization and quantization. When this characteristic re-constitutes important information for recognition of voice signal, the roles of maximum and minimum point become remarkably important. By utilizing this characteristic, numerous applications for synthesizing or coding section are possible, and especially for jamming environment, great support have been conducted for searching important factors of recognition signal by re-constituting filter and considering characteristics of peak and valley. As shown in the figures, the first one shows typical characteristics of peak and valley, and the second one displays properties of peak and valley. Final figure indicates influences which appear from clipping of sound and overload.

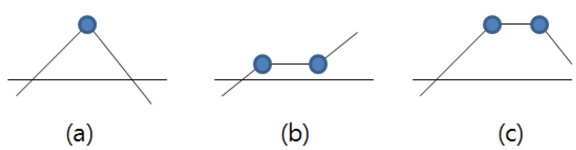


Figure 1. Quantization Influence Signal (a) Typical Peak (b) Peak, Valley Simultaneously (c) Overload

In these 3 cases, as the last one is appeared by the influence for overload, it can give the most influence to the recognizable characteristic of sound. In here, overload is the phenomenon which is appeared in the clipping environment of sound, and great influence will be affected to the recognizable section when deteriorated signal is re-constituted at this section. Therefore, in this study, the improvement method applying to clipping signal will be used by considering only the section which is generated by the influence of overload and by using this kind of important recognition signal.

3. Proposed Method

In order to make proposed method with information of peak and valley which gives influence to recognizable signal, the information is stored by dividing into non-uniform quantization information and non-uniform standardization information, in the first place. Voiced sound which is the characteristic of clipping signal is separated by dividing into information of amplitude section and time section, and then, overloaded signal is identified. For the signal detected by this, unnecessary signal will be removed by filtering with 2.75Hz Low pass filter. Only the recognition signal which gives great influence to the conspicuous peaks and valleys will be remained in the filtered signal. In this way, parameters which give influence to the recognition characteristic will be sorted out, and from the clipping non-uniform standardization information for voiced sound, information of peak and valley will be obtained. With this information, characteristic of peaks and valleys and information of time space will be searched and stored. In here, by creating signal of cosine characteristic at the peaks, clipped signal will be improved through the application to the signal of overload.

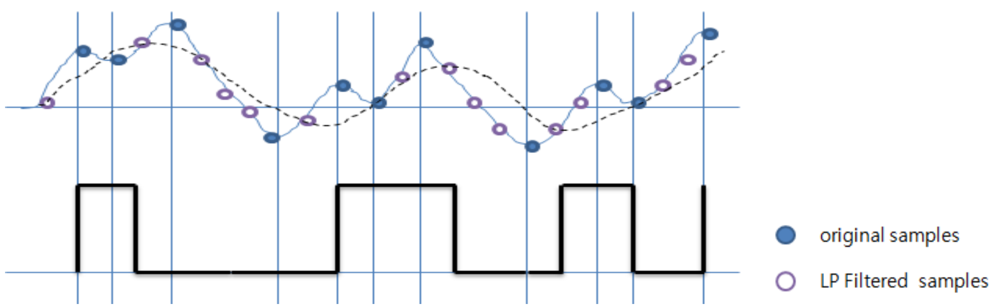


Figure 2. Proposed Method of Re-construction using Peak and Valley

As in the case of Figure 2, clipped signal will be filtered with low frequency with LPF 2.75kHz possessing recognizable characteristic information.

At this stage, non-uniform sampling and quantization information are extracted by detecting peak and valley. In here, horizontal number of clipped voiced sound which is detected from input signal is obtained and stored, and the signal is to be re-constituted by using previously obtained non-uniform sampling and quantization information to the characteristic whose horizontality is 0. The test and result have been inputted with 16 bit data through the 8KHz standardization after constructing interface with AD/DA converter for conducting input/output of voice signal. Performance evaluation of processing result has been used for the test by utilizing sentences of the man and woman of their twenties, and thirties as well. In each of these cases, data have been obtained with overload so that clipped signal can be made artificially and these have been used for the test.

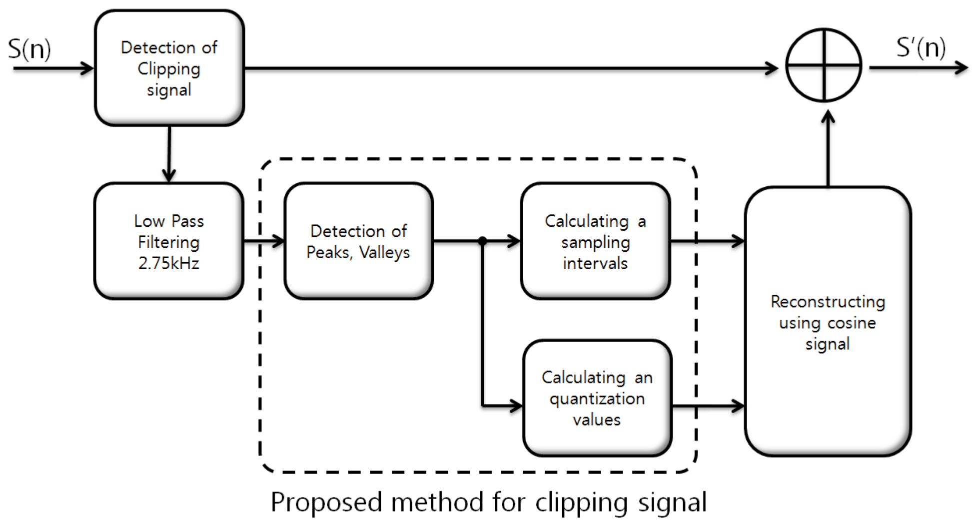


Figure 3. Proposed Method for Clipping Signal in Voiced Speech

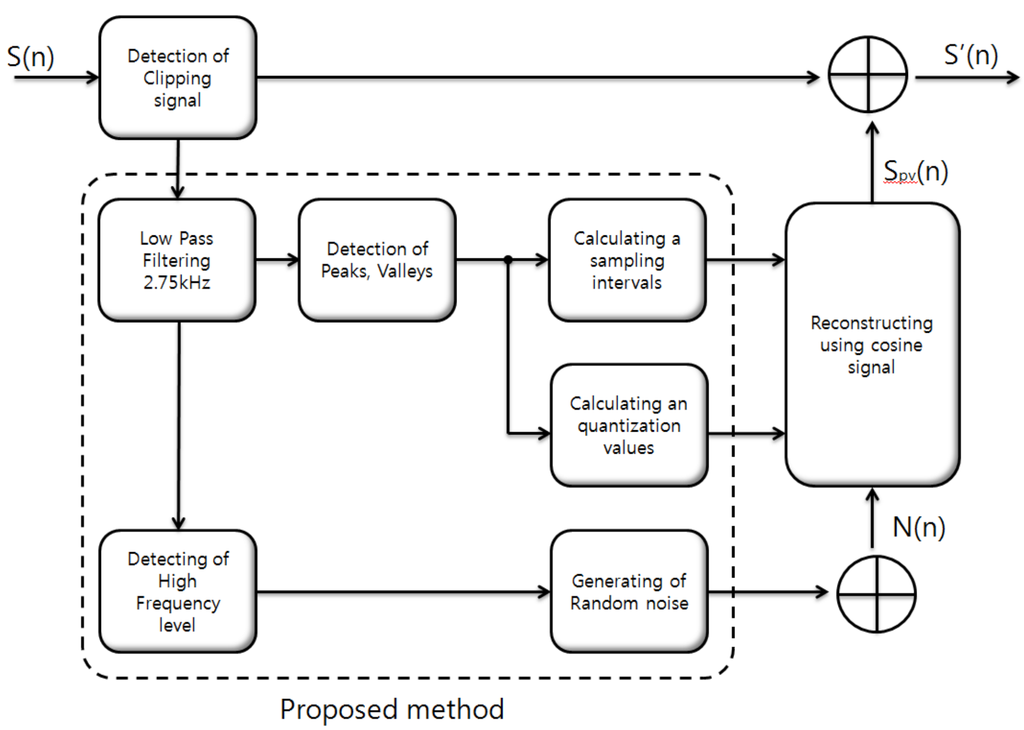
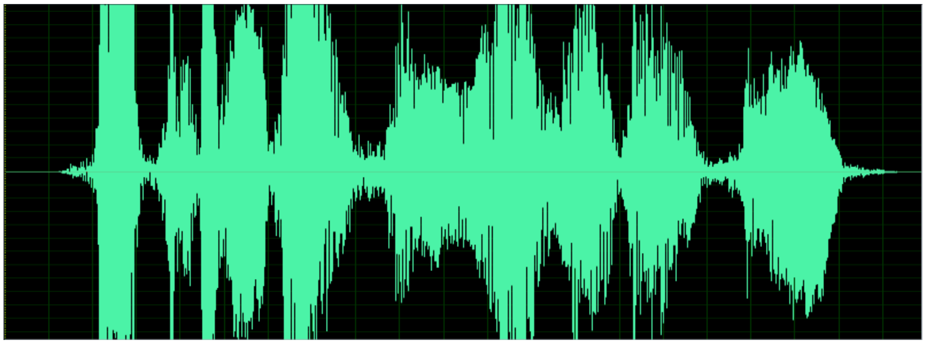
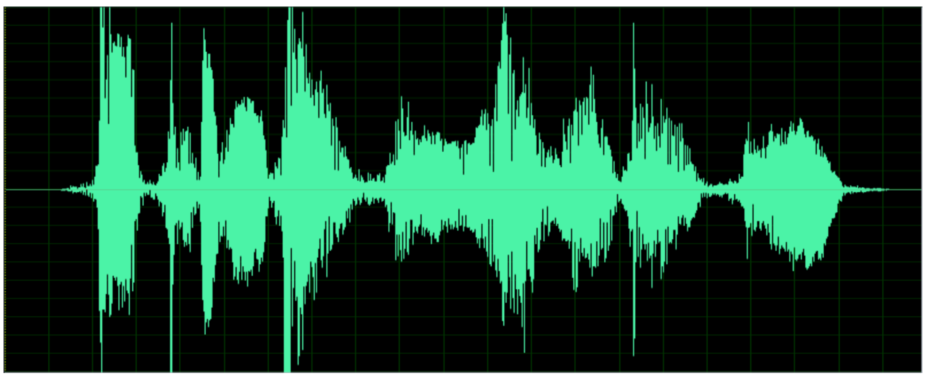


Figure 4. Proposed Method

$$S'(n) = S_{pv}(n) + N(n) \tag{3-1}$$



(a) Waveform of original.



(b) Waveform of reconstructed signal.

Figure 4. Results of Proposed Method of Test
(a) Clipping Signal (b) Re-constructed Signal

4. Conclusion

In this paper, when continuous voice signal is deteriorated by clipping, the method of restoring clipped signal at received voice signal section has been presented. In the case of voice signal, as most of the clipped signal are generated from voiced sound, the improvement method in which clipped voiced sound is used for restoration and non-uniform standardization and quantization method are used and applied by detecting peak and valley which give influence to recognizing information has been used. The method to give vividness of clarity and natural characteristic of sound from deteriorated clipping voice signal of voice recording device's input signal has been presented. If using the filter which can emphasize more of the characteristics of recognizing information and performing the process by removing voiceless sound in the first place, the performance will definitely be improved.

References

[1] 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, **(2001)** September.

[2] C. –J. Jeong and M. –J. Bae, "Analysis regarding vocalizations of teuroteu singers", The institute of electronics engineers of Korea, Summer scholarship contest of 2009, pp. 1090-1091, **(2009)**.

[3] J. Sohn, N. S. Kim and W. Sung, "A statistical model-based voice activity detector", IEEE Signal Processing Lett., Vol. 6, No. 1, pp. 1-3, **(1999)** January.

[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, **(2003)** September.

[5] T. Agarwal and P. Kabal, "Pre-processing of noisy speech for voice coders", in Proc. IEEE Workshop on Speech Coding, Tsukaba, Japan, pp.169-171, **(2002)** October.

[6] Y. –S. Lee and Y. B. Park, "An acoustic study of feeling information extracting method" IWIC, Vol. 10, No. 1, pp. 51-55, **(2010)** February.

[7] J. –Y. Ahn, Y. –S. Kim, S. –H. Kim and K. –I. Hur, "A Study on Voice Recognition Pattern matching level for Vehicle ECU control", IWIC, Vol. 10, No. 1, pp. 75-80, **(2010)** February.

[8] M. J. Bae and S. H. Lee, "Digital Voice Signal Analysis", Books Publishing Dong Young, **(1998)**, ch. 3, 6.

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.

E-mail: sgbae123@empal.com

Hyung-Woo Park

He received the M.S. degree from Soongsil University in 2007. He is currently the under Ph.D. degree at Soongsil University.

E-mail: park.hyungwoo@gmail.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