On A New Hybrid Speech Encoder using Variables LPF

Seong-geon Bae, Myung-jin Bae[1],

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

Abstract. To encode the speech quality with reduce the redundancy within samples that resulted from domain processing method like PCM and LPC, Source coding or Waveform coding methods can be considered. 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 hybrid methods is proposed

Keywords: Hybrid coding, Waveform coding, Source coding, Sampling.

1 Introduction

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 conversion of the speech encoding method for storing or transmitting in signal can largely be classified into the waveform coding, the source coding and the hybrid coding. And to maintain intelligibility and naturalness, the waveform coding method is mainly used. This coding process is the method of storing and synthesizing after eliminating repeated, unnecessary remaining components, 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.

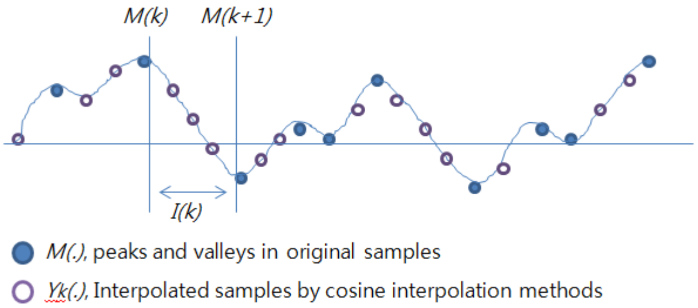
[1] Corresponding author : Myungjin Bae (mjbae@ssu.ac.kr) Soongsil University, 1-1 Sangdo 5 dong DongJak-Ku, Seoul, 156-743, Republic of Korea

© SERSC 2013

At chapter 2, the conventional waveform method which is to get recognizable characteristics of signal, and at chapter 3, the proposed method to use the hybrid coding 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 Existing Method

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.



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

Fig.1. Examples of cosine interpolation method

Figure 1 shows these two sampling methods for linear sampling and non-uniform one. where, *M(.)* are the magnitudes as peaks and valleys of non-uniformly sampled data and *I(.)* are the intervals of them. 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. The reconstructed waveform, $y_k(n)$, obtained by cosine interpolation method is represented as Eq. 2-1.

3 Proposed Method

According to the speech production mechanism, the 3rd and upper formants have broad bandwidths. Moreover, from the viewpoint of speech recognition, higher frequency band components are not significant, while the 1st and the 2nd formants are separated 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.

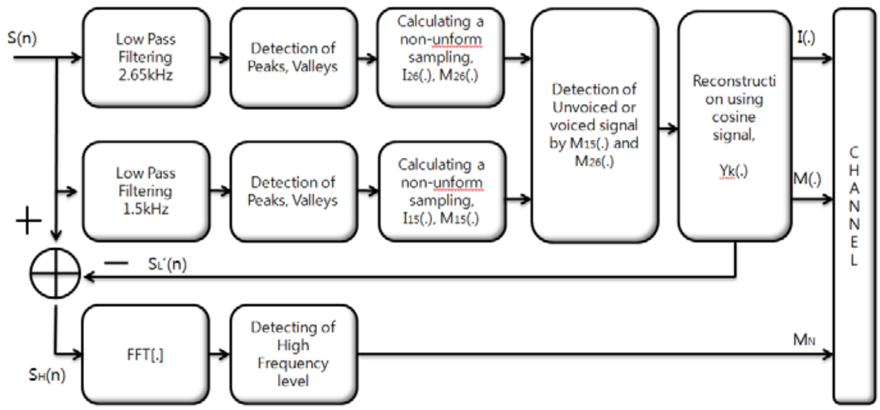


Fig.3 A new speech coding scheme

Fig. 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 $S_{L26}(n)$ is the law-pass filtered signal by 2.65 kHz as cutoff frequency, and $S_{L15}(n)$, the law-pass filtered signal by 1.5 kHz. 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. At the same time, Re-constructed speech signal, $S_L'(n)$, is reconstructed by the cosine interpolation as the non-uniform sampling.

4 Conclusion

We proposed a new hybrid speech coding technique for high quality coding. 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 overcome this problem, we proposed new sampling method using variable bandwidth LPF.

References

1. 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)
2. Jeong chan-joong, Bae myeong-jin, "Analysis regarding vocalizations of teuroteu singers", The institute of electronics engineers of Korea, Summer scholarship contest of 2009 p. 1090~1091, 2009.
3. 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