On A New Speech Synthesis method using Harmonics Filter for Naturalness

Seong-geon Bae, Myung-jin Bae

1Information 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. In speech signal processing, the speech quality with reduce the redundancy within samples get to encode that resulted from domain processing method like PCM, the 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. So we propose to overcome this problem in a new hybrid smart phone.

Keywords: Peak and valley, Waveform coding, Source coding, Sampling.

1 Introduction

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.

2 The Conventional 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.

![Fig.1. Examples of cosine interpolation method](image)

\[ y_k(n) = \left[ \frac{M(k-1) - M(k)}{2} \cos\left( \frac{\pi n}{I(k)} \right) + \frac{M(k-1) + M(k)}{2} \right], 1 \leq n \leq I(k) \]  \hspace{1cm} (2-1)

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 The 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 shows the synthesis block diagram of the proposed in this paper. In the block diagram, $S(n)$ is speech signal digitized uniformly by A/D converter, and $S_{LP}(n)$ is the low-pass filtered signal by 2.75 kHz, and $S(n)$ is a synthesis signal using Eq. 2-1. Then the harmonics filter is applied to two clipping and weighting signals and is added to $S_{LP}(n)$. At the same time, Synthesis speech signal, $S'(n)$, is made by the cosine interpolation, Harmonics filter.

4 Conclusion

We proposed a new enhancement method of speech for high quality. 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


