Title	変調分析に基づいた音声エンハンスメントのための瞬時振幅と瞬時位相の回復処理体系
Author(s)	劉，揚
Citation	
Issue Date	2016-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/13515
Rights	
Description	Supervisor:鵜木　祐史，情報科学研究科，博士

Abstract

Speech is one of the most important carriers of communications in our daily life. However, in real-world listening environments, speech signals are often smeared by various types of acoustic interferences, such as background noise and reverberation. When only monaural information is available, single-channel speech enhancement techniques are used to reduce the effects of acoustic interferences. They are especially interesting due to the simplicity in microphone installation but the major constraint of single-channel methods is that there is no reference signal for the noise available such as sound location. Therefore the performance of important applications such as hearing aids and automatic speech recognition systems, where only one microphone is available due to cost and size considerations, may severely reduce when the speech are subjected to the acoustic interferences. In order to facilitate the performance in the important applications, it is, of great necessity, to conduct some research about the single-channel speech enhancement to improve the performance of speech communication applications.

Many conventional methods of single-channel speech enhancement have been proposed in the last a half of century. These methods can suppress the effects of noise or reverberation well but they can only improve the perceived overall quality but not the intelligibility of speech. Perceived overall quality is the overall impression of the listener of how "good" the quality of the speech is and intelligibility is a measure of how comprehensible speech is. There is substantial evidence that many signals can be represented as low frequency modulators which modulate higher frequency carriers. This concept called "modulation analysis" is useful for describing, representing, and modifying acoustic signals. It has been shown that modulation frequency between 4 Hz and 16 Hz is important for speech intelligibility. Therefore, we can focus on restoring the temporal envelope for improving intelligibility of speech. Recent studies have shown that not only the amplitude spectrum but also the phase spectrum contains important information for speech enhancement, however, most of the modulation analysis based methods neglect the phase spectrum information. The most important is that these method only consider magnitude spectrum information without phase spectrum information. Recent psychoacoustical studies have reported that the temporal amplitude envelope (TAE) and temporal fine structure (TFS) are important for speech perception. TAE and TFS representations belong to complex modulation spectrum analysis and play an important role of improving intelligibility of degraded speech, instantaneous amplitude and phase by Gammatone filterbank correspond to TAE and TFS. Therefore, instantaneous amplitude and phase decomposed by Gammatone filterbank based on human hearing characteristics are used in this research for improving the perceived overall quality and intelligibility of speech.

The Kalman filter, which is an efficient computational recursive solution for estimating a signal widely used in fields related to statistical processing, is applied in our proposed methods. In the process of Kalman filter, linear prediction (LP) is utilized to obtain transition matrix. LP uses some previous values in time domain to estimate current value under principle of Minimum mean square error (MMSE), meanwhile, the Kalman filter uses the previous samples to estimate current sample and update information step by step. The Kalman filter with LP process the representations of signal using modulation analysis in time domain frame-by-frame. Cpestral Mean Normalization (CMN) was also applied as post-processing to reduce the effect of early reflection.

In summary, this thesis proposes an efficient speech enhancement method using modulation analysis for instantaneous amplitudes and phases. Instantaneous amplitude and phase are extracted from the sub-bands of Gammatone filterbank, which are representations in modulation analysis, by using the Hilbert transform. Kalman filter with LP is applied to restore the instantaneous amplitude and phase in time domain. Results of the objective and subjective experiments showed that the proposed method can improve much perceived overall quality and intelligibility of speech simultaneously in hearing aids and Automatic speech recognition (ASR) systems, compared with conventional methods such as MMSE method and Wiener filtering method.

Keywords: speech enhancement, instantaneous amplitude and phase, Kalman filter, linear prediction, modulation.