Title	振幅変調・周波数変調特性に基づく音声強調
Author(s)	Nugyen Quoc Huy
Citation	
Issue Date	2024-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19392
Rights	
Description	Supervisor: 鵜木 祐史, 先端科学技術研究科, 博士

Abstract

Clear and intelligible speech is vital for effective human communication, particularly in critical systems like airport communication, where low speech intelligibility can lead to severe consequences. Speech enhancement techniques are crucial for improving speech quality and intelligibility in various real-world applications such as telecommunication, hearing aids, and voice recognition systems, as well as in military and aviation communications where clarity is essential. Despite having a long history of studies, recent speech enhancement techniques still suffer speech over suppression and noise under suppression, distorting the enhanced speech signals, which sometimes have lower quality and intelligibility than the noisy speech itself.

Believing that the gap between mathematical/computational techniques and the nature of speech is the cause of this distortion, this study utilizes the concept of modulation for speech enhancement to build a bridge to connect this gap. The main objective of this research is to investigate the effectiveness of utilizing speech modulation characteristics for enhancement. This main objective contains three sub-objectives: to model the amplitude modulation characteristics for speech enhancement, derive the relationship between amplitude and instantaneous frequency modulation, and enhance speech using the derived relationship.

To achieve the first objective, a method to model the spectral modulation characteristics of speech in amplitude is proposed and applied for speech enhancement. In voiced speech, the speech power spectrum is amplitude-modulated, where the spectral fine structure is periodic with a period equal to the fundamental frequency. Thus, the proposed method constructs the categorical distribution of fundamental frequency to characterize spectral fine-structure characteristics of speech. Evaluating the Valentini et al. dataset, the results show that improving amplitude modulation characteristics improves speech enhancement performance.

The analytical derivative method is proposed to extract instantaneous frequency deviation (IFD) to achieve the second objective. By deriving the principal value of the logarithm of the complex time-frequency representation, an equation connecting the amplitude to the IFD is established. Via single-tone frequency-modulated signals, the proposed method is verified to work correctly, which confirms the proposed equation's validity. As the established equation indicates, this result confirms a connection between amplitude and IFD.

The findings in the second objective provide two critical perspectives on IFD. First, although defined from the phase, IFD has a multiplicative connection with the amplitude, which allows real-valued processing. Second, computationally, the IFD can be derived instantaneously without a time difference. From these findings, a method to enhance speech via IFD is proposed to modify IFD by a learnable affine transform at the frame-wise level. Evaluating the Valentini et al. dataset, the results show that the proposed method improves speech enhancement performance, especially quality. Specifically, the proposed method achieves the Perceptual Evaluation of Speech Quality of about 2.87 and Short-Time Objective Intelligibility of 0.94, outperforming many state-of-the-art techniques in speech enhancement. Significantly, the proposed method improves up to 15% in a 2.5 dB signal-to-noise ratio. These results confirm the effectiveness of using IFD in speech enhancement based on its relationship with amplitude.

All the results confirm that utilizing speech's modulation characteristics can improve speech enhancement performance, satisfying the research objective. This research has established a solid base by showing how effective modulation characteristics can improve speech quality in noisy conditions. This research has practical applications in improving user experience in mobile calls, VoIP services, and video conferencing, as well as benefiting assistive technologies such as hearing aids and cochlear implants. Additionally, it contributes to advancements in audio signal processing, machine learning, artificial intelligence, and neuroscience.

Keywords: speech enhancement, amplitude modulation, spectral modulation, frequency modulation, instantaneous frequency deviation