

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: 鷗木 祐史, 先端科学技術研究科, 博士

氏名	NGUYEN QUOC HUY		
学位の種類	博士 (情報科学)		
学位記番号	博情第 534 号		
学位授与年月日	令和 6 年 9 月 24 日		
論文題目	Speech Enhancement based on Amplitude and Frequency Modulation Characteristics		
論文審査委員	鶴木 祐史	北陸先端科学技術大学院大学	教授
	NGUYEN Le Minh	同	教授
	岡田 将吾	同	准教授
	LU Xugang	情報通信研究機構	主任研究員
	水町 光徳	九州工業大学	教授

論文の内容の要旨

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 2.5dB 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

論文審査の結果の要旨

音声は我々の日常生活で欠くことのできない重要な情報である。我々の身の回りの音環境には、背景雑音や残響など様々な妨害音があり、音響的な干渉としてこれらが音声を歪ませてしまい、さらにはその音質や明瞭性を著しく低下させてしまう。そのため、音環境に因らず明瞭で聞き取りやすい音声を提供することは、ヒューマンコミュニケーションにおいて非常に重要である。これまでに、音声強調法や雑音除去・残響除去法など数多くの音声回復法が提案されてきた。古典的には、決定論的手法にはじまり、統計的信号処理に基づく処理法、高次統計量を利用した処理法、スパースコーディングに基づく処理法などが検討されてきた。近年では、深層学習に基づく音声強調法が主流になりつつある。しかし、これら多くの方法では、音声の過度な抑圧や雑音の過度な除去により、客観的には大きな改善が得られたとしても、強調された音声信号が歪んでしまうことにより、主観的には、強調以前の音声よりも音質や音声了解度を大きく低下させてしまうことがあり、大きな問題になっている。これは、音声強調における数学的・計算的手法と音声の性質に関係する音響特徴の取り扱いの間に大きなギャップにあることに起因するものと考えられる。

本論文では、上述のギャップをつなぐ架け橋として、音の変調知覚に関連づけた音声強調のための変調概念を利用し、次の三つの点を考慮して、音声強調法の本質的な問題の解決に取り組んだ。(1) 音声のスペクトル変調特性を振幅変調モデルで表現した音声強調法を提案した。(2) 振幅変調モデルにおける搬送波情報として、振幅変調と瞬時周波数偏差 (IFD) の乗法的な関係から IFD を抽出する解析的抽出法を提案した。(3) フレーム単位で学習可能なアフィン変換による IFD の修復処理による音声強調法を提案

した。最後に、これらの方法を組み合わせた音声強調法を提案することで、様々な背景雑音下における量的な改善だけでなく、音質と音声了解度の大幅な改善も可能とした。

以上、本論文は、音声の変調特性を利用した音声強調法を確立した。振幅変調に基づく音声の振幅情報の修復方法と、振幅変調に関連づけた位相情報の修復方法を実現した点で、高い新規性と独創性が認められる。また、音声強調において量的だけではなく質的にも改善効果が認められるなど有効性も認められる。これらの成果は、国際的に定評のある学術誌に掲載されるなど学術的水準も高い。本技術は聴知覚由来の音声情報処理として検討可能であり、その応用範囲も広く、学術的に貢献するところも大きい。よって博士（情報科学）の学位論文として十分価値あるものと認めた。