氏　　名	劉　　揚		
学位の種類	博士(情報科学)		
学位記番号	博情第 340 号		
学位授与年月日	平成 28 年 3 月 24 日		
論文題目	Restoration Scheme of Instantaneous Amplitude and Phase for Speech Enhancement Based on Modulation Analysis （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）		
論文審査委員	主査　鵜木　祐史	北陸先端科学技術大学院大学	准教授
	赤木　正人	同	教授
	党　　建武	同	教授
	田中　宏和	同	准教授
	Lu Xugang	情報通信研究機構	主任研究員

論文の内容の要旨

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.

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．しかしながら，音声コミュニケーションでは，背景雑音や残響といった音響的な干渉によって音声信号が歪んでしまい，その音質や明瞭性を著しく低下させる．そのため，これまでに数多くの音声回復手法が提案されてきた．特に，これらの手法はマイクロホンアレー等の音声信号処理技術の応用問題として広く検討されてきた．しかし，単一マイクでの音声回復処理に至ってはまだ十分と呼べるだけ発展しているわけではない．この問題は，何らかの事前情報や知見が無い限り，音声とその干渉を完全に分離できないことに起因する．実際，単一マイクを利用する音声通信システムや自動音声認識システムといった重要な応用では，これらの干渉を取り除かない限り大幅な進歩は望めない．特に，補聴技術にこれらを利用するためには，雑音残響除去性能だけでなく，それらを取り除いた後の音質ならびに高い明瞭性・了解性を保持することが重要である．そのため，これらの問題点を解決し，音声アプリケーションの性能を向上させるためにも，単一チャンネルの音声回復法を実現する必要がある．

本研究では，変調伝達関数（MTF）の概念とカルマンフィルタを利用した二種類の音声回復法を提案した．いずれの方法も，ガンマトーンフィルタバンク（GTFB）を信号表現として利用したうえで，雑音・残響の影響を取り除くものである．前者は，GTFB 上の振幅包絡線情報に関して MTF とカルマンフィルタを利用して雑音残響除去（パワーエンベロープ回復法）を実現した．後者は，前者の改良版であり，再帰的なカルマンフィルタを利用することで音声の瞬時振幅だけでなく瞬時位相を回復する音声回復処理を提案した．本研究で取り扱うカルマンフィルタでは，振幅包絡線あるいは瞬時振幅・瞬時位相の時間変化を表現する遷移行列を得るために線形予測法を利用した．両者を系統的に評価するために，雑音残響除去性能を重点的に評価した．その後で，回復音声の音質・明瞭性の観点から主観評価実験等を行った．その結果，後者が最善な方法であることを明らかにした．これは，瞬時振幅と瞬時位相の時間変動に関して滑らかさの規範を設けることが音声回復ならびにその知覚にとって重要であり，振幅・位相に関する変調成分の制御の重要性を示している．

以上，本論文は，音声回復という主要な課題において，いかに雑音・残響を効果的に除去し，音質・明瞭性を回復するかという点について瞬時振幅・位相の制御の重要性を示しており，学術的に貢献するところが大きい．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．