JAIST Repository
https://dspace.jaist.ac.jp/

Title	多変量経験的モード分解を利用した音源フィルタモデルに基づく音声分析法
Author(s)	Surasak, Boonkla
Citation	
Issue Date	2018-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/15326
Rights	
Description	Supervisor:鵜木　祐史，情報科学研究科，博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

氏　　名	BOONKLA Surasak
学位の種類	博士(情報科学)
学位記番号	博情第 389 号
学位授与年月日	平成 30 年 3 月 23 日
論文題目	Speech Analysis Method Based on Source-Filter Model Using Multivariate Empirical Mode Decomposition
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　教授
	赤木正人　同　教授
	党建武　同　教授
	金子峰雄　同　教授
	Wutiwiwatchai Chai　NECTEC　教授
	Makhanov Stanislav S　タマサート大学　教授

論文の内容の要旨

The growth of speech processing technology within the last few decades enables us to communicate with each other even when we are too far apart by using speech. It is not only human-to-human but also human-to-machine communication that become important and play a vital role in our daily life. However, the speech communication is always damaged by environmental noises. Moreover, multiple echoes (reverberation) within a confined space cause severe reduction of speech intelligibility as well. These drawbacks exist since the beginning of the speech communication. To date, researchers are attempting to solve these problems because they still degrade the communication systems.

Since the availability of digital hardware, there has been much research in speech processing technology especially speech analysis which is the backbone of several applications such as voice activity detection, speech enhancement, automatic speech recognition, speaker recognition, and hearing aids. The performance of these applications degrades drastically in real environments because the speech analysis method employed by these applications is not robust against noises and reverberation. We aim to propose the robust speech analysis method by using multivariate empirical mode decomposition (MEMD). The motivation of using MEMD is that it can extract the oscillation components and make the signal sparse by reducing the degree of mixing. This ability can reduce the degree of mixing of noises in the noisy speech signals. Furthermore, MEMD can automatically separate the signals which are resulted from the addition of sub-signals. For example, automatic source-filter separation, automatic noise separation, and automatic separation of cepstrum of room impulse response. Therefore, the MEMD-based speech analysis method can ideally be able to fulfill the following requirements. (i) the source and vocal tract information are obtained simultaneously. (ii) robust against noise. (iii) robust against reverberation, and (iv) robust against both noise and reverberation.

This research aims to solve the problems of speech analysis in real environments by proposing the robust

MEMD-based speech analysis method. It exploits specific properties of MEMD as follows: (1) it can analyze the non-stationary signal. Since speech signal is the non-stationary signal, MEMD should be the appropriate approach for speech analysis. (2) It is the nonparametric and data-driven approach. MEMD does not impose any assumption regarding the input signal. (3) It can automatically separate mixtures of signals or reduce the degree of mixing. (4) It can automatically align the common component into the same index of sub signal namely intrinsic mode function (IMF). However, the challenge of using MEMD is how to correctly categorize IMFs derived from MEMD into groups of sources, vocal tract, noise, reverberation. Four main tasks would be focused on to achieve the final goal of this research. That is MEMD-based speech analysis method in (a) clean, (b) reverberant, (c) noisy, and (d) noisy reverberant environments. Then the proposed speech analysis method will be applied to some practical applications to show its effectiveness.

If estimates of speech features can be further improved by the proposed method in real environments, it would directly have a great impact on the society of speech signal processing. It would also contribute to the engineering and technology in the sense that the performance of several critical applications, for example, voice activity detection, speech recognition, hearing aids, speech enhancement, and communication systems would be enhanced. Furthermore, it would have the indirect contribution to human society when the performance of such applications is improved. Throughout this dissertation, the reader will see how our proposed speech analysis is carried out in clean, noisy, and reverberant conditions. Some applications, based on the techniques used in our speech analysis, such as voice activity detection, noise reduction, and speech dereverberation are demonstrated as well. We proposed MEMD-based speech analysis for clean speech that is superior to linear prediction and cepstrum based methods in F_0 estimation. In noisy conditions, we cooperated the MEMD-based noise reduction technique with the MEMD-based speech analysis method so that the speech analysis could be robust. In reverberant conditions, we could reduce the effects of reverberation by using MEMD so that the speech analysis could be robust. The final goal of speech analysis in noisy reverberant conditions have not yet completed and will be our future work.

Keywords: speech analysis, source-filter model, multivariate empirical mode decomposition, noise reduction, dereverberation

論文審査の結果の要旨

音声コミュニケーションは、人の営みにおいて欠くことのできない最も重要な手段である。そのため、音声を使っていつでもどこでも誰とでも安全に安心して、コミュニケーションできることが求められている。この目標を達成するために、様々な音声信号処理技術が発展してきている。例えば、音声認識技術や音声合成技術、音声符号化技術、音声変換技術といったものである。これらの技術に共通する重要な要素技術は「音声分析技術」である。この技術は半世紀前から行われており、線形予測分析やケプストラム法といった古典的なものから STRAIGHT といった最新のボコーダー型のものまで知られている。しかし

ながら、これらすべての方法はクリーンな音声信号を分析対象としており、雑音や残響といった外乱があるとその影響から音声分析性能は著しく低下してしまう。そのため、現実的な音声コミュニケーションにおいてこの技術を利用することを考えると乖離がみられる。これは、様々な音声信号処理技術を実環境で利用するときの根幹にかかわる問題点である。

本研究では、これらの問題意識のもと、クリーンな環境だけでなく、雑音や残響がある環境で収録された音声からでも、頑健で正確に音声の特徴分析を可能とする方法を提案した。一般に音声分析は音声生成の観点から音源フィルタモデル（音源信号と声道フィルタの線形モデル）に基づくが、背景雑音は発声された音声に加算され、また残響は発声された音声に畳み込まれ、複雑に音源・フィルタの特性を歪ませる。本研究では、音声に対するこれらの外乱の加法性・重畳性を適切な領域で表現し、多変量経験的モード分析（MEMD）を利用することで、これらをその領域で線形分離できるという点に着目した。まず、クリーンな環境において MEDM による音声の音源信号とフィルタ特性にモード分解し、基本周波数とフォルマント周波数、スペクトル傾斜といった特徴を正確に推定できることを示した。次に、雑音環境では加法性雑音成分をそれぞれ分解し、また残響環境では音源信号とフィルタ特性に重畳された残響成分をそれぞれ分解し、音声の特徴を頑健で正確に推定できることを示した。最後に、これらの方法が、音声の特徴分析と同時に雑音・残響除去も行うことができることを示した。また、その応用例としてロバスト音声区間検出法を実現した。

以上、本論文は、雑音の加法性・残響の重畳性と MEMD の特性とを有機的に結合させることで、外乱に頑健で正確な音声分析法を実現できたことから、音声信号処理の基盤技術として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。