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

Title	音環境バリアフリーのためのパワーエンベロープ処理体系
Author(s)	森田，翔太
Citation	
Issue Date	2017-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/14249
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	森田翔太		
学位の種類	博士(情報科学)		
学位記番号	博情第363号		
学位授与年月日	平成29年3月24日		
論文題目	音環境バリアフリーのためのパワーエンベロープ処理体系		
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	教授
	赤木 正人	同	教授
	党 建武	同	教授
	田中 宏和	同	准教授
	三好 正人	金沢大学	教授

論文の内容の要旨

Concept of Ubiquitous speech communication is speech communication for “Anytime”, “Anywhere”, “Anybody”, “Security & Safe”. “Anywhere” in Ubiquitous speech communication is important problem for smooth speech communication in sonic environments. Effects of noise and reverberation in real environments are barrier in sonic environments. I set the big goal in my research that realization of barrier-free for sonic environments. The barrier-free has some approach as speech enhancement, noise reduction and dereverberation. In this paper, the barrier-free is supposed to approach of noise reduction and dereverberation likes cancel out of barrier. Current simultaneous methods for noise reduction and dereverberation is not approach of congruous to sonic environments. Then, these methods cause over and under improvements. Noise and reverberation should be estimated as measurements and index of noise and reverberation. MMSE-STSA as noise reduction needs a priori SNR that is estimated as parameter for noise. Almost dereverberation methods, inverse filtering approach by estimated room impulse response (RIR) and preliminary measured RIR is used. In simultaneously methods of noise reduction and dereverberation, effects of noise and reverberation are reduced by noise reduction and dereverberation respectively. The method is combined methods of noise reduction and dereverberation. Therefore, the methods cannot work likes congruous to sonic environments, thus it was caused over and under improvements. Voice activity detection (VAD) is technology for speech coding and preprocessing for automatic speech recognition (ASR). Robust VAD for the barrier-free need fixing the discriminate condition for detecting signal to obtain the required performance of VAD. Noise reduction is used as pre-processing for robust VAD. VAD performance is degraded in not involving supposed sonic environments. Human has stress when they hear the over and under restored speech by the barrier free. Recognition performance of machine is degraded in that case. It is caused by incongruous between sonic environments and human with machine. A goal of my research is congruous

processing between sonic environments and human with machine as the barrier-free. Priority element to solve the problem is “congruous to sonic environments.” This paper treat noisy reverberant environments as sound field. The user for the barrier-free is set to machine in this research. The temporal power envelope of signal is treated as target signal in the barrier-free. This paper achieves that power envelope processing systems for barrier-free communication under noisy reverberant environments. As requirements for power envelope processing systems, effects of noise and reverberation are reduced as one sonic environments, and parameters for noise reduction and dereverberation should be used on a feature. Concept of modulation transfer function (MTF) is based to achieve the power envelope processing system for barrier-free communication in noisy reverberant environment. MTF can treat noise and reverberation simultaneously on the MTF. MTF is matched for congruous between sonic environments and human. On MTF concept, modulation index (MI) for power envelope of input signal is 1, MTF is decided by effects of noise and reverberation, and MI for power envelope of output signal is obtained as 1 minus MTF. The system based on MI = 1 for congruous to sonic environments. The power envelope signal processing system is based on the concept. The cohesive speech signal processing system was proposed to achieve power envelope processing system for barrier-free communication in noisy reverberant environments. The cohesive speech signal processing system is consisted of robust VAD, power envelope restoration including power envelope subtraction and MTF inverse filtering, SNR estimation, and reverberation time estimation. Most of these processing is based on MI = 1 of MTF concept. These methods were evaluated in noisy reverberant methods, then the results show these methods can work in noisy reverberant environments. For the application as speech signal processing, the cohesive speech signal processing system was applied to pre-processing of ASR system and STI estimation method. The results show that proposed approach can work as one approach of the barrier-free to sonic environments from the results.

Keywords: power envelope restoration, noisy reverberant environments, modulation transfer function, congruous to sonic environments,

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（ユビキタス音声コミュニケーション）が求められている．その中でも，「どこでも」に係る外的阻害要因を取り除くことは，音環境バリアフリーという観点からも重要な研究課題として検討されている．従来，雑音除去，残響除去，あるいは音声強調と呼ばれる音声信号処理は，音環境における外的要因を取り除く作業に相当する．これらの研究では，スペクトルサブトラクション法，MMSE-STSA

法，MINT 法，ICA 法，BSS 法など多岐にわたり様々な手法が提案されてきた．これらの成果により，定常雑音の除去や既知の残響特性の除去，独立した音源信号の分離といった諸問題が解決されてきた．しかし，音環境バリアフリーの観点から重要視すべき本質的課題（外的要因の取り残しや過剰な除去による歪の発生の低減，音環境の違いへの柔軟な対応，音環境と人の間で調和がとれた状況の創出）については，未だに解決されていない．

本研究では，上述した問題意識を持ちつつ，音環境バリアフリー技術を確立するためには，音環境と人との間に入って外的阻害要因を適切に取り除くための親和性の高い信号処理体系が必要であるという位置づけで，変調伝達関数の概念に基づいたパワーエンベロープ処理体系を提案した．このパワーエンベロープ処理体系では，雑音残響環境を変調伝達関数で表現し，原信号とそれらによって影響を受ける観測信号のパワーエンベロープの特性を線形システムとして一貫して説明することができる．また，話者（音源）の変調特性や聴取者の変調知覚の特性とも深く，直接的に関係するため，音環境と人との間を取り持つ親和性の高い信号処理体系であると考えることができる．

本論文では，このパワーエンベロープ処理体系において，(1) 信号検出理論に基づいた外的要因に頑健な音声区間検出法，(2) 雑音環境を特徴づける SNR のブラインド推定，(3) 残響環境を特徴づける残響時間のブラインド推定，ならびに (4) 音声信号のパワーエンベロープ回復処理を，変調伝達関数の概念に基づいた統合処理の仕組みとして実現した．この仕組みの応用として，(i) 頑健な音声認識のフロントエンド処理と(ii) 音環境における音声伝送指標を逆推定する方法を実現した．

以上，本論文は，音環境バリアフリーの観点から，変調伝達関数の概念に基づいたパワーエンベロープ処理体系ならび音声回復に必要な要素技術・応用技術を提案したものであり，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．