HOKKAIDO UNIVERSITY

Title	A Study on Robust Speech Recognition with Time Varying Speech Features [an abstract of dissertation and a summary of dissertation review]
Author(s)	Mufungulwa, George
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12944号
Issue Date	2017-12-25
Doc URL	https://hdl.handle.net/2115/68113
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	George_Mufungulwa_review.pdf, 審査の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文審査の要旨

博士の専攻分野の名称　博士 (情報科学)　氏名　George Mufungulwa

審査担当者　主　査　教　授　宮永 喜一
副　査　教　授　大鐘 武雄
副　査　教　授　齊藤 晋聖
副　査　准教授　筒井 弘

学位論文題名

A Study on Robust Speech Recognition with Time Varying Speech Features
（音声の時変特徴量を用いた雑音ロバスト音声認識に関する研究）

本論文は, 様々な雑音環境下において, 音声を高精度に認識できる自動音声認識システムの開発に関する研究について述べている。

近年, 音声認識や自然言語処理技術の高度化に伴い, 実用性の高い音声認識システムが実環境下において利用されている。例えば, Apple iPhone に搭載された Siri などのように, 日本でも複数の商用サービスが開始され, 特に情報検索をアシスタントするエージェント系のサービスが多い。しかしながら, 現状の音声認識システムでは, 様々な雑音環境下においての利用が難しいとの課題もある。その原因としては,(1) 室内環境において, マイクと音声の発声する音源との距離が遠いなどの理由で, SNR が 10dB 以下の環境となる場合, 誤り率が急に増大し, 高精度な音声認識が難しい。(2) 室外環境では, 騒音に対して, 5 - 7dB 程度の SNR で音声を観測する状況が多く, 高い音声認識率が望めないなど。これらの問題を解決するには, 雑音環境下においても, 音声の認識が高精度で実現できる新たな手法が必要と考えられている。

本論文は,7 つの章で構成されている。

第 1 章は, Introduction であり, 雑音に強い音声認識の必要性について述べている。

第 2 章は, 音声認識システムの基本的な手法について説明している。音声認識は, 大きく分けて, 音声分析手法と音声認識手法に分けられる。音声分析手法は, 観測された音声信号より, 音声認識に必要と考えられている, 音声の特徴量を推定する方法である。従来の手法は, 音声信号を, 10msec 程度の短いフレームに分け, そのフレーム内で観測される音声の定常スペクトルを求める。代表的な手法は, 短時間フーリエ変換であり, その他に, 線形予測モデル推定や, 最尤推定手法などがあげられる。

この章において, 従来通信などに利用されていた時変線形予測モデル (tvLPC モデル) が紹介されている。このモデルにより推定された新しい音声の特徴量を音声認識用に拡張し, それによる音声認識用時変音声分析手法を提案した。

第 3 章は, 音声の自動検出手法が提案されている。音声認識の開始時点は, 音声が観測されてからであるが, その観測された時点を決めるには, ユーザが, 事前に開始点をシステムに示す, 最も簡単な方法から, システムが, 自動的に音声と非音声を区別し, その結果を利用して, 音声波形の存在している部分のみを検出し, 音声分析の対象とする自動音声検出手法がある。本論文で対象の手法は, 自動音声検出手法であり, 様々な雑音環境下においても, 音声波形が観測できるように, 音声エ

ネルギーの時間的変化を適応的に評価したうえで, 音声の区間検出を行う手法を提案している。

第 4 章は, 雑音低減手法の紹介とその拡張として新方式の提案を行っている。雑音環境下の音声から, 雑音だけを低減する手法は, 様々な方式がすでに提案されているが, 低減効果が比較的高いと考えられている手法であるスペクトルグラム上でのバンドパスフィルタリング (Running Spectrum Filtering (RSF)) を拡張し, 変調スペクトル分析手法 (Running Spectrum Analysis (RSA)) を新たに提案している。これにより, 雑音環境下においてでも, 類似した音声 (例えば,「電気」,「元気」,「天気」など) の認識がより高精度で実現できることが示された。

第 5 章は, 上記の提案手法である, 雑音環境下における自動音声検出法, tvLPC,RSA に基づく音声分析, HMM を利用した音声認識手法を組み合わせた, 新しいロバスト音声認識システムの, 性能評価を行っている。tvLPC では, 従来の音声の特徴量に加えて,1 フレーム区間内において, スペクトルの大まかな傾きを示す第 1 次パラメータのみを導入することで, 認識性能が若干向上することを示し, RSA については, 男性発声と女性発声では, 異なる RSA 仕様で処理するほうが高効率であることや, 類似語発声フレーズとそれ以外の発声フレーズでは, RSA 仕様が異なる知見を得ている。

第 6 章及び第 7 章は, 全体のまとめと, 今後の課題について述べている。

以上より, 本論文では, 様々な雑音環境下におけるロバスト音声認識システムに関する研究・開発を進めており, 有効な音声認識システムを提案し, その性能評価を行った。これにより次世代の情報化社会に必要とされる新しいロバスト音声認識技術の研究開発を行った。

これを要するに, 筆者は, 様々な雑音環境下における高性能な音声認識システムの提案・開発・実現・評価を行った。これにより, 音声情報処理技術に関する多くの有益な知見を得ており, 情報科学の分野に貢献するところ大なるものがある。

よって筆者は, 北海道大学博士 (情報科学) の学位を授与される資格あるものと認める。