



Title	A Study of Noise-Robust Speech Recognition and Time-Varying Speech Features [an abstract of dissertation and a summary of dissertation review]
Author(s)	Ang, Federico Mendoza
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12190号
Issue Date	2016-03-24
Doc URL	https://hdl.handle.net/2115/61744
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Federico_Mendoza_Ang_review.pdf, 審査の要旨



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 Federico Mendoza Ang

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

学位論文題名

A Study of Noise-Robust Speech Recognition and Time-Varying Speech Features
(時変音声特徴抽出とロバスト音声認識に関する研究)

本論文は、雑音環境下における連続音声認識システム及びフレーズ音声認識システムに関する性能向上について述べ、主にタガル語 (フィリピンで利用されている言語) を対象とした音声認識の性能評価を行っている。

音声認識は、現在、クラウドネットワークを利用した連続音声認識システムと、オートノーマス型のフレーズ認識システムに分かれている。認識率は、認識対象をかなり限定しているオートノーマス型のフレーズ・孤立単語認識システムが高い性能を維持しているが、そこで利用されている技術は、隠れマルコフモデル (HMM) のみである。一方、言語モデルの活用も含めた連続音声認識は、雑音環境下において性能劣化が起こり、性能向上が重要な課題として現在も研究が進められている。

HMM の認識性能が高いことはすでによく知られているが、その高い認識性能を実現するには、多くのデータを事前に準備し、そのデータによる詳細な学習を行う必要がある。この学習は、認識対象が変化するたびに必要であり、多くのコストを必要としていた。例えば、大量のデータベース化が十分ではない、タガル語の音声では、良好な HMM を自動生成できるようなデータベースの準備が難しい、このような場合には、類似した言語、たとえば英語のデータベースを活用し、言語間の類似性と相違性を明示的に表現し、有効な認識システムを構築するようなことが考えられる。

さらに、雑音環境下における連続音声認識を考えた場合、雑音に埋もれてしまう音声の特徴を正しく表現することが重要な課題となる。この課題を解決するためには、雑音中の音声そのものから、有益な音声の特徴を抽出できる時変の特徴抽出手法が有効であると考えられる。時変音声の特徴抽出技術は、すでにいくつか提案されているが、本論文では、音声認識に適した新しい方式を提案し、その有効性を示している。

本論文は、下記のような構成となっている。

第 1 章は、連続音声認識システムとフレーズ音声認識システムの概要について述べている。

第 2 章は、言語に依存した音声の発話構造や特徴に基づいて設計されている連続音声認識の方式に、異なる言語でも適用可能なシステムへの拡張を行い、その性能評価を行っている。

第 3 章は、雑音環境下における音声信号の特徴について詳細に検討し、雑音音声から、音声の時変特徴を推定する手法についての調査と、音声認識に向けた時変音声分析の手法を提案している。この手法は、時変線形予測手法 (time-varying Linear Prediction Coding Method, tvLPC) と呼び、認識システムに利用することを前提とした場合における、従来手法との性能比較により、新方式の有効性を

示している。

第4章は, tvLPC の手法を利用した場合の, 雑音環境下における音声認識の性能について, 実証実験により評価している。実験では, 様々な雑音源を用いての実験や, 種々の SNR 環境による実験を行っており, 時不変とみなした音声の特徴である従来の MFCC に対して, 単純に付け加えるような形式での tvLPC 利用では, 十分な性能向上が得られないことを明確にしている。

第5章は, tvLPC から推定される時変の特徴量を, (1) 時間的平均により雑音による推定誤差を軽減するアルゴリズムを利用する方法, (2) 非線形フィルタを利用して, 時変な特徴量の中で, 認識性能に大きく寄与している特徴量を選択する方法, (3) 主成分分析などを利用して, 観測フレーム内で, 時変な特徴量の主成分を利用する方法などの, いくつか異なるアプローチを導入して, 認識システムの性能向上について調査している。

第6章は, 第5章で述べた, いくつかの手法をすべて利用し, 得られる複数の認識結果から, 多数決により最終結果を求める統合システムを構築した。その結果, 認識性能が向上し, 従来手法に比べて, 優れた音声認識を, 雑音環境下で実現できることを示している。

第7章は, 上記の各章のまとめと, 本研究の総括を行っている。

以上より, 本論文では, 雑音環境下における音声認識システムの開発について詳細に検討し, 時変な音声の特徴抽出による新しい音声認識システムを提案し, より実用的なロバスト音声認識の提案・開発・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音声認識技術の開発を行った。

これを要するに, 筆者は, 雑音環境における高性能ロバスト音声認識システムの提案・開発・評価を行った。これにより, 音声認識技術に関する多くの有益な知見を得ており, 情報科学・工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。