



Title	A Study on Efficient Robust Speech Recognition with Stochastic Dynamic Time Warping [an abstract of dissertation and a summary of dissertation review]
Author(s)	孫, 喜浩
Citation	北海道大学. 博士(情報科学) 甲第11523号
Issue Date	2014-09-25
Doc URL	<a href="http://hdl.handle.net/2115/57250">http://hdl.handle.net/2115/57250</a>
Rights(URL)	<a href="http://creativecommons.org/licenses/by-nc-sa/2.1/jp/">http://creativecommons.org/licenses/by-nc-sa/2.1/jp/</a>
Type	theses (doctoral - abstract and summary of review)
Additional Information	There are other files related to this item in HUSCAP. Check the above URL.
File Information	Xihao_Sun_review.pdf (審査の要旨)



[Instructions for use](#)

## 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 孫 喜浩

審査担当者 主 査 教 授 宮永 喜一  
副 査 特任教授 野島 俊雄  
副 査 特任教授 小川 恭孝  
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 Efficient Robust Speech Recognition with Stochastic Dynamic Time Warping  
(確率的 DTW を用いた高効率ロバスト音声認識に関する研究)

本論文は、計算量を削減したフレーズ音声認識に関する新しい方式の提案を行い、その方式実現と評価を行っている。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) である。

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、そのデータによる精密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。一方、古くから利用されている動的時間伸縮手法 (DTW) は、学習を必要としない簡易型の認識システムであり、学習コストはゼロとなる。しかし、少ない登録音声データだけでは、性能が上がらず HMM に比べ実用的な認識性能を実現しているとは言い難かった。そのため、改良型 DTW として、大量の規範データを利用することで、認識精度を向上させる方式が提案されている。しかし、これらの手法では、多くの計算量を必要としており、実用性を考慮した場合、計算量の削減が必要とされていた。本論文は、これらの課題を解決するために、計算量を削減したフレーズ音声認識に関する新しい方式の提案を行い、その方式実現と評価を行っている。

本論文は、下記のような構成となっている。

第 1 章は、音声認識システムの概要について述べている。

第 2 章は、HMM に基づく音声認識と DTW に基づく音声認識システムの説明を行っており、音声分析から、音声認識で利用されている各種技術などについての概要を説明している。

第 3 章は、観測音声波形より、認識対象としての音声波形を自動的に抽出する技術である、音声区間自動抽出法 (Voice Activity Detection(VAD)) についての必要性や、本提案方式で導入している VAD の技術に関する有効性などについて説明している。

第 4 章は、ロバスト音声処理について説明している。従来技術である、ランニングスペクトルフィルタリング (RFS)、ケプストラム平均処理 (CMA)、ダイナミックレンジアジャストメント (DRA) 等について性能評価と、DTW との連携による性能向上について説明している。

第 5 章は、従来の DTW に関する説明とその性能比較を行っている。ここで説明しているマル

チレファレンス DTW は、基本的な DTW 手法に、確率的な処理を導入し、さらに最終認識段階において、高精度化を実現できる非線形フィルタリングも導入している。これにより、ロバスト HMM 認識と同程度の性能を実現できることを示している。一方、膨大な計算量を必要としており、実用性の観点から、その計算量削減が必要であることも示している。

第 6 章は、本論文で提案する計算量を削減したフレーズ音声認識向けの DTW の説明を行い、その方式実現と評価を行っている。さらには、提案する DTW と従来手法との比較を行っている。

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実用的 DTW 音声認識システムの改良設計・開発について詳細に検討し、DTW に基づく実用的なロバストフレーズ音声認識の提案・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フレーズ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。