



Title	音響的類似単語の雑音ロバストフレーズ音声認識に関する研究 [論文内容及び審査の要旨]
Author(s)	宮崎, 善行
Citation	北海道大学. 博士(情報科学) 甲第15559号
Issue Date	2023-03-23
Doc URL	http://hdl.handle.net/2115/89606
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
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	Yoshiyuki_Miyazaki_abstract.pdf (論文内容の要旨)



[Instructions for use](#)

学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 宮崎 善行

学 位 論 文 題 名

音響的類似単語の雑音ロバストフレーズ音声認識に関する研究

(A Study on Robust Phrase Speech Recognition for Noisy Acoustically Similar Words)

第1章では研究の背景と目的について述べる。音声認識は、有用なヒューマンインターフェースであり、音声入力インターフェースとして機器の音声制御、ハンズフリー操作などの実現に役立つ。近年、音声認識を搭載した機器が商品化され、日常的に音声制御の可能な製品が開発されつつある。一般的に、大語彙音声認識が可能で、カーナビ等において地名等の固有名詞を認識できるとして高い有効性を示すシステムでは、雑音環境下での認識率を向上させるのが難しいために、音響処理段階では耐環境順応特性を向上させる手法が導入され、現時点でもその性能向上が研究されている。

本研究にて提案するフレーズ音声認識方式は、音響的な特徴をできる限り音声分析により見つけ、その特徴に基づいて認識結果を推定する。この場合、雑音環境下で判別の困難な認識語とは、音響的な発音が似ていて認識の難しい単語と考えられる。一般的に、SNRが低くなるにつれて、認識精度は低くなるが、各々の単語について、一様に低下するわけではない。低SNR環境でも認識精度が高い単語がある一方、音響的類似単語が登録単語辞書に存在する場合、SNRがそれほど低くない環境であっても認識精度が劣化する場合がある。本研究では、上記のような認識精度が劣化する場合の認識率改善手法を提案する。

第2章では雑音ロバスト音声認識システムの構成要素を紹介する。具体的には音響分析として、音声生成の仕組み、音声分析、学習、認識と共に、雑音ロバスト音声認識について述べる。

第3章では、本論文で「難認識語」と呼ぶ認識が難しい単語に対する規定について述べ、難認識語の認識精度を改善する鍵となる音響的差異区間について説明する。雑音環境下における雑音ロバスト音声認識の単語別の認識率に関するヒストグラムにおいて、交差検証法に基づいて認識率の評価を行った。SNRが低下するにつれて、他の単語よりも認識率が著しく低下する単語が存在する。例えば、「財布」と「ライス」のように音響的類似単語がそれらに相当する。さらに、共通の音素列を含む単語もそれらに含まれるような特徴を持

つ単語を「難認識語」と呼ぶ。そして、音響的類似性を定量的に評価するために、DTW 距離を適用する。DTW 距離による評価において、局所的な差異が存在する区間を音響的差異区間とし、音響的差異区間を抽出して認識することができれば、認識率の向上が期待できる。

本研究では、発話音声の中から音響的に類似しない音響的差異区間を切り出し、認識することにより、音響的類似単語の認識精度の改善を試みる。音響的差異区間の切り出しは、ビタビアルゴリズムにより、自動的に行うことができる。したがって、その区間が、単語のどの位置にあっても、そして、破裂音でなくても適用できるという利点がある。

第4章では、認識精度が最も高いと期待される音響的差異区間を決定する方法について説明し、音響的差異区間の決定を行うと共に、難認識語の認識を行う。単語を発声した教師音声を分析して得られる音声特徴量列について、単語の HMM に対する対数尤度をビタビアルゴリズムで計算すると、ビタビ経路を求めることができる。このビタビ経路を利用して、単語の音声から、単語の HMM の始端状態番号から終端状態番号までに対応する区間に存在する音声を切り出し、音響的差異区間候補と呼ぶ。

音響的差異区間候補の音声を教師音声とし、バウム・ウェルチアルゴリズムにより、音響的差異区間候補 HMM のパラメータを求める。難認識語、あるいは、その誤認時頻出語を発声した音声を最も高い精度で認識すると期待される音響的差異区間をその候補の中から決定する。音響的差異区間候補 HMM による難認識語 d の認識精度は、最大値が、従来手法による難認識語の認識精度を超えていれば認識精度の改善を期待できる。

第5章では、全話者 48 名を 8 名ずつの 6 グループに分けて、1 グループ 8 名を認識話者、残りの 5 グループ 40 名を教師話者とし、交差検証法により従来手法と提案手法の認識率を評価した。提案手法の認識率は、従来手法の認識率以上の結果となった。誤認時頻出語とペアになる難認識語の認識精度に対する改善度のほうが上回っている。尚、Speech babble、および、White noise 雑音環境下についても同様の結果が得られた。