

報告番号	※甲	第	号
------	----	---	---

主論文の要旨

論文題目 音声認識における正規化法に関する研究

氏名 磯部 俊洋

論文内容の要旨

音声認識に、HMM (Hidden Markov Model) や N -gram 言語モデルといった統計的手法が用いられるようになり、音声データやテキストデータの整備が進むとともに、音声認識の精度は飛躍的に向上した。近年では、携帯電話やカーナビの操作、電話自動応答サービス、ゲーム等で音声認識が利用されるようになってきている。しかし、現在の音声認識のレベルでは、比較的静かな環境で、新聞記事などの書き言葉を読み上げた音声は高精度に認識できるものの、あらゆる環境で、あらゆる内容の音声を全て認識することは不可能である。特に非定常な雑音環境下において、言い淀み、言い直し、助詞落ち、倒置、間投詞の挿入が生じる話し言葉を認識することは非常に困難である。このため、利用環境や発話の内容、スタイルを想定した上で、音響モデルや言語モデルを構築し、これらのモデルがカバーする範囲を条件として音声認識を利用することが多い。しかし利用条件が限定される音声認識では、条件に対するユーザの戸惑いや、条件に合致しないことによる誤認識で、かえって利便性が悪くなることも少なくない。このため、音声認識の利用条件を無くし、あらゆる環境で、様々な内容、発話スタイルの音声認識を実現することは重要な研究課題である。

音声認識が音声の内容を理解するのに対し、話者認識では音声の個人性を導き出す。特に近年では、音声で本人確認を行う音声認証が、電話自動応答サービスで利用されるケースがある。話者認識では、音声の特徴ベクトルや音響モデル等、音声認識の手法を用いることが多く、先に述べた研究課題は話者認識にも共通となる。

音響的な環境や発話内容に依存して音声認識率や話者認識率が劣化する問題は、音声認識処理の過程で得られる、音声の特徴ベクトル、音響モデルのスコア、言語モデルのスコアがこれらの条件により、大きく変動するためである。従って、特徴ベクトルやモデルのスコアに対して変動を抑制する適切な正規化処理を施せば、音声認識や話者認識の頑健性を向上できる可能性がある。

そこで、本論文では、音響的な環境や、発話内容、話題に対する音声認識、話者認識の頑健性の向上を目的として、認識処理の過程で得られる、音声の特徴ベクトル、音響モデルスコア、言語モデルスコアのそれぞれ領域における正規化法を提案し、頑健性向上の効

果を実験的に確認した。

音声の特徴ベクトルの正規化では、音声認識率が電話回線等の伝達経路の影響により低下する問題に対して、新しいケプストラム平均正規化 (CMN: Cepstrum Mean Normalization) 法を提案した。従来の CMN 法は、入力音声の時間長や発話内容に依存して、CMN に必要なケプストラム平均が変動して本来の値からずれるため、正規化の精度が劣化する問題があった。提案法は、発話に含まれる音素の頻度を考慮してケプストラム平均ベクトルを補正するため、正規化精度の劣化を抑制できるという特徴を持つ。電話音声を認識する評価において、提案法は、従来の CMN 法と比較して、認識率が有線電話で 1.4% (91.5%→92.9%)、コードレス電話で 3.2% (81.5%→84.7%) 向上し、提案手法の効果を実験的に確認した。

音響モデルのスコアの正規化では、発声内容の違いにより音声認証の精度が劣化する問題に対して、新しいスコア正規化法を提案した。音声により本人確認を行う音声認証は、話者認識のひとつであり、入力音声に対する本人の音響モデルのスコアと、事前に設定する閾値との大小関係により受理/棄却の判定を行う。このため、発話内容の違いがもたらす音響モデルのスコア変動は認証精度の劣化の原因となる。従来は、不特定話者モデルや、本人話者と音響的特徴が類似した話者の音響モデルのスコアを用いて正規化を行っていたが、不特定話者モデルは詳細な音響特徴を表現しがたく、また、話者の類似の程度は音韻によって異なるため、十分な正規化を行えない問題があった。提案法は、正規化に用いる音響モデルを HMM の音素、状態、分布が表す局所的な音響情報に基づいて、近傍話者の HMM から合成するため、音響空間において本人話者の音響モデルと正規化用の音響モデルの統計的モデル間距離を小さくでき、スコア変動を統計的に吸収できる特徴を持つ。640 人の電話音声を用いた評価では、分布に基づいて正規化用モデルを生成する提案法は、類似話者の音響モデルスコアを使用する従来法と比較して、Equal Error Rate (EER) を、5.2%から 1.6%まで削減でき、効果を確認した。

言語モデルのスコア正規化では、異なる話題の言語モデル間のスコアを比較するための新しいスコア正規化法を提案した。異なる話題の言語モデルを使用する音声認識器を複数並列駆動して、その結果を選択的に使用する音声認識システムでは、話題の拡張を行えるが、各言語モデルのスコアレンジが異なるため、音声認識スコアをそのまま結果選択に使用すると選択精度が劣化する問題がある。提案法は、スコアのばらつきを、言語モデルのエントロピーにより補正することを特徴とし、従来の N -gram 言語モデルのスコアと比較して、より「真の」言語モデルの対数確率に近いスコアを算出することが可能なため、言語モデル間のスコアのばらつきを低減でき、話題選択の精度を向上できる。レストラン検索、天気予報問い合わせ、ニュース検索の 3 つの話題を識別する実験では、各音声認識器から得られる認識スコアを用いて話題を選択する方法と比較して、提案法による補正を施したスコアを用いることにより、話題選択精度を 24% (66%から 90%) 向上でき、提案法の有効性を確認した。また話題識別を含む音声認識率は、全話題に対応する単一の音声認識器と比較して、提案法による複数認識器の並列駆動型は単語正解精度を 66%から約 68%まで向上でき、提案法の効果が実験的に示された。

本論文は次のように第 1 章から第 6 章で構成される。

第 1 章では、この論文の背景や目的について述べた。

第 2 章では、第 3 章以降の研究で使用する音声認識の基本的なアルゴリズムを述べた。

2.1 節では音声認識アルゴリズムの概要を述べた。2.2 節では音響モデルとして使用される HMM について、概要やパラメータ推定法、評価法を述べた。2.3 節では言語モデルとして

使用される *N*-gram について、概要やパラメータ推定法を述べた。

第3章では、音声認識率が電話回線等の伝達経路の影響により低下する問題と、同問題に対する新しいケプストラム平均正規化法、電話音声を用いた実験及び考察について述べた。

第4章では、発声内容の違いにより音声認証の精度が劣化する問題と、同問題に対する新しい音響モデルスコアの正規化法、同手法の効果を検証した実験、考察について述べた。

第5章では、話題の変化や拡張によって音声認識率が低下する問題と、同問題に対する新しい言語モデルのスコア正規化法、レストラン情報検索、ニュース検索、天気予報検索の3つの話題に対応させた音声認識による実験と考察を述べた。

第6章では、本論文のまとめと今後の課題について述べた。