

氏名	なん じょう ひろ あき 南 條 浩 輝
学位の種類	博士 (情報学)
学位記番号	情博第 109 号
学位授与の日付	平成 16 年 3 月 23 日
学位授与の要件	学位規則第 4 条第 1 項該当
研究科・専攻	情報学研究科知能情報学専攻
学位論文題目	Automatic Transcription of Spontaneous Presentation Speech (講演を対象とした話し言葉音声の自動認識)

論文調査委員 (主査) 教授 河原達也 教授 奥乃 博 助教授 佐藤理史

論 文 内 容 の 要 旨

本論文は、講演を主な対象とした話し言葉音声の自動認識に関する研究をまとめたものである。話し言葉音声では、発話の自発性に起因する音響・発音・言語レベルの様々な変動が音声認識を困難にするので、これらの問題に対するアプローチが述べられている。また実際に、大規模な講演音声データベースを用いて、統計的なモデル化・認識手法の実装、及び実験的評価を行っている。

第 1 章は序論で、話し言葉音声認識における問題点を整理して、本論文におけるアプローチを述べ、それらを以後の章と対応づけている。

第 2 章では、本研究で用いている大規模な日本語話し言葉コーパス (CSJ)、及びベースラインの認識システムについて述べている。また、これを用いたベンチマーク実験の結果、言語的な複雑度の指標パープレキシティと、音響的な要因として発話速度が認識精度に影響することを示している。

第 3 章では、大語彙連続音声認識システムにおける認識誤り原因の自動診断手法を提案している。本手法による診断結果は音声認識プログラム (デコーダ) の改善に貢献した他、音響モデルにおいては機能語などの区間のモデル化、言語モデルにおいては間投詞やポーズのモデル化が主たる問題であることを明らかにしている。

第 4 章では、話し言葉音声認識のための言語モデルについて述べている。まず学習データとなる話し言葉の書き起こしの不足を補うために、他のテキストコーパスを効果的に混合する方法を提案している。次に、講演のような話し言葉では、文や発話の区切りが明確でないので、従来のように事前に発話を区切るのではなく、ショートポーズを自動認識しながら発話の区分化と音声認識結果の確定を並行して行う逐次デコーディング方式を導入し、その効果を示している。

第 5 章では、講演音声認識において話者間の言語表現と発音の変動に対処するために、言語モデルの教師なし適応手法について述べている。具体的には、初期的な認識結果に基づいて、(1) 類似テキストを選択して重み付けする方法、(2) 直接的に言語モデルを更新する方法を提案し、各個別手法及びそれらの組合せの有効性を示した。15 名の話者による学会講演音声からなる評価データに対して、12.1% の認識誤り削減率を達成し、31.4% の単語誤り率を得ている。

第 6 章では、話し言葉における発話速度の変動に対処するために、適応的なデコーディング方式を提案している。これは、発話速度を自動推定し、それに応じて最適な音声分析フレーム・音響モデル・デコーディングパラメータを選択的に適用するものである。実際に、発話速度に依存して音声認識誤りに明確な差があること、異なるモデル化・手法が必要なことを示した上で、適応的に適用する提案手法の有効性を確認した。前記の評価データに対して、さらに 7-8% の認識誤り削減率を達成し、22.0% の単語誤り率を得ている。

第 7 章では、本研究の成果である話し言葉音声認識技術が、講演音声の自動アーカイブ化システムに適用できることを示している。このシステムは、与えられた講演音声を自動的に書き起こした上で、意味のまとまり (セクション) ごとに音声と対応づけてアーカイブするものである。その際に、インデックスとして重要文を自動的に抽出する。自動音声認識を用い

でも、重要文抽出が十分高い精度（再現率で人間の9割以上）で行えることを示している。

第8章では結論を述べている。

論文審査の結果の要旨

本論文は、従来から非常に困難であった自然な話し言葉の音声認識の問題に対して、これまでになかったレベルの大規模な講演音声データベースを用いて本格的に取り組んだもので、得られた成果は次の通りである。

1. 大規模な統計モデルに基づく大語彙連続音声システムに対して認識誤り原因を自動的に同定する方法を考案し、これにより話し言葉音声における問題点を明らかにした。
2. 話し言葉の音声認識において大規模な学習データを用いた統計的なモデル化がきわめて重要であることを実証するとともに、効率的な言語モデルの構成法とデコーディング手法を示した。
3. 講演音声の認識において、当該の話題や話者固有の言い回し・発音変形に対して、教師なしで言語モデルを適応する方法を考案し、認識精度を有意に改善する効果を示した。
4. 話し言葉音声認識において、発話速度の変動に適応して、音声分析・音響モデル・デコーディングのパラメータを自動的に切り替える方法を考案し、認識精度を有意に改善する効果を示した。
5. 上記の手法を組み合わせることにより、実際の学会講演音声からなる評価データに対して、78%の単語認識精度を実現した。この数字は、現在報告されている中で最高水準のものである。

よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成16年2月16日実施した論文内容とそれに関連した試問の結果合格と認めた。