

Title	Automatic speaker indexing and speech recognition for panel discussions(Abstract_要旨)
Author(s)	Akita, Yuya
Citation	Kyoto University (京都大学)
Issue Date	2005-03-23
URL	http://hdl.handle.net/2433/144802
Right	
Type	Thesis or Dissertation
Textversion	none

氏名	あき た ゆう や 秋 田 祐 哉
学位(専攻分野)	博 士 (情 報 学)
学位記番号	情 博 第 146 号
学位授与の日付	平 成 17 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 Speaker Indexing and Speech Recognition for Panel Discussions (パネル討論を対象とした話者インデキシングと音声認識)
論文調査委員	(主 査) 教 授 河 原 達 也 教 授 奥 乃 博 教 授 石 田 亨

論 文 内 容 の 要 旨

本論文は、パネル討論を主な対象とした話者インデキシングと音声認識に関する研究をまとめたものである。討論や会議は、音声を主なメディアとして意見交換や問題解決を行うもので、通常議事録が作成され、またデジタルアーカイブとしての保存も行われつつあるが、それらの効率的な構築や利用のために、発話毎の話者情報の自動付与(=話者インデキシング)や自動書き起こし(=音声認識)の技術は非常に有用であると考えられる。

第1章は序論で、本研究の位置づけ、特に話し言葉に関する音声認識研究の概観を行った上で、討論・会議音声を扱う上での問題点を整理し、本研究のアプローチを述べている。

第2章では、本研究で対象とするテレビ討論・国会の音声コーパス、及びベースラインとなる認識システムを記述した上で、これらのデータに基づく問題点の分析を行っている。

第3章では、教師なし話者インデキシングの手法を提案し、それに基づく音声認識の話者適応も検討している。提案法は、事前学習を一切必要としないものであるが、発話長が全般に短く不均一な討論音声に対処するために、大規模音声データベースから構築した多数話者モデルとの類似度を用いている。話者分離能力を向上させるための種々の統計的手法を組み合わせることで、最終的に97.1%の話者インデキシング精度を実現している。また、このインデキシング結果に基づいて音声認識の音響モデルを話者適応することができ、それにより単語認識精度が3.8%向上した。

第4章では、音声認識のための統計的言語モデルを、討論の話題と話者毎の表現様式に適應する方式を提案している。多様な話題と表現様式をそれぞれカバーするコーパスに対して、確率的潜在意味解析(PLSA)を行い、話題と表現様式を表す部分空間を構成する。そして、初期の(話者ごとの)音声認識結果をこの部分空間に射影して、当該の話題と表現様式に適應した言語確率を求める。本方式により、言語モデルの能力を示す尺度であるパープレキシティが8.5%改善され、音声認識精度においても有意な向上が得られた。

第5章では、話し言葉音声認識のための発音変動の汎用的なモデル化を提案している。大規模な話し言葉音声コーパスを用いて、標準形な読み(baseform)と実際の発音変動形(surface form)の対応を可変長の音素系列レベルで統計的に学習し、これに基づいて任意の語彙に対して、可能な発音変動形を確率付きで予測するモデルを構成している。本手法を(学習コーパスとは異なるドメインの)討論及び国会の音声認識用単語辞書に適用し、音声認識精度の有意な改善を実現している。

第6章では、それまでの章で提案した手法を用いて、パネル討論のアーカイブ化を行うシステムについて述べている。各発話から重要文を抽出したり、発話タイプ(=談話タグ)を推定する手法についても検討し、話者インデックスや書き起こしなどととも、MPEG-7の枠組みでエンコードするシステムを設計している。

第7章では、最終的な音声認識精度などについてまとめた上で、結論を述べている。

論文審査の結果の要旨

本論文は、パネル討論音声を対象とした話者インデキシングと音声認識に関して、初めて本格的に行われた研究をまとめたものであり、得られた主な成果は次の通りである。

1. 討論音声から（事前学習を必要としないで当該データのみから）教師なしで話者インデキシングを行う手法を実現した。従来手法が、話者毎の発話長が短く不均一な討論音声では十分に機能しなかったのに対して、本手法により97%を上回る話者インデキシング精度を実現した。これにより、音声認識用の音響モデルの話者適応も可能になった。
2. 音声認識のための統計的言語モデルを、討論の話題と話者毎の表現様式に適応する方式を考案し、その有効性を示した。提案方式は、確率的潜在意味解析に基づいて、当該討論の話題や話者のモデルを確率的に写像するもので、頑健性に優れていると考えられる。実際に、話題だけでなく表現様式についても適応する効果を示している。
3. 話し言葉音声認識のための発音変動の汎用的なモデル化を考案し、その有用性を示した。提案手法は、音素系列のレベルで発音の変動を確率的に予測するものであり、任意の語彙に適用できる点で汎用性に優れている。実際に、討論や国会の複数の音声認識タスクで評価を行い、認識精度の有意な改善を実現している。

以上のように本論文は、人間どうしの音声コミュニケーションから話者情報や発話内容を自動的に認識するための基盤となる技術を提供するもので、学術上・實際上寄与するところが少なくない。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17年2月17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。