

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Robust Background Models for Speaker Verification
著者(和文)	ビ ススサンギ ータ
Author(English)	SANGEETA BISWAS
出典(和文)	学位:博士（学術）, 学位授与機関:東京工業大学, 報告番号:甲第10048号, 授与年月日:2016年1月31日, 学位の種別:課程博士, 審査員:篠田 浩一,亀井 宏行,徳永 健伸,藤井 敦,石田 貴士
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第10048号, Conferred date:2016/1/31, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

(2000字程度)

報告番号	乙 第 号	学位申請者	ビスワス サンギータ BISWAS Sangeeta	
論文審査員	氏 名	職 名	氏 名	職 名
	主査 篠田 浩一	教授	石田 貴士	准教授
	徳永 健伸	教授		
	亀井 宏行	教授		
	藤井 敦	准教授		

本論文は、「Robust Background Models for Speaker Verification (話者同定のための頑健な背景モデル)」と題し、英文6章より構成されている。

第1章「Introduction」では、本研究の背景と目的について述べている。まず、現代社会におけるセキュリティ強化のために、音声により個人認証を行う自動話者同定システムが必要とされていることを述べている。そして、その性能向上のために、話者ごとの発声の特徴における差異や不適切なデータの混入に対して頑健な話者同定手法を開発することを目的とすることを述べ、最後に本論文の構成を示している。

第2章「Overview of Automatic Speaker Verification System」では、現在の自動話者同定システムとその要素技術について述べている。話者モデルとして混合ガウス分布 (GMM)、識別器としてサポートベクターマシン (SVM) がしばしば用いられることを述べた上で、高性能な話者モデルを得るためには、まず、多数話者の発話を用いて平均的な話者の特性を表す背景モデルを作成し、それを初期モデルとして当該話者の少量の登録発声を用いてその話者の特徴に適応したGMMを構築する手法が有効であると述べている。さらに、近年、GMMのパラメータを因子分析により圧縮したi-Vectorを用いる方法が同定率の向上に貢献していることを述べている。

第3章「Previous Work」では、自動話者同定の頑健性を高めるための従来手法を概観し、残された課題について述べている。特徴量や同定スコアの正規化手法、話者モデルのパラメータ変換手法、確率的線形識別分析手法、及び、各種の背景モデル構築手法について述べた上で、登録発声が少量の場合には性能がまだ不十分な話者が存在すること、背景モデルがその学習データにおける不適切なデータの存在の影響を受けること、の2点を課題として挙げている。

第4章「Robustness Against Enrollment Data Variants」では、話者の少量の登録発声を用いてモデルを当該話者に適応させることで話者同定の性能を高める技術を提案している。GMMの木構造を用いる構造的事後確率最大化 (SMAP) 法において、多数のGMM木構造からなるAcoustic Forestを構成し、個々の木構造から得られたモデルを統合して用いる。米国国立標準技術研究所 (NIST) が構築したSREデータベースを用いて評価し、従来手法に比べて話者同定における等誤り率 (equal error rate) が1.5%減少したことを述べている。

第5章「Robustness Against Irrelevant Training Data」では、背景モデルを構築するための多数話者の発話から、背景ノイズが多い、回線が適合していない、などの理由による不適切な発話を取り除くことで、頑健な背景モデルを構築する手法を提案している。特に、発話特徴の分布において近傍数を自動決定する flexible k-最近傍法 (fk-NN) を用いた発話クラスタリングを利用する方法を提案している。NIST SREデータベースにおける収録条件の異なるデータを用いて評価した結果、近傍数が適切に選択されていること、従来手法に比べて、頑健性が高いことを確認したことを述べている。

第6章「Conclusion and Future Work」では、本論文の結論及び貢献について総括し、今後の課題についてまとめている。

以上で述べたように、本論文では、自動話者同定技術において、話者特徴の分布に関する知見を活用した新しい手法を提案し、登録音声が少ない場合や学習データ中に不適切なデータがある場合に有効であることを大規模なデータセットで実証している。この成果は、話者同定システムの性能向上に有効であると同時に、機械学習を用いた同定技術全般のための基盤技術を提供するものであり、学術上、寄与するところが大きい。よって本論文は博士 (学術) の学位論文として十分価値があるものと認める。

注意：「論文審査の要旨及び審査員」は、東工大リサーチリポジトリ (T2R2) にてインターネット公表されますので、公表可能な範囲の内容で作成してください。