T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	A Study on Acoustic Modeling of Speech for Personalized Speech Interface
著者(和文)	井島勇祐
Author(English)	Yusuke Ijima
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第9881号, 授与年月日:2015年3月26日, 学位の種別:課程博士, 審査員:小林　隆夫,伊東　利哉,小池　康晴,杉野　暢彦,篠崎　隆宏
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第9881号, Conferred date:2015/3/26, 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

Powered by T2R2 (Science Tokyo Research Repository)

（博士課程）

論文審査の要旨及び審査員

報告番号	甲第　　号		学位申請者氏名		井島　勇祐	
論文審査 審査員		氏　名	職　名		氏　名	職　名
	主査	小林　隆夫	教授	審査員	篠崎　隆宏	准教授
	審査員	伊東　利哉	教授			
		小池　康晴	教授			
		杉野　暢彦	准教授			

論文審査の要旨（2000字程度）

本論文は「A Study on Acoustic Modeling of Speech for Personalized Speech Interface(音声インタフェースのパーソナライズ化のための音響モデリングに関する研究)」と題し，英文６章からなっている。

第１章「Introduction（序論）」では，本研究の背景として，音声インタフェースを用いたアプリケーションの現状について述べ，スマートフォン等のパーソナルデバイスの進化に伴い音声インタフェースにもパーソナライズ化が求められており，本研究ではその基盤となる音声認識と音声合成においてパーソナライズ化を可能とするための新たな音響モデリング手法を提案すると述べている。

第２章「Emotional Speech Recognition Based on Speaker Dependent Multiple-Regression HMM（話者依存重回帰HMMに基づく感情音声認識）」では，話者固有の発話様式・感情表現（スタイル）の変動に頑健な音声認識手法を述べている。ユーザの発話は文章あるいはフレーズ単位でスタイルが変動するため，音声認識システムが持っている音響モデルとのミスマッチが起こり，認識性能が低下する原因となる。そこで本手法では，重回帰隠れマルコフモデル（重回帰HMM）を音響モデルとして用い，入力１発話毎に入力音声のスタイルを表す低次元の説明変数であるスタイルベクトルを推定し，これに基づいて認識用音響モデルを入力音声のスタイルとマッチするように適応している。音素認識実験の結果，本手法は通常のHMMと比較して音素誤り率を約11%削減できることを示し，さらにスタイルの表出度合いの推定値が得られることも明らかにしている。

第３章「Emotional Speech Recognition Based on Speaker Adapted Multiple-Regression HMM（話者適応重回帰HMMに基づく感情音声認識）」では，第２章で提案した特定話者向け手法を任意話者用に拡張した音声認識手法を述べている。本手法では，あらかじめ多数の話者の音声データから学習した不特定話者モデルに対し，話者適応とスタイル適応を組み合せることにより，対象とする話者（目標話者）の各スタイル音声が数発話程度と少量の音声データのみの場合でも重回帰HMMの学習を可能としている。プロナレータ及び一般の話者の音声データを用いた音声認識実験の結果，提案法は通常のHMMの場合より高い音声認識性能が得られることを示している。

第４章「Average Voice Model Training Technique Based on Speaker Class（話者クラスに基づく平均声モデル学習法）」では，平均声に基づく任意話者の音声合成手法における合成音声の話者類似性向上を目的として，話者クラスに基づいた平均声モデルの学習法を述べている。本手法では，まず平均声モデルの学習に用いる全話者に対し，話者性に基づいた話者クラスタリングを行い，各話者に対して話者クラスを付与することで，複数の話者性を考慮した平均声モデルを学習する。そして，モデル適応時には，目標話者の話者クラスを推定し，モデル適応，音声パラメータ生成を行っている。客観評価実験の結果，従来法と比較してパラメータ生成誤差が減少すること，また，主観評価実験より合成音声の類似性，自然性が向上することを示している。

第５章「Similar Speaker Selection Technique Based on Distance Metric Learning（距離学習に基づく類似話者選択法）では，第４章で提案したモデル学習法の高度化を目的として，知覚的な声質類似度と音響特徴量との相関分析を行い，距離学習に基づく知覚的な類似話者選択手法を述べている。まず，多数話者の音声データを用いた主観評価実験を通して各話者間の知覚的な声質類似度を調べ，声質類似度と音響特徴量との相関分析の結果，4 kHz以下のケプストラム，非周期性指標等と高い相関があることを明らかにしている。そして，距離学習に基づく類似話者選択手法では，音響特徴量空間から声質類似性を考慮した空間へ変換するための変換行列を学習し，これを用いて入力音声の音響特徴量を変換することで，知覚的な声質類似度を考慮した話者選択を行っている。類似話者選択実験の結果，提案法では話者選択誤り率を５割程度削減できることを示している。

第６章「Conclusions and Future Work(結論と今後の課題)」では，本論文の成果をまとめ，今後の課題を述べている。

以上を要するに，本論文では音声インタフェースのパーソナライズ化に向けて，その基盤となる音声認識及び音声合成に用いられる音響モデルの新たな学習手法を提案し，従来手法との比較評価実験を通してその有効性と可能性を明らかにしたものであり，工学上ならびに工業上貢献するところが大きい。よって本論文は博士（工学）の学位論文として価値あるものと認められる。

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。