

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)	論文要旨
Type(English)	Summary

論文要旨

THESIS SUMMARY

専攻： Department of	物理情報システム	専攻	申請学位 (専攻分野)： 博士 Academic Degree Requested	Doctor of	(工学)
学生氏名： Student's Name	井島 勇祐		指導教員 (主)： Academic Advisor(main)	小林 隆夫	
			指導教員 (副)： Academic Advisor(sub)		

要旨 (和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

本論文では、音声インタフェース (音声認識、テキスト音声合成) のパーソナライズ化を目的とした音声の音響モデリング手法について述べている。

第1章「Introduction (序論)」では、本研究の背景について述べている。まず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よる音声インタフェースに求められる役割の変化について述べている。次に、音声インタフェースのパーソナライズ化における従来法の概要とその課題について述べている。最後に、本研究の概要について述べている。

第2章「Emotional Speech Recognition Based on Speaker Dependent Multiple-Regression HMM (特定話者重回帰 HMM に基づく感情音声認識)」では、音声認識結果だけでなく発話者の発話様式・感情表現 (スタイル) を得ることが可能なモデル適応手法について述べている。本手法では、重回帰隠れマルコフモデル (重回帰 HMM) の説明変数として、スタイルの表出度合いを表す低次元のベクトルであるスタイルベクトルを用いる。まず入力音声に対し、スタイル推定手法によりスタイルベクトルを推定する。その後、推定したスタイルベクトルと重回帰 HMM とを用いて HMM の出力確率分布の平均ベクトルを求めることで、入力音声のスタイルに対してモデルを適応する。音素認識実験の結果、本手法は通常の HMM と比較し、音素誤り率を約 11%削減できることを示した。また、本手法では音声認識結果だけでなく、話者のスタイルの表出度合いも得ることが可能であることも示した。

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、目標話者の各スタイル数発話程度の少量の音声データから学習した話者適応重回帰 HMM を用いた音声認識手法について述べている。第2章で提案した手法では、目標話者の重回帰 HMM を学習する際に、目標話者が発話した各スタイルの十分な量の音声データが必要のため、本手法を任意の話者へ適用することは現実的ではなかった。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を用いることで、目標話者の各スタイル数発話程度の少量の音声データから重回帰 HMM の学習を行う。音声認識実験の結果、提案法は通常の HMM と比較し、高い音声認識性能が得られることを示した。

第4章「Average Voice Model Training Technique Based on Speaker Class (話者クラスを用いた平均声モデルの学習)」では、平均声モデルに基づく任意話者の音声合成手法における合成音声の類似性向上を目的として、平均声モデル学習に用いる話者の話者性を表現する「話者クラス」を用いた平均モデルの学習法について述べている。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て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。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う。客観評価実験、主観評価実験より、従来法と比較し、合成音声の類似性、自然性が向上することを示した。

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)」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と、距離学習に基づく類似話者選択手法について述べている。まず多数の話者の音声データを用いた大規模な主観評価実験を行い、得られた各話者間の声質類似度と様々な音響特徴量との相関分析を行った。その結果、4 kHz 以下のケプストラム、非周期性指標等が高い相関が得られた。距離学習に基づく類似話者選択手法では、まず主観評価実験により得られた声質類似度と音響特徴量とを用いて、音響特徴量空間を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する。類似話者選択時には、入力音声の音響特徴量を学習した変換行列を用いて変換することで、知覚的な声質類似度を考慮した話者選択を行う。類似話者選択実験の結果、提案法では話者選択誤り率を 53.9%削減できることを示した。

第6章「Conclusions and Future Works (結論と今後の課題)」では、本論文の結論と今後の課題について述べている。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、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。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

(博士課程)
Doctoral Program

論文要旨

THESIS SUMMARY

専攻 : Department of	物理情報システム	専攻	申請学位 (専攻分野) : Academic Degree Requested	博士 Doctor of	(工学)
学生氏名 : Student's Name	井島 勇祐		指導教員 (主) : Academic Advisor(main)	小林 隆夫	
			指導教員 (副) : Academic Advisor(sub)		

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words)

This thesis presents novel approaches to acoustic modeling of speech to achieve a personalized speech interface: automatic speech recognition (ASR) and text-to-speech synthesis (TTS).

Chapter 1 describes general background of the thesis.

Chapter 2 describes a rapid model adaptation technique for emotional speech recognition. This technique utilizes the MRHMM framework for the model adaptation and the style vector which corresponds to the degree or intensity of expressivity of styles. In the recognition stage, the HMM is adapted to the input style using the estimated style vector. It has been confirmed that the proposed technique reduces the error rate by 11% of the style-independent HMM. Furthermore, the technique can obtain not only linguistic information but also the degree of expressivity of emotional speech.

Chapter 3 presents an emotional speech recognition technique based on speaker adapted MRHMM. Using a speaker-independent neutral style model, the MRHMM is trained with a small amount of target speaker's data. The experimental results indicated that the performance of the proposed technique in both speech recognition and style estimation is promising for simulated emotional speech.

Chapter 4 describes an average voice model training technique that utilizes speaker classes representing the voice characteristics of speakers. In the speaker adaptation process, the speaker class of the target speaker is estimated and used for speaker adaptation and speech parameter generation. Objective and subjective experiments showed that the proposed technique can synthesize speech that is closer to that of the target speaker than the conventional method.

Chapter 5 presents a similar speaker selection technique as the first step for further improving the similarity of the synthesized speech. The technique first trains a transform matrix based on distance metric learning using the perceptual voice quality similarity. Given an input speech, acoustic features of the input speech are transformed using the trained transform matrix. The results indicate that the proposed technique reduces the speaker selection error rate by about 53.9%.

Chapter 6 describes conclusions and future works of this thesis.

備考 : 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意 : 論文要旨は、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。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).