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)	残響および騒音のある環境下で頑健な音声認識
Title(English)	Robust Speech Recognition under Reverberant and Noisy Environments
著者(和文)	太刀岡勇気
Author(English)	Yuki Tachioka
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10885号, 授与年月日:2018年3月26日, 学位の種別:課程博士, 審査員:小林 隆夫,奥村 学,山口 雅浩,金子 寛彦,篠﨑 隆宏
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10885号, Conferred date:2018/3/26, 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

Powered by T2R2 (Science Tokyo Research Repository)

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報通信 系 情報通信 コース	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of （ 工学 ）
学生氏名： Student's Name	太刀岡 勇気	指導教員（主）： Academic Supervisor(main)	小林隆夫教授
		指導教員（副）： Academic Supervisor(sub)	

要旨（和文 2000 字程度）
Thesis Summary（approx.2000 Japanese Characters ）

本論文では音声認識の適用先拡大のため、騒音・残響がある環境で音声認識性能を向上させることを目的としている。実環境で認識性能が低下する原因としては、騒音と残響がある。音声認識を行うためには、騒音を除去して音声を強調する技術が必要である。そのためには、複数マイクを使って話者の位置を特定し、マイク間の位相差を手がかりに、音声を強調する。第 2 章では、従来法を紹介したのちに、音声区間の情報に基づき、音源の方向に関する事前分布を形成し、音源方向の推定精度を向上させる方法を提案した。また、マイクロフォンの配置誤差・残響による理論的なマイク間の到来時間差からのずれを、補正する方法を提案した。

3 章では、音声認識の前処理として重要な音声強調手法に関して詳述した。1 マイクを使う方法として、残響の程度を表す残響時間を自動的に推定し、残響を除去する方法を提案した。一方、複数のマイクを使う方法として、到来方向に基づくバイナリマスクと独立ベクトル分析を統合する方法を提案した。加えて音声強調法として有効だが初期値依存性が高いことで知られるマルチチャンネル非負値行列因子分解に対して、性能を安定させる初期値を与える方法を示した。これらの成果により、音声から騒音と残響を取り除くことができるようになり、音声認識の性能が大幅に向上することを、実験的に確かめた。音声強調の次に重要なのは、正確に音声区間を切り出してくる技術である。音声区間を取り損なえば音声認識をすることができず、騒音区間が多く混入してもこれもまた音声認識性能を下げる原因となるためである。ここでは、音声区間検出のモデルパラメータ推定に密度比推定を使うことで、従来、音声と騒音の 2 つのモデルが必要だったところを 1 つの密度比モデルに減らすことができ、頑健性が向上した。また学習環境と評価環境でミスマッチがあると、性能が低下する。本論文では、録音レベルが不適切な場合でクリッピングが起こった場合と、サンプリング周波数が学習データと評価データで異なる場合に、どの程度音声認識の性能が低下するかを検討した。サンプリング周波数のずれに関しては、深層神経回路網を用いて評価データの音声の帯域拡張を行うことで、音声認識の性能低下を抑えられる方法を提案した。

4 章では、音声認識を騒音・残響に対して頑健にする方法に関して述べている。2 章・3 章の手法により、騒音を抑圧したとはいえ、やはり騒音の影響は残っているため、騒音に頑健な音声認識システムを構築する必要がある。この際に有効なのは、音声認識の誤り率に関連する目的関数を最小化する誤り訂正学習の一種である識別学習である。これにより、音声認識の誤りを訂正するように種々のモデルを再学習する。本論文では、特徴量の線形判別分析、音声認識のモデルとして主要な 2 つのモデルである音響モデルと言語モデルに識別学習を導入し、騒音環境下で音声認識性能が向上することを確認した。併せて、音響モデルのモデル削減法と識別学習を組み合わせる方法を提案した。また、識別学習の目的関数を拡張し、複数のシステムを組み合わせた際に有効である、元のシステムと異なる仮説を出力するような補助システムを構築する枠組みを示した。3 章などで述べた音声強調は、大抵の場合音声認識性能を向上させるが、音声強調を行ったことで音声にひずみが混入し特徴量の分布が変わってしまい、性能が低下してしまうことがある。これに対処するため、起こりうるひずみを想定して学習することで、ひずみによる影響を低減する方法を提案した。

最終章である 5 章では、騒音・残響に頑健な音声認識タスクを対象とする様々なチャレンジにより、提案手法の有効性を確認した。主に家庭環境での騒音を対象とした第 2 回 CHiME チャレンジでは、識別的手法の有効性を示した。主に屋外環境の騒音を対象とした第 3 回 CHiME チャレンジでは、騒音環境が大きく異なるため複数のシステムを用意すると最良のシステムが環境ごとに異なる。この中から最良のシステムを話者認識に用いる i-vector 特徴量を用いて選択する手法を提案した。第 4 回 CHiME チャレンジは第 3 回の翌年のリピートであるが、種々の特徴量・システム、先端的な音声強調手法を用いて、第 3 回 CHiME チャレンジと同一のタスクで音声認識誤りを半減した。種々の残響環境を対象とした REVERB チャレンジでは、上述の残響除去法と識別学習を用いたシステム統合法の有効性を確認した。このうち、第 2 回 CHiME チャレンジ、REVERB チャレンジでは、参加者中トップの成績を収め、提案法の有効性が確認された。DIRHA チャレンジは、音源位置推定と音声区間検出を同時に行うタスクである。ここでは、音源位置推定法と音声区間検出のコストを統合することで両者の性能を向上させられることが分かった。

このように本論文では、騒音・残響がある環境で音声認識性能を向上させることを目的として、音源の位置推定法、音声強調法(騒音抑圧・残響除去)、音声認識の各種モデルの頑健性を向上させる識別的手法を提案し、公開チャレンジにより提案法の有効性を確認した。

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

系・コース： Department of, Graduate major in	情報通信 系 情報通信 コース	申請学位（専攻分野）： Academic Degree Requested	博士 (工学) Doctor of
学生氏名： Student's Name	太刀岡 勇気	指導教員（主）： Academic Supervisor(main)	小林隆夫教授
		指導教員（副）： Academic Supervisor(sub)	

要旨（英文 300 語程度）
Thesis Summary（approx.300 English Words ）

This paper aims to improve the automatic speech recognition (ASR) performance under noisy and reverberant environments in order to widen the applications of ASR. Noise reduction is essential, but, before noise reduction, it is necessary to detect speaker positions. Chapter 2 proposes prior distributions of a speaker direction to achieve more accurate source localization.

Chapter 3 describes front-end techniques prior to ASR. There are two types of speech enhancement: dereverberation and noise reduction. First, dereverberation method using single microphone is proposed. This method can estimate a reverberation time and eliminate reverberation from speech. Second, noise reduction methods using multi microphones are proposed: combination of binary masking based on the estimated source direction with independent vector analysis and initial value settings of multi-channel non-negative matrix factorization. After speech enhancement, it is important to detect speech activity accurately from noisy speech. Density ratio estimation is applied to model parameter estimation, which can reduce the number of models from two (speech and noise models) to one (density ratio model) and improve the robustness against noise.

Chapter 4 describes back-end techniques. The main target is discriminative training methods, which retrain ASR models to minimize their objective functions related to the number of ASR errors. This chapter applies discriminative training methods to linear discriminant analysis of acoustic features and two ASR models (acoustic model and language model). In addition, by extending their objective function, this chapter proposes to construct complementary systems that output hypotheses different from original system.

The final chapter validates the effectiveness of the proposed method on various challenges of noisy and reverberant ASR. On the second CHiME challenge, discriminative methods are effective for noisy ASR. On the third CHiME challenge, it is effective to select the best system that is suitable for the target environment from multiple systems. On the REVERB challenge, the above-mentioned dereverberation method and discriminative method for system combination are effective. For the second CHiME challenge and REVERB challenge, our team achieved the best performance.

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).