

博士学位論文審査要旨

2008年2月12日

論文題目: Studies in Signal Processing for Robust Speech Recognition in Noisy and Reverberant Environments

(雑音や残響のある環境下での頑健な音声認識のための信号処理技術に関する研究)

学位申請者: 大田 健紘

審査委員:

主査: 工学研究科

教授 柳田 益造

副査: リール中央学院 自動制御情報工学研究所 教授 Philippe VANHEEGHE

副査: 大阪工業大学大学院 情報科学研究科 教授 Mehdi NOURI-SHIRAZI

要 旨:

音声認識技術の向上により、比較的雑音の少ない環境において接話マイクロホンによる音声認識は、現在、ほぼ実用的なレベルに達している。しかし、接話マイクロホンを用いる場合は、ヘッドセットマイクなどを常に身に付けていなければならない、非常にわずらわしい。そのため、ユーザがマイクロホンに気を配ることなくシステムを利用できるようにすることが望まれる。つまり、音声認識システムの実用化には、マイクロホンから離れた位置からのユーザ発話でも正しく認識できる必要がある。しかし、話者から離れたところにマイクロホンを設置すると、認識対象音声のパワーが小さくなり、周囲の雑音や壁や天井からの反射音の影響により音声認識率は大幅に低下する。また、操作する対象の機器(例えばTV)自体が音を発する場合はその影響も受ける。それゆえ、実用的なシステムを構築するためには、雑音除去技術や残響除去技術の導入が不可欠である。

本論文は、実環境において頑健に動作する音声認識のための信号処理技術について述べている。本論文は以下の6章から構成されている。

第1章は本論文の導入であり、これまでの音声認識研究および、音声認識率改善のための信号処理技術について述べている。

第2章は、音声により操作する対象機器自体が発する音を除去する技術について述べている。ここでは、従来の適応ノイズキャンセラでは用いられていなかった音声の調波構造を考慮した雑音除去法を開発することによりこの問題を解決する手法を提案している。

第3章は、反射の周波数特性を考慮した反射音除去技術について述べている。ここでは単一マイクロホンによる各周波数ビンでの自己相関関数を基に、残響時間を周波数の関数として推定する手法を提案している。これにより壁や天井での反射の周波数特性を考慮した反射音除去処理が可能になり、過剰な減算を抑制することができる。

第4、5章では今後音声認識に期待される雑音除去技術を提案している。

第4章は、周波数領域での独立成分分析(ICA)によるブラインド音源分離(BSS)で問題となるパーミューテーション問題を抑制するICAの提案を行っている。

第5章は、粒子フィルタを用いて単一マイクロホンで音声のスペクトルを推定する手法につい

て述べている。

第6章は、本論文のまとめおよび今後の課題について述べている。

よって、本論文は、博士（工学）（同志社大学）の学位論文として十分な価値を有するものと認められる。

総合試験結果の要旨

2008年2月12日

論文題目: Studies in Signal Processing for Robust Speech Recognition in Noisy and Reverberant Environments

(雑音や残響のある環境下での頑健な音声認識のための信号処理技術に関する研究)

学位申請者: 大田 健紘

審査委員:

主査: 工学研究科

教授 柳田 益造

副査: リール中央学院 自動制御情報工学研究所 教授 Philippe VANHEEGHE

副査: 大阪工業大学大学院 情報科学研究科 教授 Mehdi NOURI-SHIRAZI

要 旨:

本論文提出者は、2005年4月より本学大学院工学研究科知識工学専攻博士課程後期課程に在学しており、各年度において優れた研究成果を挙げ、英語の語学試験に合格し、1年間のフランス留学の経験からフランス語も実用レベルに達していると認定する。

本論文の主たる内容は、電子情報通信学会論文誌(1編)および Journal of Studies in Informatics and Control (1編)、その他フル査読付き国際会議の会議録(3編)に、いずれも第1著者として掲載され、すでに十分な評価を受けており、さらにこの内容によって、論文提出者は電気関係学会関西支部連合大会奨励賞を受賞している。

本年1月19日午後1時より約1時間半にわたり、提出論文に関する学術講演会(博士論文公聴会)が開催され、種々の質疑応答が行われたが、論文提出者の説明により、いずれも十分な理解が得られた。この公聴会は、同志社大学とリール中央学院(フランス)のダブルディグリーに関する協定に基づき、リール中央学院の担当教授の出席の下で英語によって行われた。

さらに、講演会終了後、審査委員により論文に関連した諸問題につき口頭試問を行った結果、十分な学力と見識を有すると認められた。

よって、総合試験の結果は合格であると認める。

博士學位論文要旨

論文題目: Studies in Signal Processing for Robust Speech Recognition in Noisy and Reverberant Environments

氏 名: 大田 健紘

要 旨:

音声認識技術の向上により、比較的雑音の少ない環境において接話マイクロホンによる音声認識は、現在、ほぼ実用的なレベルに達している。しかし、接話マイクロホンを用いる場合は、ヘッドセットマイクなどを常に身に付けていなければならない、非常にわずらわしい。そのため、ユーザがマイクロホンに気を配ることなくシステムを利用できるようにする必要がある。つまり、音声認識システムの実用化には、マイクロホンから離れた位置からのユーザ発話でも正しく認識できる必要がある。しかし、話者から離れたところにマイクロホンを設置すると、認識対象音声のパワーが小さくなり、周囲の雑音や壁や天井からの反射音の影響により音声認識率は大幅に低下する。また、操作する対象の機器(例えばTV)自体が音を発する場合はその影響も受ける。それゆえ、実用的なシステムを構築するためには、雑音除去技術や残響除去技術の導入が不可欠である。

本論文では、実環境において頑健に動作する音声認識のための信号処理技術について述べている。本論文は以下の6章から構成されている。

第1章は本論文の導入であり、これまでの音声認識研究および、音声認識率改善のための信号処理技術について述べている。

第2章は、音声により操作する対象機器自体が発する音を除去する技術について述べている。TVのように、操作対象自体が発する機器を音声によって操作する場合、機器自体が発する音が雑音となり音声認識率が低下する。このような問題に対しては、操作対象自体が発する音は既知という条件の下で、適応ノイズキャンセラを用いることができるが、話者の発話中においてもTV-マイクロホン間の伝達特性が人の動きや室温、湿度の変化などにより変動する。また、TV音の場合、壁などからの反射音とシステム利用者の音声は相関が高いため、従来の適応ノイズキャンセラでは十分な雑音除去性能が得られないという事情もある。そこで、従来の適応ノイズキャンセラでは用いられていない音声の調波構造を考慮した雑音除去法を開発することによりこの問題を解決する手法を提案している。

提案法では、入力信号をその振幅により認識対象音声が含まれる区間であるか、そうでないかを判定する。そして認識対象音声が含まれる区間はさらにスペクトルの特徴から有声音であるか、無声音であるかを判定する。有声音の区間に関しては基本周波数を基に調波構造を推定し、推定した調波構造を入力信号から減算することによって、雑音信号を推定する。そして、この雑音の推定値と既知としている操作対象自体が発する音を用いて伝達特性の学習が行われる。無声音の区間に関しては、調波構造を用いることができないため、前フレームで推定した伝達特性を流用し、雑音信号の推定を行う。

提案法は、SN比が0~9dBの環境において、従来の単純なスペクトルサブトラクション法やWienerフィルタと音声認識率を比較しても十分よいもしくは同程度の性能を持つことを示している。ただし、SN比が12dB以上の良い音響環境での音声認識率は、改善の余地が無く、事実改善が認められない。

第3章は、反射の周波数特性を考慮した反射音除去技術について述べている。第2章の雑音除去法では、雑音源とマイクロホン間の伝達特性を考慮しているが、話者とマイクロホン間の伝達特性を考慮していない。従来の反射音除去処理では、壁や天井での反射の周波数特性が平坦であることを仮定した複数のマイクロホンによる手法が用いられていた。しかし、これらの手法では、残響時間が真の値よりも長く推定されることにより過剰な減算が起こる可能性があった。そこで、この問題を解決するために、ここでは単一マイクロホンによる各周波数ビンでの自己相関関数を基に、残響時間を周波数の関数として推定する手法を提案している。これにより壁や天井での反射の周波数特性を考慮した反射音除去処理が可能になり、過剰な減算を抑制することができる。シミュレーションおよび実際に収録した音声を用いて評価を行ったところ、過剰な減算が抑制され、SN比が改善することが明らかになっている。また付録Bに示す音声操作によるTVシステムを用いて、音声認識率の評価を行ったところ、SN比が0~12dBの範囲で認識率の改善が得られ、マイクロホンから50~100cm離れた位置からの音声認識が可能であることが明らかになった。

第2章で述べた雑音除去技術を実現するためには、雑音のみの信号が得られなければならない。また、逐次的な手法で精度良く雑音を除去するためには、伝達特性の学習のために多くの時間を要する。そのため、システムは計算量を減らさない限り連続的に音声認識をすることが困難である。しかし、逐次計算の繰り返し回数を制限するなど計算回数を減らすと音声認識率が低下する。以上のことより、第4、5章では今後音声認識に期待される雑音除去技術を提案している。

第4章は、周波数領域での独立成分分析(ICA)によるブレンインド音源分離(BSS)で問題となるパーミューテーション問題を抑制するICAの提案を行っている。BSSは音源に関する事前情報が無い状況で、複数の音源からの音を波形レベルで分離する技術である。BSSの手法の一つとして、ICAに基づく手法が近年盛んに研究されており、瞬時(畳み込みでない)混合に対しては十分な分離性能が得られている。また、畳み込み混合に対処する方法としては周波数領域ICAがある。これは信号を短時間フーリエ変換により周波数領域に変換することによって、周波数ビン毎の瞬時混合としてICAを適用するものである。しかし、BSSに周波数領域ICAを適用した場合、各チャンネルの周波数ビンの入れ違いを正しく並び替える(パーミューテーション問題を解決する)必要がある。しかし、いまだパーミューテーション問題の完全な解決には至っていない。Di Persiaらは、これまでに全周波数ビンを連結して1本のベクトルと考えて分離を行うパーミューション・フリーICA(PF-ICA)を提案している。この手法では、分離された信号の周波数ビン間の入れ替わりが生じない利点があるが、分離行列の周波数特性を考慮することができないという欠点があった。そこで、筆者は、対象とする周波数ビンの前後数本の周波数ビンを連結することによって、当該周波数ビンに関する分離行列を安定して求めるマルチビンICA(MB-ICA)を提案し、分離行列の周波数特性を考慮した分離ができるようにした。

第5章は、粒子フィルタを用いて単一マイクロホンで音声のスペクトルを推定する手法について述べている。システムの実用化を考えると、少数のマイクロホンでシステムを実現することが望ましい。それゆえ、単一マイクロホンで雑音除去、および残響除去を行う技術の研究が要請される。単一マイクロホンによって、粒子フィルタを用いた雑音のスペクトルの推定と、ガウス混合モデル(GMM)を用いた音声のスペクトルの推定法が開発されている。粒子フィルタは近年、計算機性能の向上により普及したものであり、従来のカルマンフィルタが線形・ガウス型の状態空間にしか適用できなかったものを、非線形・非ガウス型の状態空間にも適用できるように拡張したフィルタである。しかし、GMMを用いた手法は音声信号の追従性能がGMMの学習精度に

依存する。一方、音声のスペクトルのモデル化に GMM ではなくディリクレ混合過程 (DPM) に基づくモデルを用いると、混合分布数をデータから自動的に得ることができるため、予めモデルを学習する必要がなくなり、より柔軟なモデル化が可能になる。DPM とは、ディリクレ過程により生成された確率分布の混合モデルであり、ディリクレ過程は混合する各要素の確率分布の確率分布を表現するものであり、無限個の確率分布を生成するためのモデルである。また、状態空間モデルに話者とマイクロホン間の伝達特性の影響を導入し、雑音除去とともに残響抑圧も行う手法も提案している。実環境音声・音響データベースに収録されている 3 種類の雑音データ、および 4 種類の SN 比 (0, 3, 6, 9dB) を用いて、DPM に基づくモデルを用いた雑音除去法の評価を行っている。評価の結果、DPM に基づくモデルを用いた提案手法が GMM による手法よりも音声認識率を改善することを確認している。

第 6 章は、本論文のまとめおよび今後の課題について述べている。