

(続紙 1)

京都大学	博士（情報学）	氏名	吉岡 拓也
論文題目	Speech Enhancement in Reverberant Environments (残響環境下での音声強調)		
(論文内容の要旨)			
本論文は、室内で遠隔にいる話者が発した音声に含まれる雑音や残響を抑制し、音声だけを強調する研究をまとめたものである。残響や雑音が音声の品質や明瞭性を低下させるだけでなく、自動音声認識の性能を著しく劣化させてしまうので、残響環境における音声強調は、音声コミュニケーションを実用化する上で不可欠の技術である。本論文では、音声強調に対する統計的モデルに基づくアプローチが述べられている。 第1章は序論で、本研究の背景として、残響を特徴づける室内インパルス応答が部屋あるいは話者やマイクロホンの位置に依存するので、事前情報を用いない場合には、残響環境における音声強調の本質的な難しさが観測信号だけを用いた残響特性推定にあることを指摘し、次に、残響抑圧を含む音声強調システムの課題が、音声生成過程と雑音・残響重畳過程の各特徴の区別を観測信号のみから行うことにあることを述べている。 第2章では、残響抑制に関する文献、特に、統計的モデルに基づいた残響抑制及び音源分離や雑音抑制についての従来研究を概観し、そこで得られた知見を基に、統計的モデルに基づいたアプローチが妥当であることを述べ、音声強調問題を一般的な形で定式化している。 第3章では、残響抑圧のためのWPE (Weighted Prediction Error) 法について述べている。本方法では、残響重畳過程のモデル化には、時間周波数領域でAR (Auto-regressive) モデルを用いている。一方、音声生成過程のモデル化では、音声のパワースペクトルの時変性に注目している。すなわち、音声生成モデルと残響重畳モデルは、それぞれパワースペクトルの時変性とフレームを跨ぐ時間相関という相異なる側面を表現するので、これによって音声生成過程と残響重畳過程を区別している。実験により、WPE法は、数秒間の観測信号が利用可能である場合に 残響抑圧性能を示すことを確認している。 なお、本手法は、第4、5章で述べる音源分離や残響抑圧を統合した音声強調方法の基礎をなす。音源分離や雑音抑圧の従来手法は時間周波数領域で動作するので、残響重畳過程を時間周波数領域でモデル化することで、こうした他の音声強調処理を残響抑圧と統合することが容易になる。 第4章では、ブラインド音源分離と残響抑圧を統合するCSD (Conditional Separation and Dereverberation) 法について述べている。本方法では、残響抑圧フィルタに音源分離フィルタを後置させた系を用いて、音源分離と残響抑圧を行う。この系での課題は、残響抑圧効果だけでなく音源分離効果も高まるように残響抑圧フィルタのパラメータを決定することである。CSD法では、音源分離フィルタと残響抑圧フィルタのパラメータを同時に最適化することで、この解決を図っている。CSD法は残響時間が0.5秒の環境でも高精度に混合音声を分離し、残響を抑圧することを実験により示している。			

第5章では、雑音抑圧と残響抑圧を統合するCNRD (Conditional Noise Reduction and Dereverberation) 法について述べている。この方法は、CSD法と同じく、音源分離フィルタと残響抑圧フィルタのパラメータを同時に最適化する点に特徴がある。雑音抑圧と残響抑圧を順に独立して実施するよりも有意に高い音声強調効果が得られることを実験により示している。

第6章では、オンライン残響抑圧のためのAWPE (Adaptive WPE) 法について述べている。AWPE法は、第3章で述べたバッチ処理方式によるWPE法を、RLS (Recursive Least Squares) アルゴリズムを用いてオンライン方式に拡張したものである。AWPE法が高速な収束特性をもつことを実験により確認している。

第7章では、残響抑圧のためのMSPP (Maximum Sparseness Power Prediction) 法について述べている。第3章で述べたWPE法は、一定の長さの観測信号が得られる場合は高い残響抑圧効果を示すものの、極めて短時間の観測信号しか得られない場合には残響抑圧ができない。一方、MSPP法は、残響抑圧効果は限定的であるが、観測信号が短い場合にも一定の残響抑圧効果をもつことを実験により示している。すなわち、MSPP法はWPE法と相補関係にあることを実証している。

第8章では本研究のまとめを行い、実環境下での音声強調に関する関連研究に対する、本研究の主要な貢献について述べるとともに、残された課題や今後の方向性についても考察を行っている。

(論文審査の結果の要旨)

本論文は、実環境において遠隔話者の録音データから、雑音や残響を低減する音声強調の技術について、統計モデルに基づくアプローチに関する研究をまとめたものである。得られた主な成果は次の通りである。

1. 残響重畳過程を時間周波数領域で自己回帰モデルを用いてモデル化し、音声生成過程を音声のパワースペクトルの時変性でモデル化することによって、音声信号に含まれる音声生成過程と雑音・残響重畳過程とを区別する統計的モデルに基づくWPE (Weighted Prediction Error) 法を考案し、数秒間の観測信号が利用可能である場合には高い残響抑圧性能を示すことを確認していること。
2. 音源分離や雑音抑圧の従来手法の多くが周波数領域で動作するので、残響重畳過程を時間周波数領域でモデル化するWPE法が、それらの従来手法を容易に統合可能な拡張性を備えていること。さらに、同手法をRLS (Recursive Least Squares) 法によりオンライン方式に拡張したAWPE (Adaptive WPE) 法も提案し、AWPE法の高速収束特性を実験により確認していること。
3. 残響抑圧フィルタに音源分離フィルタを後置させた系で音源分離と残響抑圧を統合したCSD (Conditional Separation and Dereverberation) 法を、WPE法を基に提案し、ブラインド音源分離と残響抑圧のパラメータの同時最適化法を考案し、CSD法は残響時間が0.5秒の環境でも高精度に混合音声を分離し、残響抑圧が可能なことを実験により確認していること。
4. 上記CSD法に加えて、雑音抑圧と残響抑圧を統合するCNRD (Conditional Noise Reduction and Dereverberation) 法を、WPE法を基に提案し、CSD法と同様に音源分離フィルタと残響抑圧フィルタのパラメータを同時に最適化し、雑音抑圧と残響抑圧を順に独立して実施するよりも有意に高い音声強調効果が得られることを実験により確認していること。
5. 上記WPE法の弱点である極めて短時間の観測信号しか得られない場合の残響抑圧の低下に対して、MSPP (Maximum Sparseness Power Prediction) 法を提案し、残響抑圧効果は限定的であるものの、観測信号が短い場合にも一定の残響抑圧効果を有することを実験により確認し、MSPP法とWPE法とが相補関係にあることを実証していること。

以上本論文は、残響環境における音声強調の問題に対する包括的な統計的アプローチを示し、残響抑圧、及びそれと音源分離や雑音抑圧との統合の問題を解決したことにある。特に、残響環境における音源分離は長年の未解決問題であり、この解決策を与えたことは重要な成果と考えられる。さらに、商用システム用プラグインも開発しており、その有効性を具体的に示しており、学術上、實際上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成22年2月8日実施した論文内容とそれに関連した口頭試問を行った結果合格と認めた。