

論文内容の要旨

博士論文題目 **Barge-in Robust Spoken Dialogue Interface Using
Multichannel Sound Field Control and Array Signal Processing**
(マルチチャネル音場再現とアレー信号処理を用いた
割り込み発話に頑健な音声対話インタフェース)

氏 名 宮部 滋樹

(論文内容の要旨)

音声対話システムは、ユーザに優しいマンマシン・インタフェースとして近年盛んに研究されている。このシステムでは、ハンズフリーという音声メディアの長所を生かすため、干渉音やシステム自身が発する応答音に割り込んで入力された場合でも入力音声を受理できるのが望ましい。しかし、従来の音声認識技術はこれらの外乱に対して脆弱であり、外乱除去の技術開発が急務であった。

応答音と外部雑音の除去には、従来、音響エコーキャンセラとマイクロホンアレーによる適応ビームフォーマを用いるのが一般的であった。これらの技術では、適応フィルタの学習を適切かつ高精度に行うため、対象雑音以外の音が鳴っている時間区間を検出するダブルトーク検出という仕組みが必要となる。しかしダブルトークの完全な検出は難しく、特に応答音と外部雑音の両方が存在するような騒音下では高い精度の検出は期待できない。そこで本論文では、(1) 音響伝達系変動に対して頑健な構造を持つ固定フィルタを用いた応答音除去と、(2) 応答音と外部雑音を除去する雑音除去のダブルトーク検出不要な適応に基づく、ハンズフリー音声対話のための新しい枠組みを提案した。

まず、伝達系変動に対する頑健性を向上させるため、複数の伝達経路を用いて応答音を除去する手法について検討を行った。具体的には、音場制御理論に基づき、非常に多数のラウドスピーカ群を用いてマイクロホン位置で応答音をキャンセルさせる新しい応答音除去システムを提案した。応答音を消すための音場制御としては、応答音の高品質再現を重視したものと、再現精度を緩和し

てシステム規模を縮小する二種類の手法を提案した。前者はラウドスピーカ数を増やすことにより高い頑健性を得ることができ、ディクテーションタスクの音声認識実験では、24 個のラウドスピーカを用いることで従来型音響エコーキャンセラよりも 20%高い単語正解精度が得られた。また後者は、少ないスピーカ数でも安定な制御を実現することができ、5 個のラウドスピーカを用いて従来型音響エコーキャンセラよりも 15%高い単語正解精度が得られた。

次に、外部から到来する雑音の除去と上述の応答音除去とを融合するため、未知音源に対する教師なし適応技術であるブラインド音源分離 (BSS) を導入した。特に、本研究においては、応答音がシステムにとって既知であることに着目した。BSS を既知である応答音の情報を利用した半教師なし構造に拡張することにより、BSS よりも効率的に音源分離を行うセミブラインド音源分離アルゴリズムを提案し、前述の音場制御による応答音除去と組み合わせた。本複合システムにおいて、セミブラインド音源分離処理部は応答音の消し残りや外部雑音を除去し、音場制御による音声強調を補助することが可能である。実音響環境を模擬した実験により、提案手法の性能と音響エコーキャンセラと適応ビームフォーマの併用手法の性能限界とを比較した。その結果、単語正解精度に関して約 10%の向上が得られた。以上より、マルチチャネル音場再現技術とアレイ信号処理を融合した割り込み発話に頑健な新しい音声対話インタフェースの有効性が確認された。

(論文審査結果の要旨)

音声ユーザに優しい柔軟なヒューマンマシンインタフェースとして、ハンズフリー音声認識・対話システムが望まれている。しかし実環境下では、ユーザが発話する際に、システム応答音や種々の背景雑音等が混入してしまうため、高い音声認識率を達成することは困難とされてきた。また、従来の教師あり適応技術である音響エコーキャンセラや適応ビームフォーマ用いても、ダブルトーク検出の困難さのために、実用的な運用は不可能であるとされていた。本研究では、固定フィルタを用いた音場制御技術およびユーザ音声発話位置や区間等の事前情報が不要であるという利点を有するブラインド音源分離 (BSS) を導入することで、上記の問題の解決を図っている。本論文では、以下で述べる幾つかの新規理論提案を行い、その有効性を実環境データによって検証している。

(1) 伝達系変動に対する頑健性を向上させるため、音場制御理論に基づき、非常に多数のラウドスピーカ群を用いてマイクロホン位置で応答音をキャンセルさせる新しい応答音除去システムを提案した。ディクテーションタスクの音声認識実験では、従来型音響エコーキャンセラよりも 15~20%高い単語正解精度が得られた。

(2) 外部から到来する雑音の除去と上述の応答音除去とを融合するため、未知音源に対する教師なし適応技術であるブラインド音源分離 (BSS) を導入した。特に、本研究においては、応答音がシステムにとって既知であることに着目し、従来 BSS よりも効率的に音源分離を行うセミブラインド音源分離アルゴリズムを提案した。実音響環境を模擬した実験において、音響エコーキャンセラと適応ビームフォーマの併用手法と比較して約 10%の単語正解精度の改善が確認された。

以上より、マルチチャンネル音場再現技術とアレー信号処理を融合した新しい音声対話インタフェースの有効性が確認された。これらの手法は、従来より困難とされていた割り込みおよび外部雑音の同時抑圧問題の解決に関して、有用かつ画期的なものであったと高く評価できる。本研究成果は、2 編の英文原著論文や 9 件の査読付き国際会議にて採録されており、その研究レベルの高さが内外より評価されていることが分かる。

以上述べたように、本論文の音響・音声信号処理分野への貢献は極めて大きい。平成 19 年 7 月 30 日に開催した公聴会の結果も参考にして、本博士論文の審査を行い、本論文は博士論文 (工学) の学位論文として十分な価値があるものと判断した。