

## 論文内容の要旨

博士論文題目 Hands-free Speech Recognition Using a Microphone Array

(マイクロホンアレーを用いたハンズフリー音声認識に関する研

氏名 山田 武志

(論文内容の要旨)

(1, 200字程度)

自然で使い勝手の良い音声インタフェースを実現するためには、実環境下でのハンズフリー音声認識の技術が必要不可欠である。本論文では、マイクロホンアレーを用いたハンズフリー音声認識についての研究成果をまとめる。マイクロホンアレーによる雑音・残響抑圧は、所望の音源の方向に対して感度が高い指向性を形成することにより行われる。よって、発話者の方向が未知というハンズフリーの状況では、指向性ビームの方向と発話者の方向とのずれを防ぐために発話者の方向を正確に推定することが極めて重要となる。従来、音声の短時間パワーや長時間パワーという情報を元に発話者の方向を推定する方法が提案されている。しかし、雑音や残響が存在する環境下では十分な精度が得られないという問題がある。従来法では発話者の方向を一意に決定するので、誤りが生じた場合に認識の処理で対処することは難しい。本論文では、音声認識の処理と発話者方向推定の処理を確率的な枠組の中で統合するという新しいアプローチを導入し、3次元ビタビ探索に基づく音声認識法を提案する。3次元ビタビ法では、フレーム毎に指向性ビームを対象とする方向に順次向け、特徴ベクトルの方向・時間系列を求める。そして、この系列と音声の確率モデルとの間でパターンマッチングを行うことにより、発話者の移動軌跡と発声内容を同時に推定する。提案法では、発話者の方向の候補を複数考慮しながら認識を行うので、従来法のような方向推定の誤りに伴う問題は生じない。

3次元ビタビ法の初期評価を行うために、無響室を想定して生成したマイクロホンアレーデータを用いて特定話者の単語認識実験を行った。その結果、指向性ビームによるSNR (Signal to Noise Ratio) の改善量が不十分な場合、発話者の方向をうまく推定できず、認識精度が低下することが分かった。この問題に対処するための方法としては、マイクロホンを増やして指向性ビームを鋭くすることが考えられる。しかし、装置の規模やコストという観点から現実的ではない。本論文では、音声らしい特徴を有する方向に重み付けする方法と適応型アレーを用いて指向性ビームの形状を改善する方法を導入した。実際の環境で収録したマイクロホンアレーデータを用いて評価した結果、発話者の位置が固定の場合、SNR 21 dB のとき 93.9 %、SNR 18 dB のとき 89.8 %、SNR 10 dB のとき 83.3 %の単語認識率を得ることができた。また、発話者が移動する場合、SNR 21 dB のとき 92.5 %、SNR 18 dB のとき 88.8 %、SNR 10 dB のとき 81.0 %の単語認識率を得ることができた。最後に、実際の環境で収録したマイクロホンアレーデータを用いて不特定話者の単語認識実験により3次元ビタビ法の評価を行った。その結果、SNR 21 dB のときの単語認識率は接話マイクロホンと比べて 8.8 %の低下に抑えることができた。以上の結果から3次元ビタビ法の有効性を確認することができた。

(論文審査結果の要旨)

音声による人間と機械のインタフェースは、人間にとって最も自然なインタフェースである。しかしながら、現在の音声認識は雑音に脆弱であるため、接話マイクロホンの利用を余儀なくされている。ところが、自然なマンマシンインタフェースを実現する上では、マイクロホンを意識せずに、マイクロホンから離れて発話した音声の認識（ハンズフリー音声認識）が極めて重要な技術となる。本論文では、実環境においてマイクロホンから離れて移動しながら発話した音声の認識を、複数のマイクロホン素子から構成されるマイクロホンアレーを用いて頑健に行うための基礎研究に付いて述べている。

本論文の成果は、次のように要約できる。

(1) ハンズフリー発話の音声の認識のためにマイクロホンアレーを導入する枠組みの提案を行っている。マイクロホンアレーは、任意の指向性を信号処理により形成し、音源の空間的異なりに基づき、対象とする音源を高 SNR で受音する技術である。論文では、音声認識に応用する際に、次のような問題が有る事を明らかにしている。

- ・マイクロホンアレーの指向性を鋭くすればするほど、音源位置の推定に高い精度が要求される。
- ・SNR が低い場合には、音源の位置の推定、即ち発話者の位置、方向を抽出することが極めて困難となる。
- ・従来のようにマイクロホンアレーを音声認識装置の前段と位置づけ、対象音源の信号を高 SNR で抽出した後、音声認識を行う従来型のシステム構成では、音源の定位を誤れば致命的な認識誤りに繋がる。

(2) 上記問題を解決するため、音声認識の処理と発話者方向推定の処理を確率的な枠組みの中で統合するという新しいアプローチを提案している。この手法は、音声の分析フレームごとに対象とするすべての方向に指向性ビームを順次向け、特徴ベクトルの方向・時間系列を求める。そして、この系列と音声の確率モデルとの間で最尤復号を行うことで、発話者の移動軌跡と発話内容の音声認識を同時に行うものである（3次元ビタビ法と呼ぶ）。論文では、提案法を計算機シミュレーション、実環境音声に適用し有効性を確認し、さらに音声の確率モデルに音声源に特有の調波性の重みを導入し精度を向上する方法について述べている。マイクロホンアレー信号処理と音声認識を確率的な一つの枠組みで統合した手法については、極めて高く評価できるものである。

(3) 提案する 3次元ビタビ法においても、マイクロホンの指向性の構成が極めて重要なファクターとなる。論文では、さらに精度向上する方法として、マイクロホンアレー信号処理に適応型アレーを用い、SNR を改善する方法について提案している。実環境で移動しながら発話した音声に対し、適応型アレーを用いた提案手法を適用し、遅延和アレー型に対し著しい改善を得ることができることを明らかにしている。

以上のように、本論文はハンズフリー発話の音声認識を目的に、マイクロホンアレーと音声認識を確率的な枠組みにより統合する 3次元ビタビ法を提案し、有効性を示した。これらの研究成果は、情報科学、音情報処理学の分野における学術的な貢献として高く評価できる。平成 10 年 12 月 24 日開催の公聴会の結果を参考に、本博士論文の審査を行なった結果、本論文は博士(工学)の学位論文として十分に価値のあるものと判断する。