

論文内容の要旨

博士論文題目

Simultaneous Recognition of Distant-talking Speech of Multiple Sound Sources
(複数話者の遠隔発話同時音声認識に関する研究)

氏名

Panikos Heracleous

(論文内容の要旨)

本博士論文では、遠隔発話の音声認識アルゴリズムと実験結果について述べている。とくに、二人以上の話者が同時に発声した場合の遠隔発話の音声認識について述べている。現在の音声認識では、接話マイクや20~30cmの距離におかれたデスクトップマイクで収録されることを前提に研究開発されてきた。しかしながら、日常生活で人は、数メートル離れた声も容易に認識しているし、複数の話者の同時発話も認識することができる。

この博士論文では、複数話者の遠隔発話の同時音声認識を行うことを目的としている。このために、発話音声の発声方向の情報と遠隔発話を明瞭な受音のために、マイクロフォンアレーを用いる。従来の遠隔発話の音声認識手法では、まず発話者の方向を見つけて、次にマイクロフォンアレーでビームフォーミングを行い、最後に音声認識を行っていた。このような逐次処理の過程では、話者方向の同定誤りが致命的になる傾向があった。とくに、この博士論文のように複数話者が同時に発声している場合には、複数話者の正確な方向を見つけることには困難が伴う。

マイクロフォンアレーによる発話者の方向同定とビームフォーミングを音声認識のアルゴリズムの中で、尤度の基準で同時に行うことによって、複数話者の遠隔発話を可能にすることを考えている。このために、3-D Viterbi サーチアルゴリズムを考案した。この3-D Viterbi サーチアルゴリズムでは、従来からの音声入力フレームの時間軸の他に、あらゆる方向を新たな軸としている。この時間軸と方向のすべてのノードに、単語 HMM による N ベストの認識候補を保持している。この3次元に拡張された Viterbi サーチを One Path アルゴリズムによって実装した。さらに、認識候補のクラスターリングと認識尤度の正規化を考案して、実環境でも動作することを確認した。

最終的に、216 単語音声の認識タスクで、二人の話者が単語を発声し、32チャンネルのマイクロフォンアレーを用いて、二人の発声単語が同時に正しい単語認識率で、72.5%を達成した。上位3位以内に入る単語認識率は、86.3%であった。この認識性能は、従来の認識性能を大幅に上回っており、本博士論文のアルゴリズムの有効性を証明している。

氏名	Panikos Heracleus
----	-------------------

(論文審査結果の要旨)

本博士論文では、遠隔発話の音声認識アルゴリズムと実験結果について述べている。とくに、二人以上の話者が同時に発声した場合の遠隔発話の音声認識について述べている。現在の音声認識では、接話マイクや20～30cmの距離におかれたデスクトップマイクで収録されることを前提に研究開発されてきた。しかしながら、日常生活で人は、数メートル離れた声も容易に認識しているし、複数の話者の同時発話も認識することができる。

この博士論文では、複数話者の遠隔発話の同時音声認識を行うアルゴリズムと認識実験結果について報告している。従来の遠隔発話の音声認識手法では、まず発話者の方向を見つけて、次にマイクロフォンアレーでビームフォーミングを行い、最後に音声認識を行っていた。このような逐次処理の過程では、話者方向の同定誤りが致命的になる傾向があった。とくに、この博士論文のように複数話者が同時に発声している場合には、複数話者の正確な方向を見つけることには困難が伴うことが知られている。

マイクロフォンアレーによる発話者の方向同定とビームフォーミングを音声認識のアルゴリズムの中で、尤度の基準で同時に行うことによって、複数話者の遠隔発話を可能にすることを考えて、3-D Viterbi サーチアルゴリズムを考案している。この3-D Viterbi サーチアルゴリズムでは、従来からの音声入力フレームの時間軸の他に、あらゆる方向を新たな軸としている。この時間軸と方向のすべてのノードに、単語 HMM による N ベストの認識候補を保持している。この3次元に拡張された Viterbi サーチを One Path アルゴリズムによって実現している。さらに、認識候補のクラスターリングと認識尤度の正規化を考案して、実環境でも動作することを確認している。

最終的に、216 単語音声の認識タスクで、二人の話者が単語を発声し、32チャンネルのマイクロフォンアレーを用いて、二人の発声単語が同時に正しい単語認識率で、72.5%を達成している。この認識性能は、従来の認識性能を大幅に上回っており、本博士論文のアルゴリズムの有効性を証明している。

以上述べたように、マイクロフォンアレーを用いた複数話者の遠隔発話の音声認識のアルゴリズムの提案と、実環境での認識実験の検証までを行っており、研究としてよくまとまっている。この分野は、人と機械との円滑なコミュニケーションとして、音声インタフェースの新しい有望な分野であり、この先導的な研究は、情報科学、とくに音情報処理の分野に多大な貢献をしたと評価できる。平成14年1月21日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(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。