

氏名(本籍)	高橋 毅 (秋田県)
専攻分野の名称	博士 (工学)
学位記番号	工博甲第208号
学位授与の日付	平成25年9月25日
学位授与の要件	学位規則第4条第1項該当
研究科・専攻	工学資源学研究科 (電気電子情報システム工学)
学位論文題名	口唇の特徴に着目したコマンド識別および発話認識に関する基礎研究
論文審査委員	(主査) 教授 西田 眞 (副査) 教授 五十嵐 隆治 (副査) 教授 水戸部 一孝 (副査) 教授 景山 陽一

論文内容の要旨

近年、高機能化の著しい各種情報機器の利用を支援するために、ユーザビリティに主眼を置くヒューマンマシンインタフェースの研究・開発が行われている。その中で、日常的な動作である「発話」の音声情報に着目したインタフェースの研究・開発も行われており、カーナビゲーションなど多くのシステムに応用されている。音声による操作は、操作盤を視認する必要がなく、操作に熟達する必要もないことから、非常に利便性が高い。しかしながら、音声認識は雑音環境下では認識精度が低下すること、静粛性が要求される環境や喉の不調時などの発声が制限される状況下では使用に適さないこと、発話内容が第三者に知られてしまう可能性のあることなどの課題を有している。

一方、発話には口唇・舌の動作が必要不可欠であるため、音声情報に加えて口唇の動きという視覚情報が得られる。この口唇の動きには、発話内容に関する情報が含まれており、これをコマンド入力や発話認識システムに適用した研究事例も多数報告されている。しかしながら、多くの利用者が共用する状況下において良好な認識を可能にする要素技術、ならびに自然な発話条件を実現するための要素技術の開発は十分とはいえない。このため、口唇の動き特徴を用いたヒューマンインタフェースの要素技術として、(1)自然な発話状態での入力操作においても発話区間を良好に自動抽出する手法、(2)多数の利用者がシステムを共用する使用環境へ対応するための手法、(3)発声に起因する口唇の動き特徴変動を考慮した手法の開発が望まれている。本論文は上記項目に関わる課題について研究を行い、工

学上の進歩に寄与することを目的とするものである。

本論文は全 5 章より構成されている。

第 1 章を緒論とし、ここでは本研究の背景とその目的を述べ、本研究に対する筆者の立場を明らかにした。さらに、本論文の主題である口唇の動き特徴を用いたコマンド識別・発話認識システムならびに実用化のための要素技術について、現在までの研究状況を概観するとともに、本研究の内容について述べた。

第 2 章では、発話時の口唇画像における $L^*a^*b^*$ 色空間の色彩情報解析を行い、口裂の色彩情報および口唇形状の時系列変化を特徴量とする発話フレーム自動検出法を提案した。提案手法では、口唇領域の明度 L^* と赤み a^* を特徴量としてフレーム単位で口の開閉状態を判別し、発話区間内において生じた閉口状態を口唇形状の時系列変化に基づいて検出することで、複数の単語を任意の間隔で発話する状況においても発話フレームを検出可能とした。その有用性を評価するため、被験者 5 名を対象に 3 つの単語を任意の間隔で発話させたデータを取得し、発話フレームの検出を行った。その結果、提案手法は各単語の発話フレームを約 99% の高精度で検出可能であることを明らかにした。

第 3 章では、多数の利用者が共用する環境下でのコマンド識別・発話認識精度の向上を目的とし、非発話状態の口唇局所部位を対象とした形状特徴の統計的解析を行い、解析から得られた局所形状特徴を用いた口唇形状のグループ化手法について検討を加えた。具体的には、被験者 106 名を対象に非発話状態の口唇画像を取得し、上唇と下唇の厚さ比率特徴量、口裂の凹凸形状特徴量、ならびに口唇のアスペクト比の 3 種類の形状特徴量を算出した。得られた 106 名分の特徴量に対して統計的な解析を加え、各形状それぞれ 3 クラスから構成される 27 の形状カテゴリを構築した。さらに、口唇の局所形状に基づいて口唇形状を 27 カテゴリに自動分類するアルゴリズムを提案し、被験者 52 名を対象にした評価実験を行った。その結果、80% 以上の精度で登録データおよびその類似形状に分類可能であることを示した。また、4 位カテゴリまでに分類可能であった被験者において、照合対象を約 1/8.5 に絞り込み可能であることを示した。

第 4 章では、発話内容の識別精度向上を目的とし、発声の有無に起因する口唇の動き特徴変動に関して検討を加えた。具体的には、発話フレーム数および口唇の動き特徴（横幅、縦幅、面積、アスペクト比）のフレーム間差分の累積値に着目し、各特徴量と発声との関連について検討を加えた。さらに、同一取得日における無声発話データと有声発話データの判別についても検討を加えた。その結果、無声発話時は有声発話時と比較し、発話区間が長くなる傾向を有すること、発話全体を通した口唇の動作量が大きくなる傾向を有することが明らかになった。また、発話フレーム数および口唇動作量を用いた線形判別手法は、約 92% の精度で同一取得日における無声発話データと有声発話データを判別可能であることを明らかにした。

第 5 章は結論で、本研究で得られた主な成果と本論文の工学的意義および今後に残された課題について述べている。

論文審査結果の要旨

近年、高機能化の著しい各種情報機器の利用を支援するため、ユーザビリティの高いヒューマンマシンインタフェースの研究・開発が行われている。その中で、「発話動作」が有する「口唇の動き特徴」に着目した、利便性の高いコマンド入力や発話認識システムに関する研究成果も多数報告されている。しかしながら、多くの利用者が共用する状況下において良好な認識を可能にする要素技術、ならびに自然な発話条件を実現するための要素技術の開発は十分とは言い難い。このため、口唇の動き特徴を用いたインタフェースの要素技術として、(1) 自然な発話状態での入力においても発話区間を良好に自動抽出する手法、(2) 利用者やコマンド数の増加へ対応するための手法、(3) 発声に起因する口唇の動き特徴変動を考慮した手法の開発が望まれている。本論文は、これら課題に対する検討結果をまとめたもので、全 5 章より構成されている。

第 1 章は緒論であり、本研究の背景とその目的、研究分野における筆者の立場を明らかにしている。さらに、本論文の主題である口唇の動き特徴を用いたコマンド識別・発話認識システムについて、現状を概観するとともに、本研究の内容について述べている。

第 2 章では、より自然な発話状況下での発話区間推定を目的とし、口唇の色彩情報と形状情報に着目した発話フレーム検出法について検討を加えている。具体的には、 $L^*a^*b^*$ 表色系の明度指数 L^* 、赤みの知覚色度指数 a^* を指標として開口および閉口状態の色彩情報解析を行い、口唇の L^* 値 a^* 値ならびに口唇形状の時系列変化量に着目した発話フレーム検出法を提案している。提案手法は、被験者 5 名を対象とした発話フレーム検出実験において発話フレームを高精度 (90% 超) で検出可能であり、複数の単語を含む発話データにおける発話区間の推定に有用であることを明らかにしている。

第 3 章では、ユーザやコマンド数の増加への対応を目的とし、口唇局所部位の形状特徴に着目した口唇形状のグループ化手法について検討を加えている。具体的には、上下唇の厚さ、口裂形状、ならびに口唇のアスペクト比の 3 つの形状特徴について統計的な解析を行い、各形状それぞれ 3 クラスから構成される 27 形状の口唇形状カテゴリを提案している。さらに、口唇形状を上記 27 カテゴリに自動分類するアルゴリズムを提案し、被験者 52 名を対象にした評価実験において、80% 以上の精度で類似形状の範囲に分類可能であること、分類結果は照合対象の絞り込みに有用であることを示している。

第 4 章では、コマンド識別精度向上を目的とし、発声の有無と口唇の動き特徴の関連について検討を加えている。具体的には、発話フレーム数および口唇横幅、縦幅、面積、アスペクト比の累積変化量と発声の有無との関連について検討を加えると同時に、同一取得日における無声発話データと有声発話データの判別についても検討している。その結果、無声発話時は有声発話時と比較し、発話時間が長くなる傾向ならびに口唇の動作量が増加

する傾向を有することを明らかにしている。さらに、発話フレーム数と口唇動作量を指標とすることで、同一取得日における無声発話データと有声発話データを高精度（約 92%）で線形判別可能であることを示している。

第 5 章は結論で、本研究で得られた主な成果と本論文の工学的意義および今後に残された課題について述べている。

以上のように本論文で得られた成果は、発話に伴う口唇の動きを用いた実用的なヒューマンインタフェースの研究・開発に対して、重要な知見を与えるもので、その工学的意義は大きい。よって、博士（工学）の学位論文として十分価値あるものと認められる。