氏名（本籍）	中村　悦郎（岩手県）
専攻分野の名称	博士（工学）
学位記番号	理博甲　第　264　号
学位授与の日付	令和　　4年　3月　22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	理工学研究科　総合理工学専攻
学位論文題目（英文）	議事録自動作成システムのための画像および音声情報を用いた発話者判別手法に関する研究
論文審査委員	（主査）教授　景山　陽一 （副査）教授　有川　正俊 （副査）教授　水戸部　一孝

論文内容の要旨

近年，働き方改革の実現や新型コロナウイルスによる働き方の変化に伴い，業務の効率化が取り組まれている．特に，職場における労働環境の改善策として，会議の効率化が注目されている．会議時に作成される議事録は，情報共有やその後の会議の質を向上させること，ならびに業務の効率化に役立てられる．一方，音声認識の技術を応用し，議事録を自動作成することは，議事録作成におけるヒューマンエラーの低減や，工数の削減に寄与できる．特に，発言ごとに発話者を自動判別する技術は，議事録自動作成手法における重要な要素技術であり，その開発が急務である．

発話者判別手法として，人物ごとにマイクを割り当てる手法や，事前に取得した声紋から話者を判別する手法が提案されている．しかしながら，これらの手法は，会議環境の整備や事前の声紋登録が必要である点が課題として挙げられる．一方，発話に伴う口唇の動きは，発話内容特有の特徴を有している．したがって，画像中における口唇の動きと音声情報との類似性などを判別することで，事前準備を必要としない発話者判別手法の構築が可能であると考える．類似研究として，機械学習の1手法であるConvolutional Neural Network (CNN)およびLong Short-Term Memory (LSTM)を用い，画像と音声から発話者判別を行う手法(以降，比較手法と表記する)が提案されている．しかしながら，上記手法で

は，CNNおよびLSTMの学習のために，膨大な量の動画データを必要とする点が課題である．一方，画像の輝度勾配に基づき顔の特徴点を取得するための手法(以降，輝度勾配に基づいた手法と表記する)が提案されている．このような口唇部分の特徴点を入力特徴量として使用することは，既存の手法よりも少ない学習データ量を用いた発話者判別手法の構築に寄与すると考える．しかしながら，輝度勾配に基づいた手法は，画像中の顔の部位(目や鼻など)が隠れている場合，顔器官の検出に影響を及ぼし，口唇領域の抽出が困難になる場合がある．このため，発話者判別手法を搭載した利便性の高い議事録自動作成システムを構築するためには，①顔の部位が隠れている人物に対する口唇形状抽出手法の開発，②口唇の動きが生じている人物が1名である場合に発話者を特定するための発話区間抽出手法の開発，ならびに③複数の人物で口唇の動きが生じた場合において発話者を特定するための手法の開発が必要である．

そこで本論文では，発話者判別機能の搭載された利便性の高い議事録自動作成システムの開発に関する上記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．すなわち，被験者が発話している動画データを対象とし，①口唇形状自動抽出手法，②発話区間抽出手法，ならびに③発話者判別手法を提案し，工学上の進捗に寄与することを目的とする．本論文の内容は，5章から構成されている．

第1章を序論とし，本論文の主題である会議の効率化の必要性，実用化されている議事録自動作成システム，発話者判別手法の関連研究，ならびに本研究の目的および本研究に対する筆者の立場を述べるとともに，本論文の内容について述べている．

第2章では，口唇形状自動抽出法について検討を加えている．具体的には，口唇と肌の赤味に着目して口唇領域と肌領域をクラス分類する手法を提案した．5名の被験者の動画データを使用して評価を行ったところ，提案手法は，最大で0.9286の高いIntersection over Union (IoU)を得た．また，輝度勾配に基づいた手法と比較して，提案した口唇形状自動抽出法は，眼鏡やサングラスを掛けている場合においても平均で0.9000以上の高いIoUの値を得た．提案手法は肌の色の個人差や陰影の影響による精度低下を低減可能であること，および顔の他の部位が遮蔽されている場合においても口唇形状を抽出可能であることを明らかにした．

第3章では，発話区間の抽出手法について検討を加えた．具体的には，発話に伴う口唇の動きと音声が同時に発生したフレームを発話区間として抽出する手法に関して検討を加えた．14名の被験者の動画データを使用して評価指標(F-measure)を算出した結果，提案した発話区間抽出手法は，最大で0.99の高いF-measureの値を得た．また，上述した比較手法と比較して提案した発話区間抽出手法は，14名中13名でF-measureの向上を認めた．提案手法は機械学習を使用しないシンプルな手法を用いて，発話区間を判別することが可能であることを明らかにした．

第4章では，発話者の判別手法に関して検討を加えた．具体的には，音声を用いて推定された口唇の動きと実際の口唇の動きとの類似度を評価し，最も類似度が高い人物を発話者

として判別する手法を検討した．14名の被験者の動画データを使用して発話者判別成功率算出したところ，上述した比較手法に対する約0.05%の量の学習データを使用し，最大で93.0%，平均で87.2%の判別成功率を得た．提案手法は比較手法と比較して，単位学習データ当たりの学習効率が高く，かつ少ない学習データを用いて発話者判別のためのモデル構築が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，本論文の工学的意義，ならびに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

当学位審査委員会は，2022 年 2 月 15 日（火）10 時 30 分から 11 時 30 分まで，オンラインでの公開による論文公聴会を開催した．その後，景山陽一審査委員会主査，有川正俊審査委員，水戸部一孝審査委員出席のもと，論文内容と関連事項に関して詳細な質疑応答を行うとともに，口頭による学力の確認を行った．

審査委員からは，

・発話者と非発話者のデモにおける実際の状況への適用について
・全方位カメラから取得された顔の向きやフォーカスの程度について
・検討に用いた基準として，時間ではなくフレーム数を用いた理由について
・本研究で用いた画像の解像度について
・発話デモにおける被験者の発話内容について

などの質問が行われたが，申請者から明確な回答が得られた．

本研究に関する原著論文は 2 編である．また，国際会議における英語での発表を行っていることなどから判断して，博士（工学）としての学力が認められる．よって，中村悦郎氏は博士（工学）として十分な資格があるものと判定した．