

審査の結果の要旨

氏名 内田 秀継

本論文は「音響情報・言語情報に基づく調音運動の復元」と題し、5章より成る。音声の生成は、肺からの呼気流に対して、舌・唇などの調音器官を制御し、所望の音韻特性を与えることで行なわれる。本論文では、観測された音声信号からその信号を生成させた調音運動を推定する問題（音声生成の逆問題）に着目する。この推定問題を、音響空間から調音空間への空間写像推定問題として捉え、話者ミスマッチ問題と言語ミスマッチ問題に起因する推定精度を改善する方法を提案し、その有効性を実験的に検証している。

第一章は「序論」であり、本論文の背景、目的、及び、構成を述べている。

第二章は「音声の生成過程と逆推定問題」と題し、まず、音声の生成過程と調音運動について音声科学での知見も含め、紹介している。次に、本研究におけるデータ収集と関連する、調音運動の観測方法について現時点で可能な手法についてまとめ、それを踏まえて、本研究が対象とする逆推定問題、音声信号からの調音運動復元問題について、先行研究の成果をまとめている。音声からの調音運動復元問題は音声科学的には、「音声知覚の運動参照仮説」と密に関係する。この仮説と、本論文で検討する写像推定との関連性についてもまとめている。最終的に、これら先行研究に対する課題として、話者ミスマッチ問題と言語ミスマッチ問題を取り上げ、本研究の位置づけを明確にしている。

第三章は「話者正規化音声-調音マッピング」と題し、第一の課題である話者ミスマッチ問題に対する技術的解決策を示している。音声-調音マッピング（写像）は、ある特定話者の音声-調音パラレルコーパスから推定されるが、その写像関数は、その話者特有のモデルとなり、他話者での利用は困難である。これを回避するために、1) 前処理として話者正規化を行ない、任意の話者の声を、変換モデルが用意されている標準話者へ変換した上で音声・調音変換を行う手法、及び、2) 話者正規化処理、音声・調音変換処理の両方が写像推定問題であることを鑑み、両者を統合し、一つの写像関数として捉えて最適化する方法を提案している。特に後者においては、入力話者の音声、標準話者の音声、標準話者の調音運動の3つの特徴ストリームに対する分布を想定し、不可避な欠損デ

一タの取り扱い方を検討している。二種類の提案手法について、実験を通してその有効性を検証すると共に、両者の比較を行っている。

第四章は「音声の構造的表象を用いた未観測音素の調音運動の推定」と題し、第二の課題である言語ミスマッチ問題に対する解決策を示している。話者ミスマッチ問題が解決された場合であっても、入力話者が所望の音素を上手に生成できない場合は、その話者の調音運動の推定は困難になる。これは音声・調音変換モデルが想定する言語と、入力話者の母語とにミスマッチがあった場合に（例えば、外国語学習において母語にない音素を発音する場合など）生じる。本章では、対象言語の母語話者より得られる音素群の構造的表象を導入し、これを参照することで、入力話者がその発声に困難を抱える音素の音響的特徴を、発声可能な音素群と当該音素の関係性により推定し、最終的に、その調音運動を求める技術的解決策を提案し、有効性を検証している。また、第三章で提案した手法と本章で提案した手法の音声科学的な相違点についても述べている。

第五章は「結論」であり、本研究のまとめと今後の展望を述べている。

以上要するに本論文は、統計的な音声・調音変換問題（音声からの調音運動復元問題）において未解決であった、話者ミスマッチ問題と、言語ミスマッチ問題に対して、音響情報及び言語情報を有効に活用することで技術的解決策を提供しており、音声科学及び情報工学に貢献するところが少なくない。よって本論文は博士（工学）の学位請求論文として合格と認められる。