

## 論文審査の結果の要旨および担当者

報告番号	※ 甲 第	号
------	-------	---

氏 名 Lumbantobing Patrick

論 文 題 目

High-Quality and Flexible Voice Conversion Techniques  
based on Statistical Spectral and Waveform Modeling  
(統計的スペクトル変換・波形生成モデルに基づく高品質  
かつ柔軟な音声変換技術)

論文審査担当者

主 査 名古屋大学教授 戸田 智基

委 員 名古屋大学教授 武田 一哉

委 員 名古屋大学准教授 竹内 栄二郎

委 員 日本電信電話株式会社コミュニケーション科学基礎研究所  
メディア情報研究部特別研究員 亀岡 弘和

Lumbantobing Patrick 君提出の学位論文「High-Quality and Flexible Voice Conversion Techniques based on Statistical Spectral and Waveform Modeling (統計的スペクトル変換・波形生成モデルに基づく高品質かつ柔軟な音声変換技術)」は、身体的制約を超えた発声機能の獲得などへの応用が期待される統計的音声変換技術に対して、高品質であり柔軟性に優れた音声変換処理を実現する手法を論じた論文であり、全7章より構成されている。

第1章は序論であり、本研究の背景を説明するとともに、本論文で提案する手法間の関係性について述べている。

第2章では、従来研究として、音声生成過程における調音・音響特徴量間変換技術、統計的音声変換におけるスペクトル変換技術、深層音声波形生成技術を説明するとともに、本研究で取り組む研究課題との関連性を明らかにしている。

第3章では、柔軟性に優れた音声変換処理を実現する基盤技術として、調音動作の制御機能を備えた統計的音声変換手法について論じている。調音動作を潜在変数とした音声波形モデリング手法を提案することで、未観測の調音動作に対する加工処理を可能とし、解釈性ならびに操作性に優れた音声変換処理を実現している。

第4章では、高品質な音声変換処理を達成する基盤技術として、統計的スペクトル変換処理ならびに音声波形合成処理の改善手法について論じている。深層学習に基づく統計的スペクトル変換処理に加え、音声波形直接加工処理および深層音声波形生成処理を組み合わせた手法を提案することで、音声変換性能の改善に成功している。また、提案手法に基づき構築された統計的音声変換システムは、音声変換技術の国際的評価会 Voice Conversion Challenge 2018 において、世界中から23チームが参加する中、第2位に相当する性能が得られたことを明らかにしている。

第5章では、さらなる高品質化を達成する基盤技術として、統計的スペクトル変換処理と深層音声波形生成処理の同時最適性を加味した音声変換モデル学習手法について論じている。循環型再帰的深層ネットワークに基づく統計的スペクトル変換モデルの学習手法、ならびに、変換誤差を模倣した学習データ生成に基づく深層音声波形生成モデルの学習手法を提案することで、変換音声の音質、および、目標音声との類似度を改善できることを示している。

第6章では、任意の発話音声を用いた柔軟性の高い学習を可能とする基盤技術として、循環型変分自己符号化器に基づく統計的スペクトルモデリング手法について論じている。スペクトル特徴量の再構成に基づく従来型の変分自己符号化器学習処理に対して、循環構造をもたせた変換処理を新たに導入することで、変換処理を考慮した変分自己符号化器学習を可能とし、スペクトル特徴量からの言語情報の分離精度を改善できることを示している。これにより、教師なし学習の枠組みにおいても、高精度な統計的音声変換処理を実現できることを明らかにしている。

第7章では、本研究をまとめ、残された課題を明らかにしている。

以上、Lumbantobing Patrick 君の学位論文は、統計的音声変換処理において、従来の枠組みでは困難であった高品質な変換処理、ならびに、柔軟な変換・学習処理を実現する基盤技術を生み出している。その成果は、学術的・産業的に寄与するところが大きく、本審査委員会は、論文提出者である Lumbantobing Patrick 君が博士(情報科学)の学位を授与されるに十分な資格を有する者と判定した。