

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 HUANG Wen-Chin

論 文 題 目

Pre-training Approaches for Voice Conversion to Address Data Scarcity and Their Applications to Ground-Truth-Free Tasks
(音声変換におけるデータ不足に対処するための事前学習法と
入手不可目標データ課題への応用)

論文審査担当者

主 査 名古屋大学情報基盤センター 教 授 戸田智基

委 員 名古屋大学未来社会創造機構 教 授 武田一哉

委 員 カーネギーメロン大学 言語技術研究所 准教授 渡部晋治

HUANG Wen-Chin (黄 文勁) 君提出の学位論文「Pre-training Approaches for Voice Conversion to Address Data Scarcity and Their Applications to Ground-Truth-Free Tasks (音声変換におけるデータ不足に対処するための事前学習法と入手不可目標データ課題への応用)」は、音声に対して言語情報を保持したまま所望の情報を変換するように加工処理を与える音声変換技術において、実応用ではしばしば十分な学習データが得られないという課題を解決するために、基礎技術の構築と応用範囲の拡大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、ならびに、音声変換の課題を述べている。

第2章では、代表的な枠組みとして、入出力音声対で構成されるパラレルデータを用いた系列変換と、ノンパラレルデータを用いたフレーム変換を説明するとともに、その関係性および課題を論じている。また、音声変換性能を評価する標準的な手法について説明している。

第3章では、系列変換に基づく枠組みにおいて、大量のパラレルデータを必要とする課題に着目し、事前学習を活用することで少量のパラレルデータの利用を可能とする基礎技術について論じている。テキスト音声合成用もしくは音声認識用に学習された系列変換モデルに対し、自己符号化に基づく事前学習を行う手法を提案しており、数分程度のパラレルデータを用いた場合においても、高精度な音声変換を実現できることを実験的に示している。

第4章では、フレーム変換に基づく枠組みにおいて、高精度な音声変換用特徴表現を得るためには、事前学習において言語情報に関する教師ラベルを必要とする課題に着目し、教師ラベル情報を必要としない自己教師あり学習を活用する基礎技術について論じている。自己教師あり学習特徴表現の利用を可能とすることで、任意の話者に対応可能な音声変換の枠組みを実現するとともに、標準タスクの設計やツールキットの構築にも取り組むことで、多種多様な自己教師あり学習特徴表現の比較評価を可能とし、音声変換において有効な自己教師あり学習特徴表現を実験的に明らかにしている。

第5章では、目標データが入手不可能な応用例の一つとして、病的音声と健常音声間で変換を行う応用技術について論じている。系列変換では分節的特徴に加えて韻律的特徴も変換される点と、フレーム変換では韻律的特徴は保持されつつ分節的特徴が変換される点に着目し、これら二つの変換を段階的に適用することで、声質などの静的な音声特徴を保持したまま、話し方などの動的な音声特徴を変換する手法を提案している。本手法に基づき、音声の個人性を保持したまま、病的音声から失われた健常音声へと変換する処理や、健常音声から病気の進行から予測される病的音声へと変換する処理を実現し、その有効性を示している。

第6章では、また別の応用例として、外国語音声における訛りを変換する応用技術について論じている。先に述べた系列変換とフレーム変換を段階的に適用する手法を拡張することで、訛りのある外国語音声を流暢な外国語音声へと変換する処理を実現している。比較評価を通して、各種拡張手法の特徴を明らかにするとともに、目標データが入手不可能な応用例における評価の問題点を指摘している。

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、HUANG Wen-Chin 君の学位論文は、事前学習という観点から、音声変換において学習データに求められる条件を大幅に緩和する基礎技術を生み出すとともに、目標データが入手不可能な例に対しても音声変換の適用を可能とする応用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 HUANG Wen-Chin 君が博士(情報学)の学位を授与されるに十分な資格を有するものと判定した。