

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 呉 宜樵

論 文 題 目

Incorporating Prior Knowledge on Speech Production Mechanism
into Neural Speech Waveform Generation

(深層音声波形生成における音声生成過程に関する事前知識の導
入)

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 出口 大輔

委 員 国立情報学研究所教授 山岸 順一

呉 宜樵 (WU, YI-CHIAO) 君提出の学位論文「Incorporating Prior Knowledge on Speech Production Mechanism into Neural Speech Waveform Generation (深層音声波形生成における音声生成過程に関する事前知識の導入)」は、高品質な統計的音声変換・合成処理を実現する上で重要な役割を担う深層音声波形生成技術に対して、完全データ駆動型方式であるがゆえに生じる問題点を解決するために、音声の物理的な生成過程に基づく事前知識を導入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、および、深層音声波形生成技術の課題を述べている。

第2章では、関連研究として、従来型の音声波形生成技術、ならびに、深層音声波形生成技術とハイブリッド型音声波形生成技術を説明するとともに、その関係性を論じている。また、統計的音声変換処理における代表的な基盤技術を説明している。

第3章では、本研究においてベースラインとする技術として、深層音声波形生成技術を導入した統計的音声変換システムについて論じている。本システムは、国際的音声変換技術評価会 Voice Conversion Challenge 2018 の教師なし学習タスクにおいて、話者類似度では第2位に相当するなど、平均を大幅に上回る高い性能が得られたことを示している。

第4章では、統計的音声変換における深層音声波形生成処理がしばしば不安定になる問題に着目し、頑健に動作する音声波形生成処理を実現する手法について論じている。深層音声波形生成処理に対して、従来型音声波形生成処理に応じた制約を導入することで、ベースラインシステムの性能を大幅に改善できることを示している。

第5章では、深層音声波形生成技術の本質的な問題として、合成音声に対する操作性の低下に着目し、準周期信号のモデル化に適した深層ネットワーク構造について論じている。拡張畳み込みネットワークにおいて、音声波形の基本周波数に応じて拡張サイズを動的に変化させる基本周波数適応型拡張畳み込みネットワークを提案するとともに、本手法を自己回帰型深層音声波形生成モデル WaveNet へと導入した Quasi-Periodic WaveNet (QPNet) を構築することで、基本周波数操作性能を改善できることを示している。

第6章では、合成音声に対する高い操作性、かつ、実時間合成処理を可能とする深層音声波形生成処理を実現するために、非自己回帰型深層音声波形生成モデル Parallel WaveGAN に基本周波数適応型拡張畳み込みネットワークを導入する手法について論じている。提案モデルである Quasi-Periodic Parallel WaveGAN (QPPWG) により、基本周波数操作性能の改善が得られることを示すとともに、基本周波数適応型拡張畳み込みネットワークは、音声生成過程における音源生成機能を効果的に学習できることを示している。

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、呉 宜樵君の学位論文は、深層音声波形生成技術に対して音声生成過程に関する事前知識を導入することで、その高い品質を保持しつつ、頑健性、操作性、処理速度の面から性能改善を達成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、論文提出者である呉 宜樵君が博士（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。