

令和 2 年 6 月 21 日現在

機関番号：13903

研究種目：挑戦的研究（萌芽）

研究期間：2017～2019

課題番号：17K20004

研究課題名（和文）ディープニューラルネットワークにより学習可能な調音モデルに基づいた音声合成

研究課題名（英文）Articulatory text-to-speech synthesis based on digital waveguide mesh driven by deep neural network

研究代表者

徳田 恵一（Tokuda, Keiichi）

名古屋工業大学・工学（系）研究科（研究院）・教授

研究者番号：20217483

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：あらゆる声質を柔軟に表現可能な音声合成システムを構築するため、人間の発声機構に則した調音モデルをディープニューラルネットワークに基づいたテキスト音声合成システムに組み込んだシステムを実現した。音声品質の改善のため、WaveNetを始めとするディープニューラルネットワークに基づいた音声波形生成手法との融合を図った。さらに、敵対的学習法に基づいて合成音声の声質や感情の制御する方法の検討を行った。

研究成果の学術的意義や社会的意義

スマートフォン、スマートスピーカー等、高度な情報機器が急速に普及しつつある中で、これらの情報機器と人間との間の情報交換の方法として音声インターフェースに期待がかかっている。これらの機械と自然な会話を行うためには、出力される合成音声は自在にあらゆる声質の音声を出力し、また、様々な感情表現を行うことが必須である。本研究はこのような人間のようにしゃべる機械の実現に貢献するものである。

研究成果の概要（英文）：In order to construct a speech synthesis system that can flexibly generate expressive speech, we have developed a deep neural network-based text-to-speech synthesis system that incorporates an articulatory model based on human speech production mechanism into a text speech synthesis system based on a deep neural network. In order to improve the voice quality, we attempted to combine it with WaveNet and other voice waveform generation methods based on deep neural networks. Furthermore, we examined the method of controlling the voice quality and emotional expression based on the generative adversarial training.

研究分野：音声情報処理

キーワード：音声合成 音声情報処理

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

(1) スマートフォン、スマートスピーカー等、高度な情報機器が急速に普及しつつある中で、これらの情報機器と人間との間の情報交換の方法として音声インタフェースに期待が掛かっている。

(2) これらの機械と自然な会話を行うためには、出力される合成音声は自在にあらゆる声質の音声を出力し、また、様々な感情表現を行うことが必須である。

### 2. 研究の目的

(1) 本研究の目的は、あらゆる声質を柔軟に表現可能な音声合成システムを構築するため、実際の人間の発声機構に則した調音モデルをテキスト音声合成システムに組み込み、その有用性を検証することにある。

### 3. 研究の方法

(1) まず始めに、ディープニューラルネットワークの枠組みの中で2次元デジタル・ウェーブガイド・メッシュ調音モデルを定式化し、調音モデルをテキスト音声合成システムに組み込むアイデアを定式化する。

(2) さらに、導出した数式をもとに調音モデルを組み込んだテキスト音声合成システムを構築する。

(3) 音声品質の限界を打破するために、WaveNet を始めとするディープニューラルネットワークに基づいた音声波形生成手法との融合を試みる。

(4) 波形生成手法において、音声波形モデルは調音モデルの構造を部分的に含んでいると考えられるため、調音モデルと音声波形モデルの関係性の調査と調音モデルと音声波形モデルの融合を目指し、合成音声の声質や感情の制御の検討を行う。

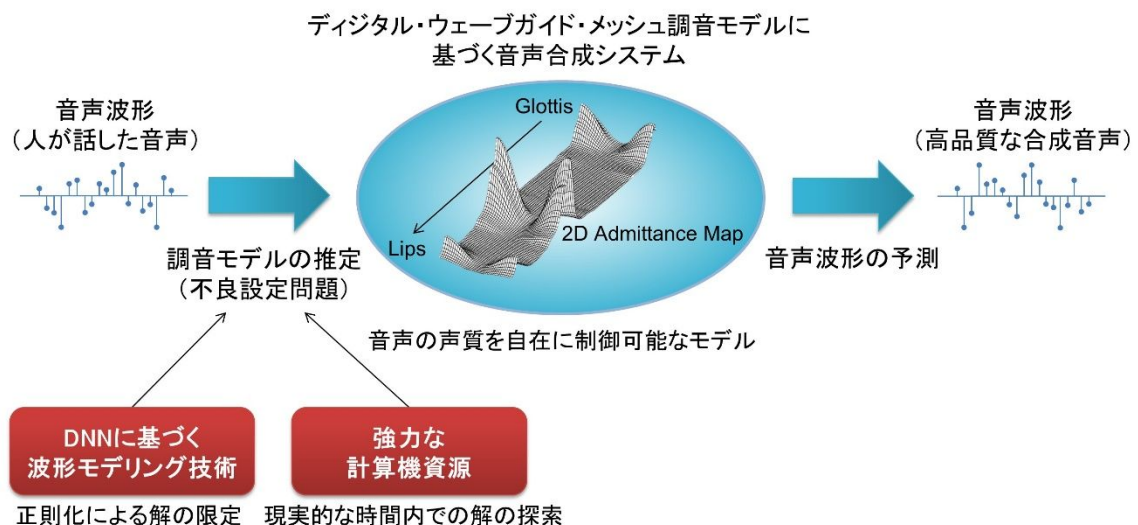


図 1 調音モデルを組み込んだディープニューラルネットワークに基づく音声合成

### 4. 研究成果

(1) ディープニューラルネットワークの枠組みの中で2次元デジタル・ウェーブガイド・メッシュ調音モデルを定式化し、調音モデルをテキスト音声合成システムに組み込むアイデアを数式として表現した。このような研究は他に例がなく、音声合成の研究を次の段階に推し進めるための1つの試みであると言える。さらに、申請者らは、導出した数式をもとに調音モデルを組み込んだテキスト音声合成システムを構築することに成功した。生成された音声のフォルマント周波数やスペクトルの大域的な構造は人間の発声のものと類似しており、明瞭な音声合成できることを確認した。また、実際に獲得された調音モデルの形状や時間的な変化を観測することにより、今後の研究を進めていく上で重要な知見を獲得することができた。観測された音声波形から直接調音モデルを推定することは、不良設定問題である。しかし、本研究ではニューラルネットワークを不良設定問題の正則化として利用することによって、これまで解くことが困難であった調音モデルの逆推定の問題を解決できる可能性を示した。

(2) これらの音声波形生成手法において、音声波形モデルは調音モデルの構造を部分的に含んでいると考えられる。このような観点から、調音モデルと音声波形モデルの関係性の調査と調音モデルと音声波形モデルの融合を目指し、合成音声の声質や感情の制御の検討を行った。ディープニューラルネットワークに基づく音響モデルの入力に話者コードやフレーズコードなどを加えることで、声質や感情を制御可能な音声合成システムを実現することができた。さらに敵対的

学習などの学習手法を適用することで、より高品質な合成音声を生成可能とした。また、発話スタイル等を表す潜在変数の階層化などにより、モデル化精度の向上を図った。

<引用文献>

Amelia J. Gully, Takenori Yoshimura, Damian T. Murphy, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda, "Articulatory text-to-speech synthesis using the digital waveguide mesh driven by a deep neural network," INTERSPEECH 2017, pp. 234-238, Stockholm, Sweden, August 2017.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda	4. 巻 26
2. 論文標題 Mel-cepstrum-based quantization noise shaping applied to neural-network-based speech waveform synthesis	5. 発行年 2018年
3. 雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	6. 最初と最後の頁 1173-1180
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2018.2818408	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計28件（うち招待講演 0件 / うち国際学会 10件）

1. 発表者名 Koki Senda, Yukiya Hono, Kei Sawada, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda
2. 発表標題 Singing Voice Conversion Using Posted Waveform Data on Music Social Media
3. 学会等名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Yukiya Hono, Shumma Murata, Kazuhiro Nakamura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda
2. 発表標題 Recent Development of the DNN-based Singing Voice Synthesis System -- Sinsy
3. 学会等名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Takato Fujimoto, Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda
2. 発表標題 Speech Synthesis Using WaveNet Vocoder Based on Periodic/Aperiodic Decomposition
3. 学会等名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Kento Nakao, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda
2. 発表標題 Speaker Adaptation for Speech Synthesis Based on Deep Neural Networks Using Hidden Semi-Markov Model Structures
3. 学会等名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Kei Sawada, Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda
2. 発表標題 The NITech text-to-speech system for the Blizzard Challenge 2018
3. 学会等名 Blizzard Challenge 2018 Workshop (国際学会)
4. 発表年 2018年

1. 発表者名 Jumpei Niwa, Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda
2. 発表標題 Statistical voice conversion based on WaveNet
3. 学会等名 2018 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (国際学会)
4. 発表年 2018年

1. 発表者名 藤本崇人, 吉村建慶, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 周期・非周期成分の分離に基づくWaveNetポコーダを用いた音声合成
3. 学会等名 日本音響学会2018年秋季研究発表会
4. 発表年 2018年

1. 発表者名 法野行哉, 村田舜馬, 中村和寛, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 Deep Neural Networkに基づく歌声合成システム -- Sinsy
3. 学会等名 日本音響学会2018年秋季研究発表会
4. 発表年 2018年

1. 発表者名 沢田慶, 吉村建慶, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 Blizzard Challenge 2018のためのNITechテキスト音声合成システム
3. 学会等名 日本音響学会2018年秋季研究発表会
4. 発表年 2018年

1. 発表者名 中尾健人, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 時間構造を考慮したニューラルネットワークに基づく音声合成における話者適応の検討
3. 学会等名 音声研究会
4. 発表年 2018年

1. 発表者名 島田基樹, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 隠れセミマルコフモデルの構造を用いたDNNに基づく音声合成における計算量削減手法の検討
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

1. 発表者名 藤本崇人, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 日本語End-to-End音声合成における入力言語特徴量の影響
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

1. 発表者名 大浦圭一郎, 中村和寛, 橋本佳, 南角吉彦, 徳田恵一
2. 発表標題 周期・非周期信号から駆動するディープニューラルネットワークに基づく音声ボコーダ
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

1. 発表者名 法野行哉, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 敵対的生成ネットワークを用いた歌声合成の検討
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

1. 発表者名 角谷健太, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 DNNに基づく感情音声合成のための敵対的学習の検討
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

1. 発表者名 Yukiya Hono, Kazuhiro Nakamura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda
2. 発表標題 Singing voice synthesis based on generative adversarial networks
3. 学会等名 2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (国際学会)
4. 発表年 2019年

1. 発表者名 吉村建慶, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 WaveNetにおけるメルケプストラムに基づくノイズシェーピング量子化法の適用
3. 学会等名 日本音響学会2017年秋季研究発表会
4. 発表年 2017年

1. 発表者名 丹羽純平, 吉村建慶, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 WaveNet-based voice conversion
3. 学会等名 日本音響学会2017年秋季研究発表会
4. 発表年 2017年

1. 発表者名 沢田慶, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 Blizzard Challenge 2017のためのNITechテキスト音声合成システム
3. 学会等名 日本音響学会2017年秋季研究発表会
4. 発表年 2017年



1. 発表者名 Amelia J. Gully, Takenori Yoshimura, Damian T. Murphy, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda
2. 発表標題 Articulatory text-to-speech synthesis using the digital waveguide mesh driven by a deep neural network
3. 学会等名 INTERSPEECH 2017 (国際学会)
4. 発表年 2017年

1. 発表者名 Kei Sawada, Kei Hashimoto, Keiichiro Oura, and Keiichi Tokuda
2. 発表標題 The NITech text-to-speech system for the Blizzard Challenge 2017
3. 学会等名 Blizzard Challenge 2017 Workshop (国際学会)
4. 発表年 2017年

1. 発表者名 Kei Sawada, Keiichi Tokuda, Simon King, and Alan W Black
2. 発表標題 The Blizzard Machine Learning Challenge 2017
3. 学会等名 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (国際学会)
4. 発表年 2017年

1. 発表者名 船戸涼平, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 DNN音声合成のためのパワーを考慮したトラジェクトリ学習
3. 学会等名 音声研究会
4. 発表年 2018年

1. 発表者名 吉村建慶, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 メルケプストラムに基づくノイズシェーピング量子化法のWaveNet音声合成への適用
3. 学会等名 音声研究会
4. 発表年 2018年

1. 発表者名 丹羽純平, 吉村建慶, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 WaveNetに基づく声質変換の検討
3. 学会等名 音声研究会
4. 発表年 2018年

1. 発表者名 沢田慶, 徳田恵一, Simon King, Alan W Black
2. 発表標題 Blizzard Machine Learning Challenge 2017の概要
3. 学会等名 日本音響学会2018年春季研究発表会
4. 発表年 2018年

1. 発表者名 法野行哉, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 隠れセミマルコフモデルの構造を利用したニューラルネットワークに基づく歌声合成
3. 学会等名 日本音響学会2018年春季研究発表会
4. 発表年 2018年

1. 発表者名 村田舜馬, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
2. 発表標題 DNNに基づく発声タイミングモデルを利用した歌声合成
3. 学会等名 日本音響学会2018年春季研究発表会
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	南角 吉彦  (Nankaku Yoshihiko)  (80397497)	名古屋工業大学・工学(系)研究科(研究院)・准教授   (13903)	
連携研究者	大浦 圭一郎  (Oura Keiichiro)  (20588579)	名古屋工業大学・工学研究科・研究員   (13903)	
連携研究者	橋本 佳  (Hashimoto Kei)  (10635907)	名古屋工業大学・工学研究科・准教授   (13903)	