様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　３ 年　 ６ 月 ２２ 日現在

機関番号：１３９０３
研究種目：基盤研究(C)（一般）
研究期間：2018～2020
課題番号：１８Ｋ１１１６３

研究課題名（和文）深層学習に基づいた新しい信号処理技術の確立と歌声および楽器音生成への応用

研究課題名（英文）Signal processing technology based on deep learning and application to singing voice and musical instrument sound generation

研究代表者

大浦　圭一郎（Oura, Keiichiro）

名古屋工業大学・工学（系）研究科（研究院）・研究員

研究者番号：２０５８８５７９

交付決定額（研究期間全体）：（直接経費）　　3,400,000円

研究成果の概要（和文）：歌声および楽器音を対象として，学習対象データの取捨選択方法の検討や，音声波形自体のモデリング手法の検討，楽譜から波形への直接変換が可能なEnd-to-End構造の検討など，従来のデジタル信号処理の枠を超えた音響モデリングの研究を進め，その一部を学術論文としてまとめ，投稿・発表することができた．その中でも，深層学習に基づいて周期・非周期信号からより自然な音声波形を生成する構成は，画像変換の分野で高い性能を示しているCycleGANのサイクル構造を応用するもので，日本音響学会の粟屋潔学術奨励賞や情報処理学会のマイクロソフト情報学研究賞を受賞するなどの成果をあげている．

研究成果の学術的意義や社会的意義
現状のほとんどの音声関連技術には，従来型のデジタル信号処理理論を基礎としており，従来型のデジタル信号処理理論は音声関連の研究分野では最も根本的な考え方として広く普及しているが，このような変換・処理で取り扱える枠組みの中に制限されていたため，モデル構造に関する過度の制約による性能限界があった．本研究は，このような状況にブレークスルーをもたらすため，近年急速に技術革新が進んでいる深層学習に基づいた音声波形の直接モデル化手法を新たに開拓しようとするものである．

研究成果の概要（英文）：For singing voices and instrument sounds, we proceeded research on acoustic modeling about automatic selection method of training data, modeling method of speech waveform itself, and end-to-end structure capable of direct conversion from musical score to waveform, etc. and publish some of them as academic papers. Among them, the waveform generation from periodic / aperiodic signals based on deep learning by applying the cycle structure of CycleGAN which show high performance in the field of image conversion has achieved results such as receiving the KIYOSHI AWAYA academic encouragement award from the acoustical society of Japan and the Microsoft informatics research award from the information processing society of Japan.

研究分野：音声合成

キーワード：信号処理　ディープラーニング　歌声合成　音声合成　楽器音生成

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

現状のほとんどの音声関連技術には，従来型のデジタル信号処理理論を基礎としている．ここでいうデジタル信号処理とは，音声のアナログ信号から変換された離散時間信号を線形時不変システムの出力と仮定することにより，フーリエ変換，z 変換等の理論に基づいて処理を行うものである．従来型のデジタル信号処理理論は音声関連の研究分野では最も根本的な考え方として広く普及しているが，このような変換・処理で取り扱える枠組みの中に制限されていたため，モデル構造に関する過度の制約による性能限界があった．2016 年，ニューラルネットワークを用いて音声波形を直接モデル化する手法として WaveNet が提案されたが，過去の音声サンプル列から次の音声サンプルを生成する自己回帰構造を持つため，合成時に並列演算ができないことから，実時間で合成できない問題があった．また，補助特徴量として与えたピッチ情報とは異なるピッチの波形が生成されることがあった．2017 年に提案された Parallel WaveNet では，自己回帰構造を持たないため実時間で合成できるものの，指定のピッチを再現できない場合がある問題はいまだ残っていた．

２．研究の目的

本研究では，このような状況にブレークスルーをもたらすため，近年急速に技術革新が進んでいる深層学習に基づいた，音声波形の直接モデル化手法を新たに提案しようとするものである．歌声合成，楽器音生成への応用により有効性を確かめるが，それだけに留まらず，従来型のデジタル信号処理の枠を超えた新しい信号処理分野の開拓を目指すものである．

３．研究の方法

あらゆる歌声・楽器音の分析・合成を可能とするための深層学習には，歌声・楽器音の声質・音質空間を埋め尽くす超巨大データが必要になる．超巨大データとして，音楽 SNS アプリにおける計１００万時間を超えるシングルトラックを利用することを考え，まずは各シングルトラックを音声，歌声，楽器音毎に選別する手法の確立に取り組む．このデータは実サービスによる音声波形であるため，歌声以外にも話し声や様々な楽器音，さらに収録時の回り込みによるエコーや伴奏が入り込んでしまっているデータや，アップロード時の接続障害により途切れているデータ等から構成されている．このような多種多様なデータから，本研究にてターゲットとする歌声・楽器音を抽出する必要がある．ここでは，一部のデータに対してのみ手作業によるラベル付与を行い，ブートストラップ的な半教師学習により全てのデータの自動選別を行う．この手法は，まず手作業によるラベルが付与された一部のデータを元に初期モデルを学習し，初期モデルにより残りのデータを自動判定することで学習データを増やし，モデルの再学習と自動判定の繰り返しにより検出精度を高める手法である．音響特徴量を DNN でモデル化することで判定に用いるが，この段階では従来のデジタル信号処理に基づくメルケプストラムや自己相関などの音響特徴量を用いる．声質空間・楽器音空間を埋め尽くす超巨大データを用いて，あらゆる歌声・楽器音を分析・合成することが可能な新しい信号処理技術の確立を目指し，深層学習を用いた歌声・楽器音の合成システムを構築する．特徴パラメータの列 o から波形 x を予測する DNN の入力層に用いる入力特徴量には過去の出力波形列を用いるが，第一段階としてはそれ以外に，従来のデジタル信号処理によるメルケプストラムや自己相関なども補助特徴量として用いる．また，分解されていたモデル構造を単一の枠組みでモデル化すること(end-to-end)が可能になってきたことから，第二段階として，DNN に基づく歌声・楽器音の合成システムにおいて，end-to-end の構造を構築する．それぞれ別々に最適化されていた，楽譜特徴 l から特徴パラメータの列 o を予測する DNN と，特徴パラメータの列 o から波形 x を予測する DNN を統合・最適化を行うことにより，楽譜特徴 l から波形 x への変換を直接的に解決する end-to-end の構造を実現するものである．単一モデルへの統合により，従来のデジタル信号処理における各種補助特徴量 o は隠れパラメータとなり，従来のデジタル信号処理の枠に捉われることのない，高精度な音声波形表現になると考えられる．以上の枠組みにより，楽譜から歌声・楽器音への変換という問題を直接的に取り組み，自然な波形の生成を目指す．

４．研究成果

歌声および楽器音を対象として，学習対象データの取捨選択方法の検討や，音声波形自体のモデリング手法の検討，楽譜から波形への直接変換が可能な End-to-End 構造の検討など，従来のデジタル信号処理の枠を超えた音響モデリングの研究を進め，その一部を学術論文としてまとめ，投稿・発表することができた．その中でも，深層学習に基づいて周期・非周期信号からより自然な音声波形を生成する構成は，日本音響学会の粟屋潔学術奨励賞（2019 年 9 月）や情報処理学

会のマイクロソフト情報学研究賞（2020 年 3 月）を受賞するなどの成果をあげている．この方式は，明示的に周期信号と非周期信号の列を入力することで，対応する音声サンプルの列を一度に生成するものである．提案モデルを学習する際，画像変換の分野で高い性能を示している CycleGAN のサイクル構造を応用することで，音声波形と周期・非周期信号の相互変換の同時学習を行った．主観評価実験から，サイクル構造を持つ学習が合成歌声の品質を向上させることを示した．この方式にて，深層学習に基づいて周期・非周期信号から音声波形を生成する構成で非常に自然な音声を生成することができたため，さらに，周期信号と非周期信号の相互依存性の検証についても行った．構成としては，周期信号と非周期信号を同時に入力・変換するものや，個別に入力・変換するもの，さらに周期波形が非周期波形に影響を及ぼすことを仮定するものなどについて，比較・検討を行った．一部の実験結果では，周期波形と非周期波形の依存性をあえて考慮しないほうがより頑健に駆動することがわかるなど，学習対象データをうまく効率的にモデリングできることを示した．

５．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計21件（うち招待講演　2件／うち国際学会　9件）

１．発表者名 村田舜馬，藤本崇人，法野行哉，高木信二，橋本佳，大浦圭一郎，南角吉彦，徳田恵一
２．発表標題 楽譜時間情報を用いたアテンション機構に基づく歌声合成の検討
３．学会等名 日本音響学会2019年秋季研究発表会
４．発表年 2019年

１．発表者名 Yukiya Hono, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, 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 Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, Keiichi Tokuda
２．発表標題 Speaker-dependent WaveNet-based delay-free ADPCM speech coding
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 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・非周期信号を用いたDNNに基づくリアルタイム音声ボコーダ
３．学会等名 情報処理学会研究報告
４．発表年 2019年

１．発表者名
大浦圭一郎，高木信二，中村和寛，橋本佳，南角吉彦，徳田恵一

２．発表標題
周期・非周期信号を用いた敵対的生成ネットワークに基づくリアルタイム音声ボコーダ

３．学会等名
日本音響学会2019年秋季研究発表会

４．発表年
2019年

１．発表者名
Keiichiro Oura, Kazuhiro Nakamura, Kei Hashimoto, Yoshihiko Nankaku, Keiichi Tokuda

２．発表標題
Deep neural network based real-time speech vocoder with periodic and aperiodic inputs

３．学会等名
10th ISCA Speech Synthesis Workshop (SSW10)（国際学会）

４．発表年
2019年

１．発表者名
和田蒼汰，法野行哉，高木信二，橋本佳，大浦圭一郎，南角吉彦，徳田恵一

２．発表標題
歌声合成におけるニューラルボコーダの比較検討

３．学会等名
音声研究会

４．発表年
2019年

１．発表者名
大浦圭一郎

２．発表標題
統計的歌声合成技術とその実用化

３．学会等名
日本AI音楽学会（招待講演）

４．発表年
2019年

１．発表者名
大浦圭一郎

２．発表標題
統計的パラメトリック音声合成技術とその実用化

３．学会等名
情報処理学会音学シンポジウム（招待講演）

４．発表年
2019年

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
APSIPA 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
APSIPA 2018（国際学会）

４．発表年
2018年

１．発表者名
Takahto 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
APSIPA 2018（国際学会）

４．発表年
2018年

１．発表者名
Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda

２．発表標題
WaveNet-based zero-delay lossless speech coding

３．学会等名
SLT 2018（国際学会）

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
日本音響学会秋季研究発表会

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ボコーダを用いた音声合成

３．学会等名
日本音響学会秋季研究発表会

４．発表年
2018年

１．発表者名
Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda

２．発表標題
Speaker-dependent WaveNet-based delay-free adpcm speech coding

３．学会等名
ICASSP 2019（国際学会）

４．発表年
2019年

１．発表者名
Yukiya Hono, 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
ICASSP 2019（国際学会）

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に基づく音声ボコーダ

３．学会等名
日本音響学会春季研究発表会

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
日本音響学会春季研究発表会

４．発表年
2019年

１．発表者名
法野行哉，高木信二，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ニューラルボコーダによる音声波形のモデル化の検討

３．学会等名
日本音響学会春季研究発表会

４．発表年
2021年

１．発表者名 法野行哉，高木信二，橋本佳，大浦圭一郎，南角吉彦，徳田恵一
２．発表標題 DNNに基づく音声ボコーダにおける周期・非周期成分のモデル化の検討
３．学会等名 日本音響学会秋季研究発表会
４．発表年 2020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

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関