

研究報告書

「適応型制限ボルツマンマシンの複素拡張に基づくボコーダー不要な非パラレル声質変換」

研究期間：平成 29 年 10 月～平成 31 年 3 月

研究者番号：50148

研究者：中鹿 亘

1. 研究のねらい

パーキンソン病や、咽頭摘出を必要とする頭頸部癌、構音障害等、会話に障害を及ぼす疾病は多数存在する。現在医療現場で広く実用されているのは喉頭摘出者に対する電気喉頭装置だが、発声の自然さや明確さが罹患前より大きく劣化する。然る中、声質変換技術を医療器具に導入することで、「自分の声」での会話や、より多様な疾患への貢献が期待されている。当技術にはパラレル声質変換と非パラレル声質変換があり、前者を適用した医療器具の臨床実験は既に行われているが、大量のパラレルデータが必要となり患者に過度の負担が掛る。一方、後者では自由発話で学習可能な為患者への負担が大幅に軽減されるが、現状の技術ではボコーダーが必須であるため、変換音声の品質が低下してしまう。本研究では、子どもから高齢者、疾患や課題を抱える人の音声コミュニケーションに貢献することを目的とし、パラレルデータ不要かつ高品質な声質変換モデルの開発を行う。

本研究により、声質変換精度が向上すれば応用範囲が広がり、医療・福祉・産業など様々な場面において新たなコミュニケーションが生まれることが予想される。例えば、保育園で保育士が母親の声となって園児に話しかける、発達障害の子どもが心を開いて積極的に会話するきっかけ作りなどが考えられる。子どもから高齢者、疾患や課題を抱えている人が、周囲と自然な音声コミュニケーションをとることを支援し、社会的孤立を防ぐことに貢献することが本研究の一つの狙いである。産業界においても本技術の応用は多岐に渡り、好きな歌手になりきって歌を歌う、オリジナルの役者の声で洋画の吹き替えを行う、亡くなった声優を復活させるなどといった、デジタルコンテンツ業界のニーズに応えることができる。

一般に音声信号処理で広く用いられているメルケプストラムは、健常者の発する音声を対象にして圧縮された音響特徴量であり、そのまま構音障害者の音声に対して用いるのは不適切である。構音障害者の音声は、メルケプストラムで表現できない調音が存在する。そこで、メルケプストラムの代わりに直接複素スペクトルを直接表現することのできる複素制限ボルツマンマシンを用いて、構音障害者音声から潜在的な特徴量を抽出することが重要となる。

2. 研究成果

(1) 概要

本研究の目的は、機械学習手法の複素数拡張モデルを用いて音声の複素スペクトルを直接表現することにより、高品質な非パラレル声質変換手法を実現することである。本研究期間では、主に、著者がこれまでに提案した複素制限ボルツマンマシン(複素 RBM)の改良と、適応型制限ボルツマンマシン(ARBM)を用いた声質変換手法の改良を行った。まず、複

素 RBM の改良について具体的に述べる。複素 RBM は可視・不可視の二層の双方向ニューラルネットワークである従来の RBM を複素数へ拡張したモデルであり、これを用いれば音声の複素スペクトルから直接特徴抽出や音声復元が可能となる。これまでの複素 RBM に関する研究では人工的に生成した複素数データのみでモデル検証を行っていたが、そのまま音声モデリングに適用すると、収束速度と音声復元精度に課題が生じていた。そこで本研究では複素 RBM による音声モデリング手法を改良するアプローチとして、複素 Adam による最適化手法、複素 PCA による次元圧縮、複素最尤系列生成による音声スペクトル生成手法をそれぞれ提案し、実験的に有効性を確認した。また、音声スペクトルの持つ時間相関を適切に表現するために、時間的に隣接する隠れ層間の接続を追加した拡張モデル(時相関複素 RBM)および可視素子の自己回帰性を表現する拡張モデル(自己回帰複素 RBM)を提案し、有効性について検証した。次に、ARBM を用いた声質変換の改良に関する研究について述べる。ARBM は上記 RBM に対して話者識別素子を追加することで、音声の中の話者の制御を可能にした拡張モデルであり、ARBM を用いると任意の話者から任意の話者へ声質変換することが可能となる。本研究ではさらに自然な変換音声を得るため、入力話者の自然音声とモデル音声の残差信号を変換音声に加える手法(差分 ARBM)の提案および ARBM の複素数拡張(複素 ARBM)を定式化した。またそれ以外にも、ボルツマンマシンの基礎拡張モデルとして、LSTM のように長・短期記憶構造を持つモデル(LSTBM)、2つの異なる特徴量間を多層にかつ双方向に変換可能なモデル(DRM)をそれぞれ提案し、実験的に有効性を確認した。

(2) 詳細

研究テーマ A「複素 RBM を用いた音声モデリング手法の確立」

これまで著者が提案した複素 RBM は、そのまま音声の複素スペクトルモデリングに適用すると、収束速度が遅く、また音声復元の品質が十分とは言えなかった。そこで本研究では、複素 Adam による最適化手法、複素 PCA による次元圧縮、複素最尤系列生成による音声スペクトル生成手法の3つの改善手法を提案し、有効性を確認した。評価実験では、人工データ・音声の実データ共に、従来の複素最急降下法と比較して、複素 Adam によるパラメータ更新時の収束速度が向上することが確認できた。また、下記表に示すように、改善された複素 RBM(CRBM)と従来の RBM を用いて音声の複素スペクトルを分析・合成すると、提案法の方が客観評価基準である PESQ 値において優れていることが確認できた。音声の分析合成では WORLD と呼ばれる手法が高品質であり一般的によく用いられているが、WORLD の PESQ 値は 2.86 であり、学習データがある程度増やせば複素 RBM の方が品質面で優れていることも確認できた。

学習データを増加させたときの精度比較 (PESQ 値)

# of sentences	CRBM	RBM
50	2.81	2.66
100	2.85	2.71
200	2.94	2.80

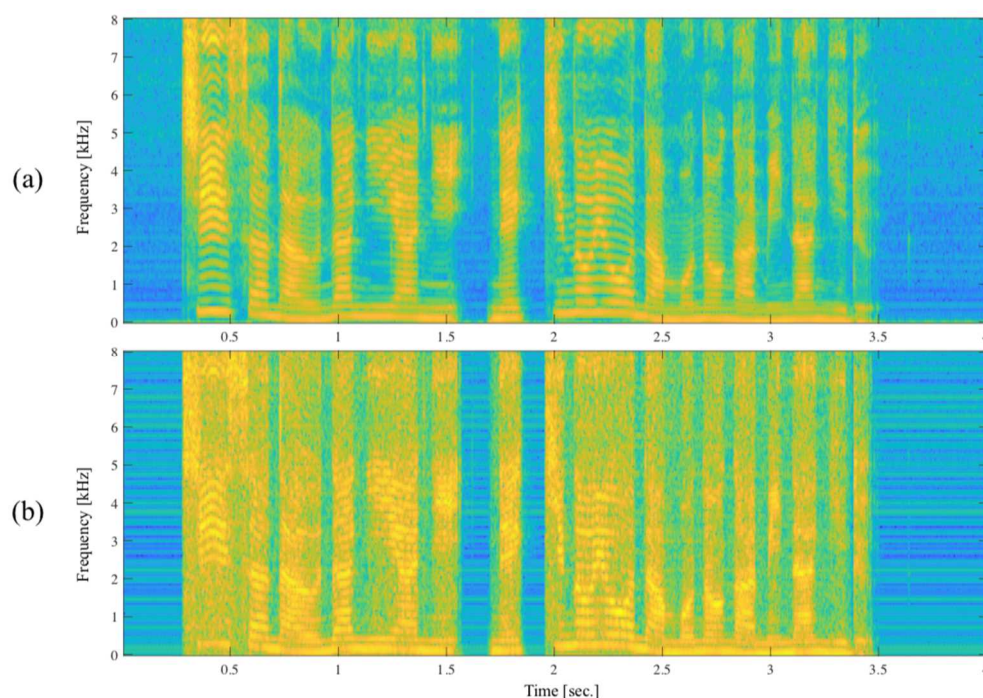
この研究成果として、英語論文誌 IEEE/ACM Transactions on Audio, Speech, and Language Processing へ論文を投稿し、受理された(主な研究成果リスト1)。

研究テーマ B「複素 RBM の潜在変数時系列表現モデル(複素 TRBM)」

これまでの複素 RBM では音声複素スペクトルの各フレームに対して独立にモデリングされていたが、複素スペクトル系列は時相関を持つため定義として不十分であった。そこで本研究では時間的に隣接する隠れ素子間の接続を考慮した拡張モデル(複素 TRBM)を新たに定義し、有効性を確認した。複素 RBM と比較して、時間接続を追加した複素 TRBM は、デコード性能はほとんど変わらないが、より収束速度が向上し、隠れ素子をよりスパースに導ける機能を持つことが確認できた。本成果は「リカレント構造を持つ複素制限ボルツマンマシンによる複素スペクトル系列モデリング」というタイトルで 2018 年 2 月に開催された第 120 回音声言語情報処理研究会にて発表済みである(主な研究成果リスト5)。

研究テーマ C「複素 RBM の自己回帰可視変数表現モデル(自己回帰複素 RBM)」

音声信号は自己回帰性を持つため、複素スペクトル系列もまた複素数の自己回帰性を持つデータとして近似できる。本研究テーマは自己回帰性を持つ複素数データ系列を表現するモデル(自己回帰複素 RBM)を提案し、実験的にその有効性を確認した。評価実験では、PESQ 値が 3.07 となり、従来の複素 RBM や静的な分析再合成手法と比較して、高品質な音声符号化・復号化が可能であることを示した。下図はオリジナル音声の対数スペクトル(a)及び自己回帰複素 RBM によって復元された対数スペクトル(b)を示しており、調波構造や微細構造が視覚的に確認され、音声の特徴を捉えられていることが分かる。



本成果は「音声スペクトル系列の自己回帰性を考慮した複素 RBM の拡張」というタイトルで 2018 年 9 月に開催された 2018 年度日本音響学会秋季研究発表会にて発表済みである (主な研究成果リスト 2)。

時間接続を追加した複素 TRBM は、デコード性能はほとんど変わらないが、より収束速度が向上し、隠れ素子をよりスパースに導ける機能を持つことが確認できた。本成果は「リカレント構造を持つ複素制限ボルツマンマシンによる複素スペクトル系列モデリング」というタイトルで 2018 年 2 月に開催された第 120 回音声言語情報処理研究会にて発表済みである (主な研究成果リスト 3)。

3. 今後の展開

本研究を通じて、従来基礎研究レベルに留まっていた複素 RBM を、実際に音声の複素スペクトルモデリングに応用し、実用的に十分使用できるレベルにまで確立させることができた。本研究のアイデアは画像の複素スペクトルや医療画像、アレー信号、衛星画像など、他の複素数データを表現する際にも適用でき、複素数データから直接潜在的な特徴を抽出させる複素 RBM を用いれば様々なアプリケーションへ応用することができる。例えば周波数を変えて撮影した衛星画像から地震予測をしたり、MRI 画像から悪性腫瘍を検出したりすることが考えられる。

4. 自己評価

・研究目的の達成状況

基礎研究レベルに留まっていた複素 RBM を応用レベルにまで改善させた点については十分な成果を達成できたと言える。一方で、当初の最終目的であった声質変換については、いくつか課題点が残された。

・研究の進め方

大規模演算が可能な計算機サーバ及び GPGPU を導入したことにより、研究を迅速に進めることができた。

・研究成果の科学技術及び学術・産業・社会・文化への波及効果

前述の通り、複素 RBM(とその派生)を応用レベルに改善させたことで、複素数を取り扱う全ての研究分野及びアプリケーションで波及効果が見られると想定している。

・研究課題の独創性・挑戦性

複素数を扱う機械学習や信号処理は実数のものと比べてほとんど研究されておらず、複素勾配の計算が必要になるなど、本研究は前提から定義を見直さないといけなため挑戦性のある研究でもある。また、数多く研究されているフィードフォワード型のニューラルネットワークを用いず、理論に直結するボルツマンマシンに着眼することは独創的である。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

- | |
|---|
| 1. 中鹿 亘、高木 信二、山岸 順一. Complex-Valued Restricted Boltzmann Machine for Direct Speech Parameterization from Complex Spectra. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2018 年, 27 巻 2 号, 244-254. |
| 2. 中鹿 亘. LSTBM: A Novel Sequence Representation of Speech Spectra Using Restricted Boltzmann Machine with Long Short-Term Memory. Proceedings of Interspeech. 2018 年, 2529-2533. |
| 3. 中鹿 亘. 音声スペクトル系列の自己回帰性を考慮した複素 RBM の拡張. 日本音響学会 2018 年秋季研究発表会. 2018 年, 1135-1138. |
| 4. 中鹿 亘. Deep Relational Model: A Joint Probabilistic Model with a Hierarchical Structure for Bidirectional Estimation of Image and Labels. IEICE Transactions on Information and Systems. 2018 年, E101-D 巻 2 号, 428-436. |
| 5. 中鹿 亘. リカレント構造を持つ複素制限ボルツマンマシンによる複素スペクトル系列モデリング. 第 120 回音声言語情報処理研究会. 2018 年, 1-5. |

(2) 特許出願

研究期間累積件数:1 件

発 明 者: 中鹿 亘

発明の名称: 符号化装置、符号化方法およびプログラム

出 願 人: 電気通信大学

出 願 番 号: 特願 2018-31875

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

日本音響学会粟屋学術奨励賞受賞、長・短期記憶構造を持つ拡張ボルツマンマシンの検討、2018/9/13.