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　５ 年　 ６ 月 １４ 日現在

機関番号：１２６１２
研究種目：若手研究
研究期間：2019～2022
課題番号：１９Ｋ２０６１８

研究課題名（和文）対面コミュニケーションと同等に感情を伝えるための音声強調処理法の開発

研究課題名（英文）Development of speech enhancement methods for conveying emotions equivalent to face-to-face communication

研究代表者

岸田　拓也（Kishida, Takuya）

電気通信大学・大学院情報理工学研究科・研究員

研究者番号：８０８２７９０７

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：音声符号化技術と通信技術を利用した音声のみによるコミュニケーションでは、視覚情報等の利用ができないために、感情・意図・態度・個人性などの非言語情報を正確に伝えることが困難となる。本研究では、非言語情報と音声の音響特徴量との関係をモデル化できるようなニューラルネットワークを考案し、音声の非言語情報における個人性や感情をを変換・強調する手法について検討した。ボルツマンマシンやその関連手法を利用して、モデルの学習に用いていない話者間で個人性の変換が可能な手法や、個人性と感情を同時に変換する手法、声質を因子に分解して因子の操作によって声質の印象を変換する手法などを提案することができた。

研究成果の学術的意義や社会的意義
本研究で得られた実験結果は、ボルツマンマシンやその関連手法が音声の音響特徴量と非言語情報との関係を表現するのに有効であることを示している。また、画像生成分野で目覚ましい成功を挙げている拡散確率モデルを声質変換課題に適用することに関する研究成果や調査結果は、音声コミュニケーションで声質変換技術をより柔軟に利用するための新たな手法の着想や知見につながった。

研究成果の概要（英文）：In the context of speech communication using communication technologies, accurately conveying paralinguistic information such as emotions, intentions, attitudes, and speaker identities becomes challenging due to the absence of visual and other relevant cues. In this study, we developed a neural network capable of modeling the relationship between paralinguistic information and acoustic features of speech. Our research focused on exploring techniques to convert and enhance speaker identities and emotions. By employing the Boltzmann machine and related models, we were able to propose several approaches. These include a method that enables speaker identity conversion between individuals not included in the model's training, a method that concurrently converts speaker identities and emotions, and a method that decomposes voice into factors, allowing for voice impression conversion through factor manipulation.

研究分野：音声信号処理

キーワード：音声信号処理　機械学習　声質変換　感情音声変換

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

言語情報を正しく伝えることがコミュニケーションの重要な要素であるが、言語情報と同時に「感情・意図・態度・個人性などの非言語情報が正しく伝わること」もコミュニケーションを成立させるために必要な要素である。音声符号化技術と通信技術の発展はコミュニケーションにおいて対面が必要であるという制約を取り除いたものの、音声のみのコミュニケーションでは表情・視線・身振り手振りによる視覚情報や身体の接触による触覚情報の利用ができないために、上述の非言語情報を伝えることがより困難となる。

しかし、音声通信に用いられる音声符号化技術の性能は、どれだけ単位時間あたりの情報量を下げられたかと、国際電気通信連合電気通信標準化部門の勧告に準じた方法で調べられた主観的な明瞭度がどれだけ高いかの２点だけで評価されることがほとんどである。つまり、現状の音声符号化技術では非言語情報がどれだけ伝わるのかを正しく評価し、改善する方法が検討されていない状況である。よって、音声符号化によってどの程度対面コミュニケーションで伝わるべき非言語情報が損なわれているのかを明らかにし、情報損失を補うための新しい音声符号化法が求められる。

２．研究の目的

本研究では、「非言語情報の一つである感情が音声の音響的特徴とどのように結びつくのかを明らかにし、対面コミュニケーションと同等に感情を伝えるための音声の強調処理法を開発する」ことを目的とした。

３．研究の方法

音声が伝える非言語情報は音声の音響特徴量がもつ時間的、周波数的な要素と複雑に関係する。そのため、その関係を効果的にモデル化できるようなニューラルネットワークを考案し、非言語情報の操作や強調が可能なニューラルネットワークベースの音声信号処理法を開発する、という方針で研究を行った。具体的には、ボルツマンマシンをはじめとする、エネルギーベースモデルやその関連手法となる生成モデルを使って、音声の非言語情報における個人性（誰が話しているかの情報）や感情をモデル化し、それらの情報を変換・強調する手法を提案した。提案したモデルの性能は客観的評価と主観的評価の両方で定量的に評価した。

４．研究成果

2019 年度の成果は大きく分けて４つ挙げられる。第１に、音声コミュニケーションは言語学的―生理学的―音響学的段階に分解されるという概念をヒントにして、ボルツマンマシンで各段階の連鎖を表現するモデルを考案し、これを話者性の声質変換技術に応用する手法 (speech chain voice conversion: speech chain VC)を提案した。第２に、speech chain VC に、話者認識分野の技術を組み合わせることで、声質変換モデルの学習には用いていない話者同士でも変換可能な手法を提案した（図１）。第３に、音声が伝える話者性と感情の情報を別の話者および感情として知覚される様に同時に変換する技術 (multi-domain ARBM) を提案した。そして第 4 に、画像のスタイル変換技術で用いられる深層ニューラルネットワークモデルである fader networks を改良し、楽譜情報を維持したまま別の楽器音に聞こえる様に変換する手法を提案した。

2020 年度は、初年度に引き続き、音声の非言語情報のモデル化の研究を行った。初年度に提案した Speech chain VC に関して追加検証を行い、Speech chain VC モデルを構成するネットワークで生理学的レベルを表現していると位置付けたユニットの一部が、実際に音声における生理学的レベルと関連の深い弁別素性に対応することを確認した。また、音声の音響的特徴をモデルで表現するときに話者と言語内容（音韻）の相互作用を考慮した方が話者性を変換する性能が高まると考え、話者・音韻の相互作用を考慮するモデルとして Cluster ARBM をベースとする声質変換モデルを提案した。この他に、音声特徴量系列内の長期的な時間依存関係を表現することを目的としてボルツマンマシンに自己注意機構を持たせた Attention RBM を新たに提案した。非言語情報のような発話全体を通して現れるような音響的特徴を表現するのに適したネットワークモデルを設計することができた。上記研究に関して、論文誌１件、国際会議１件、国内研究会２件の研究発表を行なった。

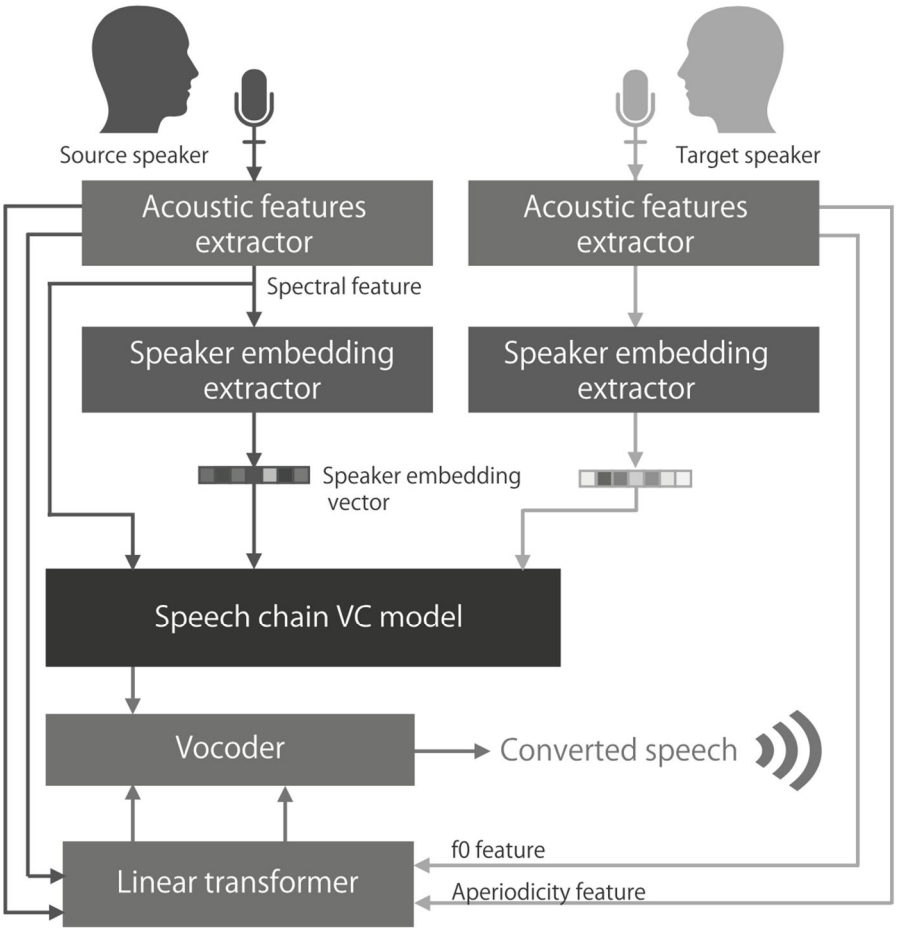

図１：学習には用いていない話者間でも個人性が変換可能な提案モデル。

2021 年度は、音声の音響的特徴をモデルで表現する際の表現力を高めるために、昨年度までに用いてきたボルツマンマシンと関連の深い、エネルギーベースモデルを用いた手法を検討した。エネルギーベースモデルとその関連モデルは高精細な画像が生成可能なモデルとして近年画像生成分野で注目を集めている。音声の音響特徴量の時系列データも、周波数と時間の二つの次元からなる空間上に描かれる画像のようなものとみなすことができる。したがって画像生成分野で成功を収めている上述の手法が有効であると考えた。提案したモデルは、音声信号においてどのような音響的特徴が生じやすいかを学習するるため、ノイズ混じりの音声からノイズの除去ができる。また、学習時に使う音声の話者の声質も学習するため、声質変換にも利用可能である。ノイズ除去および声質変換の性能を客観指標で評価したところ、従来手法を超える性能を発揮できると確認できた。この他に、音声を聴いたときに感じる「声の印象」を変える声質変換手法や、顔画像を基に、その顔画像にふさわしい声質に元音声を変換する手法などを検討した。上記研究に関して、国内研究会 8 件の研究発表を行なった。

2022 年度はこれまでの音声の非言語情報のモデル化とモデルによる非言語情報の変換と強調に関する研究をまとめ、今後の研究発展のための調査や学外発表をおこなった。具体的には、2 件の国際学会発表および 8 件の国内学会・研究会で研究の成果を発表した。

昨年度までの研究で声質変換手法として用いてきたボルツマンマシンについて、これまでとは異なる考え方に基づく方法として、ボルツマンマシンの自由エネルギーの最小化による声質変換手法を新たに提案し、国際学会にて発表した。この手法によって、変換元の話者が誰であっても、目標とする話者の声質に変換可能なモデルをボルツマンマシンで実現できるようになった。また、声質における個人性の情報はそれを構成するいくつかの因子からなるという仮定のもと、それらの因子に個人性の情報を分解し、再構築することで、因子による操作が可能な声質変換手法(VQ-SplitterNetVC)も新たに提案した（図２）。この研究成果も国際学会にて発表した。

この研究によって、音声の音響特徴量を多変量解析によって多次元空間上で表現し、その空間上の位置と音声の印象との関係を心理実験によって明らかにすることで、特定の印象が想起されやすい音声強調処理法を開発する、という当初の研究計画で想定していたものを、一つの深層生成モデルである程度まで実現できることを示せたと考える。

これらの代表的な研究成果のほかにも、高精細な画像を生成可能なモデルとして注目されている、拡散確率モデルを声質変換に利用することを検討し、音声の声質変換課題にも一定の効果

があることを確認できた。声質変換課題に拡散確率モデルを適用することに関する研究成果や調査結果は、音声コミュニケーションで声質変換技術をより柔軟に利用するための新たな手法の着想や知見につながった。

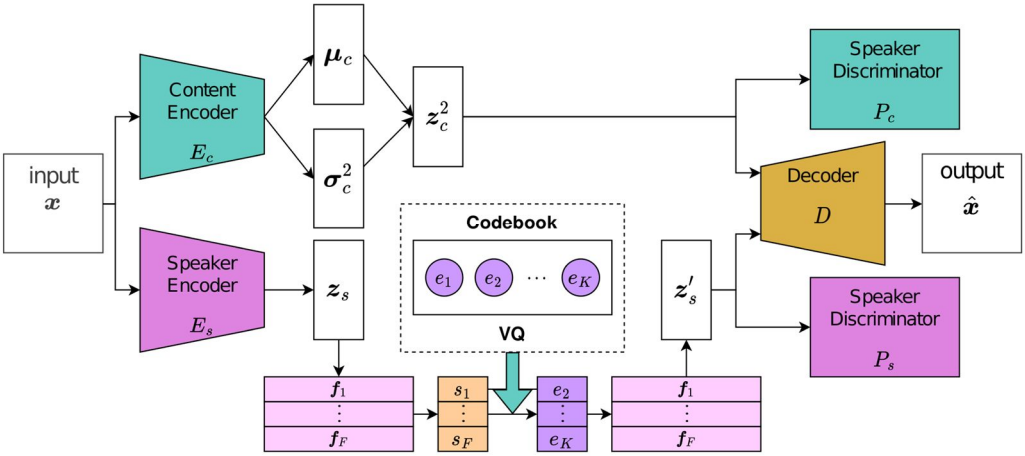

図２：VQ-SplitterNetVC **モデル**

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　1件）

１．著者名 KISHIDA Takuya、NAKASHIKA Toru	４．巻 E103.D
２．論文標題 Speech Chain VC: Linking Linguistic and Acoustic Levels via Latent Distinctive Features for RBM-Based Voice Conversion	５．発行年 2020年
３．雑誌名 IEICE Transactions on Information and Systems	６．最初と最後の頁 2340～2350
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transinf.2020EDP7032	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計29件（うち招待講演　0件／うち国際学会　4件）

１．発表者名 岸田拓也，中鹿亘
２．発表標題 入力特徴量で条件づけた拡散確率モデルによるパラレル声質変換
３．学会等名 日本音響学会音声研究会
４．発表年 2023年

１．発表者名 羽賀洋克，矢田部浩平，岸田拓也，中鹿亘
２．発表標題 振幅重み付けエネルギー関数を用いたボルツマンマシンによる位相復元
３．学会等名 日本音響学会2023年春季研究発表会
４．発表年 2023年

１．発表者名 奥田耕平 岸田拓也，中鹿
２．発表標題 Dual Diffusion Implicit Bridgesを用いた話者間の匿名性を担保した声質変換
３．学会等名 日本音響学会2023年春季研究発表会
４．発表年 2023年

１．発表者名
許誠，岸田拓也，中鹿亘

２．発表標題
Speechsplit を用いたイントネーション・リズム・発音の矯正による 外国語アクセント変換

３．学会等名
日本音響学会2023年春季研究発表会

４．発表年
2023年

１．発表者名
Kishida, T., & Nakashika, T.

２．発表標題
Non-parallel voice conversion based on free-energy minimization of speaker-conditional restricted boltzmann machine.

３．学会等名
Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) IEEE（国際学会）

４．発表年
2023年

１．発表者名
Isako, T., Onishi, K., Kishida, T., & Nakashika, T.

２．発表標題
Controllable voice conversion based on quantization of voice factor scores.

３．学会等名
Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) IEEE（国際学会）

４．発表年
2023年

１．発表者名
岸田拓也，中鹿亘

２．発表標題
条件付き制限ボルツマンマシンの平衡化傾向を利用したノンパラレル声質変換

３．学会等名
日本音響学会2022年秋季研究発表会

４．発表年
2022年

１．発表者名
井硲巧，大西弘太郎，岸田拓也，中鹿亘

２．発表標題
話者因子係数の量子化に基づく声色制御可能な話者変換

３．学会等名
日本音響学会2022年秋季研究発表会

４．発表年
2022年

１．発表者名
越森道貴，嵯峨山茂樹，岸田拓也，中鹿亘

２．発表標題
F0適応ラグ窓を用いた音声分析系の精緻化

３．学会等名
音学シンポジウム2022

４．発表年
2022年

１．発表者名
古田翔太郎，岸田拓也，中鹿亘

２．発表標題
制限ボルツマンマシンを用いた独立低ランク行列分析に基づくブラインド音源分離

３．学会等名
音学シンポジウム2022

４．発表年
2022年

１．発表者名
平本佳弘，嵯峨山茂樹，岸田拓也，中鹿亘

２．発表標題
LSP周波数間隔のクロスエントロピー誤差最小化に基づくVAE声質変換

３．学会等名
音学シンポジウム2022

４．発表年
2022年

１．発表者名
王庭輝，岸田拓也，中鹿亘

２．発表標題
リズムスタイルを考慮したFader Networksに基づく外国語学習者の発音変換

３．学会等名
日本音響学会2022年春季研究発表会

４．発表年
2022年

１．発表者名
Zhou Yujin，岸田拓也，中鹿亘

２．発表標題
TTSモデルにおけるアラインメントロバスト性向上のための非停滞化制約付きForward Attention

３．学会等名
日本音響学会2022年春季研究発表会

４．発表年
2022年

１．発表者名
岡留有希，大西弘太郎，岸田拓也，中鹿亘

２．発表標題
印象表現語ラベルを用いたFaderNetworksに基づく音声印象変換

３．学会等名
日本音響学会2022年春季研究発表会

４．発表年
2022年

１．発表者名
羽賀洋克，矢田部浩平，岸田拓也，中鹿亘

２．発表標題
時系列条件付きボルツマンマシンによる位相復元

３．学会等名
日本音響学会2022年春季研究発表会

４．発表年
2022年

１．発表者名
飯田紘崇，岸田拓也，中鹿亘

２．発表標題
マルチモーダルVAEを用いた顔画像に基づく目標話者音声不要な声質変換

３．学会等名
日本音響学会2022年春季研究発表会

４．発表年
2022年

１．発表者名
岸田拓也，中鹿亘

２．発表標題
深層エネルギーベースモデルによる音声の音響特徴量の生成

３．学会等名
日本音響学会2021年秋季研究発表会

４．発表年
2021年

１．発表者名
井硲巧，岸田拓也，中鹿亘

２．発表標題
話者依存度に応じた特徴抽出器によるdisentagleな声質変換

３．学会等名
日本音響学会2021年秋季研究発表会

４．発表年
2021年

１．発表者名
井硲巧，岸田拓也，中鹿亘

２．発表標題
話者特徴抽出器を加えたFaderNetVCによる未知話者声質変換

３．学会等名
音学シンポジウム2021

４．発表年
2021年

１．発表者名
Kishida, T., Tsukamoto, S., Nakashika, T.

２．発表標題
Simultaneous Conversion of Speaker Identity and Emotion Based on Multiple-Domain Adaptive RBM

３．学会等名
Interspeech 2020（国際学会）

４．発表年
2020年

１．発表者名
岸田 拓也, 中鹿 亘

２．発表標題
Cluster ARBM を用いた話者・音韻相互作用分類による声質変換

３．学会等名
日本音響学会2020年秋季研究発表会

４．発表年
2020年

１．発表者名
岸田 拓也, 中鹿 亘

２．発表標題
Attention RBMによる音声特徴量系列の符号化と生成

３．学会等名
日本音響学会2020年秋季研究発表会

４．発表年
2021年

１．発表者名
岸田拓也、中鹿亘

２．発表標題
Speech chain を模倣したボルツマンマシンによるワンショット多対多声質変換の検討

３．学会等名
日本音響学会2020年春季研究発表会

４．発表年
2020年

１．発表者名
荒川 賢也、岸田 拓也、中鹿 亘

２．発表標題
マルチタスクモデルを用いたdisentangleな学習による楽器音変換

３．学会等名
日本音響学会2020年春季研究発表会

４．発表年
2020年

１．発表者名
塚本 伸、岸田 拓也、中鹿 亘

２．発表標題
適応型 RBM を用いた音声情報の分離による話者と感情の同時変換

３．学会等名
日本音響学会2020年春季研究発表会

４．発表年
2020年

１．発表者名
岸田拓也、中鹿亘

２．発表標題
Speech chain VC: 音声コミュニケーションの言語-生理-音響連鎖を考慮する声質変換

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
荒川 賢也、岸田 拓也、中鹿 亘

２．発表標題
Fader Networks を用いた楽器音変換

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
塚本 伸、岸田 拓也、中鹿 亘

２．発表標題
適応型 RBM を用いたノンパラレル感情音声変換

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
Zhang, Y., Nakajima, Y., Yu, X., Remijn, G. B., Ueda, K., Kishida, T., & Elliott M. A.

２．発表標題
Acoustic analysis of word-initial consonant clusters: a perceptual basis of English syllables

３．学会等名
The 35th Annual Meeting of the International Society for Psychophysics（国際学会）

４．発表年
2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
岸田拓也 Takuya Kishida
https://kishidatakuya0119.wixsite.com/mysite

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