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　６ 年　 ６ 月 １０ 日現在

機関番号：１２６１２
研究種目：基盤研究(C)（一般）
研究期間：2021～2023
課題番号：２１Ｋ１１９５７

研究課題名（和文）音声スペクトルを対数的に表現する浅層ニューラルネットに関する研究

研究課題名（英文）Modelling Speech Spectra Based on Logarithmic Shallow Neural Networks

研究代表者
中鹿　亘（Nakashika, Toru）

電気通信大学・大学院情報理工学研究科・准教授

研究者番号：９０７４９９２０

交付決定額（研究期間全体）：（直接経費）　　4,700,000円

研究成果の概要（和文）：音声は重要なコミュニーケーションツールの一つであり，身の回りで様々な音声技術が活用されている。そのバックエンドには，特に近年世界中で注目されていることから深層学習が盲目的に利用されることが多い。深層学習は個々のタスクに対して非常に高い性能を示す反面，パラメータ数が膨大であり計算コストが高いというデメリットがある。計算資源の限られた小型デバイスにはパラメータ数の少ないコンパクトな機械学習モデルの方が望ましい。本研究では，音声データ特有の性質や構造に着目し，データを適切に表現するコンパクトな浅層モデルの方法論と枠組みを新たに提案し，複数の実験によって提案モデルの有効性を検証した。

研究成果の学術的意義や社会的意義
本研究では，音声のデータ構造に着目し，主に音声複素スペクトルを対数的に表現する複素浅層ニューラルネットを提案した。重要な本研究成果の1つとして，このモデルが，僅か800バイト程度の情報量で，最新の深層学習技術に基づく巨大なニューラルネットワークモデルと同程度の性能を示した，ということが挙げられる。このことから闇雲にパラメータ数を増やしてモデルを巨大化させるのではなく，知恵を絞って適切にデータを表現する方が得策であると言える。またこのようなコンパクトな浅層モデルは，演算による消費電力を抑えることにもなり，省エネで地球環境に配慮したグリーンコンピューティングなアプローチとして貢献することができる。

研究成果の概要（英文）：Speech is one of the most important communication tools, and various speech technologies are used around us. Especially in recent years, deep learning is often used blindly as its backend because it has been attracting worldwide attention. While deep learning shows very high performance for each task, it has the disadvantage of having a huge number of parameters and high computational cost. Compact machine learning models with a fewer number of parameters are preferable for small devices with limited computational resources. In this study, we proposed a new methodology and framework for a compact shallow-layer model that appropriately represents data, focusing on the specific properties and structures of speech data, and verified the effectiveness of the proposed model through multiple experiments.

研究分野：音声処理

キーワード：音声符号化　音声モデリング　機械学習　複素確率分布　ボルツマンマシン　ガンマ分布　フォン・ミーゼス分布　音源分離

様　式　C－１９、F－１９－１（共通）

１．研究開始当初の背景

音声は重要なコミュニーケーションツールの一つであり，スマートフォンやオンライン会議システムなど，身の回りで様々な音声技術が利用されているが，そのバックエンドには機械学習が用いられている。特に近年では深層学習が盛んに研究され，音声信号処理の様々なタスクに盲目的に利用される。深層学習は個々のタスクに対して非常に高い性能を示す反面，パラメータ数が膨大であり計算コストが高いというデメリットがある。スマートフォンやスマートウォッチなど，計算資源の限られた小型デバイスにはパラメータ数の少ないコンパクトな機械学習モデルの方が望ましい。コンパクトな浅層モデルを考えることは，演算による消費電力を抑えることにもなり，省エネで地球環境に配慮したグリーンコンピューティングなアプローチであると言える。しかし，巨大なモデルに比べて，コンパクトなモデルで性能を出すことは難しく，データの性質に関する知見や，その効果的な利用が欠かせない。本研究では，音声を適切に表現するコンパクトな浅層モデルの方法論と枠組みを考案・検証する。

２．研究の目的

本研究の目的は，音声を表現するためのコンパクトな浅層モデルとして制限ボルツマンマシン（RBM）に着目し，適切に設計した浅層モデルが巨大なモデルである深層モデルと同程度以上の性能を引き出せるかを実験で明らかにすることである。コンパクトなモデルとして他にもカーネル法や知識蒸留などがあるが，前者は明示的な確率モデルを仮定しないため音声特有の知識を組み込むことが困難であり，後者は一度大規模ネットワークを学習させる必要があり根本的な解決にはならない。RBM は観測特徴量を示す可視層と，潜在特徴量を示す隠れ層の２層からなる浅層ニューラルネットに分類される（図 1）。フィードフォワード型の一般的な浅層ニューラルネットと異なり，RBM は各素子に対して確率分布を陽に仮定でき，さらに素子の接続関係を示すグラフ構造を柔軟に設計することができる。そのため音声データが持つ特有の性質や構造を RBM で表現すれば十分な性能を発揮できる可能性がある。

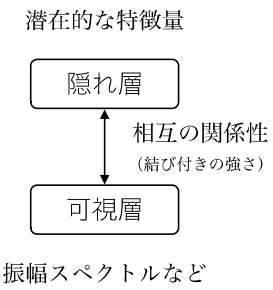

図 1：RBM の構造

３．研究の方法

スペクトルは音声の重要な情報であり，人間の聴覚的には対数振幅スペクトルが特に重要であることが知られており，メルケプストラムや MFCC など，対数振幅スペクトルから計算される音声の特徴量が広く用いられている。また，聴覚における位相の重要性が近年の研究で指摘されている。本研究では，音声を適切に表現する浅層ニューラルネットの性能を調査するため，主に 3 つの課題：(1)対数振幅スペクトルを表現する浅層ニューラルネットの基礎的な検証，(2)さらに位相を考慮した複素浅層ニューラルネットの基礎的な検証，そして(3)音源分離タスクへの応用実験を実施する。

(1) 対数振幅スペクトルを表現する浅層ニューラルネットの基礎的検証

従来の機械学習モデルでは，学習器の中で明示的に対数を取る必要がないように，入力として予め対数が取られている音声スペクトル特徴量を用いることが一般的であった。一方，本研究では，可視層が振幅スペクトルと対数振幅スペクトルで構成される RBM を検討し，学習器の中で明示的に対数を取ることで，振幅スペクトルと対数振幅スペクトルのインタラクションを実現する（図 2）。このモデル（以降，ガンマ RBM と呼ぶ）は可視素子にガンマ分布を仮定した RBM とみることができ，正値を取る振幅スペクトルと相性が良い。この課題では，複数パターン考えられるガンマ RBM の構造を実験的に検証し，音声にとって最適なガンマ RBM を明らかにする。また，学習器の中で明示的に対数を取ることによる応用上の利点についても検証し，従来の RBM や，関連性の高い深層学習手法である変分オートエンコーダと音声の分析合成性能を比較する。これにより，ヒトの聴覚を考慮した RBM の振幅スペクトルに対する表現能力を明らかにする。

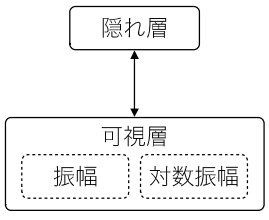

図 2：ガンマ RBM の構造

(2) 対数振幅と位相を同時に表現する複素浅層ニューラルネットの基礎的検討

従来の音声処理では，位相スペクトルを無視し，振幅スペクトルに基づく特徴量を用いることが多い。そのため，推定された振幅スペクトルから何らかの手法で位相を推定し，音声波形を生

成することが多かった。しかし，スペクトルは本来的には複素数であり，振幅のみでなく位相も重要な情報であることが近年の研究で示されている。振幅スペクトルに基づく特徴量は位相情報が欠落し，本来の複素スペクトルを完全に復元することは不可能なので，特に音声合成では位相もモデル化することが重要である。そこで本研究課題では，課題 A で扱った振幅および対数振幅スペクトルに加え，位相スペクトルについても明示的に扱う RBM モデルを検討する（図 3）。位相は複素平面における角度を表すため，代表的な方向分布であるフォン＝ミーゼス分布を仮定した RBM として位相スペクトルを定式化する。以降，このモデルを GVM（ガンマ・フォン＝ミーゼス RBM）と呼ぶ。GVM は本質的に極座標表現となるので，まずは極座標における複素スペクトルの統計的性質を調査する。その結果を元に，GVM の学習を効果的に行う方法について検討する。また，スペクトルの位相には様々な表現があるので，モデル化に適した表現についても検証する。これにより，音声スペクトルの極座標表現および振幅の対数表現を同時に実現した新たな RBM の実現を目指す。

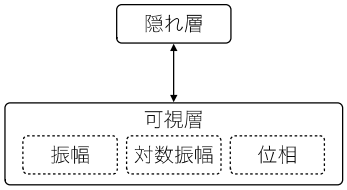

図 3：GVM の構造

(3) 浅層ニューラルネットの音源分離への応用実験

課題 A，B が音声符号化・復号化実験を通したモデルの基礎的な検証であるのに対し，本研究課題ではより実用的な音声タスクに適用可能なことを示すために，具体的な音声タスクの例として音源分離に応用することを考える。音源分離は，複数の音源が混在した信号から個々の音源信号を分離するタスクである。従来の代表的な音源分離手法として，非負値行列因子分解（NMF）による音源モデルと独立ベクトル分析による空間モデルの最適化を交互に繰り返す独立低ランク行列分析に基づく手法が挙げられる。NMF は観測されるスペクトログラムを基底行列とアクティビティ行列の積に分解するアルゴリズムであるが，基底やアクティビティに対して陽に確率分布は与えられない。しかしアクティビティは本来各基底がそれぞれの時刻において励起しているか否かを示す二値で表されるべきである。そこで本課題では，可視層に観測スペクトル，隠れ層にアクティビティ，それらの接続重みとして基底を仮定した RBM で空間モデルを代用した音源分離を考える。観測される複素スペクトルは，従来同様原点対称複素正規分布に従うと仮定する。このときフォン＝ミーゼス分布に従うと仮定した位相について周辺化すれば，振幅スペクトルはレイリー分布（パワースペクトルは指数分布）に従うことから，可視層にレイリー分布を仮定した RBM（レイリーRBM），または指数分布を仮定した RBM（指数 RBM）を定式化・実装して音源分離に応用する。これにより RBM の汎用性および具体的な応用例における実用性を明らかにし，コンパクトな浅層ニューラルネットワークモデルの可能性と有効性を示す。

４．研究成果

上述の研究課題ごとに研究成果を報告する。

(1) 対数振幅スペクトルを表現する浅層ニューラルネットの基礎的検証

ガンマ RBM の有効性を評価するため，音声振幅スペクトルを可視層としたガンマ RBM を学習させた後，テストデータを入力して推論（符号化）された隠れ層から振幅スペクトルを再構成（復号化）し，元の信号とどれほど近いのかを検証した。日本語の 50 文からなる女性音声を学習に，同じ話者の別発話 53 文の音声を評価に用いた。サンプリング周波数 16kHz にダウンサンプリングし，窓幅 256 のブラックマン窓を用いてシフト幅 64 で短時間フーリエ変換を施し，129 次元の振幅スペクトルを得た。課題 A については位相のモデル化は考慮しないため，逆フーリエ変換を用いて音声信号を復元する際の位相スペクトルは正解を与えた。比較手法として連続値可視ベクトルを表現する一般的な RBM である，可視層にガウス分布を仮定したガウス RBM を用いて振幅スペクトルまたは対数振幅スペクトルを表現する手法を用いた。精度を比較するために客観評価指標として PSEQ および STOI を，また主観評価指標として 43 人の実験参加者による５段階評価 MOS を用いた。結果をそれぞれ表 1，2 に示す。

Methods	PESQ	STOI
gamma-RBM	**4.17**	**1.00**
Gauss-RBM (log)	4.07	0.99
Gauss-RBM (amp)	2.69	0.92

表 1：各手法の客観評価

Method	MOS (±95%CI)
gamma-RBM (H800)	**4.332** ± 0.091
Gauss-RBM (log, H800)	4.199 ± 0.095
WORLD [40]	3.043 ± 0.117
Original	4.432 ± 0.085

表 2：各手法の主観評価

実験結果より，いずれの評価指標でも提案手法（gamma-RBM）が最も良い性能を示した。これは，従来のガウス RBM では振幅または対数振幅のいずれかのみを考慮した表現であるのに対し，ガンマ RBM は振幅と対数振幅を同時に考慮したモデルとなっており，このことが結果に寄与し

たと考えられる。また従来のモデルでは，正値をとる振幅スペクトルを入力特徴量として用いた場合（Gauss-RBM(amp)），-∞～+∞を定義域とするガウス分布には合致しないことから著しく精度が低下していることが分かる。各手法による振幅スペクトルの再構成例を図 4 に示す。ガンマ RBM および対数振幅スペクトルを用いたガウス RBM（Gauss-RBM(log)）はうまく元（Original）のスペクトルのフォルマントや微細構造を表現できているが、振幅スペクトルを用いたガウス RBM は本来出るはずのない負値（赤箇所）が出現してしまっている。

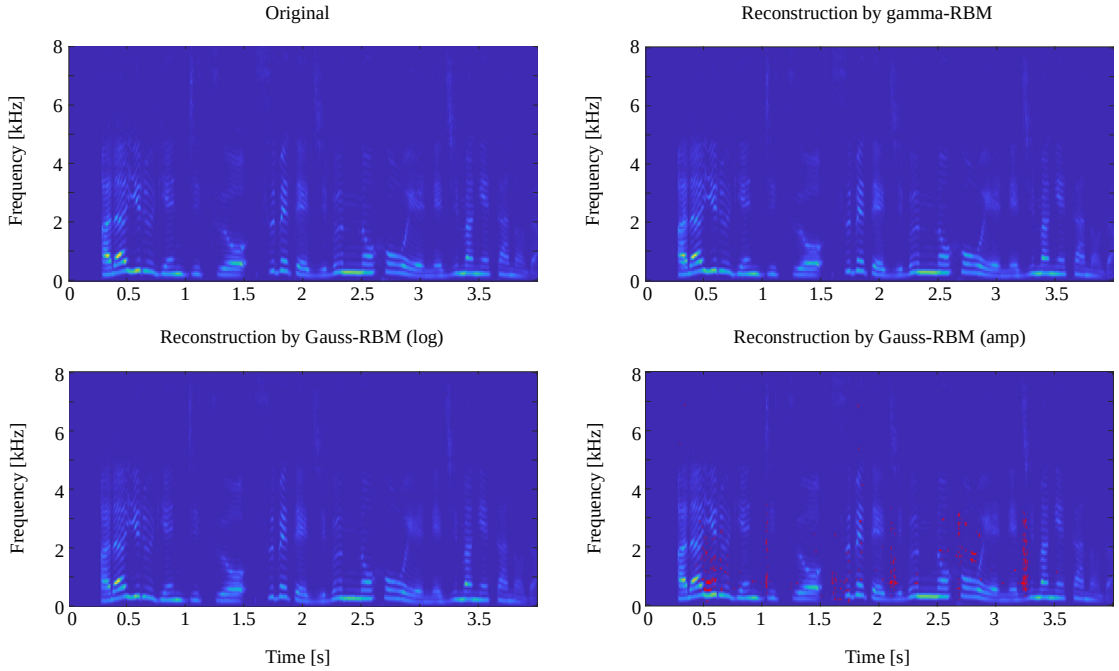

図 4：各手法による振幅スペクトルの再構成例

(2) 対数振幅と位相を同時に表現する複素浅層ニューラルネットの基礎的検討

振幅スペクトル，対数振幅スペクトルおよび位相スペクトルを同時に考慮した GVM の有効性を評価するため，課題 A と同様に，音声の複素スペクトルを可視層とした GVM を学習させた後，テストデータに対して符号化・復号化して音声信号がどれほど精度良く再構成されるのかを検証した。図 5 は提案手法の GVM と，従来手法である VAE，CVAE，CRBM，RBM について潜在変数の数を 200 から 6400 まで変えた場合の PESQ を示している。VAE および RBM は対数振幅スペクトルがモデルで表現され，位相は Griffin-Lim 法によって復元している。図 5 より，潜在変数が 800 まで CVAE が最も良い精度だが，それ以上増やすと CVAE は精度が低下しているのに対し，GVM は向上し，潜在変数 3200 以降は GVM が最も高い精度であることが分かる。また提案手法において潜在変数を二値化して符号化した場合（GVM(binary)），わずか 800(=6400/8)バイトの情報量で，最新の深層学習ベースの音声ボコーダである HiFiGAN に匹敵する性能であることが示された。

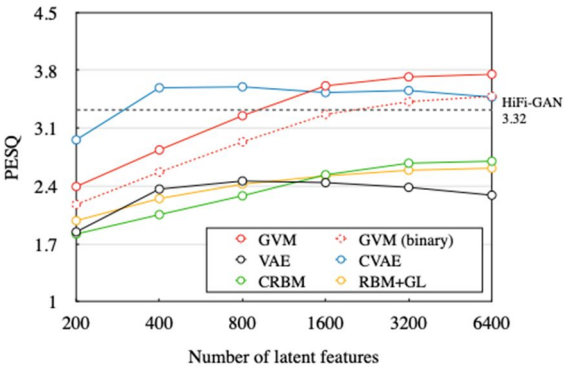

図 5：各手法の客観評価（PESQ）

また各手法の精度比較を表 3 にまとめる。再構成の精度を示す客観指標として PESQ，STOI，UTMOS，WarpQ を，主観評価指標として 14 人の被験者による 5 段階評価 MOS を，また符号化・復号化の速度を示す指標とし

Methods	PESQ	STOI	UTM	W-Q	RTF	MOS (±CI)
GVM RBM	**3.75**	**0.99**	**3.79**	**1.23**	0.09	**4.41±0.15**
CRBM [40]	2.70	0.94	2.65	2.24	0.06	3.03±0.21
CVAE [39]	3.48	**0.99**	3.67	1.40	0.01	4.00±0.20
GaussRBM+VM	3.17	0.95	2.76	1.49	0.08	3.21±0.20
GaussRBM+GL	2.62	0.95	2.26	1.28	0.48	2.39±0.23
gammaRBM+GL	2.62	0.98	2.49	1.26	0.49	2.37±0.25
VAE	2.46	0.94	1.97	2.34	0.01	1.69±0.17
HiFi-GAN [9]	3.32	0.96	3.52	1.32	0.27	**4.41**±0.17
WORLD	2.97	0.96	2.50	2.31	0.44	3.01±0.28
NATURAL	—	—	3.94	—	—	4.70±0.13

表 3：GVM の実験結果

て RTF(real time factor)を用いた。前述の実験に加えて，対数振幅スペクトルをガウス分布，位相をフォン＝ミーゼス分布とした RBM（GaussRBM+VM），対数振幅スペクトルをガンマ RBM で表現し位相を Griffin-Lim 法によって復元した手法（gammaRBM+GL），および古典的な信号処理に基づくボコーダである WORLD を比較対象とした。各手法のハイパーパラメータはそれぞれ最適化している。表 3 より，RTF を除くいずれの指標においても提案手法である GVM が最も良い性能を示していることが分かる。RTF は 0.1 未満であり，同程度の MOS である HiFiGAN よりも高速に動作することが確認できた。

(3) 浅層ニューラルネットの音源分離への応用実験

まず，振幅スペクトルを入力特徴量とするレイリーRBM（RB-RBM）およびパワースペクトルを入力特徴量とする指数 RBM（EB-RBM）による再構成の性能及び品質比較のために，女性アナウンサーの 50 文音声を用いた再構成実験を行った．比較手法としては，既存手法であるガウス RBM および ISNMF（基底数 2 および 10）を用いた。一般に，ILRMA のような瞬時混合を仮定した音源分離モデルでは窓長より残響時間が大きいことを想定しているため，窓長が短い方がより多くの実信号に対して適用できるが分離精度が下がるというトレードオフが存在する。そのため窓長を変化させて最適なパラメータを調査した。結果を表 4 に示す。表 4 より，いずれの手法でも概ね窓幅を増やすほど精度が向上するが，窓幅 256 以降はあまり変わらなかったことから，音源分離実験では窓幅 256，シフト幅 32 で実験を行った。またこのとき，提案手法の１つである RB-RBM が最も再構成の性能が高く，次いでもう１つの提案手法である EB-RBM の性能が高かった。

2*音源モデル	2*入力特徴量	Window length in STFT			
		64 ms	128 ms	256 ms	512 ms
ISNMF(K = 2)	power	1.29	1.51	1.95	1.98
ISNMF(K = 10)	power	1.74	2.05	2.22	2.50
GB-RBM	log-power	1.56	1.82	2.18	2.30
RB-RBM	amplitude	N/A	2.75	2.66	2.45
EB-RBM	power	2.24	2.46	2.46	2.17

表 4：PESQ による各音源モデルの精度比較

最後に男性２話者の混合音声および女性２話者の混合音声を用いて各手法の音源分離の性能を評価した。残響時間は 130ms または 250ms，マイク間の距離は 1m または 5cm を設定した。音源分離の客観的評価指標として信号対歪み比(SDR)の改善値を用いた。値が大きければ大きいほど分離信号の歪みが少なく，高品質であることを表す。実験結果を表 5 に示す。表 5 より，残響時間の小さいとき（残響時間 130ms），提案手法の RB-RBM が分離精度が高く，提案手法の有効性を確認することができた。一方で，マイク間の距離が小さいとき（5cm），分離性能が極端に悪くなってしまっている。この問題は，窓長やシフト長の違いによる情報量の変化が影響していることが考えられる。また，いずれの手法においても残響時間が大きくなるにつれて，あるいはマイク間の距離が近づくにつれて，分離が難しくなっていることもわかる。これは，残響によって音源が時間方向に反響することやマイク間距離が近づくことでより音源信号が複雑に混ざり合うことに起因すると考えられる。

Recording conditions	Method			
(rev. time and mic. spacing)	ILRMA	GB-RBM	RB-RBM	EB-RBM
130 ms and 1 m	11.210	11.664	**12.443**	11.797
130 ms and 5 cm	-0.839	-0.281	**0.555**	-0.794
250 ms and 1 m	**9.801**	5.1367	6.013	5.285
250 ms and 5 cm	**0.232**	-0.597	-0.867	-1.540

表 5：各音源分離手法による SDR 改善値比較

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 Nakashika Toru、Yatabe Kohei	４．巻 29
２．論文標題 Gamma Boltzmann Machine for Audio Modeling	５．発行年 2021年
３．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	６．最初と最後の頁 2591～2605
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2021.3095656	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37件（うち招待講演　0件／うち国際学会　6件）

１．発表者名 Kotaro Onishi, Toru Nakashika
２．発表標題 Consistency Regularization for GAN-Based Neural Vocoders
３．学会等名 APSIPA 2022, pp. 1132-1137, November 2022（国際学会）
４．発表年 2022年

１．発表者名 Kotaro Onishi, Toru Nakashika
２．発表標題 MoCoVC: Non-Parallel Voice Conversion With Momentum Contrastive Representation Learning
３．学会等名 APSIPA 2022, pp. 1435-1440, November 2022（国際学会）
４．発表年 2022年

１．発表者名 Takuya Kishida, Toru Nakashika
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 APSIPA 2022, November 2022（国際学会）
４．発表年 2022年

１．発表者名
Takumi Isako, Kotaro Onishi, Takuya Kishida, Toru Nakashika

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

３．学会等名
APSIPA 2022, pp. 1444-1448, November 2022（国際学会）

４．発表年
2022年

１．発表者名
奥田耕平 岸田拓也, 中鹿亘

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, 1-3Q-10, March 2023, March 2023.

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
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, 3-3-13, pp. 769-770, March 2023.

４．発表年
2022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, 1-3P-11, March 2023.

４．発表年
2022年

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
第146回研究会音声言語情報処理研究会，March 2023.

４．発表年
2022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，2-Q-48，pp. 1305-1306，September 2022.

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，2-Q-47，pp. 1301-1304，September 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，SP2022-8，pp. 26-29，June 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，SP2022-23，pp．100-103，June 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
大西弘太郎，中鹿亘

２．発表標題
非可逆圧縮を用いた敵対的ニューラルボコーダのためのデータ拡張法

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
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
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
釘本咲，中鹿亘

２．発表標題
ベータ分布に基づくFaderNetを用いた音声印象変換の性能評価

３．学会等名
日本音響学会2023年秋季研究発表会

４．発表年
2023年

１．発表者名
古田翔太郎，中鹿亘

２．発表標題
レイリー型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
日本音響学会2023年秋季研究発表会

４．発表年
2023年

１．発表者名
芦田裕飛，中鹿亘

２．発表標題
SiFiSinger: SiFi-GANを内包した歌唱音声合成

３．学会等名
日本音響学会2023年秋季研究発表会

４．発表年
2023年

１．発表者名
釘本咲，中鹿亘

２．発表標題
FaderNetを用いた未知話者に対する音声印象変換

３．学会等名
音学シンポジウム2023

４．発表年
2023年

１．発表者名
Mastuti Puspitasari, Takuya Takahashi, Gen Hori, Shigeki Sagayama, Toru Nakashika

２．発表標題
SBERT-based Chord Progression Estimation from Lyrics Trained with Imbalanced Data

３．学会等名
CMMR 2023（国際学会）

４．発表年
2023年

１．発表者名
Takuya Takahashi, Shigeki Sagayama, Toru Nakashika

２．発表標題
Controllable Automatic Melody Composition Model across Pitch/Stress-accent Languages

３．学会等名
CMMR 2023（国際学会）

４．発表年
2023年

１．発表者名
越森 道貴，嵯峨山 茂樹，中鹿 亘

２．発表標題
2 種のラグ窓によるスペクトル平滑化を用いた F0 推定

３．学会等名
日本音響学会2024年春季研究発表会

４．発表年
2024年

１．発表者名
後藤 純平，中鹿 亘

２．発表標題
FaderNetworks を用いた F0 変換による歌唱技術の付与

３．学会等名
日本音響学会2024年春季研究発表会

４．発表年
2024年

１．発表者名
芦田 裕飛，中鹿 亘

２．発表標題
歌唱音声合成における F0 の自然性向上のための Diffusion-GAN モデルの検討

３．学会等名
日本音響学会2024年春季研究発表会

４．発表年
2024年

１．発表者名
畠山 瑠一，奥田 耕平，中鹿 亘

２．発表標題
拡散確率モデルを用いたノンパラレルな Any-to-many 声質変換

３．学会等名
日本音響学会2024年春季研究発表会

４．発表年
2024年

１．発表者名
平本 佳弘，中鹿 亘

２．発表標題
事前学習済みモデルによる埋め込み表現を組み込んだ音声編集モデルの検討

３．学会等名
日本音響学会2024年春季研究発表会

４．発表年
2024年

１．発表者名
石川峻弥，中鹿亘

２．発表標題
分類型半制限ボルツマンマシンによる全音程関係を考慮した和音認識

３．学会等名
日本音響学会2024年春季研究発表会

４．発表年
2024年

１．発表者名
水野友暁，岸田拓也，吉村奈津江，中鹿亘

２．発表標題
Transformerを用いた脳波信号からの音声復元の検討

３．学会等名
第151回音声言語情報処理研究発表会

４．発表年
2024年

１．発表者名
今市夏菜子，中鹿亘

２．発表標題
潜在変数と観測データにガンマ分布を仮定したVAEによる音声振幅スペクトル表現

３．学会等名
第151回音声言語情報処理研究発表会

４．発表年
2024年

１．発表者名 越森道貴，嵯峨山茂樹，中鹿亘
２．発表標題 複数のラグ窓対を用いた音声基本周波数と周期性尺度の推定
３．学会等名 第151回音声言語情報処理研究発表
４．発表年 2024年

１．発表者名 畠山瑠一，奥田耕平，中鹿亘
２．発表標題 DDPMVC: 連続時間拡散確率モデルを用いた非パラレル声質変換と評価
３．学会等名 第151回音声言語情報処理研究発表会
４．発表年 2024年

〔図書〕　計1件

１．著者名 柳井 啓司、中鹿 亘、稲葉 通将	４．発行年 2022年
２．出版社 オーム社	５．総ページ数 288
３．書名 IT Text 深層学習	

〔出願〕　計1件

産業財産権の名称 声質変換装置、声質変換方法及びプログラム	発明者 大西弘太郎，中鹿亘	権利者 同左
産業財産権の種類、番号 特許、特願2021-026128	出願年 2021年	国内・外国の別 国内

〔取得〕　計0件

〔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	矢田部　浩平 (Kohei Yatabe) (20801278)	東京農工大学・工学（系）研究科（研究院）・准教授 (12605)	

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