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　４ 年　 ６ 月 １０ 日現在

機関番号：２２６０４
研究種目：若手研究
研究期間：2019～2021
課題番号：１９Ｋ２０２７１

研究課題名（和文）音声制御デバイスのためのセキュアな生体認証システムに関する研究

研究課題名（英文）A study on secure biometric authentication systems for voice control devices

研究代表者
塩田　さやか（Shiota, Sayaka）

東京都立大学・システムデザイン研究科・助教

研究者番号：９０７０５０３９

交付決定額（研究期間全体）：（直接経費）　　2,500,000円

研究成果の概要（和文）：本研究の目的はなりすまし検出技術の精度を向上させることで、セキュアな音声対話技術を実現することである。本研究で想定しているなりすまし攻撃はスマートスピーカーなどの声で操作する音声対話システムであり、これらのシステムを使用する際の個人情報を保護するための個人認証に関する研究及び音声情報自体のセキュリティに関する研究を行った。研究の成果として、ステレオチャンネル音声を用いたなりすまし検出に関する発表を国内外の学会で発表し、国際論文として発表したことや、世界的にも新しい音声のプライバシー保護に関する研究に関する論文発表などを行った。

研究成果の学術的意義や社会的意義
声を用いた生体認証技術の研究は盛んに行われており、実用化も期待されている技術である。しかし、生体認証技術はなりすましなどの攻撃に対する頑健性を担保することが実用化の上でも非常に重要であり、研究課題としても非常に難しい問題となっている。さらに音声のプライバシー保護に関する問題も学術的に新しい問題となっており、また今後の実用化においても避けて通れない問題となっている。これらの問題に着手した本研究の成果の学術的及び社会的意義は非常に高いといえる。

研究成果の概要（英文）：The purpose of this research is to improve the accuracy of spoofing detection technology to realize secure spoken dialogue technology. The spoofing attacks assumed in this research are spoken dialogue systems such as smart speakers, and research was conducted on personal authentication to protect personal information when using these systems, as well as on the security of the voice information itself. The results of this research include presentations on spoofing detection using stereo channel audio at domestic and international conferences, which were published as international papers, and the publication of a paper on research on privacy protection of voice, which is new to the world.

研究分野：音声信号処理

キーワード：話者認識　なりすまし検出　音声信号処理　プライバシー保護

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

## １．研究開始当初の背景

近年、スマートスピーカやロボットなど声のみで機械を制御する音声制御デバイスが普及してきている。今後は家庭内や公共機関などへの導入だけでなく、自動車やバスなどの自動運転などへの応用も期待されている。このような流れの中、人間と機械との音声対話でのやり取りには天気予報や情報検索など一般的な情報を扱うだけでなく、電話番号や住所、仕事やプライベートのスケジュールの確認など利用者に関連する個人情報を取り扱う機会も増えていくことが容易に考えられる（図 1）。そのため、音声制御デバイスを利用者以外が誤って操作しないよう、また、悪意のある第三者に情報を盗まれないように音声対話に対するセキュリティを高めることが急務となってきている。


図 1：音声制御デバイスの使用シーンの例

しかし、近年の様々な収録機材の発達やマイクの小型化などにより、誰かの声を録音することは非常に簡単になってきている。そのため、特に盗聴などにより登録者の声を録音して再生するなりすまし攻撃が非常に重大な問題として注目されている。

これまでに、スピーカでなりすまし音声を再生する「なりすまし攻撃」を検出する様々な技術が報告されてきたが、その中に『声の生体検知』という技術がある。これは図 2 に示すように、音声制御デバイスに入力された音声に対して処理を行うモジュール（音声認識、話者照合など）の前段において、入力された音声が人間による発声なのか、スピーカで再生したなりすまし音声なのかを判定する技術である。これまでに声の生体検知を実現する手法として、人間が発声する際にのみ必ず起こる現象を生体の証明として検出する方法が発表されている。その 1 つは人間が発話をする際に使う呼気による現象を発話中から検出するポップノイズ検出法である。他に人間が発話する際に口の形状を変える事で様々な音を発音していることから、口内で音が生成される音源位置の変動を検出する手法もある。これらの着目点が生体の証明として頑健であると示される一方で、実際の識別性能は収録条件や環境に強く依存することも報告もされており、十分な解決には至っていない。

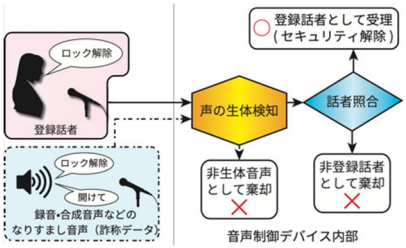


図 2：声の生体検知と話者照合

## ２．研究の目的

これまでの先行研究の知見から、人間の発声時に起こる発声音の変動よりも、スピーカ再生音声に変動が起こらないことの方がマイクやスピーカの機種に依存せず確認できるということが予測された。そこで本研究では、これまでの入力音声から生体の証明を検出する「声の生体検知」だけではなく、スピーカ再生の証明を検出する「非生体音検知」について検討する。入力信号における「スピーカ再生の証明」を示すことでなりすまし攻撃を検出する、つまり、人間の発声とスピーカ再生との決定的な要素は何かを人間側からだけでなく機械側からも「問う」ことが本研究の学術的核である。本研究課題の最終的な目的は図 3 に示すように、人間が音声制御デバイスへ問いかける場合には通常の対応を行い、スピーカ再生による問いかけには対応しないシステムの構築である。つまり、スピーカ再生による誤動作を防ぐなりすまし検出システムを実現することを指す。

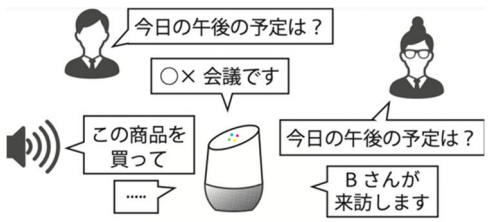


図 3：理想的な音声制御デバイスの挙動

## ３．研究の方法

【1．スピーカ再生によるなりすまし攻撃の解析】これまでに公開されたなりすまし検出のためのコーパスに対して、様々な音声特徴量を用いてスピーカ再生と実発話の差を解析する。このコーパスは実発話及び実発話をスピーカ再生して再収録したなりすまし攻撃が含まれているが、公開されているコーパスだけでは収録環境が偏ってしまうため、さらに、自身で収集したデータも用い解析を行う。独自収録するデータは、収録環境、収録に使用したマイク、再生に使用したスピーカ、再収録に用いるマイクなど様々な収録条件を用意するため、環境や機種の違いに関する詳細な分析が可能である。更には、大規模コーパスでは確認しづらい、より現実的な詐称方法を実際に再現した場合の分析も可能である。これらの分析を踏まえて、本研究で着目する複数チ

ャネル間の音声の一般化相互相関がスピーカ検出に有効であるかを確認する。ここで、一般化相互相関は信号間の類似度や遅延量などを測る電気音響の分野で用いられる特徴量の一つであるが、スピーカ検出に用いることは初めての試みとなる。
【2．スピーカ再生検出法の検討】提案するスピーカ再生検出法では 2 チャンネルマイク間の入力信号の一般化相互相関値を用いる。これは図 4 下段のピーク値を指す。図 4 中段のマイク 1、マイク 2 の例で示すとおり、2 チャンネルマイクで収録されたそれぞれの信号はほぼ同じ信号となるため、音声区間においては高い一般化相互相関値を得ることができる。人間の発声の場合には、発話の前後や発話中のショートポーズなど声を発していない無発話区間のみ、周囲の環境音が収録されてしまうため一般化相互相関値が低くなる。しかし、スピーカ再生においては、再生ボタンを押してから再生を終了するまでスピーカから無発話区間であろうと若干の音が再生されるため人間の発声とは異なり無発話区間においても一般化相互相関が高くなる事が考えられる。そこで、まず、一般化相互相関値を直接しきい値として用いた識別について提案する。この際、フレーム毎に時系列データとして抽出される一般化相互相関値の扱い方については、フレーム間平均や分散、最小値など複数パターン検討する。さらにチャンネル数を増やすことの影響や統計的な識別器の導入等も検討する。また、研究が進むうちにデータ量が不十分ということになった場合には、被験者を募集するなどでデータベースの規模を大きくすることを検討する。
【3．深層学習などの統計的機械学習による識別】研究の初期段階では、人間の発話とスピーカ再生の違いが確認できる明確な特徴である一般化相互相関値を直接用いて識別を行う。この初期調査で十分な知見を集めた上で、次に汎用化のための特徴量抽出とモデル化について検討する。モデルの表現には深層学習を用いることで、初期調査で有効と判断した特徴だけでなく非線形な空間で実発話となりすまし攻撃の違いをブラックボックスの中でどのように抽出されるかについても検討する。
【4．最先端の話者認識システムとの組み合わせによる評価】本テーマは音声制御デバイスのセキュアな使用に着目しているため、最終的にはなりすまし攻撃に対する音声制御デバイスの頑健性向上とともにその利便性を維持することが挙げられる。そのため、最先端の話者照合技術である i-vector/PLDA や x-vector と呼ばれるシステムを構築し、提案するなりすまし検出法と組み合わせた性能評価を行うことは必須となる。なりすまし検出の分野において近年、共通の指標として提唱されている t-DCF という評価尺度を用いることでシステム全体を評価し報告する予定である。それによって本研究の目的である音声制御デバイスのためのセキュアな生体認証システムの構築が達成できたといえる。

## ４．研究成果

本研究の目的は声を使った生体認証技術をより安全に使える技術を考案することにある。そのために，研究初年度はスピーカ再生特有の特徴として 2 系統の入力音声の類似度を計算する手法によって得られた特徴量を用いたなりすまし検出法について検討し、研究成果の発表を行った（図 4）。提案法では、図 4 に示すとおり実発話となりすまし音声である再生音声において、特に無音区間におけるピーク値の違いに着目して実験を行った。研究を始めた当初は類似度を用いた特徴量のみを用いたなりすまし検出法について実験を行っていた。しかし、テストする環境によって性能が非常に高い場合と不安定になる時があることが判明した。そこでテスト環境に依存しにくいシステムを検討するために更に、これまでに提案されてきた音響的な特徴を捉えた手法と組み合わせることを検討した。統合方法についても検討した結果、本実験で想定したテスト環境においては環境に依存せず高い性能を得ることを確認できた。これらの成果は査読ありの国際学会にも採択されている。

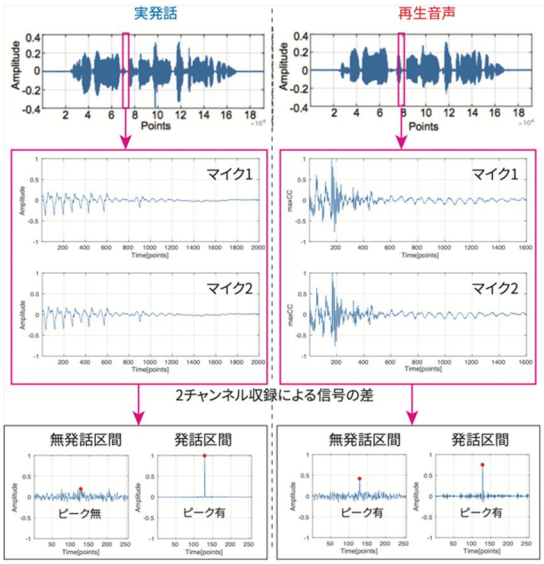

図 4：提案したステレオ信号を用いたなりすまし検出における特徴量

二年目においては、一年目の成果に対して更に実験を追加し、モデル学習や特徴量抽出についての改善を行い、テスト環境に対しての頑健性の向上を目指し、それらの結果をまとめて国際学会誌に英語論文として投稿し、採択に至っている。さらに、本研究で掲げる課題であるセキュアな生体認証システムに関する課題の一貫として、音声プライバシーに関する研究にも着手した。音声プライバシーに関してはこれまでに研究分野として確立されていなかったため、新たに立ち上がったコンペティションの条件を基に信号処理と深層学習を組み合わせた軽量な音声加工法について提案を行った。音声プライバシーについての研究実績については国内学会及び国際学会での研究発表が挙げられる。

最終年度も、前年度を引き継ぎ、なりすまし検出に関する検討と音声プライバシーに関する研究を行った。なりすまし検出においては、ステレオチャンネルで収録されたデータを用い、深層学習を用いたモデル化を行い、それらの成果を国内学会にて発表した。これは、これまでの深層学習を用いたなりすまし検出と異なり、2 系統から入力された音声信号を用いたモデル化を行っている。前年度までのモデル化においては簡素な統計モデルを用いてきており、特徴量については陽に抽出された相互相関の特徴を用いていたが、深層学習を用いた際には入力信号を 2 系統で分けて直接入力することで性能向上をはかり、特徴の違いを隠れ層において学習させるモデル化を行った。一方音声プライバシー保護については、これまでに研究分野として確立されていなかったため、新たに立ち上がったコンペティションの条件を基に信号処理と深層学習を組み合わせた軽量かつ不可逆な音声加工法について提案を行った。提案法では、信号処理に基づく複数の音声加工法を用い、それらを最適に組み合わせることで話者性を隠しつつ発話内容は損なわない程度の音質を維持するという枠組みを実現した。最適化に関しては深層学習の枠組みを用いることで、最適なパラメータの組み合わせを自動的に探索可能にしつつ、音声加工には処理の軽い信号処理に基づく手法を用いることで、高い性能を維持することができた（図 5）。さらに、音声プライバシーにおいて必要となる要件として、可逆性、つまり加工したあとの音声から攻撃者がもとに戻すことができるかどうかについての議論も行いこれらの成果をまとめて採択率の低い国際論文誌へ投稿を行い、採択されるに至った。

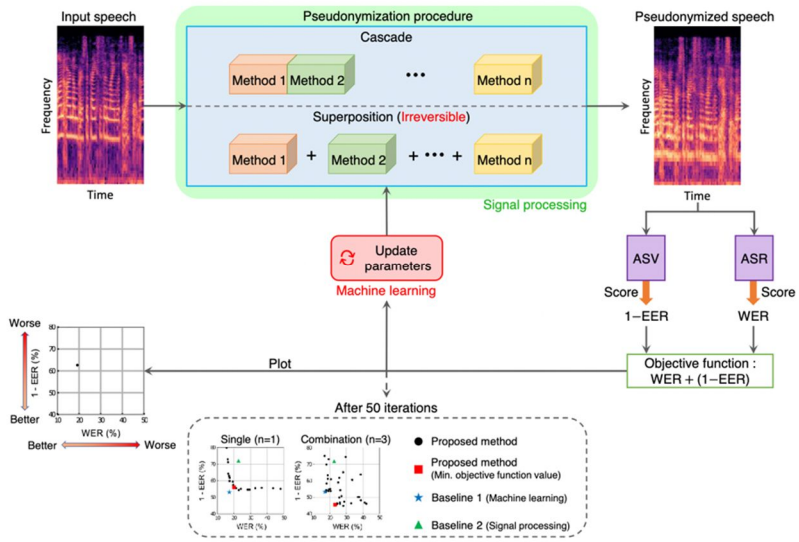


図 5：音声プライバシーのための音声加工法

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　3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<br>Kai Hiroto、Takamichi Shinnosuke、Shiota Sayaka、Kiya Hitoshi | ４．巻<br>72 |
| ２．論文標題<br>Lightweight and irreversible speech pseudonymization based on data-driven optimization of cascaded voice modification modules | ５．発行年<br>2022年 |
| ３．雑誌名<br>Computer Speech & Language | ６．最初と最後の頁<br>101315～101315 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.1016/j.csl.2021.101315 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

| | |
|---|---|
| １．著者名<br>Yaguchi Ryoya、Shiota Sayaka、Ono Nobutaka、Kiya Hitoshi | ４．巻<br>29 |
| ２．論文標題<br>Replay Attack Detection Based on Spatial and Spectral Features of Stereo Signal | ５．発行年<br>2021年 |
| ３．雑誌名<br>Journal of Information Processing | ６．最初と最後の頁<br>275～282 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.2197/ipsjjip.29.275 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

| | |
|---|---|
| １．著者名<br>KAMINISHI Ryota、MIYAMOTO Haruna、SHIOTA Sayaka、KIYA Hitoshi | ４．巻<br>E103.D |
| ２．論文標題<br>Blind Bandwidth Extension with a Non-Linear Function and Its Evaluation on Automatic Speaker Verification | ５．発行年<br>2020年 |
| ３．雑誌名<br>IEICE Transactions on Information and Systems | ６．最初と最後の頁<br>42～49 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.1587/transinf.2019MUP0008 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

〔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<br>Yuki Shiroma, Yuma KINOSHITA, Sayaka SHIOTA, Hitoshi KIYA, |
| ２．発表標題<br>Phase representation based on HSV color model for acoustic classification with convolutional neural networks |
| ３．学会等名<br>IEEE Global Conference on Consumer Electronics（国際学会） |
| ４．発表年<br>2021年 |

１．発表者名
Yuki Shiroma, Keisuke IMOTO, Sayaka SHIOTA, Nobutaka ONO, Hitoshi KIYA

２．発表標題
Investigation on Spatial and Frequency-Based Features for Asynchronous Acoustic Scene Analysis

３．学会等名
APSIPA Annual Summit and Conference（国際学会）

４．発表年
2021年

１．発表者名
城間 佑樹，木下 裕磨，塩田 さやか，貴家 仁志

２．発表標題
深層学習に基づく楽器音分類のための画像分類ネットワークを用いたファインチューニング

３．学会等名
情報処理学会音楽情報科学研究会 音学シンポジウム

４．発表年
2021年

１．発表者名
甲斐 優人，高道 慎之介，塩田 さやか，貴家 仁志

２．発表標題
音声仮名化のための加工音声重畳と非可逆性評価

３．学会等名
日本音響学会秋季大会

４．発表年
2021年

１．発表者名
荻野　天翔，今泉　遼，塩田 さやか，貴家 仁志

２．発表標題
話者照合のためのステレオ音声データを用いた深層学習に基づくなりすまし検出法

３．学会等名
日本音響学会秋季大会

４．発表年
2021年

１．発表者名
甲斐 優人，高道 慎之介，塩田 さやか，貴家 仁志

２．発表標題
パラメータ最適化を用いた信号処理による仮名化手法の復号攻撃に対するロバスト性評価

３．学会等名
電子情報通信学会 音声研究会

４．発表年
2022年

１．発表者名
城間 佑樹，木下 裕磨，井本 桂右，塩田 さやか，小野 順貴，貴家 仁志，

２．発表標題
自己符号化器を用いた多チャンネル信号の欠損復元法と 環境音分類における評価

３．学会等名
電子情報通信学会 応用音響研究会

４．発表年
2022年

１．発表者名
高道 慎之介，Kurzinger Ludwig，佐伯 高明，塩田 さやか，渡部 晋治

２．発表標題
JTubeSpeech：音声認識と話者照合のためにYouTubeから構築される日本語音声コーパス

３．学会等名
言語処理学会年次大会

４．発表年
2022年

１．発表者名
甲斐 優人，高道 慎之介，塩田 さやか，貴家 仁志，

２．発表標題
音声プライバシーのためのブラックボックス型音声加工法

３．学会等名
日本音響学会秋季大会

４．発表年
2020年

１．発表者名
今泉 遼，増村 亮，塩田 さやか，貴家 仁志，

２．発表標題
方言ラベルを補助特徴量とした End-to-End 日本語方言音声認識

３．学会等名
日本音響学会秋季大会

４．発表年
2020年

１．発表者名
奥野 桜子，塩田 さやか，貴家 仁志

２．発表標題
深層学習に基づくなりすまし検出の言語依存性に関する調査

３．学会等名
情報処理学会 音声言語情報処理研究会

４．発表年
2020年

１．発表者名
宋 裕進，塩田 さやか，高道 慎之介，村上 大輔，松井 知子，猿渡 洋

２．発表標題
短時間発話を用いた話者照合のための音声加工の効果に関する検討

３．学会等名
情報処理学会 音声言語情報処理研究会

４．発表年
2021年

１．発表者名
甲斐 優人，高道 慎之介，塩田 さやか，貴家 仁志

２．発表標題
プライバシー保護のためのカスケード型音声加工法を用いた音声仮名化

３．学会等名
日本音響学会春季大会

４．発表年
2021年

１．発表者名
今泉 遼，増村　亮，塩田 さやか，貴家 仁志

２．発表標題
マルチタスク学習による方言識別を考慮したEnd-to-End日本語方言音声認識

３．学会等名
日本音響学会春季大会

４．発表年
2021年

１．発表者名
Haruna MIYAMOTO, Sayaka SHIOTA, Hitoshi KIYA

２．発表標題
Application of Bandwidth Extension with No Learning to Data Augmentation for Speaker Verification

３．学会等名
The Speaker and Language Recognition Workshop Odyssey（国際学会）

４．発表年
2020年

１．発表者名
Ryo IMAIZUMI, Ryo MASUMURA, Sayaka SHIOTA, Hitoshi KIYA

２．発表標題
Sequence-To-One Neural Networks for Japanese Dialect Speech Classification

３．学会等名
IEEE Global Conference on Consumer Electronics（国際学会）

４．発表年
2020年

１．発表者名
Ryo IMAIZUMI, Ryo MASUMURA, Sayaka SHIOTA, Hitoshi KIYA,

２．発表標題
Dialect-Aware Modeling for End-to-End Japanese Dialect Speech Recognition

３．学会等名
APSIPA Annual Summit and Conference,（国際学会）

４．発表年
2020年

１．発表者名
Hiroto KAI, Shinnosuke Takamichi, Sayaka SHIOTA, Hitoshi KIYA

２．発表標題
Lightweight voice anonymization based on data-driven optimization of cascaded voice modification modules,

３．学会等名
IEEE Spoken Language Technology Workshop（国際学会）

４．発表年
2021年

１．発表者名
Ryoya YAGUCHI, Sayaka SHIOTA, Nobutaka ONO, and Hitoshi KIYA

２．発表標題
Replay Attack Detection Using Generalized Cross-Correlation of Stereo Signal

３．学会等名
EUSIPCO 2019（国際学会）

４．発表年
2019年

１．発表者名
Ryota KAMINISHI, Haruna MIYAMOTO, Sayaka SHIOTA, and Hitoshi KIYA

２．発表標題
Blind bandwidth extension with a non-linear function and its evaluation on x-vector-based speaker verification

３．学会等名
Interspeech 2019（国際学会）

４．発表年
2019年

１．発表者名
Haruna MIYAMOTO, Sayaka SHIOTA, and Hitoshi KIYA

２．発表標題
Investigation on Latency Issues and Objective Measurements of Non-Linear Blind Bandwidth Extension

３．学会等名
IEEE GCCE 2019（国際学会）

４．発表年
2019年

１．発表者名
Ryoya YAGUCHI, Sayaka SHIOTA, Nobutaka ONO, and Hitoshi KIYA

２．発表標題
Improving replay attack detection by combination of spatial and spectral features,

３．学会等名
APSIPA ASC 2019（国際学会）

４．発表年
2019年

１．発表者名
宮本 春奈，塩田 さやか，貴家 仁志

２．発表標題
話者照合のための非線形帯域拡張法を用いたデータ拡張の検討，

３．学会等名
音声言語情報処理研究会

４．発表年
2019年

１．発表者名
奥野 桜子，塩田 さやか，貴家 仁志

２．発表標題
ASVspoof 2019 データを用いた環境ごとにおける なりすまし検出の性能評価に関する調査

３．学会等名
音声言語情報処理研究会

４．発表年
2019年

１．発表者名
今泉　遼，塩田 さやか，貴家 仁志

２．発表標題
HMMおよびEnd-to-End音声認識における非線形帯域拡張法の性能調査

３．学会等名
音声言語情報処理研究会

４．発表年
2019年

| |
|---|
| １．発表者名<br>宮本 春奈，塩田 さやか，貴家 仁志 |
| ２．発表標題<br>深層学習に基づく話者照合システムのための非学習型帯域拡張法を用いたデータ拡張 |
| ３．学会等名<br>音声言語情報処理研究会 |
| ４．発表年<br>2020年 |

| |
|---|
| １．発表者名<br>今泉　遼，増村　亮，塩田 さやか，貴家 仁志 |
| ２．発表標題<br>系列分類型ニューラルネットワークを用いた日本語方言識別の検討 |
| ３．学会等名<br>音声研究会 |
| ４．発表年<br>2020年 |

| |
|---|
| １．発表者名<br>甲斐 優人，塩田 さやか，貴家 仁志 |
| ２．発表標題<br>ブラックボックス型敵対的攻撃に対する話者照合システムの脆弱性に関する調査 |
| ３．学会等名<br>音声研究会 |
| ４．発表年<br>2020年 |

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

| | 氏名<br>（ローマ字氏名）<br>（研究者番号） | 所属研究機関・部局・職<br>（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 |
| --- | --- |