人間と情報環境の共生インタラクション基盤技術の創出と展開
2018 年度採択研究代表者

2019 年度
実績報告書

山岸 順一

情報システム研究機構国立情報学研究所コンテンツ科学研究系
教授

VoicePersonae: 声のアイデンティティクローニングと保護

§1. 研究成果の概要

音声は手軽で、自然で、直感的なモダリティの一つである。また同時に、音声は我々のアイデンティティの一部でもあり、生体認証、音声合成、声質変換、プライバシーなど複数の分野において重要なファクターである。しかし、これらの分野では相反する目標に向けて個別に研究が進められている。

日仏共同 CREST “VoicePersonae”プロジェクトでは、声のアイデンティティに関する分野の壁を取り除き、①話者アイデンティティのモデル化技術を高精度化し、②音声による生体認証「話者認識」の安全性と頑健性を高め、③音声のプライバシー保護に関する新しい技術を実現することを目標としている。また、音声変換とライブネス検知と言った目的が相反する技術をどちらも加速させる敵対的競争型研究を実施し、分野を牽引する。さらに、顔、歩容、指紋等の他の生体情報へも研究成果を適用することで、無数のセンサーにより個人の生体情報が容易に取得され得ることが予想される社会においても、我々のアイデンティティの利活用と保護を両立するための基盤技術を確立し、アイデンティティ処理という新たな科学技術分野と研究潮流を作ることを目指している。

2019 年度には、以下の業績を達成した。

まず、話者アイデンティティのモデル化に関する要素基盤技術の高精度化のため、従来の音声信号処理と機械学習を融合させ、ソースフィルターボコーダー、Harmonic plus noise model、Code excited linear prediction といった従来の音声生成手法をニューラルネットワークで高精度化させた新たなニューラル波形モデルを提案した。また、音声合成や声質変換という異なる音声生成タスクを統合したニューラルネットワークを提案し、音声生成タスクの方法論自身を発展させることを行った。従来、音声合成や声質変換は個別に研究されてきたが、提案モデルにより統一的な枠組みで考えることが可能になり、データベース等の学習データも相互利用することが可能になった。さらに、テキスト音声合成から声質変換へニューラルネットワークを転移学習させる新たな枠組みも提案し、分野の壁を取り除き融合させることのメリットを示した。その他、声の個人性を緻密に再現する音声

合成技術の利活用例として、日本の伝統芸能である落語を音声合成で再現することに挑戦し、合成システムが聞き手を楽しませることが可能かどうかも調査した。

また、話者認識の安全性と頑健性の向上のため、音声のなりすまし攻撃を自動的に防御する「ライブネス検出」に関しても力を注いだ。大規模データベースの構築と無償提供、チャレンジの開催、国際ジャーナルにおける特集号や国際会議でのスペシャルセッション開催を行い、基礎研究分野を牽引した。個人認証と生体検知の統合スコアも開発し、学術的にも大きく発展した。更に、上記ライブネス検出の知見を、現在、社会で問題視されている deepfake ビデオに適応させることも行った。カプセルネットワークという構造を利用することにより、deepfake ビデオを高精度に識別できることを示し、また、改ざんされたピクセル領域を同時に予測するネットワークも提案した。改竄されている領域を示すことにより、deepfake であるとの根拠を示せ、説明可能性が一段と向上した。

また、音声のプライバシー研究、具体的には、音声に含まれる話者性の匿名化についても成果を挙げた。音声の自然性や音声から知覚可能な年代や性別といった話者の属性情報を保ったまま、音声の個人性を変えることを目的とする新たな話者匿名化を提案し、その有効性を示した。高品質な音声波形の再合成にはニューラル波形モデルを利用した。その他、音声のプライバシーとセキュリティーに関する Special Interest Group を設立し、ワークショップも開催した。

【代表的な原著論文】

1. Massimiliano Todisco, Xin Wang, Ville Vestman, Md Sahidullah, Héctor Delgado, Andreas Nautsch, Junichi Yamagishi, Tomi Kinnunen, Nicholas Evans, Kong Aik Lee, “ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection”, Interspeech 2019, Graz, Austria, pp.1008-1012, Sept 2019
2. Huy Nguyen, Fuming Fang, Junichi Yamagishi, Isao Echizen, “Multi-task Learning For Detecting and Segmenting Manipulated Facial Images and Videos”, The Tenth IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2019) Sept 2019
3. Fuming Fang, Xin Wang, Junichi Yamagishi, Isao Echizen, Massimiliano Todisco, Nicholas Evans, Jean-Francois Bonastre, “Speaker Anonymization Using X-vector and Neural Waveform Models”, 10th ISCA Speech Synthesis Workshop (SSW10) , pp.155-160, Sept. 2019

§2. 研究実施体制

(1)NII グループ

① 研究代表者:山岸 順一 (情報システム研究機構国立情報学研究所コンテンツ科学研究系教授)

② 研究項目

・声のアイデンティティのモデル化に関する理論的統合

・音声合成および声質変換に関する研究

・話者認識の安全性と頑健性の向上に関する研究

・音声のプライバシー保護に関する研究

・他の生体情報におけるライブネス検出の研究