

山岸 順一

国立情報学研究所コンテンツ科学研究系
准教授

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異なる分野において重要なファクターである。しかし、現在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。

2018年度には、以下の3つの主だった業績を達成した。

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、**ニューラルソースフィルタモデル**という新たなニューラル波形モデルを提案した。本モデルは、従来、音声合成、声質変換等に標準的に使われてきたソースフィルタボコーダをニューラルネットワーク拡張したものである。具体的には、音源信号を Stacked dilated convolution neural network によりフィルタリングする音声生成モデルであり、合成音声の品質および話者性の大幅向上に繋がった。Wavenet 等の類似ニューラル波形モデルと比べ、計算量が少ないというメリットがある。本モデルは今後、音声匿名化等に繋がる重要な研究成果である。

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」を共通コーパスで比較する **ASVspoof challenge 2019** を開催した。音声合成や声質変換によるなりすまし攻撃を想定した Logical access タスクと、音声の単純に再生によるリプレイ攻撃を想定した Physical access タスクの2種類を想定し、それぞれデータベースを構築した。Logical access タスクにおいては、19種類の異なるアルゴリズムによる合成音声や変換音声を、複数企業の協力のもと用意した。Physical access タスクにおいては、様々な条件によるリプレイをシミュレーションにより大量に生成した。本大規模データベースはチャレンジ参加希望の 154 組織に配布され、参加者は各々のライブネス検知モデルを構築した。これにより話者認識の安全性と頑健性は非常に高まることが確認され、ベストなライブネス検知モデルは非常に高度ななりすまし攻撃でもたった 0.22%の等価エラー率で識別ができる事が確認された。

最後に、上記ライブネス検出の知見を他の生体情報への適用することも行った。具体的には、社会問題ともなっている deepfake や face2face という技術により自動生成され、見た目は非常に自然だが**偽の顔映像**を対象とした。そして、カプセルネットワークというニューラルネットワーク技術を利用することで、高精度に識別できることを実験から示した。

【代表的な原著論文】

1. Xin Wang, Shinji Takaki, and Junichi Yamagishi, “Neural source-filter-based waveform model for statistical parametric speech synthesis”, Proc. ICASSP 2019, pp. 5916-5920, 2019年5月
2. Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen, “Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos”, Proc. ICASSP 2019, pp. 2307-2311, 2019年5月
3. Md Sahidullah, Hector Delgado, Massimiliano Todisco, Tomi Kinnunen, Nicholas Evans, Junichi Yamagishi, and Kong-Aik Lee, “Chapter 15, Introduction to Voice Presentation Attack Detection and Recent Advances” Handbook of Biometric Anti-Spoofing, 2nd edition), Springer 2019年 ISBN:978-3-319-92626-1

§ 2. 研究実施体制

(1) NII グループ

① 研究代表者: 山岸 順一(国立情報学研究所コンテンツ科学研究系 准教授)

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