

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

2021年度 年次報告書

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

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アイデンティティに関する機械学習技術を高精度化すると同時に、②生体認証の安全性と頑健性を高め、③プライバシー保護に関する新しい技術を実現する。また音声変換と生体検知、匿名化と再識別化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する。更に他のモダリティ情報へ研究成果を適用し、アイデンティティの活用と保護を両立させる技術を確立させることを目指し、研究を行っている。

本年度は、我々がこれまで取り組んできた技術の高度化を行うと同時に実データにおける評価も行なった。音声のディープフェイク検出を行う技術においては、(a)リアル・フェイクの判断に加え、改ざん可能性がある領域を推定し表示する技術の開発と、(b)異なるハイパーパラメータや特徴量に対しても安定して学習が行える学習ロスの提案を行なった。また実シナリオで評価・性能比較するため、電話越しで検出を行うシナリオ、Youtube 等の圧縮された音声に対して検出を行うシナリオ等を設定し、其々に適した追加テストデータを新たに収集、公開した。また、聞き手側の環境に存在する雑音に負けないよう音声を予め変換する「音声明瞭性強調」に関しても改良を行なった。具体的にはオンライン会議の様に話し手の環境と聞き手の環境が異なっている状況において、話し手の環境ノイズを除去しつつ、同時に、聞き手の環境で音声の明瞭性を強調する技術とそのネットワークの最適化法を提案した。その他、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「話者匿名化」の改良も行なった。洗練されたネットワーク構造へ改良すると同時に、言語依存教師データが必要なモジュールを無くし、他言語へ容易に適用可能な枠組みへと改良した。

これに加え、音声合成・声質変換による合成音声の自然性の主観評価値を予測する技術、文章の内容を知識 DB と比較して事実であるかどうかを機械学習により検証する自動ファクトチェック技術、複数顔が写る映像に対する deepfake 判別・匿名化技術、画像から Crowd counting 手法によりマスク着用率を推定する技術などの新たな研究課題にも着手し、それぞれ国際会議論文として採択された。

§ 2. 研究実施体制

(1) 日本 NII グループ (JST)

① 研究代表者: 国立情報学研究所 コンテンツ科学研究系 教授 山岸順一

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

(2) 仏 Eurecom グループ (ANR)

- ・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

(3) 仏 Avignon グループ (ANR)

- ・音声のプライバシー保護に関する研究
- ・話者認識の安全性と頑健性の向上に関する研究

【代表的な原著論文情報】

[1] Haoyu Li, Junichi Yamagishi, "Multi-Metric Optimization Using Generative Adversarial Networks for Near-End Speech Intelligibility Enhancement", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol.29, pp.3000--3011, Sept 2021

[2] Lin Zhang, Xin Wang, Erica Cooper, Junichi Yamagishi, Jose Patino, Nicholas Evans, "An Initial Investigation for Detecting Partially Spoofed Audio", *Interspeech 2021*, pp.4264-4268, Sept 2021

[3] Cooper, Erica, Wen-Chin Huang, Tomoki Toda, and Junichi Yamagishi. "Generalization ability of MOS prediction networks." *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8442-8446. IEEE, 2022.

[4] Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen, "OpenForensics: Large-Scale Challenging Dataset For Multi-Face Forgery Detection And Segmentation In-The-Wild", *International Conference on Computer Vision (ICCV) 2021*, pp. 10117 -- 10127, Oct 2021

[5] Canasai Kruengkrai, Xin Wang, Junichi Yamagishi, "A Multi-Level Attention Model for Evidence-Based Fact Checking", *Findings of ACL2021*, pp.2447-2460, Aug, 2021