2024 年度年次報告書
社会変革に向けた ICT 基盤強化
2023 年度採択研究代表者

Wang Xin

情報・システム研究機構 国立情報学研究所
特任准教授

プライバシー保護と偽音声検出を統合する音声データ処理基盤

研究成果の概要

During the financial year 2024, this project conducted research on speech deepfake detection and speaker anonymization. On speech deepfake detection, the first outcome is the creation of the ASVspoof 5 dataset – a database with a massive amount of public domain data from more than 1,000 speakers, both advanced deepfake and legacy spoofing attacks, and carefully designed protocols. This openly accessible (doi: 10.5281/zenodo.14498690) dataset is one of the largest datasets in the research community and has been downloaded more than 6,000 times since its release in December 2024. A paper describing all the technical details of the dataset is under review.

With the new dataset, this project organized the ASVspoof 5 challenge as a satellite event of ISCA Interspeech 2024. This project summarized the findings and published them in [paper 1]. Notably, the log-likelihood-ratio-based fusion of speaker verification and deepfake detection, which was proposed by this project in the previous financial year, has been used by two of the top three teams in track 2 of the ASVspoof 5 challenge. The paper has been accepted by ISCA Interspeech 2024.

On speaker anonymization, the major outcome is the pilot study on multi-speaker speaker anonymization. In collaboration with young researchers from Singapore, this project investigated speaker anonymization when the input utterance contains voices from multiple speakers. This is a novel direction as the current research community assumes that the input utterance is from a single speaker. The project proposed baseline algorithms and presented the initial findings in [paper 2].

This project will use the new dataset and other resources created during this financial year for the research in the next financial year.

【代表的な原著論文情報】

1) Xin Wang, Héctor Delgado, Hemlata Tak, Jee-weon Jung, Hye-jin Shim, Massimiliano Todisco, Ivan Kukanov, Xuechen Liu, Md Sahidullah, Tomi Kinnunen, Nicholas Evans, Kong Aik Lee, and Junichi Yamagishi. 2024. ASVspoof 5: Crowdsourced speech data, deepfakes, and adversarial attacks at scale. In Proc. *ASVspoof Workshop 2024*, 2024. 1-8.

2) Xiaoxiao Miao, Ruijie Tao, Chang Zeng, and Xin Wang. 2025. A Benchmark for Multi-speaker Anonymization. *IEEE Transactions on Information Forensic Security*, 2025 Vol.20, 3819-3833.