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　６ 年　 ４ 月 ２７ 日現在

機関番号：６２６１５
研究種目：若手研究
研究期間：2021～2023
課題番号：２１Ｋ１７７７５

研究課題名（和文）Speech privacy protection by high-quality, invertible, and extendable speech anonymization and de-anonymization

研究課題名（英文）Speech privacy protection by high-quality, invertible, and extendable speech anonymization and de-anonymization

研究代表者
Ｗａｎｇ　Ｘｉｎ（Wang, Xin）

国立情報学研究所・コンテンツ科学研究系・特任准教授

研究者番号：６０８４３１４１

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：音声データに含まれる話者識別情報の保護は現代において不可欠な技術である。深層学習に基づく話者匿名化技術は数多く存在するが、このプロジェクトは、主に三つの未解決課題に焦点を当てて研究を進めてきた。まず、自己教師あり学習モデルを用いた言語に依存しない話者匿名化システムを提案した。提案されたシステムは、英語と中国語の音声データに対して優れた匿名化性能を示した。さらに、従来のk-匿名化の問題点に着目し、話者ベクトル回転という匿名化アルゴリズムを提案した。最後に、提案手法を大規模音声データベースであるVoxCeleb2に適用し、匿名化されたデータベースの有用性とプライバシー保護性能を調査した。

研究成果の学術的意義や社会的意義
学術的成果として、現存の深層学習に基づく話者匿名化技術の言語依存性を着目し、複数の言語にも適用できる話者匿名化技術を開発した。また、従来のk-匿名化手法より、話者ベクトルの全体の分布を維持しながら匿名化が可能な手法を提案した。最後に、音声分野において初めてデータベース全体の匿名化を行い、有用性とプライバシー保護性能を調査した。いずれもの成果は音声分野のトップジャーナルや学会で発表された。そのほか、国際的なVoicePrivacyChallengeの運営にも貢献した。提案された技術はテレビ放送に使われたこともあった。

研究成果の概要（英文）：Protecting the personally identifiable information encoded in speech waveform is urgent for many SNS applications. Although there are quite a few deep-learning-based methods trying to project or anonymize the speaker identity in speech data, their solutions are not satisfying. The main contributions of this project can be summarized in three aspects. First, this project proposed language-independent speaker identity anonymization using self-supervised learning speech models. The proposed system was applied to both English and Mandarin data. Second, this project proposed a new anonymization algorithm based on vector rotation. This alleviates the issue of the k-anonymity anonymization in existing methods. Third, this project took the initiative to anonymize a real large-scale speech database called VoxCeleb2 and investigated the utility and privacy protection performance. The research outcomes were published in top journals and conferences in the speech field.

研究分野：知覚情報処理

キーワード：プライバシー　音声匿名化　話者識別　音声情報処理　深層学習

１．研究開始当初の背景

When this project started, many users uploaded multi-media data to social network software (SNS), e.g., YouTube, Facebook, and Twitter. Meanwhile, there has been growing concern that attackers can extract private information from the personal data hosted on the SNS or Internet. This led to regulations such as GDPR in Europe and APPI in Japan.

However, privacy protection faced challenges, especially in the case of speech data: 1) "Given the diversity and ubiquity of applications that now capture, store and process speech signals, … privacy has no formal, legal definition" [1, sec.1]. 2) even if we narrow the definition of speech privacy to speaker identity, i.e., the speaker's voiceprint in a speech waveform, existing technologies were limited in privacy, utility, and applications by then.

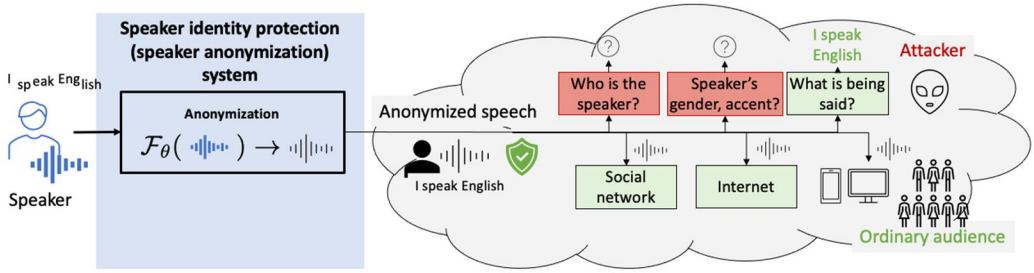


Figure 1. Application of speaker identity protection system in the cyberspace

In a more formal definition, a system that protects the speaker identity in a speech waveform can be described as a function $F_\theta: \mathbb{R}^T \to \mathbb{R}^T$ that converts the input speech $o_{1:T} = (o_1, \cdots, o_T) \in \mathbb{R}^T$ of length T to an anonymized version $\tilde{o}_{1:T} = F_\theta(o_{1:T})$. Ideally, no one can infer the speaker's identity from $\tilde{o}_{1:T}$. Meanwhile, the speech content in $\tilde{o}_{1:T}$ stays close to that in $o_{1:T}$. This idea is illustrated in Figure 1. The system is also referred to as a voice anonymization system.

The most popular framework at that time used deep neural networks (DNNs) [2] and implemented F_θ as an analysis-anonymization-synthesis pipeline (Figure 2). It extracts speech content features (BN features) using a speech recognition (ASR) model, the speaker identity representation using x-vector [2], and the F0 using an F0 extractor. After transforming the x-vector-based speaker identity with an anonymizer, it produces the protected or anonymized speech with a neural source-filter (NSF) waveform model. However, the existing method has a few limitations: 1) requires a language-dependent ASR system, 2) uses a *k*-anonymity-like anonymizer that is limited in privacy protection performance, and 3) considers only speaker identity and anonymization of a single speech waveform.

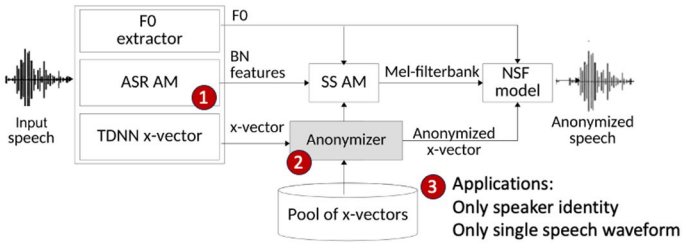


Figure 2. Typical speaker anoymization system using a analysis-anonymization-synthesis pipeline

２．研究の目的

This project aims at better solutions to privacy protection in speech data with three goals:

Goal 1: Towards language-independent

Most of the speaker anonymization systems require language-dependent modules, e.g., ASR for English. This project aims at a language-independent system that can be applied to different languages.

Goal 2: New anonymization algorithm beyond k-anonymity

The standard way to anonymize the x-vector is to replace it with a pseudo x-vector, which is averaged over *n* x-vectors randomly selected from the *k* farthest "neighbors" in the pool ($n<k$). This algorithm tends to produce similar pseudo x-vectors for different speakers [3], which degraded privacy protection and hurt the quality of anonymized speech. This project aims to define a new anonymization algorithm.

Goal 3: Beyond speaker identity in a single waveform

No study has tackled the task of anonymizing a whole speech database. Whether the anonymized speech database is useful for downstream tasks remain unknown. Furthermore, anonymization of other privacy information like gender is not well explored. This study aims to fill the gap.

３．研究の方法

For goal 1, this project leavages the self-supervised learning (SSL) speech models [4] to replace the ASR system. Although they are widely used in many fields, large-scale SSL speech models were new topics at the time. Unlike a language-specific ASR, a good SSL model works across languages. This project explored the use of an SSL model in the anonymization system. Experiments were done in English and Mandarin.

For goal 2, this project investigates theoretical conditions that strike the privacy protection and utility performance (i.e., the naturalness of the anonymized speech waveform). After that, this project produces a novel algorithm that anonymizes the speaker identity through vector rotation. This rotation is implemented using orthogonal matrices, which are further derived from the combination of Householder transformation and DNNs. Both quantitive and qualitative analyses were done to demonstrate the advantage of the proposed algorithm.

For goal 3, this project applies the proposed anonymization algorithm to anonymize the whole VoxCeleb2 dataset [5], which is widely used by the community but controversial because it contains speech data from many celebrities. This project investigated the privacy protection performance of the anonymized VoxCeleb2. This goal is to make it more challenging to detect the original speakers in the VoxCeleb2. Meanwhile, this project applied the language-independent anonymization system to the task of anonymizing the speaker's gender.

４．研究成果

✔ **Goal 1: This project built a speech anonymization system that is effective on both English and Mandarin datasets.**

This project proposed a speaker anonymization system that leverages the HuBERT-based SSL model [5] to extract speech content features. Additionally, the latest DNN-based speaker representation extractor and waveform generator are used, which is illustrated in Figure 3.

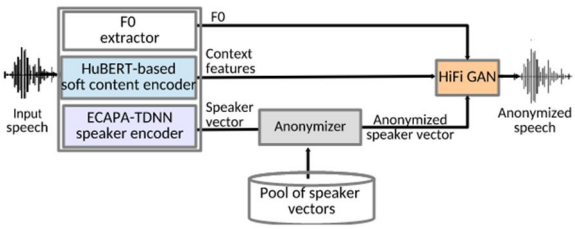


Figure 3. Proposed speaker anonymization system using HuBERT-based SSL model (**学会発表** 2022 Odyssey).

Evaluation was conducted on the VoicePrivacy challenge test sets in English and a publically available Mandarin test set. It was demonstrated that the proposed system's privacy protection performance, which is measured by speaker re-identification equal error rates (EER), is at the same level as the strong VoicePrivacy baselines. Meanwhile, the proposed system better preserved the speech content, which is shown by the lower word error rate (WER) and higher Mean-opinion-score (MOS) on the English task (Figure 4).

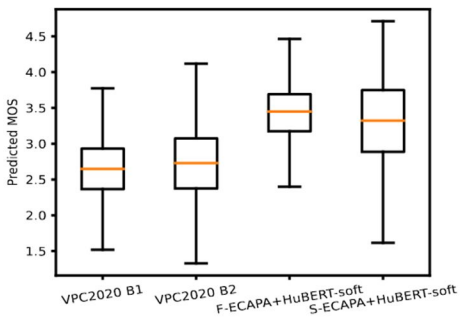


Figure 4. Boxplots of anonymized speech quality score of proposed systems (right two) and VoicePrivacy baselines (left two). (**学会発表** 2022 Odyssey).

Furthermore, the proposed system can be directly used on Mandarin data without any fine-tuning. Both EER and WER were acceptable. In contrast, the conventional systems fail to produce intelligible speech after anonymization.

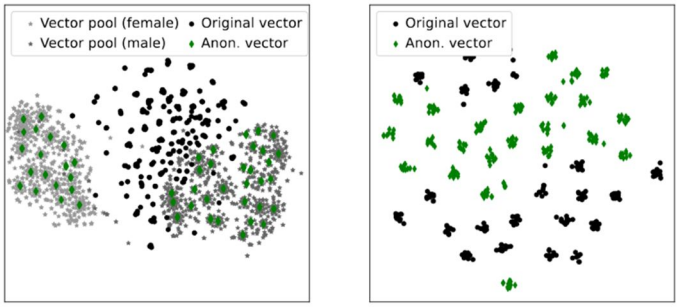
The detailed results are presented at the ISCA Odyssey Workshop 2022 (**学会発表**,2022, Odyssey). The following work analyzed the anonymization across languages and proposed strategies to mitigate the domain mismatch when applying the proposed system on a new language (**学会発表**,2022, Interspeech).

✔ **Goal 2: New anonymization algorithm beyond k-anonymity**

While the work on goal 1 addressed the language issue, the anonymizer was inherited from the conventional methods based on k-anonymity --- given an input x-vector, *k*-farthest neighbors in a speaker pool are retrieved, and *n* out of the *k* x-vectors are randomly selected and averaged to produce the anonymized x-vector.

This project empirically demonstrated that the k-anonymity-based anonymization algorithm is prone to create pseudo speakers close to each other. This not only hurts the privacy performance when facing a stronger attacker (**雑誌論文**,2023, semi-informed attacker in Fig.6) but also degrades the diversity of anonymized speakers in perception.

The proposed method anonymizes the speaker presentation through rotation. Let $x \in \mathsf{R}^d$ be the speaker representation (e.g., x-vector) extracted from the input waveform. The anonymized x-vector $\tilde{x} \in \mathsf{R}^d$ is computed as $\tilde{x} = W(x - \mu_x) + \mu_x$, where $W \in \mathsf{R}^{d \times d}$ is an orthogonal matrix predicted by a DNN and Householder transformation.



(a) S-Select (b) S-LOH*

Figure 5. Distribution of speaker representations (x-vectors) before (in black) and after anonymization (in green). Left figure is from the select-based k-anonymity method. Right figure is from the proposed anonymizer. (**雑誌論文**,2023)

This project showed that the proposed rotation generally keeps the overall distribution of speaker representations before and after anonymization. This is shown in Figure 5. Experiments in English and Mandarin data demonstrated the strong privacy protection performance against all attackers, including the strong semi-informed attackers. The WER and other utility metrics are not degraded. The detailed results are published in a top IEEE journal (**雑誌論文**,2023).

✔ **Goal 3: This project took the initiative to anonymize a whole speech database and investigated the privacy and utility of the anonymized speech database.**

This project applied the proposed system for goal 2 to anonyimizing the famous speech database called VoxCeleb2. The main contribution is to define a proper evaluation framework that measures how useful the anonymized VoxCeleb2 is for downstream tasks such as training speaker verification (ASV) systems.

This project investigated six variants of the anonymization methods and created six versions of the anonymized VoxCeleb2. The differences are whether additive noise, room reverberation, and non-speech sounds are applied to anonymized data.

Experiments showed that the EERs of detecting the original speaker in the six versions of anonymized VoxCeleb2 databases are at least ten times higher than the EER on the original VoxCeleb2 database. This indicates that it is more difficult to detect the original speaker in the anonymized VoxCeleb2 databases.

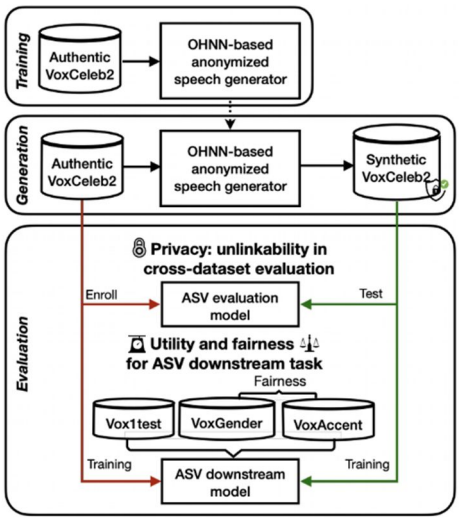


Figure 6. Flowchart of anonymizing VoxCeleb2 and evaluation. (**学会発表** 2024 ICASSP)

The detailed analysis, including additional analysis on fairness across speakers, is presented at the top IEEE conference (**学会発表**2024 ICASSP).

Contributions to the research community and the society

In addition to the research outcomes published in journals and conferences, this project also contributed to the Voice Privacy Challenges. Baseline systems of the Voice Privacy Challenges are created and is publically available: https://github.com/Voice-Privacy-Challenge/Voice-Privacy-Challenge-2022.

The proposed anonyization system was also used in TV broadcasting to anonymize the voice of the speakers. The code is also publically released: https://github.com/nii-yamagishilab/SSL-SAS.

[1] A.Nautsch, et.al., The GDPR & speech data: Reflections of legal and technology communities, first steps towards a common understanding. Proc. Interspeech, 3695-3699, 2019
[2] F.Fang, et.al., Speaker anonymization using X-vector and neural waveform models, Proc. SSW, 155-160, 2019
[3] P.Champion, Anonymizing Speech: Evaluating and Designing Speaker Anonymization Techniques, PhD thesis, LIUM, France, 2023
[4] H.Mohamed, et.al., Self-Supervised Speech Representation Learning: A Review, IEEE Journal of Selected Topics in Signal Processing, 1179-1210, 2022
[5] J.Chung, et.al., VoxCeleb2: Deep speaker recognition, Proc. Interspeech, 1086-1090, 2018

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　3件／うちオープンアクセス　3件）

１．著者名 Miao Xiaoxiao、Wang Xin、Cooper Erica、Yamagishi Junichi、Tomashenko Natalia	４．巻 31
２．論文標題 Speaker Anonymization Using Orthogonal Householder Neural Network	５．発行年 2023年
３．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	６．最初と最後の頁 3681～3695
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2023.3313429	査読の有無 有
オープンアクセス 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 該当する

１．著者名 Srivastava Brij Mohan Lal、Maouche Mohamed、Sahidullah Md、Vincent Emmanuel、Bellet Aurelien、Tommasi Marc、Tomashenko Natalia、Wang Xin、Yamagishi Junichi	４．巻 30
２．論文標題 Privacy and Utility of X-Vector Based Speaker Anonymization	５．発行年 2022年
３．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	６．最初と最後の頁 2383～2395
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2022.3190741	査読の有無 有
オープンアクセス 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 該当する

１．著者名 Tomashenko Natalia、Wang Xin、Vincent Emmanuel、Patino Jose、Srivastava Brij Mohan Lal、No? Paul-Gauthier、Nautsch Andreas、Evans Nicholas、Yamagishi Junichi、O ' Brien Benjamin、Chanclu Ana?s、Bonastre Jean-Fran?ois、Todisco Massimiliano、Maouche Mohamed	４．巻 74
２．論文標題 The VoicePrivacy 2020 Challenge: Results and findings	５．発行年 2022年
３．雑誌名 Computer Speech & Language	６．最初と最後の頁 101362～101362
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.csl.2022.101362	査読の有無 有
オープンアクセス 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 該当する

〔学会発表〕　計11件（うち招待講演　2件／うち国際学会　11件）

１．発表者名 Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, Nicholas Evans, Massimiliano Todisco, Jean-Francois Bonastre, and Mickael Rouvier
２．発表標題 SynVox2: Towards a Privacy-Friendly VoxCeleb2 Dataset
３．学会等名 ICASSP 2024（国際学会）
４．発表年 2024年

１．発表者名
Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, and Natalia Tomashenko

２．発表標題
Analyzing Language-Independent Speaker Anonymization Framework under Unseen Conditions

３．学会等名
Interspeech 2022（国際学会）

４．発表年
2022年

１．発表者名
Paul-Gauthier Noe, Xiaoxiao Miao, Xin Wang, Junichi Yamagishi, Jean-Francois Bonastre, and Driss Matrouf

２．発表標題
Hiding Speaker’s Sex in Speech Using Zero-Evidence Speaker Representation in an Analysis/Synthesis Pipeline

３．学会等名
ICASSP 2023（国際学会）

４．発表年
2023年

１．発表者名
Xin Wang

２．発表標題
Tutorial on speaker anonymization (software part)

３．学会等名
2nd Symposium on Security and Privacy in Speech Communication joined with 2nd VoicePrivacy Challenge Workshop（招待講演）（国際学会）

４．発表年
2022年

１．発表者名
Jean-Francois Bonastre, Hector Delgado, Nicholas Evans, Tomi Kinnunen, Kong Aik Lee, Xuechen Liu, Andreas Nautsch, Paul-Gauthier NoE, Jose Patino, Md Sahidullah, Brij Mohan Lal Srivastava, Massimiliano Todisco, Natalia Tomashenko, Emmanuel Vincent, Xin Wang, Junichi Yamagishi

２．発表標題
Benchmarking and challenges in security and privacy for voice biometrics

３．学会等名
2021 ISCA Symposium on Security and Privacy in Speech Communication（国際学会）

４．発表年
2021年

１．発表者名
Wang Xin

２．発表標題
Two speech security issues after the speech synthesis boom

３．学会等名
Speech Synthesis Forum, China Computer Federation（招待講演）（国際学会）

４．発表年
2021年

１．発表者名
Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, Natalia Tomashenko

２．発表標題
Language-Independent Speaker Anonymization Approach Using Self-Supervised Pre-Trained Models

３．学会等名
Proc. Odyssey 2022 The Speaker and Language Recognition Workshop（国際学会）

４．発表年
2022年

１．発表者名
Wang Xin, Yamagishi Junichi

２．発表標題
Estimating the confidence of speech spoofing countermeasure

３．学会等名
ICASSP 2022（国際学会）

４．発表年
2022年

１．発表者名
Chang Zeng,　Xin Wang, Erica Cooper, Xiaoxiao Miao, Junichi Yamagishi

２．発表標題
Attention Back-end for Automatic Speaker Verification with Multiple Enrollment Utterances

３．学会等名
ICASSP 2022（国際学会）

４．発表年
2022年

１．発表者名
Hemlata Tak, Massimiliano Todisco, Xin Wang, Jee-weon Jung, Junichi Yamagishi, Nicholas Evans

２．発表標題
Automatic speaker verification spoofing and deepfake detection using wav2vec 2.0 and data augmentation

３．学会等名
Proc. Odyssey 2022 The Speaker and Language Recognition Workshop（国際学会）

４．発表年
2022年

１．発表者名
Xin Wang, Junichi Yamagishi

２．発表標題
Investigating self-supervised front ends for speech spoofing countermeasures

３．学会等名
Proc. Odyssey 2022 The Speaker and Language Recognition Workshop（国際学会）

４．発表年
2022年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

Official page of VoicePrivacy:
https://www.voiceprivacychallenge.org/

Open-source baseline of VoicePrivacy 2022:
https://github.com/Voice-Privacy-Challenge/Voice-Privacy-Challenge-2022

Languange-independent speaker anonymization system:
https://github.com/nii-yamagishilab/SSL-SAS

Tutorial on speaker anonymization (software): https://colab.research.google.com/drive/1_zRL_f9iyDvl_5Y2RdakgOhYAl_5Rgyq

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （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			
フランス	Avignon University	Inria	EURECOM	他1機関
韓国	Naver Corporation			