Speaker Anonymization using Orthogonal Householder Neural Network

Xiaoxiao Miao[1], Xin Wang[2], Erica Cooper[2], Junichi Yamagishi[2], Natalia Tomashenko[3]
1 Singapore Institute of Technology, 2 National institute of Informatics, 3 Inria

IEEE/ACM TASLP 2023

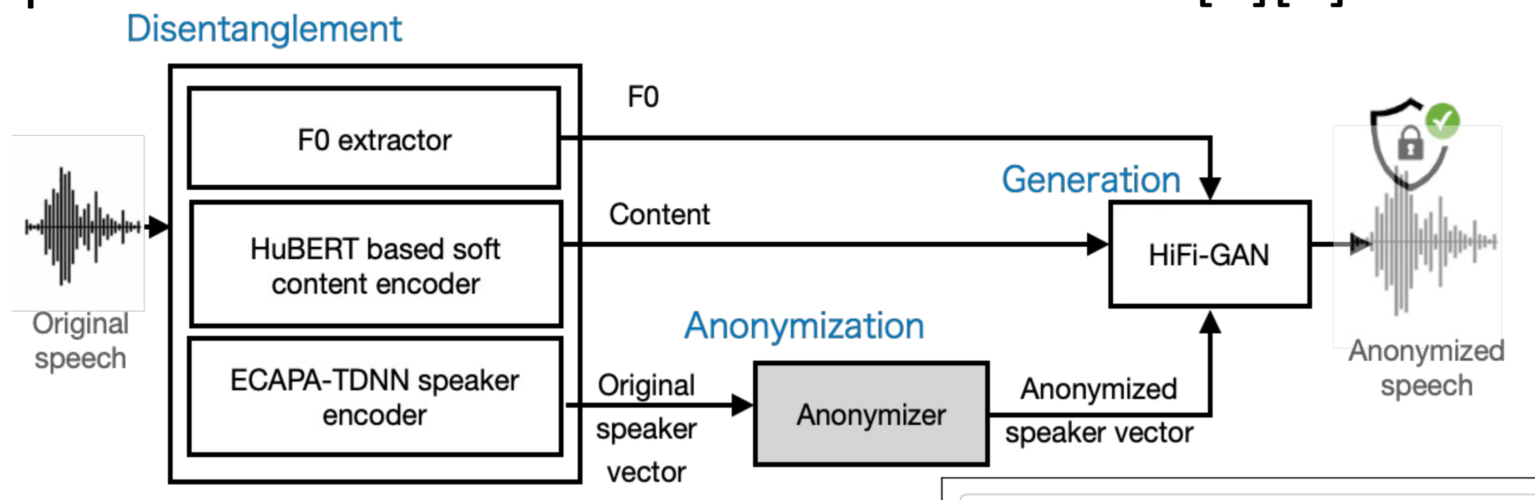
Paper   Audio samples

Introduction

Speaker anonymization: Hide speaker identity (privacy) but keep other information (utility), e.g. linguistic content, speaker diversity, allowing anonymized speech is still useful for downstream tasks [1][2].

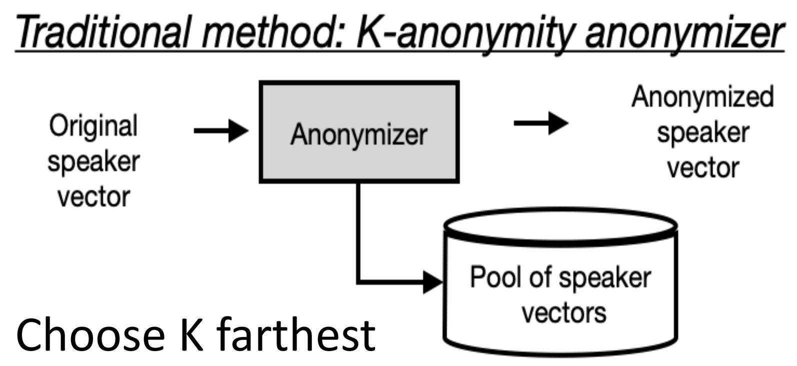
1. **Digital signal processing**: perceptually manipulate speech such as the pitch, spectral envelope, and time scaling

☹ Poor speech intelligibility

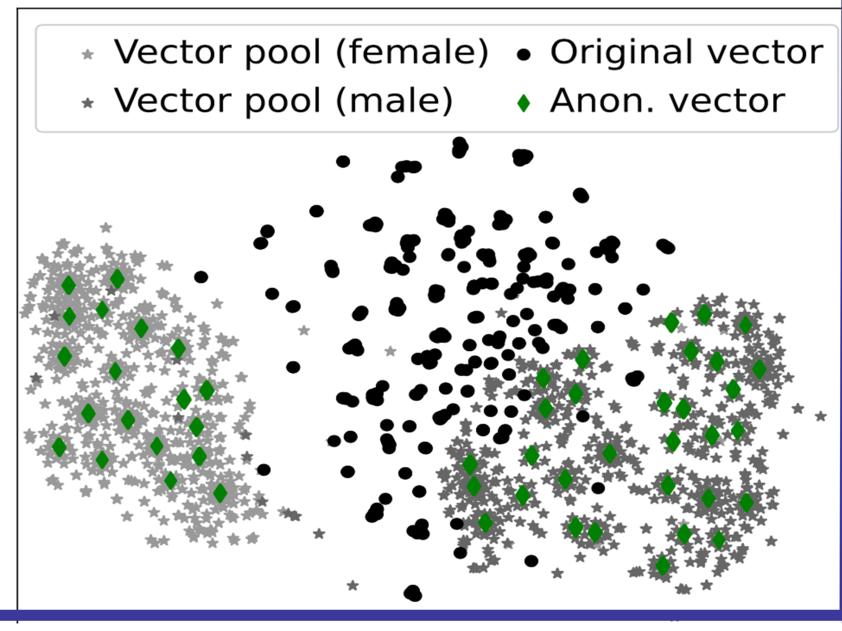


2. **Disentanglement-based speaker anonymization using K-anonymity anonymizer**
- Disentangle speech into F0, content, and original speaker vector
- Speaker vector carries most of the private information

☹ Poor speaker diversity



- Choose K farthest
- Average K* vectors



How to achieve an ideal anonymizer

C1 **Speaker privacy protection**

Anonymized speech sounds dissimilar from original speech to hide the original speaker's identity

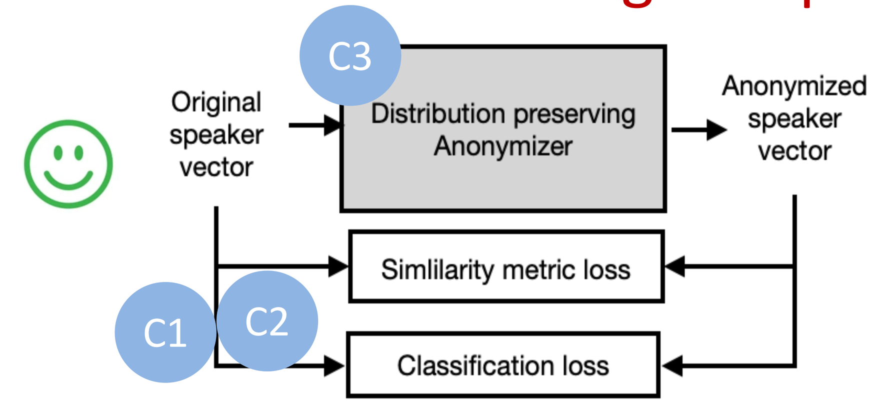
C2 **Speaker diversity**

Anonymized speech from the same speaker has a unique speaker identity to maintain the diversity of anonymized speech across different speakers, allowing anonymized speech used for different downstream tasks, e.g. speaker verification.

C3 **Speech intelligibility and naturalness**

Anonymized speech satisfies the same distribution as original speech

Solution: Learnable anonymizer on speaker embedding vector space
- **Distribution preservation** transformation
- With **constraints**



OHNN-based anonymizer

Distribution-preserving transformation: **Orthogonal Householder neural network:**

Normalization: $\mathbf{x}^o \sim \mathcal{N}(\mu_{\mathbf{x}^o}, \Sigma_{\mathbf{x}^o})$

$\tilde{\mathbf{x}}^o \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

★ **Rotation**: $\tilde{\mathbf{x}}^a = \mathbf{W}\tilde{\mathbf{x}}^o$

OHNN: $\mathbf{W} = \mathbf{W}_1\mathbf{W}_l \ldots \mathbf{W}_L$

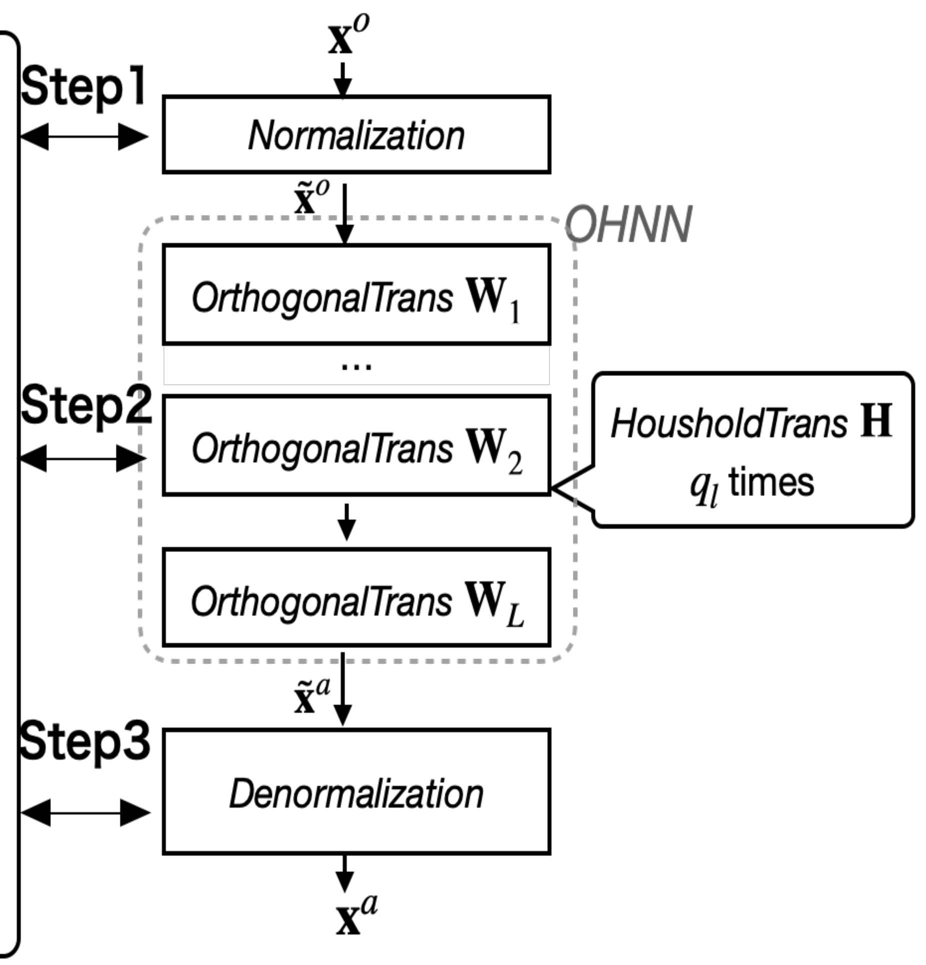
OrthogonalTrans: $\mathbf{W}_l = \mathbf{H}_{q_l}\mathbf{H}_{q_l-1}\ldots\mathbf{H}_1$

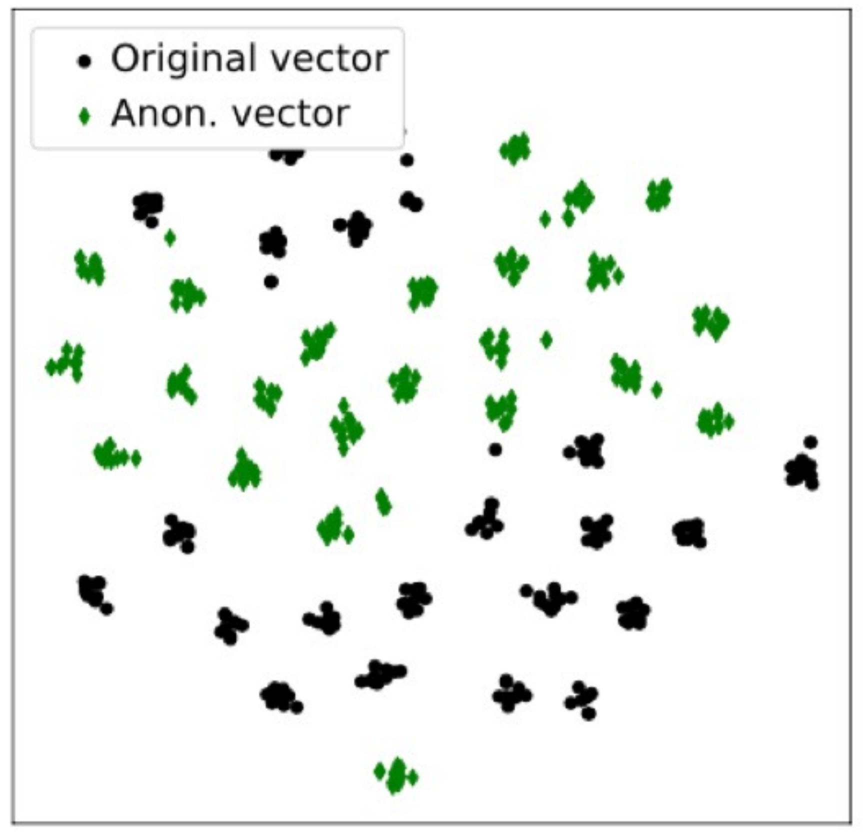
HousholdTrans: $\mathbf{H}_{q_l} = \mathbf{I} - \frac{2}{\mathbf{v}_{q_l}^\top \mathbf{v}_{q_l}} \mathbf{v}_{q_l}\mathbf{v}_{q_l}^\top$

Learnable parameters

Denormalization: $\tilde{\mathbf{x}}^a \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

$\mathbf{x}^a \sim \mathcal{N}(\mu_{\mathbf{x}^o}, \Sigma_{\mathbf{x}^o})$





Interested in speaker anonymization?
The VoicePrivacy Challenge (VPC) 2024 is open. Welcome your contribution!

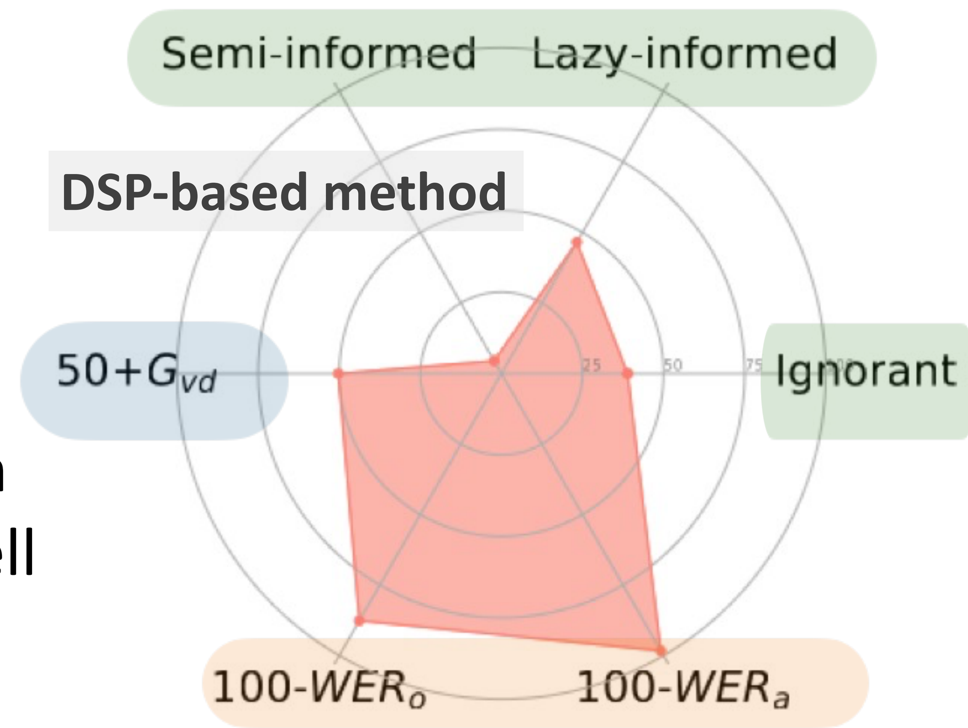
VPC2024

[1] Tomashenko N, Miao X, Champion P, Meyer S, et al The VoicePrivacy 2024 Challenge Evaluation Plan
[2] Tomashenko N, Wang X, Vincent E, et al. The voiceprivacy 2020 challenge: Results and findings[J]. Computer Speech & Language, 2022, 74: 101362.
[3] Miao X, Wang X, Cooper E, Yamagishi J, Tomashenko N. Language-independent speaker anonymization approach using self-supervised pre-trained models, Odyssey 2022.
[4] Patino J, Tomashenko N, Todisco M, Nautsch A, and Evans N. Speaker Anonymisation Using the McAdams Coefficient. Interspeech 2021.

Experiments

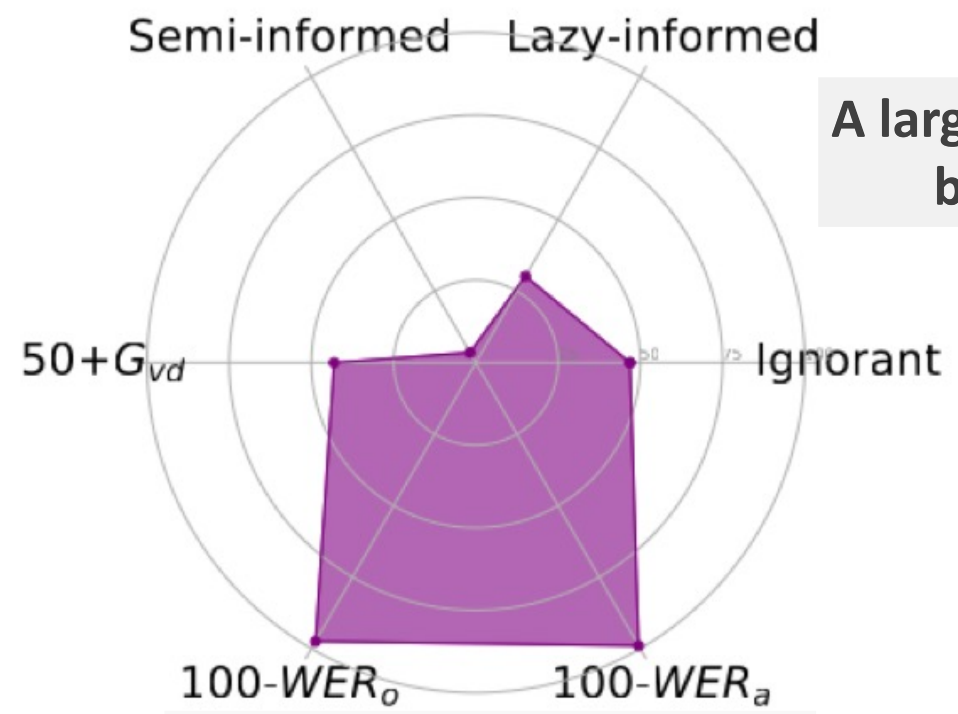
We follow the VoicePrivacy challenge protocol to conduct the experiments

- Privacy metrics (ASV-EER): assess the ability to protect speaker identity in different scenarios
- Utility metrics(Gvd): assess the preservation of voice diversity, similarity between original and anonymized speech
- Utility metrics (ASR-WER): assess how well speech content is preserved in anonymized speech

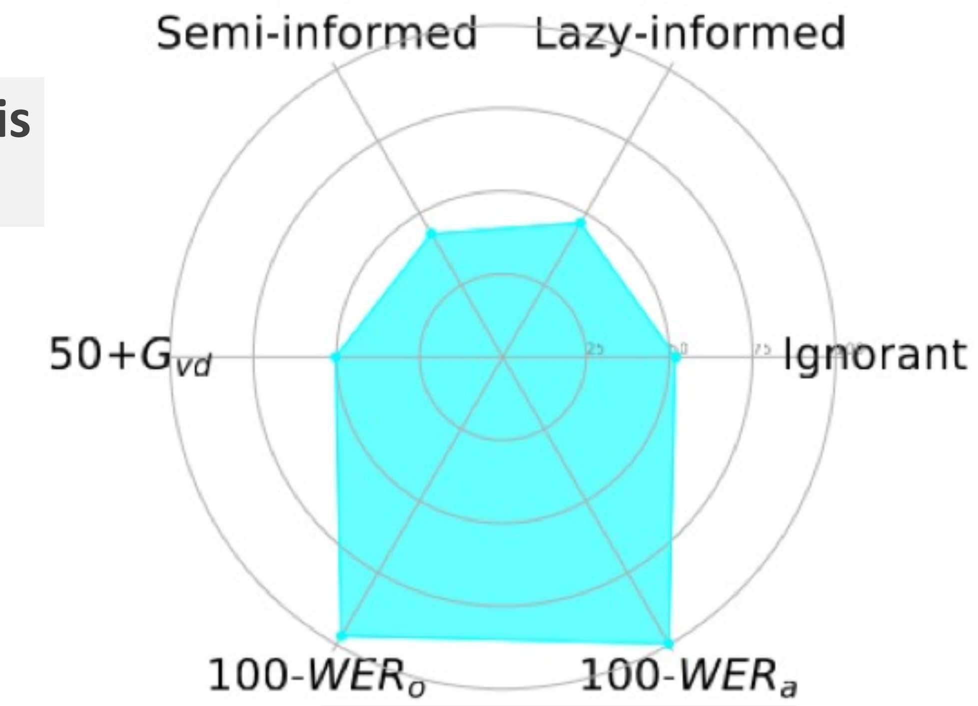


DSP-based method

A larger area is better



K-anonymity anonymizer



OHNN anonymizer