Voice-Indistinguishability

Protecting Voiceprint in Privacy-Preserving Speech Data Release

Yaowei Han, Sheng Li, Yang Cao, Qiang Ma, Masatoshi Yoshikawa
Department of Social Informatics, Kyoto University, Kyoto, Japan
National Institute of Information and Communications Technology, Kyoto, Japan

CONTENT

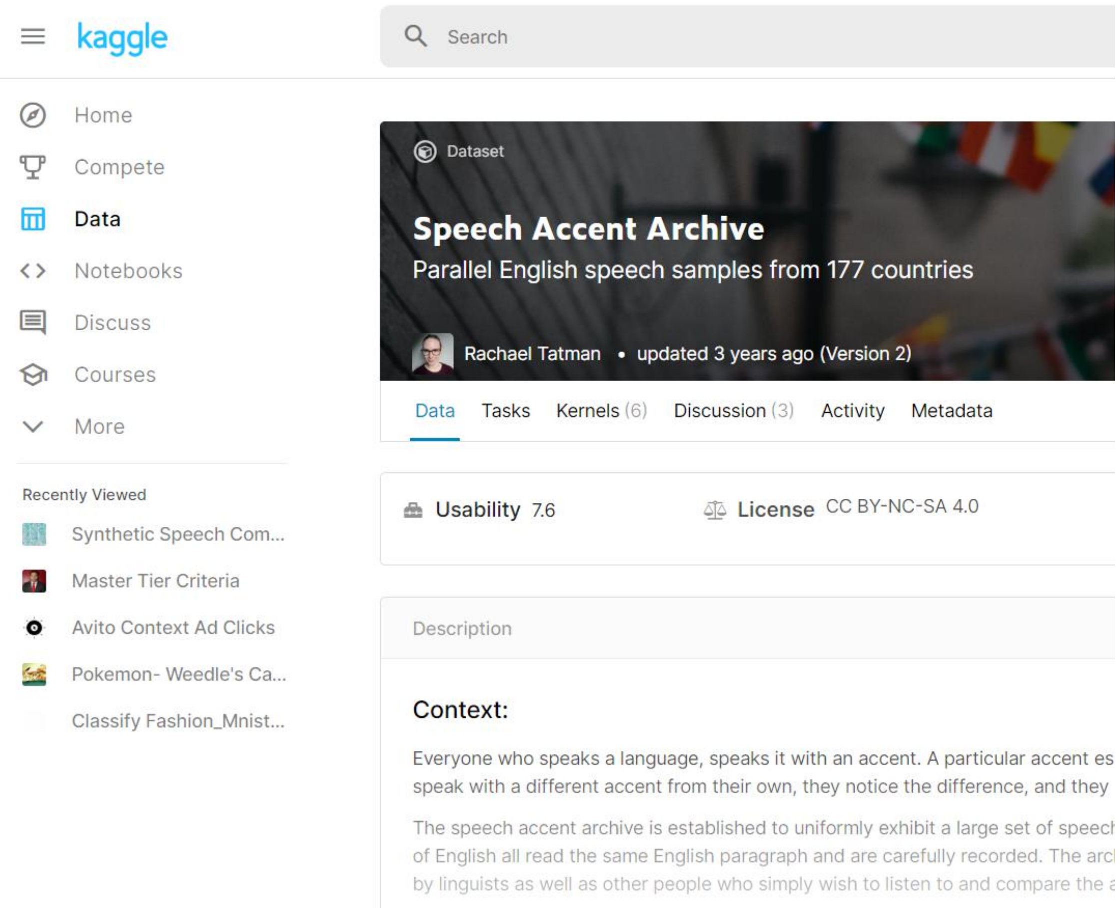
01 Motivation

Motivation - Speech Data Release

Speech Data Release

Share speech dataset with the 3rd parties

Eg. Apple collects speech data for Siri quality evaluation process, which they call grading.



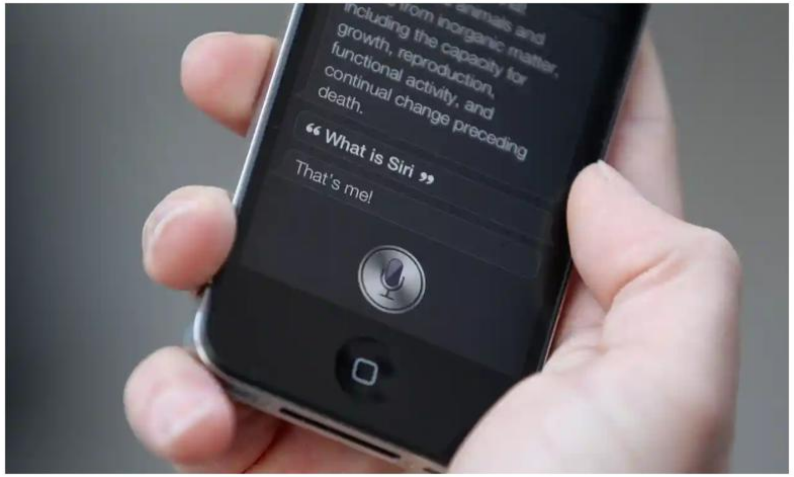
Risks of Speech Data Release

Privacy concern.

- Speech data is personal data.
- Everybody has a unique voiceprint, which is a kind of biometric identifiers.
- GDPR[1] bans the sharing of biometric identifiers.

Apple contractors 'regularly hear confidential details' on Siri recordings

Workers hear drug deals, medical details and people having sex, says whistleblower



[1] A. Nautsch and et al., “The GDPR & speech data:Reflections of legal and technology communities, firststeps towards a common understanding,” 2019.
https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings

Motivation - Risks of Speech Data Release

Risks of Speech Data Release

Security risks.

- Spoofing attacks to the voice authentication systems
- Reputation attacks (fake Obama speech[1])

How to protect privacy in speech data release?

[1] S. Suwajanakorn and et al., "Synthesizing obama: learning lip sync from audio," ACM Transactions on Graphics, 2017.

02 Related Works

Related Works

	Privacy		Voice technology
	protection level	privacy guarantee	
[1][2]	voice-level	ad-hoc	Vocal Tract Length Normalization (VTLN)
[3][4]	feature-level	k-anonymity	Speech Synthesize
[5]	model-level	ad-hoc	ASR

[1] J. Qian and et al., “Hidebehind: Enjoy voice input with voiceprint unclonability and anonymity,” in ACM SenSys 2018.
[2] B. Srivastava and et al., “Evaluating voice conversion-based privacy protection against informed attackers,” arXiv preprint arXiv:1911.03934, 2019.
[3] T. Justin and et al., “Speaker deidentification using diphone recognition and speech synthesis,” in FG 2015.
[4] F. Fang and et al., “Speaker anonymization using X-vector and neural waveform models,” in 10th ISCA Speech Synthesis Workshop, 2019.
[5] B. Srivastava and et al., “Privacy-Preserving Adversarial Representation Learning in ASR: Reality or Illusion?,” in Interspeech 2019.

Related Works - Insufficiency of Existing Methods

Existing methods for protecting speech data privacy

(1) Speech2text (2) K-anonymity

However, they are insufficient because

(1) Speech2text

not useful for speech analysis

without any formal privacy guarantee

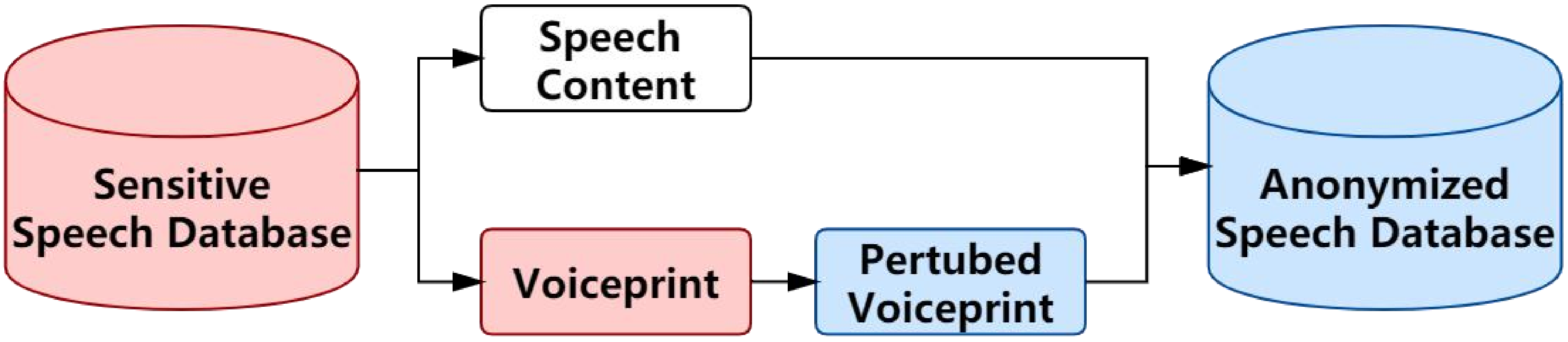
(2) K-anonymity

based on the assumption of attackers' knowledge

(= not secure under powerful attackers)

03 Problem Setting and Contributions

Problem Setting



Privacy-preserving speech data release

We focus on protecting voiceprint, i.e., user voice identity.

Contributions

1 How to formally define voiceprint privacy?

Voice-Indistinguishability

- The first formal privacy definition for voiceprint, not depend on attacker's background knowledge.

How to design a mechanism achieving our privacy definition? 2

Voiceprint perturbation mechanism

- Use voiceprint to present user voice identity
- Our mechnism output a anonymized voiceprint

3 How to implement frameworks for private speech data release?

Privacy-preserving speech synthesis

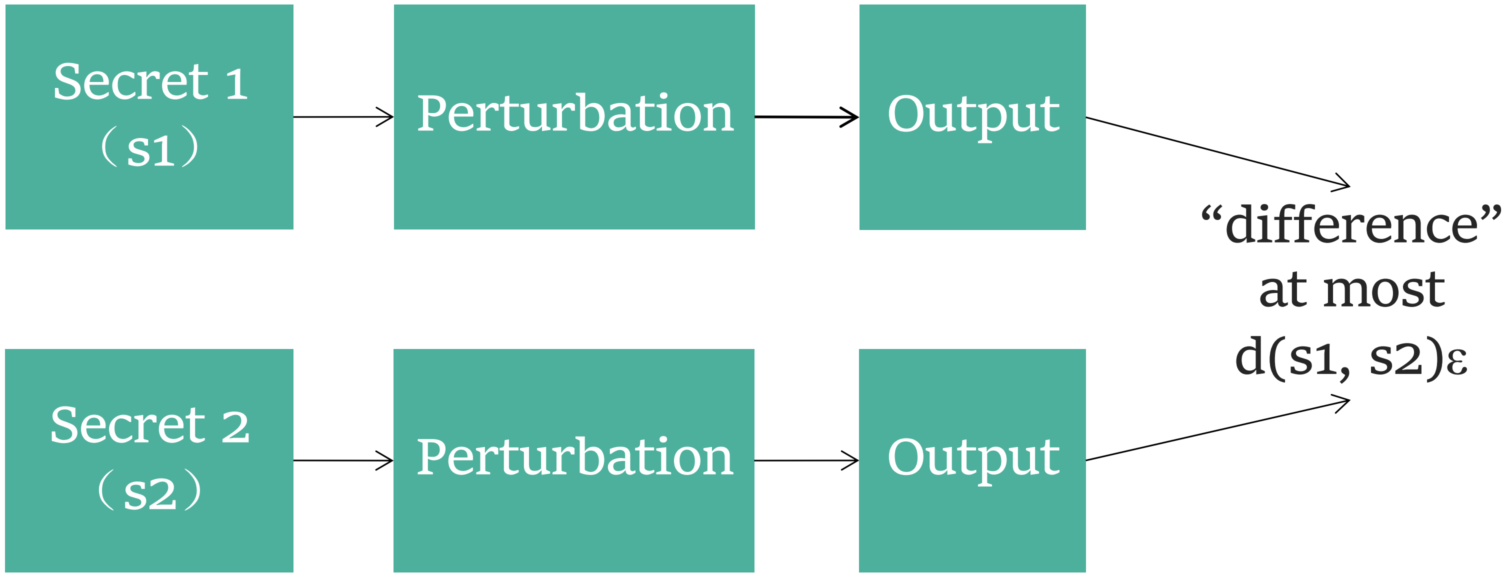
- Synthesize voice record with anonymized voiceprint

04 Our Solution

Our Solution - Metric Privacy

How to formally define voiceprint privacy?

Definition of Metric Privacy



Advantages:

1) Has no assumptions on the attackers’ background knowledge.

2) Privacy loss can be quantified.

   the bigger ε -> the better utility, the weaker privacy

3) d(s1, s2): distance metric between secrets.

Our Solution - Decision of Secrets

When applying metric privacy, we should decide secrets and distance metric.

- What's the secret?

  Voiceprint

- How to represent the voiceprint?

  x-vector[1], a widely used speaker space vector.

  For example. 512 dimensional

  [1.291081 0.9634209 ... 2.59955]

[1] D. Snyder and et al., “X-vectors: Robust dnn embeddings for speaker recognition,” inProc. IEEE-ICASSP,2018, pp. 5329–5333.

Our Solution - Decision of Distance Metric

When applying metric privacy, we should decide secrets and distance metric.

- How to define the distance metric between voiceprint?

Euclidean distance? ✕

Can not well represent the distance between two x-vectors

Cosine distance? ✕

Widely used in speaker recognition but doesn't satisfy triangle inequality

Angular distance? YES

Also a kind of cosine distance but satisfies triangle inequality

Our Solution - Voice-Indistinguishablility

How to formally define voiceprint privacy?

For single user

Voice-Indistinguishability, Voice-Ind

$$\frac{\Pr(\tilde{x}|x)}{\Pr(\tilde{x}|x')} \leq e^{\epsilon d_{\mathcal{X}}(x,x')}$$

$$d_{\mathcal{X}} = \frac{\arccos(\cos\ similarity < x, x' >)}{\pi}$$

For multiple users in a speech dataset

Speech Data Release under Voice-Ind

$$\frac{\Pr(\tilde{D}|D)}{\Pr(\tilde{D}|D')} \leq e^{\epsilon d(D,D')}$$

$$d(D, D') = d_{\mathcal{X}}(x, x')$$

ε: privacy budget
privacy-utility tradeoff

bigger ε :
(1) weaker privacy
(2) better utility

n: speech database size

larger n:
(1) stronger privacy

-> later, we will verify this

Our Solution - Mechanism

How to design a mechanism achieving our privacy definition?

$$\Pr(\tilde{x}|x_0) \propto e^{-\epsilon d_{\mathcal{X}}(x_0,\tilde{x})}$$

Pertubed / Original	A	B	C
A	$\propto e^{0}$	$\propto e^{d(A, B)}$	$\propto e^{d(A, C)}$
B	$\propto e^{d(A, B)}$	$\propto e^{0}$	$\propto e^{d(B, C)}$
C	$\propto e^{d(A, C)}$	$\propto e^{d(B, C)}$	$\propto e^{0}$

Our Solution - Privacy Guarantee

Privacy guarantee of the released private speech database.

Sensitive Speech database

Speaker	Speech Data	Attr
A	Record 1	...
B	Record 2	...
C	Record 3	...
...	...	...

Our Method →

Anonymized Speech database

Speaker	Speech Data	Attr
A	Record 1 (with C's voiceprint)	...
B	Record 2 (with A's voiceprint)	...
C	Record 3 (with B's voiceprint)	...
...	...	...

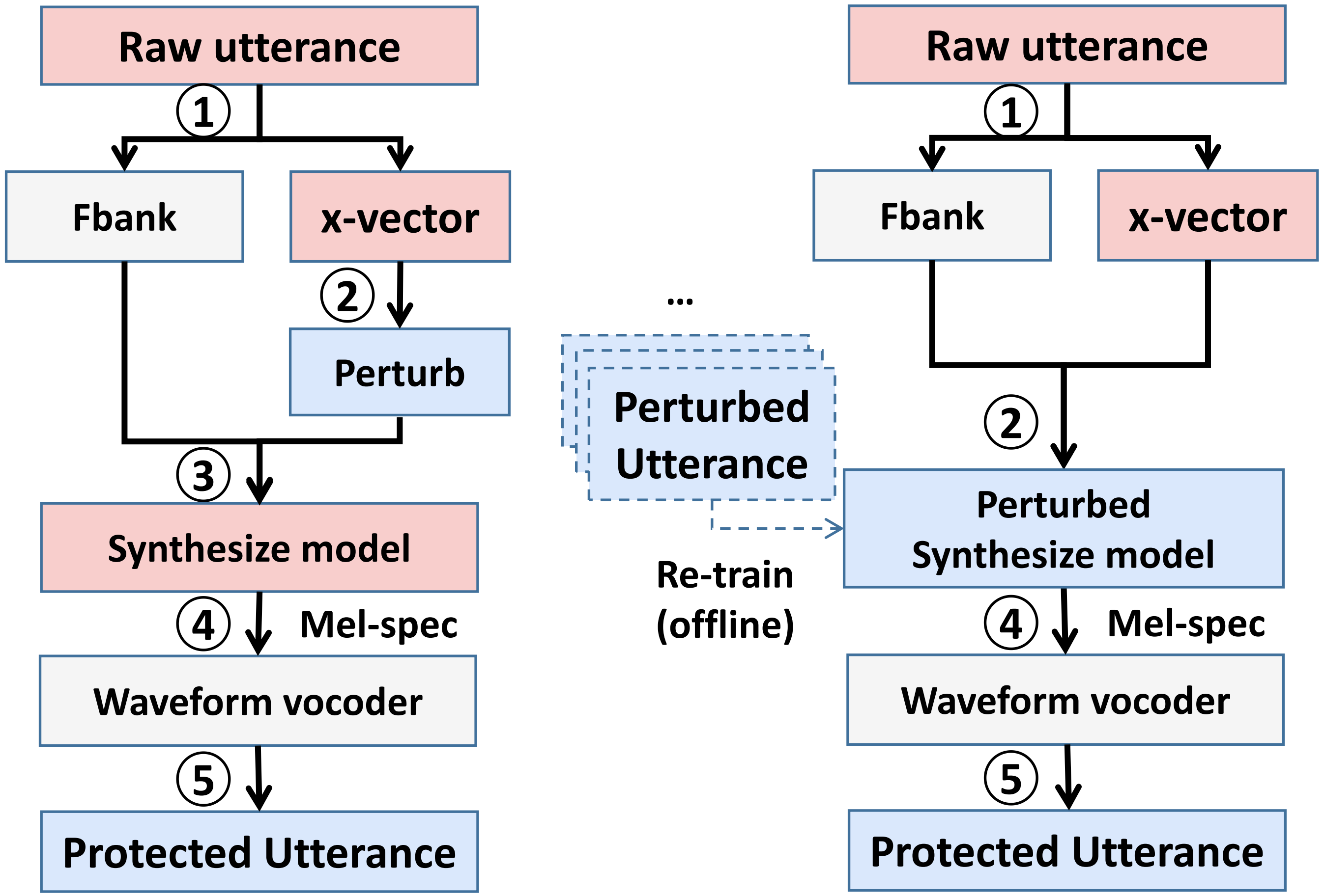
Our Solution

How to implement frameworks for private speech data release?

Voiceprint extraction (unprotected)

Protect voiceprint

Reconstruct waveform (protected)



(a) Feature-level (b) Model-level

Voiceprint extraction (unprotected)

Protect voiceprint

Reconstruct waveform (protected)

05 Experiment and Conclusion

Experiment

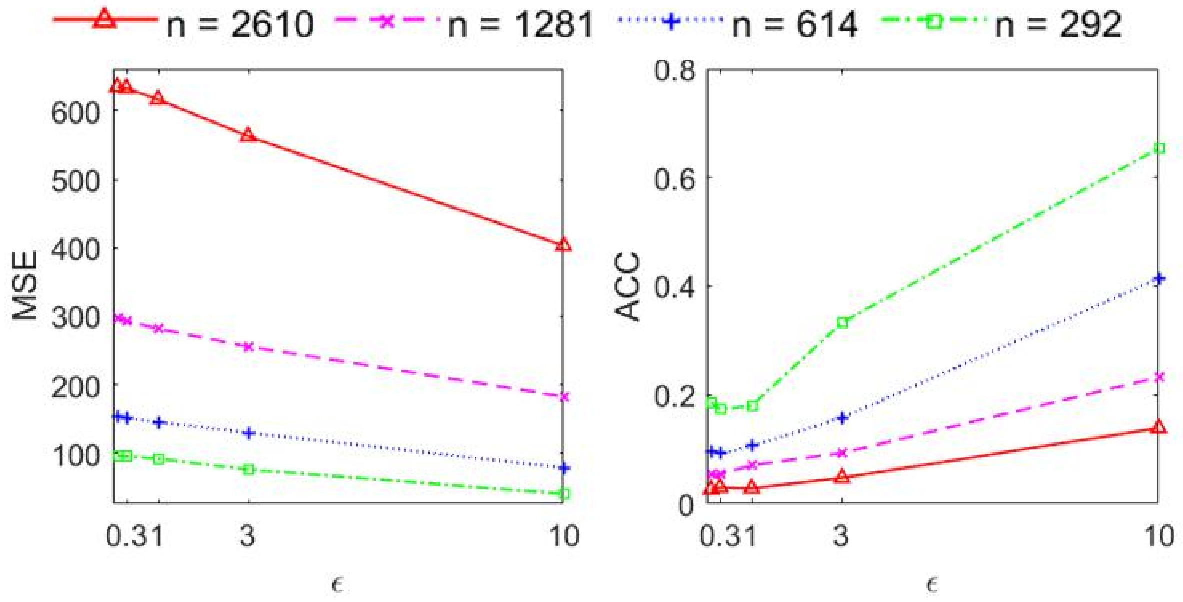
Verify the utility-privacy tradeoff of Voice-Indistinguishability.

- How does the privacy parameter ε affect the privacy and utility?
- How does the database size n affect the privacy?

Experiment

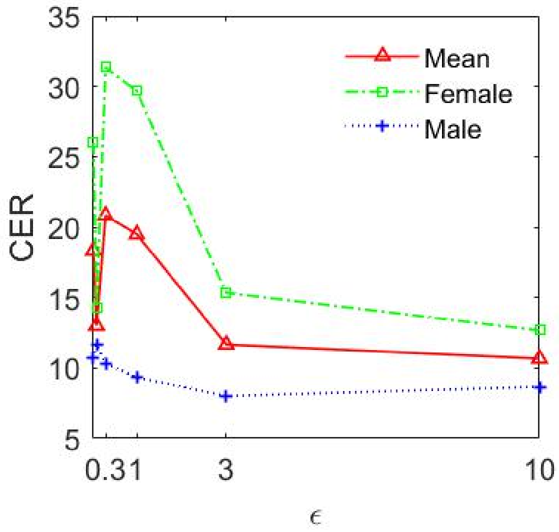
(Objective evaluation.)

Protected speech data with bigger ε -> (1) weaker privacy (2) better utility



MSE vs. ε

(PLDA) ACC vs. ε



CER vs. ε

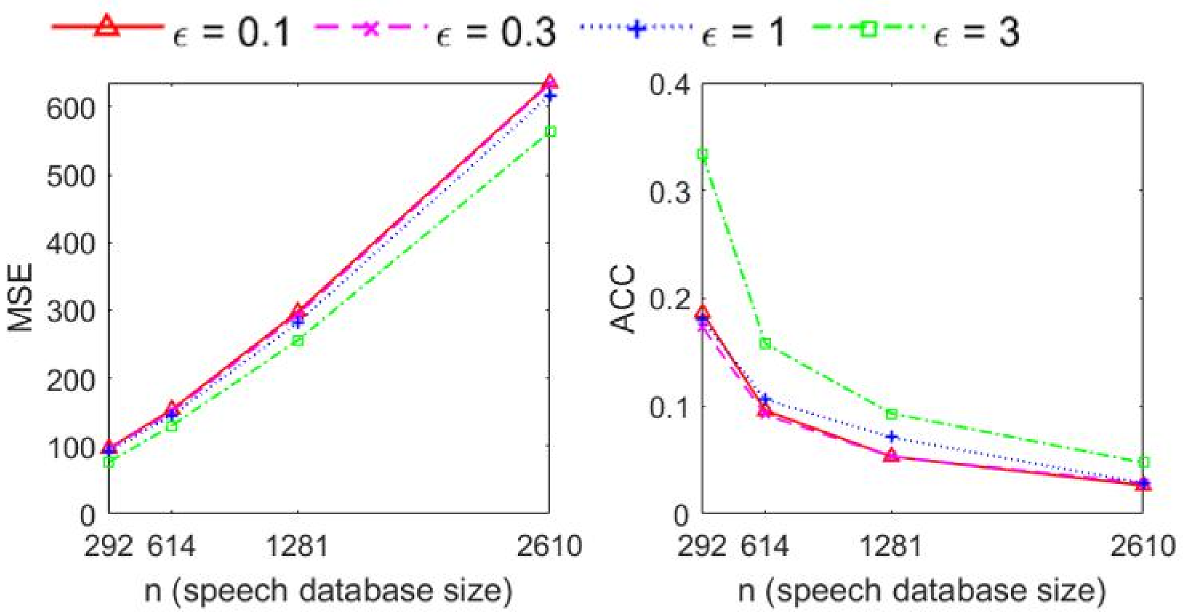
MSE: the difference before and after modification
lower MSE -> weaker privacy
(PLDA) ACC: the accuracy of speaker verification
higher ACC -> weaker privacy

CER: the performance of speech recognition
lower CER -> better utility

Experiment

(Objective evaluation.)

Protected speech data with larger n -> (1) stronger privacy



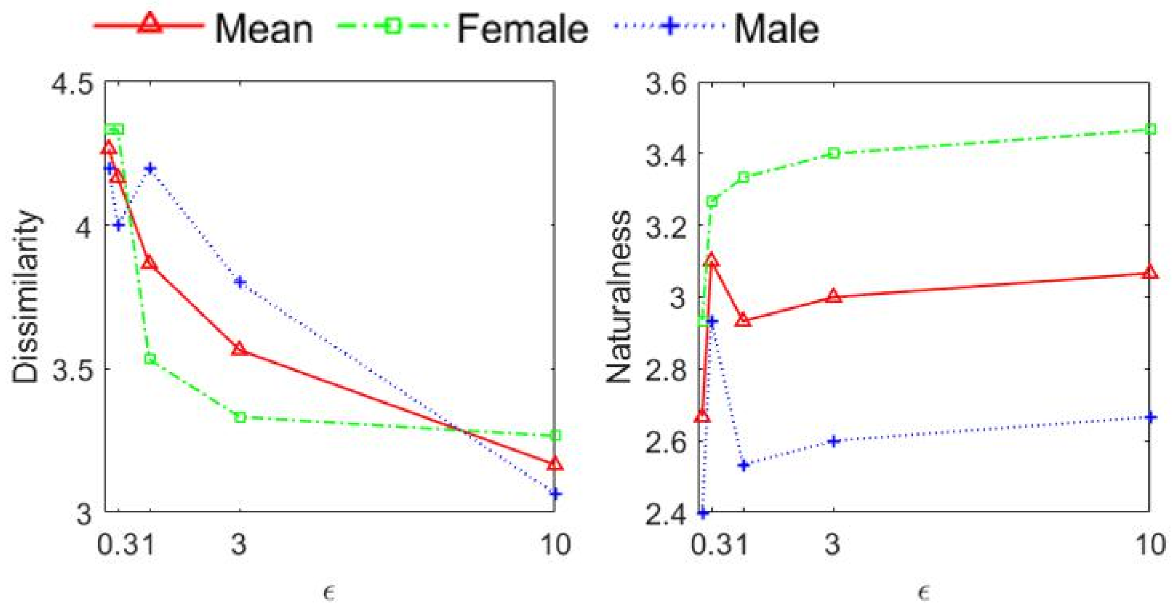
MSE vs. n (PLDA) ACC vs. n

MSE: the difference before and after modification
lower MSE -> weaker privacy
(PLDA) ACC: the accuracy of speaker verification
higher ACC -> weaker privacy

Experiment

(Subjective evaluation.) 15 speakers

Protected speech data with bigger ε -> (1) weaker privacy (2) better utility



Dissimilarity vs. ε

Naturalness vs. ε

Dissimilarity: the voice's differences between and after the modification

lower Dissimilarity -> weaker privacy

Naturalness: the naturalness of sounds that closely resemble the human voice

higher Naturalness -> better utility

Conclusion and Future work

Conclusion:

- Voice-Ind is the first formal privacy notion for voiceprint privacy.
- Our mechanism serves as a primitive to achieve voice-ind.
- Our end-to-end frameworks provide a good privacy-utility trade-off.

Future Works:

- Apply Voice-ind in Virtual Assistant, speech data processing, etc.
- Extend Voice-Ind for speech content privacy.

Thanks