SPSC-2022: 2nd Symposium on Security and Privacy in Speech Communication joined with 2nd VoicePrivacy Challenge

Speaker Anonymization by Pitch Shifting Based on Time-Scale Modification

Candy Olivia Mawalim, Shogo Okada, and Masashi Unoki
Japan Advanced Institute of Science and Technology (JAIST)
September 24, 2022

Outline

1. Introduction
2. Speaker Anonymization
3. Proposed Method
4. Experiments
5. Conclusion and Future Work

1. Introduction

Real-World Problem Examples

https://www.theverge.com/22672123/ai-voice-clone-synthesis-deepfake-applications-vergecast

PODCASTS

EVERYONE WILL BE ABLE TO CLONE THEIR VOICE IN THE FUTURE

AI speech synthesis is quick, easy, and uncannily good

By James Vincent | Sep 14, 2021, 9:04am EDT

Illustration by Alex Castro

NEWS

Home | Coronavirus | Climate | Video | World | Asia | UK | Business | Tech | Science | Stories

Business | Market Data | New Economy | New Tech Economy | Companies | Entrepreneurship | Technology of Business | Economy | Global Car Industry

https://www.bbc.com/news/business-57761873

Voice cloning of growing interest to actors and cybercriminals

By Kitti Palmai
Business reporter

12 July

NATIONAL / MEDIA | MEDIA MIX

APR 20, 2019

NHK docudrama reveals telephone scam tactics

BY PHILIP BRASOR

Thai police last month raided a residence in Pattaya where an alleged telephone swindling operation was taking place. They discovered 15 Japanese nationals suspected of calling retired people in Japan and fooling them into purchasing electronic money. Japanese police say they will arrest the ...

TECH \ ARTIFICIAL INTELLIGENCE

New AI research makes it easier to create fake footage of someone speaking

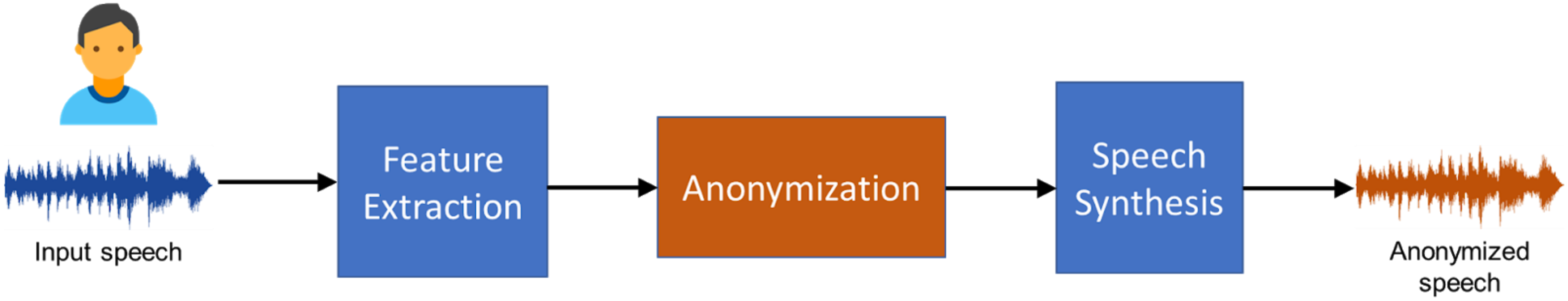
Although the scientists behind it would really rather you didn't

By James Vincent | Jul 12, 2017, 2:21pm EDT

https://www.theverge.com/2017/7/12/15957844/ai-fake-video-audio-speech-obama

2. Speaker Anonymization (Voice Privacy Challenge (VPC))

Privacy Preservation by Speaker Anonymization



Requirements:

1. Speaker identity must be hidden
2. The output anonymized speech should be natural and intelligible
3. The language information should be preserved
4. Following a speaker-to-speaker correspondence (each speaker corresponds to a pseudo-speaker)

2. Speaker Anonymization

VPC Protocols

- **Scenario**
  - Speakers want to hide their identity while allowing any desired goal to be potentially achieved.
  - The attacker has access to a single utterance and wants to identify the corresponding speaker.
- **Attack model** (the attacker has access to various amounts of data):
  - one or more anonymized trial utterances,
  - possibly, several additional utterances for each speaker, which may or may not have been anonymized and are called *enrollment* utterances

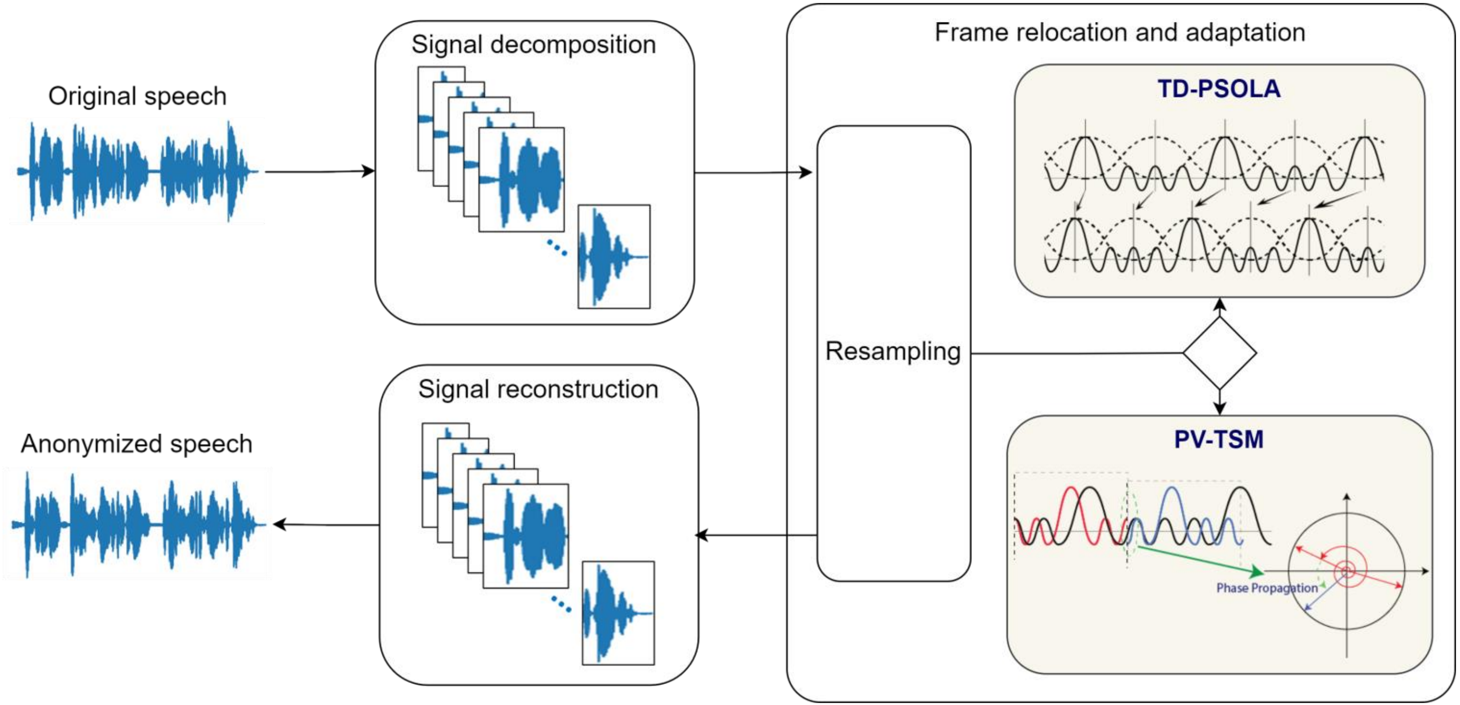
2. Speaker Anonymization

Baseline Systems

- **B1a** : primary baseline with neural source-filter (NSF) model + x-vector
- **B1b :** primary baseline with a unified HiFi-GAN NSF model + x-vector
- **B2a** : speaker anonymization using McAdams coefficients ($\alpha = 0.8$)
- **B2b** : speaker anonymization using McAdams coefficients $\left(\alpha \sim U(0.5, 0.9)\right)$

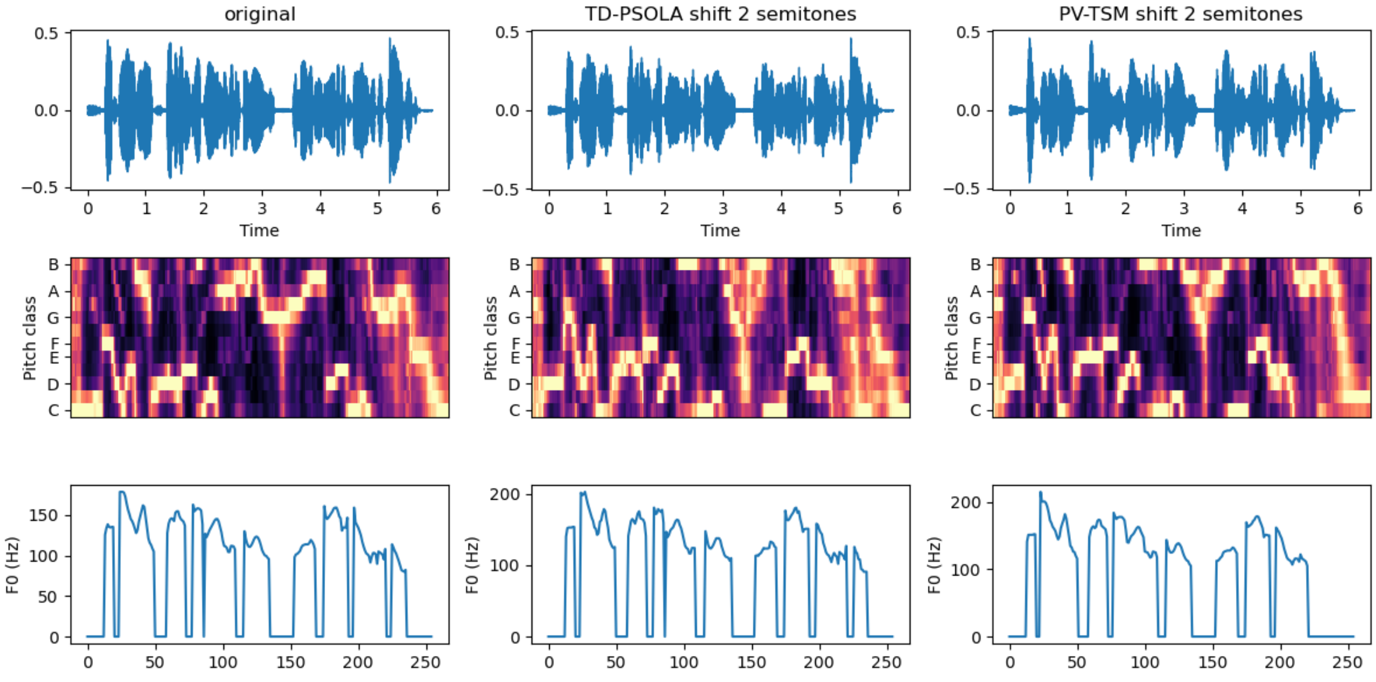
3. Proposed Methods

Time-Scale Modification (TSM) Approach



3. Proposed Methods

Pitch Shifting using TSM



F_0 of original signal

$$F_{0_y}(t) = 2^{n/12} \times F_{0_x}(t)$$

F_0 of anonymized signal

Shift parameter in semitone

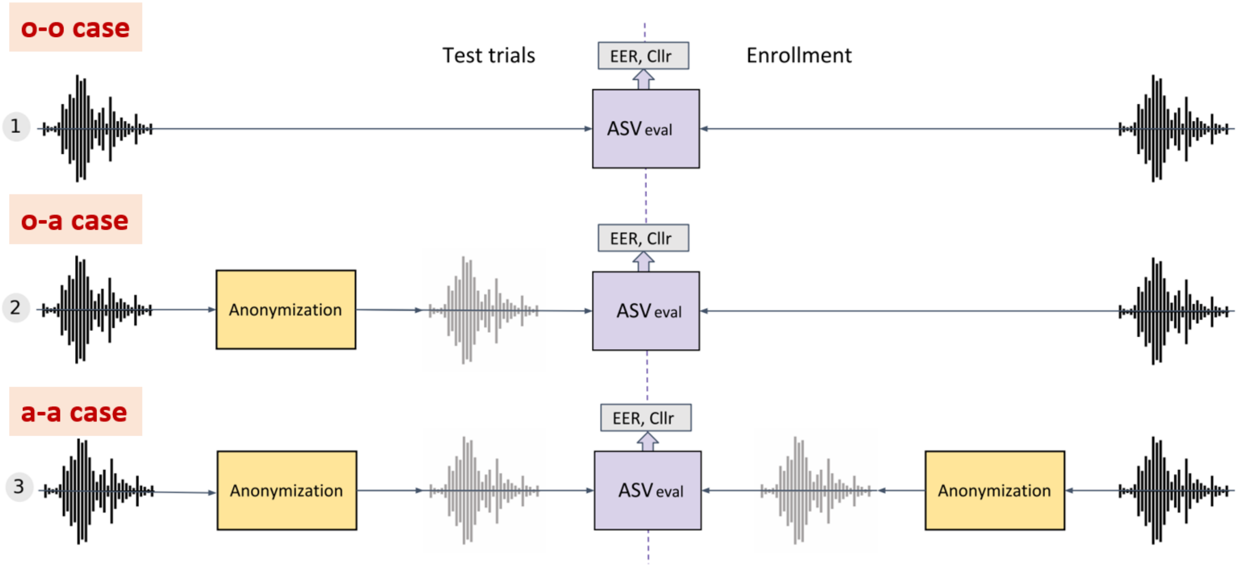
Experimental Setup

- **Dataset for development (Dev) and evaluation (Test):** LibriSpeech and VCTK
- **Evaluation**
  - **Privacy:** using an automatic speaker verification (ASV)
  - **Utility:** using an automatic speech recognition (ASR), pitch correlation, the gain of voice distinctiveness

4. Experiments

ASVeval (VPC 2020 Scenario)

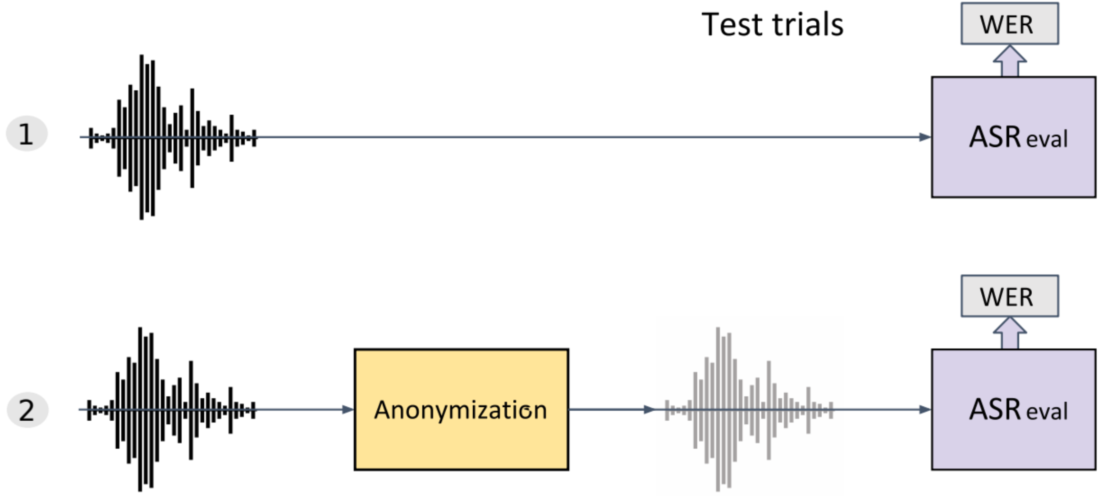
- Speaker verifiability (ASVeval) : **High EER = better privacy**



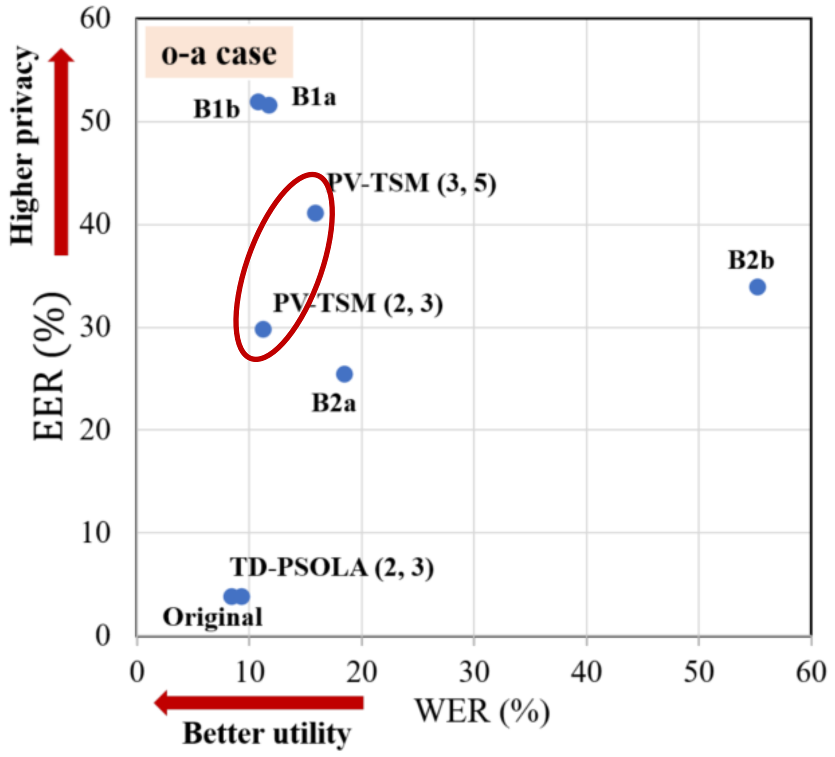
4. Experiments

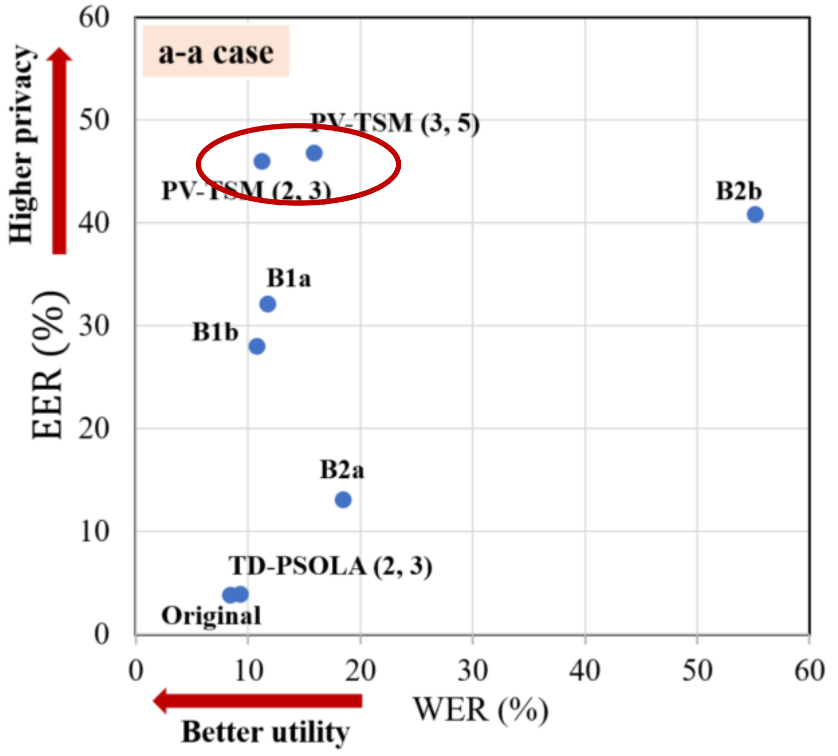
Results (VPC 2020 - ASReval)

- Speech Intelligibility : **Low WER = better utility**



Results (Privacy vs Utility)

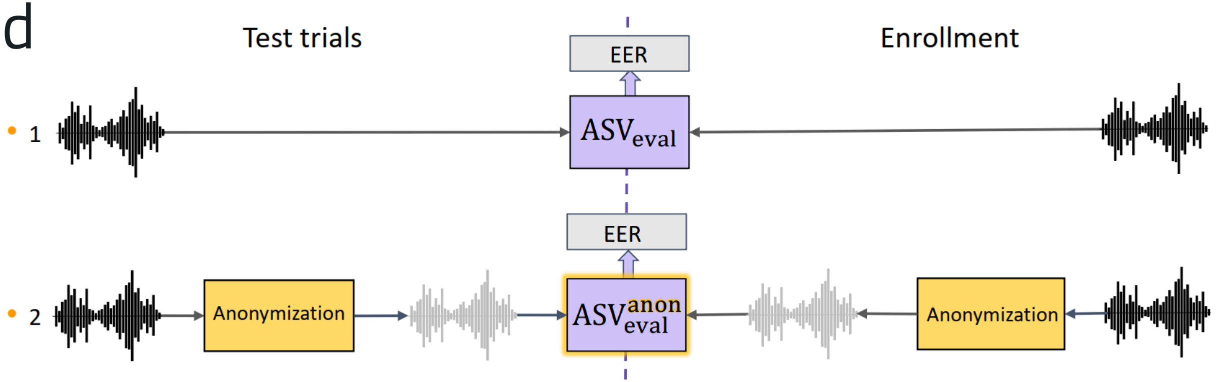




4. Experiments

Results (VPC 2022 – ASVeval)

- Semi-informed attack model

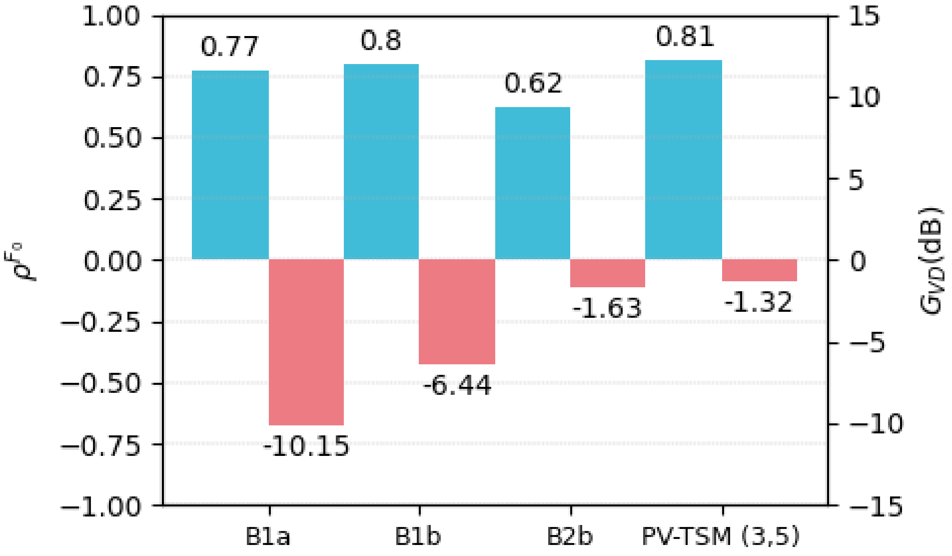


Dataset	Gender	EER (%)				
		Orig	B1a	B1b	B2b	PV-TSM (3,5)
Libri	female	7.66	12.04	9.49	7.12	19.34
	male	1.11	8.91	7.80	1.11	6.46
VCTK (diff)	female	4.94	16.00	10.91	16.92	9.77
	male	2.07	10.05	7.52	7.69	4.99
VCTK (comm)	female	2.89	17.34	15.32	10.98	6.65
	male	1.13	9.89	8.19	4.80	1.41
Weighted average test		**3.80**	**11.81**	**9.18**	**7.77**	**9.81**

4. Experiments

Results (VPC 2022 – secondary utility metrics)

pitch correlation ↑



the gain of voice distinctiveness ↑

5. Conclusion and Future Work

- Two major algorithms of TSM were investigated (TD-PSOLA and PV-TSM) for speaker anonymization based on VPC protocols.
- TD-PSOLA algorithm can be used for pitch shifting but is insufficient for privacy protection in the ASV system.
- In contrast, pitch shifting by the PV-TSM algorithm for speaker anonymization providing the highest balance of privacy-utility metrics (esp., a-a scenario/lazy-informed).
- Method of PV-TSM also can preserve secondary utility metrics.
- In future, we will investigate more the shift parameter and non-linear pitch shifting using the PV-TSM algorithm.

Thank you!