CNT 5410: Computer and Network Security

Final Project Report: Does Speaker Anonymization Really Work?

Nigama Annapurna Dendukuri
(Point of Contact)
ndendukuri@ufl.edu

Sai Ram Varma Budharaju
sbudharaju@ufl.edu

Venkata Sai Karthik Metlapalli
vmetlapalli@ufl.edu

Pavan Siva Sai Savaram
pavansivasavaram@ufl.edu

Kartheek Reddy Gade
gade.k@ufl.edu

March 17, 2024

1 Introduction

Speaker anonymization refers to protecting the identity and privacy of a person when dealing with their voice data. This is achieved by taking out the individual's identity and voiceprint but still maintaining the content of the speech. While there are numerous methods to achieve this, the security and effectiveness of these methods are still to be analyzed deeply.

Speaker anonymization has numerous real-time applications in multiple domains. The privacy of the speaker is crucial, so studying and addressing potential pitfalls in these methods will help in developing more secure techniques and increase trust in digital communications.

Our research involves thoroughly studying speaker anonymization methods, analyzing their strengths, and to address their vulnerabilities. As part of this process, we also studied various Automatic Speaker Verification (ASV) and Automatic Speech Recognition (ASR) models to evaluate the results.

Additionally, our research aims to explore the ethical implications associated with speaker anonymization methods. We will engage in discussions about potential misuse and the ethical responsibilities involved in their development and deployment. This approach aims to ensure a balanced and responsible use of speaker anonymization, respecting individual privacy and societal norms in the digital age.

2 Background & Related Work

2.1 Background

Voice Anonymization or de-identification [8] refers to the process of obscuring the identity of the speaker while preserving the linguistic content, ensuring that the data cannot be associated with a specific individual. [7]. The objective is to make it extremely difficult for malicious actors to link the voice data back to its original speaker, ensuring that even if the voice data is misused, the identity of the speaker remains protected.

In recent times voice anonymization methods have mainly focused on noise addition, voice transformation(VT), voice conversion(VC), voice synthesis(VS), and voice signal processing (SP). There are two

types of speaker anonymization: physical and logical. Physical anonymization adds noise to the original speech signal to make it more difficult to identify the speaker, while logical anonymization techniques modify the identity of speaker in recorded speech signal. Most of the state-of-the-art techniques like x-vector-based neural source-filter (NSF) [6], HiFi-GAN [13] perform logical synthesis-based anonymization on recorded audio. Some of the recent papers focus on performing real-time anonymization and also preserve timbre like VCLOAK [5].

It's crucial to understand the distinction between Automatic Speaker Verification (ASV) and Automatic Speech Recognition (ASR) in the context of voice anonymization. ASV verifies the speaker's identity, while ASR transcribes speech without identifying the speaker. Key metrics, such as EER (Equal Error Rate) for ASV and WER (Word Error Rate) for ASR, are crucial for assessing and improving the efficacy of voice anonymization methods. These metrics help measure the accuracy of verifying speakers and transcribing speech, which is essential for enhancing the performance and security of voice anonymization techniques.

2.2 Related Work

Recently, there have been multiple iterations of the VoicePrivacy Attacker Challenge [24], with a key emphasis on the development of speaker anonymization systems and the creation of attack models to test them. This involves a comprehensive assessment of voice anonymization systems through the release of open-source models and the establishment of appropriate evaluation criteria and datasets.

ASR method involves speech-to-text conversion and synthesizing with a different voice. It's known for its real-time applicability in multiple languages[9]. However, this approach may compromise speech quality and is vulnerable to adversarial attacks, potentially revealing the speaker's identity. To enhance ASR-based anonymization, one can consider employing a robust ASR system, high-quality TTS (Text-to-Speech), and voiceprint transformation systems, though the effectiveness and security of these improvements remain uncertain[22].

Automatic Speaker Verification (ASV) involves identifying a speaker's voiceprint and modifying it through techniques like voiceprint transformation or speech synthesis. This approach utilizes factor analysis to create a compact space covering speaker and communication channel variations, representing speech with i-vectors and eliminating the need for speaker enrollment. Despite its effectiveness, the method faces the possibility of skilled attackers re-identifying the speaker, and while addressing intersession variability with techniques like LDA, NAP, and WCCN, it may not be entirely robust to adversarial attacks, particularly in short audio samples [3].

.

I-vectors, commonly used in speaker recognition, face challenges in anonymization due to adversarial attacks[10]. The proposed method introduces an innovative approach by employing factor analysis, creating a compact space that covers speaker and communication channel variations. This technique eliminates the need for speaker enrollment, utilizing i-vectors for speech representation, and employs Support Vector Machines, cosine distance scoring, and additional techniques like LDA, NAP, and WCCN to enhance robustness. The cosine distance method outperforms traditional approaches, particularly in handling short audio samples, presenting a more effective solution for speaker anonymization [3].

Voice mask is a method designed to anonymize the voice while maintaining intelligibility and safeguarding speakers' identity. It leverages frequency warping method, a technique in vocal tract length normalization (VTLN) originally designed to normalize speaker individualities to increase speech recognition accuracy. This model is used especially in resource limited mobile devices for real time anonymization and is probably the first voice privacy preserving architecture on mobile devices. [19].

The language-agnostic speaker anonymization model utilizes self-supervised learning with a soft content encoder (HuBERT) to extract information, addressing mispronunciation issues. Incorporating an ECAPA-TDNN speaker encoder and an F0 extractor for fundamental frequency details, the model generates anonymized speech using an anonymization vector and HiFi-GAN neural vocoder. While demonstrating effectiveness across English and Mandarin datasets, the model's adaptability to unforeseen acoustic conditions is an ongoing challenge [14].

X-vectors, a more recent development than i-vectors, are considered more robust to noise and more discriminative in distinguishing between speakers. However, X-vector-based speaker anonymization systems face shortcomings, as they are vulnerable to adversarial attacks and can potentially invade privacy. Adversarial attacks could manipulate X-vectors to generate a speech signal that resembles the target speaker's without the original voiceprint, and X-vectors retain significant speaker information, posing a risk of identification even across different languages or accents[2].

McAdams-based speaker anonymization modifies the spectral envelope of the speaker's voice to hinder identification while preserving speech intelligibility and naturalness. This can be achieved through generative or discriminative models, adjusting McAdams coefficients. While relatively simple, effective, and robust to noise, this method is less resistant to adversarial attacks compared to alternatives like i-vector-based speaker anonymization[17].

Speech Sanitizer[20] and V-Cloak[4], are both state-of-the-art in speaker anonymization. They use different techniques to achieve their goals, but achieve high levels of anonymity without sacrificing too much speech quality. V-Cloak is a state-of-the-art real-time voice anonymization system that preserves intelligibility, naturalness, and timbre[4]. It is the first system to achieve both untargeted and targeted anonymization with high performance. Speech Sanitizer[20] uses a combination of speech content desensitization and voice anonymization.

A number of these models have been evaluated on various benchmark datasets like LibriSpeech [16], CommonVoice [1], VoxCeleb [15] that serve as the basis for performance measure across various languages and speaker profiles. Some of these extensively utilize the Kaldi[18] and the SpeechBrain[21] toolkits as foundational libraries to implement and evaluate speaker anonymization techniques.

It is still necessary to address certain issues in this area, as many anonymization techniques are not capable of providing real-time speech anonymization and are restricted to specific environments. To sum up, the level of protection of these methods against various attack vectors has yet to be thoroughly evaluated.

3 Approach: Dataset(s) & Technique(s)

3.1 Literature Survey and Analysis

While there is a lot of research done on speaker anonymization techniques, only a few deep-dive into the security of these methods. Our aim is to conduct a systematic security study and challenge the methods, designing attacks evaluate vulnerabilities, and provide a comprehensive security assessment.

Solving this problem is interesting because current voice anonymization methods may hide the speaker's identity, but they lack robust security. They rely on keeping the anonymization techniques secret - "security by obscurity", which is risky because well-informed attackers can figure them out. Recent studies have indicated that attackers, armed with enough knowledge about the anonymization methods, can

trace back the anonymized speech to the original speaker and conduct linkage attacks. So, it's essential to find more secure ways to protect people's privacy when using voice anonymization [11], [23].

3.2 Threat Models

The effectiveness of an anonymization technique depends on the threat model that is considered. A threat model defines the adversary's knowledge and capabilities. Some common threat models for voice anonymization include:

Ignorant Adversary (A1) The adversary does not know that the audio is anonymized and does not have any prior knowledge of the speaker.

Semi-Informed Adversary (A2) The adversary knows that the audio is anonymized but does not know the specific anonymizer that was used. The adversary may also have access to a limited amount of data about the speaker.

Informed Adversary (A3) The adversary knows the specific anonymizer that was used and has access to a large amount of data about the speaker.

Adversary Information

The adversary information for each technique based on Ignorant (A1), Semi-Informed (A2), and Informed (A3) adversaries is as follows:

NSF: NSF is a spectral modulation-based anonymization technique. It is effective against A1 adversaries, but it is less effective against A2 and A3 adversaries.

HFGAN: HFGAN is a generative adversarial network (GAN)-based anonymization technique. It is more effective against A2 adversaries than NSF, but it is still less effective against A3 adversaries.

McAdams: McAdams is a voice transformation-based anonymization technique. It is effective against A1 and A2 adversaries, but it is less effective against A3 adversaries.

VoiceMask: VoiceMask is a spectral modulation and temporal warping-based anonymization technique. It is more effective against A3 adversaries than NSF, HFGAN, and McAdams.

V-Cloak: V-Cloak is a two-stage anonymization technique that combines spectral modulation, temporal warping, and a GAN. It is the most effective technique against all three types of adversaries.

3.3 Model Selection

Model		B0 (%)	NSF (%)			HFGAN (%)			McAdams (%)			VoiceMask (%)			V-Cloak (%)		
		EER	MMR	WMR	EER	MMR	WMR	EER	MMR	WMR	EER	MMR	WMR	EER	MMR	WMR	EER
ASV	EP	3.72	88.89	3.89	38.09	87.33	3.89	42.21	46.53	3.89	20.69	70.15	3.89	23.40	97.90	3.89	42.21
	XV	5.74	87.33	4.73	34.47	88.89	4.73	39.05	84.28	4.73	40.19	95.73	4.73	37.79	100.0	4.73	44.73
	DP	3.72	93.97	4.05	39.70	89.39	4.05	33.13	80.15	4.05	35.00	99.24	4.05	41.37	99.77	4.05	49.47
	AVG	4.39	90.06	4.22	37.42	88.54	4.22	38.13	70.32	4.22	31.96	88.37	4.22	34.19	99.22	4.22	45.47
	WCS	-	87.33	3.89	34.47	87.33	3.89	33.13	46.53	3.89	20.69	70.15	3.89	23.40	97.90	3.89	42.21

AVG: average, **WCS**: worst-case scenario. **EP**: ECAPA-TDNN, **XV**: X-vector, **DP**: DeepSpeaker.

Figure 1: Performance results

Based on the performance results provided in Figure 1 from the reference [4], we decided to conduct a comprehensive analysis of V-Cloak. This analysis involved executing the code and delving into its design to assess its real-world effectiveness in ensuring a high level of security. The aim is to evaluate how well V-Cloak addresses the challenges associated with voice anonymization and determine if it represents the latest and most advanced approach in this domain.

3.4 Comprehensive Plan

To evaluate voice anonymization modules for anonymity and intelligibility, the proposed five-step approach is as follows:

- **Analysis:** Perform a thorough literature survey of the existing voice anonymization models, their security, and evaluation methods.
- **Replication:** Implement popular state-of-the-art models to understand the methods.
- **Evaluation:** Conduct experiments with different datasets and evaluate the performance of the datasets.
- **Security Analysis:** Design and execute attacks to identify any vulnerabilities in these methods.
- **Assessment:** With the results obtained, evaluate the anonymized audio for anonymity and intelligibility.

3.5 Dataset

The papers that we are analyzing use a diverse set of benchmark datasets like LibriSpeech [16] (English), AISHELL (Chinese), CommonVoice [1] (French) and CommonVoice [1] (Italian), VoxCeleb [15] dataset for conducting extensive experiments.

These datasets serve as the basis for our systematic study, allowing us to evaluate the performance of our speaker anonymization methods across various languages and speaker profiles.

3.6 Libraries & Toolkits

Our research analysis utilizes the Kaldi toolkit [18] and the SpeechBrain toolkit [21] as foundational libraries to implement and evaluate our speaker anonymization techniques. These powerful toolkits provide the necessary infrastructure to assess and improve the privacy and security of voice anonymization techniques.

3.7 Equipment

To evaluate for anonymity and intelligibility, we will need access to computers with GPUs, audio datasets, and audio attack tools. We have utilized resources through HiPerGator by the University of Florida.

3.8 Measurement and Evaluation

Our research aims to assess the security and anonymity features of different models, as well as their effectiveness in maintaining speech clarity following anonymization. Our overall goal is to conduct a thorough evaluation of the security and threat models of these analyzed systems. We will leverage the datasets employed in the development and assessment of speaker anonymization systems, training, development, and evaluation sets.

Furthermore, we intend to gauge the performance of model implementations using diverse techniques in Automatic Speaker Verification (ASV) and Automatic Speech Recognition (ASR) on the aforementioned datasets. Our primary focus will be on objective metrics and subjective evaluations, with a particular emphasis on assessing anonymity and speech clarity [24].

4 Results

4.1 Understanding the concepts

In the initial phase we have established a solid foundation by delving into the core concepts surrounding Automatic Speech Recognition (ASR) and Automatic Speaker Verification (ASV) along with the security concepts integral to this context. Building upon this understanding, we proceeded to gain insights from existing papers on speaker anonymization.

4.2 Insights from existing papers

We tried to understand what these studies aim to protect and the methodologies employed for protection. This served as a groundwork for a comprehensive exploration into the realm of speaker anonymization. The key aspects explored are:

- **Protection Objectives:** We sought to identify the specific aspects these studies have aimed to safeguard, and their methodologies within the realm of speaker anonymization.
- **Methodological Landscape:** A meticulous analysis was conducted to discern the various methodologies utilized to achieve voice anonymization in these studies, providing a comprehensive overview for our subsequent exploration.

4.3 Codebase Exploration of Existing Papers

In our investigation on the efficacy of speaker anonymization, we started by delving into the code bases of existing papers, particularly focusing on resources from IEEE and other reputable journals. Our preliminary work involved exploring various GitHub repositories and reaching out to authors for access to their codebases. Additionally, we examined codes from the Voice Privacy Challenge 2020 and 2022, including baseline codes and those built upon them.

Paper Name	**Repository Link**
Semi-supervised Speaker Anonymization [14]	github.com/nii-yamagishilab/ssl-sas
Voice Privacy Challenge 2020 [25]	github.com/Voice-Privacy-Challenge/Voice-Privacy-Challenge-2020
Evaluating Voice Conversion-based Privacy Protection against Informed Attackers [23]	catalyzex.com/paper/arxiv:1911.03934/code
Speaker Anonymization with Phonetic Intermediate Representations [12]	github.com/DigitalPhonetics/speaker-anonymization
V-Cloak - Advancements in Speaker Anonymization [4]	github.com/V-Cloak/V-Cloak

Table 1: Papers and Corresponding Repositories

By examining the code bases mentioned in Table 1, we gained a comprehensive understanding of the state-of-the-art techniques in speaker anonymization and assess the effectiveness of existing approaches.

4.4 Output Results

In our evaluation of the speaker verification system, we've uncovered some critical insights. Our Equal Error Rate, or EER, stands at 8.65%, reflecting a moderate level of accuracy in balancing false acceptances with false rejections—a fundamental metric for biometric systems.Turning to the Minimum Detection Cost Function, with values of 0.4826 for a target probability of 1% and 0.5412 for 0.1%, we see the system's performance in economic terms—how costly a misclassification is. These figures, while reasonable, signal a space for improvement.The Clustering Identification Rate, or ClIR, with scores of 0.3044 for minimum action and 0.4296 overall, suggests our system's proficiency in clustering and identifying speakers, which is a bespoke metric for this test.Another critical measure, the Receiver Operating Characteristic Curve Equal Error Rate, or ROCCH-EER, is closely aligned with the EER at 8.57%, confirming the system's consistent performance across different evaluation methods.

Lastly, our linkability score of 0.799043 shows a strong capacity for the system to correctly link speech samples to speakers. While this is advantageous for verification, it poses challenges for anonymization, pushing us to strike a delicate balance between verifying identity and protecting it.Together, these metrics guide us towards enhancing the speaker anonymization process, aiming to make it more difficult to verify speakers without losing the clarity and naturalness of their speech.

4.5 Challenges

The main challenges were linked to the technical needs of running/ executing the code for these speech anonymization modules:

- **Hardware Limitations:** The majority of speech anonymization and ASV models required execution on Linux systems with Graphics Processing Units (GPUs). Unfortunately, our available hardware lacked GPUs, posing a significant constraint.
- **Toolkit Dependencies:** Successful execution required the installation of various toolkits like Kaldi. Additionally, essential libraries included complex and big TensorFlow, PyTorch, and PyTorch Audio as prerequisites for running the models.
- **Software Dependencies:** Some models relied on frameworks like Speechbrain, introducing additional software dependencies that required careful handling.
- **Linux Compatibility:** Given the requirement for Linux systems, we faced the challenge of adapting existing hardware setups. Consequently, we were required to dual-boot our systems to enable Linux compatibility.

5 Conclusions

In conclusion,we evaluated the voice anonymization system's effectiveness. The Equal Error Rate, a close EER rate of 8.57%, show moderate performance in speaker verification. Our Minimum Detection Cost Function values, while reasonable, do spotlight areas for refinement to reduce misclassifications without significant cost increases. Important consideration is our linkability score. it demonstrates a strong ability to match speakers to their speech. This strength in speaker verification, however, signals a need for improved anonymization techniques to better protect speaker identity.

To address the identified challenges, future work should focus on creating more user-friendly implementations of speaker anonymization methods, reducing hardware dependencies, and enhancing compatibility with diverse systems. Collaboration between researchers and industry practitioners can facilitate the development of standardized frameworks for evaluating the security and efficacy of speaker anonymization

techniques. It is also crucial to engage in discussions about the responsible development and deployment of these technologies with their, considering potential misuse and establishing ethical guidelines.

References

[1] Rosana Ardila, Megan Branson, Kelly Davis, Michael Henretty, Michael Kohler, Josh Meyer, Reuben Morais, Lindsay Saunders, Francis M. Tyers, and Gregor Weber. Common voice: A massively-multilingual speech corpus. *CoRR*, abs/1912.06670, 2019.

[2] Pierre Champion, Denis Jouvet, and Anthony Larcher. Evaluating x-vector-based speaker anonymization under white-box assessment, 2021.

[3] Najim Dehak, Patrick J. Kenny, Réda Dehak, Pierre Dumouchel, and Pierre Ouellet. Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4):788–798, 2011.

[4] Jiangyi Deng, Fei Teng, Yanjiao Chen, Xiaofu Chen, Zhaohui Wang, and Wenyuan Xu. V-cloak: Intelligibility-, naturalness- & timbre-preserving real-time voice anonymization. *ArXiv*, abs/2210.15140, 2022.

[5] Jiangyi Deng, Fei Teng, Yanjiao Chen, Xiaofu Chen, Zhaohui Wang, and Wenyuan Xu. V-Cloak: Intelligibility-, naturalness- & Timbre-Preserving Real-Time voice anonymization. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 5181–5198, Anaheim, CA, August 2023. USENIX Association.

[6] Fuming Fang, Xin Wang, Junichi Yamagishi, Isao Echizen, Massimiliano Todisco, Nicholas Evans, and Jean-Francois Bonastre. Speaker anonymization using x-vector and neural waveform models, 2019.

[7] Marta Gomez-Barrero, Javier Galbally, Christian Rathgeb, and Christoph Busch. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Transactions on Information Forensics and Security*, 13(6):1406–1420, 2018.

[8] Anil Jain, Lin Hong, and Sharath Pankanti. Biometric identification. *Commun. ACM*, 43(2):90–98, feb 2000.

[9] Md Asif Jalal, Pablo Peso Parada, Jisi Zhang, Karthikeyan Saravanan, Mete Ozay, Myoungji Han, Jung In Lee, and Seokyeong Jung. On-device speaker anonymization of acoustic embeddings for asr based onflexible location gradient reversal layer, 2023.

[10] Carmen Magariños, Paula Lopez-Otero, Laura Docio-Fernandez, Eduardo Rodriguez-Banga, Daniel Erro, and Carmen Garcia-Mateo. Reversible speaker de-identification using pre-trained transformation functions. *Computer Speech Language*, 46:36–52, 2017.

[11] Rebecca T. Mercuri and Peter G. Neumann. Security by obscurity. *Communications of the ACM*, 46(11):160, 2003.

[12] Sarina Meyer, Florian Lux, Pavel Denisov, Julia Koch, Pascal Tilli, and Ngoc Thang Vu. Speaker anonymization with phonetic intermediate representations, 2022.

[13] Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, and Natalia Tomashenko. Language-independent speaker anonymization approach using self-supervised pre-trained models, 2022.

[14] Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, and Natalia Tomashenko. Language-independent speaker anonymization approach using self-supervised pre-trained models. *arXiv preprint arXiv:2202.13097*, 2022.

[15] A. Nagrani, J. S. Chung, and A. Zisserman. Voxceleb: a large-scale speaker identification dataset. In *INTERSPEECH*, 2017.

[16] Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. Librispeech: An asr corpus based on public domain audio books. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5206–5210, 2015.

[17] Jose Patino, Natalia Tomashenko, Massimiliano Todisco, Andreas Nautsch, and Nicholas Evans. Speaker anonymisation using the mcadams coefficient, 2021.

[18] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, and Karel Vesely. The kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, December 2011. IEEE Catalog No.: CFP11SRW-USB.

[19] Jianwei Qian, Haohua Du, Jiahui Hou, Linlin Chen, Taeho Jung, and Xiang-Yang Li. Hidebehind: Enjoy voice input with voiceprint unclonability and anonymity. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, SenSys '18, page 82–94, New York, NY, USA, 2018. Association for Computing Machinery.

[20] Jianwei Qian, Haohua Du, Jiahui Hou, Linlin Chen, Taeho Jung, and Xiangyang Li. Speech sanitizer: Speech content desensitization and voice anonymization. *IEEE Transactions on Dependable and Secure Computing*, PP:1–1, 12 2019.

[21] Mirco Ravanelli, Titouan Parcollet, Peter Plantinga, Aku Rouhe, Samuele Cornell, Loren Lugosch, Cem Subakan, Nauman Dawalatabad, Abdelwahab Heba, Jianyuan Zhong, Ju-Chieh Chou, Sung-Lin Yeh, Szu-Wei Fu, Chien-Feng Liao, Elena Rastorgueva, François Grondin, William Aris, Hwidong Na, Yan Gao, Renato De Mori, and Yoshua Bengio. SpeechBrain: A general-purpose speech toolkit, 2021. arXiv:2106.04624.

[22] Yi Ren, Chenxu Hu, Xu Tan, Tao Qin, Sheng Zhao, Zhou Zhao, and Tie-Yan Liu. Fastspeech 2: Fast and high-quality end-to-end text to speech, 2022.

[23] Brij Mohan Lal Srivastava, Nathalie Vauquier, Md Sahidullah, Aurélien Bellet, Marc Tommasi, and Emmanuel Vincent. Evaluating voice conversion-based privacy protection against informed attackers, 2020.

[24] Natalia Tomashenko, Brij Srivastava, Xin Wang, Emmanuel Vincent, Andreas Nautsch, Junichi Yamagishi, Nicholas Evans, Jose Patino, Jean-François Bonastre, Paul-Gauthier Noé, and Massimiliano Todisco. Introducing the voiceprivacy initiative. 11 2020.

[25] Natalia Tomashenko, Brij Srivastava, Xin Wang, Emmanuel Vincent, Andreas Nautsch, Junichi Yamagishi, Nicholas Evans, Jose Patino, Jean-François Bonastre, Paul-Gauthier Noé, and Massimiliano Todisco. Introducing the voiceprivacy initiative. 11 2020.