

# Automatic Audio Privacy Protection for Voice Control System

Ke Sun

University of California, San Diego  
kesun@ucsd.edu


Chen Chen

University of California, San Diego  
chenchen@ucsd.edu

PI: Xinyu Zhang

University of California, San Diego  
xyzhang@ucsd.edu

Voice-user Interface (VUI) makes spoken-user interactions with microphone enabled computing devices possible. This is achieved by the automatic speech recognition (ASR) and voice controlled systems (VCS), *e.g.* Amazon Alexa, to understand spoken commands and answer questions. Such computing devices with VCS and VUI are highly pervasive, and existing as the form of smart speakers, smartphones and heterogeneous IoT/mobile devices. Typically, these devices have the **always-on recording mechanism** [6] and keep listening to the end-users, in order to respond to trigger commands, *e.g.* “Alexa!”, and “OK Google”. Service providers will store such unintended recordings in the cloud with information containing sufficient personally identifiable information (PII) that might be shared with other companies [2]. Beside this, the private speech and PII might also be leaked in the background noise while recording intended voice commands. In both cases, users have no control on the their unintended recorded audio once being sent to remote cloud. Although there are well established regulations for protecting privacy, accidents are still hard to prevent due to unexpected causes, *e.g.* the Amazon Alexa went rogue in 2018 where a women reported the private conversation between her families was leaked to others without consent [4]. Thus, our project will answer the question: *could VCS only record the information they need other than any private sound?*

The brute-force way to protect such speech privacy is to disable the VCS listening by hitting the physical mute button  or sending a mute voice command [9]. However, users have to determine when to do this by themselves which is cumbersome. Further, once sound recordings disabled, VCS will no longer be woken up by trigger commands unless the system is unmuted manually. Alternative anti-eavesdropping approach, *e.g.* [13] used sound masking to achieve speech privacy protection. However, the masking sound will generate audible noise that will disturb dweller’s daily life. To overcome this, [3, 10] shows that using inaudible sound is also able to jam the microphone. However, their approach would disturb the VCS basic functionalities. Similar works, *e.g.* [8, 11, 14], mainly focus on the inaudible voice command attack and/or defense while did not address the implications on protecting VCS speech privacy.

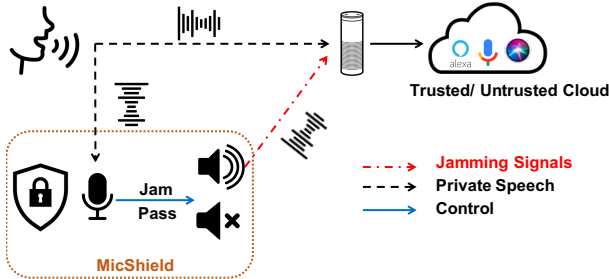
We propose MicShield, which explores the feasibility of automatic speech privacy protection against always-on microphone without disturbing either everyday life or basic system functionalities. Specifically, MicShield considers the scenario where adversaries utilize the always-on microphone to eavesdrop user private speech. To establish threat model, we assume the adversaries: 1) have the access to the raw unprocessed speech signals of the always-on microphones; 2) can use any kinds of software algorithm to enhance the sound quality; 3) can use arbitrary ASR algorithms to infer the semantic content. MicShield’s protection goal is to prevent adversaries using the always-on microphone to recognize the unintended speech while the always-on microphone can still capture the

trigger command and used to wake up the VCS. Finally, we anticipate the MicShield as a physical add-on intelligent shield without adding additional complexities to the hardware and software stacks of existing commercial available VCS.

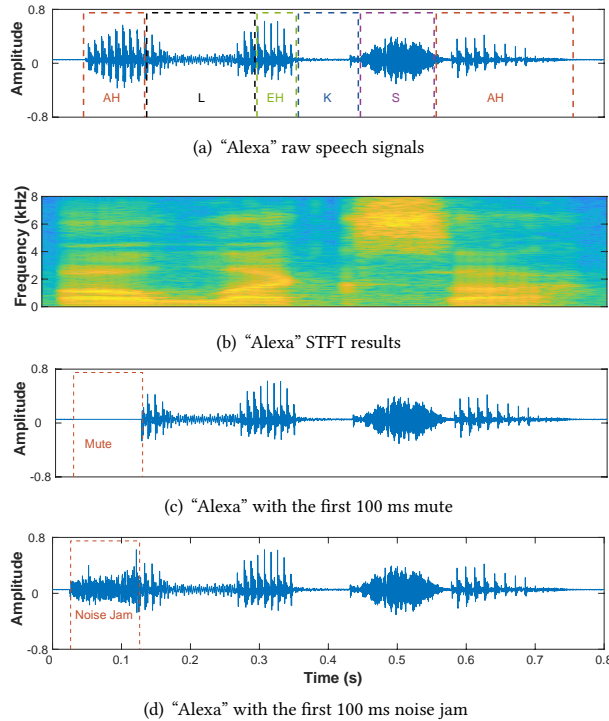
To achieve the goal, we propose three challenges. *First*, our design needs to protect the speech privacy while not disturbing either everyday life or VCS basic functionalities. While transmitting continuous inaudible sound, the jamming device will also squash the trigger command, and thus the system will not respond by trigger commands unless turning of the jamming devices. To resolve this, MicShield will dynamically control the jamming volume based on framewise wake-up word likelihood[5] that can be computed by a novel pipeline for extracting the intended speech signal from mixture of jamming and speech signal with the idea of self-cancellations. Once the wake-up word likelihood is larger than predefined threshold, the jamming will be turned off to ensure the VCS can identify the wake-up commands. Otherwise, MicShield will keep jam the speech based on pre-defined frame volume. To prevent disturbing daily life, we propose to use inaudible sound that is transmitted by ultrasonic speaker, but can be captured by regular microphones [10]. *Second*, we envision our approach being capable of preventing the private speech and PII being leaked in the background sound while recording voice commands. This motivate us to devise an additional jamming mechanism to shield the background sound during the command recording phase while the cloud side ASR is still able to extract the semantic content. *Third*, our approach needs to address the fact of multi-microphones/microphone array design in a majority of commercial available IoT/mobile devices with VCS, *e.g.* Amazon Echo, which are used for sound localizations by using beamforming and signal-noise separations [7]. A well designed MicShield should not allow the unintended speech being leaked through any microphone in the microphone array. We proposes to use a single sound source to protect the speech privacy for microphone array by leveraging a novel, compact and passive sound splitter to generate the fake environmental noise [12].

Our preliminary study used the Amazon Echo Dot [1] as the starter tool. *First*, we have designed a rapid proof-of-concept system (see Figure 1) and proven that with the beginning tens of millisecond jammed, the wake-up words can still trigger the commercial VCS (see Figure 3 and Figure 2). *Second*, we tested the feasibility that the speech signal can be successfully jammed by the inaudible noise from a distance of 30 cm by using only one ultrasonic transducer. In the long term, we see the potentialities of MicShield as an effective implication being designed as an additional compact intelligent shield, or incorporated as part of motherboard for heterogeneous mobile devices, to prevent the leak of unintentional private speech to the third parties and enable a more practical and trustable privacy control paradigm.

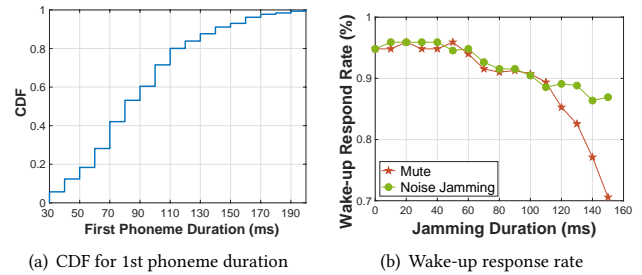
**APPENDIX: CONCEPTUAL DIAGRAMS AND FIGURES**



**Figure 1: Provisional system architecture.** We designed the MicShield as a compact add-on to prevent unintended speech being recorded by the microphone(s) of commercial available IoT/mobile devices with VCS, e.g. Amazon Echo.



**Figure 2: Example speech signal of the trigger commands - "Alexa!".** In this example, the Amazon Echo Dot can still be triggered even though the first 100ms data is jammed (c.f. figure 2(c) and 2(d)).



**Figure 3: With out collected dataset, we showed that the private audio data, which may contains sufficient PII, can be leaked to the remote trusted/untrusted cloud, e.g. Amazon Alexa Voice Service.**

**REFERENCES**

- [1] [n. d.]. Amazon.com: Echo Dot (3rd Gen) - Smart speaker with Alexa - Charcoal: Amazon Devices. <https://www.amazon.com/Echo-Dot/dp/B07FZ8S74R>. (Accessed on 12/05/2019).
- [2] 2019. Hey, Alexa: Stop recording me - The Washington Post. <https://www.washingtonpost.com/technology/2019/05/06/alexa-has-been-eavesdropping-you-this-whole-time/>. (Accessed on 12/05/2019).
- [3] Yuxin Chen, Huiying Li, Steven Nagels, Zhijing Li, Pedro Lopes, Ben Y Zhao, and Haitao Zheng. 2019. Understanding the Effectiveness of Ultrasonic Microphone Jammer. *arXiv preprint arXiv:1904.08490*.
- [4] Niraj Chokshi. 2018. Is Alexa Listening? Amazon Echo Sent Out Recording of Couple's Conversation - The New York Times. <https://www.nytimes.com/2018/05/25/business/amazon-alexa-conversation-shared-echo.html>. (Accessed on 12/04/2019).
- [5] Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM networks. In *Proceedings of IEEE International Joint Conference on Neural Networks*. IEEE.
- [6] Stacy Gray. 2016. Always on: privacy implications of microphone-enabled devices. In *Future of privacy forum*.
- [7] François Grondin and François Michaud. 2019. Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations. *Robotics and Autonomous Systems* (2019).
- [8] Yitao He, Junyu Bian, Xinyu Tong, Zihui Qian, Wei Zhu, Xiaohua Tian, and Xingbin Wang. 2019. Canceling Inaudible Voice Commands Against Voice Control Systems. In *Proceedings of ACM MobiCom*. ACM.
- [9] Heather Kelly. 2018. How to make sure your Amazon Echo doesn't send secret recordings. <https://money.cnn.com/2018/05/25/technology/amazon-alexa-stop-recording/index.html>. (Accessed on 12/04/2019).
- [10] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. 2017. Backdoor: Making microphones hear inaudible sounds. In *Proceedings of ACM MobiSys*. ACM.
- [11] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. 2018. Inaudible voice commands: The long-range attack and defense. In *Proceedings of Usenix NSDI*.
- [12] H Tijdeman. 1975. On the propagation of sound waves in cylindrical tubes. *Journal of Sound and Vibration* (1975).
- [13] Yu-Chih Tung and Kang G Shin. 2019. Exploiting Sound Masking for Audio Privacy in Smartphones. In *Proceedings of ACM AsiaCCS*.
- [14] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. Dolphinattack: Inaudible voice commands. In *Proceedings of ACM CCS*. ACM.