TagMic: Listening Through RFID Signals

Yin Li, Chunhui Duan, Xuan Ding, Cihang Liu
Tsinghua University
liyin16@tsinghua.org.cn, {duanch, dingxuan}@tsinghua.edu.cn, cihang@greenorbs.org

Abstract—**RFID is an increasingly ubiquitous technology widely adopted in both the industry and our daily life nowadays. But when it comes to eavesdropping, people usually pay attention to devices like cameras and mobile phones, instead of small-volume and battery-free RFID tags. This work shows the possibility of using prevalence RFIDs to capture and recognize the acoustic signals. To be specific, we attach an RFID tag on an object, which is located in the vicinity of the sound source. Our key innovation lies in the translation between the vibrations induced when the sound wave hits the object surface and the fluctuations in the tag's RF signals. Although the inherent sampling rate of commercial RFID devices is deficient, and the vibrations are very subtle, we still extract characteristic features from imperfect measurements by taking advantage of state-of-the-art machine learning and signal processing algorithms. We have implemented our system with commercial RFID and loudspeaker equipment and evaluated it intensively in our lab environment. Experimental results show that the average success rate in detecting single tone sounds can reach as high as 93.10%. We believe our work would raise the attention of RFID in the concern of surveillance and security.**

Keywords-**TagMic; speech recognition; RFID; privacy**

I. INTRODUCTION

Sound wave transfers energy by producing vibration in the air and it causes minimal vibrations on the nearby object's surface. Although these motions are too subtle to be visible, it indeed reflects the pattern of the sound source. Once exploited by a malicious attacker, a threat of acoustic eavesdropping could be created.

Many attempts have been made by leveraging active methods to capture the vibration. Some researchers use specialized sensors like accelerators [1] and MEMS gyroscopes [2]. Projecting laser beam on the object's surface can capture the minor vibration but requires a strict visible environment (line-of-sight). High-speed cameras can record the vibrations [3] but it is not practical due to the high cost.

In this work, we propose an RFID-based(Radio Frequency Identification) mechanism, TagMic, to capture and recognize the acoustic signals. If the ubiquitous RFID technology everywhere in daily life can be utilized for eavesdropping, it would raise a big concern of security issues.

We attach a passive RFID tag on an object (*i.e.*, a chips bag) which is located in the vicinity of the sound source. TagMic then attempts to recover the original sound by discerning and recognize the unique pattern in the collected RF signals.

Putting this idea into practice, the challenges we face are mainly the following three folds.

- **Deficient sampling rate:** Commercial-Off-The-Shelf (COTS) RFID reader can only interrogate tags approximately 40 times per second. As typical human speech contains frequencies more than a few hundred hertz. According to the sampling theorem, such a low sampling rate of RFID is insufficient for speech detection applications. We need the aliasing feature to map the patterns of high frequencies to the lower band. Recent work [4], [5] get advances in making sense of high-frequency vibrations, but can only work for periodic signals.
- **Subtle vibrations:** The tiny air pressure vibrations only cause indistinguishable fluctuations in backscatter signals. Besides, the phase is a sensitive metric that contains imperfect measurements caused by surrounding noise in a complex indoor environment.
- **Limited training data:** The RF signals are similar to sound vibrations but far more inefficient than real sound waves in speech recognition tasks. However, we could still leverage the transfer-learning model pre-trained on a large open-source audio dataset to train our RF signal model.

Summary. To combat the above challenges, this paper makes the following contributions:

- As far as we know, we are the first to show the possibility of leveraging ubiquitous small-volume and battery-free RFID tags to capture the sound wave for speech recognition purposes.
- We extract characteristic features from the tag's phase measurements using various signal processing methods, which can overcome the negative impacts of ambient noise as well as the deficient sampling rate.
- With the extracted features, we propose a speech recognition approach which achieves high classification accuracy by incorporating state-of-the-art transfer-learning-based pre-trained model in the machine learning field.
- We implement a prototype system for TagMic with COTS RFID and loudspeaker devices. Experimental results show that TagMic achieves an average success rate of 93.10% in detecting single tone sounds and the feasibility for word detection, which is meaningful for many practical applications.

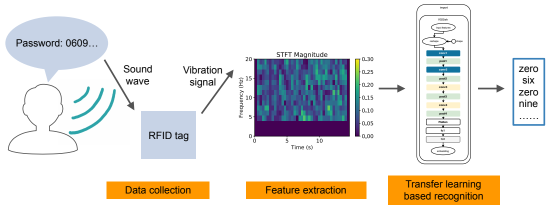


Figure 1. **System workflow**

II. SYSTEM DESIGN

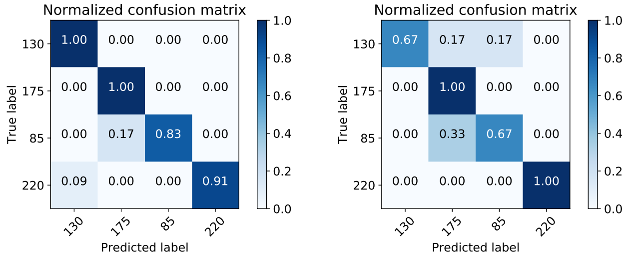
To implement our idea, we aim to handle the challenges lying in existing commercial RFID systems without any hardware modification. The system workflow is portrayed in Fig. 1. TagMic captures the sound wave by an RFID tag attached to the surface of an object, for example, a plastic bag. Then we adopt signal processing algorithms, including MFCC and STFT, to extract the patterns in backscatter signals. These algorithms and some essential intuitive features of phase and Doppler shift signals could help us handle the challenges we mentioned before. We then apply a VGGish transfer-learning model pre-trained on Google AudioSet, a large open-source audio dataset, as a feature extractor. At last, we test TagMic for multiple classification tasks from easy ones, such as single tone classification and gender/speaker identification, to speech recognition which could be used for eavesdropping private information such as bank account passwords.

III. PRELIMINARY EVALUATION

We implemented our system using a commercial UHF RFID, an ImpinJ Speedway R420 compatible with EPC Gen2 standard. 3D Hang Tag&General Tag manufactured by Aluminum etching are employed, which are widely used on files, suitcases, clothing retail, etc. We adopt Low Level Reader Protocol (LLRP) to communicate with the reader and run the TagSee system [6] on a laptop to collect data. We use a loudspeaker to play tones and records of isolated digits, i.e., 11 words recorded by 10 speakers(5 female and 5 male) twice.

Firstly, we test TagMic's ability to classify single tone sounds. They are 60-second sine sweep cut into one-second segments. We choose 85Hz, 130Hz, 175Hz and 220Hz to cover the range of male(85Hz∼180Hz) and female(166Hz∼255Hz) speech. The proportion of the dataset to include in testing set split is set from 10 to 30 percent. The best overall accuracy reaches 93.10%. And the cross-validation accuracy is 87.50%. The confusion matrix of the classification result in Fig. 2 shows that the accuracy on each single tone group ranges from 83% to even 100%.

As for the word recognition task, it gives a 56% overall accuracy in user-independent experiments using simple DTW as the classifier. It is a fair accuracy for an 11-word classification problem and shows the feasibility of eavesdropping speech conversation. We also test gender identification and reach around 76% accuracy. Although the experiment is only tested on a limited dictionary with 11 words, it is enough to show the feasibility of speech recognition on a larger dictionary. This extension relies on state-of-art speech recognition algorithms dealing with sentence structure and grammar, which is not the focus of this work.



(a) Best accuracy

(b) Accuracy of cross-validation

Figure 2. **Accuracy of single tone classification.** (a) Best accuracy reaches 93.10%. (b) Accuracy of cross-validation is about 87.50%.

IV. ACKNOWLEDGMENT

This research is supported in part by the National Key R&D Program of China (Grants No. 2018YFB1308601), and the National Natural Science Foundation of China (Grants No. 61902212 and No. 61772248).

REFERENCES

[1] L. Zhang, P. H. Pathak, M. Wu, Y. Zhao, and P. Mohapatra, "Accelword: Energy efficient hotword detection through accelerometer," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2015, pp. 301–315.

[2] Y. Michalevsky, D. Boneh, and G. Nakibly, "Gyrophone: Recognizing speech from gyroscope signals," in *USENIX Security Symposium*, 2014, pp. 1053–1067.

[3] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman, "The visual microphone: Passive recovery of sound from video," 2014.

[4] L. Yang, Y. Li, Q. Lin, X.-Y. Li, and Y. Liu, "Making sense of mechanical vibration period with sub-millisecond accuracy using backscatter signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 2016, pp. 16–28.

[5] C. Duan, L. Yang, Q. Lin, Y. Liu, and L. Xie, "Robust spinning sensing with dual-rfid-tags in noisy settings," *IEEE Transactions on Mobile Computing*, 2018.

[6] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices," in *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 2014, pp. 237–248.