社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

Auto Encoder を用いた音声認識システムへの攻撃検知手法の検討

松倉悠記† 樽谷優弥†† 福島行信††† 横平徳美††

† 岡山大学大学院ヘルスシステム統合科学研究科　〒700-8530 岡山市北区津島中 3 丁目 1 番 1 号
†† 岡山大学学術研究院ヘルスシステム統合科学学域　〒700-8530 岡山市北区津島中 3 丁目 1 番 1 号
††† 岡山大学学術研究院自然科学学域　〒700-8530 岡山市北区津島中 1 丁目 1 番 1 号
E-mail: †matsukura00net@s.okayama-u.ac.jp, ††{y-tarutn,yokohira}@okayama-u.ac.jp,
†††fukusima@okayama-u.ac.jp

あらまし　深層学習技術が音声認識に用いられるようになり、その認識精度が向上している。それに伴い、スマートスピーカや音声アシスタントをはじめとした音声認識システムの需要が高まっている。一方で、音声認識システムはセキュリティ面で脆弱であることが懸念されている。特に音声に微小なノイズを加えることで音声認識システムに誤認識を引き起こさせる Audio Adversarial Example が問題視されている。本稿では、Audio Adversarial Example による音声認識システムへの攻撃を検出する方法について検討する。提案手法では音声波形を画像として処理を行い、異常検知手法に用いられる Auto Encoder による異常検知について検討をする。検証の結果、提案手法によるって音声波形の復元を行うことはできたが、異常検知による識別が難しいことを明らかにした。

キーワード　スマートスピーカ；音声認識；セキュリティ；Audio Adversarial Example;

An attack detection method on speech recognition systems using auto encoder

Yuki MATSUKURA†, Yuya TARUTANI††, Yukinobu FUKUSHIMA†††, and Tokumi YOKOHIRA††

† Graduate School of Integrated Health Systems Science, Okayama University　3–1–1 Tsushimanaka,Kita-ku, Okayama-shi, Okayama, 700–8530 Japan
†† Interdisciplinary Science and Engineering in Health Systems, Institute of Academic and Research, Okayama University　3–1–1 Tsushimanaka,Kitaku, Okayama-shi, Okayama, 700–8530 Japan
††† Natural Science and Technology, Institute of Academic and Research, Okayama University　1–1–1 Tsushimanaka,Kita-ku, Okayama-shi, Okayama, 700–8530 Japan
E-mail: †matsukura00net@s.okayama-u.ac.jp, ††{y-tarutn,yokohira}@okayama-u.ac.jp,
†††fukusima@okayama-u.ac.jp

Abstract　Deep learning technology is now being used for speech recognition, and its recognition accuracy is improving. As a result, the demand for voice recognition systems, including smart speakers and voice assistants, is increasing. On the other hand, there is a concern that voice recognition systems are vulnerable in terms of security. In particular, the Audio Adversarial Example, which causes false recognition by speech recognition systems by adding small noises to speech, is considered to be a problem. In this paper, we investigate a method to detect attacks on speech recognition systems by Audio Adversarial Examples. In the proposed method, audio waveforms are processed as images, and anomaly detection by Auto Encoder, which is used in anomaly detection methods, is examined. As a result of verification, we found that the proposed method can recover the audio waveform, but it is difficult to identify the anomaly by anomaly detection.

Key words　Smart Speaker; Speech Recognition; Security; Audio Adversarial Example