Attacks on Neural Networks in a Lightweight Speech Anonymization Pipeline

Daan Brugmans
Radboud University
daan.brugmans@ru.nl

Abstract

1 Introduction

As advancements in the field of Automatic Speech Recognition (ASR) have accelerated with the rise of modern End-to-End neural networks, the risks associated with using such models in ASR applications has become more evident.

Modern neural ASR models are capable of parsing and producing speech to a new level of authenticity: transformers are State-of-the-Art for ASR and word recognition, and the introduction of modern unsupervised deep neural network architectures, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), has allowed for more realistic, accurate, and easier generation of speech. These modern speech generation models are capable of learning to reproduce a person's voice, and then generating new speech using the learned voice. Such synthesized speeches are called *deepfaked* speeches, or simply *deepfakes*.

The presence and influence of deepfakes has become increasingly apparent in recent years: neurally synthesized audio and video of important persons are used to spread misinformation and manipulate. One way to counteract the repercussions of deepfakes is the removal of the personalization in the learned, and thus reproduced, speech. This is called *Speaker Anonymization*. In Speaker Anonymization, we aim to maintain the ASR quality of the audio, while applying changes that make the audio untraceable to a person's likeness.

Although modern Speaker Anonymization systems, often neural in nature, have been shown to be able to anonymize speech while maintaining ASR quality, they can also be manipulated. By attacking neural Speaker Anonymization systems, we may be able to circumvent the preventative measures they provide, and generate speech to a person's likeness regardless of their presence. This paper will focus in that topic: attacking neural Speaker Anonymization systems.

2 Related Work

Meyer et al. (2023)

3 Method

4 Experiment

5 Results

6 Discussion

7 Conclusion

References

Sarina Meyer, Pascal Tilli, Pavel Denisov, Florian Lux, Julia Koch, and Ngoc Thang Vu. 2023. Anonymizing speech with generative adversarial networks to preserve speaker privacy. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, pages 912–919.