

Adversarial Anonymizing Pipeline for Batches of Speech Audio Using Automatic Speech Recognition and Speaker Recognition Models

Julian Roddeman - s1080491

June 2022

Abstract

This thesis explores a novel approach to audio anonymization by manipulating speech audio files to obscure speaker identity without significantly impacting speech intelligibility. This is achieved using a series of audio processing effects, tested against automatic speech recognition (ASR) and speaker recognition models, aiming to find the optimal balance between anonymization (loss of speaker identity) and intelligibility (retention of speech). The results show that certain combinations of audio effects can be used to effectively anonymize speech audio.

1 Introduction

Audio data anonymization has become crucial for maintaining user privacy in voice-based systems. This thesis introduces an innovative audio processing pipeline designed to alter identifiable features within speech recordings while retaining the necessary attributes for comprehensible speech. The research question focuses on how effectively different audio effects can anonymize voice data while maintaining its utility for tasks such as speech recognition.

2 Related Work

[A comprehensive review of existing literature on audio anonymization, speech synthesis, and privacy-preserving techniques in speech technologies.]

3 Method

The method section details the experimental setup, including the use of the pedalboard library to apply audio effects for modifying voice recordings. Parameters tested include distortion, pitch shifting, highpass and lowpass filtering,

bit crushing, chorus, and phaser effects. The aim is to determine the best parameter combinations that maximize anonymization while minimizing speech intelligibility loss.

Effect	Parameter Range
Distortion	0 dB to 50 dB
Pitch Shift	-10 to 0 semitones
Highpass Filter	0 Hz to 200 Hz
Lowpass Filter	2800 Hz to 4200 Hz
Bitcrush	0 to 12 bit depth
Chorus	12.5 Hz to 37.5 Hz
Phaser	12.5 Hz to 37.5 Hz
Time Stretch	0.8 to 1.2 stretch factor

Table 1: Ranges of parameters used for audio effect manipulation

4 Experiment

The experiment involved processing batches of speech audio through a series of effects configured according to parameters optimized via an Optuna framework. The setup is visualized in the included diagram (Figure ??). The effectiveness of each parameter set was evaluated based on its impact on automated speech recognition (ASR) and speaker verification accuracy.

Figure 1: Flowchart of the audio anonymization pipeline

5 Results

6 Discussion

7 Conclusion

8 Reference List

9 Appendices