Lab 1 - VocalShield Description

Joshua L. Dowell

CS 411

Professor T. Kennedy

4 September 2024

Version 1

Table of Contents

List of Figures

List of Tables

1. Introduction

With the rising growth of voice cloning technology there has been a growing threat to content creators' digital content. This poses a problem when it comes to determining the authenticity and security of online digital content. Malicious use of high-fidelity voice replicas can lead to breaches of privacy, deception of audiences, and manipulation of digital content, compromising trust in voice-based digital media.

1.2 Current Voice Protection Methods

There are currently three types of voice protection methods focusing on detection rather than prevention. The method known as "Voice Watermarking" which embeds inaudible markers in audio to detect unauthorized use. This approach is limited in preventing cloning and can be bypassed with advanced techniques. The second method is "Biometric Voice Authentication", this technique utilizes unique vocal characteristics for identification, but it is vulnerable to deepfake attacks that mimic those characteristics with high accuracy. The final method is "Anti-Spoofing Algorithms", which detect synthetic voices but often rely on machine learning models trained on limited datasets, making them less effective against sophisticated novel attacks.

1.3 Solution Overview

VocalShield aims to fill this gap by providing a proactive approach to protecting against voice cloning. Instead of detecting cloned voices, VocalShield focuses on preventing cloning efforts from the start. It employs a combination of advanced audio distortion techniques such as pitch shifting, noise injection, and frequency modulation to subtly alter the audio in a way that remains imperceptible to human listeners but disrupts AI models trained to replicate human speech. Some key benefits of using VocalShield will be prevention over detection, dynamic adaptability, and ease of use for content creators. VocalShield provides a powerful tool for content creators and individuals concerned about voice security, ensuring they can maintain control over their voice identity and protect their digital content from unauthorized exploitation.

4. Glossary

Deepfake technology involves employing advanced AI algorithms to produce media-videos, audios, images, and text.

Voice cloning: subset within deepfake technology, focusing on audio manipulation.

5. References

Burgess, Matt. "Voice Recognition Privacy & Speech Changer." Wired, 1 June 2022, https://www.wired.com/story/voice-recognition-privacy-speech-changer/.

McAfee Corp. "McAfee Unveils Advanced Deepfake Audio Detection Technology at CES 2024 to Defend Against Rise in AI-Generated Scams and Disinformation." McAfee Newsroom, 8 Jan. 2024, https://www.mcafee.com/zh-tw/consumer-corporate/newsroom/press-releases/press-release.html?news_id=509b05a3-65e9-46d4-9f17-2dbc606e111a&csrc=vanity&offerid=403203.

Hong, Tan Jian. "Uncovering the Real Voice: How to Detect and Verify Audio Deepfakes." Medium, 14 Nov. 2023, https://medium.com/htx-s-s-coe/uncovering-the-real-voice-how-to-detect-and-verify-audio-deepfakes-42e480d3f431.