

Speech Deepfakes Detection with Fast Fourier Transform using Complex Linear Algebra

Kevin Wirya Valerian - 13524019

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

kevin.wirya.valerian@gmail.com 13524019@std.stei.itb.ac.id

Abstract—Nowadays, the growing realism of speech deepfakes demands detection methods grounded in fundamental signal theory rather than purely data-driven models. A well-known speech deepfake detection approach is based on the Fast Fourier Transform (FFT), rooted in complex linear algebra and geometric signal representations. Speech signals are modeled as vectors in complex vector spaces, where the FFT acts as a structured linear transformation projecting time-domain data onto orthonormal bases of complex exponentials. From a geometric perspective, genuine and synthetic speech occupy distinct regions in high-dimensional spectral spaces. Subtle inconsistencies in phase alignment, spectral symmetry, and energy distribution are captured through algebraic and geometric measures, enabling robust and interpretable deepfake detection. The study contributes significantly to rapid AI and real speech distinction.

Index Terms—Fast fourier transform, deepfakes detection, complex linear algebra, spectral geometry

I. INTRODUCTION

Recent advances in speech synthesis and voice conversion have enabled machines to generate highly realistic human speech. Modern text-to-speech and voice cloning systems are capable of producing audio that is increasingly difficult to distinguish from genuine human recordings. A study reports that human participants were only able to accurately distinguish real from AI-generated voices with an accuracy of 70.4%. This leads to a serious risks that can be posed by speech deepfakes in areas such as biometric authentication and digital forensics. Reports indicate that voice-based fraud has increased significantly in recent years, with financial losses reaching billions of dollars globally. Deepfake-related losses have already reached \$1.56 billion, with over \$1 billion occurring in 2025 alone.

While deep learning models have achieved impressive performance in generating natural-sounding speech, detecting such synthetic audio remains a challenging task. Many existing detection methods rely heavily on large neural networks trained on specific datasets. Although effective, these approaches often lack interpretability and tend to degrade when exposed to unseen deepfake generation methods. Moreover, purely data-driven models can be computationally expensive and difficult to analyze, making them less suitable and reliable at present time.

Speech signals, however, are fundamentally mathematical objects. A digital audio signal can be represented as a finite

sequence of samples, which naturally forms a vector in a high-dimensional space. Transforming this signal into the frequency domain using the Fast Fourier Transform (FFT) reveals its spectral structure, including energy distribution and phase behavior. These properties are governed by well-established principles of linear algebra and geometry, such as vector spaces, orthonormal bases, inner products, and unitary transformations. Importantly, while deepfake speech models often reproduce spectral magnitudes accurately, subtle geometric inconsistencies in phase relationships and spectral structure may persist.

Motivated by this observation, this paper proposes a speech deepfake detection approach grounded in the linear algebraic and geometric foundations of the FFT. The proposed method aims to distinguish genuine and synthetic speech through interpretable mathematical measures. The development of this approach is supported by core theories from complex linear algebra and geometric signal analysis, which together provide a principled framework for understanding and detecting anomalies in synthetic speech. hello hi

II. THEORETICAL FRAMEWORK

III. METHODOLOGY

IV. IMPLEMENTATION

V. CASE ANALYSIS

VI. CONCLUSION

REFERENCES

- [1] Ni, Y., et al. (2024). "A Deepfake Detection Algorithm Based on Fourier Transform of Biological Signal." Tech Science Press. (Accessed on December 16, 2025) <https://www.techscience.com/cmc/v79n3/57116>
- [2] Mai, K. T., Bray, S., Davies, T., and Griffin, L. D. Warning: Humans cannot reliably detect speech deepfakes. PLOS ONE, 18(8):1–20, 2023. (Accessed on December 19, 2025) <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0285333>

STATEMENT

Hereby I declare that this paper that I have written is my own work, not a reproduction or translation of someone else's work and not plagiarized.

Bandung, 20 Juni 2025

A handwritten signature in black ink, appearing to read 'Kevin', with a stylized flourish underneath.

Kevin Wirya Valerian
13524019