Poster: Preventing Fake News through Live Speech Signature

Irtaza Shahid, Nirupam Roy
{irtaza, niruroy}@umd.edu
University of Maryland College Park

ABSTRACT

Malicious editing to alter audiovisual content has become increasingly prevalent in recent years, as it allows for targeted defamation, the dissemination of disinformation, and the incitement of political unrest. Public speeches and statements made by political leaders, public figures, and celebrities are particularly vulnerable to such attacks, as they have the power to sway public opinion. The widespread use of smart devices to record live speeches, combined with unrestricted content sharing and redistribution on social media platforms, makes it difficult to prevent the spread of manipulated media. Existing solutions, which rely on source control over the media, are not effective for live events. This paper presents *TalkLock*, a speech integrity verification system that can enable live speakers to protect their speeches from malicious alterations even when the speech is recorded by any member of the audience. The core idea is to generate meta-information from the speech signal in real-time and disseminate it through a secure QR code-based screen-camera communication. The QR code when recorded along with the speech embeds the meta-information in the content and it can be used later for independent verification in stand-alone applications or online platforms. A user study with live speech and real-world experiments with different types of voices, languages, environments, and distances show that *TalkLock* can verify fake content with 94.4% accuracy.

CCS CONCEPTS

• **Security and privacy** → **Authentication**; **Tamper-proof and tamper-resistant designs**; **Usability in security and privacy**.

KEYWORDS

Deepfake; Speech verification; Voice features; QR code

ACM Reference Format:
Irtaza Shahid, Nirupam Roy. 2023. Poster: Preventing Fake News through Live Speech Signature. In *The 21st Annual International Conference on Mobile Systems, Applications and Services (MobiSys '23), June 18–22, 2023, Helsinki, Finland*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3581791.3597369

1 INTRODUCTION

DeepFakes – a common term for intentionally deceptive synthetic audiovisual content – has become a common way to spread disinformation. The recent viral videos showing former US presidents Barack Obama and Donald Trump making misleading or inappropriate statements demonstrate the capabilities of deepfakes. Although deepfakes are difficult to produce due to the need for specialized equipment and large amounts of training data, "shallow" fakes pose a more pressing danger. These shallow fakes are often readily available audiovisual clips that have been selectively edited or manually altered to mislead the viewers and create confusion. For example, a manipulated video of the U.S. House of Representatives Speaker Nancy Pelosi, where she is portrayed as intoxicated and slurring her words during a press conference. Audiovisual media is the most influential type of content on social media, which makes synthetic videos a powerful tool for political gain. A prime example is the fake video of Ukrainian President Volodymyr Zelenskyy appearing to tell his citizens to surrender the fight against Russia. This manipulation of social attitudes, propaganda, political slander, and defamation is considered a significant threat to democracy. Despite tremendous effort, techniques to protect and verify the authenticity of audiovisual content are limited for public recordings of live events. In this poster, we ask the question – *Is it possible to protect a publicly delivered speech from alteration even when anyone from the audience can record it live and publish it?*

To detect altered audiovisual media, two main strategies are used: (a) artifact-based detection and (b) meta-information-based verification. The first strategy involves creating discriminators that can detect small abnormalities in the audio or video introduced during the editing or synthesis process, such as inconsistent eye blinking, lip movements, head poses, emotional cues, blood flow, heart rate, and facial expressions. However, artifact-based detection is limited by the current state of alteration techniques, and as these techniques become more advanced, artifact-based detection becomes less effective. This creates an ongoing battle between those producing undetectable fake content and those trying to detect it. The second strategy involves adding specific signatures to the media file to verify its integrity. This strategy requires cooperation from recording systems to append the meta-information, making it difficult for audiences to record verifiable videos without software modifications.

In this poster, we present *TalkLock*, a novel speech integrity verification system that can enable live speakers to protect their speeches from malicious alterations even when the speech is recorded by any member of the audience. The core idea is to generate real-time meta-information containing identifying features from the live speech signal and disseminate it in a way that it will automatically get included in the recording. In the verification stage, the system expects matching features from the audio data in the recording with this meta-information. It is a first step towards a broader vision of mitigating the spread of misinformation. For a detailed system description and evaluation, please refer to our MobiSys 2023 paper [4].

MobiSys '23, June 18–22, 2023, Helsinki, Finland

ACM ISBN 979-8-4007-0110-8/23/06.
https://doi.org/10.1145/3581791.3597369

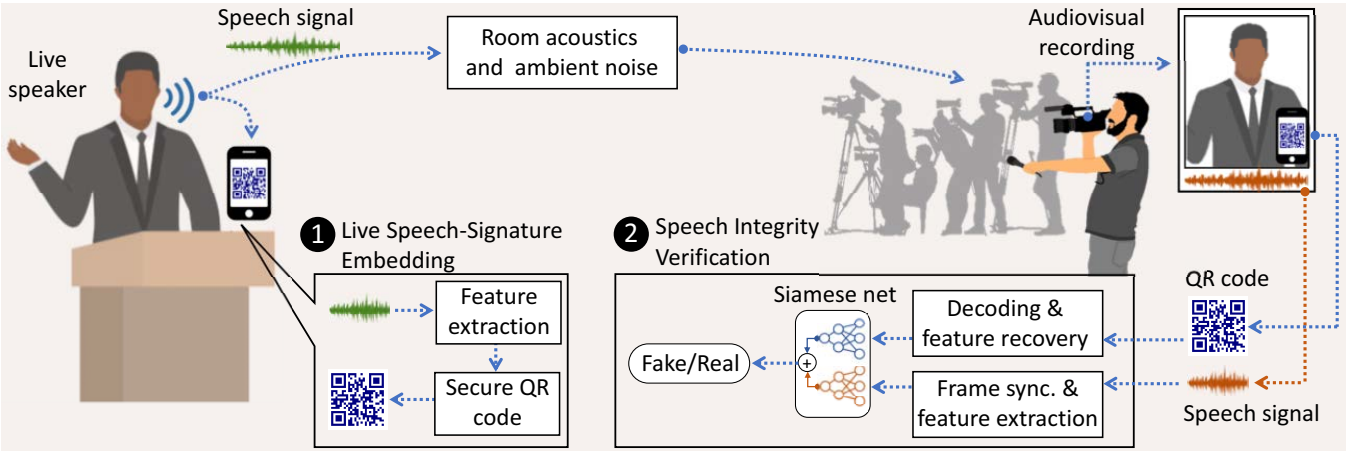

Figure 1: Design overview: ❶ The live speech-signature embedding module extracts features from speech signals in real-time and generates a sequence of cryptographically secure QR codes. ❷ The speech integrity verification module uses our algorithm to check the speech in the content under question matches with the features recovered from the QR codes visible in the video.

2 DESIGN OVERVIEW

TalkLock design has two primary modules.

❶ The **live speech-signature embedding module** extracts features from speech signals in real-time and generates a sequence of cryptographically secure Quick Response (QR) codes. The features are carefully designed to capture complete contextual information while remaining immutable to the ambient noise and multipath channel propagation. This module runs independently on a computing device that has a microphone for recording sound and a screen for displaying the QR code sequence. Any common smart device (e.g., a tablet or a smartphone) or a laptop can serve this purpose. A user who wants to protect against malicious alteration of her speech places this module with the QR code screen visible to the audience so that the audiovisual recordings capture the dynamically updated codes. The features required for verification, along with cryptographic signatures, get embedded in the live recordings of the event through the QR codes.

❷ The **speech integrity verification module** uses our algorithm to match the speech in the content under question with the features extracted from the QR codes present in the video. This module can operate on a variety of computing platforms, including standalone computers, smart devices, and online social platforms. It is important to note that the speech quality in the content may naturally vary from the source sound signal due to different propagation channels. Even if recording devices are placed a few inches apart, they can receive a distinct combination of path lengths, resulting in distinct signal qualities [3]. Although this phenomenon is beneficial for spatial sensing [1, 2], it presents a critical challenge in matching speech signals for verification. Our verification algorithm is designed using a Siamese network for robust integrity verification despite these inevitable differences in a live recording.

3 PRELIMINARY EVALUATIONS

We conducted a comprehensive evaluation of *TalkLock* in real-world scenarios. To replicate live speech recordings, we used a laptop speaker to emulate a speaker's voice. The live speech-signature embedding module was implemented on a laptop, and an iPhone 11 was used to record the speech samples. We aimed that our verification system should work effectively with any voice, without the need for additional calibration or training. Figure 2.a, shows that *TalkLock* achieves a mean classification accuracy of 94% under a diverse set of voices. Additionally, we tested *TalkLock* in seven different acoustic environments having unique multipath and ambient noise. Figure 2.b shows that *TalkLock* achieves 87.7% classification accuracy, demonstrating its robustness in these challenging scenarios.

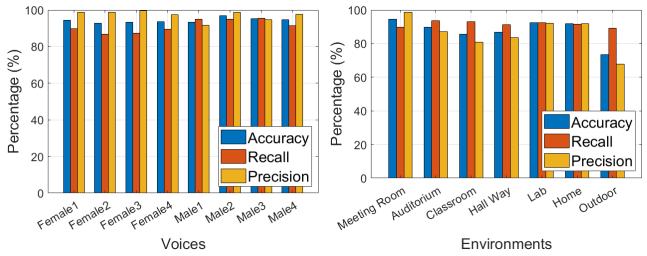

Figure 2: *TalkLock*'s identification performance under different (a) voices, and (b) environments.

REFERENCES

[1] Yang Bai, Nakul Garg, and Nirupam Roy. Spidr: Ultra-low-power acoustic spatial sensing for micro-robot navigation. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, pages 99–113, 2022.

[2] Nakul Garg, Yang Bai, and Nirupam Roy. Owlet: Enabling spatial information in ubiquitous acoustic devices. In *The 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM, 2021.

[3] Irtaza Shahid, Yang Bai, Nakul Garg, and Nirupam Roy. Voicefind: Noise-resilient speech recovery in commodity headphones. In *Proceedings of the 1st ACM International Workshop on Intelligent Acoustic Systems and Applications*, pages 13–18, 2022.

[4] Irtaza Shahid and Nirupam Roy. "*Is this my president speaking?*" tamper-proofing speech in live recordings. In *The 21st Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '23), June 18–June 22, 2023, Helsinki, Finland*. ACM.