Audio zero-watermarking method based on singular spectrum analysis

2210191 Rui Watanabe

Digital content, encompassing audio, images, and video, has become an indispensable component across numerous domains in modern society, ranging from entertainment and communication to legal evidence and security systems. However, the inherent ease with which digital audio files, in particular, can be perfectly copied, seamlessly edited, and widely redistributed presents significant challenges. Consequently, issues related to copyright infringement, unauthorized usage, and critically, malicious tampering for disinformation or evidence manipulation, have escalated into serious societal and technological concerns. Traditional audio watermarking techniques attempt to address these by directly embedding identification or authentication data into the host audio signal itself. While conceptually straightforward, this approach often introduces perceptible distortions, degrading the listening experience, and crucially, the embedded watermarks frequently lack resilience against common signal-processing operations (attacks), whether incidental or intentional, such as compression, filtering, or noise addition.

To overcome these inherent limitations of conventional embedding, zero-watermarking schemes have emerged as a promising alternative. These methods fundamentally differ by not modifying the host signal. Instead, they generate a binary pattern, often termed a watermark key or robust hash, derived from perceptually significant, intrinsic features of the audio content itself, combined with an external secret key. This generated pattern is then securely stored alongside the original audio. For verification, the process is repeated on the potentially altered audio, and the newly derived pattern is compared to the stored original. This approach intrinsically preserves the original audio fidelity while providing a mechanism for robust authenticity and integrity verification.

This study focuses on the development and evaluation of a zero-watermarking framework specifically designed for audio signals, particularly speech. The primary objective is to achieve a delicate balance: the framework must exhibit strong robustness against common, benign processing operations like lossy compression and resampling, which are frequent in standard distribution channels, yet simultaneously possess deliberate fragility, meaning it should reliably detect malicious tampering attacks intended to alter the content's meaning or authenticity. The proposed method uniquely leverages the inherent sparsity characteristic of speech signals. It employs singular spectrum analysis (SSA), a powerful technique for time series analysis, to decompose the signal and extract temporally stable,

low-rank component features that are less susceptible to noise and common distortions. These stable features form the basis for subsequent binary pattern generation, designed to capture the essential characteristics of the speech segment.

The performance of the proposed SSA based zero-watermarking system is evaluated against a notable baseline method developed by Ichikawa and Unoki, which utilizes auditory spectrum features. The primary evaluation metric employed is the bit error rate (BER) calculated between the original binary pattern and the pattern extracted from the processed or potentially tampered audio. Within this evaluation context, a BER value at or below 1% (BER $\leq 1\%$) is considered indicative of successful watermark detection, confirming the audio's authenticity despite potential benign processing. Conversely, significantly higher BER values are interpreted as evidence of potential tampering or unacceptable levels of distortion.

Experimental results demonstrate that the proposed method successfully achieves the target BER $\leq 1\%$ under several common processing conditions, including MP3 and AAC compression, resampling operations, and requantization, thereby confirming its robustness to these specific modifications. However, the system exhibits significantly higher BER, indicating watermark failure, when subjected to low-bitrate speech codecs such as G.711 and G.719, suggesting that the features extracted by SSA are sensitive to the specific type of quantization and signal representation used by these codecs. While the proposed method shows performance improvements over the baseline for MP3 and AAC robustness, its overall robustness across a wider range of operations is found to be somewhat inferior to the established conventional method. Furthermore, a critical limitation emerged regarding tampering detection: the system demonstrated excessive robustness, meaning that even after malicious modifications were applied, the extracted features did not change sufficiently. This resulted in an insufficient escalation of the BER, leading to low detection accuracy for certain types of malicious tampering.

Future research directions will refining the zero-watermarking framework to address these limitations. A key focus will be the redesign of the key-generation process, specifically targeting the binarization stage. The current binarization approach appears to lack sufficient temporal localization, making it overly resilient to localized tampering. The goal is to introduce the necessary fragility to reliably detect intentional attacks by incorporating finer temporal details or adaptive thresholding mechanisms, while carefully maintaining the achieved robustness against common, benign signal processing operations essential for practical applications.