Towards an Enhanced Audio Watermarking Approach Balancing Robustness and Imperceptibility

Jiabei Zhou, Nan Liang, Xinyan Shen

[1]School of Informatics, Xiamen University
30920241154586@stu.xmu.edu.cn, 30920241154553@stu.xmu.edu.cn, 30920241154573@stu.xmu.edu.cn

Abstract

Audio watermarking has gained increasing attention due to the rise of generative audio misuse and growing awareness of copyright protection. While DNN-based methods show promise, challenges remain in balancing robustness, imperceptibility, and payload capacity. This study proposes an approach leveraging dense and residual connections to enhance performance across these key evaluation metrics, offering a step forward in addressing existing limitations. Specifically, our method achieves a high SNR of 28.42 dB under eight main audio attack scenarios with a watermark capacity of 32 bps, and the BER for watermark extraction is nearly 0%. Its effectiveness was further validated through thorough comparison with recent state-of-the-art methods and extensive experiments.

Introduction

With the advancement of contemporary technology, the importance of audio as a multimedia broadcasting channel has become increasingly prominent. Since audio is the main medium for voice, music, broadcasting, etc., audio copyright protection is becoming increasingly important for audio creators. Audio watermarking can help audio owners confirm the legitimacy of their audio content by integrating copyright information, thereby achieving the purpose of copyright protection (Seok, Hong, and Kim 2002; Dhar and Kim 2011; Yassine, Bachir, and Aziz 2012).

In addition, due to the rapid development of generative audio technology (Kreuk et al. 2022; Borsos et al. 2023) and the public sharing of voice information, some criminals have begun to use generated audio for voice fraud. These incidents have seriously affected the development of generated audio and made people worry about the leakage of their voice information. Therefore, audio users should pay special attention to verifying the authenticity of their audio. Audio watermarking inserts watermark information, allowing users to recover these watermarks to identify the source of the audio and confirm its legitimacy, thereby preventing the abuse of generated music.

At present, some scholars are studying audio watermarking methods (Charfeddine et al. 2022) to better solve these problems. Among them, DeAR (Liu et al. 2023) implemented the anti-copying technology of audio watermarking to help audio owners better protect their copyrights, and Liu et al. (Liu et al. 2019) implemented the protection of patient information by using audio watermarks in medical audio. WavMark (Chen et al. 2023) implemented the first model for embedding watermarks on generated audio datasets by using generated audio datasets for model training, and achieved excellent performance results. However, audio watermarking research based on deep learning often focuses on enhancing a single feature while ignoring the performance improvement of other features. For example, WavMark (Chen et al. 2023) exhibits good imperceptibility, but poor performance in extraction BER under common attacks; DeAR (Liu et al. 2023) obtains relatively good anti-recording performance after embedding audio, but weak imperceptibility. To illustrate the challenge, we draw on a metaphor:

Metaphor: Designing an audio watermarking algorithm is akin to balancing a scale: one side represents robustness, the other imperceptibility, while embedding capacity remains a fixed point of reference. Leveraging deep learning techniques, the algorithm serves as a mediator, striving to harmonize these competing factors. The goal is to ensure that imperceptibility and robustness are optimized in tandem, achieving a well-rounded performance.

Building on this principle, this study introduces a novel audio watermarking technique, which seeks to excel across all three critical dimensions—embedding capacity, imperceptibility, and robustness. The primary contributions of this research are as follows:

1. Inspired by prior studies (Chen et al. 2023; Liu et al. 2023; Zhu et al. 2018), we propose an innovative audio watermark encoder-decoder framework that integrates dense and residual connections. This approach achieves superior watermark embedding and extraction, delivering an SNR exceeding 28 dB for audio with embedded watermarks and achieving a 0% bit error rate (BER) in watermark extraction under no-attack conditions.

2. To strengthen robustness, we incorporate eight conventional audio attack layers into the algorithm. Experimental results demonstrate that extraction BER remains at 0% under six attack scenarios and below 1.00% under the remaining scenarios, such as Gaussian noise and low-pass filtering attacks.

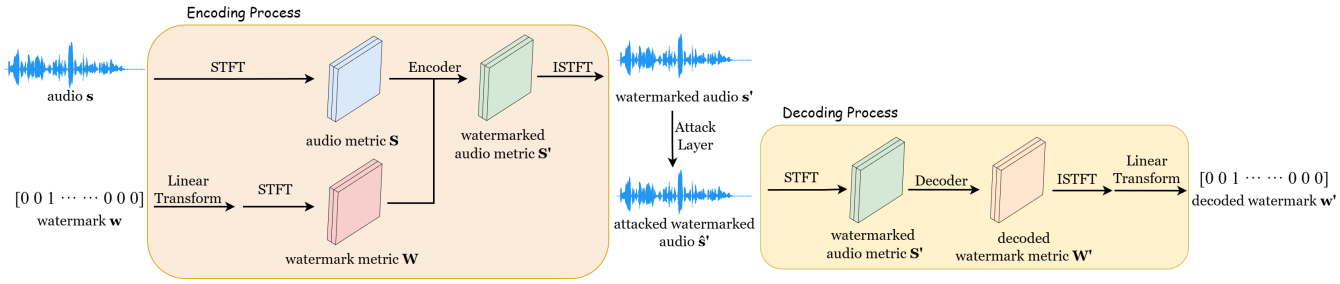

Figure 1: The Overall Architectures. This diagram shows the main framework of this algorithm, which can be divided into encoding process and decoding process. Encoding process involves a series of operations on the input audio sequence and the binary watermark sequence to produce an audio sequence with the embedded watermark. Decoding process, on the other hand, yields the decoded watermark sequence by decoding the input audio sequence with the encoded watermark.

3. Our method successfully increases the embedding capacity to 40 bps while maintaining strong imperceptibility and robustness, achieving the highest known embedding capacity in the field with balanced performance across all evaluation metrics.

By addressing the challenges of balancing these competing demands, this study aims to advance the field of audio watermarking through a powerful and high-performing approach.

Related work

The field of digital watermarking(Hartung and Kutter 1999) study was first formally introduced by Schyndel R.G(Van Schyndel, Tirkel, and Osborne 1994), who initially proposed the term in the 1994 IEEE International Conference on Image Processing. Most of the present digital watermarking research focuses on image watermarking(Begum and Uddin 2020; Wan et al. 2022), audio watermarking(Hua et al. 2016a) and video watermarking(Asikuzzaman and Pickering 2017),(Doerr and Dugelay 2003), and the growth of image watermarking is more rapid. Audio watermarking (Hua et al. 2016b) is a branch of digital watermarking (Hartung and Kutter 1999), which refers to embedding data into audio media. According to the embedding position, audio watermarking can be divided into time domain watermarking technology and transform domain watermarking technology. The initial paper (Nejad, Mosleh, and Heikalabad 2019) proposed a least significant bit (LSB) audio watermarking algorithm, which realizes the conversion of the scrambled watermark image into a quantum bit sequence, and then embeds the quantum bit sequence into the audio signal using the embedding key. Zhong (Zong et al. 2021) proposed an audio watermarking algorithm based on nonlinear echo generation, which embeds a time delay sequence in the audio signal, effectively enhancing the imperceptibility of the audio after the watermark is embedded. Due to the lack of concealment of time domain technology and its susceptibility to damage, in subsequent research, embedding watermarks in the transform domain has become more and more common. Aniruddha Kanhe and Aghila Gnanasekaran (Kanhe and Gnanasekaran 2018) proposed an audio watermarking scheme based on DCT and singular value decomposition to embed the watermark into the low-frequency components of the audio.

Traditional audio watermarking techniques usually embed the watermark into the selected frequency band coefficients by applying various transforms to the audio. M. Yamni (Yamni et al. 2022) proposed a robust audio/speech blind watermarking algorithm that combines discrete Tchebichef moment transform (DTMT), linear-nonlinear hybrid mapped lattice chaotic system (MLNCML) and discrete wavelet transform (DWT). Zhang (Zhang et al. 2023) performed discrete wavelet transform (DWT), graph-based transform (GBT), and singular value decomposition (SVD) on the audio signal to obtain the transform coefficients, and then used spread spectrum technology to embed the watermark into the audio. Although traditional audio watermarking research has made significant progress, due to its limitations and the fact that applications often rely on experience, progress has gradually slowed down in recent years.

In recent years, audio watermarking technology combined with DNN has gradually become an important direction in the field of audio watermarking research. Kosta Pavlović (Pavlović et al. 2022) proposed two adversarial neural networks as encoder and decoder to embed and extract watermarks, achieving good overall performance. DeAR (Liu et al. 2023) designed an audio re-recording watermark based on deep learning. By introducing a distortion layer to simulate the re-recording effect, the algorithm can learn to resist common distortion attacks. SilentCipher (Singh et al. 2024) combines a threshold based on a psychoacoustic model to achieve an imperceptible watermark, which effectively enhances the robustness of the watermark algorithm and the feasibility of professional applications. WavMark (Chen et al. 2023) proposed an audio watermarking algorithm based on a reversible neural network, which solves the current problem of watermark positioning in this field, while achieving an embedding rate of up to 32 bps and maintaining a high imperceptibility. AudioSeal (San Roman et al. 2024) is the first audio watermarking technology designed specifically for AI speech local detection, achieving state-of-the-art performance in terms of robustness and imperceptibility.

Proposed Solution

Architecture

The algorithm in this paper can be divided into the encoding process, attack layer, and decoding process. Figure 1 depicts the overall architecture of the algorithm.

During the encoding process, a Fourier transform is applied to the audio sequence to obtain an audio matrix. The watermark sequence undergoes a linear transformation to stretch it to the same length as the audio. Then, a Fourier transform is performed to obtain the watermark matrix. Input the audio matrix and the watermark matrix into a neural network designed with combined residual connections and dense connections for encoding operations, resulting in the encoded audio matrix. Perform an inverse Fourier transform on the audio matrix with the embedded watermark to convert it from matrix form back into sequence form, thus obtaining the audio sequence with the embedded.

During the decoding process, a Fourier transform is applied to the audio embedded with the watermark to obtain an audio matrix. This matrix is then decoded by the RDN-constructed decoder, resulting in the extracted watermark matrix. Subsequently, an inverse Fourier transform and linear transformation are performed on this watermark matrix to obtain the extracted watermark sequence.

In addition to the encoding and decoding processes, in order to enhance the model's robustness against common attacks, the algorithm also introduces an attack layer. By adding common audio attacks to the attack layer, the audio is processed to obtain the audio with embedded watermarks that have been attacked. Then, input it into the subsequent decoding process, thereby training the decoder in the decoding process to learn the ability to resist common audio attacks. The attacks introduced by the attack layer include Gaussian white noise, random cropping, low-pass filtering, resampling, amplitude scaling, MP3 compression, quantization, and echo adding.

Encoder Design

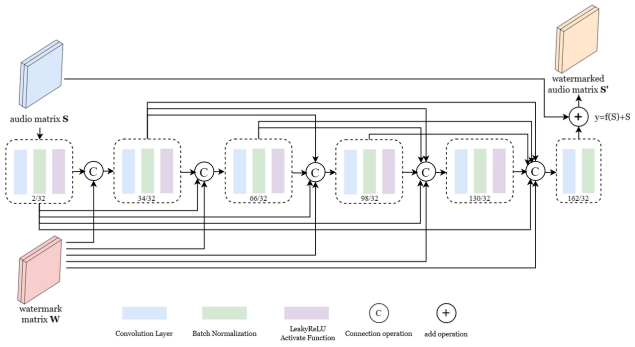

Figure 2: The Network Structure of *Encoder*.

The *Encoder* is implemented using a carefully designed neural network that takes the watermark matrix W and the audio matrix S (after applying the Short-Time Fourier Transform, STFT) as inputs to generate the embedded audio matrix S'.

In the design of this deep neural network, the number of layers is minimized while incorporating dense and residual connections to enhance performance and efficiency.

Dense connections are employed to enable the network to learn the features of the data more thoroughly with fewer layers. As a result, the *Encoder* consists of only six convolutional blocks. Each of the first five blocks includes a convolutional layer, a batch normalization layer, and a Leaky ReLU activation function, while the final block contains only a convolutional layer and a Leaky ReLU activation function.

For the residual connection, the input audio matrix is directly added to the output of the final layer, a concept also applied in the DeAR algorithm (Liu et al. 2023). This residual operation transforms the output of the convolutional blocks from generating a new audio matrix to calculating the residual result of the watermark and audio matrix. This significantly reduces the complexity of the neural network while effectively improving the imperceptibility of the watermarked audio.

Decoder Design

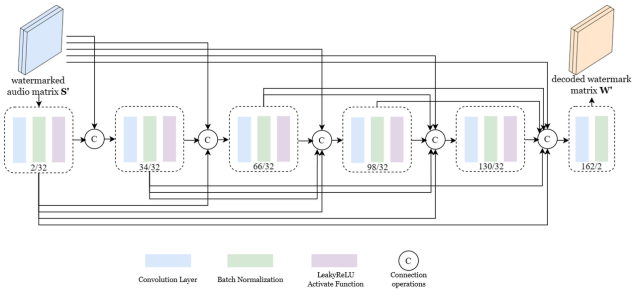

Figure 3: The Network Structure of *Decoder*.

The design of the *Decoder* is relatively straightforward, utilizing a densely connected neural network to extract the embedded watermark matrix W' from the watermarked audio matrix S'. While the convolution block structure in the *Decoder* mirrors that of the *Encoder*, it incorporates an additional dense connection mechanism to enhance learning efficiency.

To ensure the neural network fully captures the embedded watermark signal encoded in the audio matrix, the watermarked audio matrix is connected to the input of each subsequent convolution block. This dense connection design enables the network to effectively learn the embedded watermark signal, improving performance while reducing the required number of layers. By leveraging these design principles, the *Decoder* achieves accurate watermark extraction with a streamlined structure.

Loss Function

For *Encoder*, the loss function needs to reflect the difference between the audio signal before embedding and the audio signal after embedding the watermark. Therefore, this paper chooses the Mean Square Error (MSE) between the original

Table 1: Compare with Current Leading Method

Model	Capacity (↑)	SNR (↑)	BER(%)(↓)								
			No Attack	Guassian Noise	Random Cropping	Low-pass Filtering	Resampling	Amplitude Scaling	MP3 Compression	Quantization	Echo Adding
Robust DNN (Pavlović et al. 2022)	2bps	24.48	**0.00**	**0.00**	**0.00**	0.14	0.25	**0.00**	**0.00**	**0.00**	**0.00**
DeAR (Liu et al. 2023)	9bps	26.18	**0.00**	0.01	0.01	0.94	**0.00**	0.01	0.03	**0.00**	\
WavMark (Chen et al. 2023)	**32bps**	**38.32**	0.65	0.30	0.06	**0.00**	**0.00**	**0.00**	**0.00**	0.72	**0.00**
Proposed	**32bps**	28.42	**0.00**	0.06	**0.00**	0.11	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

audio matrix and the audio matrix with the embedded watermark to measure this difference. The specific calculation form ula for the MSE is as follows.

$$L_{encoder} = MSE(s, s') = \frac{1}{N}\sum_{i=1}^{N}(s(i) - s'(i))^2 \quad (1)$$

For *Decoder*, its loss function needs to reflect the gap between the embedded watermark signal and the extracted watermark signal, so in this paper, we choose the Binary Cross Entropy with Logit Loss (BCE) between the embedded watermark signal and the extracted watermark signal to measure the gap. The specific formula for the BCE is as follows.

$$BCE(w, w') = -\frac{1}{M}\sum_{i=1}^{M}(w_i log(w'_i) + (1-w_i) log(1-w'_i)) \quad (2)$$

The *Decoder* not only has audio input without attack, but also has audio input with attack. So its loss function needs to calculate the BCE of the original watermark and the extracted watermarks without attack and under attack. The loss function of the *Decoder* is as follows.

$$L_{decoder} = \mu BCE(w, w') + \upsilon \sum_{i=1}^{n} BCE(w, \hat{w}'_i) \quad (3)$$

Since the final joint loss function will be composed of *Decoder* and *Encoder* loss function together, by giving different weights to the computed loss thus affecting the final training effect of the model, the specific formula is shown in Eq. 4 .

$$L(s, s', w, w') = \alpha L_{encoder} + \beta L_{decoder} \quad (4)$$

Experiments

Audio watermarks are measured by imperceptibility, embedding capacity, and robustness. In this paper, SNR is selected as the imperceptibility evaluation index and BER is selected as the robustness evaluation index. The embedding capacity is improved based on the assumption that both have better indexes.

Comparison with Advanced Work

To demonstrate the superiority of our proposed method, we compared it against top-tier audio watermarking algorithms published in the last two years. The comparison results are summarized in Table 1.

In terms of ***embedding capacity***, our model achieves 32 bps, which matches the ideal embedding capacity of WavMark (Chen et al. 2023). This is significantly higher than the capacities of DeAR (Liu et al. 2023) (9 bps) and Robust DNN (Pavlović et al. 2022) (2 bps), showcasing its advantage in embedding efficiency. Regarding ***imperceptibility***, the algorithm delivers an SNR of 28.42 dB, ranking second only to WavMark's 38.32 dB, while still maintaining high audio quality. When evaluating ***robustness***, our method achieves a BER of 0% in the absence of attacks, outperforming WavMark's 0.65%. Under eight types of attacks, the BER of our algorithm remains consistently low, with non-zero values observed only under Gaussian white noise and low-pass filtering. In contrast, the robustness of the other algorithms significantly diminishes, as evidenced by their higher BERs across various attack scenarios.

These results highlight the balanced performance of our method, which excels in embedding capacity and robustness while maintaining competitive imperceptibility, setting it apart as a strong contender in state-of-the-art audio watermarking.

Performance on Different Datasets

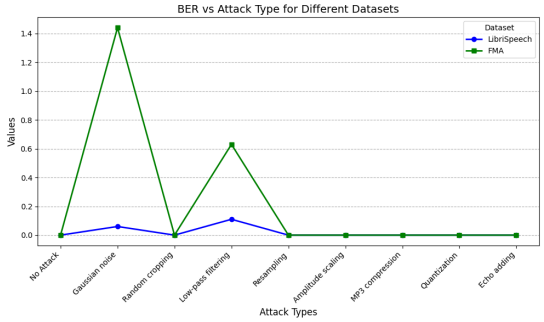

Figure 4: BER under Eight Attack Types for Different Datasets

Table 2: SNR and Average BER for Different Datasets

Dataset	**SNR**(↑)	**BER(%)**(↓)
LibriSpeech	28.42	**0.02**
FMA	**34.71**	0.23

To verify the generalizability and flexibility of the model design, we trained it on two different datasets: FMA(Ji, Luo, and Yang 2020), an audio dataset, and LibriSpeech(Panayotov et al. 2015), a speech dataset. The embedding capacity was fixed at 32 bps for both datasets. The results are presented in Figure 4 and Table 2.

In Figure 4, the robustness indicators of the two models on different datasets after training are illustrated. It is evident that the extraction BER of the model trained on the music dataset FMA increases under Gaussian noise and low-pass filtering attacks, indicating a reduced ability to resist common audio attacks. This suggests a potential trade-off between robustness and other performance metrics, which warrants further exploration.

By examining Table 2, we observe significant differences in imperceptibility. Specifically, the SNR of the model trained on the FMA dataset shows a notable improvement, reaching 34.71 dB. This improvement highlights the enhanced audio quality and imperceptibility of the model. Conversely, the average BER of the LibriSpeech dataset is lower than that of the FMA dataset, suggesting that the model trained on LibriSpeech exhibits better robustness across various scenarios. These results collectively underscore the importance of dataset selection in shaping model performance, particularly in balancing robustness and imperceptibility. Further analysis may help identify optimal configurations for specific application scenarios.

Behavior under Different Capacity

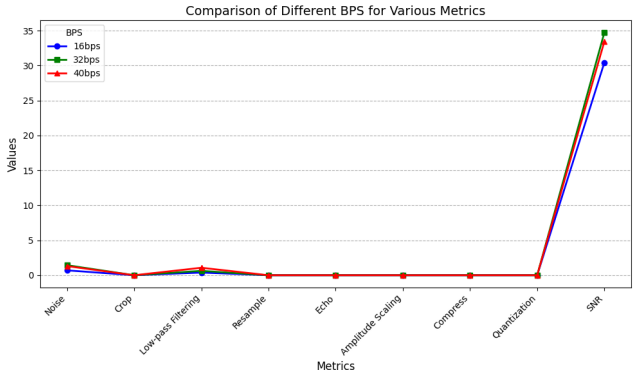

Figure 5: Comparison of Different BPS for Various Metrics

In the previous work, the algorithm set the embedding capacity of the algorithm at 32bps and achieved relatively good results. In order to better verify the flexibility of the model in terms of capacity, this paper continues to explore the performance of the model under different capacities by changing the embedding capacity to 16bps, 32bps and 40bps.

In Figure 5, you can see the results of the robustness evaluation in different embedding capacities. When the embedding capacity is reduced to 16bps, its extraction BER under Gaussian noise and low-pass filtering attacks is 0. 11% and 0. 07%, respectively, and the extraction BER under other attacks is 0%, indicating that the model with an embedding capacity of 16bps also has good robustness. As the embedding capacity increases, the BER gradually decreases, but the decrease is small. When the embedding capacity is increased to 40bps, the extracted Ber under Gaussian noise and low-pass filtering is reduced to 0.04% and 0.02%, and the robustness is improved compared with the low-capacity model. It can be seen that with the increase of embedding capacity, the imperceptibility of the method proposed in this paper decreases, but the robustness to attacks is improved.

Conclusion

In this study, we proposed an audio watermarking algorithm designed to achieve a balanced trade-off among embedding capacity, imperceptibility, and robustness. The algorithm integrates residual connections and dense connections while introducing eight types of attacks to enhance watermark embedding and extraction capabilities. Unlike many existing approaches, which primarily focus on improving a single aspect, our method emphasizes balanced optimization across all three key performance indicators. This focus necessitates careful consideration in the design of the framework and network structure.

Furthermore, we conducted a comparative analysis against state-of-the-art audio watermarking algorithms. The experimental results suggest that our algorithm demonstrates competitive advantages, achieving better overall performance compared to existing methods. By training the model on different datasets, including the music dataset FMA (Ji, Luo, and Yang 2020) and the speech dataset LibriSpeech (Panayotov et al. 2015), we observed consistently strong results. Additionally, we evaluated the algorithm's performance under varying embedding capacities and found that even at 40 bps, the model maintains excellent imperceptibility and robustness. This indicates that our method holds potential as a high-capacity audio watermarking solution with well-balanced performance across critical metrics.

References

Asikuzzaman, M.; and Pickering, M. R. 2017. An overview of digital video watermarking. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9): 2131–2153.

Begum, M.; and Uddin, M. S. 2020. Digital image watermarking techniques: a review. *Information*, 11(2): 110.

Borsos, Z.; Marinier, R.; Vincent, D.; Kharitonov, E.; Pietquin, O.; Sharifi, M.; Roblek, D.; Teboul, O.; Grangier, D.; Tagliasacchi, M.; et al. 2023. Audiolm: a language modeling approach to audio generation. *IEEE/ACM transactions on audio, speech, and language processing*, 31: 2523–2533.

Charfeddine, M.; Mezghani, E.; Masmoudi, S.; Amar, C. B.; and Alhumyani, H. 2022. Audio watermarking for security and non-security applications. *IEEE Access*, 10: 12654–12677.

Chen, G.; Wu, Y.; Liu, S.; Liu, T.; Du, X.; and Wei, F. 2023. Wavmark: Watermarking for audio generation. *arXiv preprint arXiv:2308.12770*.

Dhar, P. K.; and Kim, J.-M. 2011. Digital watermarking scheme based on fast Fourier transformation for audio copyright protection. *International Journal of Security and Its Applications*, 5(2): 33–48.

Doerr, G.; and Dugelay, J.-L. 2003. A guide tour of video watermarking. *Signal processing: Image communication*, 18(4): 263–282.

Hartung, F.; and Kutter, M. 1999. Multimedia watermarking techniques. *Proceedings of the IEEE*, 87(7): 1079–1107.

Hua, G.; Huang, J.; Shi, Y. Q.; Goh, J.; and Thing, V. L. 2016a. Twenty years of digital audio watermarking—a comprehensive review. *Signal processing*, 128: 222–242.

Hua, G.; Huang, J.; Shi, Y. Q.; Goh, J.; and Thing, V. L. 2016b. Twenty years of digital audio watermarking—a comprehensive review. *Signal processing*, 128: 222–242.

Ji, S.; Luo, J.; and Yang, X. 2020. A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions. *arXiv preprint arXiv:2011.06801*.

Kanhe, A.; and Gnanasekaran, A. 2018. Robust image-in-audio watermarking technique based on DCT-SVD transform. *EURASIP Journal on Audio, Speech, and Music Processing*, 2018(1): 16.

Kreuk, F.; Synnaeve, G.; Polyak, A.; Singer, U.; Défossez, A.; Copet, J.; Parikh, D.; Taigman, Y.; and Adi, Y. 2022. Audiogen: Textually guided audio generation. *arXiv preprint arXiv:2209.15352*.

Liu, C.; Zhang, J.; Fang, H.; Ma, Z.; Zhang, W.; and Yu, N. 2023. Dear: A deep-learning-based audio re-recording resilient watermarking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 13201–13209.

Liu, X.; Lou, J.; Fang, H.; Chen, Y.; Ouyang, P.; Wang, Y.; Zou, B.; and Wang, L. 2019. A novel robust reversible watermarking scheme for protecting authenticity and integrity of medical images. *Ieee Access*, 7: 76580–76598.

Nejad, M. Y.; Mosleh, M.; and Heikalabad, S. R. 2019. An LSB-based quantum audio watermarking using MSB as arbiter. *International Journal of Theoretical Physics*, 58(11): 3828–3851.

Panayotov, V.; Chen, G.; Povey, D.; and Khudanpur, S. 2015. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 5206–5210. IEEE.

Pavlović, K.; Kovačević, S.; Djurović, I.; and Wojciechowski, A. 2022. Robust speech watermarking by a jointly trained embedder and detector using a DNN. *Digital Signal Processing*, 122: 103381.

San Roman, R.; Fernandez, P.; Elsahar, H.; Défossez, A.; Furon, T.; and Tran, T. 2024. Proactive detection of voice cloning with localized watermarking. In *International Conference on Machine Learning*, volume 235.

Seok, J.; Hong, J.; and Kim, J. 2002. A novel audio watermarking algorithm for copyright protection of digital audio. *etri Journal*, 24(3): 181–189.

Singh, M. K.; Takahashi, N.; Liao, W.; and Mitsufuji, Y. 2024. SilentCipher: Deep Audio Watermarking. *arXiv preprint arXiv:2406.03822*.

Van Schyndel, R. G.; Tirkel, A. Z.; and Osborne, C. F. 1994. A digital watermark. In *Proceedings of 1st international conference on image processing*, volume 2, 86–90. IEEE.

Wan, W.; Wang, J.; Zhang, Y.; Li, J.; Yu, H.; and Sun, J. 2022. A comprehensive survey on robust image watermarking. *Neurocomputing*, 488: 226–247.

Yamni, M.; Karmouni, H.; Sayyouri, M.; and Qjidaa, H. 2022. Efficient watermarking algorithm for digital audio/speech signal. *Digital Signal Processing*, 120: 103251.

Yassine, H.; Bachir, B.; and Aziz, K. 2012. A secure and high robust audio watermarking system for copyright protection. *International journal of computer applications*, 53(17): 33–39.

Zhang, G.; Zheng, L.; Su, Z.; Zeng, Y.; and Wang, G. 2023. M-sequences and sliding window based audio watermarking robust against large-scale cropping attacks. *IEEE Transactions on Information Forensics and Security*, 18: 1182–1195.

Zhu, J.; Kaplan, R.; Johnson, J.; and Fei-Fei, L. 2018. HiDDeN: Hiding Data With Deep Networks. In *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XV*, 682–697. Berlin, Heidelberg: Springer-Verlag. ISBN 978-3-030-01266-3.

Zong, T.; Xiang, Y.; Natgunanathan, I.; Gao, L.; Hua, G.; and Zhou, W. 2021. Non-linear-echo based anti-collusion mechanism for audio signals. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29: 969–984.