VERİ BİLİMİ DERGİSİ
www.dergipark.gov.tr/veri

Stereo Audio Steganography based on Mid/Side Processing

Ali Erdem ALTINBAŞ[1]*, Yıldıray YALMAN[2]

[1]*Kocaeli University, Faculty of Engineering, Electronics and Communication Engineering, Kocaeli*
[2]*Piri Reis University, Faculty of Engineering, Computer Engineering, Istanbul*

Özet

Günümüzde ses steganografisi derin öğrenme, olasılık ve istatistik ve sinyal işleme tekniklerinin temel alındığı yöntemlerle geliştirilmektedir. Ancak insanların sesle olan iletişiminin büyük kısmı müzik ve haberleşme ile ilişkilidir. Bu çalışmada müzik teorisinin ve ses miksi tekniklerinin kullanıldığı yeni bir ses steganografisi algoritması geliştirilmiştir. Steganografi alanında insan duyma sisteminin hassaslığı nedeniyle taşıyıcı olarak ses dosyaları daha az tercih edilmektedir. Çünkü steganografide fark edilemezlik ilkesini sağlamak amacıyla genellikle veri gizleme kapasitesinden feragat edilir. Sunulan yöntem müzik teknolojilerinin ses steganografisinde kullanılmasının bir örneğini içerir. Özellikle ses mühendislerinin çoğu zaman kullandığı, sesin orta ve kenar bileşenlerin sinyal işleme araçlarıyla ayrı ayrı düzenlenmesi işlemi, ses steganografisine yeni bir bakış kazandırmaktadır. Taşıyıcı dosyanın tamamının kullanılması durumunda bile deneysel sonuçlar fark edilemezlik ilkesinin sağlandığını göstermektedir. Steganografideki diğer anahtar nokta olan sağlamlık konusunda ise geliştirilen yöntemin MP3 sıkıştırma algoritmasına bile dayanıklı olduğu tespit edilmiştir. Çalışmanın MATLAB kodlarına bu linkten ulaşılabilir: http://bit.ly/3R6uwZv

Anahtar Kelimeler: *Ses steganografisi, Veri gizleme, Steganografi, Dijital damgalama*

Makale Bilgisi

Başvuru:
10/10/2024
Kabul:
12/01/2025

Stereo Audio Steganography based on Mid/Side Processing

Abstract

The field of audio steganography is currently undergoing significant developments, with new methods being introduced that are based on deep learning, probability and statistics, and signal processing techniques. However, it is important to highlight that the majority of human communication with sound is related to music and communication. This paper introduces a novel audio steganography algorithm that integrates principles from music theory and audio mixing techniques. In the field of steganography, audio files are often less favored as cover media due to the heightened sensitivity of the human auditory system. To maintain the principle of imperceptibility, data-hiding capacity is typically reduced. However, the method proposed in this work demonstrates how music production technologies can be effectively applied in audio steganography, ensuring minimal perceptual impact. In particular, the process of separately mixing the mid and side components of the sound with signal processing tools, often used by mixing engineers, brings a new perspective to audio steganography. Experimental results show that even when the entire cover file is utilized for data embedding, the imperceptibility remains intact. Additionally, the proposed method exhibits strong robustness, maintaining message integrity even after MP3 compression also known as one of the common stego-attack-. MATLAB codes related to this study are available at the following link: http://bit.ly/3R6uwZv

Keywords*: audio steganography, mid/side, data hiding, steganography, watermarking*

* e-mail: alierdemaltinbas@gmail.com

1 Introduction

In today's digital world, ensuring the secure transmission of data has become a critical issue. With the rapid increase in data on the Internet, estimated to grow by 40% each year, a range of concerns has emerged, such as copyright violations, digital watermarking, forensic analysis, and authentication. Steganography, a technique that involves embedding information within a media file, offers a promising approach to addressing secure communication challenges [1]. Common media formats used for embedding hidden messages include images, videos, and audio files [2, 3, 4].

Sound is naturally linked to music, just as images are connected to photography and videos to television. Therefore, it's not surprising that techniques derived from music production are applied in audio steganography [5, 6].

The present work develops a data-hiding algorithm based on mid/side (M/S) processing a technique frequently used in music production.

M/S processing involves splitting the audio signal into two parts: the mid and the side. The mid component represents the sum of the left and right channels, while the side component captures the difference between these channels. This separation allows for more precise control and manipulation of stereo sound, particularly when adjusting stereo width, depth, and center clarity. For instance, enhancing the mid component can make center vocals or lead instruments stand out, while boosting the side component can expand the stereo image [7].

M/S processing has always been used by sound engineers to create a virtual stage experience. For example, Figure 1 shows the stage experience design of Childish Gambino's This Is America, winner of 4 Grammy Awards.

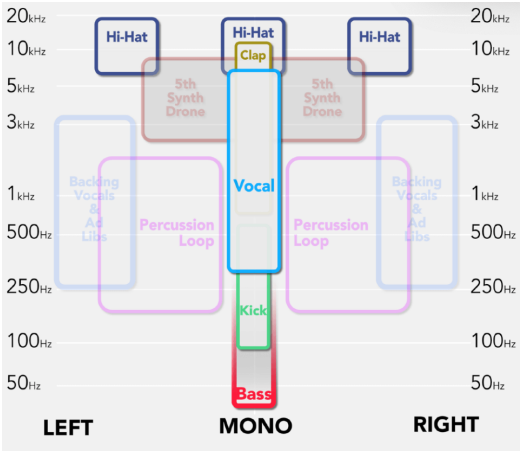

Figure 1. The virtual stage design of childish Gambino's this is America

A close examination of the stereo structure illustrated in Figure 1 reveals that the mid-range component contains the low-frequency components, which represent the fundamental elements of musical sound. Concurrently, instruments that are perceptually expected to be in the lead position are also situated within the mid-range component. This is particularly notable in the case of vocals, which are typically expected to be positioned on top of instrumental music. There is even a special phenomenon, known as the "in-your-face vocal," which exemplifies this phenomenon.

The side components are characterized by supporting instruments with relatively low volume and a predominance of mid frequencies. Consequently, songs played in mono loudspeakers often lack these components. In virtual stage, the listener's acoustic focus is directed towards the mid component. Therefore, minor alterations to the side components do not affect the listener's perception, given the nature of stereo music.

A key benefit of this technique is that it maintains mono compatibility. The mid component remains identical to the original mix in mono playback, allowing for a harmonious balance between stereo and mono sound.

As mentioned earlier, modifications to the mid components of an audio signal are often subtle and difficult to detect, making this technique particularly suitable for steganography.

2 Related works

Traditional audio steganography techniques, such as LSB, echo hiding, phase coding, and spread spectrum, are frequently adapted for various

applications [8, 9, 10, 11]. In recent years, however, machine learning approaches have also been incorporated into the field [12]. Despite this, there is still a limited number of studies that consider music theory in their methodology. To address this gap, Shiu et al. introduced a stereo audio steganography technique aimed at enhancing this aspect of the literature [13]. Their method utilizes stereo audio as the cover medium. The encoding process involves splitting the host audio into segments and embedding one bit of information within each segment. The corresponding decoding process segments the audio with embedded data and retrieves the hidden bits by analyzing the differences in domain-transferred data between the left and right audio channels.

This technique leverages a relationship between the human auditory system (HAS) and principles from music theory. Specifically, it accounts for the fact that humans perceive harmonic sounds as being nearly identical, even when subtle variations exist.

In another example of integrating music technologies, Su et al. proposed a steganography method using MIDI files. As shown in Figure 2, more advanced techniques like convolutional neural networks and generative adversarial networks were also incorporated into their research [14].

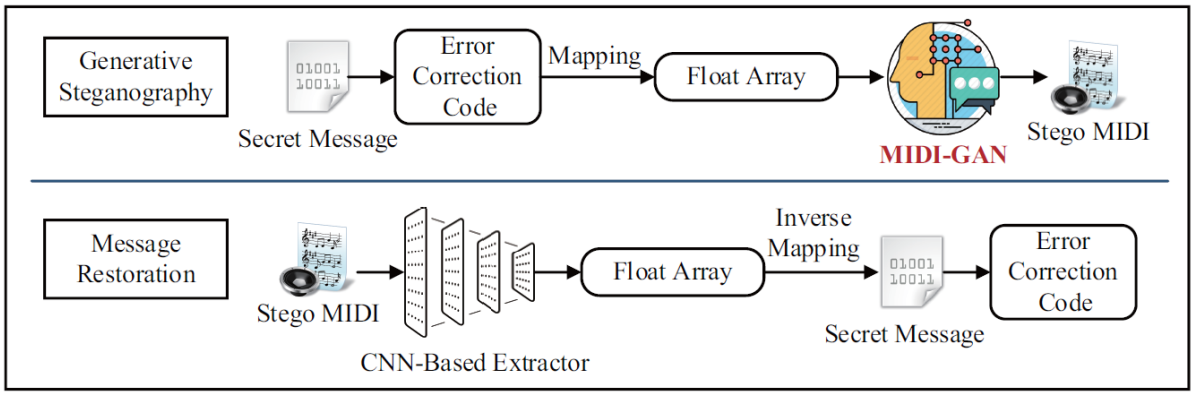

Figure 2. Block schema of MIDI-GAN audio steganography

The secret message is first encoded using error correction code (ECC), which is then transformed into a float array. This float array plays a key role in the steganographic process, as it is used to generate the stego MIDI file. This step is carried out using MIDI-GAN in what is referred to as the message-driven generative steganography phase. After the stego MIDI file is created, it is processed by a convolutional neural network (CNN), an essential component of the MIDI-GAN architecture, to extract the hidden message. At this stage, a float array is derived from either the stego MIDI file or the chord sequence. The secret message is then retrieved through an inverse mapping process, with error correction applied by ECC.

Table 1. Summary of Literature Review

Technique	Strengths	Weaknesses
Echo Hiding (9)	High capacity, robustness	Low secrecy, may degrade audio quality
LSB (8)	Simple, high capacity	Sensitive to noise and compression
Phase Coding (10)	High secrecy	Low capacity, vulnerable to compression
Spread Spectrum (11)	Robustness, hard to detect	Low capacity, complex processing
Stereo Audio Steganography (13)	Compatible with HAS	High computational cost
MIDI-GAN (14)	High undetectability, even against deep learning models	Low capacity, Sensitive to parameter choices

Table 1 indicates that there is a trade-off between capacity, imperceptibility and robustness. So, the classical methods are preferred for their intended use.

A review of the studies cited in this section, along with their references, highlights the significant need for new steganographic approaches that integrate music theory and music technology. As a result, future research is likely to focus on methods in the spatial and frequency domains, as well as techniques involving wavelet transforms or convolutional neural networks (CNNs), all designed with the human auditory system in mind.

3 The proposed mid/side steganography

The presented work uses the manipulation of the side component of a stereo audio file. The following formula (1) can be utilized to derive the mid and side components of a stereo signal:

$$M = \frac{1}{2}(L + R)$$
$$S = \frac{1}{2}(L - R) \quad (1)$$

where M, S, L, and R represent mid, side, left channel, and right channel respectively. To create a stereo audio file from the mid and side channels, the calculation of the left and right channels is provided in formula (2):

$$L = (M + S)$$
$$R = (M - S) \quad (2)$$

In a stereo audio mix, when a target signal is panned to the center, the mid-side representation offers a straightforward approach to source separation or enhancement. The mid signal strengthens the target signal, along with other sounds that are in-phase or nearly in-phase, while diminishing the presence of other elements. Conversely, the side signal entirely eliminates the center-panned target signal and may amplify components that are out of phase [15].

The side component has a lower amplitude and is inherently more difficult to detect, both due to its mathematical formulation and its characteristics [20].

As shown in Figure 3, the data-hiding process in the proposed method begins by separating the mid and side components of the stereo cover. Next, the image to be hidden is reshaped into an array and embedded into the side component by reducing its amplitude. The modified side component, along with the mid component, is then used to reconstruct the stereo sound, forming the stego-audio.

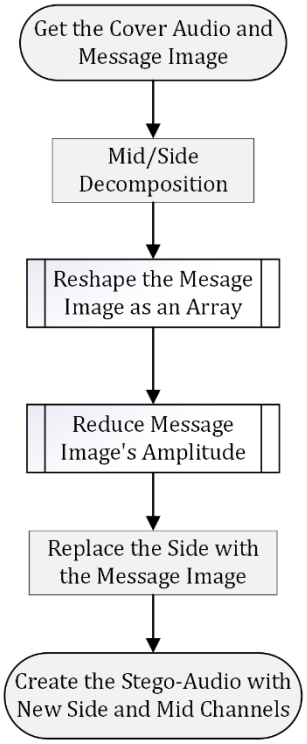

Figure 3. Data hiding phase of the proposed mid/side steganography [20]

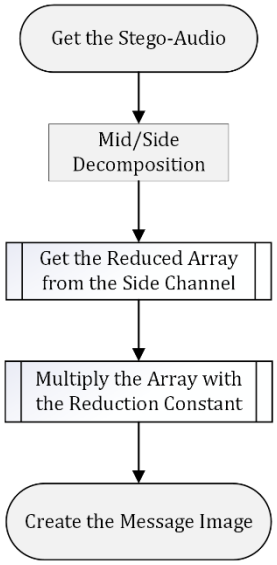

Figure 4. Data extraction phase of the proposed mid/side steganography [20]

To accurately recover the hidden message, the recipient must have knowledge of both the image dimensions and the amplitude reduction coefficient. As depicted in Figure 4, the mid and side

components of the stego-audio are calculated to facilitate data extraction. The message embedded in the side component is then retrieved, allowing the reconstruction of the hidden image.

4 Experimental results

In the experimental analysis, different payload values were tested using the ViSQOL and STOI metrics. ViSQOL (Virtual Speech Quality Objective Listener) is a signal-based, fully referenced metric that applies a spectro-temporal comparison between a reference and a test speech signal to estimate how humans perceive speech quality [16]. The Short-Time Objective Intelligibility (STOI) metric, on the other hand, is used to predict speech intelligibility by measuring the similarity between clean and processed speech signals, particularly focusing on those processed with time-frequency (TF) methods like noise reduction or speech separation [17].

For the tests, two different audio files containing both music and speech were used as cover audio for ViSQOL. These cover audio files were subjected to data embedding, with payloads ranging from 10 kilobytes to 500 kilobytes. The metrics applied in the results include NSIM (Neurogram Similarity Index Measure) and MOS (mean-opinion score) values derived from PESQ (Perceptual Evaluation of Speech Quality) [18, 19]. Figure 5 presents the results graph for music used as the cover medium.

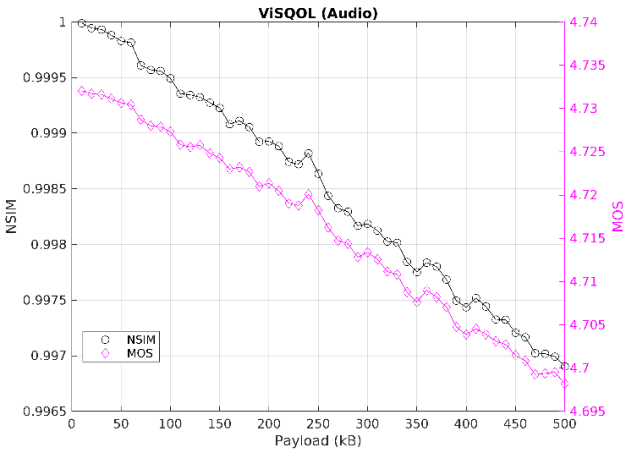

Figure 5. ViSQOL results for music as a cover with payload in 10-500 kb.

A further experiment was conducted in which the cover audio was speech, as illustrated in Figure 6.

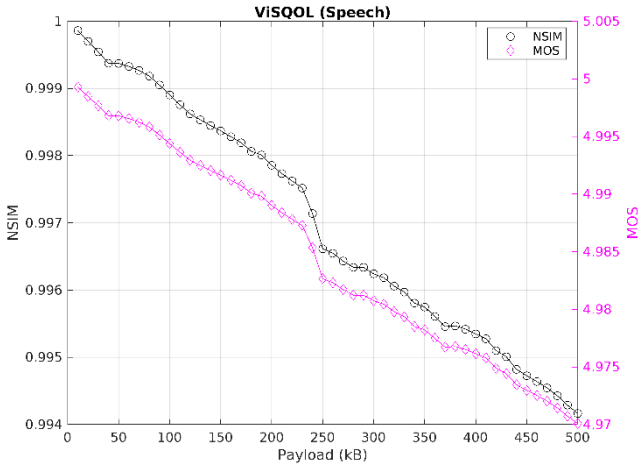

Figure 6. ViSQOL results for speech as a cover with payload in 10-500 kb.

Since STOI is specifically intended for the analysis of speech sounds, the test was conducted only when the cover audio was speech. The results, shown in Figure 7, display the similarity between the stego-audio and the original cover audio, with values ranging from 0 to 1.

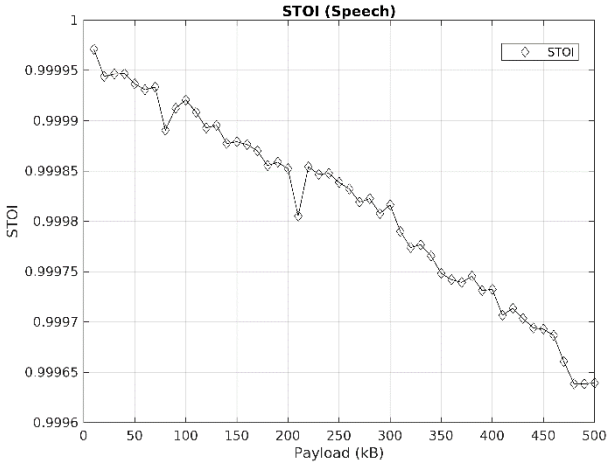

Figure 7. STOI results for speech as a cover with payload in 10-500 kb.

In a further experiment, music and speech were employed as a cover, with 500kb of data hidden in both. Upon examination of the resulting waveform, a noticeable change in the amplitude of the data hidden sections was observed, shown in Figure 8. Consequently, future studies should incorporate normalization algorithms into the studies.

In another analysis, as shown in Figure 9, no notable difference was identified when the spectrograms were compared. In other words, the presented

study did not result in a change to the frequency space.

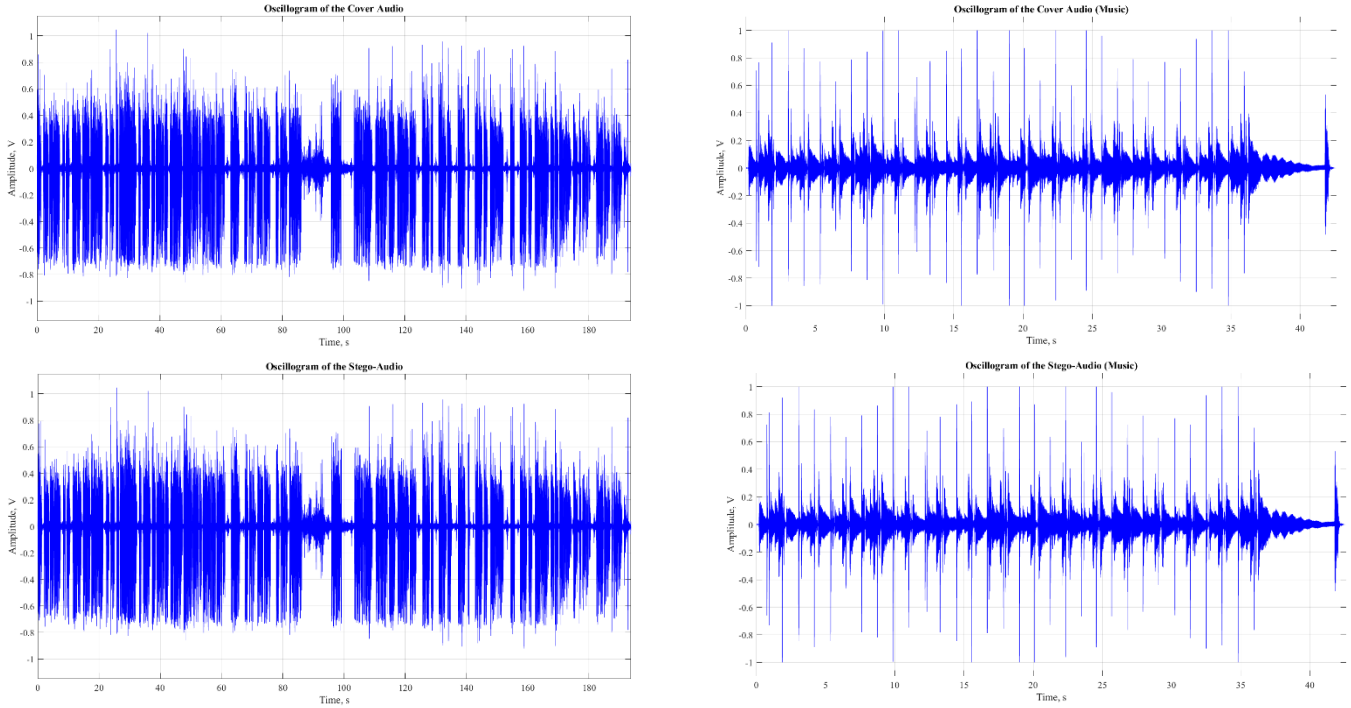

Figure 8. Waveform of cover audios and stego audios for both speech and music

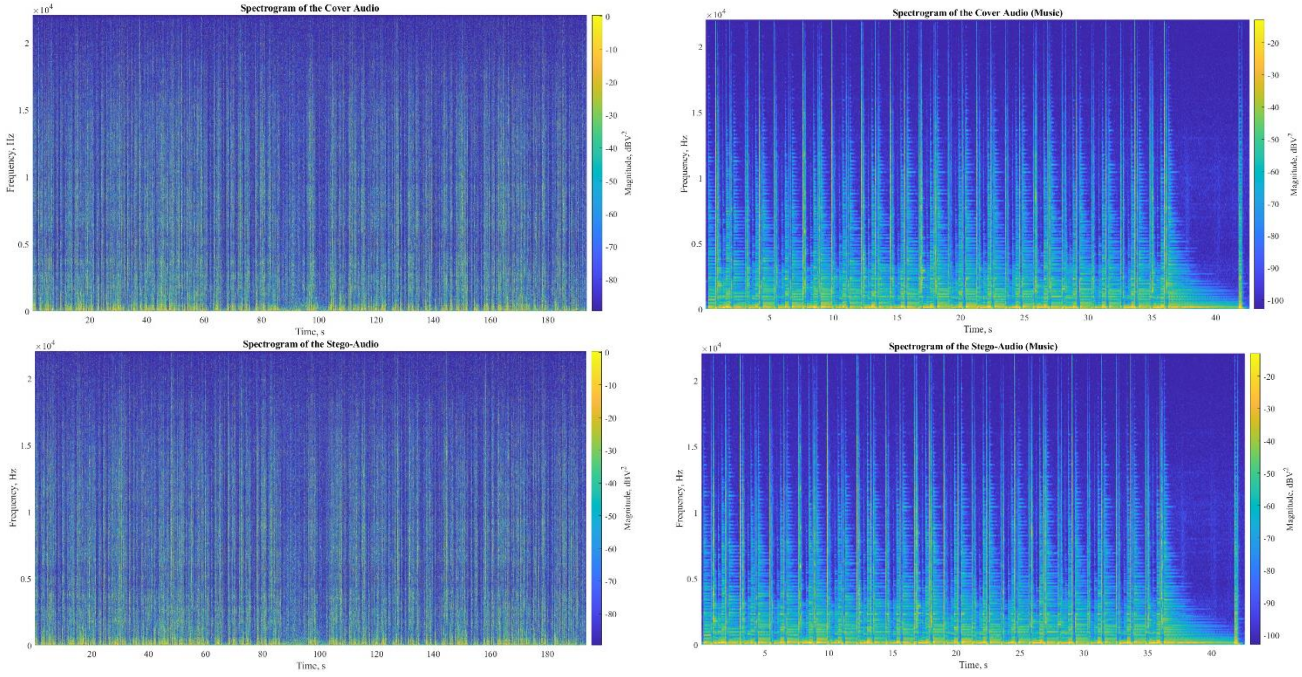

Figure 9. Spectrogram of cover audios and stego audios for both speech and music

To evaluate the robustness of the proposed method, a compression test was performed, as this is one of the most frequent types of stego-attacks. The stego-audio was compressed using the MP3 algorithm, and the results showed that the receiver could successfully extract the message in a meaningful way. Figure 10 displays the message that the receiver recovered after the stego-audio was compressed.

Figure 10. The message that the receiver can extract after a Stego-Attack.

The experimental findings indicate that the proposed steganography algorithm supports high-capacity data embedding and demonstrates robustness against compression. A significant accomplishment is its ability to extract the hidden message clearly, even in the face of stego-attacks.

5 Conclusion

This paper introduces an innovative approach to audio steganography by incorporating principles from music theory and leveraging music production technologies, while also taking into account the intricacies of the human auditory system. The proposed algorithm stands out due to its simplicity and resilience, proving to be robust even under compression—one of the most common stego-attacks. The experimental results reveal that embedding data within the side component of the audio leads to no perceptible changes, even when the entire side component is utilized for data hiding.

A comparison of the presented method with those in the literature indicates that, in the echo hiding technique, the audio distortion is particularly noticeable as the hidden data increases. In the LSB technique, it is not robust against basic stego-attacks such as compression and noise addition. In the phase coding technique, both the information-carrying capacity of the phases is limited and can be detected in case of heavy hiding. In the spread spectrum technique, the data extraction algorithm is generally complex due to the wideband usage.

The presented method provides a fresh perspective on audio steganography, offering an effective solution for covert communication. By using a music-inspired approach, it opens up new possibilities for steganography that align more closely with natural auditory perception, making it not only secure but also unobtrusive in its execution.

Looking forward, this work sets a foundation for the development of future algorithms that dive deeper into music theory while avoiding computationally intensive methods such as convolutional neural networks (CNNs) or generative adversarial networks (GANs). By streamlining the process and focusing on human perception and auditory characteristics, future research could yield more efficient, lightweight solutions that maintain both high capacity and security in data embedding, making steganography even more practical and adaptable to real-world applications.

Acknowledgement

This paper is an extended and updated version of "Audio Steganography with Stereo Audio by Using Mid/Side Processing" presented at International Conference on Cyber Security and Digital Forensics (ICONSEC'24) on September 4th, 2024.

References

[1] G. Kale, A. Joshi, I. Shukla, and A. Bhosale, "A Video Steganography Approach with Randomization Algorithm Using Image and Audio Steganography," in 2024 International Conference on Emerging Smart Computing and Informatics, ESCI 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ESCI59607.2024.10497225.

[2] Q. Li, B. Ma, X. Wang, C. Wang, and S. Gao, "Image Steganography in Color Conversion," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 71, no. 1, pp. 106–110, Jan. 2024, doi: 10.1109/TCSII.2023.3300330.

[3] S. He, D. Xu, L. Yang, H. Dai, and S. Wang, "An anti-steganalysis adaptive steganography for HEVC video based on PU partition modes," J Vis Commun Image

Represent, vol. 98, Feb. 2024, doi: 10.1016/j.jvcir.2023.103995.

[4] F. ASLANTAŞ and C. HANİLÇİ, "Comparative Analysis Of Audio Steganography Methods," Journal of Innovative Science and Engineering (JISE), Jan. 2022, doi: 10.38088/jise.932549.

[5] A. E. Altinbaç and Y. Yalman, "Bit Reduction based Audio Steganography Algorithm," in Proceedings - 6th International Conference on Computer Science and Engineering, UBMK 2021, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 703–706. doi: 10.1109/UBMK52708.2021.9558943.

[6] X. Zhang, C. Li, and L. Tian, "Advanced audio coding steganography algorithm with distortion minimization model based on audio beat," Computers and Electrical Engineering, vol. 106, Mar. 2023, doi: 10.1016/j.compeleceng.2023.108580.

[7] R. Bona, D. Fantini, G. Presti, M. Tiraboschi, J. I. Engel Alonso-Martinez, and F. Avanzini, "Automatic Parameters Tuning of Late Reverberation Algorithms for Audio Augmented Reality," in ACM International Conference Proceeding Series, Association for Computing Machinery, Sep. 2022, pp. 36–43. doi: 10.1145/3561212.3561236.

[8] J. Jezdimirovic, N. Pekez, and J. Kovacevic, "Security enhancement of LSB-based audio steganography method," in 2023 IEEE Zooming Innovation in Consumer Technologies Conference, ZINC 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 77–82. doi: 10.1109/ZINC58345.2023.10174020.

[9] H. Rafiee and M. Fakhredanesh, "Presenting a Method for Improving Echo Hiding," Journal of Computer and Knowledge Engineering, vol. 2, no. 1, 2019, doi: 10.22067/cke.v2i1.74388.

[10] M. S. Yadnya, B. Kanata, and M. K. Anwar, "Using Phase Coding Method for Audio Steganography with the Stream Cipher Encrypt Technique," in Proceedings of the First Mandalika International Multi-Conference on Science and Engineering 2022, MIMSE 2022 (Informatics and Computer Science), Atlantis Press International BV, 2022, pp. 66–75. doi: 10.2991/978-94-6463-084-8_8.

[11] A. Kuznetsov, A. Onikiychuk, O. Peshkova, T. Gancarczyk, K. Warwas, and R. Ziubina, "Direct Spread Spectrum Technology for Data Hiding in Audio," Sensors, vol. 22, no. 9, 2022, doi: 10.3390/s22093115.

[12] J. Li, K. Wang, and X. Jia, "A Coverless Audio Steganography Based on Generative Adversarial Networks," Electronics (Switzerland), vol. 12, no. 5, Mar. 2023, doi: 10.3390/electronics12051253.

[13] H. J. Shiu, B. S. Lin, B. S. Lin, W. C. Lai, C. H. Huang, and C. L. Lei, "A Stereo Audio Steganography by Inserting Low-Frequency and Octave Equivalent Pure Tones," in Advances in Intelligent Systems and Computing, Springer Verlag, 2018, pp. 244–253. doi: 10.1007/978-3-319-68527-4_27.

[14] Z. Su, G. Zhang, Z. Shi, D. Hu, and W. Zhang, "Message-Driven Generative Music Steganography Using MIDI-GAN," IEEE Trans Dependable Secure Comput, 2024, doi: 10.1109/TDSC.2024.3372139.

[15] A. S. Master, L. Lu, and N. Swedlow, "Stereo Speech Enhancement Using Custom Mid-Side Signals and Monaural Processing," AES: Journal of the Audio Engineering Society, vol. 71, no. 7–8, pp. 431–440, 2023, doi: 10.17743/jaes.2022.0070.

[16] A. Hines, J. Skoglund, A. Kokaram, and N. Harte, "ViSQOL: The Virtual Speech Quality Objective Listener," in IWAENC 2012; International Workshop on Acoustic Signal Enhancement, 2012, pp. 1–4.

[17] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 2010, pp. 4214–4217. doi: 10.1109/ICASSP.2010.5495701.

[18] A. Hines and N. Harte, "Speech intelligibility prediction using a Neurogram Similarity Index Measure," Speech Commun, vol. 54, no. 2, pp. 306–320, 2012, doi: https://doi.org/10.1016/j.specom.2011.09.004.

[19] ITU, "Mapping function for transforming P.862 raw result scores to MOS-LQO," Int. Telecomm. Union, Geneva, Switzerland, ITU-T Rec. P.862.1, 2003.

[20] A. E. Altinbas, Y. Yalman, "Audio Steganography with Stereo Audio by Using Mid/Side Processing," in ICONSEC 2024; International Conference on Cyber Security and Digital Forensics, 2024.