Received: April 8, 2024. Revised: May 4, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Data Hiding Approach to Enhance the Audio Sample Spaces Using Interpolation and Multi-layering

Daffa Tristan Firdaus[1] Ntivuguruzwa Jean De La Croix[1,2] Tohari Ahmad[1]*
Pascal Maniriho[3] Ary Mazharuddin Shiddiqi[1] Diana Purwitasari[1]

[1]*Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia*
[2]*African Center of Excellence in Internet of Things, University of Rwanda, Kigali, 3900, Rwanda*
[3]*Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia*
* Corresponding author's Email: tohari@its.ac.id

Abstract: In the face of rapid advancements in information transmission technology, ensuring the secure development of all application aspects is imperative, particularly considering the frequent data transmission over public networks. A significant vulnerability exists for data compromise during transmission without robust security measures. Thus, implementing a reliable data-hiding scheme becomes indispensable to safeguarding confidential information. Despite the introduction of various methods to address this concern, challenges persist in effectively managing data size and maintaining the resulting data quality. This research addresses these challenges with the key idea of using a combination of dynamic layering and sample interpolation to increase the samples' capacity in accommodating the secret bits. The novelty of this study is based on the dynamic use of the multi-layering paradigm during the data embedding, which contributes to maintaining the quality of the audio sample considered for concealment. The experimental results show a significant outperformance of the proposed method, which gets a maximum peak signal-to-noise ratio (PSNR) of 120.39 decibels (dB), an ideal PSNR to secure secret data transmission.

Keywords: Audio steganography, Data hiding, Information security, Interpolation, Network infrastructure.

1. Introduction

Information technology has significantly advanced in recent years, underscoring its profound impact on society across various domains. As a ubiquitous manifestation of this progress, Telecommunications employs technologies for transmitting voice, data, or multimedia, facilitating the conveyance of messages from senders to receivers. However, this mode of communication is susceptible to interception by malevolent individuals or groups who may manipulate or, in extreme cases, obliterate the conveyed messages [1].It is imperative to implement safety measures to prevent such potential tragedies. Steganography emerges as a strategic approach to evade these threats, fortifying security measures and safeguarding sensitive data [2, 3].

Steganography involves concealing secret data within a cover media file, rendering the secret message remarkable through the naked ear. The key concepts of steganography in audio covers are illustrated in Fig. 1, where the original audio, cover audio, and payload represent the input of the embedding process. The embedding process's output, the extraction process's input, is stego audio, which finally outputs the cover audio and secret message. For any steganographic method, suspicions may arise if the resulting stego audio significantly differs from the cover audio. In the general steganography paradigm, challenges arise when a cover medium cannot accommodate large amounts of data for embedding secret messages [4, 5]. This highlights that steganography is far from perfect, presenting various methods for implementation or improvement.

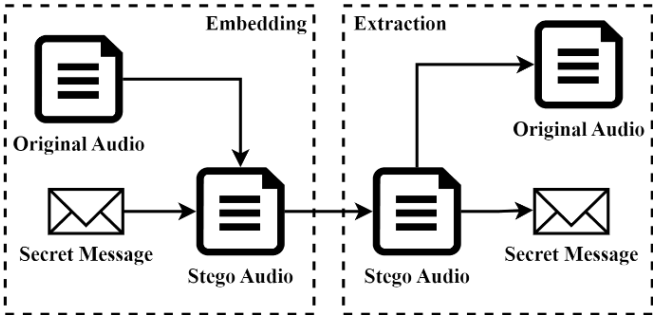

Figure. 1 The concept of audio steganography paradigm

Aligning with the steganography in audio samples, the key considerations in enhancing steganographic methods include assessing the capacity of a cover to hold secret data and evaluating the stego-audio quality [6]. Numerous methods have emerged to enhance steganography, particularly in the context of audio, aiming to address prevalent challenges [6-9]. However, to improve one aspect, navigating potential drawbacks in another becomes crucial, resulting in a delicate optimisation balance. This makes most prevalent schemes vulnerable to steganalysis attacks [10, 11]. Steganalysis to locate the secret data is known as locative steganalysis [12], [13], and a steganalysis scheme to detect the presence of the hidden data is referred to as blind steganalysis [14, 15]. Various audio steganographic techniques, including reduced difference expansion [7], interpolation [16], and multi-layering [17], have been explored to surmount these challenges. Reduced difference expansion consists of reducing the embedding capacity while simultaneously increasing the audibility quality of the audio sample. In this approach, the differences between the neighbouring samples are extended to enhance the quality of the stego audio. Conversely, the interpolation method expands the sample space, doubling the size of the original sample and leading to a distinctly altered stego-audio. Moreover, the multi-layering approach consists of augmenting the block size in the audio samples, guiding the iteration over the bits intended for embedding.

While prior research has demonstrated performance in audio steganography, utilising audio files as covers, the existing methods still exhibit room for improvement. This is evident in the observed comparison between the cover and the stego audio, which reveals residual noise, particularly reflected in the attained peak signal-to-noise ratio (PSNR). Current approaches in audio steganography reveal a trade-off between increasing sample space and compromising the quality of stego-audio, leading to a diminished PSNR value. However, when it is necessary to embed a larger payload, minimising the decline in PSNR becomes imperative. Consequently, this study aims to sustain a relatively high PSNR value while augmenting the capacity of a bit in a sample space by integrating multiple methods. The contribution of this work is highlighted as follows:

- Designing an audio steganography scheme based on interpolation and dynamic multi-layering.
- Using interpolation to increase the sample spaces, allowing for the embedding of secret messages without significantly affecting the quality of the original audio content.
- Using dynamic multi-layering to enhance the embedding capacity through an iterative fashion while minimising the impact on the perceptual quality of the stego-audio.

The next parts of this paper are structured as follows: Section 2 reviews prior works in audio steganography. Section 3 provides a detailed explanation of the proposed method. Experimental results are presented in Section 4. The paper concludes with Section 5, summarising the key findings.

2. Related works

In the study conducted by the authors in [16], a steganography method was proposed to minimise the sample space, aiming to enhance the quality of stego-audio. Addressing the trade-off between secret data size and stego-audio quality, the authors employed audio as a carrier for secret data transmission, focusing on analysing the audio sample space hosting confidential information. The proposed method aimed to optimise stego-audio quality by reducing the capacity of the audio sample space, resulting in a significant improvement of 31.3 dB in PSNR over benchmark methods. However, a notable area for improvement is identified, suggesting the need for more optimised sample spaces capable of accommodating larger amounts of secret data.

Subsequently, a steganographic scheme presented in [18] introduced a segmentation embedding method using modified interpolation to increase the capacity for adaptable and reversible audio data hiding. By modifying linear interpolation values, the proposed method offers flexibility in embedding capacity while maintaining high audio quality for substantial amounts of secret data. Experimental results demonstrated an average PSNR of 100.55 dB on 0.75 and 87.68 dB on 2.267 bits per audio data sample. However, the study identified areas for improvement, specifically in fixing dynamic distributions to enhance stego audio quality and addressing limitations in maintaining reversibility.

Furthermore, the study by the authors in [7] proposed a method for hiding messages using audio files, employing modulus operation and simple partition. The method aimed to enhance stego audio quality and increase embedding capacity by utilising a simple partition and an adaptive modulus operation calculated method. While successfully embedding relatively large payloads at a 700 kb file size, the research acknowledged low PSNR results for large-sized payloads. This suggests room for improvement to achieve higher PSNR results while expanding the capacity of the secret message.

In another approach, the authors in [9] proposed enhancing the least significant bit (LSB) method using binaries of message-sized encoding (BSME) to achieve transparent, secure, and high-capacity audio steganography. Experimental results demonstrated that LSB-BMSE outperforms existing methods in hiding capacity and imperceptibility, with an average SNR value of 99.98 dB. However, limitations were identified, particularly in its vulnerability to noise and LSB attacks.

The authors in [8] also proposed enhancing the LSB method using a piecewise linear chaotic map and a one-time pad for imperceptibility, security, and high capacity. The method utilised Huffman coding to reduce payload size and employed random numbers generated with a piecewise linear chaotic map to enhance one-time pad security. Experimental results indicated a PSNR range of 80.4408 dB to 90.6455 dB and successful inflation of payload up to 173%. Despite its resilience to re-sampling attacks, the method demonstrated vulnerability to LSB and AWGN attacks.

Lastly, the authors in [6] introduced a method for hiding payloads in multiple WAV audio cover files based on a circular secret key. This method distributed payloads randomly across cover audio files using agreed-upon keys. Experimental results demonstrated the effectiveness of hiding secret data using a secret key. However, the study suggested future improvements, such as exploring the possibility of merging different types of audio files or even images.

Based on a comprehensive examination of the previously discussed existing works in the realm of audio steganography, the focus of this study now shifts towards presenting the proposed methodology entitled using a combination of interpolation and multi-layering to enhance the audio sample spaces. This innovative approach strategically integrates the intricacies of applying interpolation and multi-layering techniques specifically tailored to optimise audio sample spaces. The primary objective is to augment the capacity and quality of concealed data within audio files by synthesising insights from preceding studies.

3. Proposed method

This proposed work will delve into audio steganography using interpolation with added multi-layering. The interpolation method generally uses a single layering technique where the embedding process is done once for each interpolated sample. However, by adding a multi-layering technique, a single interpolated sample can be embedded multiple times; hence, it can make an audio cover hold significantly larger secret data. The multi-layering process itself is a method that embeds secret data onto an audio cover a specific number of times across all samples. In this case, the interpolated method can alter that static number into a dynamic variable, which can differ for all interpolated samples.

3.1 Data embedding

The steps of the embedding process, which are illustrated in Fig. 2, are explicitly detailed as follows:

1. The Proposed Method starts by taking a cover audio file and sampling it to a 16-bit signed integer. Then, the sample value is converted to positive values for all sample spaces. The results of this normalisation process consist of values varying between 0 to 65535.
2. Considering I_i as the resulting interpolation sample at odd and even i indexes, n as the index of the original sample, and x as a variable that continues to increase for every even sample run, the sample spaces resulted from Step 1 are then interpolated using Eq. (1) and the total sample after the interpolation process are obtained using Eq. (2).

$$I_i = \begin{cases} \left\lfloor \frac{S_{\left(\frac{i}{2}\right)} + S_{(i-x)}}{2} \right\rfloor & if\ i\ is\ even \\ S_{\left(\frac{i+1}{2}\right)} & if\ i\ is\ odd \end{cases} \quad (1)$$

$$\max(i) = (\max(n) \times 2) - 1 \quad (2)$$

3. Before the embedding process, the distance of the audio signal between an interpolated sample and a normalised sample. Taking C as a variable of distance between the interpolated sample and the normalised sample in the audio signal, t as the duration of the audio, and i as the index of the interpolated sample, this value can be obtained from Eq. (3).

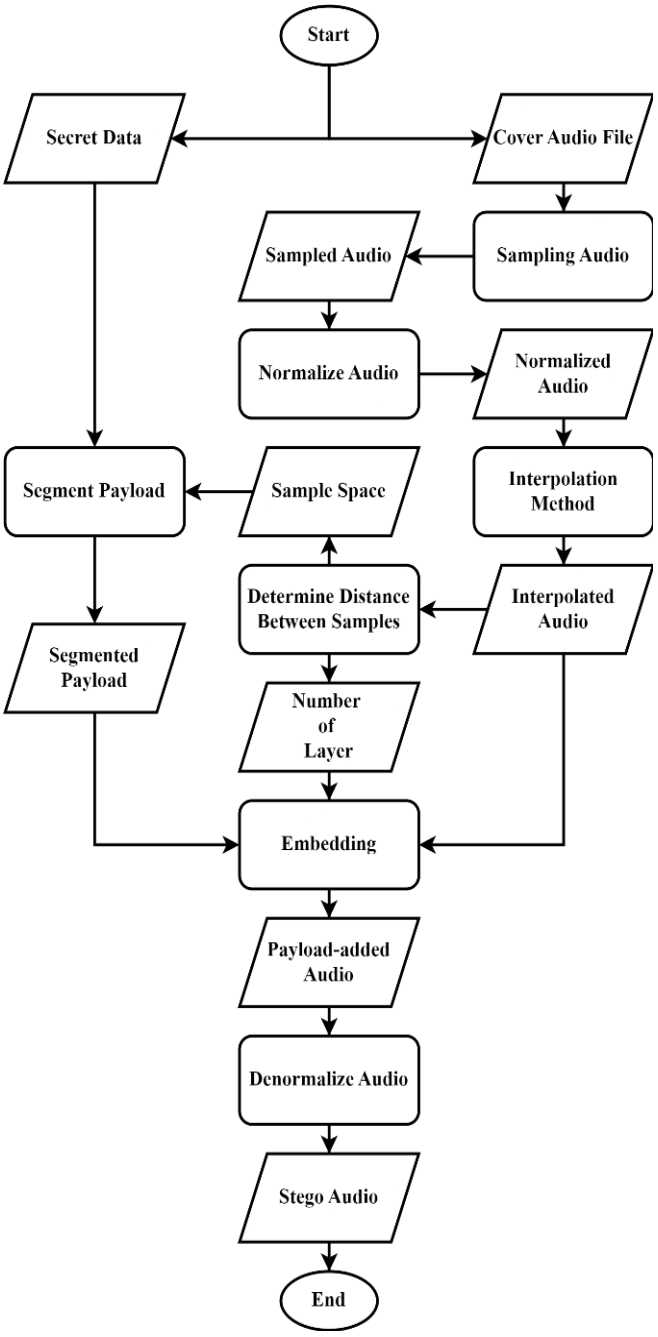

Figure. 2 Flowchart of data embedding process

$$C = \frac{t}{\max(i) \times 2} \tag{3}$$

4. Based on the value obtained for C in Step 3, the Proposed Method proceeds with Eq. (4) to obtain the sample space and the number of layering. It is important to note that the notation d_i indicates the distance between the original audio samples, x_i and y_i indicates the values at the i-index audio sample on the x and y-axis. Considering the notation N_i as the sample space and the number of layers to be processed, the sample space and the number of layering are acquired using Eq. (5).

$$d_i = \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \tag{4}$$

$$N_i = \lfloor \log_2(C \times d_i) \rfloor \tag{5}$$

5. Before the payload concealment, the bits of the secret data are segmented based on the value of N_i. To embed the bits of the secret data, there is an iteration of N_i-times within the segmented secret data using Eq. (6). Then, a key is created depending on the same equation in (6). Considering I'_i as the interpolated sample on even i after the embedding process, b as the segmented payload.

$$I'_i = \begin{cases} I_i + 1, & key = 11 \;\; if \; N_i \neq 0, b > 0 \\ I_i - 1, & key = 10 \;\; if \; N_i \neq 0, b = 0 \\ I_i, & key = 00 \;\; if \;\; N_i = 0 \end{cases} \tag{6}$$

6. After successfully concealing all the payloads into the interpolated samples, the obtained samples are then denormalised to result in values between -32768 and 32767. The stego audio is obtained from these newly obtained values.

3.2 Data and cover extraction

To extract the secret data and the original cover audio, the Proposed Method follows the steps illustrated in Fig. 3. We now explicitly clarify the steps taken in the extraction process.

1. The extraction process starts by sampling and normalising the stego audio to make all sample values positive, just like the first step of the embedding process.
2. The sample spaces acquired from Step 1 are then divided into two parts, non-changed and changed samples (non-changed to mean original samples and changed to mean the samples altered by the proposed steganographic algorithm), based on the values from indices. If the index is odd, the sample will be counted for the original sample and vice versa for the even index. The sample division follows Eqs. (7) and (8) consider S and S' as the non-changed and changed samples, respectively.

$$S_{\left(\frac{i+1}{2}\right)} = I'_i \; if \; i \; is \; odd \tag{6}$$

$$S'_{i/2} = I'_i \; if \; i \; is \; even \tag{7}$$

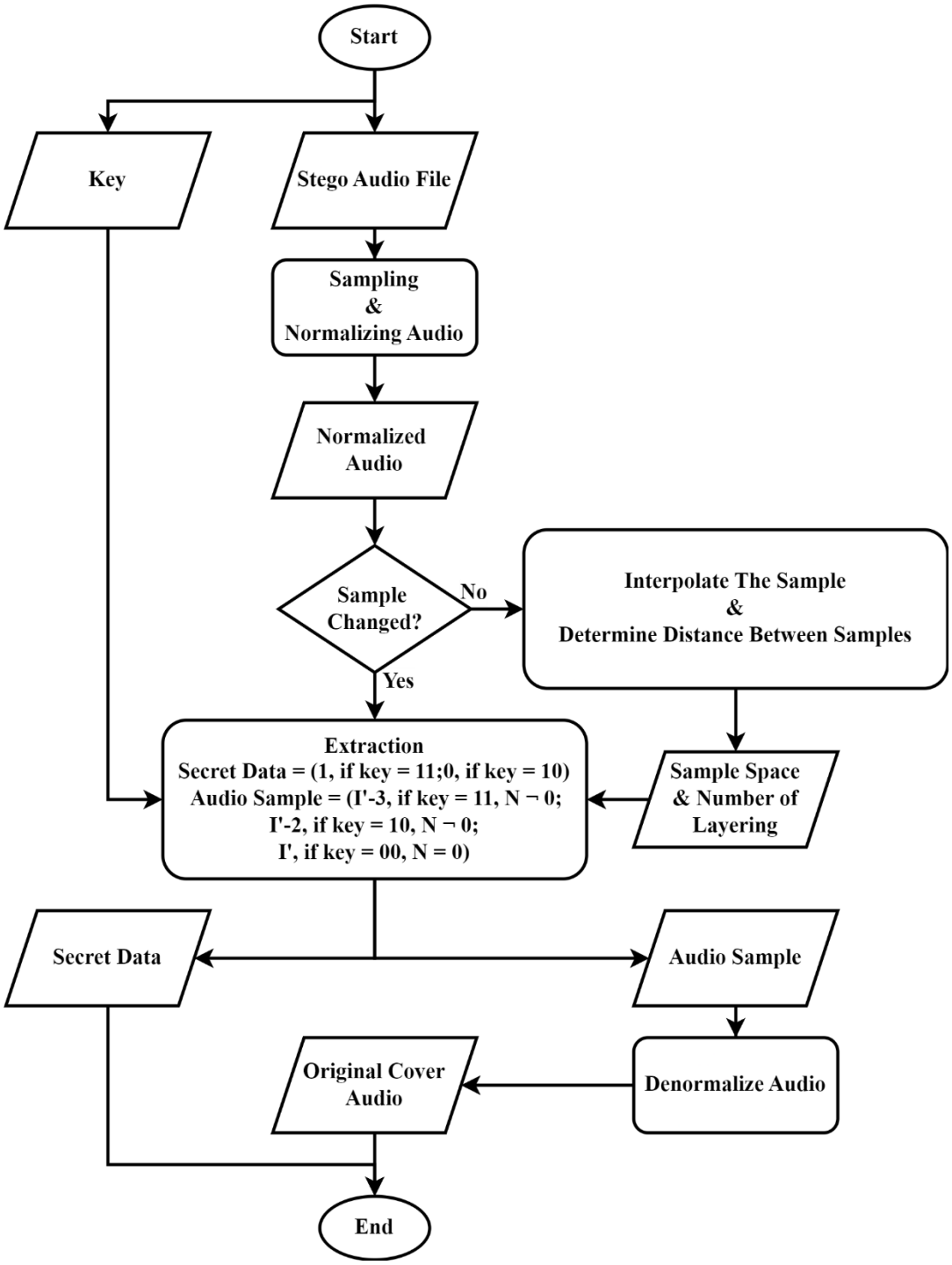

Figure. 3 Flowchart of data extraction process

3. Following Eq. (1) used in the embedding process, the samples are interpolated and based on Eqs. (3)-(5) from the embedding process, C is calculated to obtain the sample space.
4. To get the bits of the secret data, Eq. (8) is used based on the key created in Step 5 of the embedding process, considering P as the bit of the secret data. All these operations are performed on the samples obtained from Step 3.

$$P = \begin{cases} 1, if\ key\ =\ 11 \\ 0, if\ key\ =\ 10 \end{cases} \quad (8)$$

5. Considering S_i as the sample of the original audio cover, I'_i as the changed sample (at the even indices), the samples making the original audio are obtained using Eq. (9).

$$S_i = \begin{cases} I'_i - 1 & if\ key = 11, N_i \neq 0 \\ I'_i + 1 & if\ key = 10, N_i \neq 0 \\ I'_i & if\ key = 00,\ N_i = 0 \end{cases} \quad (9)$$

6. After obtaining the original audio samples from Step 5, the original audio samples are denormalised to get the original sample before normalisation.

4. Results and discussion

This section, dedicated to presenting the outcomes of the proposed methodology, is structured into four coherent subsections. The initial subsection encompasses details of the experimental setup and evaluation metrics, providing insight into the experimental procedures and the criteria employed for results assessment. Subsequently, the second subsection delves into an in-depth analysis and discussion of the acquired results, presented through tables and graphs. The third subsection entails the outcomes of an ablation study, wherein the effectiveness of the proposed method is scrutinised with and without the newly introduced components of the model. The final subsection involves a comparative analysis of the results, drawing distinctions between the proposed method and state-of-the-art works in the field.

4.1 Experimental setup and evaluation metrics

In the experiment, a set of 15 audio files was selected from the IRMAS (instrument recognition in musical audio signals) dataset [23] to serve as cover data files, each with a precisely defined sample size of 132,299. To provide a reference guide for understanding the instrumentation and genres associated with the cover audio files, Table 1, referring to the works in [16], presents information on 15 cover audio files used in the experimentation of this work. As mentioned earlier, audio files are stored in .wav format due to their originality with no compression, ensuring that no information is lost even when other data is embedded within them [24]. The Audio files are labelled "Audio 1" through "Audio 15," corresponding to the respective audio genre. The "Instrument" represents the primary instrument utilised in recording each cover audio file, including Cello, Acoustic Guitar, Piano, Saxophone, and Human Singing Voice. The "Genre" categorises each cover into Country-folk, Classical, or Pop-Rock genres, offering insights into the collection's diverse musical styles. Concurrently, an additional set of 11 text files was chosen to act as payload data. These audio files were systematically

Table 1. Cover Audio Files Specifications

Cover	Instrument	Genre
Audio 1	Cello	Country-folk
Audio 2		Classical
Audio 3		Pop-Rock
Audio 4	Acoustic Guitar	Country-folk
Audio 5		Classical
Audio 6		Pop-Rock
Audio 7	Piano	Country-folk
Audio 8		Classical
Audio 9		Pop-Rock
Audio 10	Saxophone	Country-folk
Audio 11		Classical
Audio 12		Pop-Rock
Audio 13	Human Singing Voice	Country-folk
Audio 14		Classical
Audio 15		Pop-Rock

categorised into three genres, as referred to in [22]: country folk, classical, and pop-rock, each corresponding to five distinct instruments. These genres are differentiated by the specific instruments used. Country Folk music typically features cellos, acoustic guitars, pianos, saxophones, and vocals. Similarly, Classical and Pop-Rock music also use these instruments, but with different coding schemes to distinguish them within their respective genres [22]. Furthermore, it is noteworthy that the payload comprises 11 text files of varying sizes, ranging from 1 to 100 kilobits (kb), featuring content in the form of "Lorem Ipsum" text [25].

The evaluation metrics employed in this take reference from the state-of-the-art methods evaluation, wherein mean squared error (MSE) and the PSNR serve as comprehensive benchmarks for a thorough assessment of the proposed audio steganographic algorithm. The comparison of stego-audio quality with the original audio file is facilitated through the PSNR value computed from the MSE as delineated in Eqs. (10) and (11) [6]. MSE quantifies the error between stego-audio and initial audio samples, with N representing the sample count in the audio, St_i denoting a sample from the initial audio and St'_i Signifying a sample from the stego-audio. Concurrently, the PSNR determines the similarity between the initial audio and stego-audio, with the symbol 2^{bs} denoting the highest value of the bit-depth, where bs is 16 bits.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(St_i - St_i)^2 \quad (10)$$

$$PSNR = 10 \times \log_{10} \frac{(2^{bs}-1)^2}{MSE} \quad (11)$$

4.2 **Obtained results**

Fig. 4 illustrates PSNR values across different payload sizes for various audio genres; noteworthy patterns emerge. This comprehensive figure offers critical insights into steganography within audio cover files. Analysing the average PSNR values reveals a consistent trend, depicting a gradual decline from 120.39 dB for 1kb payloads to 100.55 dB for 100kb payloads. This trend elucidates the delicate balance required in steganographic practices, emphasising that smaller payloads exhibit superior

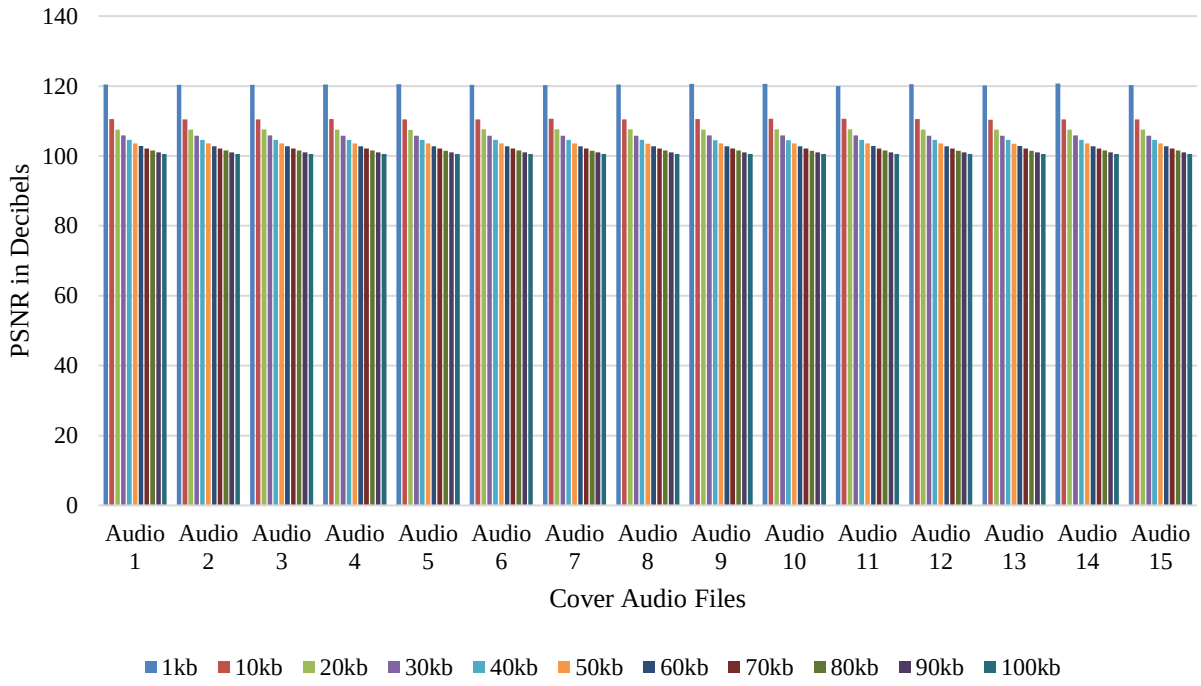

Figure. 4 PSNR obtained across various payload sizes using all tested audio samples

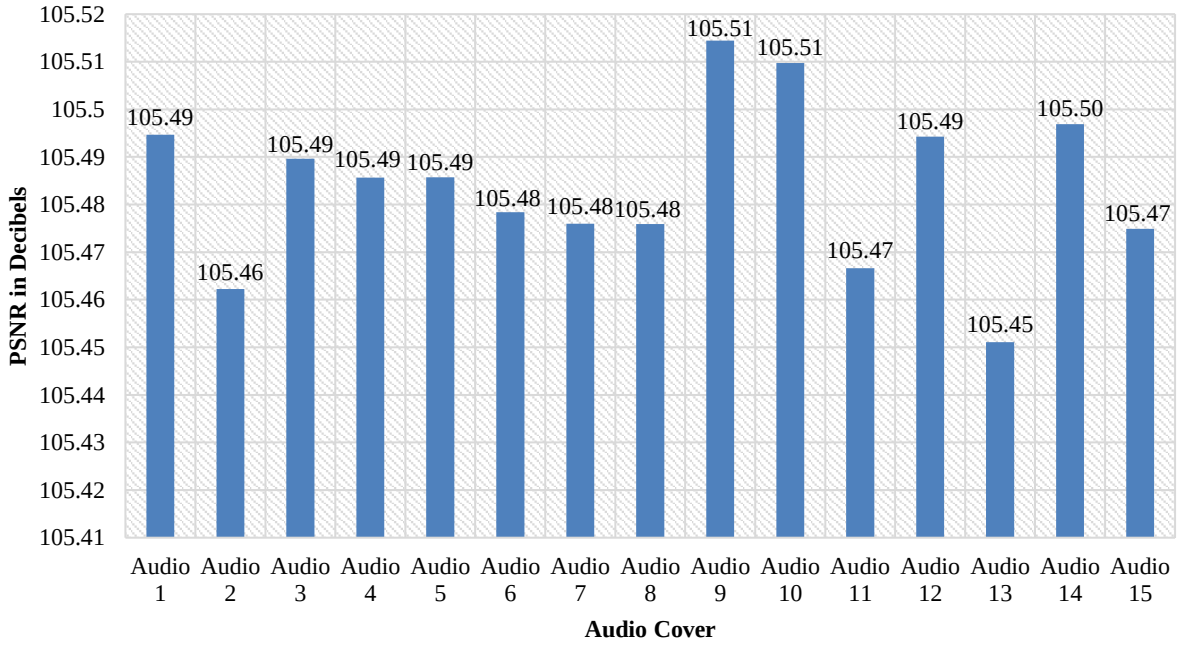

Figure. 5 Average PSNR Values for Audio Cover in Various Payload Sizes

concealment fidelity. In comparison, larger payloads may introduce perceptible distortions of the audio file used as a cover. Based on these results, practitioners in steganography can leverage these results to refine existing techniques or develop novel strategies tailored to the specific requirements of different audio genres and payload sizes. The maximum PSNR value for Audio 14 at 120.70 dB is observed, underscoring the exceptional fidelity achieved with a 1kb payload. Conversely, the minimum PSNR value is recorded for Audio 9 at 100.52 dB with a 100kb payload, indicating a notable drop in quality as the payload size increases. This variation suggests that the effectiveness of the proposed scheme to hide the secret data in audio files may vary across different audio genres, and careful consideration is required to balance file size and perceptual audio quality. A consistent trend is discernible when analysing the average PSNR values. The average PSNR begins at a high of 120.39 dB for 1kb payload sizes and gradually decreases to 100.55 dB for 100kb payloads. This decline highlights the trade-off between file size and audio quality, with larger payloads generally resulting in reduced fidelity. The average PSNR values provide a valuable overview of the overall performance of the proposed scheme across diverse audio genres.

Moreover, Fig. 5 quantitatively assesses the fidelity of audio covers across different payload sizes using the PSNR as the evaluation metric. The figure reveals a range of average PSNR values, from a minimum of 105.45 (Audio 13) to a maximum of 105.51 (Audio 9), suggesting a relatively narrow spread of values. Several genres stand out as high performers in maintaining audio quality after processing or compression. Audio 13, with the highest average PSNR of 105.45 and Audio 10, Audio 12, and Audio 14 demonstrate exceptional fidelity. Conversely, certain genres exhibit lower average PSNR values, indicating potential challenges in maintaining fidelity during secret data concealment. For instance, audio 2, Audio 11, and Audio 13 have comparatively lower average PSNR values, suggesting that these genres may experience more significant distortion or loss. The overall consistency of relatively high average PSNR values across most genres, such as Audio 1, Audio 3, Audio 4, Audio 5, Audio 6, and Audio 7, suggests that the proposed method is effective across diverse musical styles.

4.3 Ablation study results

The data illustrated in Fig. 6 present an ablation study evaluating the effectiveness of a proposed steganographic method applied to various audio genres, comparing its performance with and without the incorporation of multi-layering. The PSNR values, indicative of the quality of concealed data within the audio covers, are measured for two randomly chosen payload sizes (1 kb and 10 kb). Across all audio genres and payload sizes, the proposed method consistently outperforms its counterpart without multi-layering. Notably, in the case of Audio 1 with a 1 kb payload, the proposed method achieves a substantial PSNR of 120.44 dB, while the version

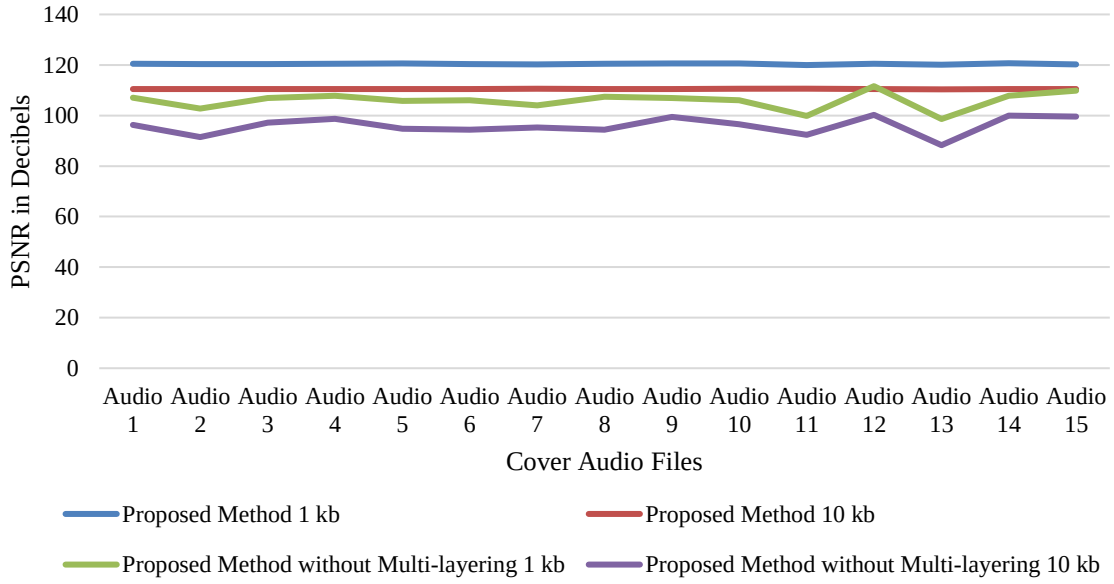

Figure. 6 PSNR yielded through an ablation experiment

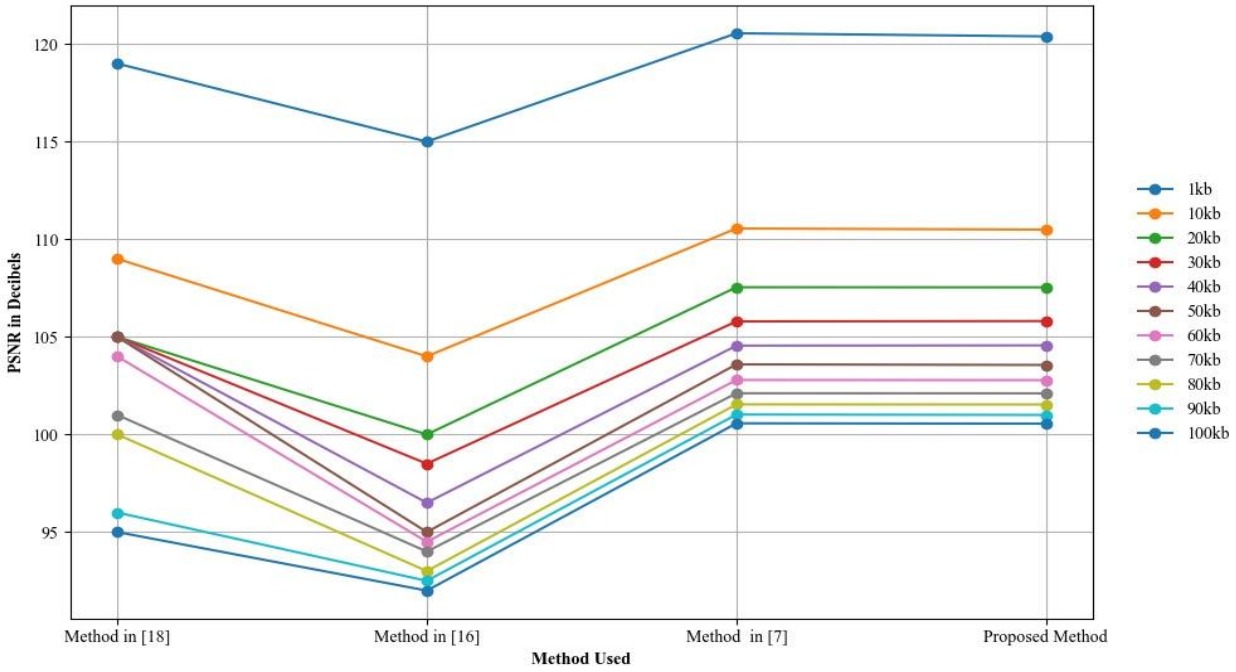

Figure. 7 The average PSNR values comparison between the Proposed Method and the existing methods

lacking multi-layering lags behind at 107.08 dB. This discrepancy underscores the significant positive impact of multi-layering in enhancing the steganographic method's ability to maintain high-quality audio despite the concealed payload. Examining the results for Audio 8 with a 10 kb payload further exemplifies the superiority of the proposed method with multi-layering. Here, the PSNR is 110.43 dB, indicating robust data concealment, whereas the version without multi-layering exhibits a notably lower PSNR of 94.44 dB. These findings emphasise the consistent trend observed across different audio genres and payload sizes, highlighting the pivotal role of multi-layering in achieving superior steganographic performance. These results imply that incorporating multi-layering enhances the steganographic method's resilience to potential attacks and increases the quality of concealed data within audio files.

4.4 Results comparison with the existing works

To highlight the performance of the proposed method in comparison to the existing works considered in this study, Fig. 7 illustrates a comparative portrait of the average PSNR values between the proposed steganographic method and two existing methods introduced [7], [16], and [18] across various payload sizes (1kb to 100kb). For the method in [18], the PSNR values consistently decrease as the payload size increases, ranging from 119 dB for 1kb to 95 dB for 100kb. Similarly, the method in [16] exhibits a declining trend, with PSNR values decreasing from 115 dB for 1kb to 92 dB for 100kb. These decreasing trends suggest that as the payload size grows, the quality of concealed data diminishes for both existing methods, which has been pinpointed in the previous section. The research work in [7] presented promising results in comparison to other related works, but it still showed a sensitive issue of being unable to hide all the secret data. Based on the data reported for the research in [7], from a payload of 800kb, their algorithm has not been able to hide all the secret data.

Comparing the method proposed in this study and the obtained results, it is identified that the proposed method, which uses a combination of sample interpolation and multi-layering, outperforms the previous works referenced in [7], [16], and [18] in terms of the quality of the stego-audio which is justified by the highest PSNR values presented with the proposed method. It is crucial to note that the PSNR results yielded with the proposed method show a slight improvement of the PSNR reported in [7] due to its new feature of using multi-layering, which can hide all the secret bits, which makes it superior and outperforming as compared to others.

Moreover, to emphasise the scientific contribution of the method proposed in this study, it is noteworthy that the obtained results in PSNR are indicators of a promising attempt to address the drawbacks showed by other previous works in [6], [8], and [9]. The PSNR values reported in [8] identified

them to range from 80.44 to 90.64 dB, which is inferior to the one yielded with the proposed method because our PSNR ranges from 100. 55 to 120.39 dB. Comparing the proposed method to the works in [6] and [9], it is also shown that the proposed method availed a promising scheme for steganography in audio covers. An indicator of the outperformance of the proposed method with sheds light on its scientific contribution to the field is based on the best results yielded as compared to the existing works considered for a benchmark.

5. Conclusion

In the realm of rapidly advancing information transmission technology, the imperative of ensuring the secure development of all application facets is undeniable, especially given the pervasive transmission of data over public networks. The absence of robust security measures leaves a significant vulnerability, creating a potential risk for data compromise during transmission. Hence, implementing a dependable data-hiding scheme becomes essential to protect confidential information. This study introduces an innovative strategy that amalgamates interpolation and multi-layering operations, aiming to elevate the security of transmitted information, mainly when dealing with a substantial payload size. The proposed method exhibits remarkable results in enhancing the payload size accommodated in audio files, achieving an overall average PSNR of 105.484 dB. It is crucial to highlight that the obtained results fall within 120.697 dB to 100.524 dB, reflecting an ideal quality benchmark for securing audio files.

For future research endeavours, further refinement of the proposed method could be achieved by incorporating paradigms from other works, such as [19-21], to enhance the quality of stego audio files, especially in scenarios involving large payload sizes.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualisation, DTF, NJDLC, TA, PM, AMS, and DP; methodology, DTF, NJDLC, TA, PM, AMS, and DP; software development, DTF, NJDLC; formal analysis, DTF, NJDLC; original draft writing, and visualisation, DTF, NJDLC; review and editing of the manuscript, TA; supervision, TA; project administration, TA; and acquisition of funding, TA.

Acknowledgments

The authors thank all laboratory and research group members for their invaluable contributions and support. Institut Teknologi Sepuluh Nopember and the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia also supported this research.

References

[1] M. Dalal and M. Juneja, “Steganography and Steganalysis (in digital forensics): a Cybersecurity guide”, *Multimed Tools Appl*, Vol. 80, No. 4, pp. 5723-5771, 2021, doi: 10.1007/s11042-020-09929-9.

[2] H. Arsyad, N. J. De La Croix, and T. Ahmad, “A Steganographic Approach to Secure Data Using Pairs-Based Difference Expansion”, In: *Proc. of 2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN)*, pp. 363-368, 2023, doi: 10.1109/CICN59264.2023.10402286.

[3] M. R. H. Aminy, N. J. De La Croix, and T. Ahmad, “A Reversible Data Hiding Approach in Medical Images Using Difference Expansion”, In: *Proc. of 2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN)*, pp. 358-362, 2023, doi: 10.1109/CICN59264.2023.10402139.

[4] A. W. Chanda D’Layla, M. Nevin, G. G. Sunardi Putra, N. J. de La Croix, and T. Ahmad, “Steganography in Grayscale Images: Improving the Quality of a Stego Image”, In: *Proc. of 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, pp. 1-6, 2023, doi: 10.1109/SMARTGENCON60755.2023.10442310.

[5] Rr. D. A. Anandha, N. J. de La Croix, and T. Ahmad, “A Steganographic Scheme to Protect Medical Data Using Radiological Images”, In: *Proc. of 2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN)*, pp. 369-374, 2023, doi: 10.1109/CICN59264.2023.10402248.

[6] A. Tahseen Suhail and H. Ghanim Ayoub, “A new method for hiding a secret file in several WAV files depends on a circular secret key”, *Egyptian Informatics Journal*, Vol. 23, No. 4, pp. 33-43, 2022, doi: 10.1016/j.eij.2022.06.003.

[7] I. B. Prayogi, T. Ahmad, N. J. De La Croix, and P. Maniriho, “Hiding Messages in Audio using Modulus Operation and Simple Partition”, In:

Proc. of Proceedings of 2021 13th International Conference on Information and Communication Technology and System, ICTS 2021, Institute of Electrical and Electronics Engineers Inc., pp. 51-55, 2021, doi: 10.1109/ICTS52701.2021.9609028.

[8] H. T. Elshoush and M. M. Mahmoud, "Ameliorating LSB Using Piecewise Linear Chaotic Map and One-Time Pad for Superlative Capacity, Imperceptibility and Secure Audio Steganography", *IEEE Access*, Vol. 11, pp. 33354-33380, 2023, doi: 10.1109/ACCESS.2023.3259902.

[9] M. M. Mahmoud and H. T. Elshoush, "Enhancing LSB Using Binary Message Size Encoding for High Capacity, Transparent and Secure Audio Steganography-An Innovative Approach", *IEEE Access*, Vol. 10, pp. 29954-29971, 2022, doi: 10.1109/ACCESS.2022.3155146.

[10] H. Tian, Y. Qiu, W. Mazurczyk, H. Li, and Z. Qian, "STFF-SM: Steganalysis Model Based on Spatial and Temporal Feature Fusion for Speech Streams", *IEEE/ACM Trans Audio Speech Lang Process*, Vol. 31, pp. 277-289, 2023, doi: 10.1109/TASLP.2022.3224295.

[11] Y. Ren, D. Liu, C. Liu, Q. Xiong, J. Fu, and L. Wang, "A Universal Audio Steganalysis Scheme Based on Multiscale Spectrograms and DeepResNet", *IEEE Trans Dependable Secure Comput*, Vol. 20, No. 1, pp. 665-679, 2023, doi: 10.1109/TDSC.2022.3141121.

[12] N. J. De La Croix and T. Ahmad, "Toward secret data location via fuzzy logic and convolutional neural network", *Egyptian Informatics Journal*, Vol. 24, No. 3, pp. 100385, 2023, doi: 10.1016/j.eij.2023.05.010.

[13] J. D. L. C. Ntivuguruzwa, A. Tohari, and M. I. Royyana, "Pixel-block-based Steganalysis Method for Hidden Data Location in Digital Images", *International Journal of Intelligent Engineering and Systems*, Vol. 16, No. 6, pp. 375-385, 2023, doi: 10.22266/ijies2023.1231.31.

[14] N. J. D. La Croix, T. Ahmad, and F. Han, "Enhancing Secret Data Detection Using Convolutional Neural Networks with Fuzzy Edge Detection", *IEEE Access*, pp. 1-1, 2023, doi: 10.1109/ACCESS.2023.3334650.

[15] J. D. L. C. Ntivuguruzwa and T. Ahmad, "A convolutional neural network to detect possible hidden data in spatial domain images", *Cybersecurity*, Vol. 6, No. 1, pp. 23, 2023, doi: 10.1186/s42400-023-00156-x.

[16] M. M. Amrulloh and T. Ahmad, "Minimising Sample Space to Optimise Quality of Stego-Audio", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 3, pp. 205-214, 2022, doi: 10.22266/ijies2022.0630.17.

[17] T. Ahmad, J. N. Faruki, R. M. Ijtihadie, and W. Wibisono, "Analysing the Effect of Block Size on the Quality of the Stego Audio", In: *Proc. of 2019 5th International Conference on Science and Technology (ICST)*, IEEE, pp. 1-6, 2019, doi: 10.1109/ICST47872.2019.9166403.

[18] Y. Samudra and T. Ahmad, "Segmentation embedding method with modified interpolation for increasing the capacity of adaptable and reversible audio data hiding", *Journal of King Saud University - Computer and Information Sciences*, Vol. 35, No. 8, 2023, doi: 10.1016/j.jksuci.2023.101636.

[19] S. Jiang, D. Ye, J. Huang, Y. Shang, and Z. Zheng, "SmartSteganogaphy: Light-weight generative audio steganography model for smart embedding application", *Journal of Network and Computer Applications*, vol. 165, pp. 102689, 2020, doi: 10.1016/j.jnca.2020.102689.

[20] X. Zhang, C. Li, and L. Tian, "Advanced audio coding steganography algorithm with distortion minimisation model based on audio beat", *Computers and Electrical Engineering*, Vol. 106, pp. 108580, 2023, doi: 10.1016/j.compeleceng.2023.108580.

[21] K. Manjunath, G. N. Kodanda Ramaiah, and M. N. GiriPrasad, "Backward movement-oriented shark smell optimisation-based audio steganography using encryption and compression strategies", *Digit Signal Process*, Vol. 122, pp. 103335, 2022, doi: 10.1016/j.dsp.2021.103335.

[22] J.J. Bosch, F. Fuhrmann, Herrera, 2012. P.: IRMAS: a dataset for instrument recognition in musical audio signals. https://www.upf.edu/web/mtg/irmas/ (Accessed 22nd February 2024).

[23] J. J. Bosch, J. Janer, F. Fuhrmann, and P. Herrera, "A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals", In: *Proc. of 13th Int. Soc. Music Inf. Retr. Conf.* ISMIR 2012, pp. 559-564, 2012.

[24] K. Sharma and K. Gupta, "Lossless data compression techniques and their performance", In: *Proc. of 2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 256-261, 2017, doi: 10.1109/CCAA.2017.8229810.

[25] "Lorem Ipsum", www.id.lipsum.com (Accessed 22nd February 2024).