Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.1968

Research Article

Audio fingerprinting for song identification using discrete wavelet transform

Swati Mukesh DIXIT*, Daulappa Guranna BHALKE

[1]*Dr. D Y Patil Institute of Technology, Pimpri, Pune, 411018, India*

ARTICLE INFO

Article history
Received: 10 April 2024
Revised: 13 May 2024
Accepted: 08 December 2024

Keywords:
Audio Fingerprint; DWT; MFCC; Song Identification

ABSTRACT

This paper aims to enhance accuracy and robustness in audio recognition through the use of the Discrete Wavelet Transform (DWT) with Daubechies wavelets for song identification. This work is important because the actual song identity can be accurately matched based on small music fingerprints that are essential for applications such copyright detection and music identification apps. This method is based on decomposing the frames of a audio sample into sub-bands using Daubechies wavelets and extracting statistical features to form a 8-bit fingerprint. This fingerprint is much more compact than that yielded by other techniques such as FFT (256 bits) and FrFT (150 bits). By comparing these fingerprints to a database, the system can accurately identify songs from short snippets even in loud environments without needing song metadata. The results show that the accuracy of Daubechies wavelet method stands at 98% trailing behind existing wavelet based methods where they were reported to have accuracies of 86.7%, 97% and 90%. The high precision and reduced fingerprint size footprint makes the approach ideal for fast, accurate song matching. The novelty of the proposed work lies in its capability to generate very accurate as well as compact fingerprints which overcomes the limitations of previous techniques (e.g. FFT, FrFT etc.). The proposed approach outperforms previous works in the literature on audio fingerprinting for short audio content and noisy environments.

Cite this article as: Dixit SM, Bhalke DG. Audio fingerprinting for song identification using discrete wavelet transform. Sigma J Eng Nat Sci 2026;44(1):74–82.

INTRODUCTION

Music and song recognition has attracted sustained interest from both academic researchers and industry, owing to the technical complexity of the task and its significant commercial value [1]. Non-invasive techniques are widely utilized in applications such as retrieval, recognition, and authentication of digital content. These techniques work by analyzing the original signal without modifying it. Among these, fingerprinting stands out as a key application, offering a fast and reliable method for content identification [2]. An audio fingerprint is a compact digital representation used to identify and index short, unlabeled audio segments in a database [3]. Audio signals are often linked to metadata like song titles and artist names, stored in databases

*Corresponding author.
*E-mail address: swaatisutar@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

via audio fingerprinting or content-based metadata [4]. Currently, the main audio fingerprinting methods incorporate the fingerprints of Shazam and Philips. Three methods are usually used in audio retrieval: key speaker identification, keyword detection, and key audio detection. While these technologies are fairly advanced, they still have several shortcomings. This paper mainly focuses on one thing to generate smaller size of fingerprint.

Audio fingerprinting extracts unique features from a piece of audio and stores them in a database. When an unidentified audio piece is presented, the system extracts its fingerprint and matches it against those in the database. For audio fingerprinting to be effective, the fingerprint must capture and describe the core attributes of the audio content [5]. This enables the system to recognize distorted versions of a recording as the same audio signal using fingerprints and matching algorithms [6,7]. Haitsma and Kalker proposed five main parameters for an audio fingerprinting system: [8,9]

Robustness: The system must be able to accurately recognize audio despite noise and distortion. It should consistently identify audio clips regardless of compression, with robustness measured by the bit error rate.

Reliability: This indicates the probability of incorrectly identifying an audio piece.

Fingerprint size: This pertains to the number of bits within a fingerprint.

Granularity: This indicates the minimal length of audio needed for identification.

Search speed and scalability: This measures how quickly a fingerprint can be found in a large database.

Database search efficiency: To maintain scalability, the fingerprint representation must allow for efficient database searches [10,11].

The different methods are used to generate audio fingerprint like FFT, FrFT, wavelet along with different feature extraction techniques like MFCC, LPCC, Receiver operating characteristics with different datasets like IRMAS and MTG-Jamendo . The summary of related previous work is arranged as follows-

Reise et al. [12] proposed a novel topological audio fingerprinting algorithm for efficient and accurate duplicate audio detection. This approach uses persistent homology on local spectral decomposition of audio signals, using filtered cubical complexes computed directly from mel-spectrograms. Using localized Betti curves to encode the audio signal, we get their guarantees that a time-shifted audio match can be accurately recovered. Experiments confirm the ability of our approach to find tracks having identical audio content even under obfuscations.This approach surpasses current methods, particularly in scenarios with topological distortions like time stretching and pitch shifting. Manjunath et al. [13] explained that audio fingerprinting technologies can monitor content without metadata or watermarks and have a wide range of applications. The algorithm is trained on a large set of short audio speech commands, where the efficiency is increased by including an audio feature extraction step at pre-processing. The accuracy of its generated model is up to 83.7%. Suhas [14] stated that over 90% of the audio signal processing domain is operated with Fourier based techniques, However, no new feature personally investigated. The present study addresses this problem by using the ROC curve to judge the performance of audio or acoustic features. Wavelet-based approaches have yielded results on par with the best-known fourier methods, therefore serving as potential alternatives for future audio processing research. Victoire Djimna Noyuma et al. [15] explained about listening to a song can inspire curiosity about the artist's biography and other works. This project centers on singer identification and includes three main phases: isolating the singer's voice from the background music, extracting vocal signal features, and using them for identification. Jaya Nameirakpam et al. [16] discussed the release of a singer recognition method, which uses DAMP-balanced dataset. It uses wavelet transform for pre-processing and MFCCs as audio features. A GMM is applied for classification. Comparison is analyzed between classification accuracy and computational time with and without use of wavelet transform in the study.

S. Preetha et al. [17] explained that the widely used Daubechies wavelet is fundamental to many modern techniques. Enhancing video copy detection can be achieved by optimizing the wavelet filter bank. This paper presents a wavelet design method using polyphase representation and the Gravitational Search Algorithm (GSA). Its effectiveness is supported by simulations as well appending. Sukanta Kumar Dash [18] The proposed system was evaluated on the IRMAS data set where n training is different from n testing . The DWT feature dimension was 250, which achieved the best performance. Aggregating micro and macro Precision, Recall, and F1scores the model reached 0.695 and 0.631. These results show gains of 12.28% and 23.0% above Han's benchmark CNN model. Mani Malekesmaeili et al. [19] introduced a novel fingerprinting scheme by extracting the local fingerprints feature from time-chroma images, which are robust against time/frequency scale transformation in audio content. This new approach outperforms significantly well known methods like SIFT. They propose also a finger-prints oriented song retrieval algorithm and display a copy error detection system which significantly outperforms state-of-the-art findings. In addition to detecting copied songs or partial copies, our system can correctly estimate pitch shift and/or tempo change into the song. Qiu-yu Zhang et al. [20] described how a feature matrix is created by combining the source speech such as the Mel-frequency cepstral coefficients (MFCC) and the linear prediction cepstrum coefficients (LPCC).Information entropy is used for columns, and an energy-based technique is used for rows, when applying dimension reduction. An audio fingerprint is formed by this reduced feature matrix. The normalized Hamming distance algorithm is used by retrieval for matching. According to experimental data, this technique

provides high recall and precision rates, produces smaller audio fingerprints, and is more robust for long speech segments. It also maintains better retrieval performance. Kamladas et al. [21] reported Fingerprint generation using DWT is robust with audio quality change for identification. This approach also yields a small number of fingerprints. Salvatore Serrano et al. [1] This paper compares a novel algorithm with various baseline methods for handling very short audio clips. Using a subset of the MTG-Jamendo dataset and a proprietary collection of 7000 songs, experimental findings indicate that the proposed algorithm outperforms others, especially for audio snippets shorter than 3 seconds. Rajnikumari et al. [4] The study presents a technique for minimizing background noise and distinguishing between voiced and unvoiced speech. This method is based on low and high-frequency subbands decomposition of voice data using the Maximal Overlap Discrete Wavelet Transform (MODWT). MODWT is robust across segments. Furthermore, in noisy conditions, PNCCs also provide much higher performance than Mel Cepstral Coefficients (MCC), resulting in a better SNR.

Although great progress has been made in many current work, limitations related to scalability, computational efficiency, the adaptability of a variety or complex dataset and need for further existence based method, explorations on reducing the fingerprint size. A new feature extraction approach, which is capable of generating a fingerprint that is robust and reliable for the audio fingerprinting system, is proposed in this paper. The proposed method tackles the challenge of minimizing audio fingerprint sizes, allowing for the storage of a large number of fingerprints in the database. As the database increases the search time increases. To overcome this, hash code of query fingerprint is generated and used to quickly locate matching entries in the hash table.

These measures help to maintain system performance as the database grows. It can recognize audio even if the speed and pitch have been changed [22,23].

Audio Fingerprint Requirements

Invariance to distortion

In the face of noise, distortion, and compression, an audio fingerprint should be able to distinguish an audio signal.

Compactness

The compact size of audio fingerprints enables the storage of a large number of them in the database.

Computational simplicity

It should take less time to extract a fingerprint.

Quicker identification

The speed of music identification through an audio fingerprint depends on the required duration of audio seconds for music recognition.

Audio Fingerprint Applications

Broadcast monitoring

This is the process of automatically creating a playlist for broadcasts on radio, television, or the internet in order to verify programs and advertisements collect royalties, and measure audience size. A central server that houses the fingerprint database and several monitoring locations make up a fingerprint-based wide-area broadcast monitoring system. At the monitoring stations, fingerprints were collected from each (local) broadcast channel. The fingerprints are gathered by the primary site from the monitoring sites. DWT is effective in identifying audio segments even when they have been slightly modified, such as those broadcasted by different radio stations using various compression formats.

Connected audio

This technology is utilized in consumer applications where audio is linked to supplementary and corroborated data. Speech coding, transmission across a cellular network, an acoustical channel between a mobile phone's loudspeaker and microphone, and FM/AM transmission can all cause an audio signal to deteriorate. Hence, this represents a challenging scenario.

Autonomous arrangement of music library

Numerous PC users possess hundreds or even thousands of tracks within their music collections. MP3 music files are compressed files. This music were downloaded from the internet, transferred via Bluetooth, and ripped from CDs, among other sources. Thus, the song library is disorganized. Since the metadata is constantly lacking and incorrect, the song library's metadata was properly organized using audio fingerprints [24].

Recommended Approach

Several techniques, including FFT, DCT, and Fractional Fourier transform or FrFT and Wavelet are used to generate audio fingerprints. Discrete wavelet transform is utilized in this paper to generate an audio fingerprint. This method uses wavelet transform as fingerprint extraction algorithm in time and frequency domain. After one and four layer wavelet decomposition of the audio signal, energy and band division are calculated. These values are then utilized as parameters to construct an 8-bit fingerprint block corresponding to each frame. Application of this algorithm reduces the amount of fingerprints datum per file, and consequently search time, system memory space. This in turn also allows for audio identification even where there is variation in the quality of audio waveforms [25].

The processes of fingerprint extraction and matching make up an audio fingerprint system. An audio signal in waveform is the system's input. The output plays back the song and contains its metadata. The suggested audio fingerprint system includes fingerprint matching and audio fingerprint extraction. Framing, overlapping, windowing, feature extraction, fingerprint modelling, and pre-processing are the

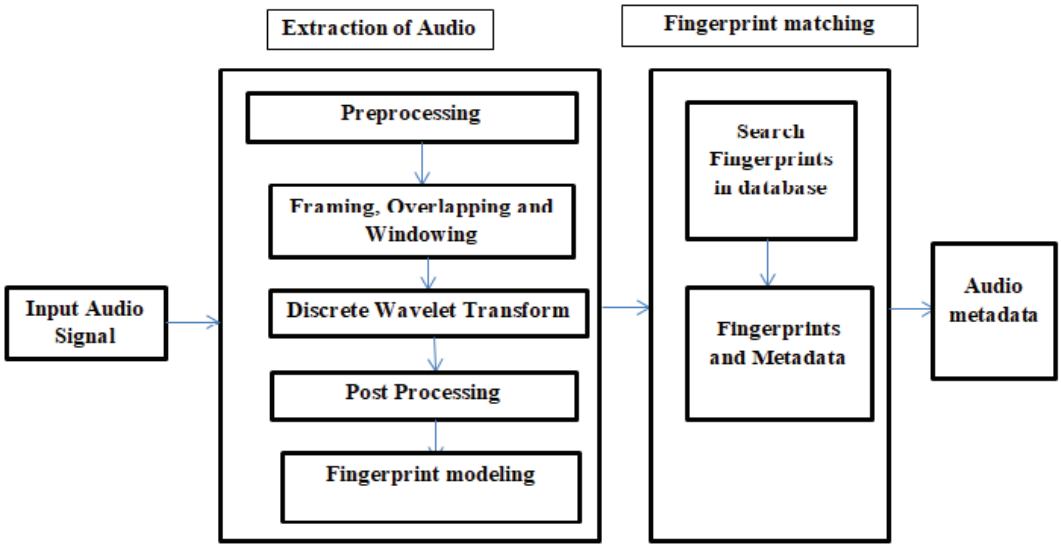

Figure 1. Recommended diagram for audio fingerprint extraction.

steps involved in audio fingerprint extraction. When two fingerprints match the best, the metadata is displayed and the song is played back. The matching process involves searching the fingerprint database [26]. Figure 1 shows the steps listed below are used to create an audio fingerprint.

Initial Preparation

This is useful to remove the silence part of an audio signal. To remove silent part, we make use of energy property of one component of the audio signal. 3. The first stage of the processing digitises the audio and converts it to a standard form (to 16-bit PCM, for example) monophonic with a sample rate varying between 5 kHz and 44.1 kHz. [27,28]. An audio fingerprinting system can usually work with various audio formats. Here, .mp3 files are converted into .wav file. It ensures that working with the original audio information without any loss due to compression. As all audio files converted into .wav, so sampling rate is same for all.

This signal is employed for down sampling to a standardized rate; in this case, 8 KHz is utilized as the down sampling rate to reduce the data rate, which is helpful for creating fingerprints.

Windowing, Framing, and Overlapping

Since audio signals are non-stationary, framing is crucial to maintaining signal stationary. For a few milliseconds, the signal is thought to be stationary; in this case, it is thought to be 20 ms. More overlap is thought to be necessary for resistance to shifting since it preserves statistical features. Overlap in this case is 90%. A Hamming window is applied to each frame to maintain coherence between its starting and ending points [29,30]. This window is chosen for its minimal sideband attenuation. Below is the representation of the Hamming window:

$$\mathrm{w(n)} = 0.54 - 0.46\cos\{\tfrac{2\pi n}{n-1}\},\ \ 0 \le n \le N-1 \qquad (1)$$

Discrete Wavelet Transform

An alternative to the short term fourier transform (STFT) in signal analysis is the Wavelet Transform (WT),which struggles with time and frequency resolution issues. The WT is effective for both stationary and non-stationary signals, with applications ranging from electrical noise reduction and sudden change detection to data compression. A specific type of WT, known as the Discrete Wavelet Transform (DWT), utilizes filters with different cutoff frequencies to analyze signals at various scales. Wavelet transforms offer a key advantage in their ability to localize signals effectively in both the time and frequency domains [31]. Wavelets break down the signal into its most basic components (detailed and approximation coefficient) and then accurately recreate it [32]. The Daubechies (db) wavelet makes audio signals more enjoyable by allowing for little degradation during de-noise and compression processes [33].

The DWT is given by the equation-

$$DWT(j,k) = \sum_j \sum_k x(k) 2^{-\frac{j}{2}} \qquad (2)$$

Since ψ(τ) is mother wavelet or transforming action.

The signal's high and low frequency components are examined using a variety of high-pass and low-pass filters. Upsampling and downsampling procedures alter the scale, while filtering refines the signal's resolution.

The processes used for filtering are by,

$$y_{\text{high}}\,[\mathrm{k}] = \sum_n X[n]\; g\;[2k-n] \qquad (3)$$

$$y_{\text{low}}\,[\mathrm{k}] = \sum_n X[n]\; h\;[2k-n] \qquad (4)$$

Thus, after being downsampled by 2, the high-pass (g) and low-pass (h) filtering yields the outputs yhigh [k] and ylow [k], respectively. Since the DWT is a decomposition method, coefficients as many as input points are produced. The decimation factor for the original signal is 2. It enables faster computing time by using DWT. In this case, the Daubechies 1 and 4 wavelet families of coefficients are used [16].

Energy Computation and Band Division

The power spectrum of the signal is increased in a range from 300 to 2000 Hz by the magnitude response of these 33 triangular band pass filters. Sub-bands are formed by the logarithmic spacing used. The bandpass filters are evenly spaced on an MEL frequency scale, which is defined as follows in terms of the standard linear frequency 'f':

$$Mel(f) = 2595 * \ln\left(1 + \frac{f}{700}\right) \quad (5)$$

Mel frequency, is directly proportional to the log-linear frequency transforming [18][25].

MFCC Sub Fingerprint Generation

Figure 2 illustrates number of steps included such as Preprocessing, framing, overlapping, windowing (cosine) frame, DWT and

Mel filter bank followed by log transformation and at the end it applies Discrete Cosine Transform (DCT). The mel-frequency cepstrum (MFC) is the representation and MFCCs are the coefficients. The DCT is then applied to the log power spectrum in order estimate the short-term audio spectrogram content across a non-linear Mel frequency scale**.**MFCCs have been widely recognized as the standard for representing acoustic features in audio processing tasks [20][33].

After obtaining MFCC coefficients discrete wavelet transform is applied to MFCC coefficients. This will decompose the MFCC values into approximation and detail coefficients at multiple levels. The statistical values has been calculated as follows-

Wavelet Decomposition

One should bear in mind that a Fourier transform produces frequency domain information and is not suitable for signals that are non-stationary (i.e., they change with time). Wavelet transform is applied since time- dependent information manifests in such signals. It decomposes signals into components that are both time and frequency localized. The signal is decomposed into wavelets through four levels, including one approximation and four details coefficients: CD1, CD2, CD3, and CD4. Features like variance, zero crossing rate, centroid, energy are also calculated once the signal is decomposed.

Variance

The variance describes how far the values in a dataset spread out from their mean. In the case of wavelet coefficients, their variance can be calculated using the following expression.

$$\sigma^2 = \frac{1}{N}\sum_i^N (c[i] - \bar{c})^2 \quad (6)$$

Since, $\bar{c}$ = Mean of the wavelet coefficients

c[i] = Wavelet coefficient at index i

N = Total number of wavelet coefficients

ii)Zero-crossing rate (ZCR)-

Another fundamental acoustic property that is simple to compute is the zero-crossing rate (ZCR). It is equivalent to the quantity of waveform zero crossings that occur inside a specific frame. For end-point identification, ZCR and volume are frequently combined. ZCR is specifically utilized to identify the beginning and ending locations of sounds that are not spoken.

$$\text{ZCR} = \frac{1}{N-1}\sum_{i=1}^{N-1} sgn(c[i]) \neq sgn(c[i+1]) \quad (7)$$

Since, sgn(x) = Sign function, which is 1 if x≥0 and −1 if x<0

iii) Centroid-

The center of energy distribution is used to represent the centroid of the wavelet domain. Since the audio signal's centroid varies over time in the wavelet domain, this property may indicate the audio signal's non-stationarity. The equation for determining the wavelet domain centroid is given as follows-

$$C = \frac{\sum_{i=1}^{N} |c[i].f[i]|}{\sum_{i=1}^{N} |c[i]|} \quad (8)$$

Since,

f[i] = Frequency associated with the coefficient c[i]

iv) Energy of wavelet domain-

An essential dynamic feature of the audio signal is its fluctuating amplitude, which can also represent variations in energy. The wavelet coefficients can capture the energy characteristics of an audio signal due to their correspondence with time-domain averages.

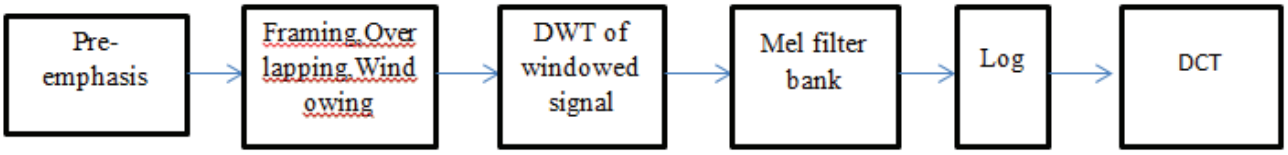

Figure 2. MFCC coefficient calculation.

$$E = \frac{1}{N}\sum_{i=1}^{N} |c[i]|^2 \quad (9)$$

Once the zero crossing rate, energy, variance and centroid were obtained for each frame of the audio stream they were used to create the fingerprints using equation (10). For each frame an 8-bit sub-fingerprint was generated. The variation of wavelet coefficients is employed to generate the final 5-bits of the 8-bit sub-fingerprints. The following equation is used to calculate hash bit values.

$$F(n,m) = \begin{cases} 1 & \text{if } E(n,m) - E(n,m+1) - (E(n-1,m) - E(n-1,m+1)) > 0 \\ 0 & \text{if } E(n,m) - E(n,m+1) - (E(n-1,m) - E(n-1,m+1)) \leq 0 \end{cases} \quad (10)$$

In this context, the m-th bit of the sub-fingerprint for frame n as F(n,m), whereas the energy of band m for frame n is denoted as E(n,m). The next 3-bits are generated by using zero crossing rate, centroid and energy of the wavelet coefficients. Therefore, for each frame, an 8-bit sub-fingerprint is generated. These bands are logarithmically in the range 300-2000 Hz. 150 additional fingerprints have been combined to form a fingerprint block. Again, slap-hand high-fives because this one has a metric shit ton of finger-pickage [21]. MFCCs are frequently utilized because they imitate human auditory perception and are less affected by noise and distortion.

Fingerprint Matching

To match fingerprints, first send a query signal to an audio fingerprint generation system. The fingerprints that were created and those that were kept in the database matched. When the best match is determined, the matching metadata is displayed and the song is played back. The matching process is based on the Hamming distance. AES256 (Advanced Encryption Standard) is a powerful encryption standard, used to protect the audio fingerprint itself and associated metadata. Once an audio fingerprint is generated, it's often sensitive data. AES256 can encrypt the fingerprint to prevent unauthorized access or modification. Audio fingerprint is a string of characters or numbers converted into a suitable format for AES256 encryption. Padding is essential to ensure data blocks align with the encryption block size because of its simplicity and effectiveness. All information stored in the database, including audio fingerprints, is protected with robust encryption methods advanced encryption standard AES-256. This ensures that even if unauthorized individuals gain access to the physical storage media, the data will remain indecipherable. AES is applied to fingerprint data before storing it. So it gives strong protection. As the key length is longer it gives strong protection.

RESULTS AND DISCUSSION

The audio files are. at 22050 Hz in.mp4 format for the input to this study. It includes three cases–i) without the wavelet db transformation, ii) applying on the transformed with the db1 db wavelets, and iii) using the db4. In each case, 9 MFCCs are computed for each audio signal. Noise is filtered by a low-pass filter. Feature extraction. The pre-processed signal is filtered and down-sampled to 8 kHz. It is then three-framed with overlap. Then each frame of the audio signal is windowed using a Hamming window. The Input signal, detail and approximation signals are shown in Figure 3.

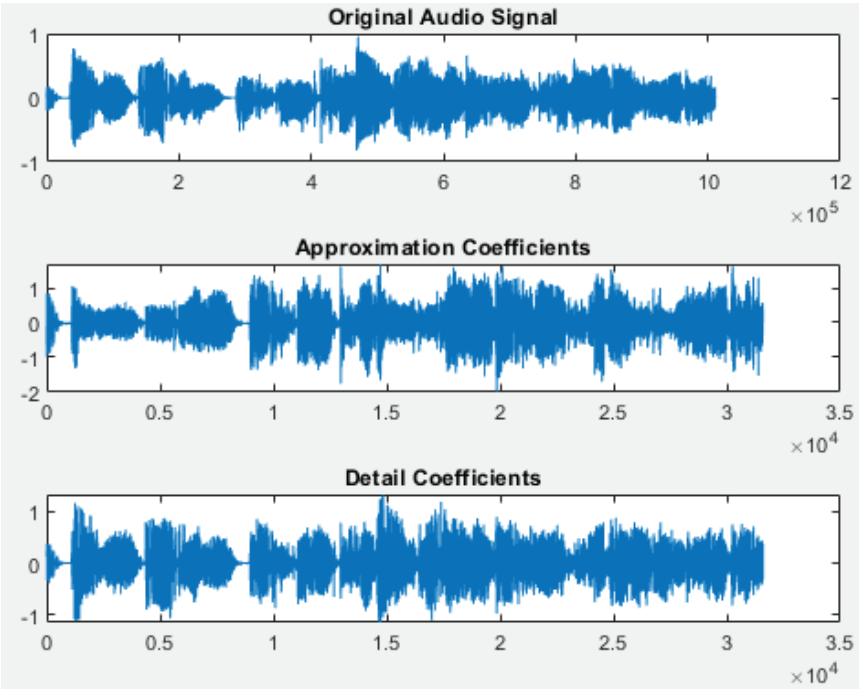

Figure 3. Original, approximation and detailed coefficient of an audio signal.

Table 1. Accuracy of song identification using different cases

Sr.No.	Cases	% Accuracy
1	Without using wavelet transform	97.2%
2	Using db1 wavelet	96%
3	Using db4 wavelet	98%

After feature extraction for MFCC coefficients, variance, zero-crossing rate, centroid and energy in the wavelet domain by using equations (6), (7), (8) and (9), an 8-bit fingerprint is obtained. The computed features are then mapped into bits: if the difference in variances between neighboring frames is positive, the bit value is 1; otherwise, it is 0. This process yields an 8-bit fingerprint for every frame in the audio signal.

While using db1 and db4, the db4 gives better accuracy i.e.98%. The comparison table of accuracy with 3 cases is shown in table1.

The proposed audio fingerprinting system using DWT (Daubechies) wavelets scales efficiently with larger databases due to the compact and robust nature of the wavelet-based fingerprints. However, as the database size increases, potential bottlenecks may arise in terms of computational resources and search speed. Here we generate a hash table, where each fingerprint is represented by a compact hash code derived from its features. These hash codes are then used as keys to store and retrieve the fingerprints efficiently. During a search, the hash code of a query fingerprint is computed and used to quickly locate matching entries in the hash table.

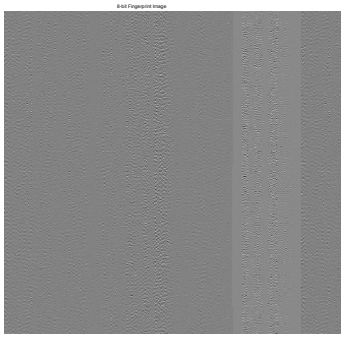

Figure 4. 8-bit fingerprint size.

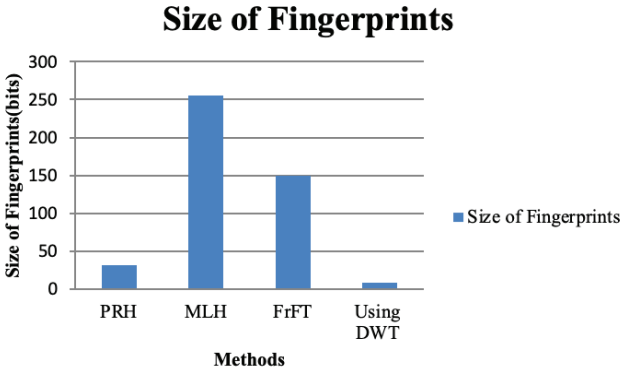

Figure 5. Comparison between different methods of fingerprint size.

Fig. 4] shows 8-bit fingerprint generated for song "Jeena Yahan marna yahan". Fig.5] shows graphical representation of the size of audio fingerprints generated by different methods like Philips Robust Hashing Method (PRH-32 bit), Multiple Hashing (MLH-256 bit) method, Fractional Fourier transform (150 bit) and proposed method (8 bit).

To better illustrate the compression potential, Table 2 presents the fingerprint size for each method. It includes the sizes for five comparison methods as well as the proposed method. Since all methods use frames of the same length and extract a fingerprint from each frame, the experiment only requires a comparison of the size of a single fingerprint

For robustness of the system this algorithm surpasses existing algorithms like PRH, MLH and FrFT when

Table 2. Fingerprint size per frame using different methods

Method	Fingerprint size (per frame)
Multiple Hashing Method	256 bits
Philips	32 bits
Shazam	32 bits
FFT	32 bits
FrFT	32 bits
DWT	8 bits

Table 3. Comparison between different methods for song identification in the presence of noise

SNR	PRH [32]	MLH[32]	FrFT[13]	MFCC+LPCC[27]	Proposed method
10	Misdetection occurred	Misdetection occurred	Misdetection occurred	Misdetection occurred	Song identified
25	Misdetection occurred	Song identified	Misdetection occurred	Song identified	Song identified
30	Song identified	Song identified	Song identified	Song identified	Song identified
40	Song identified	Song identified	Song identified	Song identified	Song identified

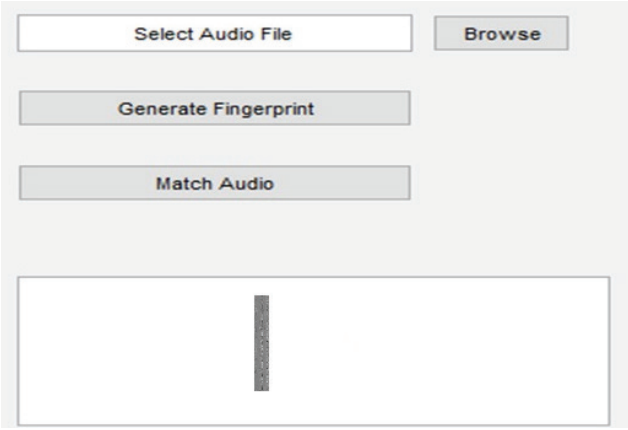

Figure 6. User interface for audio fingerprint.

handling AWGN noise at low signal-to-noise ratios (SNRs), PRH, MLH, and FrFT algorithms encounter challenges in identifying the song or may result in misdetections. Conversely, the proposed algorithm exhibits strong robustness against Additive White Gaussian Noise (AWGN), even when SNR levels are low.

Table 3 presents a comparison chart of the identification performance in the presence of AWGN noise for various algorithms, including the proposed one.

Figure 6 shows the user interface for audio fingerprint system.

As shown in fig.6] GUI is generated using MATLAB program. Select audio file field is likely where the user would input the audio file they want to process. There is a corresponding "Browse" link which can be clicked to locate and choose the file from your local computer. Create Fingerprint button simply means the action of creating the audio fingerprint for your selected audio file. The Match Audio button is used to match a new audio file to an existing fingerprint or database of fingerprints. After a match is found it shows the fingerprint.

CONCLUSION

The wavelet transform based audio fingerprinting scheme just now proposed is a milestone in audio technology. This new technique significantly limits the total number of fingerprints per file to 8 bits, quite novel from state-of-the-art wherein infinite number of fingerprints are generated. Such a reduction speeds up the search and reduces the storage of the database. With the availability of background noise in a significant amount, by involving db4 wavelet to aid fingerprints extracting process, the method improves accuracy and robustness. The Db4 wavelets are able to capture the significant features of single components with little noise interference, and it yields a comparable PF rate of 98%. This fine-grained precision is ideal for use cases that require a high-quality trigger in noisy acoustic environments, such as music recognition, audio surveillance and acoustic event detection applications. The key idea behind this work is reducing the size of fingerprint as well as improving accuracy. A smaller fingerprint size means faster fingerprints, and space for more songs in the database. The better processing of noise condition using db4 wavelets are approached over earlier methods, which makes this technique as a good number for high reliability sound recognition systems.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Serrano S, Scarpa M. Accuracy comparisons of fingerprint based song recognition approaches using very high granularity. Multimedia Tools Appl 2023;82:31591–31606. [CrossRef]

[2] Bhalke DG, Rao CBR, Bormane D. Hybridization of mel frequency cepstral coefficient and higher order spectral features for musical instruments classification. Arch Acoust 2016;41:427–436. [CrossRef]

[3] Lee S, Yook D, Chang S. An efficient audio fingerprint search algorithm for music retrieval. IEEE Trans Consum Electron 2013;59:739–745. [CrossRef]

[4] RajaniKumari, Babu K. Designing highly secured speaker identification with audio fingerprinting using MODWT and RBFNN. Int J Intell Syst Appl Eng 2024;12:25–30.

[5] Son HS, Byun SW, Lee SP. A robust audio fingerprinting using a new hashing method. IEEE Access 2020;8:168953–168963. [CrossRef]

[6] Liu Y, Cho K, Yun HS, Shin JW, Kim NS. DCT based multiple hashing technique for robust audio fingerprinting. IEEE Int Conf Acoust Speech Signal Process 2009:61–64. [CrossRef]
[7] Belletini C, Mazzini M. A framework for robust audio fingerprinting. J Commun 2010;5:401–408. [CrossRef]
[8] Haitsma J, Kalker T. A highly robust audio fingerprinting system. In: Proc Int Conf Music Inf Retrieval; 2002. p. 107–115.
[9] Unal E, Chew E, Georgiou PG, Narayanan SS. Challenging uncertainty in query by humming systems: a fingerprinting approach. IEEE Trans Audio Speech Lang Process 2008;16:359–371. [CrossRef]
[10] Doets P, Lagendijk R. Distortion estimation in compressed music using only audio fingerprints. IEEE Trans Audio Speech Lang Process 2008;16:275–286. [CrossRef]
[11] Singh A, Demuynck K, Arora V. Simultaneously learning robust audio embeddings and balanced hash codes for query-by-example. In: ICASSP 2023 - 2023 IEEE Int Conf Acoust Speech Signal Process; 2023. p. 1–5. [CrossRef]
[12] Reise W, Fernández X, Dominguez M, Harrington HA, Beguerisse-Díaz M. Topological fingerprints for audio identification. arXiv 2023;2309.03516. Preprint. doi: 10.48550/arXiv.2309.03516
[13] Arunakumar, BN, Shashidhar R, Sahana B, Jagadamba G, Manjunath AS, Roopa M. Fingerprint definition for song recognition using machine learning algorithm. In: 2023 Int Conf Smart Syst Appl Electr Sci (ICSSES); 2023. p. 1–6. [CrossRef]
[14] Suhas BN. Automatic bird sound detection in long-range field recordings using wavelets & mel filter bank features. In: 2020 IEEE Second Int Conf Cogn Mach Intell (CogMI); 2020. p. 978–1–7281-4144-2.
[15] Djimna Noyuma V, Perieukeu Mofenjoua Y, Feudjioa C, Göktugb A, Fokoué E. Boosting the predictive accuracy of singer identification using discrete wavelet transform for feature extraction. arXiv 2021;2102.00550v1. Preprint. doi: 10.48550/arXiv.2102.00550.
[16] Nameirakpam J, Biswas S, Bonjyostna A. Singer identification using wavelet transform. In: 2019 2nd Int Conf Innov Electr Signal Process Commun (IESC); 2019. p. 238–242. [CrossRef]
[17] Preetha S, Bindu VR. A wavelet optimized video copy detection using content fingerprinting. J Signal Process Syst 2023;95:363–377. [CrossRef]
[18] Dash SK, Solanki SS, Chakraborty S. Deep convolutional neural networks for predominant instrument recognition in polyphonic music using discrete wavelet transform. Circuits Syst Signal Process 2024;43:4239–4271. [CrossRef]
[19] Malekesmaeili M, Ward RK. A local fingerprinting approach for audio copy detection. Signal Process 2014;98:308–321. [CrossRef]
[20] Zhang QY, Xu FJ, Bai J. Audio fingerprint retrieval method based on feature dimension reduction and feature combination. KSII Trans Internet Inf Syst 2021;15:409–423. [CrossRef]
[21] Kamaladas M, Dialin M. Fingerprint extraction of audio signal using wavelet transform. In: 2013 Int Conf Signal Process Image Process Pattern Recognit (ICSIPR); 2013. p. 1–5. [CrossRef]
[22] Radha K, Bansal M, Pachori RB. Speech and speaker recognition using raw waveform modeling for adult and children's speech: A comprehensive review. Eng Appl Artif Intell 2024;131:107661. [CrossRef]
[23] Sutar SV, Bhalke DG. Audio fingerprinting using fractional fourier transform. IOSR J Electron Commun Eng 2015;10:30–34.
[24] Schreiber H, Muller M. Accelerating index based audio identification. IEEE Trans Multimedia 2014;16:1654–1664. [CrossRef]
[25] Bhalke DG, Rao CBR, Bormane DS. Hybridization of fractional fourier transform and acoustic features for musical instrument recognition. Int J Signal Process Image Process Pattern Recognit 2014;7:275–282. [CrossRef]
[26] Bhalke DG, Ramarao CB, Bormane DS. Automatic musical instrument classification using fractional fourier transform based-MFCC features and counter propagation neural network. J Intell Inf Syst 2015;46:425–446. [CrossRef]
[27] Son W, Cho H, Yoon K, Lee S. Sub-fingerprint masking for a robust audio fingerprinting system in real noise environment for portable consumer devices. IEEE Trans Consum Electron 2010;56:144–151. [CrossRef]
[28] Sharma G, Umapathy K, Krishnan S. Trends in audio signal feature extraction methods. Appl Acoust 2020;158:107020. [CrossRef]
[29] Hanifa RM, Isa K, Mohamad S. A review on speaker recognition: Technology and challenges. Comput Electr Eng 2021;90:107005. [CrossRef]
[30] Ravi ND, Bhalke DG. Musical instrument information retrieval using neural network. In: 2016 IEEE Int Conf Adv Electron Commun Comput Technol (ICAECCT); 2016. p. 418–422. [CrossRef]
[31] Jia M, Li T, Wang J. Audio fingerprint extraction based on locally linear embedding for audio retrieval system. Electronics 2020;9:1483. [CrossRef]
[32] Liu Y, Cho K, Yun HS, Shin JW, Kim NS. DCT based multiple hashing technique for robust audio fingerprinting. In: 2009 IEEE Int Conf Acoust Speech Signal Process; 2009. p. 61–64. [CrossRef]
[33] Osman MM, Büyük O. Parabolic filter mel frequency cepstral coefficient and fusion of features for speaker age classification. Sigma J Eng Nat Sci 2020;38:2177–2191.