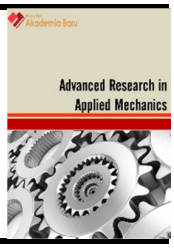
Penerbit Akademia Baru

Journal of Advanced Research in Applied Mechanics

Journal homepage: www.akademiabaru.com/aram.html
ISSN: 2289-7895



A Robust Audio Watermarking Integrating Audio Characteristics and Intrinsic Mode Functions

Open Access

Rawan Foad[1,*], Sherin Youssef[1], Marwa ElShenawy[1]

[1] Department of Computer Engineering, College of Engineering and Technology, AAST Alexandria Egypt

ARTICLE INFO

Article history:
Received 5 June 2017
Received in revised form 4 July 2017
Accepted 4 December 2017
Available online 11 March 2018

ABSTRACT

Digital watermarking technology is a process that has several applications. It is used in copyright protection, copy protection, content authentication, fingerprinting, broadcast monitoring, indication of content manipulation and information carrier. In audio watermarking, data is embedded into the audio signal. In which, embedded data can be extracted or detected from the audio signal. Digital watermarking takes place by embedding the copyright information into a cover object. In other words, digital watermarking is the embedding of an auxiliary piece of information. It takes place by embedding a perceptually transparent digital signature about a signal into the host audio signal itself. Digital signature carries a message. Investigation is held on audio watermarking that uses audio characteristics and Intrinsic Mode Functions (IMFs) for more robustness. After dividing the audio signal into segments, audio features are extracted directly from the acoustical signal that retains only the important distinctive characteristics of the intended audio classes. IMFs are calculated for each frame, which are intrinsic oscillatory components. In the last IMF, watermark is embedded. The watermark is a combination of Synchronization Code (SC) and binary image's bit. Signal to Noise Ratio (SNR) and Bit Error Rate (BER) are calculated to test the performance of the technique used. The used technique proved to be better in performance than other recent audio watermarking researches, so that the hidden watermark is robust to different image sizes and frame sizes.

Keywords:
Audio watermarking, speech, intrinsic mode function, synchronization code

1. Introduction

Copyright owner is authenticated by the knowledge of the key used to read the watermark. Audio watermarking is not only used to identify the ownership of copyright, but it is also used in fingerprinting [1]. Also, it is used in monitoring, where a secret watermark is embedded to enable the tracing of illegal copying [2]. Moreover, it can be used as an information carrier, where Public watermark embedded into the data stream.

Human Auditory System (HAS) is sensitive [3], therefore audio watermarking is a difficult process. Achievement on imperceptibility of auditory system is more difficult than that of the visual

* *Corresponding author.*
E-mail address: eng.rawan.foad@gmail.com (Rawan Foad)

system [4]. Moreover, audio files normally have size much less than that of the video files. Thus [5], the amount of data that can be hidden in video sequences is higher than the amount of data that can be embedded transparently into an audio sequence is considerably.

Audio classification is categorized into physical features and perceptual features. Physical features are directly related to the measurable properties of the acoustical signal and are not linked with human perception. Perceptual features, on the other hand, relate to the subjective perception of the sound, and therefore must be computed using auditory models.

For identifying the copyright owner, the embedded data is later detected or extracted from the multimedia. Furthermore, watermarking research is a masterful topic that attracts a lot of interests as one of the most popular approaches. To prevent the invasion of audio files, enforce owner management, etc., many audio watermarking algorithms are proposed in [5]. The phase of audio watermarking was analysed so that the embedding of the watermark can't happen live. Also watermarking should be affordable by making the extraction easier. Audio watermarking principles, features and performance evaluation techniques are thoroughly discussed. A study by [6] and [7] presented watermarking method based on Empirical Mode Decomposition (EMD). The watermark is embedded in very low frequency mode (last IMF), because the components' energy is greater than that of high frequency components. Therefore, the watermark is inaudible, doesn't alter the audible content and be easy to remove [8]. Watermark is associated with SCs and thus the synchronized watermark has the ability to resist shifting and cropping. Bits of the synchronized watermark are embedded in the extrema of the last IMF of the audio signal based on QIM. Moreover, [7] proposes a technique that includes DES algorithm, which encrypts and decrypts data to maintain optimum security. Other authors in [9] discussed a simple blind audio watermarking using Fast Fourier Transform (FFT). By using a SC, the blind audio watermarking becomes robust to de-synchronization signal processing attacks. After segmenting the audio stream, each audio segment is divided into two parts. A SC is embedded in the first part of the audio segment and watermark is embedded in the second part. Quantization Index Modulation (QIM) is used to embed the encrypted binary watermark image in FFT coefficients in each frame. Results show the comparison of the performance of various audio watermarking techniques such as EMD method.

In this paper a robust audio watermarking algorithm that merges the audio characteristics and the IMFs. Feature extraction is an important signal processing task, where the process of computing the numerical representation from the acoustical signal can be used to characterize the audio segment. After extracting features, data is embedded to each segment using an embedding equation that there parameters are omitted based on trial and error. Section II describes the proposed scheme. In Section III, algorithm is tested using different test cases. Such as frame sizes and different watermark sizes. For performance analysis, SNR and BER are calculated in each case. Section IV is concluding the method and the suggesting future work.

2. The Proposed Robust Audio Watermarking Integrating Audio Characteristic and Intrinsic Mode Functions Scheme

2.1 Embedding of Watermark

2.1.1 Form of data phase

As displayed in Figure 1, in order to apply the scheme the data needed must be prepared. First, the audio signal is assembled by reading the audio file using the desired sampling frequency and resolution. Second, the image is converted into binary image. Then, the combination is ready after concatenating the SC on both sides of the binary image to form the watermark data (SC + Watermark bit + SC).

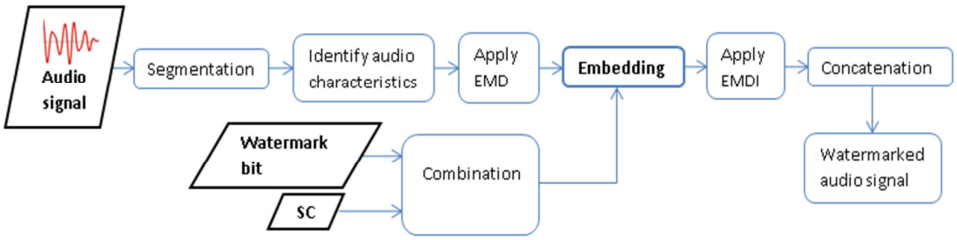


Fig. 1. Embedding technique of a watermark into audio signal

2.1.2 Segmentation phase

International Federation of the Phonographic Industry (IFPI) states that imperceptibility, robustness, payload, and security are properties and characteristics needed for an effective audio watermarking scheme [5]. In order to satisfy the capacity and complexity of watermarking properties, audio signal is divided into N non-overlapping segments with the needed block size. This happens after reading the host signal. By using segmentation the capacity and complexity of watermarking properties are fulfilled.

2.1.3 Identifying audio characteristics phase

For music signals, the physical feature Zero-Crossing Rate (ZCR) is more stable across extended time durations. ZCR is very important for the new scheme, because the number of zero-crossings is needed for the resultant IMFs. So that, the number of zero-crossings and the sum of the maxima and minima (number of IMF extrema) must be equal or differ at most by one. This is one of the requirements that should be satisfied by the retrieved IMFs in the next phase. Also, ZCR provides spectral information at a low cost. The number of times the signal waveform changes sign in the course of the current frame is measured by ZCR and is given by $ZCR = \frac{1}{2}\sum_{n=1}^{N} |sign(x_r(n)) - sign(x_{r-1}(n))|$, where $sign(x) = \begin{cases} 1, & x \geq 0; \\ -1, & x < 0. \end{cases}$. Sinusoid for example is a ZCR For narrowband signals. ZCR is directly related to the fundamental frequency. For more complex signals, the ZCR correlates well with the average frequency of the major energy concentration.

Another necessary physical feature is the Short-Time Energy, because it represents the temporal envelope of the signal. Also, it's the mean squared value of the waveform values in the data frame. More than its actual magnitude, its variation over time can be a strong indicator of underlying signal content. It is computed as $E_r = \frac{1}{N}\sum_{n=1}^{N} |x_r(n)|^2$.

Although pitch is a perceptual attribute, it is closely correlated with the physical attribute of fundamental frequency (F0). Subjective pitch changes are related to the logarithm of F0 so that a constant pitch change in music refers to a constant ratio of fundamental frequencies. Due to the higher channel bandwidths in the high frequency region, several higher harmonics get combined in the same channel. And the periodicity detected then corresponds to that of the amplitude envelope beating at the fundamental frequency. The perceptual PDAs try to emulate the ear's robustness to interference-corrupted signals, as well as to slightly harmonic signals, which still produce a strong sensation of pitch. Fundamental Frequency (F0) is computed by measuring the periodicity of the time-domain waveform. Time-domain periodicity can be estimated from the signal Auto-Correlation Function (ACF) given by $R(\tau) = \frac{1}{N}\sum_{n=0}^{N}(x_r[n]x_r[n+\tau])$. The ACF, R(τ), will exhibit local maxima at the pitch period and its multiples. The fundamental frequency of the signal is estimated as the inverse of the lag "τ" that corresponds to the maximum of R(τ) within a

predefined range. By favoring short lags over longer ones, fundamental frequency multiples are avoided. The normalized value of the lag at the estimated period represents the strength of the signal periodicity and is referred to as the harmonic coefficient.

2.1.4 Applying Empirical Mode Decomposition (EMD) method phase

On each segment, EMD method is applied to decompose the segment generating the IMFs. IMFs are retrieved by subtracting the previously extracted IMF from the original signal, the next IMF is obtained. This process is repeated continuously, until all the IMFs and the residue is extracted. For example, contains no more than two extrema. For instance, a signal of length of elements can be decomposed to five IMFs, where the last IMF is considered as the residue.

2.1.5 Embedding phase

The watermark data is embedded in the maxima and minima of the last IMF per frame. If the bit to be embedded is 1, then the watermark is embedded in the maxima. On the other hand, when the bit to be embedded is 0, the watermark is embedded in the next minima. This is accomplished using the equation $e^* = e + \frac{W * \alpha}{c}$, where c is a constant, α is the embedding factor, W is the watermark. Symbol e is the old or the original value of the maxima or minima of the last IMF in a frame. And, e* is the new value of the maxima or minima of the last IMF.

The image used can significantly affect the audibility quality of the watermarked audio. Hence, the effect of watermark strength (embedding factor) on imperceptibility is considered.

2.1.6 Applying Empirical Mode Decomposition method (EMDI) phase

After the embedding phase EMD Inverse is applied by adding the IMFs up all the IMFs extracted as a result of the decomposition.

2.1.7 Formalizing watermarked audio signal phase

In the final step is that all the frames are concatenated. Finally, write these data into a file to have the watermarked audio signal as an output.

2.2 Extraction and Detection of Watermark
2.2.1 Steps of watermark extraction

The watermarked audio signal is divided into frames, where EMD is applied on each segment to be decomposed into IMFs. Watermark is detected, followed by EMDI appliance.

2.2.2 Extraction key

The key used for a successful extraction of data, is the number of zeros and ones found and their order in the binary image. This is because the bits will be extracted correctly but not in the same order of the binary image. That's why; the watermark must be extracted exactly from the same frames that were used for embedding.

2.2.3 Empirical mode decomposition method (EMDI)

In the extraction phase, the watermark data is extracted from the maxima and minima of the last IMF of each frame using the equation $e^* = e - \frac{W * \alpha}{c}$, where c is a constant, α is the embedding factor, W is the watermark. Symbol e is the old or the original value of the maxima or minima of the last IMF in a frame. And, e* is the new value of the maxima or minima of the last IMF.

This phase occurs continuously for successive frames until all the ones and zeros are found using specific number of frames.

2.2.4 Binary image extraction phase

If the watermark was extracted from maxima, then the detected bit is 1. While, if the watermark was detected from minima, then the extracted bit is 0. As mentioned before, the watermark is the combination of SC and the image bit. Hence, the SC should be removed from both ends of the watermark. Consequently, the only bit left is recognized as the binary image's bit. Later the 1D of bits are converted back to be a normal 2D binary image.

3. Results and Discussion

The used embedding equation's parameters are generated based on trial and error of several experiments. Until the optimum values have been reached, in which, audible distortion is minimal. Based on trial and error, the proposed parameters of the embedding equation proved to have a good balance between robustness, watermarked signal's imperceptibility and payload. The value of the threshold is 1, while the value of the constant is set to 100000. The purpose is that, the original values would be increased or decreased with a reasonable value. The number of frames used in embedding must be same as the number of segments used for extraction. Subjective listening tests are held by human acoustic perception. Also Objective tests are taken by measuring the SNR and BER. Simulations are applied on different audio signals are to prove the effectiveness of our proposed scheme. Best results were found for 30sec wav audio signals with 22050 Hz sampling frequency and 16 as sample resolution (Table 1).

This technique is tested using different sizes of frames and different sizes of watermark. Each audio signal is divided into frames of sizes 64 and frame of size 100. Also, image sizes to be used in testing are 34 x 48 = 1632 bits and 60 x 60 = 3600 bits.

Table 1
Characteristics of the experimental audio signals

Audio file (wav)	Time in seconds	Sampling Frequency in Hz	Length of sampled data	Sample Resolution
Beatles	30	22050	661500	16
Blues	30	22050	661500	16
Jazz	30	22050	661500	16
Pop	30	22050	661500	16
Rock	30	22050	661500	16

In order to have a binary image, the 2D image is converted into 2D binary image, and then it is converted it into 1D sequence to embed it into the audio signal. The SC used is a 16 bit Barker sequence 1111100110101110 [6]. The SC is concatenated to the both sides of the watermark bit. The embedded watermarks are 1111100110101110 1 1111100110101110 or 1111100110101110 0

1111100110101110.

The values of the threshold and constant are set to 1 and 100000 respectively. Based on trial and error, these parameters were chosen. As they have a good compromise between robustness, watermarked signal's imperceptibility and payload.Figure 3 shows the beatles.wav signal and its watermarked version respectively, where the watermarked signal looks very similar to the original one.

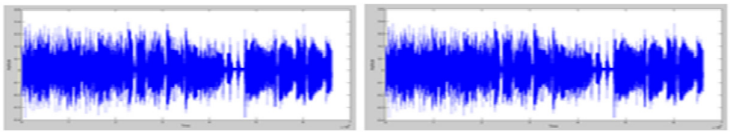


Fig. 3. Audio signal vs Watermarked audio signal of beatles.wav

At frame size 64, after embedding the watermark of size 60 x 60, the SNR and the BER between the original audio signal and the watermarked audio signal were obtained. As found in Table 2, the values of BER were between 0.0102 and 0.2477. While, when 34 x 48 in same frame size, values varied between 0.0103 and 0.2486.

Table 2
BER and SNR for different audio wav files using watermark of size 60 x 60 and 34 x 48 at frame size of 64

Audio wav	SNR using watermark 60 x 60	BER using watermark 60 x 60	SNR using watermark 34 x 48	BER using watermark 34 x 48
Beatles	90.0358	0.2477	89.8363	0.2486
Blues	83.8549	0.0135	83.2031	0.0143
Jazz	90.6992	0.0117	90.2757	0.0112
Pop	95.5532	0.0125	96.3701	0.0120
Rock	91.7706	0.0102	91.4874	0.0103

In Table 3, the results show that at frame size of 100, the BER values were between 0.0057 and 0.0432 when embedding the 60 x 60 watermark. While when embedding watermark 34 x 48 in exactly the same frame size, 0.0059 and 0.0368 were the range of BER's values. In conclusion the frame size doesn't affect the robustness of the signal.

Table 3
BER and SNR for different audio wav files using watermark of size 60 x 60 and 34 x 48 at frame size of 100

Audio wav	SNR using watermark 60 x 60	BER using watermark 60 x 60	SNR using watermark 34 x 48	BER using watermark 34 x 48
Beatles	91.5370	0.0432	91.1270	0.0368
Blues	86.3322	0.0097	85.9523	0.0084
Jazz	92.9298	0.0084	93.0204	0.0070
Pop	97.7790	0.0081	98.5393	0.0066
Rock	94.3649	0.0057	94.1340	0.0059

4. Conclusion

The proposed scheme is a robust audio watermarking that integrates the audio characteristics and the IMFs. Finding features that are invariant to irrelevant transformations and have good discriminative power across the classes was the target. Zero-Crossing Rate, Short-Time Energy and

Fundamental Frequency (F0) are physical features computed at the frame rate from windowed segments of audio. Moreover, perceptual features can be categorized as pitch. These features are essential for the proposed scheme. To decompose the signal into IMFs makes the scheme works with less computational effort. This algorithm uses the SC, in order to make the scheme withstand de-synchronization attacks. Different audio signals were used of type wav with time of 30sec. Each audio is with 16 sampling resolution, sampling frequency of 22050 Hz and 661500 sampled data. Watermark data is a combination of the SC and the binary image. Audio signal is divided into equal segments. In each segment, IMFs are resulted followed by the embedding of the watermark. Then apply the EMD Inverse and concatenate it to the rest of the signal. The subjective and objective analyses are demonstrated. Watermarking scheme offers better robustness and the accuracy of the data extracted. Objective analysis concludes that scheme is strong even when using larger frame and watermark sizes. SNR is above 20 db and follows the IFPI standard. Also, values of BER give good indication of a powerful technique. The algorithm's performance is compared with other state-of-art algorithms. The proposed method indicates that, it is superior in terms of imperceptibility, robustness and payload. For future work, Empirical Wavelet Transform (EWT) can be used instead of the EMD method. This can improve the extraction of different modes of a signal by designing an appropriate wavelet filter bank [10]. This is because experiments show that EWT gives a more consistent decomposition. Where, EMD is sometimes difficult to interpret, due to the retrieval of too much modes. Another reason is that, the classic wavelet formalism can be adapted to understand.

References

[1] G. B. Khatri and D. S. Chaudhari, "Digital Audio Watermarking Applications and Techniques", International Journal of Research in Engineering and Technology (IJRET), vol. 5, 2013, pp. 109–115.

[2] Kulkarni, Hemantkumar, and Chitra Gaikwad. "Comparative Study of some Audio Watermarking Techniques based on DCT, SVD, LWT and EMD Principals."

[3] Olanrewaju, R. F., and Othman Khalifa. "Digital audio watermarking; techniques and applications." In *Computer and Communication Engineering (ICCCE), 2012 International Conference on*, pp. 830-835. IEEE, 2012.

[4] Nosrati, Masoud, Ronak Karimi, and Mehdi Hariri. "Audio steganography: A survey on recent approaches." *world applied programming* 2, no. 3 (2012): 202-205.

[5] Electa Alice Jayarani A., D. Ane Delphin and Ambily Babu, "Analyzing the audio watermarking schemes and features", International Journal of Research in Engineering and Technology (IJRET), vol. 5, 2016, pp. 231–235.

[6] Khaldi, Kais, and A. Boudraa. "Audio watermarking via EMD." *IEEE transactions on audio, speech, and language processing*21, no. 3 (2013): 675-680.

[7] Gawale, Kunal, Harshali Chaudhari, Vasundhara Kandesar, and Smita Sakharwade. "Digital Audio Watermarking using EMD for Voice Message Encryption with Added Security."

[8] Youssef, Sherin M. "HFSA-AW: a hybrid fuzzy self-adaptive audio watermarking." In *Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on*, pp. 1-6. IEEE, 2013.

[9] Lalitha, N. V., Ch Srinivasa Rao, and PVY Jaya Sree. "Robust Audio Watermarking Scheme with Synchronization Code and QIM."

[10] Gilles, Jerome. "Empirical wavelet transform." *IEEE transactions on signal processing* 61, no. 16 (2013): 3999-4010.