IIOAB JOURNAL
ISSN: 0976-3104

ARTICLE

SYNCHRONIZED ROBUST AUDIO WATERMARKING BASED ON FWHT

N.V. Lalitha[1*], Ch. Srinivasa Rao[2], PVY. JayaSree[3]

[1]Dept. of Electronics and Communication Engineering, GMR Institute of Technology, Rajam, A.P. INDIA
[2]Dept. of Electronics and Communication Engineering, JNTU-K University, Vizianagaram, A.P. INDIA
[3]Dept. of Electronics and Communication Engineering, GIT, GITAM University, Visakhapatnam, A.P. INDIA

ABSTRACT

Digital audio watermarking is the key solution to provide copyright protection and copy protection. Many watermarking schemes are available in the literature but these fail either in providing imperceptibility or robustness or synchronization. The watermarking scheme based on Fast Walsh Hadamard Transform (FWHT) with a synchronization code is proposed in this paper. Digital audio is segmented into two parts; synchronization code is inserted in the first part and Gaussian map encrypted watermark is embedded into the FWHT coefficients of digital audio in the second part. The experimental results on a standard database demonstrate that the proposed scheme has imperceptibility of 23.30 dB to 29.83 dB, payload is 1638.4 bps and Gaussian map is used to provide security to the watermark. As well, our watermarking scheme resists desynchronization attacks such as Signal addition, Subtraction, Cropping and Time Scale Modification (TSM) attack up to ±5%. The scheme also withstands echo attack in a better manner when compared with state-of-art schemes.

KEY WORDS
Audio Watermarking, Synchronization, Fast Walsh Hadamard Transform, Time Scale Modification

Published: 2 December 2016

***Corresponding Author**
Email:
lalithanarla.ece@gmailcom
lalitha.nv@gmrit.org
Tel.: +91-9985001587

INTRODUCTION

Digital audio watermarking can be used successfully for audio copyright protection and copy protection. Based on state-of-art, audio watermarking techniques are broadly classified into two types: time domain [1] and frequency domain techniques [2]. Time domain techniques are simple but less robust to signal processing attacks. Frequency domain techniques are somewhat complex and it provides robustness to signal processing attacks [3]. Synchronization attacks are the major issue in digital audio watermarking [4], ex: cropping, signal addition, signal subtraction and Time Scale Modification (TSM). Many state of art audio watermarking algorithms are available but have not reported synchronization attack [5-8] and few algorithms have reported synchronization attack [9-13]. Major contribution of our work is to provide robustness against desynchronization attacks and mainly concentrated on TSM attack. Robustness is also provided to signal processing attacks and more light is thrown on echo-attack. In this paper, audio is segmented; synchronization code and watermark both are inserted in each segment. Fast Walsh Hadamard Transform (FWHT) [14] is used for insertion of scrambled watermark image with the help of Quantization Index Modulation (QIM) [15-16] method to provide blind extraction. Due to FWHT, computational complexity is reduced [17] and it is much faster than the FFT.

MATERIALS

Walsh-Hadamard transform (WHT) is a non-sinusoidal, orthogonal transformation technique. The signal is decomposed with a set of basis functions called Walsh functions which are rectangular or square waves with values of +1 or −1. WHT returns sequency values. The fast version of WHT is Fast Walsh-Hadamard Transform (FWHT). FWHT is faster to calculate because it uses only real additions and subtractions requires less storage space, while the FFT requires complex values and more storage space. The FWHT is able to represent signals with sharp discontinuities more accurately using fewer coefficients than the FFT. Due to its symmetric property, the calculation process of both FWHT and the inverse FWHT is similar. The FWHT and IFWHT for a signal x(n) of length N are defined in equations (1)-(2):

$$y(n)=\frac{1}{N}\sum_{i=0}^{N-1} x(i)WAL(n,i) \qquad \text{... (1)}$$

$$x(n)=\frac{1}{N}\sum_{i=0}^{N-1} y(i)WAL(n,i) \qquad \text{... (2)}$$

where i = 0,1, ..., N – 1 and $WAL(n,i)$ is Walsh function.

PROPOSED AUDIO WATERMARKING METHOD

In this proposed work, total audio ‘A’ is divided into segments for synchronization code and watermark insertion. With the help of logistic chaotic sequence, synchronization code is generated and it is inserted at the starting of each segment. For increasing the security, watermark must be pre-processed with the help of Gaussian map encryption method.

Synchronization code generation

To overcome the problem of synchronization attacks in audio watermarking process, synchronization code must be inserted before the watermark. In the literature, synchronization code is generated using Bernoulli shift map [9], Piece-Wise Affine Markov (PWAM) map [12], Barker code [18-19]. In this paper, to generate L_{syn} length synchronization code, logistic chaotic method is used.

$$L_{n+1} = \gamma L_n(1-L_n) \quad \text{... (3)}$$

Where L_n is the initial value in the range 0< S_n <1 and γ is the real parameter.

$$S_n = \begin{cases} 1 & if\ L_n > 0.5 \\ 0 & otherwise \end{cases} \quad \text{... (4)}$$

Where S_n is the synchronization code and n is varies from 1 to L_{syn}.

Watermark image pre-processing

In this paper, $M \times M$ size watermark image is pre-processed with the help of Gaussian map chaotic encryption method, to increase the security and it is defined as follows:

$$P_{n+1} = e^{(-\alpha P_n^2)} + \beta \quad \text{... (5)}$$

Where P_n is the initial value in the range of 0 to 1. α and β are the real parameters.

$$G_n = \begin{cases} 1 & if\ P_n > \frac{1}{4} \\ 0 & otherwise \end{cases} \quad \text{... (6)}$$

$M \times M$ size watermark image is converted to one-dimensional vector W_n and this is encrypted with G_n with the help of below equation.

$$E_n = G_n \oplus W_n \quad \text{... (7)}$$

Watermark embedding process

The L_{syn} length synchronization code followed by $\mathrm{M} \times \mathrm{M}$ size pre-processed watermark image is embedded into the audio signal is shown in [Fig.1].
Embedding procedure steps are given below:
Step 1: Digital audio signal is segmented and each segment length depends on size of the image and length of the synchronization code.
Step 2: Each segment is again divided into two parts.
Step 3: Synchronization code is generated and inserted directly into the first part, pre-processed watermark is embedded into FWHT coefficients of the second part, using QIM method is as follows:

$$e(i) = \begin{cases} round\left[\frac{x(i)}{Q}\right]Q, & if\ E_i = 0 \\ (floor\left[\frac{x(i)}{Q}\right]Q) + \frac{Q}{2}, & if\ E_i = 1 \end{cases} \quad \text{... (8)}$$

Where $x(i)$ is direct audio sample for part 1, Fast Walsh Hadamard transformed audio coefficient for part2, Q is the embedding strength and $e(i)$ is the embedded coefficient of corresponding audio coefficient.
Step 4: For part 2 Inverse Fast Walsh Hadamard Transform is applied, resultant part1 and part 2 segments are combined to get watermarked audio segment.
Step 5: Repeat the steps 2 to 4 for all the segments and combined them to get watermarked audio.

THE IIOAB JOURNAL
Guest editors: N. Arun Kumar & P. Mohammed Shakeel

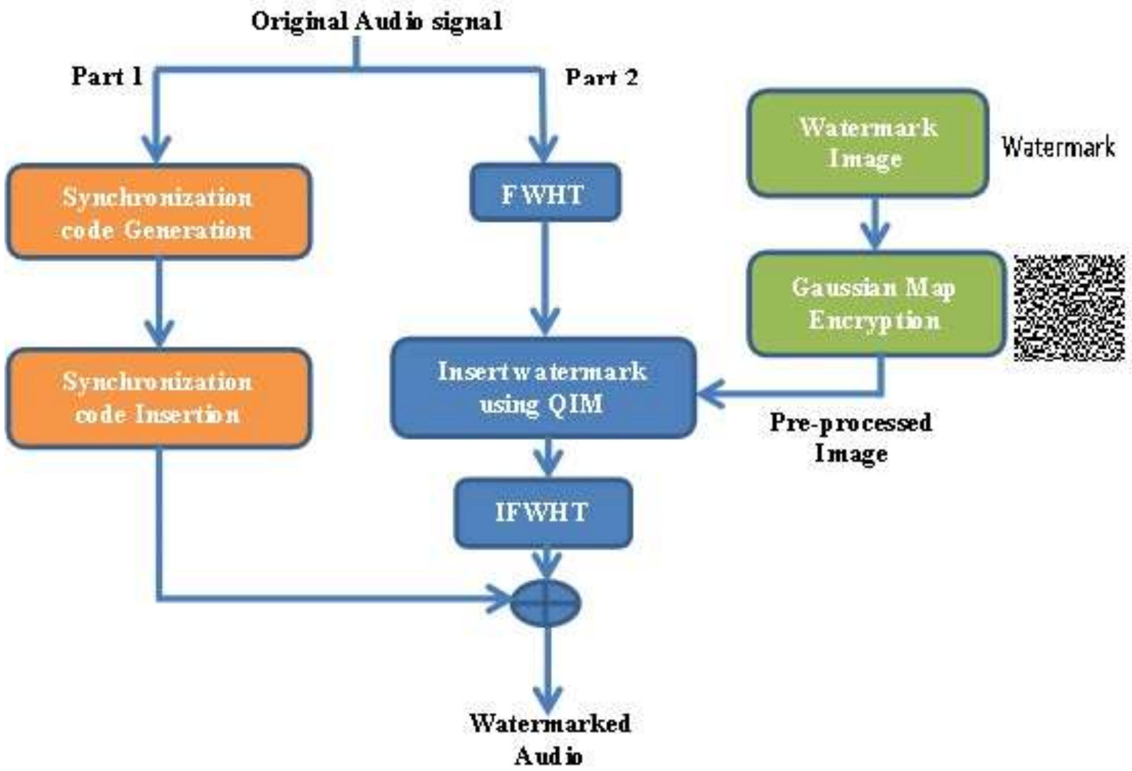

Fig. 1: Process flow of Watermark and Synchronization Code Embedding.

Watermark extracting process

De-synchronization attack dislocates the watermark. To identify the correct position of watermark, synchronization code is inserted into the time-domain. The watermark extraction steps are given below:
Step 1: Search for the synchronization code in the watermarked audio with the help of below equation.

$$S_n^{'} = \begin{cases} 1 & if \ \frac{Q}{4} \le x^{'}(n) \le \frac{3Q}{4} \\ 0 & otherwise \end{cases} \quad \text{... (9)}$$

Where $S_n^{'}$ is the extracted synchronization code and $x^{'}(n)$ is attacked watermarked audio.
Step 2: Calculate the similarity between extracted and original synchronization code with the help of below equation.

$$NC = \frac{\sum_{k=1}^{n} S_n(k) S_n^{'}(k)}{\sqrt{\sum_{k=1}^{n} S_n(k)^2} \sqrt{\sum_{k=1}^{n} S_n^{'}(k)^2}} \quad \text{... (10)}$$

If the similarity is above the threshold, the samples after the synchronization code contains valid watermark.
Step 3: Apply FWHT on those samples of size $M \times M$.
Step 4: Extract the binary sequence from the FWHT coefficients.

$$E_n^{'} = \begin{cases} 1 & if \ \frac{Q}{4} \le \mathrm{mod}(y^{'}(n), Q) < \frac{3Q}{4} \\ 0 & otherwise \end{cases} \quad \text{... (11)}$$

Where $y^{'}(n)$ is FWHT coefficient and $E^{'}(n)$ is the extracted binary sequence.
Step 5: Apply Gaussian map decryption process, to extract the binary image.

RESULTS

Four different audio signals of 10 seconds i.e. pop, rock, jazz, folkcountry [20] are considered to evaluate the performance of the proposed algorithm based on FWHT. The frequency and quantization rates of mono

Guest editors: N. Arun Kumar & P. Mohammed Shakeel

audio signals are 44,100 KHz and 16-bits per sample respectively. 128 X 128 binary image is considered as a watermark image. 128 bit synchronization code is inserted to resist the de-synchronization attacks.
To assess the performance of the proposed algorithm, two performance metrics are used i.e. imperceptibility and robustness.

Imperceptibility

Imperceptibility means perceptual quality measure. SNR is used to assess the imperceptibility between original audio and embedded audio.

$$SNR = \frac{\sum_{n=1}^{L} x^2(n)}{\sum_{n=1}^{L} (x(n) - x'(n))^2} \quad \text{... (12)}$$

Where $x(n)$ is the original audio signal, $x'(n)$ is the embedded audio signal and L is the length of the audio. It is evident from [Table 1] that SNR ranges from 23.3080 dB to 29.8364 dB for the four signals and meets the IFPI requirement.

Table 1: SNR values for different audio signals

Type of an audio signal	SNR in dB
POP Music	23.3080
ROCK Music	24.7160
JAZZ instrumental audio	24.8193
FOLKCOUNTRY	29.8364

Robustness

Robustness means ability to extract the watermark even after the attacks are applied on watermarked audio signal. Generally, these attacks are signal processing operations. Some signal processing attacks that are evaluated in this paper are: (a) Resample: Watermarked audio sampling frequency 44.1KHz is sampled to 22.05 KHz and then resampled to 44.1 KHz. (b) Lowpass filter: Watermarked signal is passing through lowpass filter with cut-off frequency 16 KHz. (c) Requantization: 16-bit watermarked audio is first quantized with 8-bit then requantized back with 16-bit. (d) Random Noise: Random noise with 40dB SNR is applied to Watermarked signal. (e) & (f) Additive White Gaussian Noise: 60dB and 50dB AWGN is applied. (g), (h) and (i) Jittering: one sample is removed in every 100000, 50000 and 10000. (j) Invert: Watermarked audio sample amplitudes are inverted.
To calculate the robustness over signal processing attacks, a performance measure BER is used with the help of below equation.

$$Bit\ Error\ Rate\ (BER) = \frac{Number\ of\ error\ bits}{Total\ number\ of\ bits} \quad \text{... (13)}$$

Table 2: BER for four classes of audio signals

Attack / Audio	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
POP	0.2363	0.2473	0.0040	0	0	0.0412	0	0	0.4542	0
ROCK	0.3397	0.3739	0.0051	0	0	0.0427	0	0	0.4915	0
JAZZ	0.2415	0.1750	0.0039	0.00006	0	0.0445	0	0	0.4238	0
FOLKCOUNTRY	0.3480	0.4333	0.0050	0	0	0.0432	0	0	0.4894	0

[Table 2] shows the robustness in terms of BER for different attacks mentioned as (a) to (j).
Some de-synchronization attacks are also applied to check the resistance of the algorithm over these. These de-synchronization attacks are signal addition, signal subtraction, cropping (starting, middle, ending), Time Scale Modification (-5% to +5%). In our work, more concentration is paid for de-synchronization attacks and the performance in terms of BER, Precision and CC are given in [Table 3].

Table 3: Robustness for Desynchronization Attacks

Type of Audio	Type of attack	Without Synchronization			With Synchronization		
		BER	Precision	CC	BER	Precision	CC
POP	Signal addition	0.2007	0.7993	0.2698	0	1	1
	Signal subtraction	0.2007	0.7993	0.2698	0	1	1
	Start cropping	0.1630	0.8370	0.3595	0	1	1

	Middle cropping	0	1	1	0	1	1
	End cropping	0	1	1	0	1	1
	TSM (-5%)	0.5049	0.4951	0.0043	0.4235	0.5765	0.0445
	TSM (-4%)	0.4980	0.5020	-0.0007	0	1	1
	TSM (-3%)	0.4969	0.5031	-0.0051	0	1	1
	TSM (-2%)	0.4969	0.5031	-0.0051	0	1	1
	TSM (-1%)	0.4969	0.5031	-0.0051	0	1	1
	TSM (+1%)	0.4969	0.5031	-0.0051	0	1	1
	TSM (+2%)	0.4997	0.5003	0.0012	0	1	1
	TSM (+3%)	0.4988	0.5012	0.0090	0	1	1
	TSM (+4%)	0.4950	0.5050	-0.0048	0	1	1
	TSM (+5%)	0.5065	0.4935	-0.0055	0.3333	0.6667	0.1385
ROCK	Signal addition	0.0356	0.9644	0.7263	0	1	1
	Signal subtraction	0.0358	0.9642	0.7257	0	1	1
	Start cropping	0.0244	0.9756	0.7816	0	1	1
	Middle cropping	0	1	1	0	1	1
	End cropping	0	1	1	0	1	1
	TSM (-5%)	0.4999	0.5001	-0.0096	0.3383	0.6617	0.1314
	TSM (-4%)	0.4999	0.5001	-0.0096	0	1	1
	TSM (-3%)	0.4999	0.5001	-0.0096	0	1	1
	TSM (-2%)	0.4999	0.5001	-0.0096	0	1	1
	TSM (-1%)	0.4999	0.5001	-0.0096	0	1	1
	TSM (+1%)	0.4981	0.5019	0.0046	0	1	1
	TSM (+2%)	0.5087	0.4913	-0.0038	0	1	1
	TSM (+3%)	0.5016	0.4984	-0.0071	0	1	1
	TSM (+4%)	0.4972	0.5028	-0.0036	0	1	1
	TSM (+5%)	0.4993	0.5007	0.0062	0.3766	0.6234	0.1082
JAZZ	Signal addition	0.0786	0.9214	0.5598	0	1	1
	Signal subtraction	0.0786	0.9214	0.5599	0	1	1
	Start cropping	0.0566	0.9434	0.6302	0	1	1
	Middle cropping	0	1	1	0	1	1
	End cropping	0	1	1	0	1	1
	TSM (-5%)	0.5043	0.4957	-0.0083	0.3998	0.6002	0.0813
	TSM (-4%)	0.4987	0.5013	-0.0075	0	1	1
	TSM (-3%)	0.4960	0.5040	-0.0089	0	1	1
	TSM (-2%)	0.4966	0.5034	-0.0028	0	1	1
	TSM (-1%)	0.4997	0.5003	-0.0052	0	1	1
	TSM (+1%)	0.4997	0.5003	-0.0052	0	1	1
	TSM (+2%)	0.4997	0.5003	-0.0052	0	1	1
	TSM (+3%)	0.4997	0.5003	-0.0052	0	1	1
	TSM (+4%)	0.4987	0.5013	-0.0075	0	1	1
	TSM (+5%)	0.4926	0.5074	-0.0021	0.3867	0.6133	0.0768
FOLKCOUNTRY	Signal addition	0.0997	0.9003	0.5156	0	1	1
	Signal subtraction	0.0997	0.9003	0.5156	0	1	1
	Start cropping	0.0908	0.9092	0.5199	0	1	1
	Middle cropping	0	1	1	0	1	1
	End cropping	0	1	1	0	1	1
	TSM (-5%)	0.5027	0.4973	0.0079	0	1	1
	TSM (-4%)	0.5016	0.4984	-0.0071	0	1	1
	TSM (-3%)	0.5016	0.4984	-0.0071	0	1	1
	TSM (-2%)	0.5016	0.4984	-0.0071	0	1	1
	TSM (-1%)	0.5016	0.4984	-0.0071	0	1	1
	TSM (+1%)	0.5016	0.4984	-0.0071	0	1	1
	TSM (+2%)	0.5001	0.4999	0.0090	0	1	1
	TSM (+3%)	0.4969	0.5031	-0.0051	0	1	1
	TSM (+4%)	0.5001	0.4999	0.0122	0	1	1
	TSM (+5%)	0.4962	0.5038	0.0065	0.3507	0.6493	0.1138

In [9], the authors analyzed TSM attack upto ± 1%, TSM attack upto +4% in [10] and upto ±1% in [12] is reported. As well the robustness against echo attack is compared with other state-of-art methods and is tabulated in Table 4.

Table 4: Echo attack analysis w.r.t BER

Echo Attack	Existing Methods	Proposed
Delay 100ms Decay 50 %	Ali Al-Haj (2014) [13] -- 0.748	0.0299
Delay 100ms Decay 40 %	Ali Al-Haj (2014) [13] -- 0.731	0.0268
Delay 10ms Decay 10 %	B.Lei et al. (2012) [9] -- 0.007	0.0118
Delay 98ms Decay 41 %	V.Bhat K at al (2010) [11] -- 0.020	0.0292
Delay 1s Decay 10 %	B.Lei et al. (2013) [21] -- 0	0

Data Payload

Data payload means the number of bits that are embedded and extracted from the audio stream. It is measured in bits per second (bps). In this paper, 128 X 128=16384 bits are embedded into 10 sec audio signal. So, data payload in this scheme is 1638.4 bps.

CONCLUSION

Many of the watermarking schemes fail either in providing imperceptibility or robustness or synchronization. To overcome this, a watermarking scheme based on Fast Walsh Hadamard Transform (FWHT) with a synchronization code is proposed in this paper. Sequency and symmetric properties of FWHT are exploited to make the scheme to be more robust with less computational effort. The experimentation is performed on a standard database and imperceptibility of 23.30 dB to 29.83 dB and a payload of 1638.4 bps are achieved. Gaussian map is used to provide security to the watermark. The use of synchronization code can resist desynchronization attacks such as Signal addition, subtraction, Cropping and Time Scale Modification (TSM) attack up to ±5%. The scheme is also robust to echo attack in a better manner when compared with state-of-art schemes.

CONFLICT OF INTEREST
None

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None

REFERENCES

[1] Bassia P, Pitas I, Nikolaidis N. [2001] Robust audio watermarking in the time domain, IEEE Transactions on Multimedia, 3(2):232-241.
[2] Wu S, Huang J, Huang D, Shi YQ. [2005] Efficiently self-synchronized audio watermarking for assured audio data transmission, IEEE Trans. Broadcasting, 51(1):69-76.
[3] Al-Haj A, Mohammad A, Bata L. [2011] DWT-Based Audio Watermarking, The International Arab Journal of Information Technology, 8(3):326-333.
[4] Bhat KV, Sengupta I, Das A. [2010] An adaptive audio watermarking based on the singular value decomposition in the wavelet domain, Digital Signal Processing, 20:1547-1558.
[5] Dhar PK, Shimamura T. [2014] Audio watermarking in transform domain based on singular value decomposition and Cartesian-polar transformation, International Journal of Speech Technology, 17:133-144.
[6] Fallahpour M, Megias D. [2014] Secure logarithmic audio watermarking scheme based on the human auditory system, Multimedia Systems, 20:155-164.
[7] Wang J, Healy R, Timoney J. [2011] A robust audio watermarking scheme based on reduced singular value decomposition and distortion removal, Signal Processing, 91:1693-1708.
[8] Hu HT, Hsu LY. [2016] Incorporating Spectral Shaping Filtering into DWT-Based Vector Modulation to Improve Blind Audio Watermarking, Wireless Personal Communications, :1-20.
[9] Lei B, Soon IY, Zhou F, Li Z, Lei H. [2012] A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition, Signal Processing, 92:1985-2001.
[10] Lei BY, Soon IY, Li. [2011] Blind and robust audio watermarking scheme based on SVD-DCT, Signal Processing, 91:1973-1984.
[11] M.Fallahpour, D.Megias, [2015] Audio Watermarking Based on Fibonacci Numbers, IEEE/ACM Transactions on Audio, Speech and Language Processing, 23(8):1273-1282.
[12] Lei B, IYSoon, Tan E, [2013] Robust SVD-Based Audio Watermarking Scheme With Differential Evolution Optimization, IEEE Transactions on Audio, Speech and Language processing, 21(11):2368-2378.
[13] Ali Al-Haj, [2014] A dual transform audio watermarking algorithm, Multimedia Tools and Applications 73: 1897-1912.
[14] Beauchamp KG. [1984] Applications of Walsh and Related Functions: with an Introduction to Sequency Theory, Academic Press, :295–300, London
[15] Bhat KV, Sengupta I, Das A.[2008] Audio Watermarking Based on Quantization in Wavelet Domain. In: Lecture Notes in Computer Science,5352:235–242.
[16] Chen B, Wornell G. [2001.] Quantization Index Modulation: A class of provably good methods for digital watermarking and information embedding, IEEE Transactions on Information Theory, 47:1423-1443,
[17] Dhavale SV, Deodhar RS, Patnaik LM. [2011] Walsh Hadamard Transform Based Robust Blind Watermarking for Digital Audio Copyright Protection, Communications in Computer and Information Science, 250:469-475.
[18] Zhao H, Wang F, Chen Z, Liu J. [2014] A Robust Audio Watermarking Based on SVD-DWT, Elektronika Ir Elektrotechnika, Vol.20, No.1, pp. 75-80,.
[19] Wang XY, Zhao H, [2006] A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT. IEEE Transactions on Signal Processing, 54(12):4835-4840.
[20] http://www-ai.cs.uni-dortmund.de/audio.html
[21] B Lei, IY Soon, Ee-Leng T. [2013] Robust SVD-Based Audio Watermarking Scheme with Differential Evolution Optimization. IEEE Transactions on Audio, Speech and Language Processing, 21(11).