Integration of Quantization Watermarking and Amplitude-Thresholding Compression for Digital Audio Signal in the Wavelet Domain

Ming Zhao[1], Xindi Tong[2], and Jie Li[1]
(Corresponding author: Jie Li)

School of computer science, Yangtze University[1]
Jingzhou, Hubei province, China
Department of Mathematics and Information Engineering, The Chinese University of Hong Kong[2]
Email: 1104112257@qq.com

(Received Sept. 2, 2020; Revised and Accepted Jan. 10, 2021; First Online Apr. 24, 2021)

Abstract

Due to the advancement of technology and the rapid development of the Internet, digital information transmission has skyrocketed, and its acquisition and dissemination have become easier. Without the legal owner's permission, digital information is often stolen or turned into profit by illegal persons. This study proposes a combination of quantization watermarking and amplitude-thresholding compression technology for digital music (or audio) based on discrete wavelet transform (DWT). This technology is expected to reduce the carrying capacity of network transmission while protecting personal copyrights. Moreover, it is resistant to various malicious attacks.

Keywords: Digital Audio Watermarking; Compression; Discrete Wavelet Transform (DWT)

1 Introduction

An audio watermarking technology usually consists of the embedding and extraction techniques and satisfies three minimum requirements of audio watermarking standards set by the International Federation of Phonographic Industry (IFPI) requirements. According to the requirements of the IFPI, an audio watermarking technology should have three specifications [4, 5, 7, 8, 10, 11, 15, 16]:

1) Audio watermark should be imperceptible of original signal;

2) Signal-to-noise ratio (SNR) needs to be higher than 20 dB and the embedding capacity should be more than 20 bits-per-second (bps);

3) Watermark should be capable of resisting common attacks.

Internet development not only brings a lot of convenience but also relative risk. How to reduce the carrying amount of nature data and the hidden information in network transmission at the same time is an important issue. Audio compression technology is to sample or quantize digital audio to reduce the amount of audio data in order to save the time required for file storage and the communication bandwidth required for data transmission. The compressed audio quality must also be at a certain level. There are many types of audio compression technologies, including: MP3, WMA, WAV, turnpoint, threshold compression and so on. Some of these compression technologies can also be implemented in combination with transform domain based on their characteristics. In [18], authors proposed a new idea of integrating electrocardiogram watermarking and compression approach, which has never been researched before. ECG watermarking can ensure the confidentiality and reliability of a user's data while reducing the amount of data. In the evaluation, they apply the embedding capacity, bit error rate (BER), signal-to-noise ratio (SNR), compression ratio (CR), and compressed-signal to noise ratio (CNR) methods to assess the proposed algorithm. After comprehensive evaluation, the final results show that their algorithm is robust and feasible.

In this study, we integrate quantization watermarking technology with amplitude-thresholding compression for digital music (or audio) based on discrete wavelet transform (DWT). First of all, we perform DWT on each audio signal to embed private information into DWT lowest coefficients. Then, we obtain watermarked audios by inverse discrete wavelet transform (IDWT). At the same time, we adopt the amplitude-threshold compression to reduce the data amount of the embedded audio signal. In addition, the hidden information can be extracted without the original audio signal and the recovery of the compressed audio signal adopts cubic spline. In experiments, we evaluate the appropriate threshold ε, embedding strength Q, and the robustness against various malicious attacks.

The rest of this paper is organized as follows. Section 2 reviews some preliminaries for later use. Section 3 presents the proposed integration of quantization watermarking and amplitude-thresholding compression for digital audio signal in the wavelet domain. Section 4 shows experimental results. Conclusions are drawn in Section 5.

2 Preliminaries

In this section, we review some preliminaries including Discrete Wavelet Transform, Cubic spline, and performance measurement for audio watermarking and signal compression.

2.1 Discrete Wavelet Transform

The wavelet transform maps a function in $L^2(R)$ onto a scale-space plane. Wavelets are obtained by a single prototype function (mother wavelet) $\psi(x)$ which is regulated with a scaling parameter and a shift parameter [3,12,13]. The discrete normalized scaling and wavelet basis function are defined as

$$\begin{aligned}\varphi_{i,n}(t) &= 2^{i/2} h_i \varphi(2^i t - n).\\ \psi_{i,n}(t) &= 2^{i/2} g_i \psi(2^i t - n).\end{aligned}$$

where i and n are the dilation and translation parameters; h_i and g_i are the low-pass and high-pass filters. Orthogonal wavelet basis functions not only provide simple calculation in coefficients expansion but also span $L^2(R)$ in signal processing. As a result, any audio signal $S(t) \in L^2(R)$ can be expressed as a series expansion of orthogonal scaling functions and wavelets. More specifically,

$$S(t) = \sum_{\ell} c_{j_0}(\ell)\varphi_{j_0,\ell}(t) + \sum_{k}\sum_{j=j_0}^{\infty} d_j(k)\psi_{j,k}(t),$$

where

$$c_j(\ell) = \int_R S(t)\varphi_{j,\ell}(t)dt$$

and

$$d_j(k) = \int_R S(t)\psi_{j,k}(t)dt$$

denote the sequences of low-pass and high-pass coefficients, respectively; j_0 be the integer to define an interval on which $S(t)$ is piecewise constant. Throughout this paper, the host digital audio signal $S(n), n \in N$, denoting samples of the original audio signal $S(t)$ at the n-th sample time, is cut into segments where DWT will be preformed. This can be done by exploiting orthogonal basis to implement DWT through filter bank. Figure 1 demonstrates how the input digital audio signal $S(n)$ is segmented into eight non-overlapping multi-resolution subbands by the seven-level DWT decomposition.

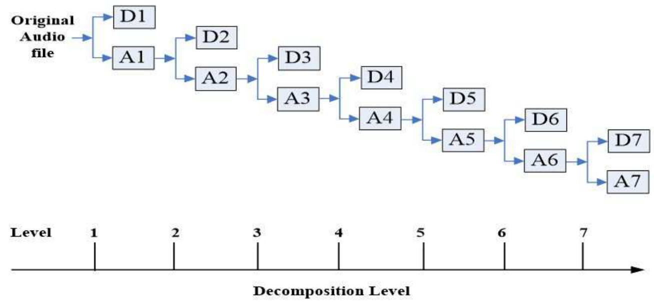

Figure 1: Seven-level discrete wavelet transformation

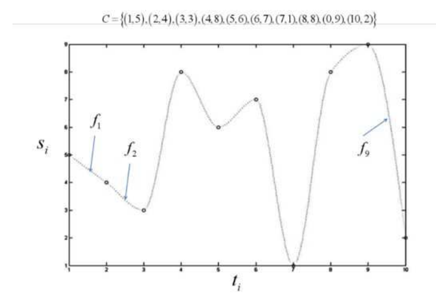

Figure 2: Cubic spline interpolation

2.2 Cubic Spline Interpolation

For a given dataset $C = \{(t_0, s_0), (t_1, s_1), ..., (t_N, s_N)\}$, the cubic function is formulated as [1,9,14]

$$f_i(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3$$

Found that the N cloud gauge line collection of functions $\{f_i(t) | i = 1, ..., N\}$ as shown in Figure 2 to describe the entire set of data, where $f_i(t)$ must satisfy

$$\begin{aligned} f_i(t_i) &= s_i = f_{i-1}(t_i),\\ f'_i(t_i) &= f'_{i-1}(t_i),\\ f''_i(t_i) &= f''_{i-1}(t_i),\\ f''_1(t) &= f''_N(t) = 0\end{aligned}$$

2.3 Performance Measurement of Audio Watermarking and Signal Compression

In general, transparency is the key performance of audio watermarking. It is measured by signal-to-noise ratio (SNR) which are defined as follows [4,5,7,8,10,11,15,16]:

$$SNR = -10\log\left(\frac{\sum_{i=1}^{N}(\bar{s}_i - s_i)^2}{\sum_{i=1}^{N} {s_i}^2}\right)$$

In addition, we also apply relative root mean square error (rRMSE) and root mean square error (RMSE) to

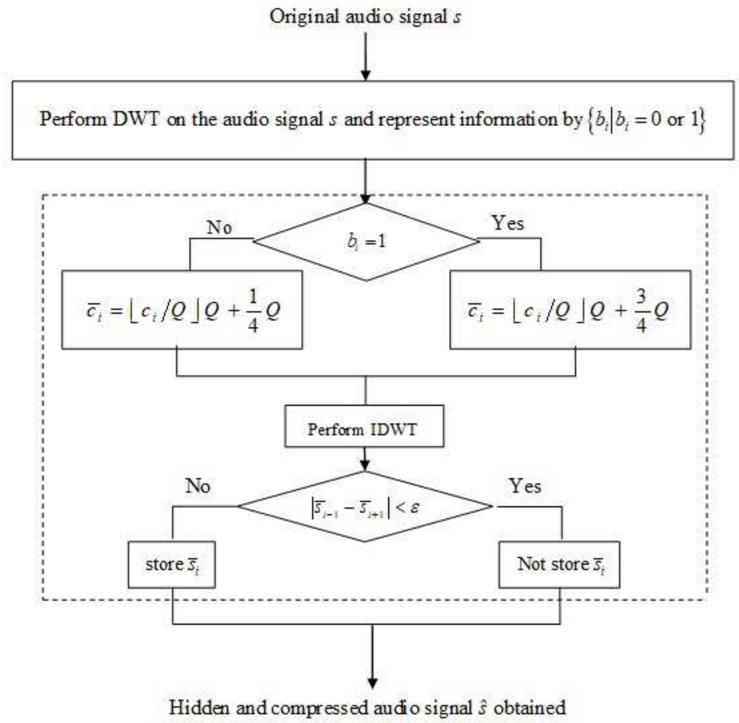

Figure 3: Flow chart of integration of watermarking and compression

judge the transparency, which is defined as follows:

$$\text{rRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\bar{s}_i - s_i}{s_i}\right)^2}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\bar{s}_i - s_i)^2}$$

where $\{s_i\}$ represents the original ECG signal for ECG, $\{\bar{s}_i\}$ represents the hidden (or modified) ECG signal.

In order to evaluate the quality of signal compression, compression ratio (CR) and percentage ratio difference (PRD) are utilized. Assume s is the original signal and $\breve{s}$ is the reconstruction from the compressed signal, then CR and PRD are defined as [2,6,17]

$$\text{CR} = \frac{\text{Datasizebeforecompression}}{\text{Datasizeaftercompression}}$$

$$\text{PRD} = \sqrt{\frac{\sum_{i=1}^{N}\left(\breve{s}_i - s_i\right)^2}{\sum_{i=1}^{N}{s_i}^2}} \cdot 100$$

where N is the number of testing samples in the signal s.

3 Proposed Method

In order to hide information into an audio and reduce the carrying amount on it when transmitting in the Internet, this section presents the proposed method integrating quantization watermarking and amplitude-thresholding compression for digital audio signal in the wavelet domain. Figures 3 and 4 show the flowchart of the proposed integration. The detail is introduced in the following.

3.1 Watermarking and Compression

In order to embed private information into audio signals conveniently, we utilize binary bits $\{b_i|b_i = 0 or b_i = 1\}$ to represent information that will be hidden, and then embed these binary bits to DWT coefficients of audio signals by quantization embedding technique which is proposed as follows. We use 7-level Haar DWT to decompose an ECG signal into eight non-overlapping sub-bands. Figure 1 shows the structure of the 7-level DWT decomposition. Taking into account the robust performance of the low-pass filtering, we embedded the watermark (binary bits $\{b_i|b_i = 0 or b_i = 1\}$) into the sub-band coefficients in the 7th level, which are the lowest-frequency coefficients. The embedding rule is based on the quantization technique [14,17].

$$\bar{c}_i = \begin{cases} \left\lfloor {}^{c_i}\!/_{Q} \right\rfloor Q + {}^{3Q}\!/_{4}, \text{if} b_i = 1 \\ \left\lfloor {}^{c_i}\!/_{Q} \right\rfloor Q + {}^{Q}\!/_{4}, \text{if} b_i = 0 \end{cases}$$

where c_i and $\{\bar{c}_i\}$ are the 7th low-frequency DWT coefficients before and after embedding, respectively; Q is the embedding strength; By applying the IDWT, the watermarked audio signal $\bar{S}$ is obtained.

To reduce the amount of data when transmitting in the Internet, we compress the embedded audio signal $\bar{s}$ by a

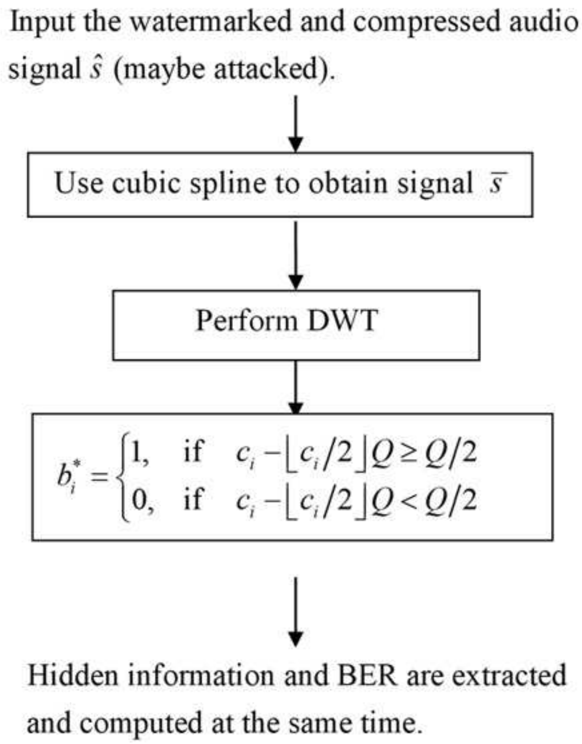

Figure 4: Flow chart of integration of watermarking and compression

threshold compression method which is as follows:

$$\hat{s}_0 = \bar{s}_0, \hat{s}_N = \bar{s}_N$$

$$\hat{s}_i = \begin{cases} \phi & \text{if } |\bar{s}_{i-1} - \bar{s}_{i+1}| < \varepsilon \\ \bar{s}_i & \text{otherwise} \end{cases}, i = \{1, \cdots N-1\}$$

where ε represents the threshold.

For the extraction of the hidden confidential information, we first recover the signal $\{\bar{s}_i\}_{i=0}^N$ from the compressed signal $\{\hat{s}_i\}_{i=0}^N$ by using the cubic function which is formulated as

$$f_i(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3$$

Found that the N cloud gauge line collection of the functions $\{f_i(t) \,|i = 1, ..., N\}$ to describe the entire set of data, where $f_i(t)$ must satisfy

$$f_i(t_i) = \hat{s}_i = f_{i-1}(t_i), f'_i(t_i) = f'_{i-1}(t_i),$$
$$f''_i(t_i) = f''_{i-1}(t_i), f''_1(t) = f''_N(t) = 0$$

Next, we extract the hidden information from the DWT coefficients $\{\bar{c}_i\}_{i=0}^N$ of the recovered audio signal $\{\bar{s}_i\}_{i=0}^N$ according to the following rules:

$$b_i = \begin{cases} 1, \text{if} \bar{c}_i - \left\lfloor {}^{\bar{c}_i}\!/_{Q} \right\rfloor Q \geq {}^{Q}\!/_{2} \\ \\ 0, \text{if} \bar{c}_i - \left\lfloor {}^{\bar{c}_i}\!/_{Q} \right\rfloor Q < {}^{Q}\!/_{2} \end{cases}$$

where Q is the same as the embedding strength (or secret key) in embedding; $b_i = 1$ or $b_i = 0$ is extracted binary bits or the embedded information equivalently.

4 Experimental Results

The evaluation of the proposed method is discussed in this section. Two types of audio signals, love song, folklore, and dance, are to be tested. These audio signals are 16-bit mono-type of length 11.6 seconds and sampling rate 44.1kHz.

From the results in Table 1, two observations are discussed. First, for the same threshold value, strong embedding strength has better compression due to the fact that the variation of the overall audio is small when the embedding strength is strong. As a result, the SNR, rRMSE and RMSE are worse. Second, for different thresholds, we found that CR value is increased when the threshold value is greater than the embedding strength. Restate, compression is better when the threshold value is greater than the embedding strength.

Common attacks are carried out after the embedding process with $Q = 6500$ and compression with $\varepsilon = 100, 500, 1000$. Based on the robustness is evaluated with the BER, three forms of attacks that apply to the audio signals will be explained in detail below.

1) Re-sampling: The sample rate, the number of samples of audio carried per second. The procedure converts an audio signal from a given sample rate to a different sample rate. In the proposed algorithm, watermarked audio signals are first decimated from 44.1kHz to 22.05kHz, and then interpolated to the original 44.1kHz. This step repeated two more times from 44.1kHz to 11.025kHz and 8kHz, and then back up to 44.1kHz. The BER under the re-sampling at-

Table 1: Performance test by different thresholds and quantization sizes

Audio	ε	Q	CR	PDR	SNR	BER	rRMSE	RMSE
love song	100	100	1.3889	1.2777	37.8712	1.0986	0.3735	70.3929
		500	2.0398	3.7400	28.5427	3.4668	0.5296	206.1828
		1000	3.2663	6.8321	23.3089	3.7119	0.6330	377.9388
		3000	5.6264	12.6983	17.9251	4.0049	0.7711	711.0507
		6500	12.1732	24.5075	10.5972	4.5974	0.9972	1.3591e+003
	500	100	3.7372	4.2549	27.4223	3.6387	1.0696	234.4087
		500	2.0398	3.7400	28.5427	3.4668	0.5296	206.1828
		1000	3.2663	6.8321	23.3089	3.7119	0.6330	377.9388
		3000	5.6264	12.6983	17.9251	4.0049	0.7711	711.0507
		6500	9.1636	23.5035	11.5974	4.5176	0.9972	1.3591e+003
	1000	100	15.8760	19.6754	14.1215	4.7373	5.4751	1.0840e+003
		500	8.0630	8.5770	21.333	4.4932	1.2893	472.8477
		1000	3.2663	6.8321	23.3089	4.7119	0.6330	377.9388
		3000	5.6264	12.6983	17.9251	5.0049	0.7711	711.0507
		6500	9.4136	23.5035	11.5174	5.5176	0.9972	1.3591e+003
dance	100	100	1.3676	0.9204	40.7209	1.0254	0.3558	48.5834
		500	2.2934	3.3630	29.4656	2.3701	0.4141	177.6400
		1000	3.6868	6.3754	23.9098	2.5176	0.5336	337.8759
		3000	6.7927	12.8518	17.8207	3.1279	0.7803	686.8794
		6500	9.8462	23.7338	11.4627	3.3965	0.9554	1.3129e+003
	500	100	5.0135	6.3998	23.8767	3.2734	2.0738	337.8303
		500	2.2934	3.3630	29.4656	4.3701	0.4141	177.6400
		1000	3.6868	6.3754	23.9098	5.5176	0.5336	337.8759
		3000	6.7927	12.8518	17.8207	5.1279	0.7803	686.8794
		6500	9.2482	23.7338	11.4327	5.3965	0.9554	1.3129e+003
	1000	100	13.8847	43.4587	7.2385	5.3965	12.3850	2.2941e+003
		500	9.7062	16.2784	15.7678	5.8105	2.2083	859.8685
		1000	3.6868	6.3754	23.9098	5.5176	0.5336	337.8759
		2048	6.7927	12.8518	17.8207	6.1279	0.7803	686.8794
		6500	9.2462	23.7338	11.4925	6.3965	0.9554	1.3129e+003

Table 2: BER (%) IN THE RE-SAMPLING ATTACK

Audio Type		Love Song			Dance		
Re-sampling Rate (kHz)		22.05	11.03	8	22.05	11.03	8
BER(%)	$\varepsilon = 100$	7.75	19.04	18.28	14.25	28.35	27.55
	$\varepsilon = 500$	9.75	20.04	20.28	16.25	29.85	29.76
	$\varepsilon = 1000$	10.84	23.12	23.18	16.97	30.47	30.65

Table 3: BER (%) IN THE LOW-PASS FILTERING ATTACK

Audio Type		Love Song		Dance	
Cut-off frequency (kHz)		3	6	3	6
BER(%)	$\varepsilon = 100$	40.64	25.13	36.24	21.22
	$\varepsilon = 500$	41.25	27.83	37.52	24.67
	$\varepsilon = 1000$	41.13	30.62	37.54	29.73

Table 4: BER (%) IN NOISE ATTACK

Audio Type		Love Song				Dance			
Noise in dB		-40	-30	-20	-15	-40	-30	-20	-15
BER(%)	$\varepsilon = 100$	6.84	9.14	14.26	17.78	5.91	9.21	14.46	17.52
	$\varepsilon = 500$	9.54	13.58	17.39	19.13	8.46	12.72	17.42	18.92
	$\varepsilon = 1000$	12.84	15.62	20.15	22.43	11.79	16.38	19.56	21.83

tacks are shown in Table 2. The data confirms that the proposed design results in a lower BER.

2) Low-pass filtering: A low-pass filter is a circuit that provides easy passage to low-frequency signals and difficult passage to high-frequency signals. Table 3 shows the BER information with a low-pass filter and cutoff frequencies at 3kHz and 6kHz.

3) MP3 compression: MP3 compression is generally used to reduce file sizes. The bit rate of an MP3 file is a measure of the audio signal in a given period of time. Usually, the larger the bit rate, the better the sound quality. Table 4 contain experimental data while applying MP3 compression at different bit rates to the watermarked audio.

5 Conclusions

In this study, we propose the integration technology of the audio-signal quantization watermarking and amplitude-thresholding compression. The integration technology not only protect the security of private information but also reduce the amount of audio data transmission. We evaluate the appropriate the relationship between threshold and embedding strength Q. Furthermore, we test the robustness against common attacks. The future work is to find the optimal and Q between the CR and SNR.

References

[1] D. Boor and Carl, "A practical guide to splines," *Applied Mathematical Sciences New York Springer*, vol. 27, no. 149, pp. 157–157(1), 1978.

[2] M. Brito, J. Henriques, P. Gil, and M. Antunes, "A predictive adaptive approach to generic ecg data compression," in *IEEE International Workshop on Intelligent Signal Processing*, pp. 32–37, 2005.

[3] C. S. Burrus, R. A. Gopinath, and H. Gao, *Introduction to Wavelet Theory and Its Application*, New Jersey: Prentice-Hall, 1998.

[4] S. T. Chen, H. N. Huang, C. J. Chen, and G. D. Wu, "Energy-proportion based scheme for audio watermarking," *IET Signal Processing*, vol. 4, no. 5, pp. 576–587, 2010.

[5] S. T. Chen, G. D. Wu, and H. N. Huang, "Wavelet-domain audio watermarking scheme using optimization-based quantization," *IET Signal Processing*, vol. 4, no. 6, pp. 720–727, 2010.

[6] H. H. Chou, Y. J. Chen, Y. C. Shiau, and T. S. Kuo, "A high performance compression algorithm for ecg with irregular periods," in *IEEE International Workshop on Biomedical Circuits and Systems*, pp. S2/4–9–12, 2004.

[7] F. Djebbar and B. Ayad, "Energy and entropy based features for WAV audio steganalysis," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 1, pp. 168–181, 2017.

[8] S. Gupta and N. Dhanda, "Audio steganography using discrete wavelet transformation (DWT) and discrete cosine transformation (DCT)," *IOSR Journal of Computer Engineering*, vol. 17, no. 2, pp. 32–44, 2015.

[9] C. A. Hall and W. W. Meyer, "Optimal error bounds for cubic spline interpolation," *Journal of Approximation Theory*, vol. 16, no. 2, pp. 105–122, 1976.

[10] M. C. Lee and C. Y. Lau, "Three orders mixture algorithm of audio steganography combining cryptography," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 4, pp. 959–969, 2018.

[11] H. Liu, J. Liu, X. Yan, P. Xue, S. Wan, and L. Li, "Centroid-based audio steganography scheme in wavelet domain," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 5, pp. 1222–1232, 2018.

[12] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 4, 1989.

[13] MathWords, *Dwt: Single-level 1-D Discrete Wavelet Transform*, Apr. 18, 2021. (`https://www.mathworks.com/help/wavelet/ref/dwt.html`)

[14] MathWords, *Spline: Cubic Spline Data Interpolation*, Apr. 18, 2021. (`https://www.mathworks.com/help/matlab/ref/spline.html`)

[15] R. M. Noriega, M. Nakano, B. Kurkoski, and K. Yamaguchi, "High payload audio watermarking: Toward channel characterization of MP3 compression," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 91–107, 2011.

[16] P. Parnami, K. Niwariya, and M. Jain, "Performance evaluation of DWT and LSB based audio steganography," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 3, no. 12, pp. 167–170, 2016.

[17] S. G. Miaou, S. N. Chao, "Wavelet-based lossy-to-lossless ECG compression in a unified vector quantization framework.," *IEEE Transactions on Bio Medical Engineering*, vol. 52, pp. 539–543, 2005.

[18] K. K. Tseng, X. He, W. M. Kung, S. T. Chen, M. Liao, and H. N. Huang, "Wavelet-based watermarking and compression for ecg signals with verification evaluation," *Sensors*, vol. 14, no. 2, pp. 3721–3736, 2014.

Biography

Ming Zhao received the Master degrees in computer aided Instruction from Huazhong Normal University, China in 2006, and he got Ph.D. in computer science and technology from Harbin Institute of Technology, China in 2015. He is currently an associate professor in Yangtze University, China. His research interests include computational intelligence, image and signal processing, pattern recognition etc. He is an IEEE Senior Member.

Xindi Tong started her college life in 2019. She is currently a year-2 student in The Chinese University of Hong Kong. Her research interests includes Natural language processing, signal processing, artificial intelligence etc.

Jie Li received the Bachelor Degree in School of Computer Science from Yangtze University, China in 2020, and He is currently a graduate student at Yangtze University. His research interests include computational intelligence, few-shot learning, pattern recognition etc.