Multi-bit Grouping Audio Watermarking Algorithm Based on Phase Shifting

Shengbei Wang, Yuqing Yan

Abstract— **Watermarking technology has been widely used in copyright protection of video, audio, and image. In this paper, a multi-bit grouping audio watermarking method based on a frame-wise framework is proposed, which utilizes the property of phase spectrum distribution. At first, the phase spectrum of each frame is calculated by applying Fourier Transform. In according with the embedded bit number of each frame, the phase bins are segmented into several subsets. A specific phase pattern is constructed in one of these subsets to carry a specific multi-bit watermark. In the process of watermark extraction, a specific phase pattern is identified in each audio frame to extract the watermark information during the watermark embedding. We evaluated the performance of the proposed method and the results showed that the proposed method achieved satisfactory performance.**

Index Terms— **Grouping audio watermarking, Phase distribution, Phase shifting, Multi-bit embedding.**

I. INTRODUCTION

With the development of digital multimedia devices and t- echnology, a large number of multimedia files such as audio and video are circulated over the Internet [1][2]. Watermark- ing is a technique through which the secure information can be carried without degrading the quality of the original signal [1]. The audio watermarking technique consists of two steps: watermark embedding and watermark extract. In the process of watermark embedding, a unique binary watermark is inserted into the audio signals in a manner that is difficult to perceive. In the process of watermark extract, watermarks are calculated to verify copyright information [3].

The design of watermark algorithms should consider the following aspects: robustness, inaudibility, capacity and security [3]. It can be considered a good watermark algorithm when meeting the above four metric. Robustness entails sensitivity to common attacks and lossy compression, ensuring that watermark information remains intact when under attack. Inaudibility requires that the audio signal with a watermark is nearly indistinguishable from the original audio signal, making it difficult for the human visual system (HVS) [4] to perceive any differences. Capacity refers to the amount of watermark information that can be embedded in the audio signal. Security involves the watermark's ability to remain undetectable to unauthorized parties and resist hostile attacks. [4].

Manuscript received December 15, 2023

Shengbei Wang, School of software, Tiangong University, Tianjin, China.

Yuqing Yan, School of software, Tiangong University, Tianjin, China.

II. LITERATURE REVIEW

In recent years, many watermarking methods have been proposed, which mainly include two categories: time-do main [5] methods and frequency-domain methods [6]. In ti- me domain, there are lots of methods, such as least significa- nt bit replacement [5] and echo hiding [7][8]. In frequency domain, the methods usually use some transformations such as discrete cosine transform, discrete wavelet transform and fast Fourier transform for watermark embedding. For exam- ple, Chen et al. [9] proposed an audio watermarking scheme using minimum-amplitude scaling on the optimized lowest -frequency coefficients in the wavelet domain. This method utilized the Karush-Kuhn-Tucker (KKT) theorem to mini- mize the difference between the original and the water- marked coefficients to modify the low-frequency amplitude for watermark embedding [9]. The Fourier transfrom is a widely employed technique that facilitates the decompos- ition of audio signals into frequency components. This transformative method holds substantial significance in the design of audio watermark algorithm. e.g., a method takes advantage of the translation-invariant property of the FFT coefficients [10] to embed watermarks, which resists small distortions in the time domain. Better perceptual quality and less computational burden are achieved in this method. In addition, the phase information is found to be effective in designing robust and inaudible watermarking in the audio. Adaptive magnitude modulation and phase modulation [11] were combined to improve the robustness of the watermarking. An audio watermarking method based on phase modulation was proposed to resist band-pass filtering attack [12]. According to the characteristic of cochlear delay, the phase of audio signal was controlled to embed inaudible watermarks [13][14]. A method such as direct spread spectrum (DSS) [15][16] is also included in frequency domain.

III. PROPOSED METHOD

The majority audio watermarking methods are imple- mented on the basis of frame-wise framework. In this paper, we proposed a multi-bit grouping audio watermarking method based on this framework. Our approach enables the embedding of multiple bits of watermarks in a single frame by applying grouping mechanism. To achieve this, we divide the audio signal into non-overlapping frames and apply the fast Fourier transform to calculate its Fourier spectrum. As a result, the magnitude spectrum and phase spectrum can be obtained, where all phase bins fall within the range of $(-\pi,\pi]$. We have plotted the phase distribution and cumulative

distribution function (CDF) of phase bins of varying frame

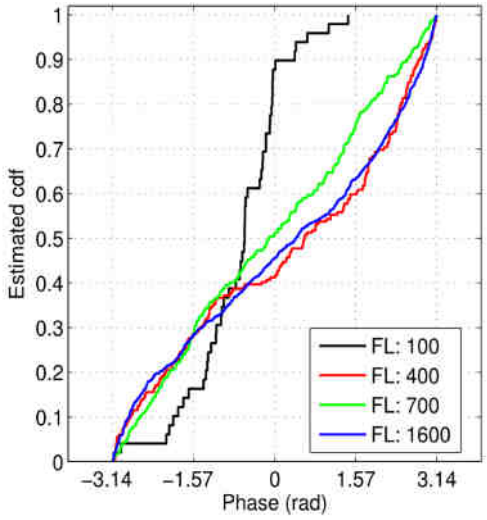

Figure 1. Cumulative distribution function(CDF) of phase bins for audio signals of different frame lengths

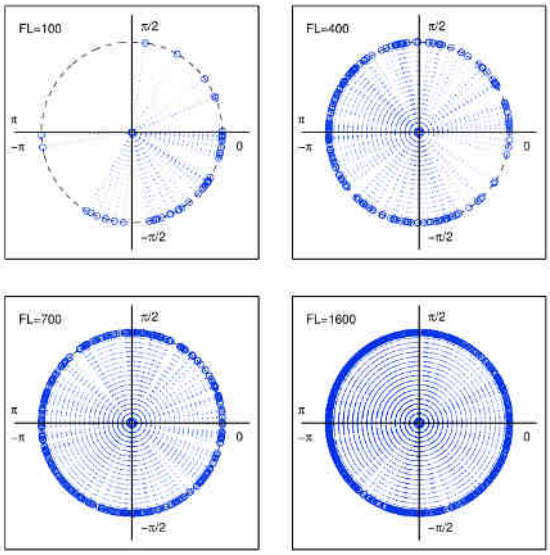

Figure 2. Phase distribution of phase bins for audio signals of different frame lengths

lengths. As depicted in Fig. 1, the cumulative distribution becomes more stable as the frame length increases. In Fig. 2, the distribution of phase bins appears highly random for short audio frames but approaches uniformly as the frame length increases. From our analysis, it can be observed that the phase bins of all frequency components are randomly scattered between $-\pi$ and π in a natural audio signal. Furthermore, when the audio frame is sufficiently long, all phase bins tend to approximate a uniform distribution.

We took advantage of this phase distribution property to construct a specific phase pattern for multi-bit grouping watermark embedding.

A. *Watermark embedding*

We embed M-bit ($M > 1$) watermarks into each frame. The M binary bits, denoted as $\{b_{M-1},\cdots,b_m,\cdots,b_0\}$, $b_{\{\cdot\}} \in \{0, 1\}$, indicate K possible bit combinations, where $K = 2^M$. Next, we divide the watermarks into two parts. The first part is $\{b_{M-1},\cdots,b_m,\cdots,b_1\}$, which determines the number of the phase subsets. The full phase range $(-\pi, \pi]$ is equally divided into T subranges, where $T = 2^{\wedge(M-1)}$). i.e., each subranges covers $2\pi / T$ [rad]. The T subranges are denoted as $R = \{R_0, \cdots, R_t, \cdots, R_{T-1}\}$, respectively, where R_t is

$$R_t \equiv (t\times\frac{2\pi}{T},(t+1)\times\frac{2\pi}{T}],0\le t\le\frac{T}{2} \quad (1)$$

$$R_t \equiv (t\times\frac{2\pi}{T}-2\pi,(t+1)\times\frac{2\pi}{T}-2\pi],\ \frac{T}{2}\le t\le T-1 \quad (2)$$

The T phase subsets are able to carry T possible bit combinations indicated by M-1 bits. The second part of the watermarks is b_0, which determines how to shift the phase bins in phase subsets.

To embed watermark b_0, at first, we convert the (M-1) bit watermarks, $\{b_{M-1},\cdots,b_m,\cdots,b_1\}$, to a decimal number using $t = \sum_{m=0}^{M-2} b_m \times 2^m$. This decimal number decides in which phase subsets for the phase shifting. In fact, when shifting the phase bins, the phase subsets are reduced to half of the previous phase subsets. And each new phase subset are divided into two parts, denoted as $\delta1$ and $\delta2$. $\delta1$ and $\delta2$ ($\delta1 = \delta2$), two smaller subsets, are used to select the phase bins to shift. The degree size of $\delta1$ and $\delta2$ are half of the new phase subset. $\delta1$ and $\delta2$ are calculated as

$$\delta_1 = \delta_2 = \frac{2\pi}{T}\times\frac{1}{4} \quad (3)$$

The new phase subsets, denoted as R_t^*, are calculated as

$$R_t^* \equiv (t\times\frac{2\pi}{T}+\delta_1,(t+1)\times\frac{2\pi}{T}-\delta_2],0\le t\le\frac{T}{2} \quad (4)$$

$$R_t^* \equiv (t\times\frac{2\pi}{T}-2\pi+\delta_2,(t+1)\times\frac{2\pi}{T}-2\pi-\delta_1],\frac{T}{2}\le t\le T-1 \quad (5)$$

We set two fixed phase values $\{C_1, C_2\}$ within R_t*. C_1 is located in $\delta1$. And C_2 is located in $\delta2$. For the convenience of the calculation for C_1 and C_2, we firstly define three variables, which are λ_1, λ_2 and λ_3. λ_1. λ_2 and λ_3 are determined values. The size of λ_1 is a quarter of the phase subset R_t. The size of λ_2 is half of the phase subset R_t. The size of λ_3 is equal to the phase subset R_t. And the three variables are related to M, denoting as $\pi / 2^{\wedge(M+1)}$, $\pi / 2^{\wedge(M-1)}$, $\pi / 2^{\wedge(M-2)}$, respectively. kk is a numerical value that we define. If C_1 is calculated, then $kk = 2$. If C_2 is calculated, then $kk = 1$. Based on the above parameter regulations, C_1 and C_2 are calculated as

$$C_1 = (kk+1)\times\lambda_1+(t-1)\times\lambda_3,0\le t<T/2 \quad (6)$$

$$C_2 = \lambda_1+kk\times\lambda_2+(t-1)\times\lambda_3,0\le t<T/2 \quad (7)$$

$$C_1 = (kk+1)\times\lambda_1+(t-1)\times\lambda_3-2\pi,T/2\le t<T \quad (8)$$

$$C_2 = \lambda_1+kk\times\lambda_2+(t-1)\times\lambda_3-2\pi,T/2\le t<T \quad (9)$$

The division of phase subsets is shown in Fig. 3(b), Fig. 3(c), and Fig. 3(d). And we need to adjust the phase bins between $\delta1$ and $\delta2$ to achieve the phase shift. As shown in Fig. 4(b), Fig. 4(c), and Fig. 4(d). The phase bins in a subset of $\delta1$ are shifted to match the fixed phase value C_2, which are present in a subset of $\delta2$, represented as '0'. Similarly, the phase bins in a subset of $\delta2$ are shifted to match the fixed phase value C_1, found in a subset of $\delta1$, represented as '1'. By performing the mutual shifting of phase bins in $\delta1$ and $\delta2$, it

can present two different phase patterns. Therefore, the watermark b0 can be embedded.

To increase robustness, only phase bins that have relatively high magnitude are selected from the new phase subsets of R_t^* for watermark embedding. A ratio, $\rho (0 < \rho < 1)$, is used to calculate the percentage of selected phase bins, e.g., if the total phase bins in R_t^* is T, then $\lceil T \times \rho \rceil$ high-magnitude phase bins will be selected, where $\lceil . \rceil$ is the ceiling function.

To increase inaudibility, only phase bins within the phase subrange of R_t^* are selected. The change of phase subsets' range is shown in Fig. 4. (a). The phase subsets R_t^* after changing become half of the original phase subsets R_t. We can summarize the contrast for the change of phase subsets by referring to Fig. 3(a) and Fig. 4(a).

After manipulation, the phase spectrum is combined with the original magnitude spectrum to construct the water-marked audio frame.

Example of phase shifting for (b) $M = 2$, (c) $M = 3$, (d) $M = 4$ are shown in Fig. 3. In Fig. 3(c), the phase range $(-\pi,\pi]$ is d-ivided into $T = 4$ subranges, i.e., $R = \{R_0, R_1, R_2, R_3\}$, where $R_0 = (0,\pi/2]$, $R_1 = (\pi/2,\pi]$, $R_2 = (-\pi,-\pi/2]$, and $R_3 = (-\pi/2,0]$. Each subrange is responsible for carrying one type of bit combination (“00”, “01”, “10”, “11”). Each phase subset is evenly divided into two smaller subsets, $\delta 1$ and $\delta 2$. In accordance with the embedded watermark bit $\{b_0\}$, the selected phase bins in a determined subrange are shifted to C_1 or C_2. If b_0 is ‘0’, the phase bins will be shifted to the fixed point of C_2, if b_0 is ‘1’, the phase bins will be shifted to the fixed point of C_1. Note that the above manipulations change the natural distribution of phase bins. However, such artifacts can hardly be discovered from watermarked audio signals when ρ is very small, which ensures the security of the proposed method.

B. *Watermark extraction*

Watermark extraction is affected with not only the number of phase bins in subsets but also the ratio ρ. The extract method is as follows.

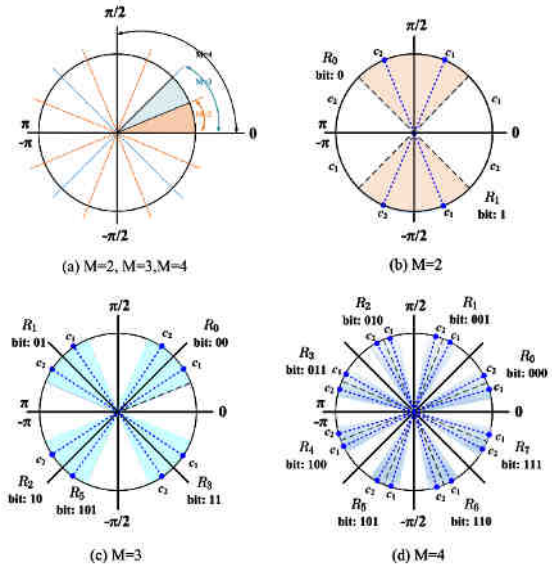

Fig. 3. Phase subsets and the setting of fixed phase values in each subset for multi-bit grouping watermark embedding.

We need to sequentially traverse and detect each phase subset.

We calculated the number of high-magnitude phase bins in every subset, denoted as β_t, where t represents the index of the phase subset. t starts from 0. The number of high magnitude phase bins in two smaller subsets $\delta 1$ and $\delta 2$ is calculated, denoted as β^1_t and β^2_t. This calculation is applied to every quadrant. Subsequently, we can obtain two ratios in each subset: β^1_t / β_t and β^2_t / β_t. In the end, we obtain a number of ratios that are twice the number of phase subsets. The maximum ratio is selected to determine the embedding position of the watermarks, which can ensure the watermarks $\{b_{M-1}, \cdots, b_m, \cdots, b_0\}$. The maximal ratio, i.e.,

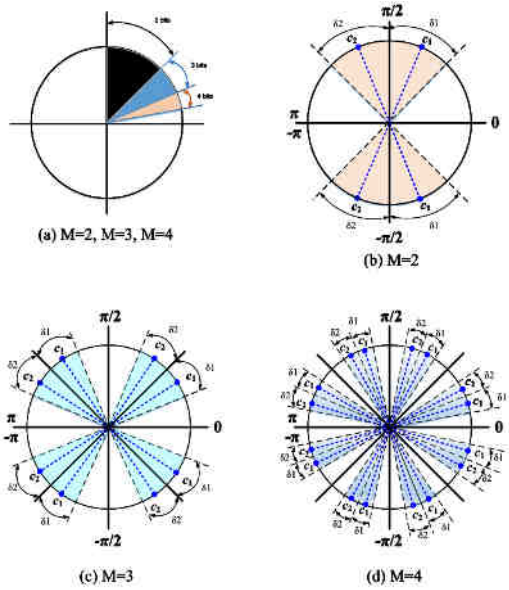

Fig. 4. Phase subsets of calculating the number of phase bins and the division of the phase subsets for $M = 2$, $M = 3$, $M = 4$.

$$W^* = \arg \max_{0 \le t \le T-1} \frac{(\bigcup_{i=1}^{2} \beta_t^i)}{\beta_t} \tag{10}$$

Watermarks can be decoded by converting $W^* / 2$ into binary code.

IV. EXPERIMENTAL RESULT

The proposed method was evaluated with respect to the inaudibility and robustness. Twenty audio signals (44.1-kHz and 16-bit) drawn from the Real World Computing dataset [17] were used in the experiment. Inaudibility was measured by log-spectrum distortion (LSD) [18] and perceptual evaluation of audio quality (PEAQ) [Objective Difference Grade, ODG] [19]. The LSD threshold was 1.0 dB, where a lower value indicates lower distortion. The PEAQ threshold was 1.0 ODG (perceptible but not annoying), where a higher value indicates better quality. Robustness was measured by bit detection rate (BDR) [%]. A higher BDR indicates stronger robustness.

A. *Robustness performance*

Robustness was affected by M and ρ. By conducting experimental verification, we find the best parameters.

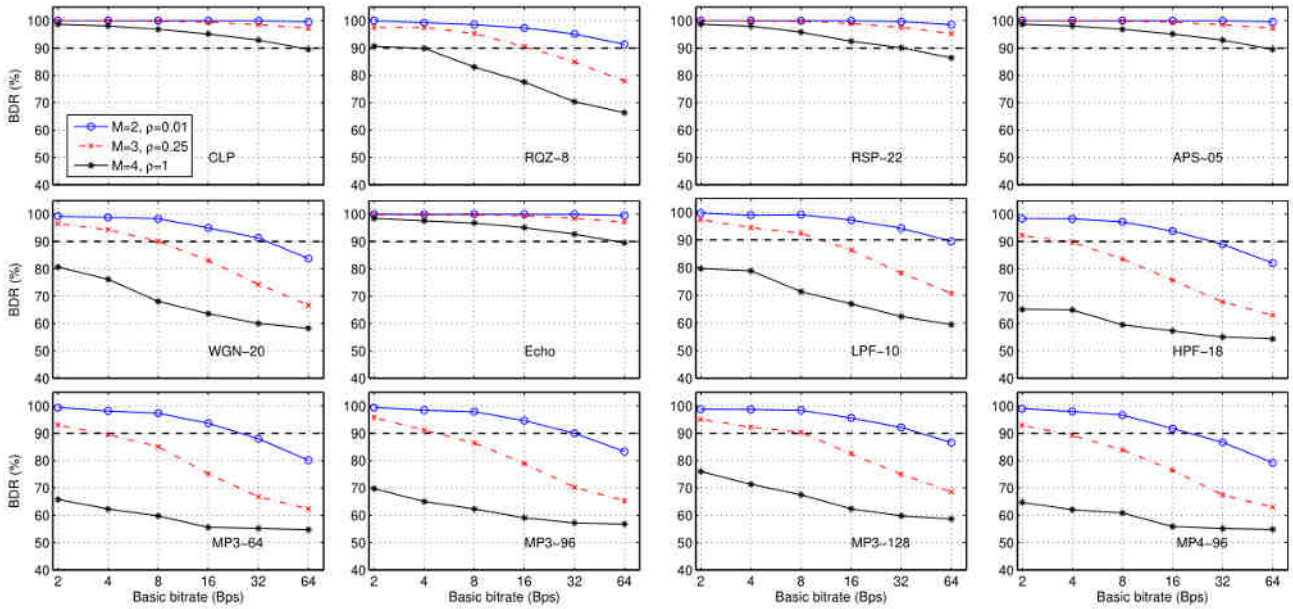

Fig. 5. Robustness of proposed method for $M = 2$, $\rho = 0.01$, $M = 3$; $\rho = 0.25$ and $M = 4$, $\rho = 1$, where straight dashed line indicates bit detection rate (BDR) of 90%. Embedding capacity was given in basic bitrate and real embedding capacities for M =2, M =3 and M =4 were two, three and four times that of basic bitrate, respectively.

Robustness was investigated from three aspects: (1) normal extraction, i.e., closed-loop (CLP), (2) common attacks: requantization with 8 bits (RQZ-8), resampling at 22 kHz (RSP-22), amplitude scaling by 0.5 (APS-05), WGN of 20 dB (WGN-20), echo addition with decay of 0.05 and delay of 50 ms (Echo), lowpass filtering with a cutoff frequency of 10 kHz (LPF-10), and high pass filtering with a cutoff frequency of 18 kHz (HPF-18), and (3) lossy compression: MP3 64 kbps (MP3-64), 96 kbps (MP3-96), 128 kbps(MP3-128), and MP4 96 kbps (MP4-96).

The robustness results for settings ($M = 2$, $\rho = 0.01$; $M = 3$, $\rho = 0.25$; $M = 4$, $\rho = 1$) are plotted in Fig. 5. The proposed method had a high BDR in CLP for all embedding capacity. It also showed satisfactory robustness against most operations with $M = 2$. In contrast, the robustness for $M = 3$ and $M = 4$ at a high bitrate degraded after several operations, especially for $M = 4$. Overall, when $M = 2$, $\rho = 0.01$, the robustness performance is better than $M = 3$ and $M = 4$;

B. *Inaudibility performance*

The inaudibility is primarily influenced by the range of the phase subsets and the shifting degree for the phase bins.

The inaudibility results for $M = 2$, $M = 3$, and $M = 4$ are plotted in Fig. 5, where the straight dashed lines indicate the thresholds of LSD (1.0 dB) and PEAQ (-1.0 ODG). The inaudibility was enhanced as M increased, and the shifting degree of phase bins weakened with the increase of M. It can be seen from the Fig. 6(e) and Fig. 6(f), the inaudibility results of $M = 4$ had an overall improvement compared with those of $M = 2$ in Fig.6 (a) and Fig. 6(b). From these results, the proposed method exhibited satisfactory inaudibility performance, except for $M = 2$, $\rho = 0.01$, at a high bitrate. Especially, regardless of the value of ρ, the inaudibility performance is excellent when $M = 4$.

There is a good trade-off between robustness and inaudibility when $M = 3$.

V. CONCLUSION

We proposed a multi-bit grouping audio watermarking method based on phase shifting. The grouping of watermarks plays a crucial role in the process. Additionally, the range of phase subsets is important when shifting the phase bins. We leverage the inherent distribution characteristics of phase bins in each frame to construct a specific phase pattern for watermark embedding. Our future work aims to enhance both inaudibility and robustness in higher-bit watermark embedding methods. We strive to achieve a better balance between inaudibility and robustness through an even more refined mechanism.

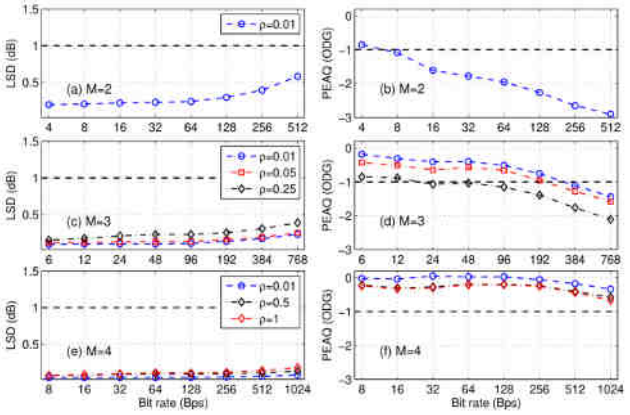

Fig. 6. Inaudibility of proposed method for $M = 2$, $M = 3$ and $M = 4$.

REFERENCES

[1] RD Shelke and Milind U Nemade, "Audio watermarking techniques for copyright protection: A review," in 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC). IEEE, 2016, pp. 634 – 640.
[2] V Neethu and R Kalaivani, "Efficient and robust audio watermarking for content authentication and copyright protection," in 2016 international conference on circuit, power and computing technologies (ICCPCT). IEEE, 2016, pp. 1 – 6.
[3] In-Kwon Yeo and Hyoung Joong Kim, "Modified patchwork algorithm: A novel audio watermarking scheme," IEEE Transactions on speech and audio processing, vol. 11, no. 4, pp.381–386, 2003.
[4] Martin Kutter and Fabien AP Petitcolas, "Fair benchmark for image watermarking systems," in Security and watermarking of multimedia contents. SPIE, 1999, vol. 3657, pp. 226–239.
[5] Paraskevi Bassia, Ioannis Pitas, and Nikos Nikolaidis, "Robust audio watermarking in the time domain," IEEE Transactions on multimedia, vol. 3, no. 2, pp. 232 – 241, 2001.
[6] Mehdi Fallahpour and David Megias, "Robust high-capacity audio watermarking based on fft amplitude modification," IEICE TRANSACTIONS on Information and Systems, vol. 93, no.1, pp. 87 – 93, 2010.
[7] Byeong-Seob Ko, Ryouichi Nishimura, and Yôiti Suzuki, "Time-spread echo method for digital audio watermarking," IEEE Transactions on Multimedia, vol. 7, no. 2, pp. 212 – 221, 2005.
[8] Guang Hua, Jonathan Goh, and Vrizlynn LL Thing, "Time-spread echo-based audio watermarking with optimized imperceptibility and robustness," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 2, pp. 227 – 239, 2015.
[9] chur-Jen Chen, Huang-Nan Huang, Shu-Yi Tu, Che-Hao Lin,and Shuo-Tsung Chen, "Digital audio watermarking using minimum-amplitude scaling on optimized dwt low-frequency coefficients," Multimedia Tools and Applications, vol. 80, pp.
[10] Mehdi Fallahpour and David Megias, "Audio watermarking based on fibonacci numbers," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 8, pp.1273 – 1282, 2015.
[11] Hwai-Tsu Hu, Shyi-Tsong Wu, and Tung-Tsun Lee, "FFT-based dual-mode blind watermarking for hiding binary logos and color images in audio," IEEE Access, 2023.
[12] Haruka Sakai and Mamoru Iwaki, "Audio watermarking method based on phase-shifting having robustness against band-pass filtering attacks," in IEEE 7th Global Conference on Consumer Electronics, GCCE 2018, Nara, Japan, October 9-12, 2018, pp. 343–346.
[13] Candy Olivia Mawalim and Masashi Unoki, "Audio information hiding based on cochlear delay characteristics with optimized segment selection," in Security with Intelligent Computing and Big-Data Services 2019: Proceedings of the 3rd International Conference on Security with Intelligent Computing and Big-data Services (SICBS), 4 – 6 December 2019, New Taipei City, Taiwan. Springer, 2020, pp. 128 – 138.
[14] Masashi Unoki, Kuniaki Imabeppu, Daiki Hamada, Atsushi Haniu, and Ryota Miyauchi, "Embedding limitations with digital-audio watermarking method based on cochlear delay Characteristics.," J. Inf. Hiding Multim. Signal Process., vol.2, no. 1, pp. 1 – 23, 2011.
[15] Yiming Xue, Kai Mu, Yan Li, Juan Wen, Ping Zhong, and ShaoZhang Niu, "Improved high capacity spread spectrum-based audio watermarking by hadamard matrices," in Digital Forensics and Watermarking: 17th International Workshop, IWDW 2018, Jeju Island, Korea, October 22-24, 2018, Proceedings 17. Springer, 2019, pp. 124 – 136.
[16] Yong Xiang, Iynkaran Natgunanathan, Dezhong Peng, Guang Hua, and Bo Liu, "Spread spectrum audio watermarking using multiple orthogonal PN sequences and variable embedding strengths and polarities," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 3, pp. 529 – 539, 2017.
[17] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka, "Rwc music database: Music genre database and musical instrument sound database," 2003.
[18] Eugen Hoffmann, Dorothea Kolossa, Bert-Uwe Köhler, and Reinhold Orglmeister, "Using information theoretic distance measures for solving the permutation problem of blind source separation of speech signals," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2012, no. 1, pp. 1 – 14, 2012.
[19] Yiqing Lin and Waleed H Abdulla, "Perceptual evaluation of audio watermarking using objective quality measures," in 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2008, pp. 1745 – 1748.

Shengbei Wang, received the B.S. and M.S. degrees from Tiangong University, Tianjin, China, in 2009 and 2012, respectively, both in signal processing. She received the Ph.D. degree in information Science from the Japan Advanced Institute of Science and Technology (JAIST), Nomi, Japan, in 2015. Her main research interests are signal processing, digital audio/speech watermarking,and deep learning-based source separation.

Yuqing Yan, received the B.S. degree from Shijiazhuang University, Shijiazhuang, China, in 2022. She is currently pursuing the M.S. degree in software engineering at Tiangong University. Her research interests include audio watermarking and speech signal processing.