A robust zero-watermarking scheme based on non-negative matrix factorization for audio protection

Xing Guo, Daiyu Huang, Longting Xu*

College of Information Science and Technology, Donghua University, Shanghai, China

* xlt@dhu.edu.cn

Abstract

The copyright problem of digital products is becoming more and more prominent. In this case, digital watermarking technology has attracted the attention of many experts and scholars in the field of information security. Among the proposed technologies, zero-watermarking technology has been favored greatly with its excellent imperceptibility. In this paper, a novel robust audio zero-watermarking scheme is designed by applying non-negative matrix decomposition algorithm to zero-watermarking technology. Firstly, the proposed scheme divides the input audio signal into fixed frames, then applies fast Fourier transform (FFT) and non-negative matrix factorization (NMF) algorithm to extract the feature vector of the original audio signal. Finally, XOR the feature vector and the digital watermark sequence to achieve the embedding of zero-watermarking. The experimental results show that the proposed scheme performs more effectively in resisting common and frame-desynchronization attacks than the existing zero-watermarking schemes.

Introduction

Nowadays, due to the rapid development of network and computer communication technology, multimedia files can be easily distributed, shared and manipulated by people [1, 2]. Undoubtedly, this dramatically increases the demand for copyright protection. While digital watermarking is a promising technology to protect the digital audio products from tampering [3]. Specifically, zero-watermarking technology hides the copyright information in the original audio signal in an imperceptible way. It has become the main focus of researchers. However, a good zero-watermarking scheme also needs to be robust enough to resist many intentional or unintentional attacks [4, 5]. So this paper makes efforts to improve robust performance.

Watermarking technology can be divided into non-blind (including semi-blind) and blind watermarking according to whether the original audio signal and watermark are needed for watermark extraction. Non-blind watermarking requires the original audio and watermark to participate in watermark extraction, while blind watermark requires not [6]. Obviously, blind watermarking is more valuable in practical application. The audio zero-watermarking scheme
designed in this paper belongs to blind watermarking. Until now, audio blind watermarking has been explored a lot.

All existed zero-watermarking schemes can be broadly categorized into two groups: time domain based and transform domain based. Methods in transform domain are more popular than in time domain due to its simplicity and effectiveness. At present, the advanced zero-watermarking schemes are mainly based on discrete cosine transform(DCT) [7], discrete wavelet transform(DWT) [8], singular value decomposition(SVD) [9] or the combination of multiple transform domain [10, 11].

The zero-watermarking schemes based on the transform domain utilize the essential characteristics of the original audio signal. Specifically, an audio zero-watermarking scheme based on dct coefficients symbol is proposed in paper [12]. It performs DCT transformation on the average value of the absolute value of each frame, then selects the maximum value of the DCT coefficient to realize the embedding of watermark. In [13], DWT transformation is executed on the host signal to get the power approximation or energy of the audio segment. Then, the watermark is extracted according to the relative energy of the continuous segment. This scheme is a suitable candidate for audio copyright protection. However, it does have weak point. Because the watermark is constructed by relationship between adjacent fragments, so this scheme is not robust against frame-desynchronization attacks. Min L et al. [11] proposes a scheme based on DCT-DWT-SVD which can effectively resist common attacks. The watermark is registered by performing SVD on the coefficients which are generated after DWT and DCT transformation. However, it still has room for further improvement. For example, when it suffers low-pass filter attack, its NC value is less than 0.96.

Furthermore, as for some state-of-the-art techniques, researchers have explored the phase spectrum of Short Time Fourier Transform (STFT) [14] to generate watermark. But it just studies three types of attacks. The various other attacks are not considered. L.Xu et al. [15] propose an audio zero-watermarking method based on sparse representation, the OMP algorithm and K-SVD algorithm are adopted. After that, authors propose a novel zero-watermarking technique based on the GFT [16]. They use the combined shift operator to construct the graph signal, and then the stable graph Fourier coefficients are selected for encoding. These two zero-watermarking schemes have good robustness against various attacks, but ours is not bad either. From the experimental results, we can find that the NC values in our scheme can reach more than 99% when resisting common attacks. So the robustness of proposed scheme and above two schemes [15, 16] is comparable.

This paper proposes an audio zero watermarking scheme based on non-negative matrix factorization so as to further improve the robust performance. The reasons why NMF algorithm used are explained below.

- Audios and images are generally represented by high-dimensional data matrix in the fields of image recognition and speech signal processing. So how to deal with multi-dimensional data has become an urgent problem for researchers. While non-negative matrix decomposition is a simple, intuitive and effective matrix decomposition technique to decompose the original high-dimensional data matrix into the basis matrix and coefficient matrix of a lower dimension [17, 18].

- More importantly, all elements in the matrix are not negative, which not only reduces the storage space, but also makes the decomposed results sparse. It is worth mentioning that this sparse feature can better represent the essential characteristics of the original audio signal [19]. Furthermore, it overcomes the limitation of SVD with non-negative restriction.
In this paper, we use the maximum value of each group of coefficients to represent each frame of audio signal, and then encode all the maximum values. Through conducting comparative experiments with some zero-watermarking schemes, experiments results show that the proposed scheme can resist attacks better and enhance the robustness performance.

**Work contributions**

Due to the rapid development of network technology, multimedia files can be easily distributed, shared and manipulated, so the demand for copyright protection is increasing day by day. A key contribution of this work is to apply NMF algorithm to zero-watermarking technology and propose a new scheme for audio copyright protection.

Audio zero-watermarking technology is a promising audio copyright protection technology with its excellent imperceptibility. Furthermore, improving its robustness is still the goal pursued by researchers. A second key contribution of this work is that we have conducted experiments on a number of known data sets and compared the results with those of three related works. Experimental results show the proposed scheme can achieve satisfactory robust performance against common attacks and frame-desynchronization attacks.

**Materials and methods**

The basic concept of non-negative matrix factorization

NMF algorithm is a matrix factorization algorithm with non-negative constraints. Given a matrix \( V \in \mathbb{R}^{n \times m} \), looking for the non-negative basis matrix \( W \in \mathbb{R}^{r \times n} \) and the non-negative coefficient matrix \( H \in \mathbb{R}^{r \times m} \) to satisfy \( V \simeq WH \). A non-negative matrix is thus decomposed into the product of two non-negative matrices. By replacing the original data matrix with the coefficient matrix, the dimensionality reduction matrix of the data feature can be obtained.

Convert the matrix factorization problem into a problem of minimizing the error between two matrices. This theory can be better explained by the following formula:

\[
\min \| V - V' \| = \sum_{ij} (V_{ij} - V'_{ij})^2 
\]

where \( V \) is the original matrix and \( V' \) is the matrix to be updated \( (V' = WH) \). \( V_{ij} \) is the atom in row \( i \) and column \( j \) of the \( V \) matrix. In order to obtain the optimal value, the iterative formulas used in the experiment for the matrices \( W \) and \( H \) are given below.

\[
W_{ij}^{k+1} = W_{ij}^k \frac{(H^k)^T \ast V_y}{((H^k)^T \ast W^k \ast H^k)_y} 
\]

\[
H_{ij}^{k+1} = H_{ij}^k \frac{(W^k)^T \ast V_y}{((W^k)^T \ast W^k \ast H^k)_y} 
\]

The NMF algorithm is divided into two parts: training process and testing process. When training, the \( W \) matrix and \( H \) matrix are initialized randomly, where \( k \) represents the number of iterations. When Eq (1) converges and approaches 0 through iteration, it indicates that the input \( V \) matrix has been decomposed into \( W \) and \( H \) matrices. The \( W \) matrix will be stored. When testing, input \( V \) matrix and \( W \) matrix obtained from the training process, according to the formula in algorithm 2 to obtain the \( H \) matrix. Algorithm 1 and 2 describes NMF training and testing process clearly and concisely.
**Algorithm 1**: NMF-training process

Input: $V_{n \times m}$

Initialize the random matrix $W_{n \times r}$ and matrix $H_{r \times m}$.

$W, H = \min_{W,H} D(V||WH)$

for $k \leftarrow 1$ to iteration number do

$W_{k+1} \leftarrow W_k \left( \frac{\langle W_k^T V \rangle}{\langle W_k^T W_k \rangle} \right)$

$H_{k+1} \leftarrow H_k \left( \frac{\langle H_k^T V \rangle}{\langle H_k^T H_k \rangle} \right)$

$V = WH$

end

Output: $W_{n \times r}, H_{r \times m}$

**Algorithm 2**: NMF-testing process

Input: $V_{n \times m}, W_{n \times r}$

Output: $H_{r \times m}$

The proposed zero-watermarking scheme based on non-negative matrix factorization

The scheme designed in this paper includes two modules: watermark embedding and watermark extraction. The detailed analysis is as follows. The simple and comprehensible watermark embedding and extraction process can be seen in Fig 1.

1. Watermark embedding process
   (1) Watermark image preprocessing: In this paper, the binary image $C$ of $N \times N (N = 32)$ is used as copyright information, and the binary image is transformed into a one-dimensional vector $C'$. Then, the binary image $C'$ is embedded into the audio signal $Y$. The algorithm for watermark embedding is shown in Fig 1.

![Watermark embedding and extraction process](https://doi.org/10.1371/journal.pone.0270579.g001)
signal vector by dimensionality reduction processing. Each pixel in the binary image is represented by a one-dimensional signal vector, namely:

\[
C = \{c(i) = c(1), c(2), \ldots, c(M), 1 \leq i \leq M, M = N^2\}
\]

(4) Obtaining the watermark key: XOR the polarity vector \(B\) and the one-dimensional watermark signal \(C\) to obtain the watermark key. So that we achieve the embedding of watermark.

\[
K = B \oplus C
\]

2. Watermark extraction process
(1) Suppose the audio signal after attacks is \(Y'\), then repeat step (2) of the watermark embedding above to conduct audio preprocessing.
(2) Get the polarity vector \(B'\) according to the same encoding method (3) as the watermark embedding above.
(3) XOR the polarity vector and the key to obtain the extracted watermark signal \(C'\).

\[
C' = B' \oplus K
\]

(4) The one-dimensional watermark signal is converted to a two-dimensional watermark signal, and the binary watermark image is obtained after image restoration.

**Experimental results and analyses**

**Experiment preparation**

The input audio signal of this experiment is from the LibriSpeech corpus [20]. It is composed of 585 hours of real voice data and corresponding text collected by 2456 speakers at a sampling
rate of 24kHz. In this experiment, we set the resample point of the speech signal to 256. In order to prove the validity and the feasibility of the proposed scheme, the selected audio data set is divided into training data set and testing data set. The training data set consists of 250 sentences selected from three randomly selected people for about 20 minutes. The testing data type is divided into in-domain and out-of-domain conditions for the experiment, each testing data set contains 50 sentences and takes about 3 minutes. In-domain means that the testing data is within the range of the training data set, and out-of-domain means that the testing data is outside the range of the training data set. The detailed information is shown in Table 1. In this paper, for evaluating the robustness of the proposed scheme, we choose six different types of attacks. Below is a brief introduction to each attack.

- **Noise:** the SNR of the additive white Gaussian noise (AWGN) which is added to the original audio signal is 10dB, 20dB, 30dB.
- **Low pass filter:** a low-pass filter with cut-off frequency 5kHz is used.
- **Resampling:** the sampling frequency is first changed to one half of the original sampling frequency, and then changed to the original sampling frequency.
- **MP3 compression:** the original signal is compressed in MP3 format (128kbps).
- **Re-quantization:** the audio signal is quantized from 16 bits to 8 bits, and then quantized from 8 bits to 16 bits.
- **Frame-desynchronization:** the beginning or the end of audio signal will be cropped to 0.5 second, 1 second and 2 seconds, respectively.

In this paper, the bit error rate (BER) is used to measure the difference between the extracted watermark and the original watermark, and the normalized correlation coefficient (NC) is used to measure the similarity between the extracted watermark and the original watermark, both BER and NC are the evaluation metrics to measure the robustness of the proposed scheme under various attacks. The specific calculation formulas are as follows.

\[
BER = \frac{Nb_e}{Nb_o}
\]  
\[
NC(C, C') = \frac{\sum_i^N C(i)C'(i)}{\sqrt{\sum_i^N C'(i)^2} \sqrt{\sum_i^N (C(i))^2}}
\]

where \(Nb_e\) is the number of error bits, \(Nb_o\) means the total bits of the original watermark sequence, \(C\) represents the embedded watermark sequence and \(C'\) represents the extracted watermark sequence, \(N\) is the length of watermark sequence. BER is closer to 0 and NC is closer to 1, the performance of the proposed scheme against attacks is better.

Table 1. The Detailed experiment introduction and classification. (a,x,y and z represent a set of statements for a particular speaker).

| Testing data type | Testing data | Training data set |
|-------------------|--------------|------------------|
| In-domain         | x            | x, y, z          |
| Out-of-domain     | a            | x, y, z          |

https://doi.org/10.1371/journal.pone.0270579.t001
Imperceptibility analysis
Zero-watermarking technology does not modify the original data. There is no perceptible difference between the watermarked carrier and the original carrier, so it has excellent imperceptibility.

Robustness analysis

Robustness comparison between the proposed scheme and other zero-watermarking schemes under common attacks. Through BER and NC values, Table 2 compares the robust performance of the proposed scheme and some related zero-watermarking schemes [10, 12, 13] against various common attacks.

As for the proposed scheme, We observe that whether it is in-domain or out-of-domain, excluding noise attack (SNR = 10dB), the BER values under all other attacks are all below 1%, and the range of all NC values is close to or greater than 99%. This clearly illustrates the good robustness of the scheme proposed in this paper against different common attacks. For the noise attack and re-quantization attack, compared with other schemes, the proposed scheme clearly shows better BER and NC values. For the low pass filter attack, compared with the schemes [10, 12], the proposed scheme provides better BER and NC values while maintaining robustness comparable to the scheme proposed in [13]. In terms of re-sampling attack, the proposed scheme has the same robust performance as schemes [10, 12, 13], but it is also slightly better than schemes [10, 12, 13] from the values of BER and NC. The robust performance of the proposed scheme against MP3 compression attack is similar to that of scheme [10], but better than that of scheme [12, 13]. Generally speaking, the NC values of the proposed scheme are almost all around 99%, while the NC values of schemes [10, 12, 13] all range from 90% to 99%.

When resisting different types of common attacks, the original watermark image and watermark images extracted by the proposed scheme and other zero-watermarking schemes are shown in Figs 2–5. A–F represents the different attack types. Specifically, A-no attack, B-

![Table 2. Robustness comparison of the proposed scheme and other zero-watermarking schemes under common attacks, where “/” indicates average metrics BER/NC and the bold mark indicates the best number across all the schemes under each attack.](https://doi.org/10.1371/journal.pone.0270579.t002)

![Fig 2. The proposed scheme used to extract the watermark images.](https://doi.org/10.1371/journal.pone.0270579.g002)
noise attack, C-low pass filter attack, D-resampling attack, E-MP3 compression attack and F-re-quantization attack. We can observe that the watermark images we extracted are all clearly visible. Specifically, When resisting noise attack (10dB), the extracted watermark image of DCT [12] scheme, DWT [13] scheme and DWT-DCT [10] scheme are a bit fuzzy, showing poor robustness. This directly demonstrates better robustness of our scheme.

Robustness comparison between the proposed scheme and other zero-watermarking schemes under frame-desynchronization attacks.

Table 3 shows BER and NC values of the proposed scheme and other zero-watermarking schemes under frame-desynchronization attacks. We can see that the proposed scheme based on blind extraction demonstrates good robustness for slight frame-desynchronization attacks, but poor robustness for serious frame-desynchronization attacks. When the duration of removed frames is less than or equal to 1 second, the BER and NC values are less than 0.3 and greater than 0.8. As the removed frames increase, the results show worse robustness.

But on the whole, compared with other zero-watermarking schemes, the proposed scheme clearly indicates better BER and NC values. As the frame shift increases, the proposed scheme still shows better performance than other zero-watermarking schemes.

In addition, we select one situation where the audio signal is clipped to 0.5 second at the beginning and make a comparison with other zero-watermarking schemes. We can intuitively see from Fig 6 that When the duration of removed frames is 0.5 second, the extracted watermark image of the proposed scheme is relatively clear. While the extracted watermark images of the other three zero-watermarking schemes are very blurry. So we can conclude that the proposed scheme can resist slight frame-desynchronization attacks effectively.

The complexity comparison and analysis. Table 4 lists the average time cost for an audio segment and memory cost in the watermark embedding process for different schemes. As can
be seen from the Table 4, compared with most schemes, the proposed scheme is more time efficient.

For the proposed scheme, the core process of constructing zero-watermarking information is extracting the coefficient matrix to represent the original audio signal. Firstly, we will have a training process for input audio signal to get a base matrix which will occupy the storage space of 73,728 bytes. Then we will obtain the coefficient matrix according to Algorithm 2 which is essentially a simple matrix operation process.

**Conclusion**

In this paper, a robust zero-watermarking scheme based on NMF is proposed. The audio signal is divided into fixed frames, then each frame of audio signal applies FFT and NMF to obtain the basis matrix and coefficient matrix. The maximum coefficient value of each frame

| Methods | Test Data | Metric | Frame-desynchronization attacks |
|---------|-----------|--------|--------------------------------|
|         |           |        | 0.5 second (begin) | 1 second (begin) | 2 seconds (begin) | 0.5 second (end) | 1 second (end) | 2 seconds (end) |
| Our     | In-domain | BER    | 0.1616 | 0.2573 | 0.3280 | 0.1073 | 0.1986 | 0.3275 |
|         | NC        |        | 0.8839 | 0.8107 | 0.7546 | 0.9231 | 0.8559 | 0.7544 |
|         | Out-of-domain | BER    | 0.1219 | 0.2421 | 0.3140 | 0.1731 | 0.2095 | 0.3193 |
|         | NC        |        | 0.9137 | 0.8227 | 0.7669 | 0.8762 | 0.8483 | 0.7620 |
| DCT [12] | In-domain | BER    | 0.4956 | 0.5060 | 0.5008 | 0.3415 | 0.5046 | 0.5035 |
|         | NC        |        | 0.6076 | 0.5992 | 0.6042 | 0.7299 | 0.5997 | 0.6015 |
|         | Out-of-domain | BER    | 0.4906 | 0.4928 | 0.4995 | 0.4417 | 0.4945 | 0.4990 |
|         | NC        |        | 0.6248 | 0.6124 | 0.6047 | 0.6507 | 0.6111 | 0.6047 |
| DWT [13] | In-domain | BER    | 0.3457 | 0.4158 | 0.4728 | 0.1479 | 0.3762 | 0.4600 |
|         | NC        |        | 0.7377 | 0.6789 | 0.6314 | 0.8885 | 0.7137 | 0.6425 |
|         | Out-of-domain | BER    | 0.3128 | 0.4392 | 0.4740 | 0.3183 | 0.3974 | 0.4778 |
|         | NC        |        | 0.7659 | 0.6584 | 0.6310 | 0.7591 | 0.6965 | 0.6272 |
| DWT-DCT [10] | In-domain | BER    | 0.2981 | 0.3730 | 0.4668 | 0.1190 | 0.3389 | 0.4590 |
|         | NC        |        | 0.7757 | 0.7153 | 0.6369 | 0.9117 | 0.7437 | 0.6435 |
|         | Out-of-domain | BER    | 0.2320 | 0.4025 | 0.4586 | 0.2714 | 0.3617 | 0.4597 |
|         | NC        |        | 0.8309 | 0.6919 | 0.6441 | 0.7981 | 0.7253 | 0.6432 |

![Fig 6. Extracted watermark images when the audio signal is clipped to 0.5 second at the beginning.](https://doi.org/10.1371/journal.pone.0270579.g006)
signal after NMF decomposition is taken as the feature information about this frame, then the feature sequence is clustered into two classes to encode as binary sequence 0 and 1 respectively. The zero-watermarking sequence is created by performing XOR operation between the encoded sequence and the watermark sequence. Through comparative experiments, it can be observed that the proposed scheme can extract watermark image well and has good robustness under the common attacks and slight frame-desynchronization attacks. However, it still has limitations such as a need for a trusted third party. The key generated by the audio will be registered in the third party in a one to one correspondence. When extracting the watermark image, you will need to find the required key according to the mapping relationship.

Author Contributions
Investigation: Xing Guo.
Project administration: Xing Guo.
Software: Daiyu Huang.
Writing – original draft: Xing Guo.
Writing – review & editing: Longting Xu.

References
1. Zong T., Xiang Y., Natgunanathan I., Gao L., Hua G. and Zhou W., “Non-Linear-Echo Based Anti-Collusion Mechanism for Audio Signals,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 969–984, 2021. https://doi.org/10.1109/TASLP.2021.3058892
2. A. Fierro-Radilla, M. Nakano-Miyatake, M. Cedillo-Hernandez, L. Cleofas-Sanchez and H. Perez-Meana, “A Robust Image Zero-watermarking using Convolutional Neural Networks,” 2019 7th International Workshop on Biometrics and Forensics (IWBF), Cancun, Mexico, 2019, pp. 1-5.
3. Xiang Y., Natgunanathan I., Peng D., Hua G. and Liu B., “Spread Spectrum Audio Watermarking Using Multiple Orthogonal PN Sequences and Variable Embedding Strengths and Polarities,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 3, pp. 529–539, March 2018. https://doi.org/10.1109/TASLP.2017.2782487
4. Hua G., Goh J. and Thing V. L. L., “Time-Spread Echo-Based Audio Watermarking With Optimized Imperceptibility and Robustness,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 2, pp. 227–239, Feb. 2015. https://doi.org/10.1109/TASLP.2014.2387385
5. J. Fan and K. U, “A Novel Image Zero-Watermarking Scheme Based on Non-Uniform Triangular Partition,” 2019 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Kobe, Japan, 2019, pp. 1-6.
6. Xiang Y., Natgunanathan I., Rong Y. and Guo S., “Spread Spectrum-Based High Embedding Capacity Watermarking Method for Audio Signals,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 12, pp. 2228–2237, Dec. 2015. https://doi.org/10.1109/TASLP.2015.2476755
7. Z. Zhang, “Zero-watermarking algorithm based on DC component in DCT domain,” 2021 International Conference on Electronic Information Engineering and Computer Science (IEECS), 2021, pp. 475-478.
8. J. Panda, S. Choudhary, K. Nath and S. Kumar, “Audio zero watermarking scheme based on sub band mean energy comparison using DWT-DCT,” 2016 International Conference on Signal Processing and Communication (ICSCC), 2016, pp. 352-357.
9. Farzaneh M, Toroghi R M. Robust Audio Watermarking Using Graph-based Transform and Singular Value Decomposition[C]// 2020 10th International Symposium on Telecommunications (IST); 2020.
10. Yu Y., Lei M., Liu Xiaoming, Qu Zhiguo and Cheng Wang, “Novel zero-watermarking scheme based on DWT-DCT,” in China Communications, vol. 13, no. 7, pp. 122–126, July 2016. https://doi.org/10.1109/CC.2016.7559084
11. Lei Min, Yang Yu, Liu Xiaoming, Cheng MingZhi and Wang Rui, “Audio zero-watermark scheme based on discrete cosine transform-discrete wavelet transform-singular value decomposition,” in China Communications, vol. 13, no. 7, pp. 117–121, July 2016. https://doi.org/10.1109/CC.2016.7559083
12. RUIPENG S, HAITAO X. An audio zerowatermarking algorithm based on dct coefficients symbol[J]. Computer Technology And Development, 2014, 24(05):146–149.
13. Yang Y., Lei M., Cheng M., Liu B., Lin G. and Xiao D., “An audio zero-watermark scheme based on energy comparing,” in China Communications, vol. 11, no. 7, pp. 110–116, July 2014. https://doi.org/10.1109/CC.2014.6895390

14. A. Electa Alice Jayarani, M. R. Bhatt and D. D. Geetha, “Zero Watermarking on Audio Based on STFT,” 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), 2018, pp. 253-256.

15. Xu L. et al., “A novel robust zero-watermarking algorithm for audio based on sparse representation,” in China Communications, vol. 18, no. 8, pp. 237–248, Aug. 2021. https://doi.org/10.23919/JCC.2021.08.017

16. Xu L., Huang D., Zaidi S. F. A., Rauf A. and Das R. K., “Graph Fourier Transform Based Audio Zero-Watermarking,” in IEEE Signal Processing Letters, vol. 28, pp. 1943–1947, 2021. https://doi.org/10.1109/LSP.2021.3112338

17. Byun J. and Shin J. W., “Initialization for NMF-based audio source separation using priors on encoding vectors,” in China Communications, vol. 16, no. 9, pp. 177–186, Sept. 2019. https://doi.org/10.23919/JCC.2019.09.013

18. H. Murata and A. Ogihara, “Note Estimation by Contaminated Normal Distribution for Audio Watermarking Method Using Non-negative Matrix Factorization,” 2018 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 2018, pp. 469-474.

19. Chen M., Gong M. and Li X., “Feature Weighted Non-Negative Matrix Factorization,” in IEEE Transactions on Cybernetics. PMID: 34437084

20. ZEN H, DANG V, CLARK R, et al. Libritts: A corpus derived from librispeech for text-to-speech[J/OL]. CoRR, 2019, abs/1904.02882. http://arxiv.org/abs/1904.02882.