

## Digital Audio Watermarking Based on Artificial Neural Networks

<sup>a</sup>Liangbin Zheng, <sup>b</sup>Ruqi Chen, <sup>c</sup>Xiaojin Cheng, <sup>d</sup>Linhong Li

<sup>a, b, c, d</sup>Department of Computer Science Beijing Institute of Graphic Communication Beijing, 102600, China

---

### Abstract

A digital audio watermarking based on artificial neural networks is proposed in this paper. Utilizing the learning and adaptive capabilities of artificial neural networks, the relationship between audio signals and embedded watermark is established by using the important characters of audio signals as the input vector of artificial neural networks, and the watermark were embedded into original audio signals without modifying the audio data. The experimental results show that the embedded watermark is robust to audio signal processing, and the watermarking method does not require the original audio signals for watermarking extraction.

**Index Terms:** digital audio watermarking; artificial neural networks; discrete cosine transform

© 2012 Published by MECS Publisher. Selection and/or peer review under responsibility of the International Conference on E-Business System and Education Technology

---

### 1. Introduction

With the rapid develop of computer network and multimedia information processing technology, it is more easily to copy digital product illegally, thus intellectual property protection for digital products is required more urgently. Digital watermarking technology provides a method of copyright protection for its security, robustness and imperceptibility. Digital audio watermarking technology uses the features of human auditory system to embed the secret information into the audio files and achieves the purpose of copyright protection for digital audio products.

According to the different ways of embedding the watermark into the audio files, current digital audio watermarking technologies can be divided into time domain algorithms and transform domain algorithms. The former directly embeds the watermark information into the selected time domain of audio signal, such as the least significant bit algorithm[1], the phase coding algorithm[2], and the echo hiding method[3]; the latter performs some kind of transformation on the audio signal, and then embeds the watermark into the transformed coefficients. Finally it recovers the watermarked audio signal through the corresponding inverse transform. Transformation algorithms are mainly the discrete Fourier transform(DFT), the discrete cosine transform(DCT), the discrete wavelet transform(DWT). In general speaking, the transform domain algorithm has more transparency, security, and capacity in embedding watermark into audio signal. Both the time domain algorithms and the transform domain algorithms achieve the purpose of embedding watermark information into audio signal by modifying the original audio signal. Although it makes full use of the characteristics of human auditory

model systems, the perception quality of audio signal is changed to a certain extent.

This paper makes full use of the learning and adaptive capabilities of artificial neural networks and uses the important characters of audio signal as the input vector of artificial neural networks. The relationship between audio signal and watermark information is established through the learning of artificial neural networks, which achieves embedding watermark into the original audio signal without modifying the audio data. This method does not change the perception quality of audio signal and improve the imperceptibility of watermark. In addition, this proposed watermarking method does not require the original audio signal for watermarking extraction.

## 2. Back propagation algorithm of artificial neural networks

Back propagation(BP) algorithm of artificial neural network[4] belongs to  $\delta$  algorithm, which is a supervised machine learning algorithm. The main idea of BP algorithm is to propagate the output layer error from back to front and indirectly calculate the hidden layer error. BP algorithm is divided into two phases. In the first phase, the value of each unit of output layer is obtained by calculating the input vector from input layer to output layer. In the second stage (back propagation), using the vector of output layer, the error of each hidden layer is calculated which is used to modify the connection weights of neural network. BP algorithm usually uses gradient method to modify the weights of neural network that minimize the sum of squared errors. In addition, back propagation algorithm often uses Sigmoid function as output function. Fig. 1 shows the conventional symbols in back propagation algorithm. For the calculating unit  $j$ , the subscript  $i$  is on behalf of calculating unit  $i$  of its former layer and the subscript  $k$  is on behalf of calculating unit  $k$  of its later layer. The  $O_j$  represents the output value of current layer and the  $W_{ij}$  represents the weight from the former layer to current layer. When sample data is inputted to the neural network, each calculating unit from input layer to output layer performs the following calculation:

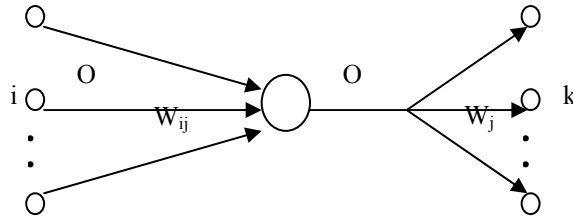


Figure 1. the conventional symbols in back-propagation algorithm

$$net_j = \sum_i W_{ij} O_i \quad (1)$$

$$o_j = f(net_j) \quad (2)$$

For the output layer, the  $\hat{y}_j$  ( $\hat{y}_j = o_j$ ) is the actual output value and the  $y_j$  is the ideal output value. The output error  $E$  is as follow:

$$E = \frac{1}{2} \sum_j (y_j - \hat{y}_j)^2 \quad (3)$$

In order to simplify the formula, the gradient is defined as follow:

$$\delta_j = \frac{\partial E}{\partial net_j} \quad (4)$$

Considering the impact of weights  $W_{ij}$  to the error, the following formula can be obtained:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} = \delta_j O_i \quad (5)$$

Weights modifying should be made to reduce error fast. The weights to be modified as follow:

$$\Delta W_{ij} = -\eta \delta_j O_i$$

$$W_{ij}(t+1) = W_{ij}(t) + \Delta W_{ij}(t) \quad (6)$$

If the node  $j$  is output unit,  $\delta_j$  can be obtained as follow:

$$o_j = \hat{y}_j$$

$$\delta_j = \frac{\partial E}{\partial \hat{y}_j} \frac{\partial \hat{y}_j}{\partial net_j} = (y_j - \hat{y}_j) f'(net_j) \quad (7)$$

If the node is not output unit  $j$ ,  $\delta_j$  can be obtained as follow:

$$\delta_j = \frac{\partial E}{\partial net_j} = \sum_k \frac{\partial E}{\partial net_k} \cdot \frac{\partial net_k}{\partial O_j} \cdot \frac{\partial O_j}{\partial net_j} = \sum_k \delta_k w_{jk} f'(net_j) \quad (8)$$

In practice, in order to accelerate the convergence rate, the weights to be modified is usually added the previous modified value which is commonly called inertia item. Then the weights to be modified as follow:

$$\Delta W_{ij}(t) = -\eta \delta_j O_i + \alpha \Delta W_{ij}(t-1)$$

As for the Sigmoid function  $y = f(x) = \frac{1}{1+e^{-x}}$ ,

$$f'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = y(1-y), \text{ the back propagation algorithm performs the following steps:}$$

- (1) Initialize the connection weights of artificial neural network.
- (2) Repeat the following process until the artificial neural network is convergent.
  - ① Calculate the output value ( $O_j$ ) of each calculating unit from input layer to output layer as follow:

$$net_j = \sum_i W_{ij} O_i, O_j = \frac{1}{1+e^{-net_j}}$$

② Calculate the  $\delta_j$  of output layer as follow:

$$\delta_j = (y - O_j)O_j(1 - O_j)$$

③ Calculate the  $\delta_j$  of hidden layer from back to front as follow:

$$\delta_j = O_j(1 - O_j) \sum_k w_{jk} \delta_k$$

④ Calculate and save the modified weights as follow:

$$\Delta W_{ij}(t) = \alpha \Delta W_{ij}(t-1) + \eta \delta_j O_i$$

⑤ Modify the weights as follow:

$$W_{ij}(t+1) = W_{ij}(t) + \Delta W_{ij}(t)$$

### 3. Audio watermark embedding and extraction

Watermark embedding should take advantage of the most important features of audio signal, which can make the watermark information more robustness. Low frequency signal is the most concentrated energy part of audio signal. In general, the processing of audio files is as much as possible to avoid the loss of low frequency information, or would cause grate damage to audio quality. The method proposed in this paper performs discrete cosine transform (DCT) on the digital audio signal and selects the low frequency coefficients to construct the characteristics of audio signal which is as the input vector of artificial neural network. Through the training of artificial neural network, the relationship between audio signal and watermark information is established. So the proposed scheme achieves embedding watermark into audio signal without modifying the original audio files.

#### 3.1 Watermark generation

The watermark signal in this proposed method is not traditional one dimensional pseudorandom sequence, but rather a meaningful binary image, which has better practical value. Supposing  $W$  is  $M \times N$  image, which can be expressed as  $W = \{w(i, j), 0 \leq i < M, 0 \leq j < N\}$ . The  $w(i, j)$  represents the gray value of the  $i, j$  entry pixel of watermark image and  $w(i, j)$  belongs to the set of 0 to 1, write it as:  $w(i, j) \in \{0, 1\}$ . In order to embed the two dimensional image into the one dimensional digital audio signal, the dimensions of watermark image should be reduced. Watermark image can be changed to one dimensional sequence as follow:

$$V = \{v(r) = w(i, j), 0 \leq i < M, 0 \leq j < N, r = i * M + j\}.$$

In order to eliminate the spatial correlation of binary watermark image and improve the robustness of watermarking algorithm, the one dimensional binary sequence  $v$  is executed scrambling transformation. In addition, in order to strengthen the confidentiality of watermark, the watermark sequence is encrypted. So even if someone can extract the watermark from audio signal, he can not restore the watermark correctly without the decryption key. A random sequence is generated by a secret key[5], which is used to readjust the watermark sequence  $v$ , and then a new one dimensional binary sequence is obtained, write it as:  $X = \{x(i), 0 \leq i < Nw, Nw = M \times N\}$ .

#### 3.2 embedding watermark information

Assuming the audio data is defined as:  $A = \{a(i), 0 \leq i < n\}$ , the watermark embedding process can be described as follows:

(1) Dividing the entire audio signal into  $n/L$  audio segments, write it as follow:

$A = \{A(k), 0 \leq k < n/L\}, A(k) = \{a(kL+i), 0 \leq i < L\}.$

(2) Dividing the watermark information into  $n/L$  segments, write it as follow:

$X = \{X(s), 0 \leq s < Nw/q, Nw = M \times N\}, X(s) = \{x(sq+i), 0 \leq i < q\}.$

(3) Each segment of audio signal is performed discrete cosine transform, and then the low frequency coefficients are selected. Write it as follow:

$CA = \{CA(k), 0 \leq k < n/L\}, CA(k) = \{ca(kL+i), 0 \leq i < m\}.$

(4) Creating an artificial neural network with  $CA(k)$  as its input vector,  $X(s)$  as its output vector. Through the training of artificial neural network, the relationship between  $CA(k)$  and  $X(s)$  is established. That means the corresponding relationship between the important features of audio signal and the watermark information is established.

### 3.3 the extraction of Watermark information

(1) Dividing the entire digital audio signal into  $n/L$  audio segments, write it as follow:

$A = \{A(k), 0 \leq k < n/L\}, A(k) = \{a(kL+i), 0 \leq i < L\}.$

(2) Performing discrete cosine transform on each segment of audio signal and selecting the low frequency coefficients, write it as follow:

$CA = \{CA(k), 0 \leq k < n/L\}, CA(k) = \{ca(kL+i), 0 \leq i < m\}.$

(3) Using the  $CA(k)$  as input vector of the trained artificial neural network and recording the output vector  $X(s)$ .

(4) Using the secret key to decrypt the watermark information  $X$  getting the original one dimensional watermark binary sequence  $V$ , write is as follow:

$V = \{v(r), 0 \leq r < Nw, v(r) \in \{0,1\}\}.$

(5) Converting the one dimensional binary sequence  $V$  to two dimensional sequence restoring the watermark image  $W$ , write is as:  $W = \{w(i,j), 0 \leq i < M, 0 \leq j < N\}.$

The original audio signal is not required while using the trained artificial neural network to extract watermark information. In order to eliminate the influence of subjective or objective factors, such as testers experience, physical condition, experimental conditions, equipments and so on, the normalized correlation coefficient (NC)[6] should be considered to evaluate the similarity between the original watermark and the extracted watermark and to determine the copyright of audio files.

## 4. the results of experiment

This proposed method is implemented on the Windows XP platform using Visual C++. The sample frequency of original digital audio signal is 44.1kHz. The selected watermark information is a  $32 \times 32$  binary image.

Experimental results show that the proposed watermarking scheme is inaudible and robust against various signals processing such as noising adding, low pass filtering, lossy compression, re-sampling, re-quantifying. The extracted watermark information from the watermarked audio signal which has been attacked is shown in Fig. 2.

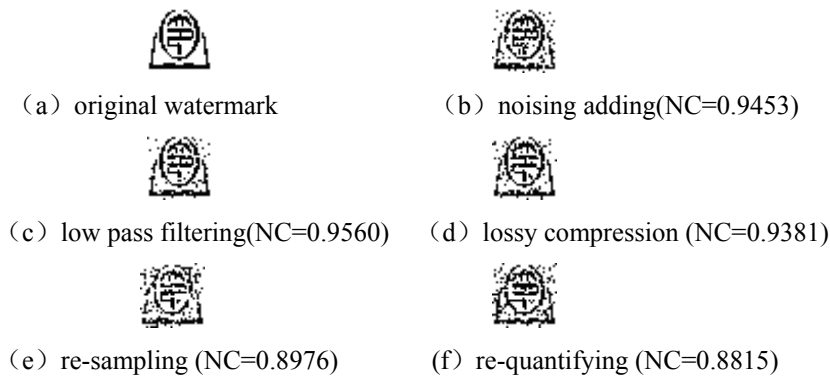


Figure 2. the results of Experiment

## 5. Conclusion

This paper presents a digital audio watermarking method based on artificial neural networks. Utilizing the learning and adaptive capabilities of artificial neural networks to establish the relationship between audio signal and watermark information. The machine learning method is applied to the audio watermarking technology. Because the watermarked audio signal is also the original signal, it is difficult to discover whether the audio is protected or not, which is a great advantage of this proposed method. This method has many advantages such as easy in calculation and implementation, robust against many signal processing. Especially the original audio signal do not required in the watermark extraction, which expand its scope of application greatly.

## ACKNOWLEDGMENT

This research is supported by the Scientific Research Common Program of Beijing Municipal Commission of Education (Grant No. KM201010015003) and the Funding Project for Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality (Grant No. PHR201108349, PXM201001 4223095557).

## References

- [1] Bassia P,Pitas I.Robust Audio Watermarking in the Time-domain.Proceedings of EUSIPCO'98, Rhodesk, Greece,1998:25-28.
- [2] Bender W,Gruhl D,Morimoto N.Techniques for Data Hiding.IBM System Journal,1996,35:313-336.
- [3] Caoyang Zhao,Zhenghua Liu,Ting Wang.Echo hiding techniques for digital audio signals. Application Research of Computers, 2000, 17(7):42-44(in Chinese).
- [4] Zhaoqi bian,Xuegong Zhang.Pattern Recognition.Beijing: Tsinghua University Press, 2003, 254-257(in Chinese).
- [5] Hsu Chiou-Ting,Wu Ja-Ling.Hidden digital watermarks in images.IEEE Trans. on Image Processing,1999,8(1):58-68.
- [6] Martin K,Fabien A,Petitcolas P.A fair benchmark for image watermarking systems.Security and Watermarking of Multimedia Continents:Proceedings of Electronic Imaging,1999:226-239