

基于心理声学模型的自适应音频数字水印算法^{*}

李云剑, 陈 亮, 张雄伟

(解放军理工大学通信工程学院, 江苏 南京 210007)

摘 要: 提出一种基于心理声学模型的频域音频数字水印算法。首先对音频信号分段进行离散傅立叶变换 (DFT), 再利用心理声学模型估算各段掩蔽阈值, 使水印信号嵌入 DFT 域低频子带区时自适应处于掩蔽阈值以下, 对音频品质冲击小, 含水印音频信号与原始信号感知相似。实验表明本算法具有良好的不可觉察性和稳健性。

关键词: 数字水印; 掩蔽效应; 心理声学模型

中图分类号: TP391 **文献标识码:** A **文章编号:** 0529-6579 (2004) S2-0204-05

近年来, 信息隐藏技术已成为国内外学者研究的热点之一。信息隐藏技术用于版权保护称为数字水印。数字水印具有隐形性、稳健性和确定性的主要特征^[1]。音频数字水印技术可用于数字音频产品的版权保护, 因此研究前景广阔。对音频信号嵌入水印是利用音频信号中的冗余信息和人耳的主观听觉特性。目前常用的音频水印算法有: 最低有效位 (LSB) 方法^[2], 相位编码法^[3], 扩展频谱法^[4]和回声隐藏法^[5]。本研究提出一种基于心理声学模型的自适应音频数字水印算法, 将水印信号自适应嵌入音频信号频域低频子带内掩蔽阈值以下, 人耳感知不到; 水印检测采用非盲检测方法。算法仿真结果表明, 这种方法隐蔽性好, 具有较强的稳健性, 能较好的抵抗噪声、压缩、滤波等攻击。

1 心理声学模型

心理声学模型常用于感知压缩编码, 基本原理是利用人的听觉阈特性和掩蔽特性, 把噪声控制在听觉阈值以下, 使人耳觉察不到。听觉的阈值特性是指听觉对不同频率的声音有不同的灵敏度, 在听觉阈值以下的声音人耳感知不到, 这给实现隐形水印嵌入提供了理论基础。人耳的掩蔽效应分为频域掩蔽和时域掩蔽。心理声学模型还把人的耳蜗看成为一个有重叠的带通滤波器组, 其幅度响应不对称、非线性, 且滤波器带宽非均匀, 带宽随频率增加而增大。所以根据人耳的听觉特性, 将 10 ~ 22 kHz 声音分成 25 个临界带^[6]。

本文通过估算各子带的最小掩蔽阈值来自适应

控制水印嵌入的强度, 减少对音频品质的冲击。具体步骤^[6]如下。

1.1 频谱分析和声压级标准化

将输入信号 $s(n)$ 归一化 $x(n)$ 后分成每帧 512 个样点, 用 1/16 重叠的汉宁窗 $w(n)$ 降低边缘效应, 再进行 512 点快速傅立叶变换, 计算声压级得标准化功率谱密度 $P(k)$ 。

$$P(k) = PN + 20\log_{10} \left| \sum_{n=0}^{N-1} w(n) x(n) \exp\left[-\frac{j2\pi kn}{N}\right] \right| \quad (1)$$
$$0 \leq k \leq N/2$$

PN 是参考声压级 96 dB 与每帧频谱中最大声压级之差。

1.2 识别音调和非音调掩蔽音

在功率谱密度样值中, 找出比相邻值大的局部最大值 local maxima, 再从中找出音调 (类似正弦波) 成分 S_T ;

$$S_T = \left\{ P(k) \left| \begin{array}{l} P(k) > P(k \pm 1) \\ P(k) > P(k \pm \Delta_k) + 7 \text{ (dB)} \end{array} \right. \right\} \quad (2)$$

从音调成分中计算连续 3 个频谱值为音调掩蔽音声压级 $P_{TM}(k)$:

$$P_{TM}(k) = 10\log_{10} \sum_{j=1}^3 10^{0.1P(k+j)} \text{ (dB)} \quad (3)$$

再从剩下频谱中计算出非音调 (类似噪声) 掩蔽音声压级 $P_{NM}(k)$ 。

$$P_{NM}(k) = 10\log_{10} \sum_j 10^{0.1P(j)} \text{ (dB)}$$
$$\forall P(j) \notin \{P_{TM}(k, k \pm 1, k \pm \Delta_k)\} \quad (4)$$

* 收稿日期: 2004—09—11

基金项目: 国防预研基金资助项目

作者简介: 李云剑 (1977 年生), 女, 硕士研究生; E-mail: liyunjian77@21cn.com

1.3 抽取和重识掩蔽音

去除小于安静阈值的掩蔽音后，用带宽为 0.5 临界带宽的滑动窗抽取出窗内的掩蔽音。

$$P_{TM,NM}(i) = P_{TM,NM}(k)$$
$$P_{TM,NM}(k) = 0$$
$$P_{TM,NM}(k) \geq T_q(k)$$

$$i = \begin{cases} k & 1 \leq k \leq 48 \\ k + (k \bmod 2) & 49 \leq k \leq 96 \\ k + 3 - ((k - 1) \bmod 4) & 97 \leq k \leq 232 \end{cases}$$

(5)

1.4 计算单独掩蔽阈值

对不同临界带的信号施加掩蔽函数 $SF(i, j)$ ，决定每个单独频率块的噪声掩蔽阈。

$$T_{NM}(i, j) = P_{TM}(j) - 0.275z(j) + SF(i, j) - 6\ 025$$
$$T_{NM}(i, j) = P_{NM}(j) - 0.175z(j) + SF(i, j) - 2\ 025$$

(6)

$T_{TM}(i, j)$, $T_{NM}(i, j)$ 单位为 dB SPL。

$T_{TM}(i, j)$ 表示频率 i 处音调对频率块 j 的掩蔽阈值, $T_{NM}(i, j)$ 表示频率 i 处非音调对频率块 j 掩蔽阈值, $P_{TM}(j)$ 表示音调掩蔽音在频率块 j 处的声压级, $P_{NM}(j)$ 表示非音调掩蔽音在频率块 j 处的声压级, $SF(i, j)$ 表示频率 i 处声音对频率块 j 处声音的掩蔽函数。

1.5 计算总体掩蔽阈值

组合各单个掩蔽阈值来估计各个临界带的全局掩蔽阈值 $T_g(i)$ 。

$$T_g(i) = 10\log_{10} \left[\frac{10^{0.1T_q(i)} + \sum_{l=1}^L 10^{0.1T_{TM}(i, l)} + \sum_{m=1}^M 10^{0.1T_{NM}(i, m)}}{3} \right],$$
$$i \in 25 \text{ 个临界带}$$

(7)

1.6 决定最低掩蔽阈值

心理声学模型 1 将 256 点频率谱点平均分为 32 个子带，在每个子带中，根据总体掩蔽阈值的情况，决定各个子带的最小掩蔽阈值 $T_{\min}(n)$ 。

$$T_{\min}(n) = \min(T_g(i)), i \in \text{第 } n \text{ 子带}$$

(8)

2 水印的嵌入

本算法分帧对数字音频信号嵌入水印，在每帧的频域嵌入水印信号的 1 bit，虽然音频信号幅度有微小变化，但不产生被感知的效果。水印信号是文本信息，其比特数不大于音频信号的帧数。假设输入音频信号含 M 个样值，表示为 $x(i)$, $i = 0, \dots, M - 1$ 。

(1) 将时域音频信号样值分为每帧 384 点，然

后对每帧信号 $X_k(i)$ 进行 512 点快速傅立叶变换 $Y_k(j\omega)$ ，将 256 点频域样值平均分为 32 个子带，每个子带 8 个样点。

(2) 由心理声学模型 1 估算出每个子带的最小掩蔽域值 $T_{\min}(n)$ 。

(3) 为有效抵抗攻击，选择感知重要的低频子带作为嵌入区域，如 2~6 子带，计算嵌入区域最小掩蔽阈值作为水印嵌入系数 α_k 的上界，使嵌入的水印信号处于掩蔽阈值内，人耳感知不到。

(4) 在音频信号频域的嵌入区域嵌入水印信号，水印信号 w 为双极性信号 $\{-1, +1\}$ ，这样含水印信号的频谱为

$$Y_k^w(j\omega) = \begin{cases} Y_k(j\omega) + \alpha_k w_k & \omega \in \text{嵌入区域} \\ Y_k(j\omega) & \omega \notin \text{嵌入区域} \end{cases}$$

(9)

(5) 对修改后的频域系数做傅立叶逆变换，得到含水印的音频信号 $X_k^w(i)$ 对应不同嵌入强度，原始音频信号与水印信号信噪比不同，计算公式为：

$$snr = 10\log_{10} \frac{\sum_n x^2(n)}{\sum_n [x(n) - x^w(n)]^2}$$

(10)

$x(n)$ 为原始音频信号, $x^w(n)$ 为含水印音频信号, n 为时域样点数。

(6) 综上所述，水印嵌入过程如图 1 所示。

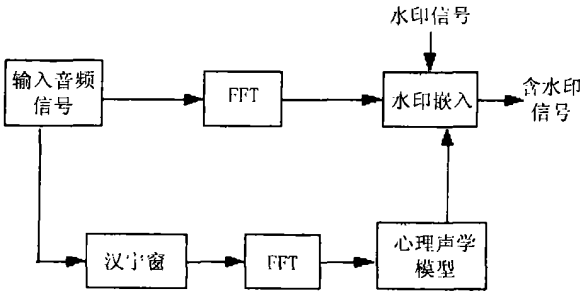


图 1 水印嵌入图
Fig 1 Embedding of the watermark

3 水印的提取

水印提取采用非盲水印检测方法。对应于嵌入过程，在音频信号的频域进行水印提取。假设接收信号为 $X^w(i)$, $i = 0, \dots, N - 1$, 比较 M 和 N , 取小值作为提取操作长度。具体过程为：

(1) 分别将原始信号 $X(i)$ 和接收信号 $X^w(i)$, $i = 0, \dots, \min(M, N) - 1$, 分帧后进行快速傅立叶变换 $Y_k(j\omega)$, $Y_k^w(j\omega)$ 。

(2) 在每帧嵌入子带频域内进行幅度比较，对应频谱差值小于 0，赋值为 -1，差值大于 0，赋值

为+1; 对所有值求和, 以 0 作为判决门限。和为 0, 判断没有水印嵌入; 和大于 0, 判断嵌入水印为+1; 和小于 0, 判断嵌入水印为-1。

$$\sum_{j\omega \in \text{嵌入子带区}} (Y_k^w(j\omega) - Y_k(j\omega)) = \begin{cases} 0 & \text{无水印嵌入} \\ < 0 & \text{嵌入水印为 } -1 \\ > 0 & \text{嵌入水印为 } +1 \end{cases}$$

(11)

水印检测过程如图 2 所示。

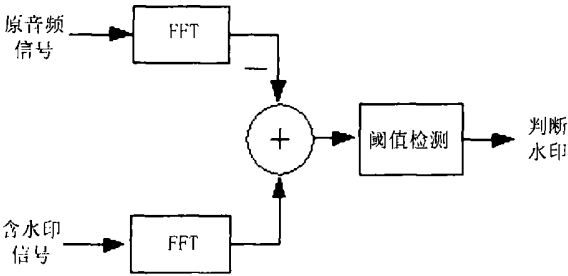


图 2 水印检测图
Fig. 2 Extracting of the watermark

4 仿 真

测试信号为单声道, 44.1 kHz, 16 bit, 19.559 s 长的音频信号, 含 86 2464 个样点数, 分为 2246 帧; 采用一段文本作为水印信号。仿真实现工具为 Matlab6.5, 实验过程: 首先将每帧信号依据最小掩蔽阈值自适应嵌入不同强度的水印, 嵌入区域为 2~6 子带。然后用 cooleditpro2.0 对含水印信号进行加噪、压缩, 滤波等处理, 进行水印检测, 将提取水印信号与嵌入水印信号比较, 得出检测正确率。不同嵌入强度对应于不同信噪比含水印信号, 主观听觉效果也不同, 表 1 所示。原音频信号、含水印信号 SNR=32.787 dB 及两者误差信号分别如图 3 中 a、b、c 所示; 原始音频信号、含水印音频信号一帧的频谱图及两者误差分别如图 4 中 a、b、c 所示。

表 1 不同信噪比信号主观感受

Tah 1 The subjective hearing reception of the watermarked audio signal

嵌入强度	SNR/dB	主观听觉感受
$T_{\min}(n) * 1.0$	32.7875	轻微失真
$T_{\min}(n) * 0.8$	34.7257	无失真
$T_{\min}(n) * 0.5$	38.8076	无失真
$T_{\min}(n) * 0.1$	52.7693	无失真

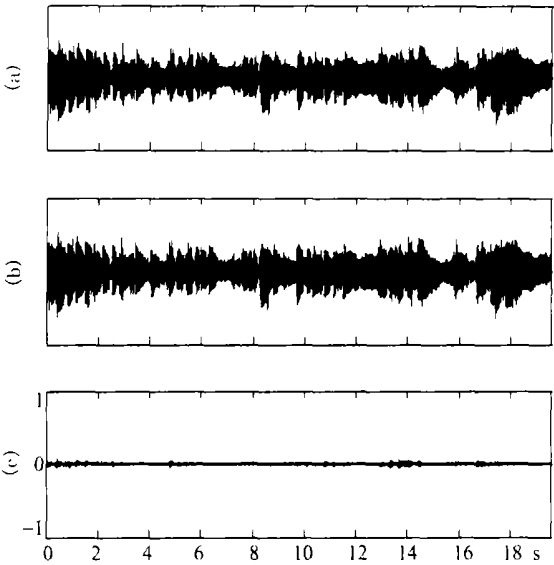


图 3 原始信号、含水印信号及两者误差
Fig. 3 The original audio signal, the watermarked audio signal and the difference between them
(a) 原始音频信号; (b) 含水印信号; (c) 误差信号

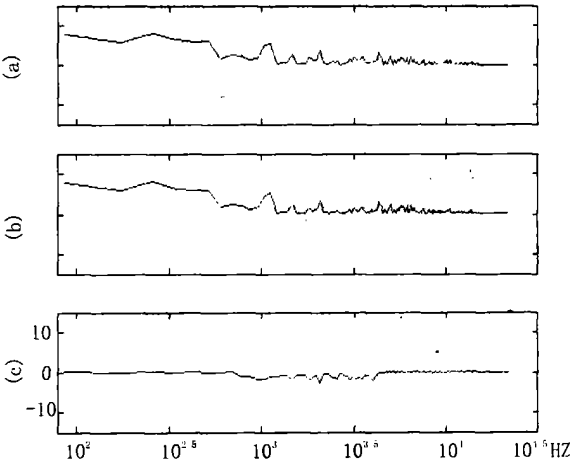


图 4 原始信号、含水印信号一帧频谱及两者误差
Fig. 4 Frequency spectrum of one frame original audio signal, frequency spectrum of one frame watermarked audio signal and the difference between them
(a) 原始音频信号一帧频谱图;
(b) 含水印信号一帧频谱图; (c) 误差信号

差别, 通过主观听觉测试也证实了这一点。为检测算法的稳健性, 对不同嵌入强度含水印信号分别进行下面的处理, 比较实验结果。

4.1 加入高斯白噪声

通过检测算法提取水印信号结果显示, 随着嵌入强度减弱, 检测正确率降低。随着信号信噪比降低, 检测正确率降低 (图 5)。

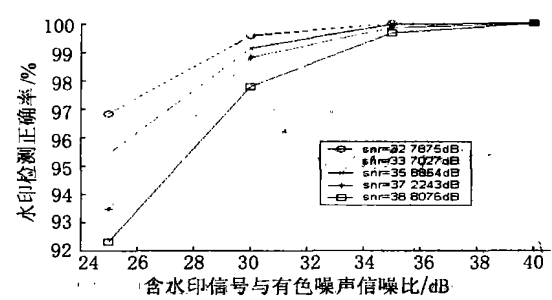


图 5 加不同强度白噪声检测正确率

Fig 5 The correctness rate of extracting in white-noised environment

4.2 加入有色噪声

有色噪声为产生的 Pinknoise，不同信噪比信号水印检测正确率如图 6 所示。

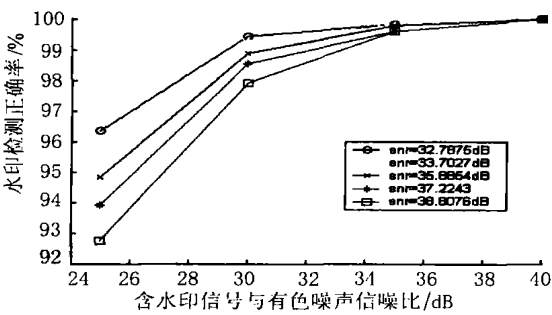


图 6 加不同强度有色噪声检测正确率

Fig 6 The correctness rate of extracting in color-noised environment

4.3 低通滤波

将不同嵌入强度含水印信号分别通过截止频率为 8 kHz 和 4 kHz 的低通滤波器，检测正确率如表 2 所示。

表 2 抗低通滤波检测结果

Tah 2 The extracting result after low pass filter		
信噪比/dB	低通截止频率/4 kHz	低通截止频率/8 kHz
32 7875	99. 8219	99. 9110
34 7257	99. 3644	99. 8219
37. 2243	99. 3321	99. 5993
38 8076	98. 8869	99. 4657

4.4 中值滤波

将含水印信号分别进行 3 点、5 点中值滤波后，水印检测正确率如表 3 所示。

4.5 有损压缩

信号压缩至 128 kbit/s MP3 信号，然后进行

MP3 解压缩，结果显示，水印检测正确率为 100%。

表 3 抗中值滤波检测结果

Tah 3 The extracting result after median filter		
信噪比/dB	3 点/%	5 点/%
32 7875	99. 8219	98. 2636
33. 7027	99. 7774	97. 9074
34. 7257	99. 5548	97. 5512
35. 8854	99. 4212	97. 0169
37. 2243	99. 2431	96. 0374

4.6 重新量化

信号由 16 bit 量化为 8 bit，再量化为 16 bit，水印检测错误率如表 4 所示。

表 4 抗重新量化检测结果

Tah 4 The extracting result after resampling			
信噪比/dB	32 7875	35. 8854	38 8076
检测错误率/%	0	0. 1336	0. 48

5 结 论

从仿真结果可以看出，水印嵌入强度越强，攻击能力越强；但同时，信噪比降低，音频音质也有所降低，所以可根据需要取舍。由于利用了心理声学模型 1 估算出的最小掩蔽阈值来自适应地控制水印嵌入强度，使得水印隐蔽性好，具有较强的鲁棒性。此方法的缺点是采用非盲水印检测方法，水印检测需要原始音频信号作为参考。因此，研究不需原始音频的盲检测方法是下一步的工作目标。

参考文献:

[1] 刘振华, 尹萍. 信息隐藏技术及其应用[M] . 北京: 科学出版社, 2000.

[2] PETITCOLAS A P, ANDERSON R J, KUHN M G. Information hiding —— a survey[J] . Proceedings IEEE, 1999, 87(7): 1062—1077.

[3] CHANG L, MOSKOWITZ S. Critical analysis of security in voice hiding techniques[A] . Proceedings of the International Conference on Informatin and Communications Security, Vol 1334 of Lecture Notes in Computer Science [C] . Springer, 1997: 203—216.

[4] COOPERMAN M, MOSKWITZ S. Steganographic Method and Device[M] . USA: Patent, 1997.

[5] BENDER W, GRUHL D, MORIMOTO N, et al. Techniques for data hiding[J] . IBM System J, 1996, 35(3—4): 313—336.

[6] TED P, ANDREAS S. A review of algorithms for perceptual coding of digital audio signals, 2000.

**Adaptive Audio Digital Watermark Algorithm
Based on Psychoacoustic Auditory Model**

LI Yun jian, CHEN Liang, ZHANG Xiong wei

(Institute of communications engineering, PLA University of Sciences and Technology, Nanjing 210007, China)

Abstract: A digital watermark algorithm based on psychoacoustic model is presented. Watermark is embedded in frequency domain. The audio signal is divided into segments first, then each segment is performed discrete fourier transform (DFT) . The masking thresholds of all segments are estimated by psychoacoustic analysis. And the watermark is embedded into the low frequency subbands and below the masking thresholds adaptively. The acoustic impact is little and the audio signal embedded watermark is perceptually similar to the original. The simulated results demonstrate that the algorithm is reliable and robust.

Key words: digital watermark; masking effect; psychoacoustic model

(上接第 203 页)

Research on Watermark Signal of Public key Digital Watermarking Scheme

HU Yan Jun, MA Xiao ping, CHEN Ying, GAO Li

(School of Communication & Electronic Engineering,
China University of Mining and Technology, Xuzhou 221008, Jiangsu, China)

Abstract: A public key robust watermarking scheme has more potential value of application. A signal with with special auto correlation characteristic is proposed. Also a public key robust watermarking scheme based on this signal is designed and experimented. The experiment result shows the scheme is valid and robust to common signal distortions and malicious attacks.

Key words: robust watermarking; public key; auto correlation