

# 基于听觉频率掩蔽效应的音频数字水印算法

桑鹤平, 赵鹤鸣

SANG He-ping, ZHAO He-ming

苏州大学 电子信息学院, 江苏 苏州 215021

School of Electron Information, Suzhou University, Suzhou, Jiangsu 215021, China

E-mail: twocakes725@yahoo.com.cn

SANG He-ping, ZHAO He-ming. Audio digital watermarking algorithm based on auditory frequency masking. Computer Engineering and Applications, 2008, 44(9): 101-103.

**Abstract:** Watermarking embedding algorithm based on auditory masking is proposed in this paper, which embeds watermarking based on masking threshold in low-band coefficients of DCT transformation. Calculating masking threshold of every critical band by auditory masking model, then quantization steps are controlled by masking threshold and value of audio carrier of the embedding points. At the same time, for meeting the require of fast embedding, we can find the best balance between time of embedding and robustness by algorithm proposed in this paper, and get the best repeating times of embedding watermark in one frame. Results of the experiment show that the embedded watermark are inaudible and have good robustness against kinds of attacks such as white noise, low-pass filtering, re-sampling and compression.

**Key words:** audio watermarking; masking model; quantization; robustness

**摘 要:** 文中提出一种基于听觉掩蔽效应的音频数字水印算法, 根据掩蔽阈值在 DCT 低频系数中嵌入水印。依据听觉掩蔽模型计算每个临界频带的掩蔽阈值, 再根据水印嵌入点处的掩蔽阈值和音频载体值的大小共同控制量化步长。同时为满足水印快速嵌入的要求, 通过文中提出的算法可以寻找到嵌入时间(计算量)与鲁棒性的最佳平衡点, 得出一帧语音中嵌入水印的最佳重复次数。实验结果证明所嵌入的水印具有很好的不可感知性, 对加白噪声、低通滤波、重采样和压缩等各种具有良好的鲁棒性。

**关键词:** 音频水印; 掩蔽模型; 量化; 鲁棒性

**文章编号:** 1002-8331(2008)09-0101-03 **文献标识码:** A **中图分类号:** TP309

## 1 引言

数字水印技术是指在数字化信息(通常为图像、音频、视频)中通过某种算法嵌入能代表作者身份或版权的记号, 但要求水印的嵌入对这些载体的影响是不可察觉的, 并且在经历一些不破坏原数据使用价值或商用价值的操作或攻击后水印仍能被提取出来。数字水印的提出较好地解决了数字化信息的版权保护问题。数字水印是 20 世纪 90 年代提出的一门新兴技术, 目前国内外对数字水印的研究主要集中在图像上, 而对在音频中嵌入水印的研究较少, 但网络上越来越多的音频传输同样需要得到版权保护。

近年来提出的音频水印算法主要分为以下两类: 时域算法(包括最低比特位算法、基于回声延时的算法等); 变换域算法(包括基于 DFT、DCT、DWT 变换的算法)。但是最低比特位法抗噪声和低通滤波的能力差, 回声延时法信号利用率低, 嵌入速度慢, 变换域算法嵌入强度无法客观控制, 只能靠实验结果主观控制。本文依据人耳听觉掩蔽效应, 在保证人耳不可感知的前提下, 最大程度地嵌入水印, 在水印的鲁棒性、嵌入时间和不可感知性之间找到最好的平衡。

## 2 掩蔽阈值的计算

两个信号同时发生, 且频率相近, 那么较强的信号使较弱的信号变得不可感知, 这就是人耳的频率掩蔽效应。在一定宽度频带内掩蔽效应相同, 依据这个, 人耳听觉掩蔽模型可以划分为若干个临界频带, 同一临界频带内, 各个点上的噪声掩蔽阈值相同。

计算掩蔽阈值时, 首先划分临界频带, 现在一般将 0 kHz~16 kHz 划分成不等的 24 个临界频带, 由于音调与噪声成分的掩蔽模型不同, 所以首先要区分音调和噪声成分。先根据式(1)计算音频载体的功率谱。

$$X(f) = 10 \lg \left| \frac{1}{N} \sum h(i) s(i) \exp(-j2\pi fi/N) \right|^2 \quad (1)$$

将  $X(f)$  最大值归一化到 96 dB, 然后依据功率谱区分音调和非音调,

$$X_{\text{tone}} = \begin{cases} X(f) & X(f) > X(f \pm 1) \\ X(f) & X(f) > X(f \pm \Delta_k) + 7 \text{ dB} \end{cases} \quad (2)$$

$$\Delta_k \in \begin{cases} 2 & 2 < k < 63 & (0.17 \text{ kHz} \sim 5.5 \text{ kHz}) \\ [2, 3] & 63 \leq k < 127 & (5.5 \text{ kHz} \sim 11 \text{ kHz}) \\ [2, \dots, 6] & 127 \leq k < 256 & (11 \text{ kHz} \sim 20 \text{ kHz}) \end{cases} \quad (3)$$

**作者简介:** 桑鹤平(1982-), 男, 硕士研究生, 研究方向: 语音信号处理; 赵鹤鸣(1957-), 男, 教授, 博导, 研究方向: 语音信号处理。

**收稿日期:** 2007-08-01 **修回日期:** 2007-10-19

满足式(2)要求的  $X_{tone}$  即为音调成分,由此可计算音调成分的声压级  $Y_{tone}(f)$

$$Y_{tone}(f) = 10 \lg \left( 10^{\frac{X(f-1)}{10}} + 10^{\frac{X(f)}{10}} + 10^{\frac{X(f+1)}{10}} \right) \quad (4)$$

而非音调(噪声)成分的声压级为每个临界频带内残留信号功率谱的总和,即:

$$Y_{ns}(f) = \sum 10^{\frac{X(f)}{10}} \text{ (dB)}, \forall X(f) \notin \{X_{tone}(f, f \pm 1, f \pm \Delta_k)\} \quad (5)$$

$$\bar{f} = \left( \prod_{j=1}^u j \right)^{1/(l-u+1)} \quad (6)$$

其中,  $u$  和  $l$  为临界频带的上下界。

计算绝对掩蔽阈值  $T_{ab}$ , 删除  $Y_{tone}(f) < T_{ab}$  与  $Y_{ns} < T_{ab}$  的分量, 再计算单个音调或噪声的掩蔽阈值, 单个音调的掩蔽阈值定义为:

$$Y_{tone}(i, j) = Y_{tone}(j) - 0.275z(j) + Y_s(i, j) - 6.025 \quad (7)$$

其中  $Y_{tone}(j)$  表示音调掩蔽者在  $j$  的声压级(SPL),  $z(j)$  为频率的巴克表示,  $Y_s(i, j)$  表示掩蔽者  $j$  到掩蔽对象  $i$  的扩展门限函数。单个噪声掩蔽者的阈值  $Y_{ns}(i, j)$  为:

$$Y_{ns}(i, j) = Y_{ns}(j) - 0.175z(j) + Y_s(i, j) - 2.025 \quad (8)$$

音调和噪声掩蔽门限加上绝对掩蔽门限便得到全局掩蔽门限:

$$T_{end}(i) = 10 \lg \left( 10^{\frac{T_{ab}(i)}{10}} + \sum_{j=1}^L 10^{\frac{Y_{tone}(i, j)}{10}} + \sum_{m=1}^M 10^{\frac{Y_{ns}(i, m)}{10}} \right) \text{ dB} \quad (9)$$

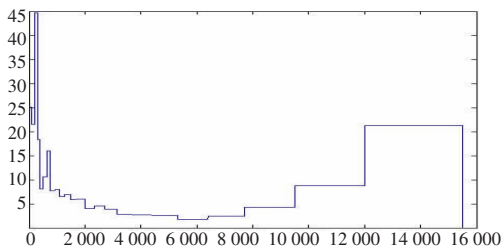


图1 一帧语音的频域掩蔽阈值

### 3 水印的嵌入

#### 3.1 水印的降维

$$W = \{w(i * M + j) = p(i, j), 0 \leq i \leq M, 0 \leq j \leq N\} \quad (10)$$

$P$  为待嵌入图像 ( $M \times N$ )。将音频载体分帧, 做 DCT 变换, 每一帧中嵌入 1 bit 水印信息, 将纵坐标划分成如图 2 所示的  $A, B$  区间。设待量化系数为  $x$ , 区间的长度为  $\Delta$ ,  $\Delta$  即为量化间隔。

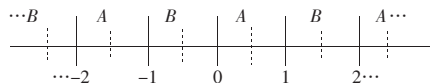


图2 量化区间的划分

#### 3.2 量化间隔 $\Delta$ 的选取

根据以上求得的掩蔽阈值决定  $\Delta$  的大小, 在每个点上的掩蔽阈值乘一个系数  $\alpha$  便得到这个点上的  $\Delta$ , 但由于存在某些点, 本身值很小, 而这个点所在临界频带的掩蔽阈值较大, 这样如果完全由它所处的临界频带的掩蔽阈值决定该点的  $\Delta$ , 那么嵌入水印时的改变量与该值本身的比值将较大, 这样就对水印的不可感知性有较大影响, 所以在以前的研究中只根据掩蔽阈值决定  $\Delta$  的大小的基础上, 再加入音频载体  $X$  本身的大小, 即:

$$T_{end} \times X \times \beta = \Delta \quad (11)$$

其中  $\beta$  为衰减系数。

#### 3.3 量化规则

当  $W=1$  时, 将  $x$  的值量化为与  $x$  最接近的  $A$  区间的中间值, 当  $W=0$  时, 将  $x$  的值量化为与  $x$  最接近的  $B$  区间的中间值。

#### 3.4 最佳重复嵌入数

如果在每一帧中选取一个系数嵌入水印, 那水印的鲁棒性不强。所以从低频端开始选取多个系数将一比特水印重复嵌入。但是如果在低于掩蔽阈值的分量中嵌入水印, 难以抵抗压缩攻击, 所以选取高于掩蔽阈值的分量嵌入水印。一帧中重复嵌入次数越多, 鲁棒性越好, 但由于高频段能量较小, 对加噪声, 低通滤波攻击的能力差, 所以在高频端嵌入水印的话提取时容易产生错误, 而且选取系数的多少与嵌入水印的速度(计算量)也是一对矛盾, 因此本文设计以下算法计算出一帧中被用于重复嵌入水印的系数的最佳个数  $I$ , 首先定义水印提取的位错误率  $BER$ :

$$BER = \frac{1}{L} \sum_{i=1}^L B_i \times 100\% \quad (12)$$

其中  $B_i = \begin{cases} 1 & wt_i \neq w_i \\ 0 & wt_i = w_i \end{cases}$ ,  $L$  为水印长度,  $w_i$  为提取出的水印。

算法如下:

(1) 每一帧音频做 DCT 变换, 去除所有低于掩蔽阈值的部分, 并将去除的值另外保存起来, 并且记录他们在每一帧音频中所处位置。

(2) 由第一步得到已去除掩蔽阈值部分的音频中, 每一帧从低频端起的  $i(i < 5)$  个系数重复嵌入一比特水印信息, 加入低于掩蔽阈值的部分, 得到已加入水印的音频。经各种攻击后提取水印, 计算位错误率  $BER_i$ , 计算嵌入水印的计算量  $M_i$ 。

(3)  $i=i+1$ , 在已去除低于掩蔽阈值的音频中, 从低频端起的  $i$  个系数中重复嵌入一比特水印, 加入低于掩蔽阈值的部分, 得到已加入水印的音频。相同攻击后再提取水印计算位错误率  $BER_i$ , 计算嵌入水印的计算量  $M_i$ 。

(4) 如果位错误率减小则转(5), 若增加则结束,  $i-1$  即为最佳重复嵌入系数个数。

$$(5) \text{ 如果 } \frac{BER_{i-1} - BER_i}{BER_i} > \frac{M_i - M_{i-1}}{M_i} * \lambda, \text{ 则转(3), 否则 } i-1 \text{ 即为最佳重复嵌入系数个数的大小, } \lambda \text{ 为衰减系数。}$$

加入低于掩蔽阈值的部分, 再做 IDCT 变换便得到嵌入水印后的音频。

### 4 水印的提取

首先对已嵌入水印的音频分帧, 然后对每一帧做 DCT, 得到  $X'$ , 根据该帧掩蔽阈值和嵌入系数的个数  $I$  确定嵌入水印的点, 对这些点按照以下公式提取水印

$$wt_{ji} = \begin{cases} 1 & x'_i \in A \text{ 区间} \\ 0 & x'_i \in B \text{ 区间} \end{cases} \quad (13)$$

$$wm_j = \frac{\sum_{i=1}^I wt_{ji}}{I} \quad (14)$$

$$wt_j = \begin{cases} 1 & wm_j > 0.5 \\ 0 & wm_j < 0.5 \end{cases} \quad (15)$$

再将  $wt$  升维便得到提取出来的水印图像。

### 5 实验结果与分析

本文实验选取的音频为一段歌曲, 长度 23 s, 采样频率为

32 kHz,分辨率为 16 bit,水印为一副二值图像(65×21)。图 3 为原始音频时域图,图 4 为嵌入水印后的音频时域图,直观可以看出嵌入水印后对原始音频改变很小。

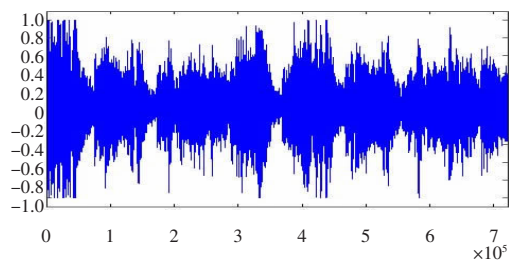


图 3 原始音频

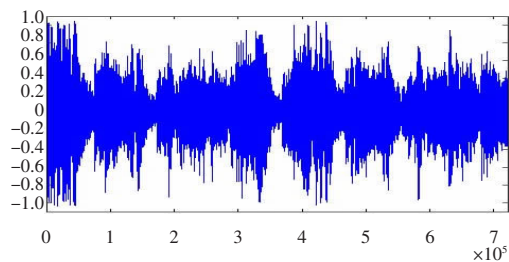


图 4 嵌入水印后的音频

为检测算法的鲁棒性,对含水印的音频进行以下攻击:

- (1)加入高斯白噪声;
- (2)低通滤波,取长度为 6,截止频率为 8 kHz 的低通滤波器;
- (3)重新采样,先进行下抽样,从 32 kHz 到 16 kHz,再进行内插,从 16 kHz 到 32 kHz;
- (4)压缩攻击。

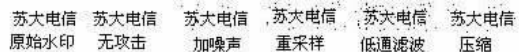


图 5 受攻击后提取的水印

图 5 是原始水印和攻击后提取的水印对比

为从客观上评价水印,用文中定义的水印提取位错误率 BER 来评价水印。

表 1 本文算法与两种常用算法鲁棒性的比较

BER	加噪声 SNR=20 dB	低通滤波	重采样	压缩
DFT 算法	5.36	7.65	3.14	3.13
DWT 算法	0.63	3.91	3.54	4.32
本文算法	0.33	4.61	1.32	1.11

## 6 结束语

本文提出一种基于人耳听觉掩蔽效应的数字音频水印算

法,用量化和重复编码的方法将水印重复嵌入到音频中,并且找到一帧音频中的最佳重复嵌入次数,从而既满足了快速嵌入的要求又满足了水印的鲁棒性,同时有很好的不可感知性。但是由于现在对音频水印的攻击没有一个完整的标准,所以本文算法中的攻击只能是靠实验主观设定,再者文中也不能计算出每一帧的最佳重复嵌入次数,只能算出对于一段语音总体的一个值,这些将是下一步的研究重点。

## 参考文献:

- [1] Garcia R A. Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory[C]//Proc of 107th AES Convention, New York, USA, 1999:1-42.
- [2] He Xing, Scordilis M S. An enhanced psychoacoustic model based on the discrete wavelet packet transform[J]. Journal of the Franklin Institute, 2006, 343(7): 738-755.
- [3] Wang Rang-Ding. Audio watermarking algorithm based on wavelet packet and psychoacoustic model[C]//Proceedings-Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies, 2005: 812-814.
- [4] Nakayama A, Lu Jinlin, Nakamura S, et al. Digital watermarks for audio signal based on psychoacoustic masking model[J]. Electronics and Communications in Japan: Part 3 (Fundamental Electronic Science), 2003, 86(12): 65-75.
- [5] Virag N. Single channel speech enhancement based on masking properties of the human auditory system[J]. IEEE Transactions on Speech and Audio Processing, 1999, 17(2): 126-137.
- [6] Malik H, Khokhar S, Rashid A. Robust audio watermarking using frequency selective spread spectrum theory[C]//2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004.
- [7] Quan Xiaomei, Zhang Hongbin. Audio watermarking based on psychoacoustic model and adaptive wavelet packets[C]//2004 7th International Conference on Signal Processing Proceedings (IEEE Cat. No. 04TH8739), 2004.
- [8] 王秋生, 孙圣和. 基于量化数字音频信号频域参数的水印嵌入算法[J]. 声学学报, 2002, 27(4): 379-385.
- [9] 陶智, 赵鹤鸣, 顾济华, 等. 基于心理声学模型和临界频带子波变换的数字音频水印[J]. 声学学报, 2006, 31(2): 114-119.
- [10] 王让定, 徐达文, 陈金儿. 基于频率掩蔽效应的自适应音频数字水印技术[J]. 计算机工程与应用, 2004, 40(15): 31-34.
- [11] 王泳, 李斌. 基于 MPEG 心理声学模型 II 的自适应音频水印算法[J]. 中山大学学报: 自然科学版, 2006, 45(2): 29-32.

(上接 72 页)

- [2] Atanassov K. Operations over interval-valued intuitionistic fuzzy sets[J]. Fuzzy Sets and Systems, 1994, 64: 159-174.
- [3] Mondal T, Samanta S. Topology of interval-valued intuitionistic fuzzy

sets[J]. Fuzzy Sets and Systems, 2001, 119: 483-194.

- [4] 郭效芝. 模糊不确定性度量的探讨及扩展[D]. 西安: 西北大学, 2004: 48-50.
- [5] Liu X. Entropy distance measures and similarity measure of fuzzy sets and their relations[J]. Fuzzy sets and systems, 1992, 72: 331-348.