

# 基于心理声学模型的自适应量化同步 数字音频水印算法\*

王永琦<sup>1</sup>, 杨洋<sup>2</sup>, 邓琛<sup>1</sup>

(1. 上海工程技术大学 电子电气学院, 上海 201620;

2. 上海工程技术大学 高职学院, 上海 200437)

**摘要:** 针对传统的固定步长量化方法对不同种类音频载体适应性较差的缺点, 基于心理声学模型, 在获取的语音信号的 DCT 域的低频系数中量化地嵌入水印信息, 量化步长由掩蔽阈值自适应控制, 以满足水印的盲检测。同时引入了同步信号的思想, 利用同步信号和水印序列号定位水印的嵌入位置。实验结果表明, 该算法不仅具有较好的不可感知性, 而且对诸如低通滤波、有损压缩、重采样和剪裁等攻击均具有较好的顽健性。

**关键词:** 数字音频水印; 心理声学模型; 自适应量化; 同步信号

中图分类号: TP393

文献标识码: A

## Adaptive quantization and synchronous digital audio watermarking algorithm based on psychoacoustic model

WANG Yong Qi<sup>1</sup>, YANG Yang<sup>2</sup>, DENG Chen<sup>1</sup>

(1. College of Electrical and Electronic Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2. Vocational of Technical College, Shanghai University of Engineering Science, Shanghai 200437, China)

**Abstract:** The existing quantization audio watermarking algorithms have generally adopted the fixed quantizations strategy, which is unavailable for variety kinds of audio. Watermarking was embedded by quantizing DCT low-band coefficients of digital audio signal. The embedding strength was adaptively controlled by the masking threshold. The idea of embedded synchronic signal is proposed, which uses the synchronic signal for locating the watermark location. Experiment results show that the proposed watermarking scheme is inaudible and robust against various signal processing, such as lowpass filtering, noise adding, re-sampling and cutting.

**Key words:** digital audio watermarking; psychoacoustic model; adaptive quantization; synchronic signal

近年来, 音频水印技术正逐渐成为数字水印研究的一个重要方向。由于人类听觉系统比视觉系统更加敏感, 因而在音频信号中嵌入水印具有很大的挑战性。

现有的音频数字水印技术, 根据嵌入水印时对信号处理方式不同, 可分为时域算法和变换域算法。时域算法是直接将水印信息嵌入到音频信号的时域选定的采样数据中, 其算法简单、运行速度快, 但抗干扰能力差; 变换域算法是将一个域内的数字信号映射成另一个域内的信号, 然后加入水印, 再逆变换为原来域中的数字信号, 其复杂度较高, 但鲁棒性好, 因此成为目前水印的主流技术。参考文献[1]提出的基于量化 DCT 域音频水印新算

法, 通过对 DCT 系数中 DC 分量和第 4 AC 分量进行轮流调制来完成水印嵌入, 未能结合听觉掩蔽特性确定水印嵌入深度, 影响了数字水印的不可感知性与鲁棒性。参考文献[2]提出了一种新的基于小波变换的自适应量化音频水印算法, 其鲁棒性有一定的保证, 但由于没有涉及到同步信号, 故不能抵抗剪切等攻击。参考文献[3]提出了一种基于混沌序列的小波域的音频信息隐藏算法。该算法由于没有设置密钥, 其安全性不能得到保证。

本文结合人类听觉系统掩蔽效应及同步码技术, 提出了一种可抵抗去同步攻击的自适应量化 DCT 域数字音频水印算法。为了提高算法的随机性和安全性, 将混沌理论引入算法。仿真实验表明, 该算法具有较好的抗攻击性能。

\* 基金项目: 上海高校选拔培养优秀青年教师科研专项基金 (05XPYQ50)

## 1 基于听觉心理声学模型的听觉掩蔽效应

在基于量化的音频水印算法中,量化步长是增强与平衡数字水印的鲁棒性和不可感知性的决定因素,因此,将心理声学模型中的听觉掩蔽效应应用在音频水印中,使用全局掩蔽阈值来自适应地计算出量化步长,取消传统量化算法中的基础量化步长,从而实现完全自适应。在基于心理声学模型中的数字音频水印算法中,如果某帧的听觉掩蔽阈值比较大,则对该帧的 DCT 变换系数量化步长可以相应地取大,使水印具有更好的不可感知性和鲁棒性。本文采用的是 MPEG-1 听觉心理声学模型获取全局掩蔽阈值<sup>[4]</sup>。其全局掩蔽阈值计算过程如下:

(1) 首先将原始音频信号分成互不重叠的帧,计算每帧能量谱  $P(k)$ ,变换之前用标准海宁窗加权,减少变换窗的边界效应,将其划分为 25 个临界频率,然后测定其声压级。

(2) 根据短时能量谱的分布情况确定音频信号的特性,判断有调成分和无调成分。一个掩蔽成分的音调性影响到掩蔽阈值,在有调成分和无调成分大于绝对阈值的范围内,且在小于 0.5 bark 的距离内抽取最高的功率成分,移去较小的功率成分。

(3) 单独掩蔽阈值的计算,有调和无调的单独掩蔽阈值表示如下:

$$LT_m[Z(j),Z(i)]=X_m[Z(j)]+av_m[Z(j)]+vf[Z(j),Z(i)]$$

$$LT_{tm}[Z(j),Z(i)]=P_{tm}[Z(j)]+av_{tm}[Z(j)]+vf[Z(j),Z(i)] \quad (1)$$

式中,  $LT_m$  和  $LT_{tm}$  分别代表了有调成分和无调成分的单独掩蔽阈值,  $X_m[Z(j)]$  标号为  $j$ 、相应临界频带率  $Z(j)$  的声压级,  $av$  为掩蔽标记,  $vf$  为掩蔽成分的掩蔽函数。

(4) 总掩蔽阈值的  $LT_g$  的计算:把相应于单独掩蔽阈值的功率和安静状态阈值的功率加起来即可求得总掩蔽阈值。

$$LT_g(i)=10 \times \log_{10} \left( 10^{LT_q(i)/10} + \sum_{j=1}^m 10^{LT_m(Z(j),Z(i))/10} + \sum_{j=1}^n 10^{LT_{tm}(Z(j),Z(i))/10} \right) \quad (2)$$

式中,  $m$ 、 $n$  分别是有调掩蔽器和无调掩蔽器的数量,  $LT_q(i)$  表示频带  $i$  处的绝对听觉阈值。

(5) 最小掩蔽阈值的确定:子带  $n$  中最小掩蔽电平  $LT_{\min}(n)$  表示为:

$$LT_{\min}(n)=\text{MIN}[LT_g(i)] \quad (3)$$

## 2 数字水印信息的嵌入

参考文献[5-6]提出在音频信号中引入同步信号的方法。同步信号的作用是对每个嵌入的水印定位,在时域内嵌入同步码作为同步信号,当音频作品遭到剪裁攻击时,该方法虽能提取出部分正确的水印信息,但无法确定这些值在水印图像中的位置,因而无法恢复图像,而且该算法对噪声攻击非常敏感。

本文提出的算法在加入同步信号的基础上,还添加了每个水印值的序号,在检测时通过同步信号确定水印嵌入在音频作品中的位置,在恢复图像时利用水印序号确定每个水印值在图像中的位置。同时为了增强嵌入数据的鲁棒性,将同步信号、水印序号、水印信号均嵌入 DCT 域的低频系数中。图 1 给出了嵌入数据的结构。



图 1 嵌入数据结构

### 2.1 同步信号

文中采用  $m$  序列作为同步信号,设  $\{a_n\}$ 、 $\{b_n\}$  是具有 2 个相同周期  $P$  的 2 个  $m$  序列,  $a_n, b_n \in \{-1, 1\}$ , 则它们的相关函数被定义为:

$$C_{a,b}(t)=(1/p) \sum_{n=1}^p a_n b_{n-t} \quad (4)$$

$m$  序列  $\{a_n\}$  的自相关函数有如下的性质:

$$C_{a,b}(t)= \begin{cases} 1 & t=0 \\ -1/p & t \neq 0 \end{cases} \quad (5)$$

设  $\{a_n\}$  是作为同步信号的原始  $m$  序列,  $b_n \in \{-1, 1\}$  是一待检测的序列,如果  $C_{a,b}(0) \geq T/p$ , 则认为  $\{b_n\}$  是一个同步信号。阈值  $T$  的意义是:如果待检验序列最多只有  $(P-T)/2$  bit 与  $\{a_n\}$  不同,则认为  $\{b_n\}$  是一个同步信号。如果同步码组长度过大,则会导致计算量大幅度增加。因此,在满足同步要求的前提下应尽量选用较短的同步码。本文选用长度为 12 bit 的 bark 码 111110011010 作为同步码。

### 2.2 水印预处理

文中采用的水印信号是一幅大小为  $H(M_1 \times M_2)$  的二值灰度图像,要将其嵌入到一维音频信号中,必须进行降维处理,即将二维图像转换成一维序列:

$$V=\{V(k)=H(m_1, m_2), 0 \leq m_1 < M_1, 0 \leq m_2 < M_2, k=m_1 \times M_2 + m_2\} \quad (6)$$

通过降维操作,水印  $W$  中的像素  $H(m_1, m_2)$  由序列  $V$  中的元素  $V(k)$  表示。然后先对一维水印序列进行置乱,因为若直接将水印信息嵌入,在遭受大面积剪裁后,提取的水印图像会出现整块区域无法恢复的情况。为了避免水印图像成块损失,使提取的水印图像易于辨认,需要将此缺失区域分散到整个水印图像内。本文采用 Logistic 映射,首先产生一个混沌序列,对水印图像与混沌序列进行“异或”运算,即可生成置乱加密后的安全水印信息  $V_1$ 。图 2 为二值水印图像混沌加密及还原过程。用混沌序列排序的方法对图像进行置乱有很高的保密度,只有确切知道初值即密钥才能恢复原图像。

### 2.3 水印嵌入

对每个水印元素按照同步码、水印序号、水印的顺序组成一个水印的数据组,首先要将数字音频信号进行



图2 二值水印图像混沌加密及还原

分段处理, 然后再将每个音频数据段分割成  $L_1$ 、 $L_2$ 、 $L_3$  3 部分, 分别进行 DCT 变换, 每帧选取频率较低的交流分量  $X_e(k)(m_w)$  ( $1 \leq m_w \leq 64$ ), 嵌入同步码、水印序号和水印信息。本文采用双极性参数量化法对所选低频系数进行量化处理, 其嵌入过程如下:

(1) 将每帧中选择好的中低频系数  $X_e(k)(m_w)$  划分区间集: 设定合适的量化步长  $\Delta$  将坐标轴分割成如图 3 所示的 A 区间集和 B 区间集。

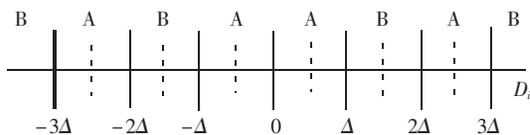


图3 量化系数嵌入水印信息原理图

(2) 取整数商和余数运算: 根据确定好的量化步长  $\Delta$  对待量化系数进行取整数商和余数运算。

(3) 量化参数: 根据水印的取值和待量化系数  $X_e(k)(m_w)$  所在区间集进行量化, 若量化结果  $X'_e(k)(m_w)$  处在 A 区间集内, 则  $X'_e(k)(m_w)$  代表水印信息“1”; 反之, 量化结果  $X'_e(k)(m_w)$  处于 B 区间集内, 则  $X'_e(k)(m_w)$  代表水印信息“0”, 即:

$$X'_e(k)(m_w) \in \begin{cases} A, & V_p(k)=1 \\ B, & V_p(k)=0 \end{cases} \quad (7)$$

根据实验结果可知, 若量化的步长取得大一些, 则水印的抗攻击能力会强一些; 若量化的步长取得小一些, 则抗攻击的能力会弱一些, 但是量化步长不能无限增大, 增大到一定程度, 就会听到加水印对声音音质的改变。而且, 该量化步长值依赖于不同的原始的语音数据, 如果人工地调整量化步长  $\Delta$ , 将使得效率低下, 无法成为一种成熟的商业算法。为了能使计算机自适应地得到合适的参数, 利用听觉心理声学模型, 为每个待嵌入水印信息的数据帧 DCT 系数自适应地控制其量化步长, 并能保证水印信息不被听到, 则有:

$$\Delta' = \Delta + \log_2 LT_{\min} / 50 \quad (8)$$

式中,  $\Delta$  是基础量化步长,  $LT_{\min}$  是待嵌入水印信息的数据帧的最小掩蔽阈值。显然该数据帧的听觉掩蔽阈值越大, 则可取比较大的量化步长, 因此可以实现水印信息嵌入强度的自适应。

### 3 数字水印的提取

本文采用量化策略将水印数据组嵌入音频信号中。该算法属于盲水印检测, 即检测数字水印信号时不需要原始音频载体, 整个数字水印检测过程包括 3 个关键环

节: 同步码、水印序号和水印信号的提取。

#### 3.1 同步码的检测

将音频信号  $S$  分段, 每段的数据信号再化分为  $L_1 + L_2 + L_3$  个子段, 假设各个子段的长度为  $k$ 。检测同步码时, 将信号  $S(1) \sim S(L_1 \times k)$  分成  $L_1$  段, 分别进行 DCT 变换, 从中提取到  $L_1$  位待检测同步码  $\{b_n\}$  与原始的 bark 码  $\{a_n\}$ , 根据公式(4)进行自相关运算, 得到自相关系数  $C_{a,b}(0)$ , 若  $C_{a,b}(0) \geq K/L_1$ , 则认为  $\{b_n\}$  为同步码 ( $K$  为同步码的判断阈值)。

#### 3.2 水印信号的提取

通过同步码对水印元素定位, 1 个同步码对应 1 个水印序号和 1 个水印元素, 将水印序号与水印元素对应起来, 水印序号代表水印元素在序列中的位置, 这样所有水印元素的位置都被确定下来。若受到随机裁剪攻击或其他攻击, 必然会丢失部分水印信息, 若原水印总数为  $P$ , 如果只提取到  $m$  个水印, 说明有  $P-m$  个水印丢失。将缺失的水印元素被置 0, 使提取的水印与原水印有相同的长度, 以此来恢复水印图像。这  $P-m$  个黑点将均匀分布在整个图像上而不会形成整块区域。

#### 3.3 水印图像的恢复

将所有的语音数据段处理后, 得到一个水印混沌序列  $V'_1$ , 再对  $V'_1$  与原混沌序列进行“异或”运算, 得到解密后的水印信息  $V'$ , 对其进行升维处理, 最终可以把图像水印恢复出来。

$$H(m_1, m_2) = V'(k), \quad 1 \leq m_1 \leq M_1, \quad 1 \leq m_2 \leq M_2 \quad (9)$$

由于嵌入的水印是视觉可辩别的二值图像, 可在视觉上直接判断, 另外也可通过比较归一化系数和位误码率, 很容易地对数字音频信号是否发生篡改或破坏作出判断。

$$NC(V, V') = \sum_{i=1}^{n \times n} V(i) \times V'(i) / \sum_{i=1}^{n \times n} V^2(i) \quad (10)$$

$$BER = \frac{\text{错误 bit 数}}{\text{总 bit 数}} \times 100\% \quad (11)$$

### 4 实验结果与分析

在本实验中, 原始信号采用的量化精度为 8 bit、采样率为 22.05 kHz、长度为 26 s 的数字音频信号, 采用  $L_1=12$  bark 码作为同步码, 水印图像是  $42 \times 42$  的“程”二值图像, 总的水印元素为 1764 个。由于  $\log_2(1764) \approx 10.78$ , 所以水印序号长度定为  $L_2=11$ , 仅在每个子段的 DCT 系数的第 2 个位置上嵌入水印信息, 所以  $L_3=1$ , 每个子段的数据长度取为 8, 量化步长  $\Delta$  取值为 0.68,  $k=8$ 。水印的嵌入结果如图 4 所示。

图 4(a)、(b) 两部分感觉不到任何的差异性, 经过试听后, 从听觉上也感觉不到任何的差异, 说明嵌入的水印具有较好的透明性。为测试方案的性能, 在 MATLAB 实验环境下对水印系统进行了多组攻击仿真测试: (1) 加入高斯白噪声; (2) 低通滤波(采用 10 阶截止频率为 6 kHz 的巴特沃斯滤波器); (3) 重采样(先采样 44 100 Hz, 然后再

