

利用抖动量化和信噪比的自适应鲁棒音频盲水印算法

黄雄华^{1,2} 王宏霞² 蒋伟贞³ 崔更申¹

(1 西南交通大学信息科学与技术学院,成都市二环路北一段 111 号 610031)

(2 桂林电子科技大学广西可信软件重点实验室,桂林市金鸡路 1 号 541004)

(3 暨南大学信息科学技术学院,广州市黄埔大道西 601 号 510084)

摘要:提出了基于抖动量化和信噪比(SNR)的自适应鲁棒音频盲水印算法。在分析经典量化方法存在的安全问题的基础上,提出了抖动量化嵌入和提取水印的方法,抖动量由密钥控制的混沌通过映射产生,加强了系统的安全性。针对现有水印算法需通过反复实验确定合适嵌入强度的问题,算法基于音频内容的信噪比公式自适应地计算水印嵌入强度。实验表明,本算法是不可听的,对常规信号处理操作是鲁棒的。

关键词:音频水印;信噪比;抖动量化;混沌;小波变换

中图分类号: TN91

自适应水印嵌入是水印的嵌入强度、嵌入位置、嵌入信息量随着音频数据各部分特性的不同而自适应改变的嵌入方法,自适应音频水印是数字水印技术发展的重要趋势^[1]。现有的自适应嵌入方法很多基于心理声学模型和掩蔽效应,如文献[2]中的水印方法。基于心理声学模型的听觉门限计算开销大;也有不少利用信噪比设计的自适应水印算法^[3-4]。事实上,对非线性的复倒谱变换而言,复倒谱域是不能应用信噪比公式的,并且量化过程只考虑了嵌入水印后的参数与嵌入水印前在相同区间的情况。

也有利用音频的能量等自身特性设计的自适应水印方法。其量化步长为^[5-6]:

$$s = 2^{\lfloor \log_2(|k| \cdot m) \rfloor} \quad (1)$$

嵌入强度即量化步长与 k 、 m 相关,实现了自适应的水印嵌入。相比用实验的方法确定最佳量化长度,自适应确定水印嵌入强度能避免反复实验的缺陷,因而更具实际使用意义。

本文设计了一种基于 SNR 和抖动量化的音频水印方案,通过 SNR 自适应确定水印嵌入强度,实现不可听性和鲁棒性折衷的同时避免反复实验确定合适水印强度的不足;通过抖动量化实现水印嵌入和检测,避免经典量化方法的安全隐

患;水印序列通过混沌产生,可以避免采用二值图像作为水印引起的传输过程中被篡改或者替换的安全问题^[7-8]。

1 经典量化原理和抖动量化

1.1 经典量化原理

经典量化算法表示为:

$$s(x) = q(x; m; \Delta) \quad (2)$$

信号 x 可以分为双边量化和单边量化。

1) 双边量化。双边量化信号 x 的取值范围属于整个实轴,即 $x \in R$ 。如图 1 所示, Δ 表示量化步长。量化过程首先将信号 x 映射到 A 区或者 B 区,然后根据消息 m 将 x 调制到 A 区或者 B 区的中点,量化误差 $|e| \in [0, \Delta)$ 。



图 1 双边量化原理

Fig. 1 Dual Side Quantization Principle

2) 单边量化。若信号 x 的取值范围仅属于实轴的正半轴,如图 2,首先将信号 x 映射到 A 区

或者B区,然后根据消息 m 将 x 调制到A区或者B区的两端,量化的误差范围 $|e| \in [0, \Delta)$ 。

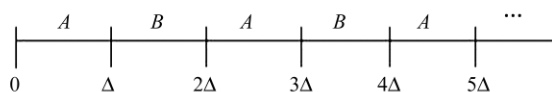


图2 单边量化原理

Fig. 2 Single Side Quantization Principle

1.2 经典量化方法的安全性分析

假设 x_1, x_2, x_3 分别是经过量化后的值,且 $x_1 > x_2 > x_3$, 根据 Kerchhoff 算法公开的原则,只要水印嵌入位置是非加密的, x_1, x_2, x_3 为已知量。假设量化步长为 $\Delta, k_1, k_2, k_3 \in \mathbf{z}$, 且存在下面的关系式:

$$x_1 = (k_1 + 0.5) \times \Delta \quad (3)$$

$$x_2 = (k_2 + 0.5) \times \Delta \quad (4)$$

$$x_3 = (k_3 + 0.5) \times \Delta \quad (5)$$

$y_1 = x_1 - x_2, y_2 = x_2 - x_3, u_1 = k_1 - k_2, u_2 = k_2 - k_3, u_1, u_2 \in \mathbf{z}$, 由式(3)~(5), 有:

$$y_1 = u_1 \times \Delta \quad (6)$$

$$y_2 = u_2 \times \Delta \quad (7)$$

由式(6)和式(7), 得:

$$\frac{u_1}{u_2} = \frac{y_1}{y_2} = \frac{\rho_1}{\rho_2} \quad (8)$$

式中, ρ_1 与 ρ_2 是互质正整数。假设 $u_1 = \rho_1, u_2 = \rho_2$, 利用式(6)和式(7)检验 Δ 的值是否一致, 如果一致, 假设成立; 否则, 假设 $u_1 = 2\rho_1, u_2 = 3\rho_2$, 重新利用式(6)和式(7)检验 Δ 的值, 如此反复, 直到找到满足式(6)和式(7)的 Δ 值, 重复计算的次数小于 $\min((x_1 - x_2)/\Delta, (x_2 - x_3)/\Delta)$, 运算开销不高。经典量化方法一旦确定量化长度, 如果水印位置没有加密, 就面临受攻击的风险, 存在安全隐患。

1.3 抖动量化

抖动量化可用公式表示为:

$$s(x) = q(x + d(k, m); m; \Delta) - d(k, m) \quad (9)$$

其中, $d(k, m)$ 表示抖动量, 为了实现盲检测, 使 $d(k, 0) = d(k, 1)$, 标记为 $d(k)$, 为使抖动量化和经典的量化方法具有相同的误差范围, 约束 $d(k) \in [-\Delta/2, \Delta/2]$, 为使 $d(k)$ 在 $[-\Delta/2, \Delta/2]$ 上均匀分布, 采用如下的方法生成抖动向量 $d(k)$ 。

首先, 利用 Logistic 动力方程生成混沌序列:

$$x_{k+1} = ux_k(1 - x_k) \quad (10)$$

式中, $x_k \in (0, 1), 0.369\ 945\ 6 < u \leq 4$ 。然后, 作如下的映射:

$$d(k) = -\Delta \times (x_k - 0.5) \quad (11)$$

容易验证, $d(k)$ 满足设计要求, 且 $(x + d(k)) - q(x + d(k); m; \Delta) \in [0, \Delta)$ 。

式(9)中等式右边是一个以原信号所在区间为中心, 在其左右相邻区间随机分布的随机量, 避免了经典量化的安全隐患。

2 基于 SNR 的自适应量化步长与水印的生成

本文提出的算法在变换域需嵌入水印。首先对信号作 n 级小波分解, 然后量化其近似分量系数。假设时域信号为 $x_i, i \in [1, l], l$ 是信号样本数, n 级小波分解的近似分量为 ca^n , 细节分量分别为 $cd^n, cd^{n-1}, \dots, cd^2, cd^1$, 量化步长为 Δ ; 若量化的系数个数为 L , 根据量化原理, 每个系数的量化误差 $e_c \in [0, \Delta), c \in [1, L]$ 。假设 $y_i (i \in [1, l])$ 是小波逆变换后信号的样本, 则 DWT 变换满足式(12)^[9]:

$$\sum_i (x_i^2 - y_i^2) = \sum_c e_c^2 \quad (12)$$

信噪比可表示为:

$$\text{SNR} = 20 \lg \left[\frac{\sqrt{\sum_i x_i^2}}{\sqrt{\sum_i (x_i^2 - y_i^2)}} \right] \quad (13)$$

综合式(12)、式(13), 有:

$$\text{SNR} = 20 \lg \left(\frac{\sqrt{\sum_i x_i^2}}{\sqrt{\sum_c e_c^2}} \right) \quad (14)$$

根据式(14), 对设定的信噪比初值 SNR_0 , 可计算出最大水印嵌入强度, 即最大量化步长 Δ :

$$\Delta \leq \sqrt{E/L} \times 10^{\text{SNR}_0/10} \quad (15)$$

式中, $E = \sum_i x_i^2$ 表示帧信号能量。

为避免二值图像作为水印在传输中的安全问题, 采用混沌序列排序后再取索引值的二进制表示的 LSB (least significant bit) 作为水印比特, 具体如下。

1) 利用 Logistic 方程(10)生成混沌序列 x ;

2) 按升序排序 x , 排序后序列为向量 x^T , 排序下标索引向量为 Ind ;

3) 用二进制比特表示每个索引, 即 $\text{Ind}_i = (b_{i1}, b_{i2}, \dots, b_{i1})_2$, 取每个索引的最后一比特 b_{i1} 为水印比特 w_i 。

3 水印嵌入算法和检测算法

3.1 水印的嵌入方法

用 $A = \{a_i, i \in [1, N]\}$ 表示音频信号, 其中 N 是样本数。水印的嵌入算法具体步骤如下。

1) 对音频作分帧处理,每帧长为 l , 音频帧记为 f_g , 帧内各样本表示为 $f_g(h)$, $g \in [1, \lfloor N/l \rfloor]$, $h \in [1, l]$, g 表示帧序号, h 表示帧中样本序号。

2) 计算每帧的能量 E_i , 并按从大到小排序 E_i , I 是排序结果下标向量。

3) 对步骤 2) 中的 I_1, I_2, \dots, I_z 所对应的帧 $f_{I_1}, f_{I_2}, \dots, f_{I_z}$ 分别作 n 级小波变换, z 为水印重复嵌入的次数。采用重复嵌入是为了降低嵌入水印后帧能量大小顺序变化带来的水印检测误码率。

4) 按式(15)和设定的信噪比初值 SNR_0 , 计算 f_{I_z} 的量化步长 Δ 。

5) 生成水印比特序列 $w_i (i \in [1, L])$, 生成抖动向量 $d(i), i \in [1, L]$, L 是水印序列长。

6) 选择近似分量的前 L 个低频系数、量化步长 Δ 、抖动向量 $d(i)$, 并利用抖动量化法嵌入步骤 5) 生成的水印。

7) 嵌入水印后的小波系数作相应的逆小波变换, 并和没嵌入水印的音频样本连接, 从而得到水印音频。

3.2 水印的检测方法

1) 待检测音频分帧。帧长为 l , 音频帧记为 f'_g , 样本记为 $f'_g(h)$, $g \in [1, \lfloor N/l \rfloor]$, $h \in [1, l]$, g 表示帧序号, h 表示样本序号;

2) 计算每帧的能量 E'_g , 并按从大到小排序 E'_g , 其中 I' 表示排序结果下标向量;

3) 对步骤 2) 中的 I'_1, I'_2, \dots, I'_z 所对应的帧 $f'_{I'_1}, f'_{I'_2}, \dots, f'_{I'_z}$ 分别作 n 级小波变换, 各帧对应的近似分量分别为 $cA(I'_1, j), cA(I'_2, j), \dots, cA(I'_z, j)$, 其中 $j \in [1, l/2^n]$, 表示帧近似分量系数序号;

4) 生成水印比特序列 $w_i (i \in [1, L])$, 按第 1 节的方法生成抖动向量 $d(i) (i \in [1, L])$, 其中 L 是水印序列长;

5) 对步骤 3) 中每帧的近似分量的前 L 个近似系数, 采用嵌入水印时的量化步长 Δ , 采用步骤 4) 的抖动向量 $d(i)$, 采用抖动量化的逆过程提取比特序列 $b(\alpha, \beta), \alpha \in \{I'_1, I'_2, \dots, I'_z\}, \beta \in [1, L]$, 其中 $b(\alpha, \beta)$ 表示提取的第 α 帧的第 β 比特;

6) 判断水印比特, 重复出现的次数多的比特值作为提取的水印比特:

$$\begin{cases} w'_i = 1, b(I'_1, i) + b(I'_2, i) + \dots + b(I'_z, i) \geq z/2 \\ w'_i = 0, b(I'_1, i) + b(I'_2, i) + \dots + b(I'_z, i) < z/2 \end{cases} \quad (16)$$

7) 用生成的水印与提取的水印作归一化运算并与设定的门限 ρ 比较, 判断水印存在与否。

4 算法性能分析

4.1 水印容量分析

根据算法, 帧长为 l , 经过 n 级小波变换, 重复嵌入率为 z , 采样率为 f_s , 水印容量的最大值为:

$$C = \frac{f_s}{z \times 2^n} \quad (17)$$

式中, f_s 是指在一定的情况下, 水印容量与重复嵌入率成反比, 与小波分解级的指数成反比。

4.2 不可听性分析

本文采用信噪比来评价音频的听觉质量。

根据量化原理, 系数被量化在本区间的概率为 0.5, 误差 $|\epsilon_1| \in [0, 0.5\Delta]$; 系数被量化在相邻区间的概率为 0.5, 此时误差为 $|\epsilon_2| \in (0.5\Delta, \Delta)$ 。则可认为误差是均匀分布的, 误差的期望值为:

$$E(|e|) = 0.5 \times (1/2 \times 0.5\Delta) + 0.5 \times [1/2 \times (\Delta + 0.5\Delta)] = 0.5\Delta \quad (18)$$

根据量化长度公式(15), 最大量化长度为:

$$\Delta = \sqrt{E_0/L \times 10^{\text{SNR}_0/10}} \quad (19)$$

式中, $E_0 = \sum_h f_k^2(h)$ 表示参考帧能量, $k \in [1, \lfloor N/l \rfloor]$, 对任意嵌入水印的帧:

$$\text{SNR} = 10\lg[E_g / (e_1^2 + e_2^2 + \dots + e_L^2)] \leq 10\lg[E_g \cdot L / (\sum_{k=1}^L |e_k|^2)] = 10\lg\{E_g / [(E(|e|))^2 \times L]\} \quad (20)$$

联立式(18)~式(20)可得:

$$\text{SNR} \leq \text{SNR}_0 + 6.021 + 10\lg(E_g/E_0) \quad (21)$$

即嵌入水印的每帧信噪比范围在 $(\text{SNR}_0, \text{SNR}_0 + 6.021 + 10\lg(E_g/E_0))$ 间。式(21)可用于确定含水音频帧的 SNR 理论范围。

为了较为全面地评估音频的不可听性, 本文也采用听觉质量主观区分度 SDG (subjective difference grades) 作为评价指标, SDG 分值的含义如表 1 所示。

表 1 听觉质量主观区分度 SDG 标准

Tab. 1 Aural Quality Subjective Difference Grades

Criterion	
SDG	描述
0	不可感觉
-1	可感觉但不刺耳
-2	轻微刺耳
-3	刺耳
-4	非常刺耳

4.3 安全性分析

算法中抖动量化的量化长度空间等价于混沌的密钥空间,考虑到计算机的字长效应,其密钥空间为 2^φ , φ 为计算机字长;同时,水印由混沌序列映射生成,其密钥空间也为 2^φ ,因此,总的密钥空间为 4^φ ,对现在 32 位、64 位甚至字长更高的计算机是难以破译的,保证了算法的安全性。

5 实验结果

5.1 不可听性测试

为了验证提出算法的不可听性,本文分别选择了 1 806 336 个样本,采样率为 44.1 kHz,16 bit 量化的单声道高节奏的摇滚乐剪辑,记为 type1;中等节奏的流行乐剪辑,记为 type2;节奏平缓的古典乐剪辑,记为 type3,进行一系列实验,水印重复嵌入率为 40,帧样本为 4 096,其信号波形如图 3(a)~3(c)所示,嵌入水印后的波形如图 4(a)~4(c)所示,波形上难以觉察到变化。

设置 SNR 的初值为 20 dB,IFPI(international federation of the phonographic industry)规定嵌入水印音频的信噪比大于 20 dB,按照式(19)计算的量化步长嵌入水印后,3 种类型音频按式(21)计算的理论 SNR 值上限和实测 SNR 值分别如图 5(a)~5(c)所示。图 5 表明,实测 SNR 值均在理论门限的下方,并且大于 20 dB。SDG 听觉实验表明,当 $SNR > 22$ dB 时,基本难以觉察到音频的变化,从图 5 的 SNR 曲线可以看出,SNR 均在 22 dB 之上。SDG 实验由 10 个人一起完成,实验结果为零,证明了水印的不可听性。

5.2 鲁棒性测试

为检验算法对常规信号处理的鲁棒性,本文进行了包括添加噪声、重采样、剪切、低通滤波和有损压缩的测试实验。截止频率为 8 kHz 的低通滤波,信噪比 38 dB 的高斯白噪声,下采样至 22.05 kHz 后再上采样至 44.1 kHz 的重采样,16~8~16 bit 的重量化,将音频信号进行 MP3 压缩至 64 KB/s 再解压缩。

测试结果采用式(22)的归一化相关值(normalized cross-correlation, NC)来度量:

$$NC = \frac{\sum_i w_i \cdot w'_i}{\sqrt{\sum_i w_i^2} \sqrt{\sum_i w_i'^2}} \quad (22)$$

式中, w_i 是由混沌序列生成的水印序列; w'_i 是提取的水印序列。NC 值越大,水印算法的鲁棒性越好。测量结果如表 2 所示。

表 2 的 NC 值表明,总体上本文算法抵抗常

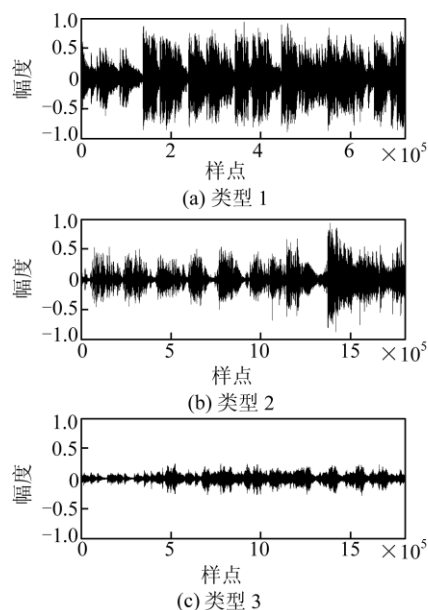


图 3 原始音频波形

Fig. 3 Original Audio Waveform

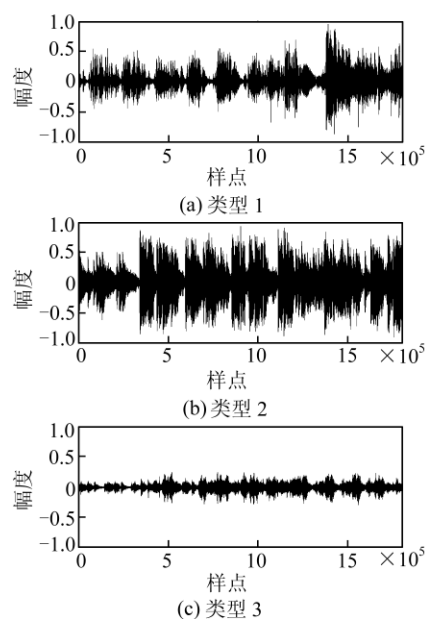


图 4 含水印音频波形

Fig. 4 Watermarked Audio Waveform

规信号处理的能力比文献[3,10]强。

表 2 常规信号处理下的 NC 值

Tab. 2 NC Value in Common Signal Process

操作类型	本文音频类型			文献[3]	文献[10]
	1	2	3		
无操作	1	1	1	1	1
MP3 压缩	1	1	1	0.79	0.99
低通滤波	1	1	1	0.90	0.82
高斯噪声	1	0.98	0.96	0.92	0.99
重量化	1	1	0.98	-	0.99
重采样	1	1	1	0.95	0.99

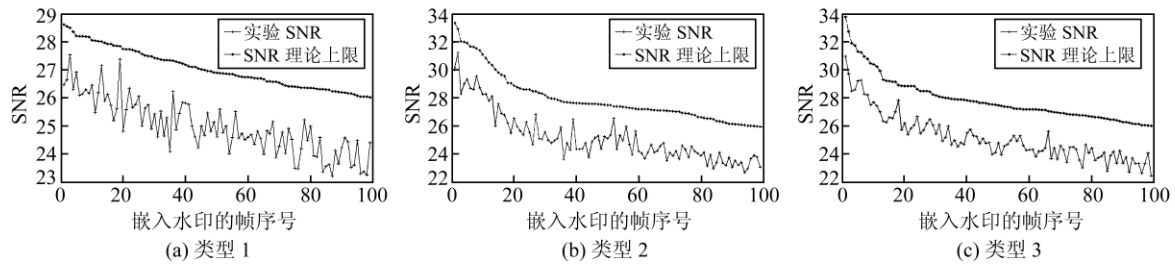


图 5 信噪比理论和实验曲线

Fig. 5 Curse of SNR Theory and Experiment

参 考 文 献

- [1] 王宏霞. 基于内容的自适应音频水印研究综述[J]. 西南交通大学学报, 2009, 44(3): 430-437
- [2] 冯涛, 韩纪庆. 基于听觉特性的音频水印能量估计及自适应嵌入方法研究[J]. 声学学报, 2006, 31(1): 48-54
- [3] 付永庆, 孙滢. 一种新的基于小波变换的自适应量化音频水印算法[J]. 哈尔滨工程大学学报, 2007, 28(9): 1 052-1 055
- [4] 黄雄华, 王宏霞. 一种 DCT 域自适应音频水印算法[J]. 计算机应用研究, 2009, 26(8): 2 989-2 992
- [5] 王向阳, 杨红颖, 牛盼盼. 性能优越的混合域数字音频盲水印算法[J]. 通信学报, 2007, 28(2): 109-114
- [6] 王向阳, 祁薇. 用于版权保护与内容认证的半脆弱音频水印算法[J]. 自动化学报, 2007, 33(9): 936-940
- [7] 范明泉, 王宏霞. 基于音频内容的混合域脆弱水印算法[J]. 铁道学报, 2010, 32(1): 118-122
- [8] 王宏霞, 范明泉. 基于质心的混合域半脆弱音频水印算法[J]. 中国科学(信息科学), 2010, 40(2): 313-326
- [9] 刘瑞桢, 谭铁牛. 水印能量估计的一般性框架[J]. 计算机学报, 2001, 24(3): 242-246
- [10] 谭良, 吴波, 刘震, 等. 一种基于混沌和小波变换的大容量音频信息隐藏算法[J]. 电子学报, 2010, 38(8): 1 812-1 818

第一作者简介: 黄雄华, 博士生, 研究方向: 数字水印。
E-mail: 494777@163.com

An Adaptive Robust Audio Blind Watermarking Algorithm Using Dithering Quantization and SNR

HUANG Xionghua^{1,2} WANG Hongxia¹ JIANG Weizhen³ CUI Gengshen²

(1 School of Information Science and Technology, Southwest Jiaotong University, 111 Section 1, North Second Ring Road, Chengdu 610031, China)

(2 Guangxi Key Lab of Trusted Software, Guilin University of Electronic Technology, 1 Jinji Road, Guilin 541004, China)

(3 School of Information Science and Technology, Jinan University, 601 West Huangpu Road, Guangzhou 510084, China)

Abstract: An adaptive robust audio blind watermarking algorithm based dither quantization and signal to noise ratio is proposed. Based on analyzing open secure question of classical quantization means, dithering quantization is proposed to embed and extract watermarking. Dithering vector is generated by mapping chaos sequences which is controlled by keys, so system security is improved. To aiming at the problem of existing watermarking algorithms to ascertain the appropriate strength by repeating experiment or experiential data, the proposed algorithm calculates adaptively watermarking strength by SNR formula based on audio content. Experimental results prove the proposed algorithm is inaudible and robust for common signal process manipulations.

Key words: audio watermarking; SNR; dither quantization; chaos; wavelet transform