

基于重心的抗同步攻击的音频水印算法

黄雄华^{1,2} 王宏霞¹ 蒋伟贞³ 崔更申²

摘要 抗同步攻击是音频水印领域公开的难题. 相对重心是音频的一个稳定的全局特征量, 对 TSM (Time scale modification) 等同步攻击不敏感. 本文提出一种基于相对重心的抗同步攻击的音频水印算法, 提出的算法通过量化音频的相对重心生成水印, 生成的水印由混沌序列加密后以量化的方法嵌入载体音频三级小波近似分量均值中. 为了降低同步攻击后同步位置的连续偏移, 水印嵌入提取过程采用了均匀分帧的方法. 实验结果表明提出的算法能抵抗常规信号处理操作, 同时能抵抗 20% 的 TSM 操作 (包含音调不变 TSM 和重采样操作) 和 1/10 的抖动操作, 有很好的鲁棒性.

关键词 音频水印, 相对重心, TSM, 均匀分帧

引用格式 黄雄华, 王宏霞, 蒋伟贞, 崔更申. 基于重心的抗同步攻击的音频水印算法. 自动化学报, 2013, 39(8): 1321–1329

DOI 10.3724/SP.J.1004.2013.01321

Audio Watermarking Algorithm Against Synchronization Attacks Based on Gravity of Center

HUANG Xiong-Hua^{1,2} WANG Hong-Xia¹ JIANG Wei-Zhen³ CUI Geng-Shen²

Abstract It is an open difficult problem in audio watermarking field to resist synchronization attacks. An audio watermarking algorithm against synchronization attacks is proposed based on the relative gravity of center, which is a globally steady feature parameter of audio and is insensitive to time scale modification (TSM) operation and other synchronization attacks. The watermark is generated by quantifying relative gravity of center in our method. Firstly, the generated watermark is encrypted by chaos sequence and then the encrypted watermark is embedded into the third level wavelet approach coefficients of audio by quantification method. The method of dividing frame evenly is applied to embedding and extracting watermarking in this paper in order to reduce synchronization place offset continuously after synchronization attacks. Experiment results demonstrate that the proposed algorithm can resist common signal processing and 20% TSM (including pitch-invariant TSM and resampling manipulation), 1/10 dithering manipulation and has very good robustness.

Key words Audio watermarking, normalized gravity of center, time scale modification (TSM), dividing frame evenly

Citation Huang Xiong-Hua, Wang Hong-Xia, Jiang Wei-Zhen, Cui Geng-Shen. Audio watermarking algorithm against synchronization attacks based on gravity of center. *Acta Automatica Sinica*, 2013, 39(8): 1321–1329

第 1 代数字音频水印算法, 大多将水印嵌入时域样本或者变换域系数, 水印嵌入没有考虑音频内容和特性, 且数据没有嵌入感知重要区域. 时域算法存在对信号处理操作鲁棒性差的问题, 变换域算法常利用人类听觉系统 (Human audio system, HAS) 特性的音频掩蔽效应、心理声学模型^[1–4] 及扩频技术^[5–6] 的思想以提高鲁棒性, 但遭受同步攻击后重新定位水印的嵌入位置困难、抗同步攻击能力差.

自 1999 年 Kutter 等^[7] 提出第 2 代数字水印的

概念以来, 已有一些基于音频内容的水印算法^[8] 的报道, 其基本思想是利用音频载体自身特性生成水印或者水印嵌入考虑音频自身的特性^[9–14].

采用二值图像作为水印会带来因传输过程中二值图像被篡改引起的虚警和认证无法判别的问题^[15–16]. 因此, 我们通过量化相对重心的方法生成水印, 使水印与音频的内容密切相关.

音频即使遭受 10% 的 TSM 操作、1/10 的抖动等同步攻击, 也难以觉察音频质量的变化, 并且攻击成本很低^[17]. 因此抵抗同步攻击是音频水印领域的公开难题^[13, 18–19].

现有抗同步攻击的音频水印研究成果可以归为 4 类: 穷举搜索、显示同步、恒定水印、隐含同步.

穷举搜索的缺陷是计算开销大. 显示同步在水印前嵌入巴克码^[20–21]、混沌序列、 m 序列^[22]、PN 序列^[23] 等作为同步信号. 按 Kerchhoff 算法公开原则, 同步信号容易遭受攻击. 同时, 因引入了同步信

收稿日期 2011-10-24 录用日期 2012-07-22
Manuscript received October 24, 2011; accepted July 22, 2012
本文责任编辑 戴琼海

Recommended by Associate Editor DAI Qiong-Hai
1. 西南交通大学信息科学与技术学院 成都 610031 2. 桂林电子科技大学广西可信软件重点实验室 桂林 541004 3. 暨南大学信息科学技术学院 广州 510084

1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031 2. Guangxi Key Lab of Trusted Software, Guilin University of Electronic Technology, Guilin 541004 3. School of Information Science and Technology, Jinan University, Guangzhou 510084

号,降低了水印的容量和不可听性.

文献[24]利用3级小波近似系数均值的稳定性设计恒定水印,水印能抵抗2%的TSM攻击;文献[25]利用倒谱均值的稳定性设计恒定水印,水印能抵抗1%的TSM攻击.这与IFPI(International Federation of the Phonographic Industry)规定算法能抵抗10%的TSM操作还有较大的差距.

文献[26]利用Mel频率倒谱系数的稳定性实现隐含同步;文献[11]以能量迅速爬升到极值的点作为同步点;文献[12]通过基音实现同步;Wei等^[27]以能量的峰值点作为同步点;Mansour等^[9]把音频包络线小波分解系数的极值点作为显著点.上述隐含同步的方法都是以提取特征点为同步点,其缺陷是特征点是否明显与音频类型相关,算法只适用于部分类型音频;并且特征点在受攻击后容易消失,方案应用受限.

上述同步方案还存在如下共同问题:1)针对某一种同步攻击设计,一般不能同时抵抗多种类型的同步攻击.如文献[20-21]的方法用于抵抗剪切攻击,文献[9,23]的方法用于抵抗TSM攻击,文献[28-29]的方案仅能抵抗重采样操作;2)满足不了IFPI能抵抗10%的TSM攻击的规定,如文献[23]的算法只能抵抗4%的TSM操作,文献[25]的算法只能抵抗1%的TSM操作,文献[11]的方法只能抵抗2%的TSM攻击;3)少有文献研究既改变音调又改变节奏的重采样操作的攻击;4)现有同步方案多基于样本严格对齐的同步策略,而遭受TSM、抖动、随机剪切等同步操作后保持样本严格对齐是十分困难的.

基于上述抗同步攻击水印方案的问题,我们提出用均匀分帧的方法解决水印音频遭受同步攻击后同步位置连续偏移带来的失步,所提出的方法不需要嵌入同步码,不需要严格的样本对齐;利用现有的小波近似系数均值具有稳定性的研究成果^[24],通过量化3级小波近似系数均值的方法嵌入和提取水印.

实验结果表明,该算法不但对常规的信号处理操作具有很好的鲁棒性,而且能同时抵抗 $\pm 20\%$ 的音调不变TSM操作、 $\pm 20\%$ 的重采样TSM操作、1/10的抖动操作和 $\pm 20\%$ 的PSM(Pitch scale modification)操作.

1 相对重心

地球上的物体受到地心的引力,其合力大小称为物体的重量,合力的作用点称为物体的重心.更进一步,厚度均匀的平面物体在 x 轴方向的重心位置由式(1)计算.

$$x_c = \frac{\int_A x dA}{\int_A dA} \quad (1)$$

其中, dA 表示微分面积.对于规则形状组合而成的物体,每个规则形状的形心(重心)已知,可用如下的公式计算物体在 x 轴上的重心:

$$x_c = \frac{\sum_{i=1}^n A_i x_i}{\sum_{i=1}^n A_i} = \frac{\sum_{i=1}^n \rho A_i x_i}{\sum_{i=1}^n \rho A_i} = \frac{\sum_{i=1}^n G_i x_i}{\sum_{i=1}^n G_i} \quad (2)$$

其中, ρ 表示匀质物体单位体积的重量.

对于数字音频而言,可以把音频曲线和坐标轴组成的曲线图形看作均匀的平面物体,其样本值 m_i 看作平面上一个个相邻的规则部分的重心,其重量积聚于重心之上,在坐标系中,依次将各个重心的坐标标记为 $(1, y_1), (2, y_2), \dots, (n, y_n)$.借用力学上重心的概念,定义音频样本的 x 轴重心坐标 x_c^A :

$$x_c^A = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i} \quad (3)$$

再定义重心的位置与物体在对应的数轴上的最大坐标位置的比为归一化重心,即相对重心:

$$x_c^r = \frac{x_c^A}{n} \quad (4)$$

2 相对重心对 TSM 攻击稳定的证明

数字音频标记为 $F_d = \{f(i) | i = 1, 2, \dots, n\}$,其对应模拟形式为 $F_a = \{f(t) | t \in [0, T]\}$,其重心

$$t_c = \frac{\int_0^T t |f(t)| dt}{\int_0^T |f(t)| dt} \quad (5)$$

假设在时间轴上线性的伸缩因子为 α ,经过操作后,重心为

$$\hat{t}_c = \frac{\int_0^{\hat{T}} t |\hat{f}(t)| dt}{\int_0^{\hat{T}} |\hat{f}(t)| dt} = \frac{\int_0^{\alpha T} t |f(\frac{t}{\alpha})| dt}{\int_0^{\alpha T} |f(\frac{t}{\alpha})| dt} = \frac{\int_0^T \alpha t |f(t)| d\alpha t}{\int_0^T |f(t)| d\alpha t} = \alpha \cdot t_c \quad (6)$$

经过线性操作后的相对坐标为

$$\hat{t}_c^r = \frac{\hat{t}_c}{\hat{T}} = \frac{\alpha \cdot t_c}{\alpha \cdot T} = \frac{t_c}{T} = t_c^r \quad (7)$$

其中, t_c^r 表示线性操作前音频在 x 轴上的相对重心.式(7)表明线性操作后音频在 x 轴上的相对重心保

持不变, 具有稳定性. 数字音频样本是模拟音频的数字化表示形式, 忽略采样值的量化误差等, 式 (7) 也成立. TSM 操作可以近似看作时间线性尺度操作^[19]. 因此理论上经过 TSM 操作后相对重心是稳定的.

为进一步从实验上证明 TSM 操作后音频相对重心的稳定性, 选取三种不同类型的音乐: 波形变化明显的乐曲, 标记为类型 1; 有一定波形变换的乐曲, 标记为类型 2; 波形变换平缓的乐曲, 标记为类型 3. 其波形分别如图 1 (a) ~ 1 (c) 所示.

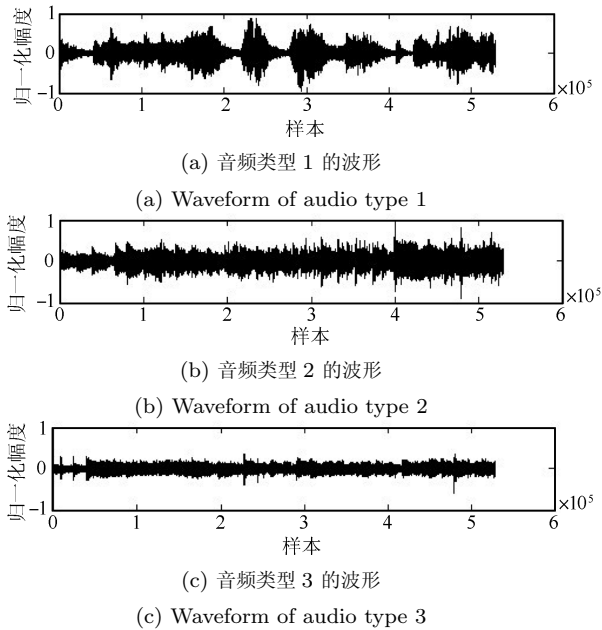


图 1 三种类型音频的原始波形图

Fig. 1 Original waveform of three type audios

分别对三种音频作 $\pm 10\%$ 的 TSM 操作 (含音调不变 TSM 操作和重采样 TSM 操作), 实验表明其相对重心变化很小. 考虑篇幅, 仅绘制音频类型 1 分为 50 帧的重心变化曲线, 如图 2 所示. 图中带 “*” 号的曲线表示原始音频相对重心的曲线, 带 “○” 号的曲线表示经过 TSM 操作后相对重心的曲线. 其他两种音频类型曲线图与图 2 类似.

图 2 表明经过 TSM 操作 (包括音调不变 TSM 和重采样 TSM) 后, 音频的相对重心位置只发生很小的变化, 具有很好的稳定性. 实验也进一步证明了式 (7) 的合理性. 第 4 节将利用音频相对重心的稳定性, 设计一种具有抵抗多种同步攻击的音频水印算法.

3 均匀分帧

现有音频水印方案, 绝大多数采用固定帧长严格样本对齐的同步方法. 但音调不变 TSM 操作、重采样 TSM 操作、抖动、随机剪切等同步攻击后, 音

频长度会发生改变, 难以重新实现严格的样本对齐. 固定帧长分帧对齐同步的方法必然引起同步位置连续偏移, 导致水印检测器失效. 本文通过均匀分帧的方法降低同步位置的连续偏移. 如果音频没有受同步攻击, 按均匀分帧的方法, 水印嵌入和检测过程中帧的样本相等; 如果受到同步攻击, 水印嵌入和检测过程中帧样本数发生变化.

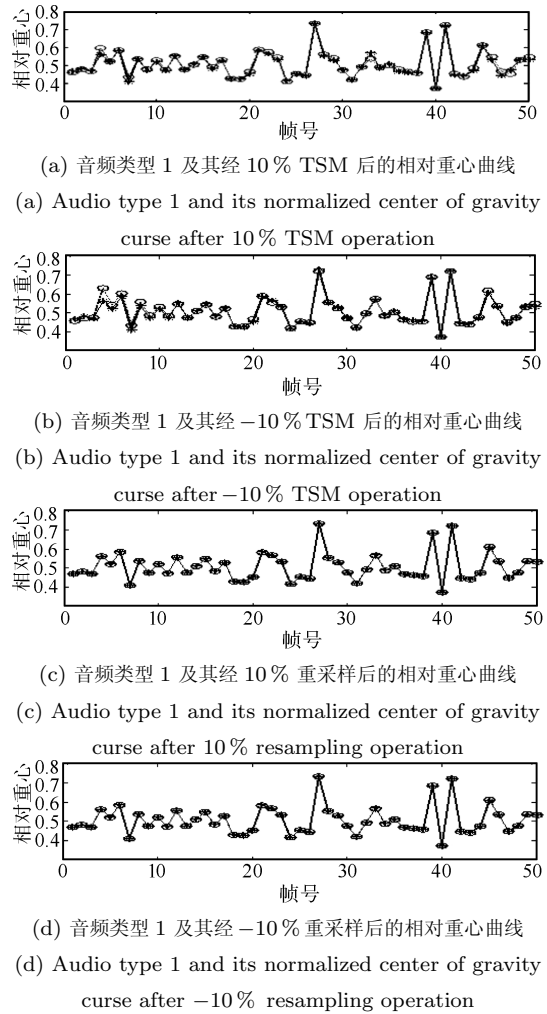


图 2 音频类型 1 及其经过 $\pm 10\%$ 的时间伸缩后的相对重心曲线图

Fig. 2 Audio type 1 and its normalized center of gravity curve after $\pm 10\%$ time scale stretch

为说明均匀分帧的合理性, 下面分别分析抖动、随机剪切、音调不变 TSM 操作、重采样 TSM 操作和 PSM 操作下均匀分帧后同步位置变化情况.

3.1 抖动

抖动是一种均匀地在每固定长样本中插入或者删去一个或者几个样本的操作.

首先分析理想状态模型. 假设音频被分为 n 帧, 每帧长 l_f , 音频尾部不够一帧长的样本数记

为 l_r ($l_r \in [0, n-1]$). 每帧相同的位置被均匀地剪切 1 个或 k 个样本. 按固定帧长的方法, 从第 1 帧开始, 剪切后每帧中偏移的样本数依次是 1 或 k , 2 或 $2k$, 3 或 $3k, \dots, n$ 或 $n \times k$, 也就是偏移成线性变化, 帧样本偏移按线性规律累积. 如果均匀分帧, 每帧有 1 个或者 k 个样本被剪切, 不存在样本的偏移累积.

其次考虑抖动被剪切的样本小于帧数 n 的情况. 假设抖动过程中剪切 k 个样本, 并且 $k < n$. 如果 $l_r \geq k$, 按均匀分帧的方法, 帧长不会改变, 帧的样本偏移最小为 1, 最大为 k , 逐步增加, 此时同步偏移和固定帧长方法相同. 如果 $l_r < k < n$, 按均匀分帧的方法, 帧长减少 1, 偏移最小为 1, 最大为 $n-k+1$; 按固定帧长的方法, 因样本被剪切后样本数量不足, 出现检测错误.

如果抖动中被剪切的样本 k 大于帧数 n , 可以看作前两种情况的组合. 首先每帧被均匀地剪切 $\text{floor}(k/n)$ 个样本, 然后在整个音频中均匀地剪切掉 $k - n \times \text{floor}(k/n)$ 个样本, 其最大偏移为 $k - n \times \text{floor}(k/n)$ 或 $n + n \times \text{floor}(k/n) - k$. 帧的样本偏移的情况类似于两种情况的组合; 如果按固定帧长分帧, 最大偏移为 k (对应于均匀分帧偏移为 $k - n \times \text{floor}(k/n)$ 的情况) 或者出现检测错误. 其中 $\text{floor}(\cdot)$ 表示向下取整函数.

综合上述 3 种情况, 对删去样本的抖动操作而言, 均匀分帧引起的同步位置的偏移不会大于固定帧长分帧方法引起的同步位置的偏移, 可以降低同步位置偏移引起的水印检测误差; 添加样本的抖动操作可作类似分析.

3.2 随机剪切

如果被剪切的样本 $k \leq l_r$, 此时, 按均匀分帧方法, 帧长不发生变化, 其水印提取情况完全与固定帧长相同.

其次, 考虑被剪切的样本 $l_r < k < n$ 时的情况. 此时, 按均匀分帧的方法, 帧长发生变化, 帧长 $l_f^T = l_f - 1$, 其剪切位置如果在音频起始部位, 偏移从最大值 k 开始, 逐步递减; 如果剪切位置在音频中部位置, 剪切之前的帧偏移样本数从 1 开始, 逐步增加到 t ($t < k$), 到剪切处突变为 $k-t$, 然后开始逐步递减. 如果按固定帧长方法, 自剪切之处起, 偏移即为 k , 显然均匀分帧的方法可以降低偏移.

对 $l_r < n < k$ 的情况, 与第 2 种情况类似, 不同之处在于帧偏移递变速度为 $\text{floor}((k+l_r)/n)$.

3.3 重采样 TSM

重采样同时改变了音频的音调和节奏, 样本的数量和幅度同时发生改变. 在不考虑样本幅度变化的情况下, 可以认为重采样是一种理想的抖动操作,

每帧被均匀地删掉若干个样本或添加若干个样本. 均匀分帧不会引起样本的偏移, 只是每帧被剪切或者添加若干个样本, 并且样本幅度发生微小变化. 由于音频信号的连续性, 重采样后的音频样本和原始音频样本有高度的相关性, 因此样本幅度的变化可以忽略.

3.4 音调不变 TSM

音调不变 TSM 是一种周期性的按某种方法对部分样本按某种技术进行叠加处理的操作, 抖动可以看作音调不变 TSM 操作的一种特殊形式. 音调不变 TSM 可以看作一种非理想状态的抖动. 按均匀分帧引起的偏移类似于抖动.

3.5 PSM

PSM 和 TSM 间具有双向性^[30]: 改变音调的信号可以在重采样后通过音调不变 TSM 获得, 而音调不变 TSM 信号可以在 PSM 后通过重采样获得. 其关系如图 3 所示 (注: 图 3 来自文献 [31]).

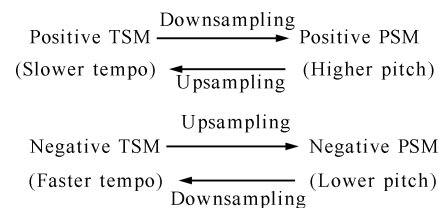


图 3 PSM 和 TSM 的相互转化关系图
Fig. 3 Mutual conversion relation graph of PSM and TSM

因此, PSM 可以看作是音调不变 TSM 和重采样 TSM 的组合操作, 均匀分帧的方法不会引起同步位置的连续偏移, 并且保留了原帧的最大信息.

综上, 抖动、随机剪切、音调不变 TSM 和重采样 TSM 操作及 PSM 操作按均匀分帧的方法相对于固定帧长的样本严格对齐而言, 同步位置的偏移不大于固定分帧的方法.

4 基于相对重心和均匀分帧的抗同步攻击的音频水印算法

本节提出一种基于重心和均匀分帧的音频水印算法. 利用相对重心生成水印、小波近似系数均值具有稳定性的现有研究成果, 将提取的水印通过混沌比特序列加密后嵌入 3 级小波分解近似系数的均值, 嵌入方法是量化近似系数的均值. 水印的嵌入和提取采用均匀分帧的方法, 帧长是保密的. 水印检测时通过计算帧 3 级小波近似系数均值并按量化的方法提取水印, 不必通过样本严格对齐方法同步.

4.1 水印嵌入算法

水印的嵌入过程如图 4 所示.

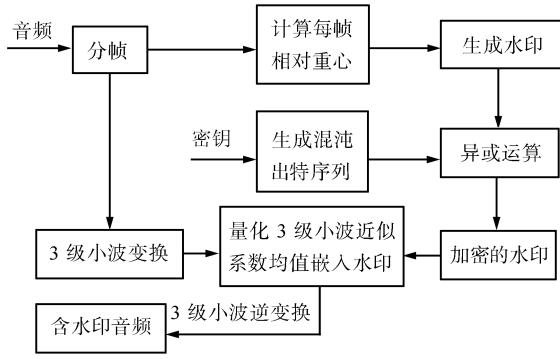


图4 水印嵌入过程

Fig. 4 Watermark embedding process

具体步骤如下:

步骤 1. 将原始音频 $F = \{f(i) | i \in [1, l_a]\}$ 均匀地分为 l_w 帧. l_a 表示音频长度, $f(i)$ 是音频样本幅度, l_w 是水印长度. 帧长 $l_f = \text{floor}(l_a/l_w)$.

步骤 2. 对各音频帧按式 (3) 和式 (4) 计算其相对重心 $x_c^r(m)$, 并将其按式 (8) 作映射:

$$\hat{x}_c^r(m) = \frac{x_c^r(m)}{l_m} \quad (8)$$

其中, l_m 表示映射长度, 因为 $x_c^r(m) \in (0, 1)$, 为方便后面的量化, l_m 取值小于 1. 根据实验的经验值, 一般在 $(0.02, 0.1]$ 间比较合适.

步骤 3. 由相对重心 $\hat{x}_c^r(m)$, 根据式 (9) 生成水印.

$$w(m) = \text{mod}(\text{round}(\hat{x}_c^r(m)), 2) \quad (9)$$

其中, $\text{round}(\cdot)$ 表示取整函数, 该运算可使相对重心有 $(-l_m/2, l_m/2)$ 的容错能力.

为提高算法的安全性, 采用混沌比特序列加密生成的水印比特序列 $w(m)$, 具体如下:

混沌序列 x_m 由 Logistic 映射生成

$$x_{n+1} = ux_n(1 - x_n) \quad (10)$$

其中, $x_n \in (0, 1)$, $3.699456 \leq u \leq 4$, 初值 x_0 和 u 作为密钥.

按升序排列 x_m , 排序后的序列为 y_m , 并取其下标

$$z(m) = \text{index}(y_m) \quad (11)$$

将 $z(m)$ 表示为二进制比特形式 $b_u^m \dots b_3^m b_2^m b_1^m$, 取其次高位 b_{u-1}^m 作为混沌比特序列. 用混沌比特序列与提取的水印序列 $w(m)$ 作异或运算加密水印序列.

$$w_c(m) = b_{u-1}^m \oplus w(m) \quad (12)$$

其中, \oplus 表示异或运算.

步骤 4. 嵌入加密后的水印序列 $w_c(m)$. 嵌入水印利用文献 [25] 的研究成果, 按式 (13) 调制各帧 3 级小波近似系数, 嵌入水印.

$$ca3^T(m, k) = ca3(m, k) - M_{ca3}(m) + M_{ca3}^q(m) \quad (13)$$

其中, $ca3(m, k)$ 表示第 m 帧第 k 个 3 级小波近似分量系数, $M_{ca3}(m)$ 表示第 m 帧的 3 级小波近似分量的均值, $M_{ca3}^q(m)$ 是 $M_{ca3}(m)$ 的量化值.

步骤 5. 作 3 级小波逆变换并依次连接含水印的帧和不含水印的音频样本得到含水印的音频.

4.2 水印的检测算法

检测水印基本上是嵌入水印的逆过程, 其过程如图 5 所示.

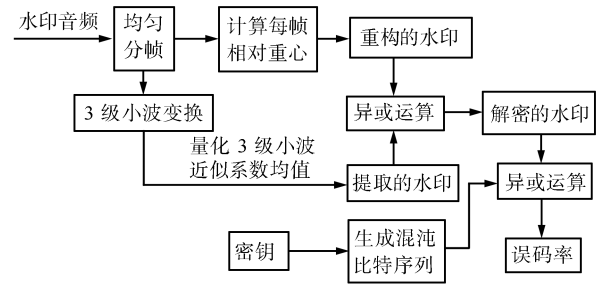


图5 水印检测过程

Fig. 5 Watermark detection process

具体步骤如下:

步骤 1. 将音频 $F^T = \{f^T(i) | i \in [1, l_a^T]\}$ 均匀地分为 l_w 帧. l_w 表示水印长度, l_a^T 表示受攻击后音频的长度, $f^T(i)$ 表示音频样本的幅度. 帧长 $l_f^T = \text{floor}(l_a^T/l_w)$.

步骤 2. 对步骤 1 中的各帧作 3 级小波变换, 并计算每帧的 3 级小波近似系数均值 $M_{ca3}^T(m)$, 用与嵌入过程相同的量化步长 l_q , 量化 $M_{ca3}^T(m)$, 提取水印比特序列 $w_e^T(m)$.

步骤 3. 采用和嵌入相同的方法, 生成混沌比特序列 B_m , $m \in [1, l_w]$. 并按下式解密步骤 2 中提取的水印序列 $w_e^T(m)$, 得到解密的水印序列 $w_c^T(m)$.

$$w_c^T(m) = B_m \oplus w_e^T(m) \quad (14)$$

步骤 4. 利用式 (3) 和式 (4) 计算各帧的相对重心 $x_c^r(m)$ 并作映射 $\bar{x}_c^r(m) = \hat{x}_c^r(m)/l_m$. 其中 l_m 与水印嵌入过程取值相同. 对 $\bar{x}_c^r(m)$, 按下式重构水印:

$$w^T(m) = \text{mod}(\text{round}(\bar{x}_c^r(m)), 2) \quad (15)$$

步骤 5. 将步骤 3 中解密的水印 w_c^T 与步骤 4 中重构的水印按下式计算误码率, 与设定的门限 ρ 比

较, 判别水印存在与否.

$$BER = \left[\frac{\sum_m w_c^T(m) \oplus w^T(m)}{l_w} \right] \times 100\% \quad (16)$$

一般, 误码率在 20% 范围内是可以接受的^[19]. 可以根据应用需要的不同, 设定可接受的误码率范围.

5 实验结果

为检验信噪比 SNR 与量化长度的关系, 我们测试了不同量化长度下的信噪比, 其关系曲线如图 6 所示. 图 6 表明 SNR 值随着量化长度 l_a 的增大而减小. IFPI 规定嵌入水印的音频的信噪比大于 20 dB, 从图 6 可以看出, 嵌入水印的音频信噪比均符合要求. 经验表明, 当 SNR 大于 22 dB 时, 人耳难以感觉到音频听觉质量的变异, 图 6 中的 SNR 都在 22 dB 之上. 水印嵌入和提取过程中的量化长度 l_a 选择 0.02, 嵌入水印后的波形分别如图 7(a)~7(c) 所示. 图 7 从波形上与图 1 对应的原始波形看不出差异.

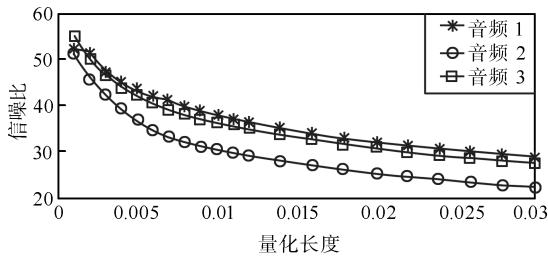


图 6 信噪比随量化长度变化曲线
Fig. 6 Curves of SNR alteration following quantization length

为验证算法的不可听性和对常规信号处理操作的鲁棒性及对抖动、TSM 等同步操作的鲁棒性, 选取图 1 中的三种类型的音频信号进行系列的实验. 三种类型音频均为采样率 44.1 kHz, 16 比特量化的 1M 大小的单声道音乐剪辑.

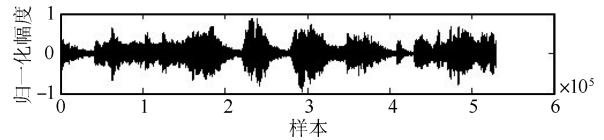
实验中水印长度 l_w 选择 50 比特, 相对重心的映射过程的变换长度 l_m 选择 0.1.

5.1 不可听测试

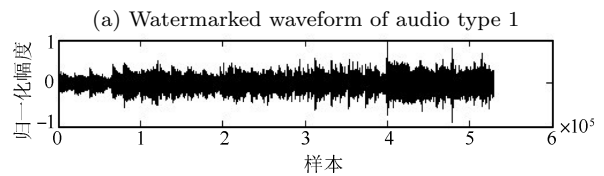
利用 ITU (International Telecommunication Union) 推荐的 BS.1116 定义的主观评价标准即听觉质量主观区分度 SDG (Subjective difference grades) 对音频听觉质量进行评价. 该标准由听众对比原始音频和水印音频给出主观评价分值, 然后取平均. SDG 分值含义如表 1 所示. 在实验过程中, SDG 实验测试人数为 10, SDG 分值测试均为 0. SDG 的实验结果表明了水印的不可听性.

SNR 评价标准没考虑人耳听觉的情况.

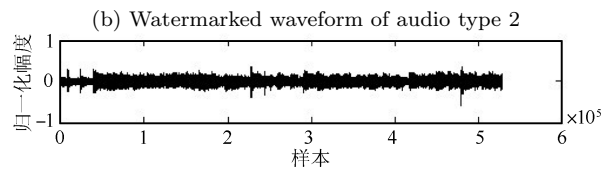
ITU 推荐的 BS.1387^[32] 音频质量听觉评价标准中的 PEAQ (Perceptual evaluation of audio quality) 测试工具考虑到人类听觉系统的一些特性, 其用听觉质量客观区分度 ODG (Objective difference grades) 参数指标作为评价标准对水印音频进行客观的评价, ODG 分值含义参见表 2. 3 种音频的 ODG 测试值分别为 -0.025, -0.159, -0.003. ODG 测试结果进一步说明了水印的不可听性.



(a) 音频类型 1 含水印波形



(b) 音频类型 2 含水印波形



(c) 音频类型 3 含水印波形

(c) Watermarked waveform of audio type 3

图 7 嵌入水印后的音频波形

Fig. 7 Watermarked audio waveforms

表 1 听觉质量主观区分度 SDG 测评标准

Table 1 Aural quality SDG evaluation criterion

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

表 2 听觉质量客观区分度 ODG 评价标准

Table 2 Aural quality ODG evaluation criterion

ODG	音频质量	水印不可听性
0	优	完全不可听
-1	良	可听, 但不明显
-2	中	可听, 比较明显
-3	差	可听, 明显
-4	不能分辨	可听, 非常明显

5.2 鲁棒性测试

为全面评价提出的水印算法的鲁棒性, 我们根据操作对音频信号同步结构的影响, 模拟了 2 种类

型的攻击: 1) 不破坏同步结构的常规信号处理操作, 包括: MP3 压缩、低通滤波、高斯白噪声、重采样、重量化等; 2) 同步攻击, 包括: 抖动 (Dithering)、音调不变 TSM、重采样 TSM、PSM 操作. 测试的参数选择了式 (16) 定义的误码率 BER, 误码率越低表明检测的水印与原始水印越接近, 算法的鲁棒性越好.

5.2.1 对常规信号处理操作的鲁棒性测试

1) MP3 压缩: 音频文件常进行 MP3 有损压缩, 以去除一些冗余信息后进行传输, 降低传输带宽和提高传输效率. 我们首先将比特率为 705 kbps 的水印音频分别压缩至 128 kbps、96 kbps、64 kbps, 然后再解压还原为 wav 格式文件, 压缩后音频质量发生不同程度的下降, 水印检测情况如表 3 所示, 表中“—”表示实验缺省. 结果表明提出的算法对 MP3 压缩的压缩率不敏感, 有很好的抗 MP3 压缩的能力.

表 3 常规信号处理后水印的误码率 BER (%)
Table 3 BER after common signal processing (%)

操作类型	本文 音频类型			文献 [19] 音频类型		
	1	2	3	1	2	3
MP3/128 kbps	0	0	2	1.67	15	6.7
MP3/96 kbps	2	0	2	—	—	—
MP3/64 kbps	2	0	0	—	—	—
35 dB 白噪声	0	0	0	0	0	0
40 dB 白噪声	0	0	0	—	—	—
9 kHz 低通滤波	0	0	0	0	3.33	8.33
8 kHz 低通滤波	0	0	0	0	10	16.7
4 kHz 低通滤波	0	0	0	—	—	—
重采样	0	0	0	0	0	0
重量化	0	0	0	0	0	0

2) 低通滤波: 对不少水印算法, 低通滤波能有效地去除水印. 为了检测算法对低通滤波的鲁棒性, 我们模拟了最低截止频率为 4 kHz 的低通滤波, 在该频率滤波作用下, 音频的听觉质量已经明显下降. 表 3 列出了不同类型的水印音频滤波后提取水印的误码率情况. 结果表明本文算法对低通滤波具有很好的鲁棒性. 经过 4 kHz 截止频率的低通滤波, 音频信号已经明显失真的情况下, 水印仍能准确地提取出来.

3) 添加高斯白噪声: 噪声污染是常见的信号处理操作, 仿真实验分别模拟 35 dB、40 dB 的高斯白噪声污染, 其水印的误码率值如表 3 所示. 结果表明提出的算法能较好的抵抗噪声污染.

4) 重采样: 先分别将音频信号下采样到 22.05 kHz, 再上采样到 44.1 kHz, 误码率见表 3. 结果表明算法对重采样具有很强的鲁棒性.

5) 重量化: 将音频先量化至 8 比特, 然后再量

化至 16 比特, 其误码率如表 3 所示. 结果表明提出的算法对重量化有好的鲁棒性.

表 3 中的误码率值表明本文水印算法相比文献 [19] 的方法对 MP3 压缩、低通滤波等操作有更好的鲁棒性.

5.2.2 对抖动和 TSM 处理操作的鲁棒性测试

本文算法对抖动攻击的仿真结果如表 4 所示. 表 4 的结果表明提出的算法对抖动具有很好的鲁棒性.

表 4 抖动攻击后的误码率 BER (%)
Table 4 BER after dithering attack (%)

操作类型	本文 音频类型			文献 [19] 音频类型		
	1	2	3	1	2	3
抖动 1/1000	2	0	0	0	0	0
抖动 1/100	2	0	2	0	0	0
抖动 1/10	2	0	2	0	0	0

TSM 操作包括音调不变 TSM 操作和重采样 TSM. 水印对音调不变 TSM 操作后的误码率仿真结果见表 5, 水印对重采样操作后的误码率仿真结果见表 6, 其中用“×”表示误码率大于 20% 的情况. 文献 [19] 将误码率大于 20% 的情况作为不可接受范围.

表 5 音频受音调不变 TSM 操作后的误码率 (%)
Table 5 BER after pitch invariant TSM (%)

操作类型	本文 音频类型			文献 [19] 音频类型		
	1	2	3	1	2	3
5%	4	2	0	0	0	0
10%	6	2	4	0	0	0
15%	6	2	4	0	0	0
20%	12	2	8	0	0	0
-5%	10	0	4	0	0	0
-10%	6	0	2	0	0	0
-15%	2	0	2	0	0	0
-20%	14	0	2	0	0	0

表 5 的实验数据表明, 本文算法能抵抗 ±20% 的音调不变 TSM 操作, 远高于 IFPI 规定大于 10% 的要求, 但文献 [19] 的方法抵抗音调不变 TSM 性能更好.

表 6 的实验结果表明, 本文算法能抵抗 ±20% 的重采样 TSM 操作, 并且比文献 [19] 的方法有好得多的鲁棒性. 从表 6 的实验结果还可以看出, 文献 [19] 的方法对音频的类型有选择性, 其对音频类型 2 的效果不是很好, 而我们的方法具有通用性, 其抵抗重采样的能力与音频类型无关.

表 6 音频受重采样 TSM 操作后的误码率 (%)

Table 6 BER after re-sampling TSM (%)

操作类型	本文 音频类型			文献 [19] 音频类型		
	1	2	3	1	2	3
5%	0	0	0	0	10	0
10%	0	0	0	0	16.7	0
15%	0	0	0	0	×	0
20%	2	0	0	0	×	0
-5%	0	0	0	0	11.7	0
-10%	2	0	2	0	18.7	0
-15%	2	0	2	0	×	0
-20%	4	0	2	0	×	0

5.2.3 对 PSM 处理操作的鲁棒性测试

PSM 操作可以看作音调不变 TSM 和重采样 TSM 的组合, 表 7 给出了水印受 PSM 操作后的误码率情况, 表中的实验结果表明我们提出的方法能抵抗 $\pm 20\%$ 的 PSM 操作。

表 7 音频受 PSM 操作后的误码率 (%)

Table 7 BER after PSM operation (%)

PSM 比率	音频类型 1	音频类型 2	音频类型 2
5%	0	0	0
10%	0	0	0
15%	0	0	0
20%	2	0	0
-5%	0	0	0
-10%	2	0	2
-15%	2	0	2
-20%	4	0	2

6 结论

抵抗同步攻击是音频水印的一个公开难题, 本文首先证明了基于音频内容特征的相对重心的 TSM 攻击稳定性, 然后分析了均匀分帧降低同步位置偏移优点, 接着设计了基于相对重心的均匀分帧的音频水印算法。实验的结果表明, 我们提出的算法在抵抗常规信号处理操作的同时, 能很好地抵抗抖动、音调不变 TSM 和重采样 TSM、PSM 等同步攻击, 算法不需嵌入同步码, 可以抵抗多种去同步操作, 尤其是能抵抗大于 10% 的 TSM 攻击, 具有很好的鲁棒性。算法的鲁棒性源自两个方面的原因: 1) 均匀分帧避免了同步攻击引起的同步位置连续偏移累积问题; 2) 小波域近似分量均值和定义的相对重心是两个对同步攻击稳定的量。反过来, 算法的鲁棒性也是相对重心对同步攻击稳定性的实验证明, 我们进一步的研究工作将围绕如何更好地有效重同步和寻找受攻击稳定的特征量展开, 以便设计更鲁棒的抗同步攻击的音频水印算法。

References

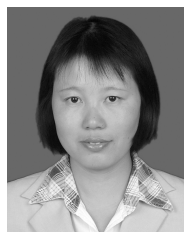
- Maha C, Maher E, Chokri B A. A blind audio watermarking scheme based on neural network and psychoacoustic model with error correcting code in wavelet domain. In: Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing. St Julians, France: IEEE, 2008. 1138–1143
- Swanson M D, Zhu B, Tewfik A H, Boney L. Robust audio watermarking using perceptual masking. *IEEE Transactions on Signal Processing*, 1998, **66**(3): 337–355
- Sun Rui, Jiang Jian-Guo, Yang Xing-Ming. Audio watermarking algorithm based on wavelet packet and psychoacoustic model. *Journal of System Simulation*, 2007, **19**(10): 2248–2255
(孙锐, 蒋建国, 杨兴明. 基于小波包和心理声学模型的音频水印算法. 系统仿真学报, 2007, **19**(10): 2248–2255)
- Quan Xiao-Mei, Zhang Hong-Bin. Statistical audio watermarking algorithm based on auditory analysis in wavelet packet domain. *Acta Electronica Sinica*, 2007, **35**(4): 673–678
(全笑梅, 张鸿宾. 基于小波包听觉感知分析的统计音频水印算法. 电子学报, 2007, **35**(4): 673–678)
- Darko K, Henrique S M. Spread-spectrum watermarking of audio signals. *IEEE Transactions on Signal Processing: Special Issue on Data Hiding*, 2007: 1–13
- Kirovski D, Malvar H S. Spread-spectrum watermarking of audio signals. *IEEE Transactions on Signal Processing*, 2003, **51**(4): 1020–1033
- Kutter M, Bhattacharjee S K, Ebrahimi T. Towards second generation watermarking schemes. In: Proceedings of the 1990 International Conference on Image Processing. Kobe, Japan: IEEE, 1999. 320–323
- Wang Hong-Xia. Overview of content-based adaptive audio watermarking. *Journal of Southwest Jiaotong University*, 2009, **44**(3): 430–447
(王宏霞. 基于内容的自适应音频水印研究综述. 西南交通大学学报, 2007, **44**(3): 430–447)
- Mansour M F, Tewfik A H. Data embedding in audio using time-scale modification. *IEEE Transactions on Speech and Audio Processing*, 2005, **13**(3): 432–439
- Fan M Q, Wang H X. Centroid-based robust audio watermarking scheme. In: Proceedings of the 2008 IEEE International Conference on Audio, Language and Image. Shanghai, China: IEEE, 2008. 476–479
- Bao De-Wang, Yang Hong-Ying, Qi Wei, Wang Xiang-Yang. A content based audio watermarking against desynchronization attacks. *Journal of Image and Graphics*, 2009, **14**(12): 2619–2622
(鲍德旺, 杨红颖, 祁薇, 王向阳. 基于音频特征的抗去同步攻击数字水印算法. 中国图象图形学报, 2009, **14**(12): 2619–2622)
- Li Yan-Ping, Tang Zhen-Min, Qian Bo. Audio watermark algorithm robust to desynchronization based on speech parameter model. *Computer Engineering*, 2008, **34**(9): 177–179
(李燕萍, 唐振民, 钱博. 基于语音参数模型的抗同步攻击音频水印算法. 计算机工程, 2008, **34**(9): 177–179)
- Li W, Xue X Y, Lu P Z. Localized audio watermarking technique robust against time-scale modification. *IEEE Transactions on Multimedia*, 2006, **8**(1): 60–69
- Mansour M F, Tewfik A H. Audio watermarking by time-scale modification. In: Proceedings of the 2001 International Conference on Acoustics, Speech, and Signal. Salt Lake City, USA: IEEE, 2001. 1353–1356
- Fan Ming-Quan, Wang Hong-Xia. Content-based fragile audio watermarking in hybrid domain. *Journal of the China Railway Society*, 2010, **32**(1): 118–122
(范明泉, 王宏霞. 基于音频内容的混合域脆弱水印算法. 铁道学报, 2010, **32**(1): 118–122)

- 16 Wang Hong-Xia, Fan Ming-Quan. Centroid-based semi-fragile audio watermarking in hybrid domain. *Science China Information Sciences*, 2010, **40**(1): 313–326
(王宏霞, 范明泉. 基于质心的混合域半脆弱音频水印算法. *中国科学信息科学*, 2010, **40**(1): 313–326)
- 17 Zhao Hong, Shen Dong-Sheng, Zhu Yuan-Hui. A semi-fragile audio watermarking against shearing. *Acta Automatica Sinica*, 2008, **34**(6): 647–651
(赵红, 沈东升, 朱元辉. 一种抗裁剪的半脆弱音频水印算法. *自动化学报*, 2008, **34**(6): 647–651)
- 18 Wang X Y, Wu J. A feature-based robust digital image watermarking against desynchronization attacks. *International Journal of Automation and Computing*, 2007, **4**(4): 428–432
- 19 Xiang S J, Huang J W. Histogram-based audio watermarking against time-scale modification and cropping attacks. *IEEE Transactions on Multimedia*, 2007, **9**(7): 1357–1372
- 20 Wang X Y, Zhao H. A novel synchronization invariant audio watermarking scheme based on DWT and DCT. *IEEE Transactions on Signal Processing*, 2006, **54**(12): 4835–4840
- 21 Wang Hong, Sun Yi. Anti cropping synchronization audio digital watermark algorithm based on watermark sequence number. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, 2008, **32**(3): 481–484
(王虹, 孙毅. 基于水印序号的抗裁剪音频数字水印算法. *武汉理工大学学报 (交通科学与工程版)*, 2008, **32**(3): 481–484)
- 22 Wu S Q, Huang J W, Huang D R, Shi Y Q. Efficiently self-synchronized audio watermarking for assured audio data transmission. *IEEE Transactions on Broadcasting*, 2005, **51**(1): 69–76
- 23 Tachibana R, Shimizu S, Nakamura T, Kobayashi S. An audio watermarking method robust against time and frequency fluctuation. In: Proceedings of the 2001 Conference on Security and Watermarking of Multimedia Contents. California, USA: SPIE, 2001. 104–115
- 24 Li Wei. Research on roust digital audio watermarking [Ph.D. dissertation], Fudan University, China, 2005
(李伟. 鲁棒性数字音频水印算法研究 [博士学位论文], 复旦大学, 中国, 2005)
- 25 Nian Gui-Jun, Wang Shu-Xun, Tang Xiao-Nian. Robust adaptive audio watermarking algorithm in cepstrum. *Journal of Jilin University (Information Science Edition)*, 2008, **26**(1): 55–61
(年桂君, 王树勋, 唐笑年. 倒谱域自适应音频水印算法. *吉林大学学报 (信息科学版)*, 2008, **26**(1): 55–61)
- 26 Feng Tao, Han Ji-Qing. Synchronization and blind detect algorithm for dual channel audio watermark. *Journal of Communications*, 2006, **27**(10): 62–68
(冯涛, 韩纪庆. 双声道音频水印的同步及盲检测算法. *通信学报*, 2006, **27**(10): 62–68)
- 27 Wei F S, Xue F, Li M Y. A blind audio watermarking scheme using peak point extraction. In: Proceedings of the 2005 IEEE International Symposium on Circuit and Systems. Kobe, Japan: IEEE, 2005. 4409–4412
- 28 Chen S T, Wu G D, Huang H N. Wavelet-domain audio watermarking scheme using optimisation-based quantisation. *IET Signal Processing*, 2010, **4**(6): 720–727
- 29 Chen S T, Huang H N, Chen C J, Wu G D. Energy proportion based scheme for audio watermarking. *IET Signal Processing*, 2010, **4**(5): 576–587
- 30 Lin Y Q, Abdulla W H, Ma Y D. Audio watermarking detection resistant to time and pitch scale modification. In: Proceedings of the 2007 IEEE International Conference on Signal Processing and Communications. Dubai, The United Arab Emirates: IEEE, 2007. 1379–1382
- 31 Lin Y Q, Abdulla W H. A secure and robust audio watermarking scheme using multiple scrambling and adaptive synchronization. In: Proceedings of the 2007 IEEE Conference on Information, Communications and Signal Processing. Singapore: IEEE, 2007. 1–5
- 32 International Telecommunication Union. Method for Objective Measurements of Perceived Audio Quality (PEAQ), Technical Report ITU-R BS.1387-1, Geneva, 1998–2001



黄雄华 西南交通大学博士研究生, 主要研究方向为信息隐藏与数字水印。本文通信作者。E-mail: 494777@163.com

(HUANG Xiong-Hua Ph.D. candidate at the School of Information Science and Technology, Southwest Jiaotong University. His research interest covers information hiding and digital watermarking. Corresponding author of this paper.)



王宏霞 西南交通大学信息科学与技术学院教授, 主要研究方向为数字水印, 信号处理。

E-mail: hxwang@home.swjtu.edu

(WANG Hong-Xia Professor at the School of Information Science and Technology, Southwest Jiaotong University. Her research interest covers digital watermarking and signal process.)



崔更申 桂林电子科技大学副教授, 主要研究方向为数字水印, 嵌入式系统。

E-mail: cgs@guet.edu.cn

(CUI Geng-Shen Associate professor in Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology. His research interest covers digital watermarking and embedding system.)



蒋伟贞 西南交通大学博士研究生, 暨南大学讲师, 主要研究方向为数字水印。

E-mail: 65274397@qq.com

(JIANG Wei-Zhen Ph.D. candidate at the School of Information Science and Technology, lecturer at the School of Information Science and Technology, Jinan University. Her main research interest is digital watermarking.)