

一种应用心理声学模型的鲁棒音频水印算法

王 卓, 赵千川

(清华大学自动化系, 北京 100084)

摘要: 数字水印技术作为一种版权保护的重要手段得到了广泛的研究和应用。文中提出了应用心理声学模型对音频进行逐帧分析、自适应地选择水印嵌入位置的算法框架, 并在此基础上应用了一种音频频域子带能量量化的水印算法。这一框架对有效地控制听觉透明性和 MP3 鲁棒性有指导意义。实验证明在保证较高的听觉保真度的同时, 算法对各种攻击手段具有满意的鲁棒性。

关键词: 版权保护; 数字水印; 心理声学模型; 能量量化; MP3 压缩

中图分类号: TP391

文献标识码: A

文章编号: 1000-3630(2006)-03-0248-05

A robust audio watermarking algorithm using psychoacoustic model

WANG Zhuo, ZHAO Qian-chuan

(Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: Digital watermarking techniques have been widely studied and applied as a main method for copyright protection. This paper proposes a framework that analyses each audio frame and selects positions for embedding adaptively using a psychoacoustic model. A new method for embedding digital watermarks into audio signals in the frequency domain based on the critical sub-band energy quantization is introduced. This framework can control the trade-off between the perceptual transparency and robustness against MP3 attacks more efficiently. Experimental results show that the algorithm is robust to the most popular attacks such as MP3 compression, etc.

Key words: copy protection; digital watermarking; psychoacoustic model; energy quantization; MP3 compression

1 引 言

近年来, 多媒体素材的丰富化、数字化, 大容量存储设备和网络的迅速发展, 以及功能强大的多媒体处理软件的出现, 使非法使用、伪造和传播数字媒体成为一个不能回避的问题。作为数字媒体作品知识产权保护的一种有效手段, 数字水印 (Digital Watermarking) 技术得到了广泛关注, 成为国际学术界的一个研究热点。音频水印技术作为数字水印研究

的一项重要内容, 与静态图像和视频水印的研究相比, 成果相对较少, 其面临的挑战主要有: (1) 由于人的听觉系统 (HAS) 与视觉系统相比更加敏感, 使透明的水印嵌入更加困难。(2) 一般情况下音频中能够嵌入的水印比特通常比图像和视频少。(3) 相对于图像这种二维信号, 音频水印所面临的攻击手段有一定的不同。

音频水印算法主要可分成时域嵌入 (例如回声隐藏算法)^[1,2] 和变换域中嵌入 (例如 DCT 域、小波域和 FFT 变换域嵌入算法)^[3-5] 两大类, 另外还有一些直接在压缩域上进行的水印隐藏算法^[6]。无论是基于时域还是频域的嵌入算法, 选择恰当的嵌入位置都是最为核心的课题。这是有效地控制水印系统的听觉透明性和鲁棒性的关键, 以达到所需的信息隐藏效果。例如, 在网络上传播的音频数据, 面临最多

收稿日期: 2005-03-30; 修回日期: 2005-06-27

基金项目: 教育部新世纪优秀人才支持计划 (NCET-04-0094) 和清华大学“九八五”基础研究基金 (985 信息-07-基金-07)。

作者简介: 王卓 (1980-), 男, 河北辛吉人, 硕士研究生, 研究方向: 数字媒体和无线通讯。

的是 MP3 压缩攻击,这就需对 MP3 压缩过程进行预分析和预判断,选择在这一过程中能够较大程度上保持不变性的频域位置进行水印的隐藏;而对声音保真度要求较高的场合,在一些被掩蔽、相对听觉不敏感的位置上进行水印嵌入更为合适。目前,许多研究成果中^[3,4,7],在嵌入位置的选择上基本依靠经验推断和实验判断,尚无一个指导性框架。

本文提出应用心理声学模型对音频帧进行预分析,从而恰当地选择水印嵌入位置的算法框架。心理声学模型是一种尽可能逼真地模拟人类听觉系统特性的算法,是 MP3 等音频压缩算法在高保真度的要求下实现大的压缩比率的关键。文[3,8]中利用心理声学模型在水印嵌入前整形水印信号,降低了水印嵌入造成的听觉失真。本文提出了一种应用心理声学模型的新思路,将其引入水印嵌入和提取过程,作为选择和寻找合适的频域子带进行水印嵌入和提取的指导,并在此基础上应用一种音频频域子带能量量化的水印算法。文献[9]中提出了能量量化的方法,并证明了基于能量的量化相对于采样点量化的优越性。但是这种方法直接在时域上进行,并且对每段音频中所有的时域采样点做修改,会对信号的整个频带造成影响,降低水印的隐蔽性。本文的能量计算和量化在频域进行,相对于时域修改具有隐蔽性强的特点。并且,这些修改仅发生在心理声学分析模块选择到的某几个频带上,保证了不给原始信号造成听觉意义上的巨大干扰。有目的地选择这些频带,有利于有效地控制水印的听觉透明性和对某些攻击(如 MP3 压缩)的鲁棒程度。

2 水印嵌入

2.1 算法框架简述

本文提出的算法属于频域嵌入算法。图1为水印嵌入算法原理图,虚线代表可选的过程。整个算法按照以下步骤进行:

(1)首先将原始音频信号分成互不重叠的、长度为 1152 的帧,经过 FFT 把每帧的时域信号变换到频域。

(2)把第 k 帧中的频率采样点划分成 25 个关键频带 $F_{k,j}(i)$, k 为帧序号, j 为子带序号 $0 \leq j \leq 25$, i 为子带内频率采样点序号, $0 \leq i \leq N$, N 为这个子带中包含的频率采样点个数。应用心理声学模型^[13]计算每帧

的子带能量 $S(j)$ 和全局掩蔽域值 $T(j)$, $0 \leq j \leq 25$ 。

(3)计算 $r(j)=T(j)/S(j)$, j 为子带序号。选择满足条件: $r < 1$ 的 r 最大的 3 个子带作为水印信号的嵌入位置,并可以选择把这些位置序列保存起来。

(4)依照经过预处理的水印信号位,对选择的子带的能量应用 Energy QIM^[9]算法进行量化,修改这些子带上频率采样点的值,实现水印嵌入的目的。

(5)子带综合,经过 IFFT 得到时域帧。连接各帧,组成嵌入水印后的音频文件。

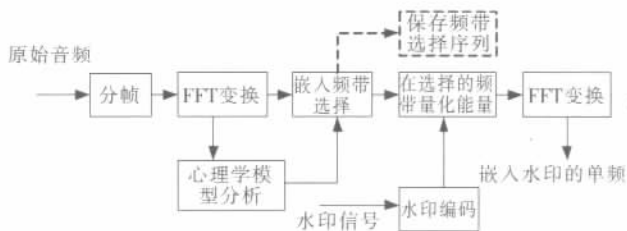


图1 水印嵌入算法框架

Fig.1 Framework for watermark embedding.

以上步骤中,水印的预处理、心理声学模型分析和频带选择、子带能量量化组成了整个水印算法的核心。

2.2 水印预处理

水印预处理分成乱序和编码两步进行。如果采用二维图像作为水印信号,可以首先对二维图像做降维处理。设得到的水印数据: $V=(v_1, v_2, \dots, v_m)$, m 为水印长度。为了避免未授权者能够还原水印图像,将水印序列 V 根据一个随机序列 R 重新排序得到 $W=(w_1, w_2, \dots, w_m)$ 。 R 成为水印检测时重排还原水印数据的密钥。

在通信系统中,对信息序列纠错编码可以使传输更加可靠和有效。水印系统可以看成是一个信息传输系统^[7],因此,可以应用纠错编码技术来提高检测正确率和水印鲁棒性。本文使用最简单的重复编码,将长度为 m 的水印重复嵌入 r 次,在满足 $m \cdot r \leq K$ 的条件下, K 为载体水印容量,水印比特检测错误率 $P_{rep} = \sum C_r^n P^n (1-P)^{r-n}$, 其中 P 为每位的错误率, C 表示组合。在 P 较低的情况下,重复编码能较地降低水印比特的检测错误率。

2.3 心理声学模型分析和频带选择

心理声学模型是一种尽可能逼真地模拟人类听觉系统特性的算法。该模型计算以频率为自变量的噪声掩蔽阈值新罗马,确定每个子带里的信号能量

与掩蔽阈值的比率。心理声学模型是新罗马等音频压缩算法在高保真度的要求下实现大的压缩比率的关键。心理声学模型首先计算音频帧能量谱: $P_{\text{ob}}(i) = |F_{k,j}(i)|^2$, i 为频率采样点序号。进而得到每个关键频带的能量: $S(j) = \sum_{k=LB}^{UB} P_{\text{ob}}(k)/N$, $j=1, 2, \dots, 25$ 。LB、UB 分别表示第 j 个子带所包含的频率上界和下界。每个关键频带能量被用来计算本帧音频内各个子带的全局掩蔽阈值 $T(j)$, $j=1, 2, \dots, 25$ (参考心理声学模型 [10])。

按照式 (1) 计算全局掩蔽阈值 $T(j)$ 与关键频带能量 $S(j)$ 的比值 $r(j)$:

$$r(j) = \frac{T(j)}{S(j)}, 0 < j \leq 25 \quad (1)$$

定义集合 $B = \{j | r(j) < 1, 0 < j \leq 25\}$, 在 B 中选择 r 值最大的三个频带作为嵌入水印的位置。

根据不同的实际需要, 我们可以有不同的选择方案。本文选择了满足条件 $r(j) < 1$ 的频带中 r 值最大的 3 个子带作为嵌入位置, 是一种兼顾听觉透明性和 MP3 攻击鲁棒性的方案。首先, 分析心理声学模型的实现过程, 可以发现 r 值较大的频带一般出现在被附近的强势子带掩蔽的边缘。这是因为: 每个频带的掩蔽阈值是由附近频带的掩蔽函数值叠加得到的, 如果其附近存在某个能量值很大的频带, 我们称为强势子带, 这个强势子带的掩蔽效果将抬升此频带的掩蔽阈值 $T(j)$, 而这个频带的能量值 $S(j)$ 是固定的, 那么其比值 r 就会较大。根据人耳听觉掩蔽特性, 由于附近强势子带的存在, 在这些频带上进行的微小修改是相对不敏感的。同时, 频带选择的另一个要求 $r < 1$, 保证了此频带的能量大于掩蔽阈值。根据 MP3 编码原理, 此频带中的绝大部分频率采样点在 MP3 编码过程中是不能被丢弃的。因此, 在这样的频带选择方案保证了水印不会被 MP3 编解码过程完全剔除。图 2 显示了某帧的子带能量、掩蔽域值以及选择的嵌入点。

2.4 子带能量的量化

本文水印嵌入方式采用了 QIM (Quantization Index Modulation) 算法 [11]。值得注意的是, 这种量化调制是作用在能量上的。文献 [9] 证明, 基于能量的量化调制 Energy QIM 相对于采样点 QIM [4] 更加鲁棒。设某帧第 j 个子带的能量为: $S(j)$, 并且这个子带已经被选择进行水印嵌入, 量化步长为 Δ , 见图

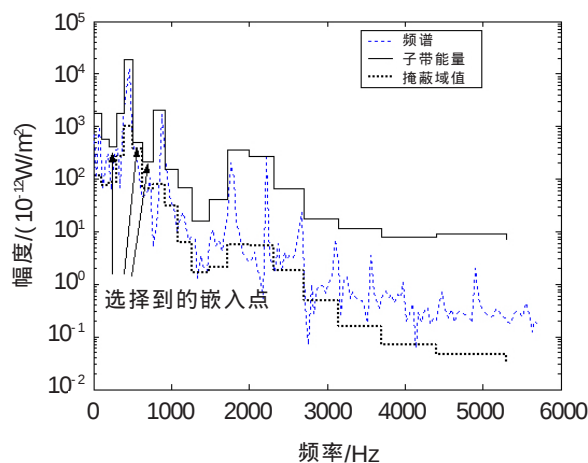


图 2 子带能量、掩蔽域值以及选择的嵌入点

Fig.2 Subband energy, masking threshold and selected embedding point.

3. 当水印比特 $w_i=1$ 时, $S(j)$ 被量化到最近的标记 $\times$ 上; 当 $w_i=0$ 时, $S(j)$ 被量化到最近的标记 $\circ$ 上。称以标记 $\circ$ 和标记 $\times$ 为量化聚集点集的量化器分别: Q_0 和 Q_1 。例如, 设 $w_i=0$, $S(j)$ 经过量化后 $S(j)=Q_0(S(j))$, 如图 3。然后按比例修正包含在这个子带中的频率点的幅值, 完成 1 比特水印的嵌入。

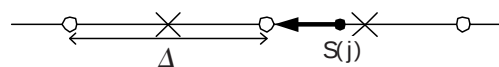


图 3 能量的量化

Fig.3 Energy quantization

对音频频率采样点的修改一定程度上造成了原始音频的失真, 当这种失真足够小时, 人耳并不能察觉, $S(j)$ 的量化步长 Δ 控制了嵌入强度与听觉透明性的折衷, 当 Δ 大时, 水印更容易检测, 同时对原始音频质量的影响也越大。图 4 显示了各个频带能量量化前后的变化情况。

3 水印检测

设待检信号为 $s_n=s+n$, s , n 分别为嵌入水印的音频和传输、音频处理过程中引入的噪声。同嵌入过程类似, 检测过程经过音频分段、频域变换, 通过心理声学模型分析和与嵌入时相同的频带选择方案, 寻找到承载了水印比特的子带 (或者直接应用保存的频带选择序列, 可以达到更高的监测率和更快的监测过程, 但需要嵌入时保存频带选择序列信息), 得到这些频率子带的能量 $S'(j)$ 。根据 $S'(j)$ 最近邻的

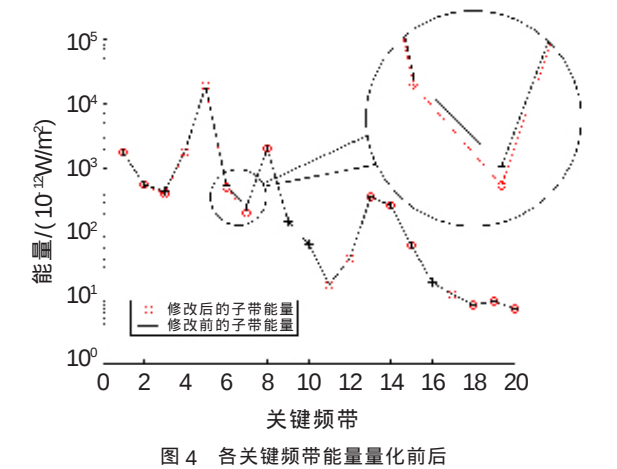


Fig. 4 Before and after energy quantization of key frequency bands.

标记 $\circ$ 或标记 $\times$ 所属的量化器聚集点集判断此段音频嵌入的水印比特。当得到整个的水印 W 后, 计算其与嵌入的水印各位相同的比率 ρ , 如果大于预先设定的某一阈值 T , 则认为待检信号中有此水印, 如果检测不到同步头或者 $\rho < T$, 则认为没有水印。 ρ 的计算按照式 (2) 进行:

$$\rho = \frac{\sum_{i=1}^m W(i) \otimes W(i)}{m}$$

$\otimes$ 代表“同或”。整个检测过程不需要原始音频的参与, 如图 5 所示。

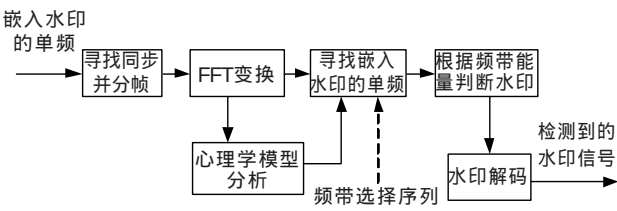


图 5 水印检测算法框架
Fig.5 Framework for watermark detection

4 实验结果

本文选取了一段单声道、44.1kHz 采样频率、16bit、27s 的小提琴曲 *Thais.wav* 进行测试。音频幅值进行了归一化。采用一段长度为 1024bit 的伪随机序列作为水印。嵌入算法采用上文提出的应用心理声学模型的频域子带能量量化算法。图 6 显示了小提琴曲原始音频与嵌入水印后的音频信号。与原始信号相比, 嵌入水印的音频的信噪比 SNR 为 34.45dB。经过多人听力测试, 感受不到二者的差别。

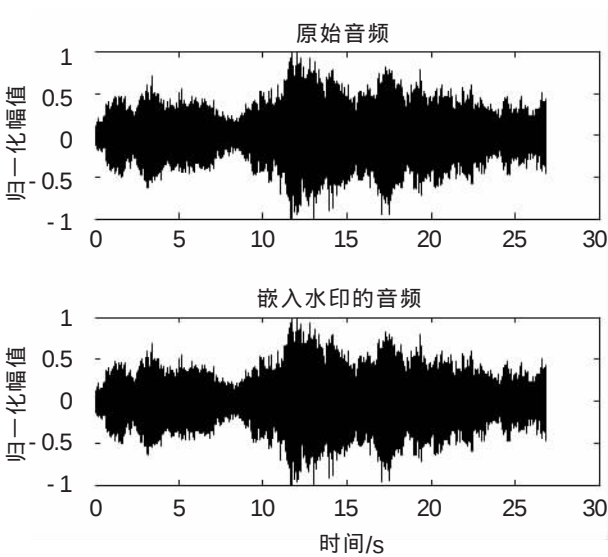


图 6 原始音频信号和嵌入水印的音频信号
Fig.6 Original audio signal and audio signal with watermark

实验检验了水印在以下攻击手段下的检测正确率:

(1) 添加随机白噪声。给幅值归一化的音频上叠加一系列不同强度的噪声 (对严重影响音频质量的加噪攻击不作考虑), 如图 7 显示了水印检测正确率与叠加噪声产生的信噪比的关系:

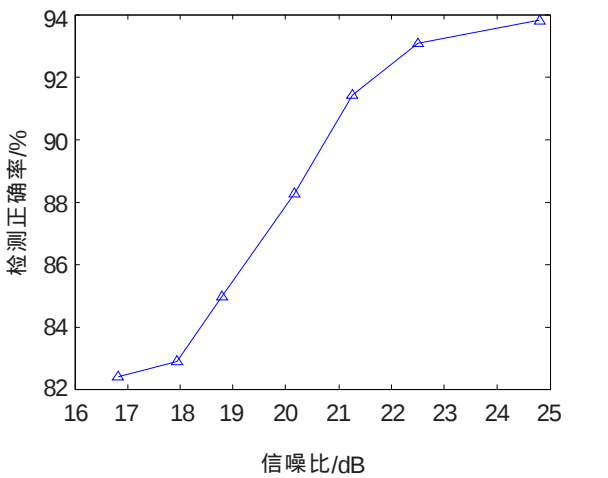


图 7 检测正确率随噪声攻击强度的变化曲线
Fig.7 Detection rate versus SNR

- (2) MP3 压缩攻击。将嵌入水印的音频压缩成比特率为 128kbps 的 MP3 文件再解压, 从中提取水印。
- (3) MP3 压缩攻击。将嵌入水印的音频压缩成比特率为 96kbps 的 MP3 文件再解压, 从中提取水印。
- (4) 低通滤波。将嵌入水印的音频通过一个截止频率为 12kHz 的低通滤波器后提取水印。

(5) 变换采样速率。将嵌入水印的音频先下采样从 44.1kHz 到 11.025kHz, 再内插得到 44.1kHz 的音频, 从中提取水印信号。

(6) 变换采样分辨率。将嵌入水印的音频重采样, 采样点分辨率由 16bits 变为 8bits 的音频中提取水印。

经过以上攻击后, 得到的检测正确率如表 1。

表 1 Thais 经过各种攻击后检测到的水印与原始水印的互相关值
Table 1 The correlation between original watermark and those detected under several attacks for Thais

攻击手段	检测正确率/%
加噪声	93.81
MP3 压缩(128kbps)	90.04
MP3 压缩(96kbps)	85.64
低通滤波	94.53
变换采样速率	1
变换采样分辨率	99.92

从表中可以看出, 经过以上各种攻击后, 提取的水印同嵌入的水印有极大的相关性。该算法在很多音频上的实验都取得了满意的效果, 能够满足音频数字水印鲁棒性的要求。算法没有采用同步机制, 因此不具有抵抗同步攻击的能力。

5 结 论

数字水印是一种很有研究价值的数字媒体版权保护方案。本文提出了应用心理声学模型选择水印嵌入位置的水印嵌入框架。此框架有利于有目的地、有效地控制水印系统的听觉透明性与攻击鲁棒性, 有很大的灵活性和拓展性。在此框架下, 本文提出了一种嵌入位置选择方案, 并在选择的子带上应用了能量量化的算法嵌入水印。仿真实验验证了该算法在满足保真度要求的情况下, 对随机噪声、MP3 压缩、低通滤波、重采样等攻击手段的鲁棒性。算法检测过程不需要原始音频的参与。算法的进一步拓展包括在水印嵌入和提取时增加同步机制来抵抗同步攻击等。这套水印框架和水印方案有着广阔的应用前景。

参 考 文 献

[1] Gruhl D, Lu A, Bender W. Echo Hiding[A]. Lecture

Notes in Computer Science, Information Hiding, Springer, 1996, 11174, 295-315.

[2] Paraskevi Bassia Pitas I. Robust audio watermarking in the time domain[A]. Proc. Of EUSIP-98: Signal Processing IX, Theories and Applications[C]. Greece, 1998, 25-28.

[3] Nakayama A, Lu J, Nakamura S, Shikano K. Digital watermarks for audio signal based on psychoacoustic masking model[J]. Electronics and Communications in Japan(Part III: Fundamental Electronic Science), 2003, 86(12): 65-75.

[4] 王秋生, 孙圣和. 一种在数字音频信号中嵌入水印的新算法[J]. 声学学报, 2001, 26(5): 464-467.
WANG Qiusheng, SUN Shenghe. Novel algorithm for embedding watermarks into digital audio signals[J]. Shengxue Xuebao/Acta Acustica, 2001, 26(5): 464-467.

[5] Arttameeyanant P, Kumhom P, Chamnongthai K. Audio watermarking for Internet[J]. Proc IEEE, 2002, 976-979.

[6] QIAO L, Nahrstedt K. Non-invertible watermarking methods for MPEG encoded audio[J]. Proceedings of SPIE-The Internation Society for Optical Engineering, 1999, 3657: 194-202.

[7] 王向阳, 杨红颖, 赵红. 一种新的自适应量化数字音频水印算法[J]. 声学技术, 2004, 23(2): 117-120.
WANG Xiangyang, YANG Hongying, ZHAO Hong. Digital audio watermarking algorithm based on adaptive quantization[J]. Technical Acoustics, 2004, 23(2): 117-120.

[8] Ricardo A Garcia. Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory[M]. Master of Science Thesis, Music Engineering Technology, University of Miami, Coral Gables, FL, 1999.

[9] 王卓, 赵千川. 基于能量量化的音频水印算法[J]. 计算机工程与应用, 2004, 40(26): 48-51.
WANG Zhuo, ZHAO Qianchuan. Energy quantization based audio watermarking algorithm[J]. Computer Engineering and Applications, 2004, 40(26): 48-51.

[10] Ted Painter, Andreas Spanias. A review of algorithms for perceptual coding of digital audio signals[A]. Proceedings of International Conference on Digital Signal Processing(DSP)[C]. 1997. 179-205.

[11] Chen B, Wornell G W. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding[J]. IEEE Transactions on Information Theory, 47(4): 1423-1443.