

无感知音频信号数字水印技术*

刘兵¹, 杨鉴²

(1. 云南大学 软件学院, 云南 昆明 650091; 2. 云南大学 信息学院, 云南 昆明 650091)

摘要: 从时域统计的音频信号数字水印技术出发, 提出一种无感知音频信号数字水印方法, 该方法以 ISO/IEC11172-3(MPEG-Audio) 中定义的心理声学模型作为数字水印产生的模型, 使嵌入到音频信号中的水印信息最大可能地被掩蔽, 保证声信号的质量. 根据统计模型, 检测时可在不需要原声信号的情况下对数字水印的有无作出判决.

关键词: 数字水印; 人耳听觉系统(HAS); 心理声学模型; 听觉掩蔽

中图分类号: TN 911.7 **文献标识码:** A **文章编号:** 0258-7971(2003)05-0408-04

在版权保护中, 数字水印技术被认为是一种有效的技术. Mitchell D. Swason^[1], Cox^[2] 等人已对数字水印所应具有的特性作了定义. 不论是何种数字水印技术, 首先需要解决的问题是, 从人的感知角度上评价, 数据的嵌入不会使原始信号失真为前提, 进一步要求能够对嵌入数据准确地恢复. 对于音频信号的数字水印研究, D. Gruhl, W. Bender 和 A. Liu 提出了 Echo hiding 的方法^[3], 利用人耳时间分辨率的特性, 在宿主信号中引入“回声”的机制嵌入数据. 该方法针对 16 bit 量化的音频信号在有损压缩、轻度噪声叠加和 D/A 变换后可以检测出水印, 而对于音频信号中存在安静状态的间隔, 例如话音间的停顿的情况出现了问题. Laurence Boney, Ahmed H. Tewfik 和 Khaled N. Hamdy 提出了利用心理声学模型来对音频信号作数字水印处理的方法^[4], 这一方法对原信号质量的保持较好, 但是在检测时需要原始信号从而限制了应用. 本文先从简单的基于统计分析的时域水印方法入手, 结合 Laurence Boney 的方法, 发展了一种基于心理声学模型的无感知音频信号数字水印技术, 它检测时不需要原声信号.

1 基于统计分析的时域水印方法

时域中, 利用人耳对声信号的感知特性产生一

个不易被感知的水印信号, 然后将它叠加到原声信号中生成携带有水印信息, 但从人耳听感上质量没有变化的声信号. 检测时, 根据假设检验理论, 定义一个合适的判决函数对数字水印的有无作出判决.

假设长为 N 的音频信号采样序列 $x(i)$, $i = 0, 1, \dots, N-1$, $w(i)$ 为同长度的水印信号, 它是在 $[-\beta, \beta]$ 范围内均匀分布的伪随机数序列. 根据人耳听觉感知特性, 定义函数^[3] $f(x(i), w(i))$. 为了将水印信号嵌入原声信号中, 对声信号的每个采样点按(1)式作修改, 得携带了水印信息的声信号 $y(i)$,

$$y(i) = x(i) + \alpha f(x(i), w(i)), \quad (1)$$

其中, α 是比例因子, 用于动态调节水印信号嵌入原声信号的强度.

在检测端, 为了判决某一经过传输或修改的音频信号 $y'(i)$ 中水印的有无, 建立二元假设检验模型^[5]: 2 个互斥的假设表示为

$$H_0: y'(i) = x(i) \circ n(i), \quad (2)$$

$$H_1: y'(i) = (x(i) + f(x(i), w(i))) \circ n(i), \quad (3)$$

式中, $y'(i)$ 代表观测数据, $n(i)$ 表示某种失真, 运算符 \circ 代表失真的运算, 它表示了声信号在各种传播渠道中可能受到的失真处理.

为定义判决函数, 首先定义 S 如(4)式

* 收稿日期: 2002-05-31

基金项目: 云南省自然科学基金资助项目(2002D0003Q); 云南省教育厅自然科学基金资助项目(99111119; 0142174).

作者简介: 刘兵(1976-), 男, 云南人, 讲师, 主要从事信息处理方面的研究.

$$S = \sum_{i=0}^{N-1} y(i) w(i). \quad (4)$$

将(1)式代入(4)式中,得

$$S = \sum_{i=0}^{N-1} [x(i) w(i) + f(x(i), w(i)) w(i)], \quad (5)$$

若伪随机数序列 $w(i)$ 和声信号 $x(i)$ 是零均值的,则(5)式的第一项和近似为零.若伪随机数序列 $w(i)$ 存在偏差,假设其中有 Δw 个随机数引起偏差,则这一因素也应被考虑在内,则

$$S = \sum_{i=0}^{N-\Delta w-1} x(i) w(i) + \sum_{i=1}^{\Delta w} x(i) w(i) + \sum_{i=0}^{N-1} f(x(i), w(i)) w(i), \quad (6)$$

式中第一项 $\sum_{i=0}^{N-\Delta w-1} x(i) w(i)$ 近似为零.若信号中无水印,(6)中无第三项,则 S 近似等于

$$S = \frac{\Delta w}{N} \sum_{i=0}^{N-1} x(i) w(i). \quad (7)$$

若声信号中有水印,则 S 近似为

$$S = \frac{\Delta w}{N} \sum_{i=0}^{N-1} x(i) w(i) + \sum_{i=0}^{N-1} f(x(i), w(i)) w(i). \quad (8)$$

假设在检测端无法获得原声信号,即 $x(i)$ 未知,用 $y(i)$ 代替式中的 $x(i)$,则(8)式变为

$$S = \frac{\Delta w}{N} S + \sum_{i=0}^{N-1} f(y(i), w(i)) w(i). \quad (9)$$

由(9)式定义判决函数

$$r = \frac{S - \frac{\Delta w}{N} | S |}{\sum_{i=0}^{N-1} f(y(i), w(i)) w(i)}, \quad (10)$$

则 r 是一个归一化为(0, 1)间的值.若检测端音频信号含有水印,即假设 H_1 成立,则从而 $r \approx 1$.若检测端音频信号不含水印,即假设 H_0 成立,则从而 $r \approx 0$.判决水印有无的阈值通过实验测定为0.5,当 r 的值大于0.5时可判断为音频信号中含有水印信号,反之则不含水印信号.为了提高判断的准确度,该阈值可以适当提高.

2 无感知音频信号水印技术

我们考虑由 CCITT 和 ISO 的运动图像专家组(简称 MPEG)提出的一种运动图像及其伴音的压缩与编码标准 MPEG1 中对音频信号编码的方

式来确定(1)式中的函数 $f(\cdot)$,利用心理声学理论来构建一种音频信号的数字水印处理模型.MPEG1 中,通过心理声学模型来计算声信号各子带刚可察觉的噪声电平,这些噪声电平用于比特或噪声分配,以确定实际的量化器和量化级^[6].通过这种心理声学模型对噪声分配的控制,MPEG 可以达到很高的压缩率同时保持极高的声音质量,因此,在我们的音频信号数字水印处理中,引入心理声学模型,首先对声信号进行分析,计算声信号的频域掩蔽阈值,用该掩蔽阈值来修整噪声,再将修整后的噪声叠加到声信号中,实现水印不可感知的嵌入.

2.1 心理声学模型及水印处理 频域掩蔽模型 已经被广泛地用于音频信号的感知编码.我们采用 MPEG 心理声学模型 I 第一层所定义的掩蔽模型来计算音频信号各帧的最小掩蔽阈值.如下计算掩蔽阈值的方法主要针对 44.1 kHz 的采样信号,同时该方法也支持 32 kHz 和 48 kHz.

最小掩蔽阈值是根据心理声学模型从输入 PCM 信号的 FFT 推导出来的,按以下步骤计算:
① FFT 分析,计算信号短时功率密度谱;
② 找出声信号中的有调和无调成分;
③ 去除被掩蔽的有调和无调成分;
④ 单独掩蔽阈值的计算;
⑤ 总体掩蔽阈值的计算;
⑥ 每个子带中最小掩蔽阈值的决定^[6].

每一个待处理的音频信号都必须由一个唯一确定的水印信号与之对应.用一高质量的伪随机数产生算法来获得需要的水印信号,同时为了利用人耳听觉系统的频率掩蔽特性,可以用上述讨论的最小掩蔽阈值来修整伪随机数序列,产生一个人耳听觉上无法觉察的噪声,然后叠加到原声信号上,实现水印的嵌入.

为了产生最终用于叠加的水印信号,我们首先按照如上所讨论的关于 MPEG 音频信号心理声学模型 I 的内容计算声信号的最小掩蔽阈值.该掩蔽阈值由声信号经海宁窗加权的 512 个采样点,叠加 50% 的区域计算而来.而后,计算得到的掩蔽阈值采用最小均方误差准则,构建一 20 阶的线性相位有限冲击响应滤波器来近似,该滤波器表示为 $M(n)$.伪随机数发生器产生的伪随机数序列 $e(n)$ 经 $M(n)$ 滤波形成用于叠加到声信号上的水印信号 $w(n)$.最后将 $w(n)$ 与声信号 $x(n)$ 叠加形成嵌入了水印的声信号 $y(n)$

$$y(n) = x(n) + w(n). \quad (12)$$

如图 1 是滤波后的伪随机数序列的频谱,可看出,它的频谱包络上类似于最小掩蔽阈值的形状.根据以上的讨论,利用心理声学模型进行水印嵌入的过程可以用框图 2 表示.

在检测端,对接收的音频信号计算判决函数 r

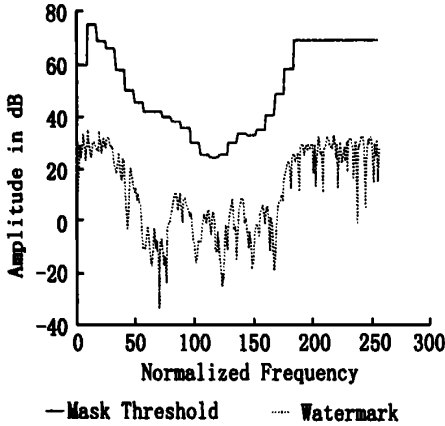


图 1 滤波后的伪随机数序列谱

Fig. 1 Filtered random sequence

的值,对水印的有无作出判决.为了方便,现将 r 的表示式重写在这里

$$r = \frac{S - \frac{\Delta w}{N} |S|}{\sum_{i=0}^{N-1} f(y(i), e(i)) e(i)}, \quad (13)$$

式中的 $f(y(i), e(i))$ 是对接收到的音频信号,利用心理声学模型计算出的水印信号, $e(i)$ 是由伪随机数发生器产生的唯一序列.通过检测值 r 可对水印的有无作出判决.

2.2 实验结果及讨论 按照以上讨论的方法对大量音频信号作了水印处理.这里给出一个具有代表性的音乐信号作为实验处理对象给出结果,该声信号既具有持续音段,中间又具有一段足够长的静音段.如图 3(a) 是原声信号谱图, (b) 是利用心理声学模型进行水印处理后携带了水印信息的谱图.从谱图上看基本上没有差别,从听感上也无法分辨携带了水印信息的声信号质量上的改变,特别在静音段,依然能很好地保持原声信号的纯净听感.对水印后的声信号在无发生失真的情况下作水印检测,判决函数检测值 $r = 1.1520$, 大于阈值 0.5, 表明利用上面讨论的方法是可行而有效的.

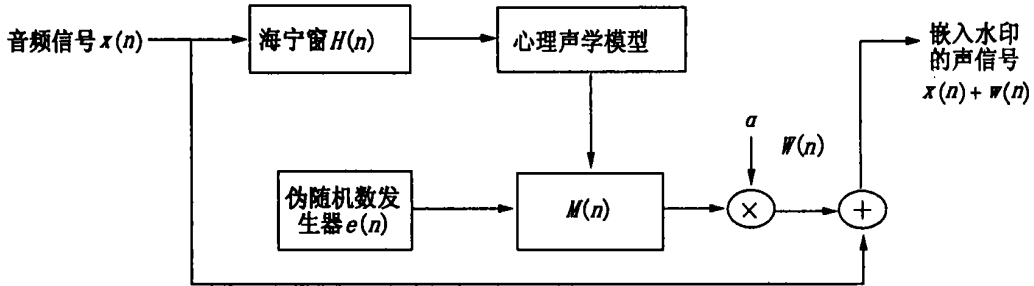


图 2 基于心理声学模型的数字水印嵌入框图

Fig. 2 Framework of watermarking

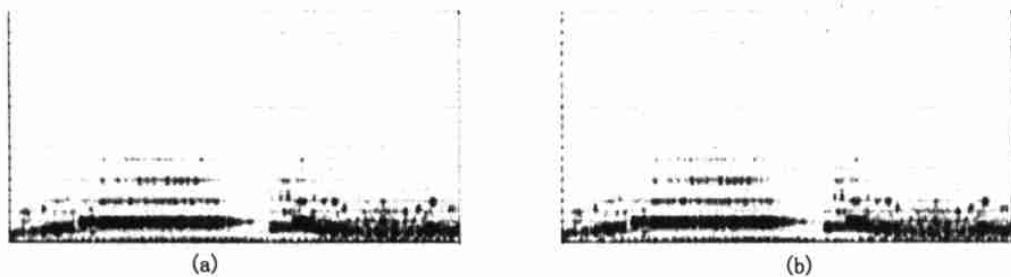


图 3 (a) 原声信号谱图, (b) 水印声信号谱图

Fig. 3 (a) Original spectrum and (b) watermarked spectrum

为验证该方法的鲁棒性, 对水印后的声信号作失真处理后再进行水印检测. 首先对水印后的声信号加入白噪声, 在信噪比 $SNR = 26.4988$ 时, 水印判决函数的检测值 $r = 1.3333$, 大于阈值 0.5 , 表明在噪声环境一定的信噪比情况下, 水印信息可被检测出.

再对水印后的声信号进行重采样处理, 即将采样率从原来的 44.1 kHz 降为原来的一半, 22.05 kHz , 而后再恢复到 44.1 kHz 的采样率, 此时信噪比 $SNR = 29.5207$, 水印判决函数的检测值 $r = 1.0027$, 大于阈值 0.5 , 表明该方法对重采样的失真具有一定的抵抗力.

接下来对水印后的声信号进行重量化处理, 将量化位数变为原来的一半, 即从原来的 16 bit 变为 8 bit 量化, 然后又恢复到 16 bit 进行水印检测, 此时信噪比 $SNR = 20.6582$. 整个过程引入了大量高频噪声, 丢失了许多声信号细节, 从听感上虽然能辨析声信号的内容, 但质量上已无法接受. 水印判决函数的检测值 $r = 0.1282$, 小于阈值 0.5 , 水印检测失败, 表明该方法无法抵抗重量化引入的失真.

3 结 论

本文利用心理声学模型对音频信号分帧进行分析, 计算每一帧的最小掩蔽阈值, 由求出的最小掩蔽阈值来修整一个唯一的伪随机序列, 并和原声信号叠加以实现水印的嵌入. 检测过程基本上是嵌入过程的逆过程, 在检测端, 对接收的音频信号同样利用心理声学模型对音频信号作分帧处理, 计算

出最小掩蔽阈值, 同时修整同一密钥控制下产生的伪随机数序列, 最后计算水印检测判决函数的值, 通过与阈值相比对水印的有无作出判决. 携带了水印信息的声信号和原声信号相比无听感上的差别, 充分保证了处理后无法感知察觉的要求; 检测过程中不需要原声信号, 极大地节省了存储空间; 另一方面, 该方法只对水印的有无作出判决, 因此所能表达的信息量较少. 为了得到声信号的最小掩蔽阈值, 需要对声信号分帧进行频域分析, 处理速度比较慢. 同时由于在检测端不需要原声信号, 其抵抗失真的能力还有待提高.

参考文献:

- [1] SWANSON M D, KOBAYASHI M, TEWFIK A H. Multimedia data embedding and watermarking technologies[J]. Proceedings of the IEEE, 1998, 86(6): 1064—1087.
- [2] COX I J, KILIAN J, LEIGHTON T, et al. A secure, robust watermark for multimedia[A]. Proc Inform[C]. Cambridge, England: Hiding Workshop, 1996, 147—158.
- [3] GRUHL D. Echo Hiding. Proc. of First Int[A]. Workshop on Information Hiding[C]. Cambridge, UK, 1996, 1174: 295—316.
- [4] BONEY L, TEWFIK A, TEWFIK H. Digital watermarks for audio signals[J]. IEEE Proceedings Multimedia, 1996, 473—480.
- [5] 钟玉琢, 蔡莲红, 李树青, 等. 多媒体计算机技术基础及应用[M]. 北京: 高等教育出版社, 1999.
- [6] 余崇智, 李存珠, 张志力. 运动图像及其伴音的压缩与编码——MPEG1[M]. 南京: 南京大学出版社, 1995.

Inaudible watermarking technique for audio signal

LIU Bing¹, YANG Jian²

(1. College of Software, Yunnan University, Kunming 650091, China;

2. College of Information, Yunnan University, Kunming 650091, China)

Abstract: An inaudible watermarking technique for audio signal is developed. The audio signal watermarking method is modeled based on MPEG-Audio psychoacoustic model. It can preserve the quality of audio signal while embedded the watermark signal. Furthermore, the original audio signal is not necessary at the detect edge.

Key words: digital watermarking; human auditory system (HAS); psychoacoustic model; auditory masking