基于 MPEG心理声学模型 II的自适应音频水印算法*

王 泳，李 斌
（中山大学电子与通信工程系，广东 广州 510275）

摘 要：音频水印作为保护音频作品的版权和秘密通信的可行方法，成为近年数字水印研究领域中的一个热点。针对目前音频水印算法研究中自适应性能没有得到重视的问题，提出一种基于心理声学模型 II（以下称模型 II）的自适应盲检测音频水印算法，令算法具备良好的自适应性和稳健性。

关键词：音频水印；自适应；稳健性；MPEG心理声学模型 II
中图分类号：TP309 文献标识码：A 文章编号：0529-6579（2006）02-0029-04

音频水印是数字水印研究中的一个研究热点。其算法主要分为时域的算法和变换域的算法（包括压缩域中算法）[1~7]。已报道算法的主要缺点在于：①人类听觉系统 HAS没有应用于水印算法中，从而无法达到自适应；②在多数自适应算法中仅嵌入 1比特的信息量（随机序列）；③在多数自适应算法中不能达到盲检测。针对这些问题，本文提出了一种基于模型 II的自适应音频水印算法。其特点是利用模型 II对音频信号进行分析，通过其分析结果控制水印的嵌入强度，从而达到水印的自适应嵌入。实验结果表明水印具有良好的抵抗加性高斯噪声、抵抗 MPEG压缩和抗裁剪的性能。

1 算法原理

音频信号是时间轴上的函数，剪裁等攻击会引起严重的同步错误。为了在检测时取得水印的快速同步，本文采用 [8] 的同步技术，在时域上嵌入同步信号。由于变换域上的水印能量能较均匀地扩散到时域上，对水印的不可感知性和稳健性比较有利，水印嵌入于原始音频信号的 DCT系数中。本文算法利用模型 II对音频信号进行分析，获取音频信号在频域（DCT域）上在满足不可感知要求下的最大改变量 P 值。利用该改变量及 DCT系数的位置控制水印的嵌入强度，达到自适应的目的。音频在不同频域段上的屏蔽性质不同，故亦须对同一段音频信号的频域进行分段。每段频域的嵌入强度由各自的 P 值控制。为了降低检测水印的差错率，从而提高水印的稳健性，本文算法对水印先进行 BCH纠错编码。图 1是所提出算法的原理框图。

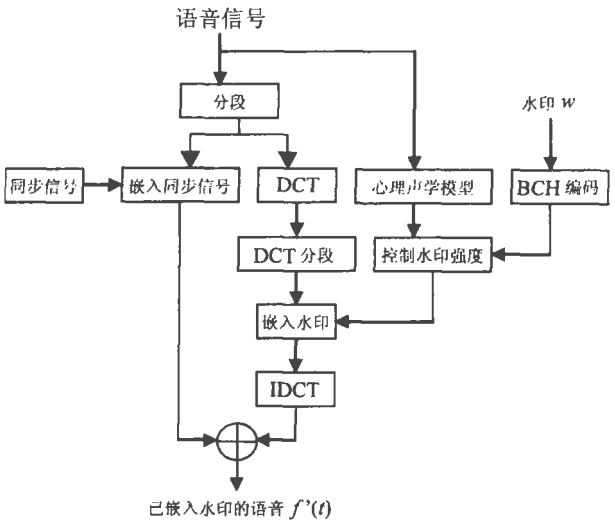

图 1 嵌入算法

Fig. 1 Embedding Algorithm

提取水印时，首先在待测音频信号中沿时间轴搜索同步信号，并利用数字通信的帧同步原理实现快速重同步。然后根据同步信号对音频数据分段。对每段音频信号进行心理模型分析，得出每段信号的频域上的 P^* 值。利用该 P^* 值，提取该段信号 DCT域上的 BCH码。提取出来的隐藏数据经过 BCH译码，得到一个估计的隐藏水印 W^*。

2 心理声学模型和控制水印强度

模型 II对每帧 1 152个 PCM音频数据分成两个粒度组（每组 576个 PCM数据）进行分析。每个粒度组的频域划分为 21个伸缩因子带。该模型

* 收稿日期：2005-07-04
基金项目：国家自然科学基金资助项目（60172067）
作者简介：王泳（1976年生），男，博士研究生；E-mail：jacobwong1220@126.com

计算出每个伸缩因子带在不可感知性下的最大可改变量 P 值。

本文算法对每帧 1 152 个 PCM 数据分成两组，每组 576个数据分别进行 DCT 变换。依照该模型的伸缩因子带的划分，把 DCT 域划分成 21 个频带，与 21个伸缩因子频带一一对应。

假设第 i帧第 j个粒度组 21 个频带中第 k个频带的最大可改变量为 P_k，希望把 1bit水印嵌入到该频带的能量中，且最终嵌入强度为 α_k，则

$$\alpha_k = f(P_k, k) \tag{1}$$

其中，P_k 作为经心理声学模型分析得出的结果，控制嵌入强度。但同时必须考虑嵌入信息的频段的位置。我们以多段音频片断作实验材料进行统计，发现 P值的大小与频段高低没有关系。但由于高频成分容易在各种信号处理中丢失，因此拥有大 P_k 值的高频段并非是嵌入水印的最佳位置。而低频信息在各类攻击及处理中比较容易保留下来，因此本算法的水印嵌入在低频段中，取得自适应、不可感知性能和稳健性的良好平衡。但如果嵌入的位置在高频段或具有小 P_k 值的低频段，本算法的 f 函数会调整嵌入强度于 P_k 值之上。这样在理论上，一定程度破坏了不可感知性能，但换来一定程度的稳健性能。总之，本算法把频率高低（即频段位置 k）亦作为控制嵌入强度的因子，从而达到自适应和稳健性的最大程度上的平衡。嵌入过程中基于模型 II对音频信号的分析，以及分析结果应用于水印嵌入的控制环节中的过程如图 2和图 3所示。

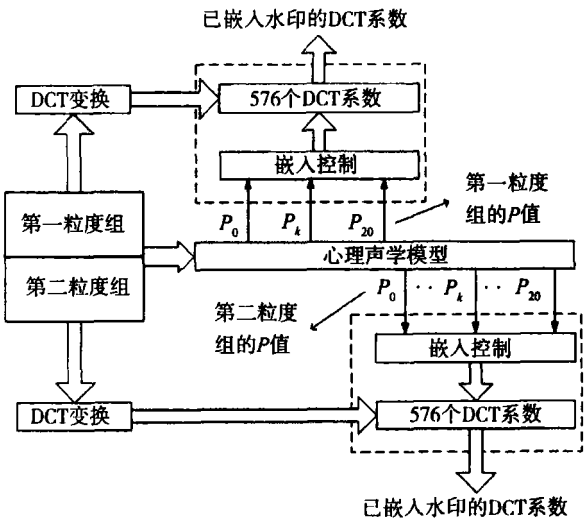

图 2 嵌入过程

Fig. 2 Embedding procedure

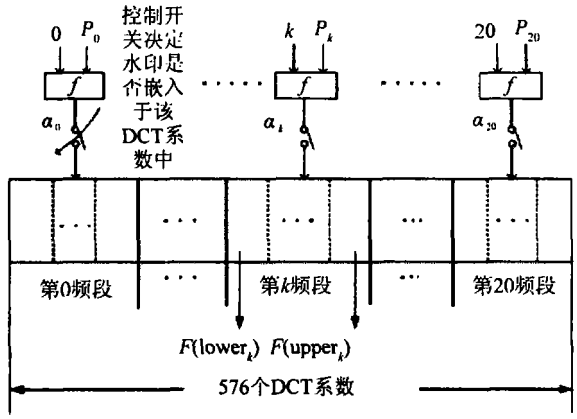

图 3 嵌入控制环节

Fig. 3 Embedding Strength Control

3 水印的嵌入

音频信号每 M 个抽样点为一帧，记其第 i帧信号为 A_i；同步码嵌入于每帧信号的第 1至第 N 的抽样点中。本文取 $M=1\ 164$，$N=12$。每帧余下的 1 152个抽样点分成两个粒度组，每个粒度组 576个抽样点用来隐藏水印数据。不失一般性，记其某一粒度组为 $FA_i(j)$，$0 \leqslant j < 576$。抽样点 $A_i(j)$ 在音频数据中的位置为 $i*M+N+j+576$（$A_i(j)$ 在第一粒度组内）或（在第二粒度组内）。对每个粒度组的进行 DCT变换，得到 DCT系数 $F_i(j)$，$0 \leqslant j < 576$。把 $F_i(j)$ 分成 21个频段，记第 k个频段为 $\mathrm{bank}_i(k) = \{F_i(j) \mid \mathrm{lower}_k \leqslant j < \mathrm{upper}_k\}$。$\mathrm{upper}_k$ 与 lower_k 为该频段的上下限。$\mathrm{bank}_i(k)$ 的能量记为 $En_i(k) = \sum_{j=\mathrm{lower}_k}^{\mathrm{upper}_k} F_i^2(j)$。

记嵌入于该频段的 1bit BCH 码为 $D_i(j)$，嵌入强度为 α_k。令嵌入后该频段的能量 $En_i'(k)$ 为

$$\begin{cases} En_i'(k) = En_i(k) - (En_i(k) \bmod \alpha_k) + \frac{3}{4}\alpha_k, \\ \text{if } D_i(j) = 1 \text{ and } (En_i(k) \bmod \alpha_k) \geqslant \frac{1}{4}\alpha_k \\ En_i'(k) = \left[En_i(k) - \frac{1}{4}\alpha_k\right] - \left[\left(En_i(k) - \frac{1}{4}\alpha_k\right) \bmod \alpha_k\right] + \frac{3}{4}\alpha_k \\ \text{if } D_i(j) = 1 \text{ and } (En_i(k) \bmod \alpha_k) < \frac{1}{4}\alpha_k \\ En_i'(k) = En_i(k) - (En_i(k) \bmod \alpha_k) + \frac{1}{4}\alpha_k \\ \text{if } D_i(j) = 0 \text{ and } (En_i(k) \bmod \alpha_k) \leqslant \frac{3}{4}\alpha_k \\ En_i'(k) = \left[En_i(k) - \frac{1}{2}\alpha_k\right] - \left[\left(En_i(k) + \frac{1}{2}\alpha_k\right) \bmod \alpha_k\right] + \frac{1}{4}\alpha_k \\ \text{if } D_i(j) = 0 \text{ and } (En_i(k) \bmod \alpha_k) > \frac{3}{4}\alpha_k \end{cases} \tag{2}$$

由公式（2）得该频段能量的改变量为：$\Delta_i(k) = En_i'(k) - En_i(k)$。把该改变量分散到频段中的每

一条频线 $F_i(j)$ 的能量中，则该频段的频线的能量改变量为 $\Delta_i(k,j)=\Delta_i(k)*\frac{F_i^2(j)}{En_i(k)}$。于是得到嵌入了水印的频线 $F_i'(j)=\mathrm{sgn}(F_i(j))*\sqrt{F_i^2(j)+\Delta_i(k,j)}$。

4　水印的提取

4.1　水印提取的原理和过程

水印的检测无需原始音频信号。设待检测音频信号为 $f^*(t)$，提取水印时，先检测出 $f^*(t)$ 中的同步码，从而建立水印的重同步。根据同步码的位置，对待检测的音频信号进行分段，每段有 $M=1\ 164$个抽样点。每段的后 $M-N=1\ 152$个抽样点分成前后两个粒度组分别进行 DCT变换，如图 4所示。

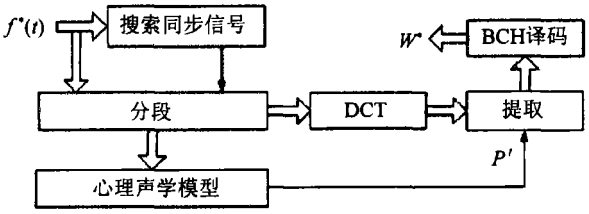

图 4　水印提取原理框图

Fig. 4　Watermark Extraction

对两个粒度组的 DCT系数经心理声学模型分析后得出的 P^*值，按照图 3的算法原理，在强度检测控制环节得到嵌入强度 α^*。

4.2　提取水印公式

记某一粒度组的 DCT系数为 F_i^*，其中 i与 j含义与上节相同。假设 1bit信息隐藏在该粒度组的第 k个频段 $\mathrm{bank}^*{}_i(k)=\{F_I^*(j)\mid \mathrm{up}_k\leqslant j\leqslant \mathrm{low}_k\}$中。该粒度组的能量为 $En_i^*(k)=\sum_{j=\mathrm{low}_k}^{\mathrm{up}_k}(F_i^*(j))^2$。记从 $\mathrm{bank}_i^*(k)$提取出来的二进制 BCH码为 X_i^*。则提取公式为

$$X_i^*(j)=\begin{cases}1, En_i^*(k)\bmod\alpha^*>\dfrac{\alpha^*}{2}\\ 0, \text{otherwise}\end{cases}\qquad(3)$$

从各段粒度组提取出来的 X_i^*连接起来，得出 BCH码 X^*。对 X^*进行 BCH译码，获得水印 W^*。

提取结果取决于 $F_i^*(j)$与 $F_i'(j)$、P^*与 P之间的误差。前者由 $f^*(t)$经受各类攻击或信号处理。而后者由①水印嵌入②$f^*(t)$经受各类攻击或信号处理这两个因素引起。水印能正确提取的前提之一是由①所引起的误差必须足够小，我们将在实验结果中分析这种误差；而②则是本文算法要抵抗的攻击。

5　实验结果

本实验采用 $N=12$的同步码，音频信号为两段 9 s的 WAVE格式的音频片段 Dlg.wav（人声）和 Piano. wav（钢琴声），44.1 kHz抽样频率，16 bits量化值，单声道。分段为 1 s 嵌入的水印为 40个字符的字符串。加入的噪声是服从分布的加性高斯噪声。

首先我们测试 P^*与 P的差别。实验结果显示，对于两段音频信号，水印嵌入对 P值的改变量最大值仅为为 1.7%和 1.86%，即水印音频无经过处理或攻击的情况下，P与 P^*几乎相等。这表明：把 $f^*(t)$输入心理声学模型进行分析，得出的 P^*值可以直接运用于提取过程，且本算法非常好的满足了不可感知性能。

然后我们测试水印抗高斯白噪声和 MP3压缩的稳健性。结果在表 1和表 2中给出。其中 SER表示字符错误概率。结果表明，水印即使在 SNR低于 10 dB时对噪声攻击仍具有良好的稳健性；同时也具有良好的抗 MP3压缩性能。

表 1　水印抗加性噪声的稳健性

Tab. 1　Robustness to Gaussian Noise

Dlg.wav				Piano.wav			
δ	SNR/dB	BER	SER	δ	SNR/dB	BER	SER
0	30.3	0	0	0	28.3	0	0
100	29.1	0	0	100	27.9	0	0
500	20.9	1.3	0	500	19.8	1.0	0
1000	15.7	2.1	0	1000	11.2	2.5	0
1500	12.4	12.7	0	1500	8.6	14.0	0
1800	11.3	19.0	0	1800	7.2	19.2	0
1900	10.2	20.2	0	1900	6.0	20.6	0
2000	10.0	21.3	21.2	2000	5.7	21.4	21.2
2100	9.2	21.5	21.2	2100	5.2	21.7	21.2

表 2　水印抗 MP3压缩的稳健性

Tab. 2　Robustness to MP3 compression

Dlg.wav				Piano.wav			
Bit rate/kb·s^{-1}	CR	SNR/dB	BER	Bit rate/kb·s^{-1}	CR	SNR/dB	BER
320	2.2	26.5	0	320	2.2	20.4	0
256	2.8	26.2	0	256	2.8	20.2	0
112	6.3	25.6	0	112	6.3	19.6	0
64	11.0	18.7	0	64	11.0	19.1	0
48	14.7	13.3	0	48	14.7	17.8	0
32	22.1	8.9	0	32	22.1	14.5	0

6 总 结

本文提出了一种同时结合时域和变换域的有意义音频水印自适应嵌入和盲检测算法。利用模型 II对音频进行分析，并应用于水印算法中。实现了自适应水印算法，在相同的不可感知性能下获得了更高的稳健性。

参考文献：

[1] BASSIA P, PITAS I. Robust audio watermarking in the time domain[C]. Proc. of EUSIP-98, Signal Processing IX, Theories and Applications, 1998, 25-28.

[2] LIE W N, CHANG L C. Robust and high-quality time-domain audio watermarking subject to psychoacoustic masking[C]. Proc. of IEEE Int. Sym. on Circuits and System, 2001, 2, 45-48.

[3] SWANSON M D, ZHU B, TEWFIK A H. Robust audio watermarking using perceptual masking[J]. Signal Processing, 1998, 66(3): 337-355.

[4] LEE S K, HO Y S. Digital audio watermarking in the spectrum domain[J]. IEEE Trans. on Consumer Electronics, 2000, 46(3): 744-750.

[5] WU C P, SU P C, JAY KUO C C. Robust and efficient digital audio watermarking using audio content analysis[J]. Proc. of SPIE: Security and Watermarking of Multimedia Contents II, 2000, 3971: 382-392.

[6] QIAO L, KLARA N. Non-invertible watermarking methods for MPEG encoded audio[J]. Proc. of SPIE, 1999, 3657: 194-202.

[7] WANG, CHAI P Q. A new adaptive audio watermarking algorithm for copyright protection Rangding[C]. Proc. of Natural Language Processing and Knowledge Engineering, 2003: 281-286.

[8] HUANG J W, WANG Y, SHI Y Q. A blind audio watermarking algorithm with self-synchronization[C]. IEEE International Symposium on Circuits and Systems, 2002, 3: 627-630.

An Adaptive Algorithm of Audio Watermarking Based on MPEG Psychoacoustics Model II

WANG Yong, LI Bin

(Department of Electronics and Communication Enineering, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: Digital audio watermarking has been a hot issue in recent years. Current research of audio watermarking mainly focuses on robustness and ignores adaptability. An adaptive algorithm of audio watermarking based on MPEG psychoacoustics model II is proposed to solve this problem. The algorithm is robust to noise attack and MP3 compression.

Key words: audio watermarking; robustness; adaptability; MPEG psychoacoustics model II