

# DCT 域音频水印算法的嵌入位置选择策略<sup>\*</sup>

马翼平 韩纪庆

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

**摘要** 本文在 DCT 最大误差模型研究的基础上提出了一种新的音频水印嵌入位置选择策略。建立了水印嵌入位置和嵌入水印后的音频文件的听觉感知性之间的关系,根据音频水印的不可听性的要求选择最优的嵌入位置,然后选择不同的水印强度调整鲁棒性,从而最大程度地保持音频水印的不可听性和鲁棒性,为解决音频水印嵌入过程中不可听性和鲁棒性之间的矛盾提供了一种策略。

**关键词** 音频水印, DCT, 信噪比, 嵌入位置

## Choice Approach for Embeddable Position of Audio Watermarking in the DCT Domain

MA Yi-Ping HAN Ji-Qing

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

**Abstract** A new choice approach for embeddable position of audio watermarking scheme in the DCT domain based on research of DCT maximum error modal is proposed. A math relationship between the watermark embeddable position and the audible quality of the audio file with watermark is set up. First the optimal position to be embedded based on the inaudible requirement is selected. Second the power coefficient of watermark is adjusted to satisfy the requirement of robustness. So the utmost need of inaudible and robust of audio watermarking scheme was filled. It is a good approach to settle the trade-off between the two requirements of audio watermarking: inaudible and robustness.

**Keywords** Digital audio watermarking, Discrete cosine transform, SNR, Embeddable position

## 1 引言

多媒体技术及计算机网络的日益普及,使得数字多媒体信息的广泛传播变得非常容易,如何既充分利用因特网的便利,又能有效地保护知识产权,已受到人们的高度重视<sup>[1~3]</sup>。因此,数字水印技术在近年来获得了很大的发展。

数字水印技术,就是通过修改媒体信号而有意识地在其中隐藏一些秘密信息(即数字水印),并且这些隐藏的信息在视觉和听觉上不可感知,从而来标识该多媒体信息的所有权。人们已经提出了许多种数字水印的算法<sup>[2]</sup>,有的技术是在空间域上进行嵌入,有的算法则是在变换域如 DCT,或小波域进行。有的算法嵌入的是二进制序列,也有的作者提出使用图像作为水印能够达到更好的版权保护作用<sup>[4]</sup>。

DCT 域的水印算法由于其计算量较小,且与国际数据压缩标准兼容,目前对该领域的研究比较多。传统的基于 DCT 的水印算法,如 Cox<sup>[5]</sup>的基于全局 DCT 的方法,随着嵌入比特数的增加,进行 MPEG 攻击后,其检出率大大下降。而许多基于分块 DCT 的方法,对于不同的分块,引入了 HAS 的概念,采用不同的嵌入强度,但对于每一分块则采用固定数目的系数用于嵌入,例如一定数量的中频系数,但对于中频系数却没有一个统一的定义,从而忽略了不同分块之间可供嵌入的系数之间的区别,也即用于嵌入的系数中存在有冗余,影响其鲁棒性。

在音频水印中,还没有研究者提出一个定量的标准来确

定嵌入位置的选择策略。文[4]提出了将视觉可辨的二值图像作为水印的 DCT 域分段音频水印算法,将水印嵌入到每一段的中频系数上,但文中算法将中频系数定义为每一段的第 2 个 DCT 系数,没有足够的说服力。

不可听性(或者说含音频文件的听觉质量)是音频水印用户特别关心的问题之一,也是音频水印测评的重要技术指标。虽然从直观上讲,音频文件的听觉质量随着水印嵌入位置的不同而不同,但由于数字音频文件的听觉质量直接关系其商业价值,因此对嵌入水印所导致的听觉质量损失必须有一个科学的、定量的估计。在音频压缩领域中信噪比是客观评价音频听觉质量的常用指标之一,这个指标对于数字音频水印也同样适用。

本文针对典型的分块 DCT 域数字水印算法,利用 DCT 变换的最大误差模型,建立了水印嵌入位置与信噪比和峰值信噪比的理论关系,从而提出了音频水印算法的嵌入位置选择策略,为解决水印算法的不可听性和鲁棒性之间的折中提供了一种解决方法。

## 2 DCT 的最大误差模型

DCT 是一种将时域强度信息转换成频域系数信息的方法(式 1),具有很好的能量压缩性,仅用部分变换系数即可代表原始序列能量的总体,在数据通讯中具有广泛的应用。

$$F = f \times C \quad (1-a)$$

$$f = F \times C' \quad (1-b)$$

<sup>\*</sup> 本课题受国家自然科学基金项目(60173044)、教育部跨世纪优秀人才培养计划基金资助。马翼平 博士生,研究方向:音频水印,音频信息伪装;韩纪庆 博士生导师,博士,教授。

其中  $f$  为时域序列  $1 \times N$  行向量,  $F$  为频域系数  $1 \times N$  行向量,  $C$  为  $N \times N$  转换矩阵,  $C'$  为  $C$  的逆矩阵, 转换矩阵  $C$  可表示如下:

$$C(i, j) = \begin{cases} \frac{1}{\sqrt{N}}, & j=0 \\ \sqrt{\frac{2}{N}} \cos \frac{(2i+1)j\pi}{2N}, & j \neq 0 \end{cases} \quad (2)$$

由于 DCT 是一种正交变换,  $C$  为正交矩阵, 显然有  $C'(i, j) = C(j, i)$ .

写成解析形式即有

$$F(k) = c(k) \sum_{n=0}^{N-1} f(n) \cos \left[ \frac{(2n+1)k\pi}{2N} \right] \quad (3-a)$$

$$f(n) = \sum_{k=0}^{N-1} c(k) F(k) \cos \left[ \frac{(2n+1)k\pi}{2N} \right] \quad (3-b)$$

$$\text{其中, } c(u) = \begin{cases} \frac{1}{\sqrt{N}}, & u=0 \\ \sqrt{\frac{2}{N}}, & u \neq 0 \end{cases} \quad (4)$$

假设在频率系数向量  $F$  上添加一个误差  $E(E = [E(0) \dots E(k) \dots E(N-1)])$  向量, 则转换成时域序列后的结果成为:

$$\begin{aligned} f'n &= \sum_{k=0}^{N-1} c(k) [F(k) + E(k)] \cos \left[ \frac{(2n+1)k\pi}{2N} \right] \\ &= f(n) + \sum_{k=0}^{N-1} c(k) E(k) \cos \left[ \frac{(2n+1)k\pi}{2N} \right] \\ &= f(n) + e(n) \end{aligned} \quad (5)$$

其中  $e(n)$  是误差向量  $E$  衍生出来的对时间域上第  $n$  个点(即  $f(n)$ ) 的影响. 事实上, 由于 DCT 变换为线性变换, 有

$$e(n) = \sum_{k=0}^{N-1} e_k(n) \quad (6)$$

其中

$$e_k(n) = c(k) E(k) \cos \left[ \frac{(2n+1)k\pi}{2N} \right] \quad (7)$$

表示频域第  $k$  个点(即  $F(k)$ ) 上的误差  $E(k)$  对时间域上第  $n$  个点(即  $f(n)$ ) 的影响.

考虑误差向量  $E$  对时域序列  $f$  所形成的最大误差  $\max(e)$ , 即

$$\max(e) = \max_n |e(n)| = \max_n \left| \sum_{k=0}^{N-1} e_k(n) \right| \quad (8)$$

从而有

$$\max(e) \leq N \max_n \left( \max_k |e_k(n)| \right) \quad (9)$$

令

$$\max e_k = \max_n |e_k(n)| = \max_n \left| c(k) E(k) \cos \left[ \frac{(2n+1)k\pi}{2N} \right] \right| \quad (10)$$

表示  $F(k)$  上的误差  $E(k)$  对时域序列  $f$  的最大影响.

式(10)可以写成下面的形式, 即

$$\max e_k = \max_n \left| \sqrt{\frac{2}{N}} E(k) t(n, k) \right| = \sqrt{\frac{2}{N}} |E(k)| \max_n |t(n, k)| \quad (11)$$

$$\text{其中, } t(n, k) = \begin{cases} \cos \frac{\pi}{4}, & k=0 \\ \cos \frac{(2n+1)k\pi}{2N}, & k \neq 0 \end{cases}$$

定义误差向量  $E$  对  $f$  所形成的最大误差  $e_f = \sup \max(e)$ , 则有

$$e_f = N \max_n (\max_k e_k) = \sqrt{2N} \max_k (|E(k)| \max_n |t(n, k)|) \quad (12)$$

### 3 音频质量评价准则

作为性能指标而言, 如何评价由于水印的嵌入而导致原始声音退化的程度是一个需要着重考虑但实际上却常常被忽视的一个问题. 在音频信息处理中大多数的退化程度或听觉质量评价均属于差别失真评测准则 (Difference Distortion Measures), 这些准则都是从原始音频与退化或失真音频的数学对比得到的<sup>[5]</sup>. 目前水印研究中对这种音频降质的度量均借鉴音频编码及压缩中的方法, 即通过信噪比 (Signal to Noise Ratio, SNR) 或峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 来衡量嵌入水印后音频文件的听觉质量.

假设  $A$  是含有  $L$  个采样数据的原始数字音频信号, 它可以表示为  $A = \{a(l), 0 \leq l < L\}$ , 其中  $a(l)$  是第  $l$  个数据的幅值.  $B = \{b(l), 0 \leq l < L\}$  表示嵌入水印后的数字音频信号, 则信号  $A$  与信号  $B$  之间的 SNR 和 PSNR 分别定义为

$$\text{SNR} = 20 \log \frac{\sqrt{\sum_l a^2(l)}}{\sqrt{\sum_l [a(l) - b(l)]^2}} \quad (13-a)$$

$$\text{PSNR} = 20 \log \frac{b(m)}{\sqrt{\frac{1}{L} \sum_l [a(l) - b(l)]^2}} \quad (13-b)$$

其中  $b(m)$  为音频信号  $B$  采样所得的最大值.

如果在时域音频信号  $A$  对应的频域信号  $F_A = \text{DCT}(A) = \{f_A(l), 0 \leq l < L\}$  上添加一个水印向量  $E = \{E(l), 0 \leq l < L\}$  后, 再转换成时域信号  $C = \text{IDCT}(F_A) = \{c(l), 0 \leq l < L\}$ , 不考虑其它的误差, 考虑式(12), 则信号  $A$  与信号  $C$  之间的 SNR 和 PSNR 变为如下形式

$$\text{SNR} \geq 20 \log \frac{\sqrt{\sum_l a^2(l)}}{\sqrt{L \times e_f^2}} = 20 \log \frac{1}{\sqrt{L} e_f} \sqrt{\sum_l a^2(l)} \quad (14-a)$$

$$\text{PSNR} \geq 20 \log \frac{\sqrt{L} c(m)}{\sqrt{L \times e_f^2}} = 20 \log \frac{c(m)}{e_f} \quad (14-b)$$

其中  $c(m)$  为音频信号  $C$  采样所得的最大值.

### 4 水印的嵌入策略

一个有效的音频数字水印系统, 必须满足不可听性和鲁棒性两个特性, 这是一对互相矛盾的因素, 水印算法必须折中考虑不可听性和鲁棒性. 水印的不可听性和鲁棒性直接与嵌入水印的位置和嵌入水印的强度有关, 从而也影响到音频载体的听觉质量退化程度.

本文提出这样一种音频水印嵌入策略, 首先由音频水印载体 PSNR 或 SNR 的要求(即式(14-a)或(14-b))决定水印的最优嵌入位置, 从而保证水印的不可听性, 然后在此基础上参考文献[6]调整水印嵌入的强度因子, 从而满足音频水印的鲁棒性要求, 这样能够最大程度地解决音频水印算法不可听性和鲁棒性之间的矛盾.

### 5 实验结果及讨论

以典型的分块 DCT 域音频数字水印算法为例, 每个音频数据块的数据个数为 8, 每个数据块嵌入一个水印信息. 下面的表 1 即为当音频数据块的大小为 8(即  $0 \leq n, k < 8$ ) 时的 |

$t(n, k)$  数值表。

表 1  $|t(n, k)|$  数值表

$ t(n, k) $		n							
		0	1	2	3	4	5	6	7
k	0	0.707	0.707	0.707	0.707	0.707	0.707	0.707	0.707
	1	0.981	0.832	0.556	0.195	0.195	0.556	0.832	0.981
	2	0.924	0.383	0.383	0.924	0.924	0.383	0.383	0.924
	3	0.832	0.195	0.981	0.556	0.556	0.981	0.195	0.832
	4	0.707	0.707	0.707	0.707	0.707	0.707	0.707	0.707
	5	0.556	0.981	0.195	0.832	0.832	0.195	0.981	0.556
	6	0.383	0.924	0.924	0.383	0.383	0.924	0.924	0.383
	7	0.195	0.556	0.832	0.981	0.981	0.832	0.556	0.195

因为每一个数据块只嵌入一个水印信息,则水印向量  $E$  将大大简化,变成一个只有一个非零元的向量,从而式(12)被式(11)取代,式(14-a)和式(14-b)也将被大大简化。

从表 1 可以得到以下结论:

(1)在频域的某一位嵌入水印对时域的影响有对称关系,即对时域第 0 位的影响和对第 7 位的影响相同,同样,对时域第 1 位和第 6 位的影响相同,对时域第 2 位和第 5 位的影响相同,对时域第 3 位和第 4 位的影响相同;

(2)考虑对时域第 7 位的影响,随着嵌入位置频率的增加(从低频到高频,即从频域第 1 位到第 7 位),对时域第 7 位的影响呈下降趋势;

(3)在频域第 0 位嵌入水印对时域各个采样点的影响相同。频域第 0 位系数是时域幅值数据的一种平均值,一般称作直流系数,改变直流系数相当于改变了时域幅值的整个分布情况,但对时域每个采样点的影响却相同;

(4)在频域第 4 位嵌入水印对时域各个采样点的影响相同,并且与在频域第 0 位即直流系数上嵌入水印对时域各个采样点的影响分别相同;

(5)记  $T = \max_n |t(n, k)|$ , 可见当  $k=4$  时,  $T=0.707$  取最小值,根据式(11)知  $\max e_k$  取最小值,从而相应的 SNR 和 PSNR 取最大值,即音频载体具有最大的听觉质量。所以在典型的分块 DCT 音频水印算法中,将水印嵌入到频域的第 4 个 DCT 系数上能够最大程度地保证水印算法的不可听性,从

而间接地保证水印算法的鲁棒性。

**结论** 本文在 DCT 最大误差模型研究的基础上针对典型的分块 DCT 域数字水印算法,建立了水印嵌入位置和嵌入水印后的音频文件的听觉感知性之间的理论关系,从而提出了一种新的音频水印算法的嵌入位置选择策略,根据音频水印算法的不可听性的要求选择最优的嵌入位置,从而最大程度地保持音频水印算法的不可听性和鲁棒性,为解决音频水印嵌入过程中不可听性和鲁棒性的矛盾提供了一种解决方法。

研究如何将本文提出的音频水印嵌入位置选择策略与具体的音频水印算法结合是今后的工作重点。

### 参考文献

- 1 Voyatzis G, Pitas I. The use of watermarks in the protection of digital multimedia products. Proceedings of the IEEE, 1999, 87(7): 1197~1207
- 2 Podilcbuk C I, Delp E J. Digital watermarking: algorithms and applications. IEEE Signal Processing Magazine, 2001(7): 33~46
- 3 Cox I J, Kilian J, Leighton T. Secure spread spectrum watermarking for multimedia. IEEE Transaction on Image processing, 1997, 6(12): 1673~1687
- 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. 声学学报, 2001, 26(5): 464~467
- 5 韩纪庆, 张磊, 郑铁然. 语音信号处理(第 1 版). 北京: 清华大学出版社, 2004
- 6 刘瑞祺, 谭铁牛. 水印能量估计的一般性框架. 计算机学报, 2001, 24(3): 242~246

## 中国计算机学会 全国第五次程序设计语言发展与教学学术会议 征文通知

全国第五次程序设计语言发展与教学学术会议定于 2006 年 10 月在南京召开。本次会议由东南大学主办, 将就程序设计语言现状与发展、各类程序设计语言及其设计与实现、程序设计语言教学与教材等方面的内容进行广泛的学术交流。会议论文集将由国家出版社正式出版, 其中部分优秀论文将在会后推荐给国内权威核心杂志发表。现将有关事项通知如下:

### 一、征文范围(包括但不限于)

程序设计语言历史、现状与发展; 面向对象语言及相关技术; 各类建模语言及其设计、实现与应用面向网络应用的程序设计语言; 程序设计语言分析、评价与比较程序设计语言语法、语义与语用; 程序设计语言教学、教材与课件软件开发过程中各类描述语言; 并发、并行与实时程序设计语言; 面向网络与 Web 的各类语言; 第四代语言与数据库语言; 程序设计语言形式化描述技术与方法程序设计语言标准化; 其它各种新型程序设计语言(包括逻辑型语言、函数型语言等); 写作语言与工具

### 二、来稿要求

1. 本次会议只接受 Email 投稿。
2. 中英文稿均可, 一般不超过 6000 字。为了便于出版论文集, 来稿必须附中英文摘要、关键词、资助基金与主要参考文献, 注明作者及主要联系人姓名、工作单位、详细通信地址(包括 Email 地址)与作者简介。稿件要求采用 WORD 或 PDF 格式。

### 三、联系信息

投稿地址: 东南大学计算机科学与工程系 周晓宇(zhouxy@seu.edu.cn)  
会务情况: 东南大学计算机科学与工程系 徐宝文 周晓宇(zhouxy@seu.edu.cn)

四、重要日期 征文截止: 2006 年 4 月 10 日 录用通知发出: 2006 年 4 月 25 日 正式论文提交: 2006 年 5 月 10 日