

用于版权管理的数字音频水印算法

赵学敏* 郭宇弘 邹学强 刘建 颜永红
(中国科学院语言声学内容与理解重点实验室 北京 100190)

摘要: 该文提出了一种用于多媒体资料版权管理的自同步音频盲水印算法。由于版权应用需要通过解码结果得到素材的使用情况,对解码结果时间精度有一定要求,同时对鲁棒性和不可感知性要求也较高。针对之前算法存在鲁棒性不够,时间精度低的问题。该文提出的算法利用扩频的方法和音频信号的短时相关的特性提高嵌入水印的鲁棒性的同时,通过心理声学模型的应用改善算法的不可感知性。算法通过合理设计了水印信息帧结构和同步策略提高了解码结果的时间精度。实验结果表明,该文的方法在具有很高不可感知性的情况下,能够有效地抵抗各种常见的音频编辑操作。

关键词: 数字水印; 心理声学; 扩频; 版权保护

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2011)10-2384-06

DOI: 10.3724/SP.J.1146.2011.00009

Digital Audio Watermarking Algorithm for Media Copyright Management

Zhao Xue-min Guo Yu-hong Zou Xue-qiang Liu Jian Yan Yong-hong
(Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics,
Chinese Academy of Sciences, Beijing 100190, China)

Abstract: A blind audio watermarking algorithm with self-synchronization for copyright protection is proposed. In the application of media copyright management, the decoding results should have a time precision, better robustness and fidelity. However, most of algorithms can not provide enough robustness and time precision. To deal with these problems, in this paper, spread spectrum theory and the short-time relevance of audio signal are used to generate a watermark. To improve the fidelity of audio, a psychoacoustic auditory model is further used. Moreover, special synchronization strategy and watermark message frame structure are used to get a high time precision decoding results. Experimental results show that the watermarking scheme is robust against common audio editing attacks while keeping the watermarked audio a high perceptual quality.

Key words: Digital watermarking; Psychoacoustic model; Spread spectrum; Copyright protection

1 引言

数字水印技术作为信息安全技术领域的一个重要方向,在认证多媒体来源、进行所有权认证和追踪非授权行为等方面得到广泛的应用^[1,2],具有十分广阔的应用前景。数字水印算法一般应具有以下特点:不可感知、不可擦除、鲁棒、易于提取^[3]。其中,鲁棒性和不可感知性是数字水印算法中的一对主要矛盾,算法的设计要兼顾两方面的需求。数字音频水印算法一般会根据心理声学模型^[4]调整嵌入的水印信号,使得嵌入的水印在保证一定鲁棒性的前提下尽可能不为人所察觉。

数字音频水印算法一般采取以下的方法,相位

编码^[5]、回声隐藏^[6]、扩频水印^[7,8]等。相位编码和回声隐藏技术具有良好的不可感知性,但缺点是嵌入的水印信息量很小且鲁棒性较差。在Cox等人^[7]首先将扩频的方法引入到数字水印之后,扩频水印的方法得到了广泛的应用,成为数字水印算法中的主流方法。Kirovski等人^[8]提出了改进的扩频音频水印方法,并对扩频类算法作了具体的性能分析,详细地讨论了扩频算法的同步问题和提高鲁棒性的方法。当前应用于版权保护的一些算法^[9,10],主要通过扩频的应用^[9]和设法减少载体信号的影响^[10],但由于算法缺乏同步保护的机制,无法抵御编辑中剪切、拼接操作。所以,如何在保证算法鲁棒性和不可感知性的前提下,进一步提高算法的时间精度,成为有待解决的问题。

针对多媒体资料和版权管理中编辑操作类型多而复杂,且要求解码结果时间精度高的需求。本文

2011-01-06 收到, 2011-06-24 改回

国家自然科学基金(10925419, 90920302, 10874203, 60875014, 61072124, 11074275)资助课题

*通信作者: 赵学敏 zhaoxuemin@hcl.ioa.ac.cn

提出了一种扩频的自同步音频盲水印的算法。算法利用音频信号的短时平稳特征，减少载体音频对水印的影响，同时使用 MPEG 心理声学模型来确保嵌入信息的不可感知性。通过合理地设计水印信息帧结构和多扩频码的应用，将解码结果时长统计的时间精度提高到秒级。最后，还通过同步相关时的一些特征使得同步位置的确定更为精准，减少误同步的影响。实验结果表明算法在具有很高不可感知性的同时具有很强的鲁棒性，解码结果有更准确的时间精度。

论文按照下列结构组织：首先介绍水印的嵌入算法，介绍水印信息帧的组织结构和心理声学模型的应用；第 3 节是解码算法的介绍，介绍了同步的策略，并讨论了提高同步和解码准确率的方法；第 4 节是具体的实验结果和分析；最后给出结论。

分段 1		分段 2		...	分段 N	
同步位 S_1 (扩频码 $P_{s1}, 1 \text{ bit}$)	信息 M_1 (扩频码 $P_m, F \text{ bit}$)	同步位 S_2 (扩频码 $P_{s2}, 1 \text{ bit}$)	信息 M_2 (扩频码 $P_m, F \text{ bit}$)	...	同步位 S_N (扩频码 $P_{sN}, 1 \text{ bit}$)	信息 M_N (扩频码 $P_m, F \text{ bit}$)

图 1 信息帧结构图

2.2 平衡调制

由于音频信号一般具有短时平稳的特性，所以，相邻几帧之间的频域的幅值相对稳定。可以利用该特性减少载体音频对水印的影响。嵌入时，将一个比特的水印嵌入信息隐藏在连续几帧的音频信号之中。如图 2 所示，当嵌入的比特为 1 时，将前 K 帧的基带信号 S 设为 +1，接下来的 K 帧的基带信号设为 -1，当嵌入比特为 0 时， S 的取值顺序相反。解码时，将前 K 帧的信号减去后 K 帧的信号，可以在一定程度上减弱载体信号对水印的影响。

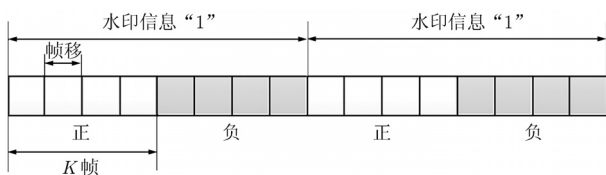


图 2 平衡调制(嵌入比特为 1 的情况)

2.3 利用扩频码嵌入过程

将输入的音频信号做分帧处理，加汉宁窗后做 FFT 变换，变换的前 N 个系数用于水印嵌入， N 等于扩频码的长度。设待嵌入水印的频域信号的幅值为 $\{X(n), 0 < n < N\}$ ，扩频码为 $\{w(n), 0 < n < N\}$ 。扩频码 $w(n)$ 由嵌入比特的类型是信息位还是同步位来确定。将 $X(n)$ 转为 dB 值 $x(n) = \log(X(n))$ ，然后对通过式(1)嵌入水印。

$$y(n) = x(n) + w(n) \cdot S \cdot \alpha(n) \quad (1)$$

其中 S 为该帧的基带码信号，取 +1 或 -1， $\alpha(n)$ 通过

2 水印嵌入算法

2.1 水印信息帧组织结构

为提高统计结果的时间精度，将水印信息分为 N 段，每段信息长度为 $F=L/N$ 。 L 为水印嵌入信息的位数，选择 N 使得表示该长度水印信息的载体音频长度小于 1 s。在每个分段之前，插入一位作为同步信息，同步位与信息位采用不同的扩频码嵌入，同时，不同分段的同步位之间，也用不同的扩频码表示。解码时，可以通过扩频码的序号确定该分段在信息码中位于第几段。由此得出的信息帧结构如图 1 所示。当某一分段丢失时，如果通过一定解码策略使得其他分段的解出不受其影响，可以由此提高时长统计的时间分辨率。

心理声学模型计算得出，用于控制嵌入能量，使之不影响听感。设 $Y(n)$ 是 $y(n)$ 的频域幅值，则

$$Y(n) = X(n) \cdot 10^{w(n)S\alpha(n)} \quad (2)$$

将嵌入水印的信号 $Y(n)$ 再结合原来的相位信息做 IFFT 变换，完成一帧的嵌入。完成水印嵌入的各帧之间再做重叠相加，恢复时域信号，得到嵌水印的音频，完成水印的嵌入过程，水印嵌入的过程见图 3。

2.4 心理声学模型的应用

本算法利用 ISO/IEC 11172-3^[11] 定义的心理声学模型计算掩蔽模型，有关计算的细节在文献[11]有详细描述。在 MPEG 压缩过程中，待处理的频域信号被划分为若干子带，心理声学模型计算出每一个子带的信掩比(Signal to Mask Ratios, SMR)，反映出该子带的在音频中的可听程度。计算好的 SMR

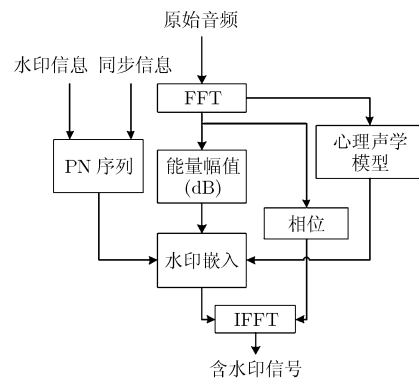


图 3 水印嵌入过程

为比特分配提供依据。对于 SMR 较高的子带, 将分配给更多的比特, 这样会使该子带有更小的量化噪声。在进行水印嵌入的过程中, 将某一频点嵌入的能量按照 SMR 进行调节, 使得水印更多的嵌入到不为察觉的部分中去。这样, 嵌入的水印就能根据音频本身的特性动态调整, 从而减少对音频引起的整体主观失真。

3 水印算法解码原理

3.1 相关解码

根据嵌入公式(2), 对第 k 帧嵌入水印的信号 $Y^k(n)$ 和扩频码序列 $w(n)$ 做相关的结果为

$$\begin{aligned} C^k &= \sum_n w(n) \cdot \lg(Y^k(n)) \\ &= S^k \sum_n w^2(n) \cdot \alpha^k(n) + \sum_n w(n) \cdot x^k(n) \end{aligned} \quad (3)$$

式(3)中的第 1 部分为 $w(n)$ 的自相关项, 正负取决于 S^k 的取值, 第 2 部分为互相关项, 理论上两者不相关, 因而应当远小于自相关项。

设 $Y^{k'}(n)$ 为嵌入信号 $Y^k(n)$ 经过缩放 M 倍的信号, 则

$$\begin{aligned} C^{k'} &= \sum_n w(n) \cdot \lg(Y^{k'}(n)) = \sum_n w(n) \cdot (\lg MY^k(n)) \\ &= C^k + \lg M \cdot \sum_n w(n) \end{aligned} \quad (4)$$

当 $\sum_n w(n) = 0$ 时, 相关得分与原信号的幅值无关, 从而使水印算法可以抵抗对时域信号的幅值调整等操作。所以在选取扩频码时, 需选取平衡的扩频码, 即扩频码的 +1, -1 数目相等。

水印嵌入时还采用了平衡调制的结构, 前 K 帧嵌入的符号与后 K 帧反向, 分别计算这 $2K$ 帧的结果, 并根据以下式子确定最终的比特信息。

$$I = \sum_{k=0}^{K-1} C^k - \sum_{k=K}^{2K-1} C^k \quad (5)$$

当 $I > 0$ 时, 解码结果为 1, 当 $I < 0$ 时, 解码结果为 0。

将式(3)代入式(5), 得

$$\begin{aligned} I &= \sum_{k=0}^{K-1} C^k - \sum_{k=K}^{2K-1} C^k \\ &= S \cdot \sum_n \sum_k \alpha^k(n) + \sum_n w(n) \cdot \left(\sum_{k=0}^{K-1} x^k(n) - \sum_{k=K}^{2K-1} x^k(n) \right) \\ &= S \cdot \sum_n \sum_k \alpha^k(n) + \sum_n w(n) \cdot x'(n) \end{aligned} \quad (6)$$

与式(3)相比, 互相关项中的 $x(n)$ 变为 $\sum_{k=0}^{K-1} x^k(n) - \sum_{k=K}^{2K-1} x^k(n)$, 设为 $x'(n)$ 。由于信号短时平稳, 当 $\sum_{k=0}^{K-1} x^k(n)$ 与 $\sum_{k=K}^{2K-1} x^k(n)$ 相关时, $x'(n)$ 的方差小于 $x(n)$ 。设变量 $x_1(n), x_2(n)$ 为服从相

同高斯分布的随机变量, 即 $x_1(n) \sim N(0, \sigma_x), x_2(n) \sim N(0, \sigma_x)$, 则

$$\begin{aligned} E(x_1 - x_2)^2 &= E(x_1)^2 + E(x_2)^2 - 2E(x_1 \cdot x_2) \\ &= \sigma_x^2 + \sigma_x^2 - 2\rho_{x_1 x_2} \cdot \sigma_x \cdot \sigma_x \\ &= 2(1 - \rho_{x_1 x_2})\sigma_x^2 \end{aligned} \quad (7)$$

当 $x_1(n), x_2(n)$ 相关, $\rho_{x_1 x_2}$ 接近于 1, 所以根据以上分析, $x'(n)$ 的方差要远小于 $x(n)$ 。根据 Kirovski^[8]文中给出的分析, 扩频水印系统的虚警 P_{FA} 和漏警 P_{MD} 概率分别为

$$P_{FA} = \frac{1}{2} \operatorname{erfc} \left(\frac{\tau \sqrt{N}}{\sigma_x \sqrt{2}} \right) \quad (8)$$

$$P_{MD} = \frac{1}{2} \operatorname{erfc} \left(\frac{(E[xw] - \tau) \sqrt{N}}{\sigma_x \sqrt{2}} \right) \quad (9)$$

因为互补误差函数 $\operatorname{erfc}(x)$ 为减函数, 式(8), 式(9)中方差项的减少使得错误概率减少。这样, 利用短时平稳的相关性可以降低解码错误率。

3.2 同步过程

同步过程分为两步, 即帧同步和数据同步。帧同步就是确定每帧解码位置的过程。通过事先在水印信号中加入同步信号的方法来确定解码位置。由于同步位采用了不同的扩频码用于标识不同分段的结构, 得到该同步位对应的扩频码序号即可确定分段位置信息。

为保证不同扩频码能正确标识其所在分段, 要求不同扩频码之间的互相关得分尽可能小。本系统选取的扩频码为平衡 Gold 码^[12], 同一组的 Gold 码之间具有良好的互相关特性。同时, Gold 码的 +1, -1 数目相等, 满足本系统对扩频码所要求的平衡特性。

在搜索同步位的过程中, 我们采用滑动相关的方法, 取相关得分的最大的位置作为帧同步的位置。在找到合适的帧同步位置后, 还要进行数据同步, 确定数据的起始位置。首先要检查同步位的解码结果是否和嵌入的同步信息一致, 如不一致, 则这一分段的同步失败。在开始同步之前, 需要能够找到一个完整的信息周期才能进行信息位的解码, 否则信息码不完整, 解码结果无意义。此时先进入同步模式 1, 该模式要求一个周期内各个分段都符合帧同步条件, 且分段之间的连续性符合预先设置的规律。在找到一个周期之后, 进入同步模式 2。如果接下来找到一个分段符合下列条件, 则认为同步: (1)得到的扩频码序号与同步位信息与前一个周期一致; (2)设帧同步位置为 Pos_2 与上一个分段的帧同步位置为 Pos_1 , 一个分段理论上的长度为 L_f , 即满足条件 $|\text{Pos}_2 - \text{Pos}_1 - L_f| < L_s$, L_s 根据实验确定。当解码经过了一个分段的距离之后还没有同步, 设此

时解码位置为 Pos_3 ，即满足条件 $|Pos_3 - Pos_1 - L_f| > L_m$ ， L_m 也根据实验确定。则退出同步模式 2，重新搜索同步模式 1 状态。

同步模式的流程如图 4 所示，同步后，即可解出该分段的信息码。

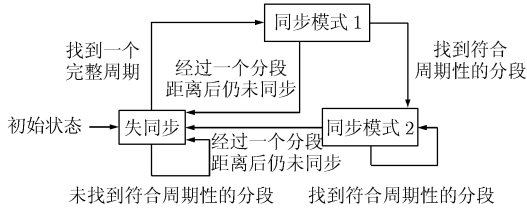


图 4 同步模式

3.3 利用相关特性减少误同步

在进行相关时，相关得分的曲线图应呈现尖峰的特性，如图 5 所示。该特性可以用来减少误同步。由于同步位与信息位采用的扩频码不同，所以，同步时，信息位可以看作没有嵌入水印的情况。由于平衡调制的结构，解码时，前后未嵌水印的 N 帧信号与一个比特前后 N 帧的信号在分别相减时，也会有较大的得分，该得分与比特解码得分反向。统计得到正向峰与反向峰的平均距离为 L_p ，将帧同步位置处左右的解码得分 $C(k)$ 乘以下列的窗函数 $W(k)$

$$W(k) = \begin{cases} 1, & (|k - k_{\text{sync}}| \leq L_p/2) \\ -1, & (L_p/2 < |k - k_{\text{sync}}| \leq L_p) \end{cases} \quad (10)$$

累加后得到一个分数

$$S_w = \sum_k C(k) \cdot W(k) \quad (11)$$

进行同步判定时，要求得分 S_w 超过某一阈值。由于未同步位置随机出现的高峰值周围不会有此特征，将该条件考虑进来可以有效地减少误同步。

3.4 分段信息拼接

同步并解码之后，可以通过解码结果和其位置信息得到音频的使用时间信息。我们需要将连续出

现且符合周期性规律的一些分段拼接在一起。判定拼接的条件需同时满足：(1) 帧同步位置 Pos_2 与上一个分段的帧同步位置 Pos_1 符合 $|Pos_2 - Pos_1 - L_f| < L_s$ ， L_f 为一个分段理论上的长度， L_s 根据实验确定；(2) 得到的扩频码序号为下一个应该出现的序号；(3) 解码结果与之前一个周期分段的结果相同。

对于某些孤立出现的分段，即无法直接通过以上条件拼接到一个完整周期内的分段，当同时符合以下条件时，也被拼接到一起。(1) 帧同步位置 Pos_2 与上一个分段的帧同步位置 Pos_1 符合 $|Pos_2 - Pos_1 - N \cdot L_f| < L_s$ ， L_f 与 L_s 定义与之前相同，且 $N <$ 分段数；(2) 得到的扩频码序号为后 N 个应该出现的序号；(3) 解码结果与之前一个周期分段的结果相同。

计算拼接好的分段第 1 个分段到最后一个分段的采样点距离。最终，通过采样率得到时间信息。

4 实验结果与分析

本实验选取的音频素材均为 48 k 采样率，16 bit 的双声道 wav 格式，测试素材包括 6 种类型，具体见表 1。每类素材选取 3 个文件，共 18 个文件，文件长度从十几秒到几分钟不等。然后在每类素材中选取 10-20 s 的片段进行不可感知性测试，所有的文件都参与鲁棒性测试。本实验嵌入的信息码为 32 bit 的信息。听感测试过程严格执行国际标准。

表 1 音质评价结果

素材种类	SDG	ODG
流行音乐(有人声)	0.13	-0.35
流行音乐(无人声)	0.36	-0.67
交响乐	-0.71	-0.52
语音	0.04	-0.48
中国民乐	-0.50	-0.70
歌剧	-0.36	-0.60

4.1 不可感知性分析

测试同时采用主观评价和客观评价的方法，主观评价标准测试标准采用国际电联的 ITU-R BS.1116^[13] 建议书对嵌入数字水印的音频的音质进行分析。该标准主要适用于在音质损伤比较小的情况下进行主观评价测试。测试结果称为 SDG (Subjective Difference Grade)，打分采用 ITU-R BS.562^[14] 的五分制标准，0 分表示音质损伤小，而 -4 分表示损伤大。客观测试采用 PEAQ (Perceived Audio Quality) 算法^[15] 进行测试，该算法也为国际标准。算法的结果为 ODG (Objective Difference

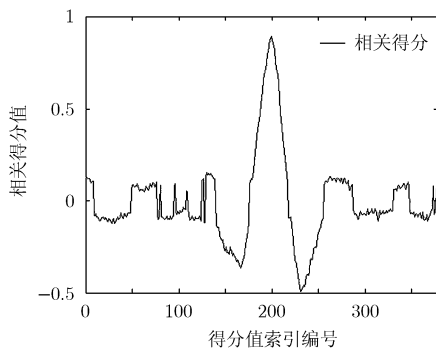


图 5 相关得分出现的峰值

Grade), 该分数对音质的描述与 SDG 相同。主观测试的被试者均为具有专业音乐工作背景的人员, 包括音频工程技术人员、录音师等, 共计 24 人。统计结果如表 1 所示。

根据主观测试结果, 在嵌入水印以后, 3 种素材(流行音乐带人声、流行音乐无人声、语音)为“损伤不可察觉”, 另 3 种素材(交响乐、中国民乐、歌剧)属于“损伤可察觉、但不讨厌”。而所有的客观测试结果均属于“损伤可察觉、但不讨厌”。测试结果表明素材在嵌入水印后音质的损伤极小。

4.2 鲁棒性分析

为测试水印算法的鲁棒性, 我们对嵌入水印的素材进行了多种处理, 其中多数是在广播电视与节目制作过程中常见的操作。测试包括以下几种: MPEG 压缩(连续两次 256 kbps 转换)、幅值调整、随机剪切、随机拼接、随机插入无水印的素材、以及对素材连续进行 MPEG 压缩、幅值调整、随机剪切的混合操作。

在对素材进行处理以后, 然后进行水印解码, 解码结果为水印信息码和该信息码的时长信息, 精度按秒计。统计含有正确解码信息的时长, 经过汇总后除以参与测试素材的全部时长, 得到解码正确率 P_{dec} 。在进行解码测试时, 同时测试了没有采用 2.3 节中的相关特征进行解码的准确率。测试结果见表 2。

表 2 鲁棒性测试结果(%)

操作类型	不使用相关特征	使用相关特征
MPEG 压缩	95.1	95.7
幅值调整(+20%)	99.6	99.6
幅值调整(-20%)	99.6	99.6
随机剪切	95.5	95.9
随机拼接	94.5	99.4
随机插入	92.0	98.4
混合操作	96.4	97.1

测试过程中虽然有一些水印在操作后丢失, 但在可以接受的范围之内。另外, 与没有相关特征的结果相比, 算法的性能有显著提高。测试结果说明, 算法对各种常见的音频编辑处理具有一定的鲁棒性, 并且相关特征的应用提高了算法的性能。

5 结束语

本文针对媒体资料和版权管理应用提出了一种基于心理声学模型的扩频水印算法。算法利用扩频

的原理嵌入水印信息, 利用音频信号的短时平稳特征, 减少载体音频对水印的影响, 同时使用 MPEG 心理声学模型来确保嵌入信息的不可感知性。通过合理的设计水印信息帧结构和多扩频码的应用, 提高了解码结果时长统计的时间精度到秒级。最后, 还通过同步相关时的一些特征使得同步位置的确定更为精准, 减少误同步的影响。正式的听感测试表明, 算法具有很高的不可感知性, 同时进行的鲁棒性测试又验证了水印算法能有效抵抗随机剪切、拼接、音频压缩等音频编辑过程中的常见操作。

参考文献

- [1] Lei B and Soon I Y. Audio watermarking for DRM based on chaotic map[C]. Proceedings of Conference on Computer Engineering and Technology, Singapore, 2009, Vol. 1: 55-59.
- [2] Xu T and Yang Z. Blind audio watermarking scheme for copyright protection of multi-user[C]. Proceedings of Conference on Multimedia Information Networking and Security, Wuhan, 2009, Vol. 2: 159-162.
- [3] Cox I, Miller M L, and Bloom J A. Digital watermarking[M]. Massachusetts, USA, Morgan Kaufman Publishers, 2002: 36-46.
- [4] Yamamoto K and Iwakiri M. Robust audio watermarking with time and frequency division[C]. Proceedings of Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kyoto, 2009: 76-81.
- [5] Imabepu K, Hamada D, and Unoki M. Embedding limitations with audio-watermarking method based on cochlear-delay characteristics[C]. Proceedings of Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kyoto, 2009: 82-85.
- [6] Wei F S and Qi D. Audio watermarking of stereo signals based on echo-hiding method[C]. Proceedings of Conference on Information Communications and Signal Processing, Macau, 2009: 1-4.
- [7] Cox I, Kilian J, Leighton T, et al. Secure spread spectrum watermarking for multimedia[J]. *IEEE Transactions on Image Processing*, 1997, 6(12): 1673-1687.
- [8] Kirovski D and Malvar H S. Spread-spectrum watermarking of audio signals[J]. *IEEE Transaction on Signal Processing*, 2003, 51(4): 1020-1033.
- [9] Wang R and Xiong Y. A novel audio aggregation watermarking for copyright protection[C]. Proceedings of Conference on Intelligent Systems and Knowledge Engineering, Hangzhou, 2010: 166-169.
- [10] Li L and Fang X. New detection scheme for spread spectrum audio watermarking[C]. Proceedings of Conference on Wireless Communications, Networking and Information Security, Beijing, 2010: 330-333.

- [11] ISO/IEC 11172-3. Coding of moving pictures and associated audio for digital storage up to about 1.5 Mbits/s[S]. International Organization for Standardization, Geneva, Switzerland, 1993.
- [12] Gold R. Optimal binary sequences for spread spectrum multiplexing[J]. *IEEE Transactions on Information Theory*, 1967, 13(4): 619–621.
- [13] ITU-R BS.1116. Method for the subjective assessment of small impairments in audio systems including multichannel sound systems[S]. International Telecommunication Union, Geneva, Switzerland, 1994.
- [14] ITU-R BS.562. Subjective assessment of sound quality[S]. International Telecommunication Union, Geneva, Switzerland, 1990.
- [15] ITU-R BS.1387. Method for objective measurements of perceived audio quality (PEAQ)[S]. International Telecommunication Union, Geneva, Switzerland, 1998.
- 赵学敏：男，1984年生，博士生，研究方向为数字音频水印。
- 郭宇弘：男，1985年生，博士生，研究方向为数字音频水印。
- 邹学强：男，1978年生，工程师，研究方向为网络、信息安全。
- 刘建：男，1971年生，副研究员，研究领域包括嵌入式语音识别与合成、口语语言学习算法、数字音频水印技术等。
- 颜永红：男，1967年生，研究员，博士生导师，研究方向为语音识别等。