

中图分类号：TN253

单位代号：10280

密 级：公开

学 号：22721244

上海大学



硕士学位论文

SHANGHAI UNIVERSITY

MASTER'S DISSERTATION

题
目

抵御去同步攻击
的鲁棒音频水印技术研究

作 者：

朱伟南

学科专业：

通信与信息系统

导 师：

吴汉舟

完成日期：

2025 年 5 月

姓 名：朱伟南

学号：22721244

论文题目：抵御去同步攻击的鲁棒音频水印技术研究

上海大学

本论文经答辩委员会全体委员审查，
确认符合上海大学硕士学位论文质量要求。

答辩委员会签名：

主 席：陈延利 贵州师范大学

委 员：王子驰 上海大学
熊玲 北京语言大学

导 师：吴俊

答辩日期：2025 年 10 月 16 日


姓 名：朱伟南

学号：22721244

论文题目：抵御去同步攻击的鲁棒音频水印技术研究

上海大学学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师指导下，独立进行研究工作所取得的成果。除了文中特别加以标注和致谢的内容外，论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他研究者对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。


学位论文作者签名：

日 期：2025 年 10 月 11 日

上海大学学位论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

（保密论文在解密后应遵守此规定）

学位论文作者签名：

导师签名：

日 期：2025 年 10 月 11 日

日 期：2025 年 10 月 11 日

上海大学工学硕士学位论文

抵御去同步攻击的
鲁棒音频水印技术研究

作者：朱伟南

学科专业：通信与信息系统

导师：吴汉舟

上海大学通信与信息工程学院

2025 年 09 月

A Dissertation Submitted to Shanghai University for the Degree
of Master in Engineering

Research on Robust Audio Watermarking Technology Against Desynchronization Attacks

Candidate: Weinan Zhu

Major: Communication and
Information System

Supervisor: Hanzhou Wu

School of Communication and Information Engineering

Shanghai University

September, 2025

摘 要

随着多媒体和网络技术的快速发展，音频数据的复制与传播变得极为便捷。然而，这一便利性也为音频的版权鉴定和内容认证带来了安全挑战。为了应对这些安全挑战，研究人员采用数字水印技术对音频内容进行标识与保护，通过在音频中嵌入水印，可有效地保护音频的版权，并对侵权行为进行追踪与溯源。但在实际应用中，各类攻击可能擦除或破坏水印，削弱其验证与溯源的能力。现有方法针对常见的信号处理攻击具有较强的鲁棒性，然而在面对去同步攻击时仍存在鲁棒性不足的问题。在此背景下，本文围绕音频信号开展了抵御去同步攻击的水印技术研究，取得的主要成果如下：

(1) 针对现有的音频水印算法在去同步攻击下鲁棒性较弱的问题，本文提出了一种基于帧内能量关系调制的鲁棒音频水印算法。首先，将原始音频划分为不相交的帧，并进一步细分为三个不相交的片段，提取中频区域的平均能量作为特征。随后，在每一帧内，采用三角调制策略对能量特征进行调制，以承载水印信息。为了平衡水印的隐蔽性和鲁棒性，设计了优化策略和缓冲补偿机制。在水印提取阶段，通过分析帧内关系实现水印重构。实验结果表明，所提算法在抵抗常规信号处理攻击、时间尺度修改和音高尺度修改方面均表现出较强的鲁棒性，同时能够保持良好的不可感知性。

(2) 为了进一步提升音频水印算法针对去同步攻击的鲁棒性，本文在传统拼凑方法的基础上提出了一种改进策略。该方法将每一帧的音频信号划分为两个不相交的片段，提取中频段的能量作为特征，并通过能量关系调制嵌入水印信息。为了兼顾水印的隐蔽性和鲁棒性，设计了线性递减缓冲补偿机制。针对最具挑战性的裁剪攻击，设计了一种对称的同步码嵌入与检测机制。在嵌入阶段，通过对称且重复地嵌入同步码组，为后续检测提供依据；在检测阶段，通过同步码判断音频是否遭受裁剪攻击，若检测为裁剪攻击，则使用滑动窗口来定位水印区域并通过帧内关系来重构水印；否则，仅依赖两个片段的能量关系来完成水印的提取。实验结果表明，该方案不仅具有良好的不可感知性，而且能够有效抵御去同步攻击（尤其是裁剪攻击）以及各种常规信号处理操作。

关键词：版权鉴定；音频水印；去同步攻击；补偿机制；关系调制

ABSTRACT

With the rapid development of multimedia and network technologies, the replication and dissemination of audio data have become extremely convenient. However, it also poses significant challenges for copyright identification and content authentication. To address these security issues, researchers have adopted digital watermarking techniques to embed identifying information into audio signals, thereby protecting copyright and enabling infringement tracing and source tracking. Nonetheless, in practical applications, various attacks may erase or degrade the watermark, undermining its capacity for verification and traceability. Although many methods are robust to common signal processing, they remain vulnerable to desynchronization attacks. Against this backdrop, this thesis focuses on robust audio watermarking techniques designed to resist desynchronization attacks, and the main contributions are as follows:

(1) To address the insufficient robustness of existing audio watermarking algorithms under desynchronization attacks, a robust audio watermarking algorithm based on intra-frame energy relationship modulation is proposed. The original audio is first segmented into non-overlapping frames, each further divided into three non-overlapping sub-segments. The average energy in the mid-frequency band is extracted as a feature. Subsequently, a triangular modulation strategy is applied to modulate the energy features within each frame to embed watermark information. To balance imperceptibility and robustness, an optimization strategy and a buffer compensation mechanism are introduced. During the extraction phase, the watermark is reconstructed by analyzing the intra-frame relationships. Experimental results demonstrate that the proposed algorithm offers high robustness against common signal processing attacks, time-scale modification and pitch-scale modification, while maintaining excellent imperceptibility.

(2) To further enhance robustness against desynchronization attacks, an improved strategy based on the conventional patchwork method is presented. In this method, the audio signal in each frame is divided into two non-overlapping segments, and the mid-frequency energy is extracted as a feature. The watermark is embedded by modulating the energy relationship between these segments. A linearly decreasing buffer

compensation mechanism is designed to ensure a balance between imperceptibility and robustness. To tackle the highly challenging cropping attack, a symmetric synchronization-code embedding and detection mechanism is developed. In the embedding phase, synchronization code groups are embedded symmetrically and redundantly to provide a basis for subsequent detection. In the detection phase, these synchronization codes are used to determine if a cropping attack has occurred. If cropping is detected, a sliding window locates the watermark region for reconstruction via intra-frame relations; otherwise, extraction relies on the two-segment energy relation alone. Experimental results show that the proposed scheme not only ensures high imperceptibility but also effectively resists desynchronization attacks (especially cropping) and various conventional signal processing operations.

Keywords: copyright identification; audio watermarking; desynchronization attack; buffer compensation; relation modulation

目 录

摘 要.....	III
ABSTRACT.....	IV
第一章 绪论.....	1
1.1 音频水印技术研究的背景与意义	1
1.2 音频水印技术的国内外研究现状	2
1.3 论文的主要研究内容	7
1.4 论文组织结构	7
1.5 本章小结	8
第二章 音频水印基础	9
2.1 音频水印技术的框架	9
2.2 音频水印的分类与应用	10
2.2.1 音频水印的分类.....	10
2.2.2 音频水印的应用.....	12
2.3 音频水印算法的相关技术	12
2.3.1 混沌加密.....	12
2.3.2 离散余弦变换.....	13
2.3.3 拼凑算法.....	14
2.4 攻击类型	14
2.4.1 常见的信号处理攻击.....	14
2.4.2 去同步攻击.....	16
2.5 音频水印的评价指标	18
2.6 本章小结	19
第三章 抵抗 TSM 与 PSM 攻击的三角能量调制音频水印算法	20
3.1 引言	20
3.2 基于三角能量调制的水印嵌入方案	21
3.2.1 特征提取.....	21
3.2.2 水印嵌入和优化策略.....	22
3.2.3 水印提取策略.....	25
3.2.4 参数优化分析.....	26
3.2.5 鲁棒性分析.....	28
3.3 实验结果与分析	30
3.3.1 实验设置.....	31
3.3.2 水印隐蔽性测试.....	31
3.3.3 水印鲁棒性测试.....	32
3.3.4 消融实验与分析.....	34
3.4 本章小结	35
第四章 抵抗去同步攻击的双段式能量调制音频水印算法	36
4.1 引言	36
4.2 基于双段式能量调制的水印嵌入方案	37
4.2.1 特征提取.....	37

4.2.2 音频水印的嵌入.....	38
4.2.3 音频水印的提取.....	41
4.2.4 参数优化分析.....	41
4.2.5 鲁棒性分析.....	43
4.3 实验结果与分析	45
4.3.1 水印隐蔽性测试.....	45
4.3.2 水印鲁棒性测试.....	46
4.3.3 消融实验.....	49
4.4 本章小结	50
第五章 总结与展望	51
5.1 全文总结	51
5.2 未来研究展望	52
参考文献.....	53
攻读硕士学位期间取得的研究成果	59
致 谢.....	60

第一章 绪论

随着数字与网络技术的普及，音频内容的获取与传播极为便捷，但由此引发的盗版复制与非法传播愈发突出，严重侵害创作者与平台权益。学术界与工业界不断推进数字水印等内容保护技术的研究与应用，在不影响感知质量的前提下提升音频的安全性与可追溯性。

1.1 音频水印技术研究的背景与意义

早期的多媒体内容多以磁带、光盘等模拟格式存储，由于复制依赖专用翻录设备且音质容易受损，技术上的限制遏制了盗版的传播。然而，随着数字化进程的加快与网络技术的广泛普及，音频等多媒体内容以数字的形式呈现，音频的复制与传播不仅更为高效，而且几乎无任何质量损耗，极大助长了盗版的隐蔽性与扩散速度。与此同时，数字音频文件普遍缺乏可见的版权标识，版权信息难以随文件附带传播，这使得版权所有者在维权过程中面临诸多挑战。

音频作为数字时代信息交互的重要媒介，凭借其轻量化存储、便于传输、易于复制等特点，已成为跨平台内容传播的核心载体^[1]。然而，技术便利性也间接导致了侵权的风险。例如，使用音频编辑软件（如 CoolEdit、Adobe Audition 等）工具对音频进行时域裁剪、频域篡改或深度伪造的操作，易造成原创内容的完整性受损，并引发版权归属模糊的问题。因此，有意识地对音频媒体采取安全防护措施，显得尤为重要^[2,3]。

传统的数字音频版权保护策略，如加密传输与序列号认证，主要立足于访问控制机制。然而，此类机制在现实中暴露了其内在的局限性：尽管它们能有效阻止未经授权的访问并限定合法用户的范围，但其效力仅限于内容解密之前。一旦音频文件被授权用户合法播放，其数字信号流依然可以通过内部录制（如截取声卡输出）或模拟转换等方式被再次捕获与复制。此过程导致版权保护链条在消费终端发生断裂，使得初始的访问控制措施最终失效^[4,5]。针对这一问题，研究者逐步将目光转向音频内容本身，提出将版权信息以不可感知的形式嵌入到音频信号

中，使其在复制、传播甚至格式转换的过程中仍随内容一同存在^[6]。这种以内容作为载体的版权保护方式即为数字音频水印技术，其核心思想是通过对音频信号进行微小调整，嵌入具有识别性和追踪性的数据信息^[7]。近年来，该技术已被广泛应用于广播电视、流媒体服务及光盘防拷贝等领域，部分方案甚至被纳入国际标准体系，如 ATSC 3.0 标准和蓝光影碟的 Cinavia 系统^[8]。音频水印技术的发展不仅提升了数字音频版权保护的有效性，也为后续的司法维权和版权溯源提供了强有力的技术支撑。

在数字音频的传播过程中，去同步攻击是一类常见且具有破坏性的干扰方式，主要通过对音频信号进行剪切、插入、时间尺度修改（Time-Scale Modification, TSM）或音调尺度修改（Pitch-Scale Modification, PSM）等操作，改变其时间轴或频率轴的结构。这类操作会破坏水印嵌入时的参考位置，导致水印无法正确地同步和提取，从而严重影响音频水印算法的鲁棒性^[9]。此外，由于人耳系统对音频信号的变化比较敏感，水印设计需兼顾不可感知性与嵌入强度，这进一步压缩了抵御去同步攻击的优化空间。鉴于现有算法在应对多样化时频扰动时普遍表现出鲁棒性不足的问题，研究能有效抵御去同步攻击的音频水印技术已成为该领域的关键挑战。本文的研究工作正聚焦于此，旨在为现有方法提供新的设计思路与技术支持。

1.2 音频水印技术的国内外研究现状

数字音频水印技术可分为时域水印和变换域水印两大类。时域方法直接对音频采样点进行修改进而嵌入水印信息，而变换域方法先对音频做频域变换，再在特定系数上嵌入水印，以提高水印的鲁棒性。下面将基于音频水印技术的嵌入域划分，系统梳理并阐述不同技术的迭代演进历程。

早期音频水印技术以时域方法为主，其中最低有效位法（Least Significant Bit, LSB）最为基础。它通过将水印嵌入到音频采样值的最低比特位，利用人耳对细微信号变化的不敏感特性，在保证音质的前提下实现版权信息的隐蔽嵌入^[10,11]。该方法实现容易且容量高，但极易被攻击所移除。例如，只要将音频采样最低位

清零或加入微弱的噪声，LSB 水印即被破坏。因此该方法的鲁棒性很差，仅适合不可见性要求高但安全性要求低的场景。

回声隐藏法是由 Gruhl 和 Bender 率先提出，作为音频水印技术的重要分支，其核心原理是在原始音频中嵌入极短延迟的回声信号来承载水印比特。通过精准调控回声的强度、延迟时长与衰减参数，使回声干扰控制在人耳听觉阈值之下，既能隐蔽传递版权信息，又不影响音频的原始听觉效果^[12]。由于回声叠加在音频中，水印不易被简单移除，能较好抵抗常规信号处理。为了增强鲁棒性，Kim 等人^[13]提出改进的前向/后向双回声内核方法。随后 Ko 等人^[14]将传统的单回声扩展为时扩回声，采用伪随机序列控制的一串微弱回声替代单个回声，以提升水印在面对噪声、滤波、压缩等常见攻击时的鲁棒性与安全性。Xiang 等人^[15]利用优化的伪噪声序列和改进的解码函数，实现了更高的不可感知性和稳健性。总体而言，回声法嵌入简单，对常规处理有一定抵抗力，但受到去同步攻击时仍容易失效，因为回声位置被扰乱后提取较为困难。

直接序列扩频法同样可在时域中实现。其基本原理是将水印比特调制至伪随机序列上，再将调制后的序列叠加至原始音频信号中，从而完成水印信息的嵌入^[16]。典型工作如 Boney 等人^[6]在 1996 年率先将扩频水印应用于音频领域，通过结合人耳听觉掩蔽模型对水印序列进行能量调整后嵌入音频，实现了数字音频的版权标记。该方法通过将水印信息扩展至大量的采样点中，使攻击者难以在不破坏音质的前提下完全去除水印信息。相对于单点嵌入，扩频法的鲁棒性更好，检测则是采用相关运算来提取水印。然而，该方法仍存在一定局限性。一方面，受限于嵌入强度与掩蔽阈值，其嵌入容量较小；另一方面，水印提取过程依赖于序列的精确同步，若音频信号经历整体时轴拉伸、压缩或局部裁剪等攻击，容易导致水印失去同步，从而影响提取的可靠性^[17]。

时域直方图水印是一种基于音频信号时域统计特征的嵌入方法。该方法通过调整音频幅度直方图的形状或特定区间的采样分布来嵌入水印，比传统时域技术具有更强的鲁棒性，尤其在抵御去同步攻击（如 TSM、PSM、裁剪）方面表现出色。Xiang 等人^[18]提出了一种基于直方图形状与均值特征的音频水印算法，该方法利用时域直方图中相邻三个桶采样数的相对关系嵌入多比特水印，同时结合帧均值的稳定性进行帧定位，从而实现盲提取。在面对 TSM、裁剪等典型去同步攻

击时，音频直方图的统计结构表现出较强的不变性。接着，Xiang 等人^[19]提出了一种基于统计特征修改的可逆水印算法。该方法在每一帧中，以三个样本为一组生成预测误差，并将帧内所有预测误差之和计算为统计特征值。通过对这些统计特征值的直方图进行移位操作，将水印比特嵌入到音频帧中，并利用帧内最小值的稳定性附加辅助信息，用于支持水印提取后的音频恢复。该算法不仅支持盲提取，而且在保证水印鲁棒性的同时实现了原始音频的可逆恢复。在面对加噪、裁剪、重采样等典型攻击时，该方法依然保持了较低的误码率和较高的音频质量。尽管上述方法在鲁棒性与可逆性方面取得了良好的效果，但仍存在一定局限性。首先，该方法在设计上依赖于直方图特定幅度区间内的采样分布，因此在嵌入容量较高时，可能导致每个直方图箱中的采样数量不足，进而削弱统计结构的稳定性，影响整体鲁棒性。

总的来说，时域水印方法实现简单、嵌入容量大，但整体鲁棒性较差。在对抗裁剪、TSM、PSM 这类去同步攻击时更为困难，往往需要附加同步措施。隐蔽性方面，回声在适当参数下人耳难以察觉，而 LSB 和扩频法若功率控制不当容易造成音质劣化。相较而言，直方图水印虽在抵御去同步攻击方面具有一定优势，但其性能高度依赖于统计特征的稳定性，且在嵌入容量增加时易受到限制。随着嵌入信息增多，直方图桶内采样数量减少，鲁棒性随之下降。由于上述局限，研究重心很快转向变换域以提高鲁棒性。

变换域方法自 90 年代末兴起，显著提升了水印对抗压缩、滤波等攻击的能力。典型频域包括离散余弦变换（Discrete Cosine Transform, DCT）^[20-22]、离散小波变换（Discrete Wavelet Transform, DWT）^[23-25]、离散傅里叶变换（Discrete Fourier Transform, DFT）^[26,27]等。通过修改特定范围内的频域系数，可将水印嵌入到音频之中，使其更难以移除。以下按主要嵌入技术分类介绍：

相位编码法利用人耳对绝对相位不敏感的特性，将水印嵌入音频片段的相位信息中。1996 年，Bender 等人^[28]首先提出了相位水印方案。该方案先将音频划分为短帧，随后对每帧施加 DFT 变换，把水印信息嵌入到各帧的绝对相位中，在此过程中保留相位差，如此一来，水印嵌入操作对音质的影响极小。尽管相位编码具备优异的不可感知性，但其鲁棒性较弱。很多音频处理操作，如有损压缩、

裁剪、回声攻击都会改变信号相位，导致水印难以检出。此外相位法需严格的帧同步，一旦音频时轴发生变化，相位水印的对齐和解码将难以实现。

扩频法在频域的应用是将水印序列分散添加到频谱的多个系数或整个频带上，以实现信息的隐蔽嵌入。Kirovski 和 Malvar^[16]率先将扩频技术应用于音频水印的频域嵌入，利用复数调制重叠变换转换到频域。该方法结合心理声学频率掩蔽、协方差测试修正、块重复编码及倒谱滤波等技术，显著提升了水印检测的性能。在此基础上，该方案在对抗 TSM、PSM、噪声干扰等攻击时表现出较强鲁棒性，同时通过棋盘水印与动态块检测保证水印的不可感知性。然而，扩频法一般属于低容量水印。并且为兼顾音质，水印信号幅度必须很小，否则会导致感知上的音频失真。随后，部分研究者引入启发式算法优化扩频水印的嵌入强度，在满足信噪比约束的前提下，他们致力于在听觉感知性和鲁棒性之间找到平衡，让水印既难以被人耳察觉，又能有效抵御常见的信号处理攻击。例如 Su 等人^[29]提出了一种在信噪比约束下优化扩频因子的算法，并进一步采用多目标遗传算法权衡鲁棒性与感知性。总体而言，频域扩频算法在抗常规攻击（如加噪、重采样、滤波）方面表现优秀。但和时域方法类似，如果未经特殊处理，频域扩频水印对去同步攻击（如裁剪攻击、抖动）仍比较脆弱。

量化索引调制法（Quantization Index Modulation, QIM）是 Chen 和 Wornell^[30]在 2001 年提出的一类数字水印嵌入策略。QIM 在音频中的典型做法是选取若干变换域系数，将这些系数量化到不同间隔，以此分别代表比特 0 或 1，进而实现水印的嵌入^[31-33]。由于对每个系数进行的量化扰动都经过精心设计，所以相应产生的听觉失真能够被控制在阈值范围内，使得水印具有良好的不可感知性。QIM 水印鲁棒性也较高，轻微的噪声或压缩不会将系数拉离其量化区间外，提取时通过检查系数所在区间即可还原水印比特。近年来，基于 QIM 的音频水印方法受到了广泛关注。Zhao 等人^[34]结合心理声学模型动态调整量化步长，在保证高音质的同时增强了系统的鲁棒性；此外，Wu 等人^[35]通过引入冗余结构实现了 QIM 的自同步提取机制，有效抵御裁剪与 TSM 攻击；Jiang 等人^[36]基于全局帧特征设计了自适应嵌入强度策略，进一步提升了系统对去同步攻击的抵抗能力。这些研究从不同角度扩展和优化了 QIM 模型，体现出其良好的鲁棒性与透明性。

拼凑法 (Patchwork) [37,38] 最初是一种统计水印思想, 2003 年 Yeo 和 Kim [38] 将其改进用于音频中。拼凑算法随机选择两组频域系数, 略微提升一组系数值而降低另一组, 从而嵌入一个二值水印。接收端通过统计两组系数差异来判断水印比特。拼凑法的特点是不需要原始音频的参与且对全局失真有一定抵抗能力, 这是由于多数攻击方法作用于音频时, 对相邻两段系数的统计特征影响较为相似, 能够保留水印差值, 确保水印信息的有效检测。近年来拼凑法被用于抵御去同步攻击的音频水印研究。Xiang 等人 [39] 提出一种改进的拼凑法, 在 DCT 域中嵌入水印, 并于对数 DCT 域插入同步码, 用以估计攻击导致的缩放因子并进行重采样校正, 有效抵御 TSM、PSM 及抖动攻击。Natgunanathan 等人 [40] 进一步提出多层的拼凑水印嵌入结构, 通过有序排列 DCT 系数及引入自适应误差缓冲区, 实现多层水印的独立嵌入与提取, 兼顾了感知质量与鲁棒性。此外, Zhao 等人 [41] 提出基于相邻频段奇异值比的拼凑式水印嵌入方法, 进一步增强了系统在高嵌入速率下抵御去同步攻击的能力。综上所述, 这些研究表明拼凑法具备良好的可扩展性和适应性, 在应对多种类型的去同步攻击时表现出强大的实用价值。

为提升对去同步攻击的鲁棒性, Li 等人 [42] 将水印嵌入 DWT 变换后的低频分量直方图中, 并将鲁棒水印嵌入引起的失真作为补偿信息进行可逆隐藏, 以实现原始音频文件的恢复。该方法在 TSM、PSM 及随机裁剪下表现出良好鲁棒性, 且对常规攻击亦具稳定性, 但由于依赖全局统计特征, 嵌入容量较低。针对大规模裁剪攻击, Zhang 等人 [43] 提出一种结合 m 序列与滑动窗口机制的水印方案, 在 DWT、图变换、奇异值分解后通过扩频法嵌入加密后的水印, 并借助 m 序列的周期性实现自恢复, 同时摆脱了对同步码的依赖。该方法在高裁剪比例下依然能够提取并重构完整水印, 展现出较强的恢复能力与嵌入容量。然而, 由于其设计主要面向裁剪攻击, 对于 TSM 与 PSM 等去同步攻击的鲁棒性仍有待进一步提升。

尽管现有基于时域和频域的鲁棒音频水印方法在抵御常见信号处理攻击方面取得了一定成效, 但在应对不同类型攻击时仍面临鲁棒性权衡的难题。当前方法普遍存在三方面问题: 一是去同步攻击的抵抗能力不足: 去同步攻击通过改音频的时域或频域结构, 破坏水印的同步性, 导致传统水印方法失效, 特别是在裁剪攻击中, 水印嵌入位置发生偏移, 无法恢复水印信息。二是常规攻击下鲁棒性

有限：现有方法也无法全面应对各种攻击，涵盖从简单的噪声到复杂的时频扰动，导致在不同攻击场景下的鲁棒性存在明显的不足。三是水印容量偏低，难以满足实际需求：为了增强鲁棒性，很多方法不得不在水印容量上做出妥协，导致水印容量较低，难以满足实际应用中的需求。尤其在需要高容量的场景下，现有方法往往难以提供足够的水印容量来保证高效的版权保护。

1.3 论文的主要研究内容

本文主要针对 TSM、PSM 和裁剪攻击进行研究，提出了两种抵御去同步攻击的鲁棒性音频水印算法，其主要研究内容如下：

针对 TSM 和 PSM 等去同步攻击，本文提出了一种基于三角能量调制与缓冲补偿机制的音频水印算法。该算法引入一种高效的缓冲补偿策略，通过减小水印嵌入区域与非嵌入区域之间的系数差异，有效提升了水印在去同步攻击下的鲁棒性。同时，采用最小优化算法进一步增强水印的隐蔽性。在提取阶段，通过分析帧内的相对关系实现水印信息的准确恢复。

为了进一步提升抵御 TSM 和 PSM 等去同步攻击的鲁棒性，本文基于拼凑方法，通过调整一帧内相邻两段的能量来自适应的嵌入水印。使用改进的缓冲补偿机制和优化算法来提升水印的鲁棒性与隐蔽性。在水印提取阶段，比较相邻两段的平均能量来提取水印。针对难以抵抗的裁剪攻击，本文设计了一种基于对称同步码的嵌入与检测机制。在检测阶段，首先利用对称同步码判断是否遭受裁剪攻击；若检测为裁剪攻击，则采用固定长度的滑动窗口进行水印提取；若未检测到裁剪攻击，则通过分析帧内能量关系完成水印提取。

1.4 论文组织结构

本论文是作者在攻读硕士学位期间，围绕音频水印抵御去同步攻击这一研究方向开展的相关工作的成果总结。旨在保障音频水印隐蔽性的同时，提高算法在各种去同步攻击下的鲁棒性。本文结合音频信号的特点，分别针对 TSM、PSM

及裁剪等去同步攻击，提出两种鲁棒音频水印算法，并通过实验对其性能进行验证。本文的各章节安排如下：

第一章为绪论，首先介绍了本课题的研究背景与意义，然后综述了音频水印技术的国内外研究现状，并明确了本文的研究目标与研究内容。最后，对全文的结构进行了说明。

第二章为音频水印的基础部分，介绍了音频信号的基本框架以及音频水印技术的分类与应用领域。同时，详细阐述了音频水印中的关键相关技术，包括混沌理论、DCT 变换、拼凑方法等，接着分析了常见的信号处理攻击与去同步攻击的类型，并介绍了水印系统的评价指标，为后续算法设计提供理论支撑。

第三章提出了一种面向 TSM 和 PSM 攻击的鲁棒音频水印算法，采用基于三角能量调制与缓冲补偿机制的嵌入方式，并通过优化算法提升水印的隐蔽性，显著增强了水印系统在去同步攻击下的鲁棒性。同时，通过设计缓冲补偿机制平滑嵌入区域与非嵌入区域的系数过渡，减少因攻击导致的特征提取误差。随后，通过理论推导与实验验证对算法的有效性进行了评估。

第四章提出了一种针对裁剪等去同步攻击的鲁棒音频水印算法。该算法通过双段式能量调制与对称同步码机制，提高了水印提取在裁剪等复杂攻击下的鲁棒性。并引入线性递减缓冲补偿策略和优化算法，在嵌入区域边缘实现平滑过渡，有效提升听觉质量。实验部分验证了该方法在鲁棒性与隐蔽性方面的综合性能。

第五章对全文的研究内容进行了总结，并对未来音频水印技术在抗攻击方向的研究趋势和可进一步改进的方向进行了展望。

1.5 本章小结

本章为论文的绪论，首先介绍了当前音频内容在数字化传播中的广泛应用以及由此引发的盗版与侵权问题，阐明了研究鲁棒音频水印技术的现实背景与重要意义。随后，对国内外音频水印技术的发展现状进行了综述，重点分析了时域与变换域方法的技术特点与演进趋势，并指出现有方法在去同步攻击下的鲁棒性仍有待提升。在此基础上，明确了本文的研究内容与技术路线。最后，介绍了论文的主要研究内容及成果，并阐述了论文的结构安排。

第二章 音频水印基础

本章将首先简要介绍数字音频水印框架的一般流程，随后对实现音频水印过程中涉及的关键技术进行说明，包括混沌加密、DCT 变换以及拼凑方法等。接着，将对数字音频水印的性能评价指标及常见攻击类型进行系统阐述。

2.1 音频水印技术的框架

音频水印技术是一种通过在音频信号中嵌入特定水印信息，以实现版权保护、身份认证、内容追踪或隐秘通信等目的的信息隐藏技术。其本质是在不影响原始音频质量的前提下，将机密信息嵌入到音频载体中，使其与音频内容紧密融合，并在必要时实现水印信息的检测与提取。为了实现鲁棒性、不可感知性、隐藏容量与安全性的平衡，一个完整的音频水印系统通常包含以下四个关键组成部分：水印预处理、水印嵌入、攻击处理、水印提取等模块，如图 2.1 所示。

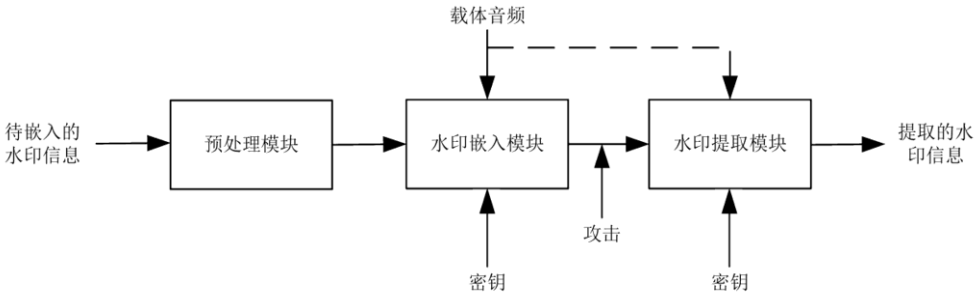


图 2.1 数字音频水印基本框架

Figure 2.1 The basic framework of digital audio watermarking

1) 水印预处理模块是确保水印鲁棒性和安全性的关键环节，主要包括水印编码和特征提取两大核心步骤。水印编码的主要任务是对原始水印信息（如文本、图片、序列码等）进行必要的编码与加密处理，以增强水印的安全性与纠错能力。常见的操作包括混沌加密、纠错编码（如 BCH 码或 RS 码）^[44,45]等。这一步骤不仅保障了水印的抗篡改能力，还为后续的盲提取提供了基础。此外，由于音频信号通常具有非平稳性和冗余性，因此在水印嵌入前，需要通过分帧、窗函数处理或变换（如 DWT、DCT、FFT 等）来提取其稳定的特征区域作为嵌入位置。合理的特征选择不仅能够增强算法的鲁棒性，还能减少对原音频的感知影响。例如，

可选取能量较强的频段、小波系数中低频部分或音频的调制包络特征作为嵌入区域。

2) 水印嵌入模块：水印嵌入是系统的核心环节。根据不同的嵌入策略，可将水印信息嵌入到音频的时域、频域、或混合域中。嵌入方式通常包括 QIM、拼凑方法或扩频技术。此外，为增强同步能力，部分系统还会设计冗余嵌入机制或引入同步码结构。嵌入后生成的音频应满足高透明性，即人耳难以察觉音质变化。

3) 攻击处理模块：嵌入水印的音频文件在通过互联网、社交媒体、音频平台或移动终端传播过程中，易遭受多种外部攻击，主要包括常规信号处理攻击以及去同步攻击。这些操作通常由用户通过音频编辑软件直接完成。为了确保水印在复杂环境下的可提取性与鲁棒性，音频水印系统需具备良好的抗攻击能力。

4) 水印提取模块：水印提取模块负责判断目标音频中是否存在水印信息，并根据嵌入策略执行提取操作。若采用盲水印方式，系统无需原始音频即可完成提取；而若为非盲方式，则需要与原始载体做对比。提取过程包括同步码定位、特征区恢复及纠错解码等步骤。若攻击导致部分水印信息缺失，系统还可利用冗余编码或序列相关性进行部分恢复。

2.2 音频水印的分类与应用

2.2.1 音频水印的分类

1993 年，Tirkel 等学者^[46]首次提出了数字水印的概念，自此引发了国内外研究者的广泛关注，相关技术也迅速发展，涌现出大量研究成果。近年来，音频数字水印技术得到了深入研究，不仅在单声道音频中应用广泛，也逐渐扩展至双声道音频的嵌入与处理^[47]。此外，音频水印技术依据不同的标准可以进行多种分类，如图 2.2 所示，常见的分类方式包括嵌入位置、提取方式以及应用目的等。其中，按照嵌入位置进行划分是较为常见的方法，通常可分为时域嵌入和变换域嵌入两大类。在 1.2 节，已经对这两种嵌入方式展开了详细介绍，为避免重复，本节将不再赘述。

若基于提取方式分类，音频水印技术可大致分为非盲提取、半盲提取和盲提取三类方法。其中，非盲提取是指在提取水印信息时需借助原始音频信号作为参考，通过对原始音频与含水印音频的对比分析来还原嵌入信息。尽管该方法通常具有较高的检测精度和鲁棒性，但由于在实际应用中原始音频往往难以获取，并且其存储和传输开销较大，因此应用范围受到限制。为提高实际可行性，研究者提出了盲提取方案。盲提取方法无需原始音频即可完成水印的提取，具备良好的适应性和实用性，是当前研究和应用中的主流方向。该类方法通常基于音频信号的某些统计特征进行提取，因此在保证鲁棒性的同时，简化了系统结构和使用流程。介于两者之间的半盲提取方法在提取过程中不需要完整的原始音频数据，但需提供原始水印或某些辅助信息。该类方法在保持一定提取精度的基础上，降低了对原始信号的依赖程度，是非盲和盲提取之间的一种折中方案。

根据水印使用目的不同，音频水印技术可划分为鲁棒音频水印、脆弱音频水印以及半脆弱音频水印。其中，鲁棒音频水印强调在各种信号处理操作下仍能够准确提取水印信息，因而广泛应用于版权保护、内容追踪及广播监测等场景，对抗攻击能力是其衡量的核心指标。相比之下，脆弱音频水印主要用于内容完整性验证，其嵌入信息在音频数据发生微小变动时就会被破坏，因此适合于检测音频内容是否遭到篡改。半脆弱音频水印则结合了上述两类方法的优点，在设计时既具备对部分常见信号处理攻击的容忍能力，又能够识别恶意修改行为。综上所述，音频水印技术已形成多维分类体系，涵盖嵌入方式、提取方式与应用目的。其中，盲提取因实用性强而备受关注，鲁棒水印因抗攻击性强而得到广泛应用。

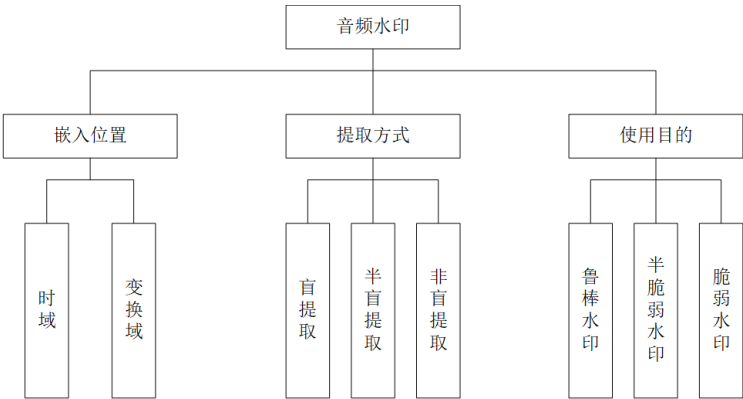


图 2.2 音频水印技术分类
Figure 2.2 Classification of audio watermarking technology

2.2.2 音频水印的应用

音频水印作为数字水印技术的重要分支，目前已在多个实际场景中得到广泛应用，涵盖版权标识、广播监管、内容认证、拷贝控制、信息标注以及隐秘通信等领域^[48,49]。版权保护是其最核心的应用之一，通过将标志着版权信息的水印嵌入音频信号中，实现音频作品的版权归属认定和侵权追溯，对水印的鲁棒性和隐蔽性要求较高。在广播监控中，水印技术可记录音频内容的播出时间与次数，替代传统的人工监听方式，有效提高监测效率^[50,51]。内容认证与篡改检测一般利用水印对音频数据的敏感性来识别异常篡改，在数据安全、司法鉴定等场景中具有重要价值。拷贝控制则结合设备识别机制，对非法录制和复制行为进行限制，从源头降低盗版风险。在信息标注方面，音频水印可嵌入作者信息、使用说明、歌词等辅助数据，提升内容管理的完整性与智能化水平。此外，水印还可用于隐秘通信，将加密信息隐蔽地嵌入音频载体，实现通信内容与行为的双重隐藏，为军事信息传输及网络安全提供了新型手段^[52,53]。总体而言，音频水印技术在保障音频数据安全、追踪内容传播路径方面展现出强大的应用潜力。

2.3 音频水印算法的相关技术

2.3.1 混沌加密

1975 年，美国数学家 James A. Yorke 与华裔学者李天岩^[54] 首次为‘混沌’赋予了严格的数学定义，为这一学科的兴起奠定了里程碑式的基础。自此，学术界对这一现象的探索愈加深入。混沌现象本质上是确定性非线性动力系统在初始条件极度敏感的作用下所展现出的准随机、难以预测的复杂演化。它并非真正的随机无序，而是在确定规则之下呈现出的内在的有序混乱。

为了提升系统安全性，许多音频水印在嵌入前都会先进行加密操作。本节以 Logistic 映射为例介绍混沌加密系统^[55]，其定义如下：

$$x_{k+1} = \mu x_k (1 - x_k) \quad (2.1)$$

其中 k 为迭代时间步, $x_0 \in (0,1)$ 为初始值, μ 为系统控制参数。映射的动态行为与 μ 密切相关: 当 $0 < \mu < 3.57$ 时, 随着 $k \rightarrow \infty$, 序列 $\{x_k\}$ 将会收敛到某个定点或周期轨道, 系统表现出稳定的周期性。当 k 和 μ 同时满足于 $x_0 \in (0,1)$ 并且 $\mu \in (3.57, 4.4]$ 时, 映射进入典型的混沌状态, 此时 $\{x_k\}$ 既不收敛也不重复, 轨迹如同随机噪声般复杂, 对初始条件具有极强的敏感性。即便是极其微小的 x_0 变化也会在后续迭代中被指数级放大, 导致完全不同的演化结果。

以图 2.3(a)所示的一幅 256×256 的原始水印图像为例, 选取 Logistic 映射作为加密手段, 设定初始参数为初值 $x_0 = 0.382$, 系统控制参数 $\mu = 3.99$ 。通过迭代 Logistic 映射生成长度为 65536 的混沌序列, 并对其进行归一化和量化处理后, 作为混沌密钥对水印图像进行像素级加密。加密后的图像如图 2.3(b)所示。从图中可以看出, 加密后的图像呈现出明显的乱码特征, 视觉上难以辨认原始图像内容, 说明 Logistic 映射生成的混沌序列有效地扰乱了水印图像的像素分布, 从而提升了水印在传输过程中的安全性。



图 2.3 加密前后的水印图像示意图

Figure 2.3 Schematic diagram of the watermarked image before and after encryption

2.3.2 离散余弦变换

在数字水印技术中, DCT 常作为重要的嵌入域, 广泛应用于图像和音频水印的嵌入与提取过程中, 其优势在于良好的能量集中性和频域稳定性。DCT 水印的基本思想是将信号从时域转换到频域, 使得信号能量主要集中在少数低频系数上, 修改特定范围内的系数, 从而在保证感知质量的前提下便于信息嵌入。以一维 DCT 为例, 其变换形式可表示为:

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) k \right], \quad k = 0, 1, \dots, N-1 \quad (2.2)$$

其中, $x(n)$ 表示长度为 N 的输入信号序列, $X(k)$ 为变换后的 DCT 系数, n 和 k 分别表示时域与频域中的离散索引。通过选择适当的频域系数进行调制, 可在增强水印鲁棒性的同时保持较高的感知透明性。

2.3.3 拼凑算法

拼凑方法是一种典型的基于统计特征调制的盲水印算法, 该方法具有良好的鲁棒性, 能够有效抵抗加噪声、滤波、压缩、重新量化和重采样等常见攻击。假设以均值为特征为例, 其基本思想是在宿主信号中随机选取两组样本集合 $A = \{a_1, a_2, \dots, a_M\}$ 和 $B = \{b_1, b_2, \dots, b_M\}$, 每组包含 M 个样本。通过对其均值进行微小调整嵌入水印。具体地, 若嵌入水印比特为 0, 则执行:

$$a'_i = a_i + \delta, \quad b'_i = b_i - \delta \quad \forall i \in \{1, \dots, M\} \quad (2.3)$$

若嵌入水印比特为 1, 则执行:

$$a'_i = a_i - \delta, \quad b'_i = b_i + \delta \quad \forall i \in \{1, \dots, M\} \quad (2.4)$$

检测时, 计算统计量:

$$T = \sum_{i=1}^M (a_i - b_i) \quad (2.5)$$

若 T 大于零, 则判定水印为 0; 否则判定为 1。该方法无需原始信号即可检测水印, 适用于图像、音频等多种场景。

2.4 攻击类型

在实际应用环境中, 数字音频水印在传输或存储过程中不可避免地会受到各类信号处理操作或攻击, 这些攻击可能导致水印信息提取的准确性受到严重影响, 甚至完全丢失。因此, 对水印算法抵抗常见信号处理攻击的鲁棒性分析至关重要。本节将对常见的信号处理攻击和去同步攻击进行详细介绍, 以便于后续实验中鲁棒性的测试与分析。

2.4.1 常见的信号处理攻击

常见信号处理攻击旨在破坏水印的完整性或干扰其提取, 以降低水印技术的

有效性。以下先简单概述常见类型：

(1) 无攻击

无攻击指水印嵌入后的音频信号在不受任何外界干扰的情况下直接进行水印提取，以此作为水印算法在理想环境中的基准性能。

(2) 噪声攻击

噪声攻击指向嵌入水印的音频信号中叠加一定量的随机生成的噪声，一般为加性高斯白噪声（Additive White Gaussian Noise, AWGN）是通信领域中一种常见的噪声模型，用来模拟实际传输中可能出现的环境干扰。

(3) 量化攻击

量化攻击是指对已嵌入水印的音频信号进行重新量化，通常表现为降低信号的量化位数，如从 16 位降至 8 位。该攻击可能导致嵌入的水印信息因量化位数的影响从而受到严重损坏。

(4) 幅度缩放攻击

幅度缩放攻击指在水印提取前对含水印音频信号进行幅度的线性缩放，例如将音频信号的振幅乘以一个小于或大于 1 的系数，从而模拟音频在传输或播放过程中音量的变化。

(5) 回声攻击

回声攻击是指在音频信号中加入具有一定时间延迟和特定幅度的回声信号，从而影响音频的时域和频域特性，干扰水印信息的提取。

(6) MP3 压缩攻击

MP3 压缩攻击指采用 MPEG-1 Layer III 标准对嵌入水印后的音频信号进行压缩处理，以测试水印对有损压缩的鲁棒性。该攻击在实际音频传输、存储和网络共享中极为常见。

(7) AAC 压缩攻击

AAC 压缩攻击指使用 MPEG-4 高级音频编码算法对含水印音频进行压缩。这种攻击方式具有更高的压缩效率，广泛应用于现代音频流媒体环境中。

(8) 重采样攻击

重采样攻击是指在水印提取前对音频信号进行采样率的降低再提升，使音频信号经历频域的变化。会影响水印信息的正确提取。

（9）低通滤波攻击

低通滤波攻击指对水印音频信号施加低通滤波器，从而去除信号中的高频成分，模拟音频信号处理或传输中的频率衰减现象。

（10）高通滤波攻击

高通滤波攻击指对含水印音频信号施加高通滤波器，去除信号中的低频部分，模拟实际环境中音频处理设备或传输媒介造成的低频丢失。

上述常见攻击方式涵盖了实际音频处理环境中可能出现的主要干扰因素，为后续水印算法的鲁棒性实验和评估提供了可靠的参考依据。

2.4.2 去同步攻击

与普通信号处理攻击相比，去同步攻击破坏了宿主音频数据的同步结构，导致水印的嵌入位置发生偏移，严重影响水印提取的准确性。常见的去同步攻击包括抖动攻击、TSM、PSM 和裁剪攻击。其中，抖动攻击是指在音频信号中以固定的间隔均匀地插入或删除少量样本数据，对于基于平均分段的水印方案，这种攻击的影响有限，因此本节将重点分析 TSM 攻击、PSM 攻击以及裁剪攻击对音频水印的影响。



图 2.4 TSM 和 PSM 之间的转换关系

Figure 2.4 The conversion relationship between TSM and PSM

音频变速变调在不同的场景中可以分为三种形式：变速不变调、变调不变速以及变调又变速。音频变速指的是在时域上拉长或缩短语音信号的时长，而语音的采样率、基频及共振峰不发生变化，即 TSM。语音变调指的是改变语音的基频，使其升高或降低，并相应地调整共振峰，但采样频率保持不变，即 PSM。根据 Lin 等人^[56]的研究，TSM 和 PSM 之间具有双重性。具体而言，可以通过首先修改时间尺度，然后调整采样速率来实现基频的缩放；另一方面，也可以通过首先修改音高尺度，再调整采样率来实现时间的缩放，如图 2.4 所示。

根据 Zhao 等人^[41]所述, TSM 和 PSM 对音频信号的影响可以视为频率轴的拉伸或收缩操作, 其中时间和音高的缩放攻击可以简洁地表示为以下公式。

$$f' = \sigma \cdot f \quad (2.6)$$

其中 σ 为缩放因子, f 为信号的频率分量。由于时间缩放和音高缩放效应会影响水印提取的区域, 这种频率缩放在水印提取过程中引入了去同步攻击, 导致水印无法正确提取。通常, 通过同步码来定位水印位置, 但当同步码无法完整提取时, 水印也无法正确提取。

裁剪攻击是一种常见的去同步攻击方式, 它是指攻击者对受版权保护的音频信号随机或有针对性地删除音频信号的片段, 导致水印的嵌入位置发生偏移或丢失, 从而严重干扰水印的准确提取, 如图 2.5 所示。具体而言, 裁剪攻击会直接删除宿主音频的部分时域信息, 造成音频信号的整体长度缩短, 原有嵌入水印的位置和对应的参考点发生变化, 从而导致传统依赖于固定位置或同步码的水印提取方法失效。

现阶段针对裁剪攻击的方法多采用滑动窗口与重复嵌入的策略, 但这类方法普遍存在一些问题。一方面, 滑动窗口的长度与步长难以有效确定: 因为一些攻击会导致帧长的变化, 所以要设计步长可变的滑动窗口。步长过小会显著增加算法的计算复杂度, 而步长过大则可能导致水印信息遗漏。另一方面, 重复嵌入的方法虽然提升了抗裁剪能力, 却在很大程度上降低了水印的嵌入率, 难以达到理想的效果。

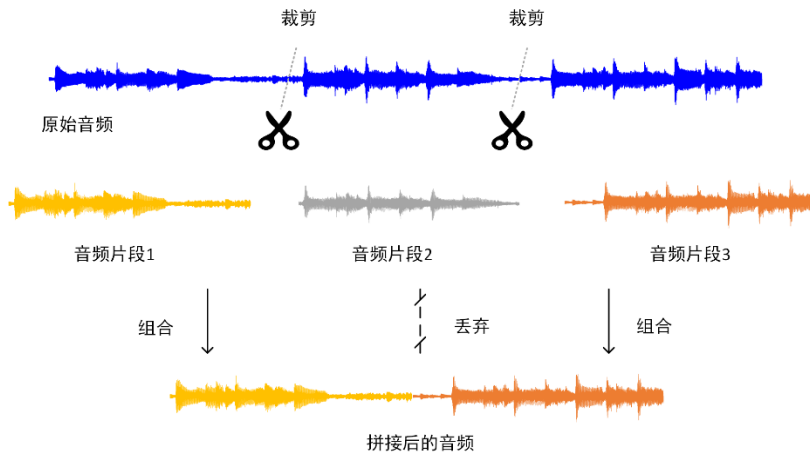


图 2.5 裁剪攻击对音频信号的影响

Figure 2.5 The impact of cropping attacks on audio signals

2.5 音频水印的评价指标

在音频水印领域中，评价嵌入性能通常涉及感知透明性、鲁棒性和嵌入容量三个维度，以下介绍常用的评价指标。

1. 听觉感知质量评价

听觉感知性是衡量音频水印系统性能的重要指标之一，主要反映水印嵌入对原始音频听感质量造成的影响程度。对携带水印音频质量的评估通常包括信噪比（Signal-to-Noise Ratio, SNR）和客观差异等级（Objective Difference Grade, ODG）测量两种方法。

SNR 用于衡量水印嵌入前后音频信号的差异程度，其数值越大代表嵌入水印后音频的失真越小，感知质量越好。计算公式为：

$$\text{SNR} = 10 \log_{10} \frac{\sum_{n=1}^N s(n)^2}{\sum_{n=1}^N [s(n) - s'(n)]^2} \quad (2.7)$$

其中， $s(n)$ 为原始音频信号， $s'(n)$ 为嵌入水印后的音频信号， n 为音频样本长度。

ODG 是一种基于人耳听觉感知模型的客观评价方法，通常的取值范围为-4 到 0，数值越接近 0 表示音频质量越接近原始音频，感知透明性越高。其评估更贴近人类实际听觉感受，因此在音频水印领域中广泛使用。

2. 鲁棒性评价指标

鲁棒性是衡量音频水印系统可靠性的重要性能指标之一，用于评估嵌入水印在经过各种信号处理攻击后仍能被正确提取的能力^[62,63]。常用的鲁棒性评价指标包括误码率（Bit Error Rate, BER）和归一化互相关系数（Normalized Cross-Correlation, NCC）^[57]，其中 BER 用于衡量提取水印与原始水印的一致性，NCC 则反映两者在整体结构上的相关程度。二者从不同角度共同反映了水印系统在各种攻击条件下的抗干扰能力。

BER 衡量水印提取过程中错误的比特占总比特数的比例。BER 越低表示水印恢复的准确性越高，鲁棒性越强。计算公式为：

$$\text{BER} = \frac{E}{T} \quad (2.8)$$

其中, E 为提取过程中错误的比特数, T 为水印的总比特数。NCC 是一种常用于衡量两个信号、图像或水印序列之间相似程度的数学指标。它广泛应用于水印提取精度评估、图像匹配等领域。计算公式为:

$$\text{NCC} = \frac{\sum_{i=1}^N W(i) \cdot \hat{W}(i)}{\sqrt{\sum_{i=1}^N W(i)^2} \cdot \sqrt{\sum_{i=1}^N \hat{W}(i)^2}} \quad (2.9)$$

其中 W 为水印的原始序列, \hat{W} 为提取出的水印序列, N 为序列长度。

3. 嵌入率 (Embedding Rate)

嵌入率用于评价水印嵌入算法的容量, 即单位时间内能够嵌入的水印信息量。通常以比特每秒 (Bits Per Second, bps) 为单位, 嵌入率越高意味着算法能够在保持一定感知质量和鲁棒性的前提下嵌入更多的水印信息, 实用性更强。

2.6 本章小结

本章围绕音频数字水印技术进行展开, 首先介绍了音频水印的技术框架, 随后从嵌入域、提取方式与使用目的三个方面对音频水印方法进行了分类, 并结合实际场景探讨了其典型应用。在此基础上, 系统梳理了水印系统在实际传输与处理过程中可能遭遇的常见信号处理攻击与去同步攻击类型。最后, 重点分析了用于衡量音频水印算法性能的核心评价指标, 包括隐蔽性、鲁棒性与实用性等, 为后续章节中水印算法的设计与性能分析提供了理论支撑与参考依据。

第三章 抵抗 TSM 与 PSM 攻击的三角能量调制音频水印算法

本章节首先提出了一种基于三角调制技术的音频水印方案。然后，从理论角度探讨了该方案在抵抗常见信号处理攻击以及 TSM 和 PSM 攻击方面的鲁棒性。最后的对比实验结果显示，该方法在不可感知性与鲁棒性方面均优于现有先进方法，特别是在应对 TSM 和 PSM 攻击时展现出更强的鲁棒性。

3.1 引言

随着音频内容在各类网络平台上的广泛传播，数字音频的版权保护问题日益凸显。去同步攻击作为一种常见的攻击方式，会通过破坏音频信号的时域或频域结构，严重威胁水印系统的稳健性，尤其是 TSM 与 PSM 攻击，可显著改变音频帧的结构，进而导致水印提取失败。尽管部分现有鲁棒音频水印算法在一定程度上具备抵抗常见信号处理攻击的能力，但在面对上述去同步攻击时仍表现出较高的误码率。因此，如何提升水印系统在 TSM 与 PSM 攻击下的鲁棒性，已成为当前亟需解决的关键问题。

为增强音频水印系统在去同步攻击下的鲁棒性，本文提出了一种基于三角调制策略的鲁棒音频水印算法。该方法将每帧音频划分为三个互不重叠的子段，提取其中频区域的能量特征，并通过构建三角能量关系来调制水印信息，实现水印的嵌入。通过以能量之间的相对关系为条件，提出了最小化失真功率的优化策略，有效提升了水印的隐蔽性。针对 TSM 与 PSM 攻击可能造成的频率线性缩放问题，本章算法设计了缓冲补偿机制，以平滑水印嵌入区域与非嵌入区域之间的能量差异，从而进一步提升系统的鲁棒性与音频感知质量。

实验结果表明，所提算法在保持良好听觉感知质量的同时，展现出显著的鲁棒性优势。与现有主流方法相比，该算法不仅在面对常见信号处理攻击时表现稳定，而且在 TSM 与 PSM 等典型去同步攻击下依然能够有效提取水印，在保障音频质量的同时，实现了水印鲁棒性与不可察觉性的良好平衡。

3.2 基于三角能量调制的水印嵌入方案

本节所提出的基于三角调制和缓冲补偿的音频水印方案整体框图如图 3.1 所示。

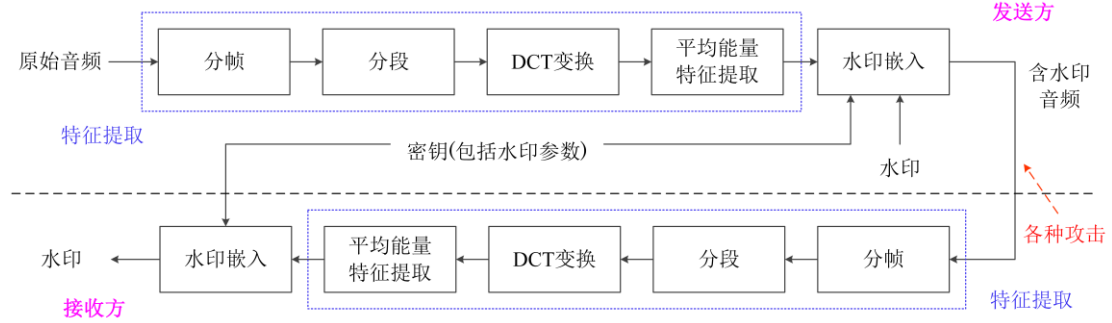


图 3.1 音频水印方案整体框图

Figure 3.1 Overall framework of audio watermarking scheme

在嵌入端，首先根据加密后的水印长度将信号均匀分帧，然后将每一帧进一步划分为三个子段进行处理。接着，分别对每一段应用 DCT 变换，并结合缓冲补偿和最小化失真技术，调制三段的平均能量关系来嵌入水印。最后，针对嵌入水印的系数执行逆离散余弦变换，从而得到包含水印的音频信号。

在水印提取阶段，类似于水印的嵌入阶段，在提取完一帧内的三段能量后，通过比较三段能量的能量关系来恢复出水印信息。

3.2.1 特征提取

给定原始音频 $x = \{x_1, x_2, \dots, x_k, \dots, x_L\}$ 其中 x_k 为第 k 个采样点， L 为总采样点数。首先，将音频信号 x 均匀划分为 N 个不相交的帧，每一帧可以表示为 $x_i = \{x_{(i-1)M+1}, x_{(i-1)M+2}, \dots, x_{(i-1)M+M}\}$ ，其中 $i \in [1, N]$ ，并且满足 $NM = L$ 。为了简化处理，假设 N 可以整除 L 。紧接着，每一帧 x_i 被进一步划分为三个不相交的子段，即 $x_i = [x_{i,1}, x_{i,2}, x_{i,3}]$ ，其中每一个子段 $x_{i,j}$ 的数学形式可以表达为 $\{x_{(i-1)M+(j-1)S+1}, x_{(i-1)M+(j-1)S+2}, \dots, x_{(i-1)M+(j-1)S+S}\}$ ，其中 $j \in \{1, 2, 3\}$ ，且 $S = M/3$ 。假设 S 整除 M 。随后，对每一段 $x_{i,j}$ 进行 DCT 变换，得到相应的 DCT 系数 $X_{i,j} = \{X_{i,j}(1), X_{i,j}(2), \dots, X_{i,j}(S)\}$ 。

根据人类听觉系统的特性，DCT 系数可以分为三个频段：低频段、中频段和高频段。修改低频段会显著影响音频信号的感知质量，而修改高频段虽能提高隐蔽性，但容易在信号处理过程中被干扰，从而降低水印的鲁棒性。因此，为了平衡鲁棒性与感知质量，我们选取中频段为水印的嵌入区域。本研究选取 $[f_{low}, f_{high}]$ 范围内的系数作为主要嵌入的区域。其中 f_{low} 和 f_{high} 分别是该范围的下限和上限。对应范围内的系数可以表示为： $X_{i,j}[f_{low}, f_{high}] = \{X_{i,j}(\eta_1 S), X_{i,j}(\eta_1 S + 1), \dots, X_{i,j}(\eta_2 S)\}$ 其中， $\eta_1 = 2 \cdot f_{low} / f_{sampling}$ 和 $\eta_2 = 2 \cdot f_{high} / f_{sampling}$ 。这里 $f_{sampling}$ 为音频信号的采样频率，本文默认值为 44.1 kHz。对于每一段的 $X_{i,j}[f_{low}, f_{high}]$ ，其平均能量特征可以用下式进行表示：

$$E_{i,j} = \frac{1}{|X_{i,j}[f_{low}, f_{high}]|} \sum_{X_{i,j}(k) \in X_{i,j}[f_{low}, f_{high}]} X_{i,j}^2(k) \quad (2.10)$$

其中 $|\cdot|$ 表示集合的大小。由于 TSM 和 PSM 近似于线性缩放操作。在频率线性缩放过程中，平均能量呈线性变化，这使得水印能够更好地抵抗此类攻击。此外，平均能量特征具有良好的稳定性和抗噪能力，作为水印嵌入依据，有助于提升系统在恶意攻击下的鲁棒性。

3.2.2 水印嵌入和优化策略

令 $w = \{w_1, w_2, \dots, w_N\} \in \{0, 1\}^N$ 表示水印，假设该水印是原始明文的混沌加密版本，以确保安全性。接着通过基于三角形策略的系数修改来嵌入水印 w 。具体来说，将三段不相交子段所确定的平均能量特征视为三角形的边长。这三条边的长度之间的关系被用来承载水印信息。直观上，如图 3.2 所示。假设将 w_i 嵌入到 x_i 中，本文将修改 $E_{i,1}, E_{i,2}, E_{i,3}$ 为 $E'_{i,1}, E'_{i,2}, E'_{i,3}$ 来携带 w_i 。其中的原则是，当 $w_i = 0$ 时， $E'_{i,1}, E'_{i,2}, E'_{i,3}$ 不能形成三角形，否则它们可以构成等边三角形。对于满足该要求的 $E_{i,1}, E_{i,2}, E_{i,3}$ 可以通过多种方式进行修改。从水印嵌入的角度来看，期望该修改操作能够在鲁棒性和听觉感知性之间取得良好的平衡。接下来，本节将详细介绍所提出的方法。

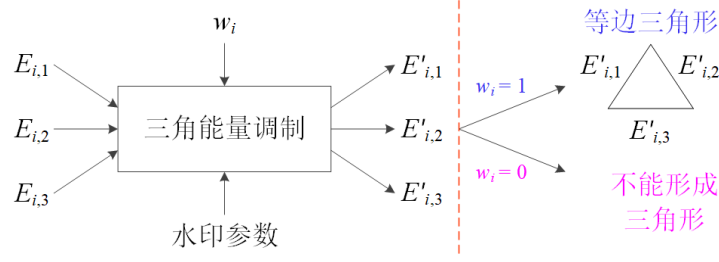


图 3.2 三角能量调制的直观说明

Figure 3.2 An intuitive illustration for triangular energy modulation

令 $E_i = \max\{E_{i,1}, E_{i,2}, E_{i,3}\}$ 且 $F_i = \frac{3E_i}{\sum_j E_{i,j}}$ 。在 $w_i = 0$ 的情况下，如果 F_i 不小于预

设的阈值 α ，则保持 $E_{i,1}$, $E_{i,2}$, $E_{i,3}$ 不变，即：

$$E'_{i,j} = E_{i,j}, \forall j \in [1,3], \text{ if } F_i \geq \alpha \quad (2.11)$$

这也表明， $[f_{low}, f_{high}]$ 范围内的相应频率系数保持不变。否则对于 $\forall j \in [1,3]$ ，将每段的平均能量做以下的修改来满足定义的三角能量关系，即：

$$E'_{i,j} = \begin{cases} \beta^2 E_{i,j} & \text{if } E_{i,j} = E_i \\ \gamma^2 E_{i,j} & \text{if } E_{i,j} \neq E_i \end{cases} \quad (2.12)$$

假设最大值 E_i 有且只有一个，若有特殊情况，可以任取其中之一作为最大值乘以 β^2 ，其余 $E_{i,j}$ 乘以 γ^2 。上述修改能量的操作是通过修改 $[f_{low}, f_{high}]$ 范围内的系数来实现的，即用 β 和 γ 分别乘以相应频率系数便得到公式 (3.3) 的修改关系，即：

$$X'_{i,j}(k) = \sqrt{\frac{E'_{i,j}}{E_{i,j}}} X_{i,j}(k), \forall k \in [f_{low}, f_{high}] \quad (2.13)$$

注意， β 和 γ 是由以下方式确定的水印参数：

$$\begin{cases} \beta = \frac{1}{\rho+1} \left(\rho + \sqrt{\rho(-E_i + \sum_{j \in [1,3]} E_{i,j}) / E_i} \right) \\ \gamma = \frac{1}{\rho+1} \left(1 + \sqrt{\rho E_i / (-E_i + \sum_{j \in [1,3]} E_{i,j})} \right) \end{cases} \quad (2.14)$$

其中 $\rho = \alpha / (3 - \alpha)$ 。因此修改后的关系满足 $F_i = \alpha$ ，即三段的能量形成不了三角形。关于 β 和 γ 的理论分析将在 3.2.4 小节中提供。当 $w_i = 1$ 的情况下，可以令 $E'_{i,1}$, $E'_{i,2}$, $E'_{i,3}$ 的值彼此相等，即：

$$E'_{i,1} = E'_{i,2} = E'_{i,3} = (E_{i,1} + E_{i,2} + E_{i,3}) / 3 \quad (2.15)$$

如上所述，平均能量将用于水印的嵌入。在基于变换域的音频水印技术中，水印通常嵌入特定区域的频率分量，然后进行逆变换以生成带水印的音频。在水印检测过程中，将对带水印音频进行频域变换，以提取其相应的频率分量，从而实现隐藏水印的提取。为了增强鲁棒性，水印的嵌入强度通常会增加，这导致特定范围内的频率系数与范围外的系数差异较大。经过一些攻击后，特征提取过程可能会涉及到超出指定范围的系数，这可能会影响水印提取的准确性。为了提高所提出方法对 PSM 和 TSM 攻击的鲁棒性，本节采用了缓冲补偿机制进行嵌入，具体如图 3.3 所示。

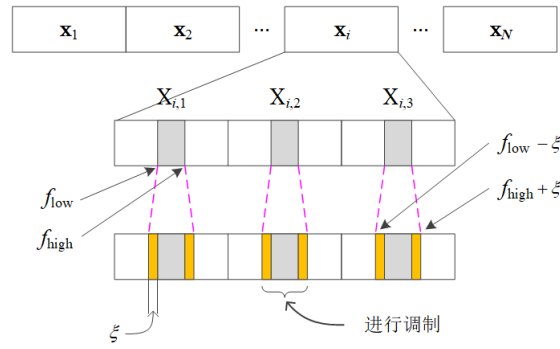


图 3.3 缓冲补偿策略的直观说明

Figure 3.3 An intuitive illustration for the proposed buffer compensation strategy

具体而言，通过一个特定值 ξ 手动扩展用于水印嵌入的系数范围，以确保水印系数与非水印系数之间的过渡更加平滑。具体的操作如下：

$$X'_{i,j}(k) = \sqrt{\frac{E'_{i,j}}{E_{i,j}}} X_{i,j}(k), \forall k \in [f_{\text{low}} - \xi, f_{\text{high}} + \xi] \quad (2.16)$$

请注意， $E'_{i,j}$ 和 $E_{i,j}$ 是从 $[f_{\text{low}}, f_{\text{high}}]$ 范围内的系数进行计算的，在提取过程中，仅使用 $[f_{\text{low}}, f_{\text{high}}]$ 范围内的系数计算平均能量特征，因此，该范围内的系数仍然满足上述嵌入规则。总体而言，缓冲补偿扩展了水印嵌入所使用的系数范围，使水印与非水印系数之间的过渡更加平滑，从而降低了特征提取过程可能会涉及到超出指定范围的系数的影响。

通过公式 (3.7) 计算得到修改后的 DCT 系数，随后将其低频和高频的 DCT 系数合并，并执行逆 DCT 变换，从而获得嵌入水印后的音频片段。将该嵌入操作应用于所有帧后，即可生成完整的含水印音频。算法 3.1 给出了水印嵌入过程的伪代码。

算法 3.1 水印嵌入算法

输入：原始音频 x ，水印 w ，密钥 k (其中包含所有预定义的参数，例如 α 等)

输出：含水印音频 x'

1. for $i=1$ to N :
 2. 通过帧划分、分段以及 DCT 变换，提取 $X_{i,j}$
 3. 根据公式 (3.1) 计算平均能量 $E_{i,1}$, $E_{i,2}$ 和 $E_{i,3}$
 4. 令 $E_i = \max\{E_{i,1}, E_{i,2}, E_{i,3}\}$ ，并计算能量比例因子 $F_i = 3E_i / \sum_j E_{i,j}$
 5. if $w_i = 0$
 6. if $F_i \geq \alpha$
 7. 根据公式 (3.2) 计算调制后能量 $E'_{i,1}$, $E'_{i,2}$ 和 $E'_{i,3}$
 8. else
 9. 根据公式 (3.3) 计算调制后能量 $E'_{i,1}$, $E'_{i,2}$ 和 $E'_{i,3}$
 10. else
 11. 根据公式 (3.6) 计算调制后能量 $E'_{i,1}$, $E'_{i,2}$ 和 $E'_{i,3}$
 12. end if
 13. end for
 14. 合并所有帧，输出含水印音频 x'
-

3.2.3 水印提取策略

水印的提取过程本质上是嵌入过程的逆操作，其核心在于通过提取算法，从含水印音频中准确还原出嵌入的秘密信息。本文所提出的算法在提取阶段遵循该原则，通过对嵌入时的调制关系执行相应的逆运算，实现水印的有效提取。由于该方法在提取过程中无需原始音频参与，故属于盲提取算法。水印提取的具体流程如图 3.4 所示，其中 $E''_{i,j}$ 表示解码器接收到信号的第 i 帧中第 j 段的平均能量特征，并且 $E_i = \max\{E_{i,1}, E_{i,2}, E_{i,3}\}$ 。

令 x'' 为解码器收到的水印信号。其中 x'' 可能是 x' 被攻击后接收到的信号，接着用 x'' 来提取水印信息，将接收到的信号进行平均分帧处理，将第 i 帧记为 x''_i 。随后对每一帧进行平均分割成三段，并对每段分别进行 DCT 变换，以提取

$[f_{low}, f_{high}]$ 范围内的频率系数。然后再用公式 (3.1) 计算每一段的特征, 并依据公式 (3.8) 提取第 i 帧嵌入的水印信息, 记为 w_i'' 。

$$w_i'' = \begin{cases} 0, & 2E_i'' \geq \sum_{j \in [1,3]} E_{i,j}'' \\ 1, & 2E_i'' < \sum_{j \in [1,3]} E_{i,j}'' \end{cases} \quad (2.17)$$

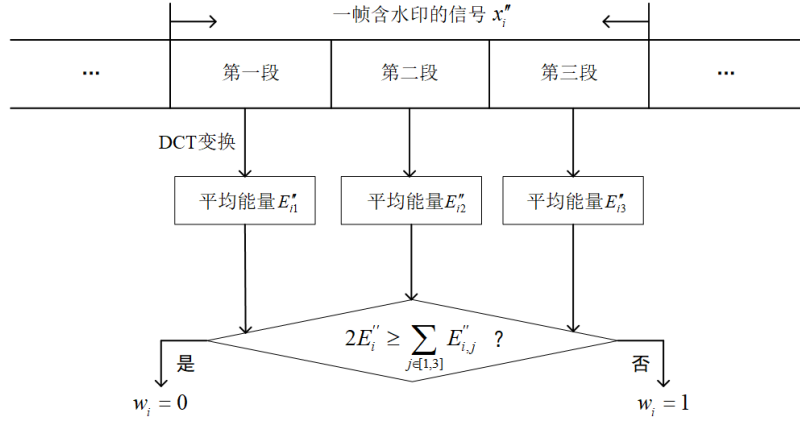


图 3.4 水印提取流程图

Figure 3.4 Flowchart of the watermark extraction procedure

3.2.4 参数优化分析

公式 (3.3) 和公式 (3.6) 分别对应水印比特 0 和 1 的嵌入操作。本节将从理论的角度分析两种嵌入方式对各种攻击的鲁棒性。

(1) 当待嵌入的水印比特为 0 时

显然, 对于公式 (3.2), 均有 $F_i' \geq \alpha$, 意味着不需要做改变。因此, 只需对公式 (3.3) 进行分析。应用公式 (3.3) 后得到:

$$F_i' = \frac{3E_i'}{\sum_{j \in [1,3]} E_{i,j}'} = \frac{3 \max\{E_{i,1}', E_{i,2}', E_{i,3}'\}}{\sum_{j \in [1,3]} E_{i,j}'} = \alpha \quad (2.18)$$

假设 $E_{i,c}'$ 对应于 E_i' , 而 $E_{i,a}'$ 和 $E_{i,b}'$ 分别对应于其余两个, 其中 a, b, c 均属于区间 $[1,3]$ 且互不相同, 则公式 (3.9) 可重写为:

$$E_{i,c}' = \rho(E_{i,a}' + E_{i,b}') \quad (2.19)$$

其中 $\rho = \alpha / (3 - \alpha)$ 。根据公式 (3.3) 可以得到:

$$\beta^2 E_{i,c} = \rho(\gamma^2 E_{i,a} + \gamma^2 E_{i,b}) \quad (2.20)$$

从而得到一个关于 β 和 γ 的关系表达式：

$$\beta = \gamma \sqrt{\frac{\rho(E_{i,a} + E_{i,b})}{E_{i,c}}} \quad (2.21)$$

考虑到一帧内总的失能量 D_i 可以表达为下式：

$$\begin{aligned} D_i &= \sum_{j \in [1,3]} \sum_{k \in [f_{\text{low}}, f_{\text{high}}]} \frac{[X'_{i,j}(k) - X_{i,j}(k)]^2}{|X_{i,j}[f_{\text{low}}, f_{\text{high}}]|} = \sum_{j \in [1,3]} \sum_{k \in [f_{\text{low}}, f_{\text{high}}]} \frac{(\sqrt{E'_{i,j}/E_{i,j}} - 1)^2 X_{i,j}^2(k)}{|X_{i,j}[f_{\text{low}}, f_{\text{high}}]|} \\ &= \sum_{j \in [1,3]} \left(\sqrt{E'_{i,j}/E_{i,j}} - 1 \right)^2 E_{i,j} = (\gamma - 1)^2 (E_{i,a} + E_{i,b}) + (\beta - 1)^2 E_{i,c} \\ &= (\gamma - 1)^2 (E_{i,a} + E_{i,b}) + (\gamma \sqrt{\frac{\rho(E_{i,a} + E_{i,b})}{E_{i,c}}} - 1)^2 E_{i,c} \end{aligned} \quad (2.22)$$

为了在公式 (3.12) 成立的情况下，使得总的失能量 D_i 最小，对 D_i 关于参数 γ 求导并置 0 得到：

$$\frac{dD_i}{d\gamma} = 2(\gamma - 1)(E_{i,a} + E_{i,b}) + 2\sqrt{\frac{\rho(E_{i,a} + E_{i,b})}{E_{i,c}}} (\gamma \sqrt{\frac{\rho(E_{i,a} + E_{i,b})}{E_{i,c}}} - 1) E_{i,c} = 0 \quad (2.23)$$

解得：

$$\gamma = \frac{1}{\rho + 1} \left(1 + \sqrt{\frac{\rho E_{i,c}}{E_{i,a} + E_{i,b}}} \right) \quad (2.24)$$

结合公式 (3.12) 可以得到：

$$\beta = \frac{1}{\rho + 1} \left(\rho + \sqrt{\frac{\rho(E_{i,a} + E_{i,b})}{E_{i,c}}} \right) \quad (2.25)$$

(2) 当待嵌入的水印比特为 1 时

若选择使 $E'_{i,1}$, $E'_{i,2}$ 和 $E'_{i,3}$ 三者相等，则公式 (3.6) 可达到最优状态，这一点可通过与上述类似的方法进行证明。为了进一步验证该优化策略的有效性，分别进行了有优化和无优化策略的对比实验，并使用 SNR 和 ODG 作为指标，以评估不同 ρ 取值下水印音频的质量。如图 3.5 所示，实验选取了 25 段不同风格的音频，结果表明，使用了优化策略相较于未使用优化策略在 SNR 和 ODG 指标上均表现出明显优势，尤其在较大的 ρ 时更为突出。这种性能提升主要源于优化过程对能量失真的有效降低。

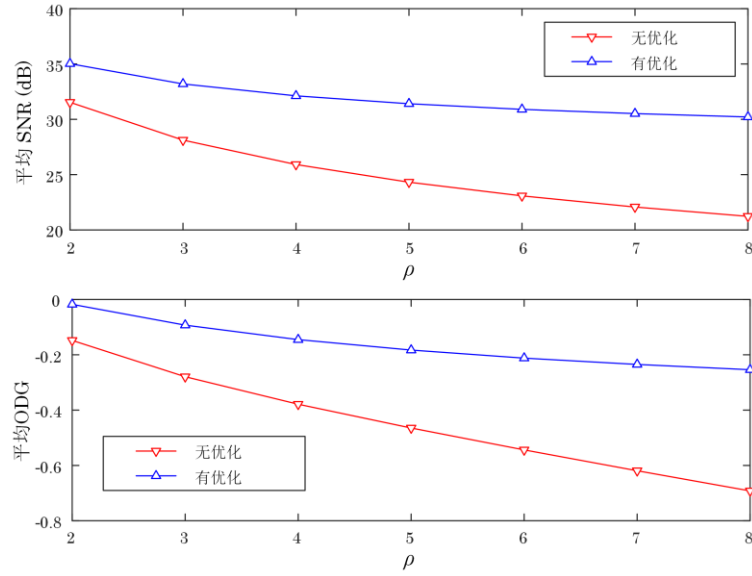


图 3.5 有优化与无优化的平均 SNR 和 ODG 值

Figure 3.5 The mean SNRs and ODGs with and without optimization

3.2.5 鲁棒性分析

公式 (3.2)、公式 (3.3) 和公式 (3.6) 分别对应于不同的水印嵌入操作。接下来，将从理论上分析这些嵌入方式对各种攻击的鲁棒性。

(1) 对于公式 (3.2) 和公式 (3.3) 的情况

当取 $\beta = \gamma = 1$ 时，公式 (3.3) 可简化为公式 (3.2)，即公式 (3.2) 是公式 (3.3) 的特例。显然，对于公式 (3.2) 而言，均有 $F'_i \geq \alpha$ 。因此，只需对公式 (3.3) 进行分析。我们可以将公式 (3.10) 重写为：

$$\rho = \frac{E'_{i,c}}{E'_{i,a} + E'_{i,b}} = \frac{\sum_{k \in [f_{low}, f_{high}]} X'_{i,c}(k)^2}{\sum_{k \in [f_{low}, f_{high}]} (X'_{i,a}(k)^2 + X'_{i,b}(k)^2)} \quad (2.26)$$

经过各种攻击后，可以将公式 (3.17) 建模为：

$$\rho' = \frac{E''_{i,c}}{E''_{i,a} + E''_{i,b}} = \frac{E'_{i,c} + \Delta_c}{E'_{i,a} + E'_{i,b} + \Delta_a + \Delta_b} \quad (2.27)$$

其中， $E''_{i,a}$ ， $E''_{i,b}$ ， $E''_{i,c}$ 分别表示含水印信号在攻击后相应频带区间内的平均能量， $\Delta_a = E''_{i,a} - E'_{i,a}$ ， $\Delta_b = E''_{i,b} - E'_{i,b}$ ， $\Delta_c = E''_{i,c} - E'_{i,c}$ 分别表示三部分能量的变化量。本文所提出的方法本质上对闭环攻击具有天然的抵抗性，即在闭环攻击下有 $\Delta_a = \Delta_b = \Delta_c = 0$ 。此外，低通滤波、MP3 压缩和 AAC 压缩攻击主要影响的是音

频的高频成分，而所提出的特征嵌入方法作用于音频的中频范围，因此受这些攻击的影响极小。因此，可以合理地认为 $\Delta_a \approx \Delta_b \approx \Delta_c \approx 0$ 。由此可见，本文提出的方法在理论上对低通滤波、MP3 压缩及 AAC 压缩攻击具有较强的鲁棒性。当攻击类型为幅度缩放（缩放因子记为 σ ）时，公式（3.17）经过攻击后的关系可建模为：

$$\rho' = \frac{E''_{i,c}}{E''_{i,a} + E''_{i,b}} = \frac{\sigma^2 E'_{i,c}}{\sigma^2 E'_{i,a} + \sigma^2 E'_{i,b}} = \rho \quad (2.28)$$

上式表明，本章的方法对幅度缩放攻击具有较强的鲁棒性。此外，对于添加回声、重新采样、量化和噪声攻击，该式可进一步改写为：

$$\rho' = \frac{E''_{i,c}}{E''_{i,a} + E''_{i,b}} = \frac{\rho(E'_{i,a} + E'_{i,b}) + \Delta_c}{E'_{i,a} + E'_{i,b} + \Delta_a + \Delta_b} \quad (2.29)$$

根据式（3.20），为了成功提取比特 0，需满足条件： $E''_{i,c} \geq E''_{i,a} + E''_{i,b}$ ，即 $\rho' \geq 1$ ，因此，可得到：

$$E'_{i,a} + E'_{i,b} \geq (\Delta_a + \Delta_b - \Delta_c) / (\rho - 1) \quad (2.30)$$

其中， $E'_{i,a}$ 和 $E'_{i,b}$ 表示着相应频带的主要信息，从经验上看，攻击者很难在不明显损害水印音频质量的情况下，产生较大的 $\Delta_a + \Delta_b - \Delta_c$ 。通过调节参数 α 可产生一定的误差缓冲区，从而使公式（3.21）更易满足。因此，本章提出的方法对添加回声、重采样、量化和噪声攻击具有良好的鲁棒性。

（2）对于公式（3.6）的情况

显然， $(E'_{i,a} + E'_{i,b})/2 = E'_{i,c}$ 。该方案本质上对闭环攻击具有抗性，同时对低通滤波、MP3 压缩和 AAC 压缩攻击也具备良好的鲁棒性。对于添加回声、重采样、量化和噪声攻击，根据公式（3.8），若要成功提取比特 1，需满足：

$$E''_{i,c} < E''_{i,a} + E''_{i,b} \quad (2.31)$$

即：

$$E'_{i,c} + \Delta_c < E'_{i,a} + \Delta_a + E'_{i,b} + \Delta_b \quad (2.32)$$

进一步可改写为：

$$\frac{E_{i,1} + E_{i,2} + E_{i,3}}{3} > \Delta_c - (\Delta_a + \Delta_b) \quad (2.33)$$

可以看出公式（3.24）的成立与水印参数无关，即方法的鲁棒性完全取决于基于帧内信号能量强度的自适应控制。并从后续的实验可以观察到，帧内的能量的自适应控制足以抵抗这些攻击。

如文献[41]所述，TSM 和 PSM 可看作是在频率轴上的拉伸或压缩操作。经过 TSM 与 PSM 攻击后，未嵌入水印的系数可能被错误地参与到水印提取过程中，而部分已嵌入水印的系数则可能被错误排除。因此，为了应对这些未嵌入但被错误纳入提取过程的系数所引入的干扰，通过设计缓冲补偿来降低这些系数的影响，从而有效减小提取误差。其数学模型可表示为：

$$\rho' = \frac{E_{i,c}}{E_{i,a} + E_{i,b}} = \frac{E'_{i,c} + \Delta_c(\xi)}{E'_{i,a} + E'_{i,b} + \Delta_a(\xi) + \Delta_b(\xi)} \quad (2.34)$$

其中， $\Delta_a(\xi)$, $\Delta_b(\xi)$ 和 $\Delta_c(\xi)$ 对应的是未嵌入水印信息但参与水印提取的频率系数。由于缓冲窗口的缩放因子是由嵌入操作确定的，因此我们可以近似为：

$$\Delta_a(\xi) \approx \rho(\Delta_a(\xi) + \Delta_b(\xi)) \quad (2.35)$$

因此可以将式（3.25）改写为：

$$\rho' = \frac{E_{i,c}}{E_{i,a} + E_{i,b}} \approx \frac{E'_{i,c} + \Delta_c(\xi)}{E'_{i,c} / \rho + \Delta_c(\xi) / \rho} = \rho \quad (2.36)$$

理论上所提出的方法能够有效抵抗 TSM 和 PSM 攻击。在适度强度的抖动攻击情况下，可以假设 $\Delta_a(\xi) \approx \Delta_b(\xi) \approx \Delta_c(\xi) \approx 0$ 。这表明本文提出的方法对于抖动攻击同样具有良好的鲁棒性。

3.3 实验结果与分析

本节将展示和分析所提方法的实验结果。首先介绍实验的整体设置，包括测试所用数据集、对比的音频水印方法以及关键参数的优化配置。接着，从水印的不可感知性和鲁棒性两个方面分别进行实验验证与性能分析，以全面评估所提方法的实用性和稳定性。最后，通过消融实验深入探讨关键模块与参数设置对水印性能的影响，进一步验证方法设计的合理性与有效性。通过上述实验设计，可以全面验证所提方法在实际应用中的可行性与优势。

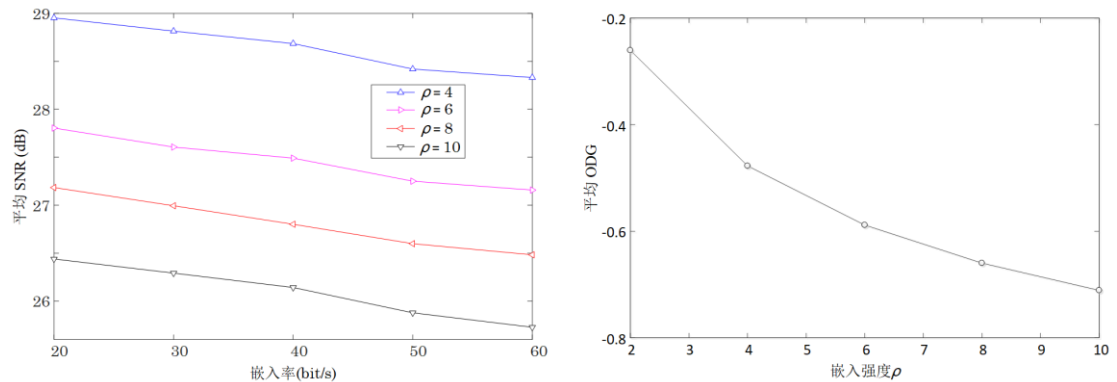
3.3.1 实验设置

为了公平起见，本研究采用文献[24]广泛使用的数据库进行后续实验。从该数据库中选择 25 个不同音乐流派的音频片段，包括古典音乐、钢琴曲、爵士乐、说唱音乐、流行音乐和民谣等。所有音频均统一处理为长度 20 秒，采样频率为 44.1 kHz，量化参数为 16。此外，本实验设置低频边界 $f_{\text{low}} = 3000\text{Hz}$ ，高频边界 $f_{\text{high}} = 4000\text{Hz}$ ，参数 ξ 设为 0.5 kHz， α 取 2.667，对应的 ρ 为 8。需要说明的是，上述参数可根据实际需求进行调整以进一步优化性能。实验结果表明，这些参数设置已能确保水印方案取得良好的性能表现。

3.3.2 水印隐蔽性测试

为了客观地评估本文提出水印方法的隐蔽性，本研究选取了来自 25 个不同音乐流派的音频片段作为测试样本，并采用 SNR 和 ODG 两个指标对嵌入水印后的音频质量进行测量与分析。

图 3.6(a)展示了不同嵌入强度 ρ 和嵌入速率对应的 SNR。如预期所示，更高的嵌入速率和更强的嵌入强度会对信号的信噪比产生负面影响，即随着嵌入强度和速率的提高，SNR 逐渐降低。然而，由图 3.6(a)可见，即使在嵌入速率高达 60 bps（比特每秒）时，本文所提出的方法仍然能够维持较高的感知质量。换句话说，本文方法在提供较高容量的同时，有效保证了水印音频的感知质量。



(a) 不同嵌入率和 ρ 值下的平均 SNR 值

(b) 不同嵌入强度下的平均 ODG 值

图 3.6 不同嵌入率与嵌入强度对音频质量的影响

Figure 3.6 Impact of embedding rate and strength on audio quality

图 3.6(b)进一步展示了在 30 bps 的固定嵌入速率下, 不同嵌入强度对应的 ODG 结果。可以看出, 随着嵌入强度的增加, ODG 值逐渐降低。但在各种嵌入强度下, 平均 ODG 值始终高于-1, 并且在较小嵌入强度时接近 0, 这表明本文方法具有较好的音频感知质量。综合考虑鲁棒性与感知透明性两个方面, 本实验经验性地选择参数 $\rho=8$, 默认嵌入速率设置为 30 bps (除非另有特别说明)。需要指出的是, 由于音频信号本身具有多样性, 针对特定音频, 上述参数设置可能并非最优。然而, 实验结果总体表明, 本文所建议的参数设置能够有效兼顾鲁棒性和良好的音频感知质量。

3.3.3 水印鲁棒性测试

为了公平地进行评估, 针对各种攻击选取了典型的参数设置。具体而言, 对于噪声攻击, 本实验采用 AWGN, 使得水印音频与噪声之间的 SNR 分别为 20 dB 或 30 dB; 对于量化攻击, 音频信号的量化位数从 16 位降低至 8 位; 对于幅度缩放攻击, 缩放系数分别设为 0.8 和 1.2; 对于回声攻击, 设置回声延迟为 0.5 秒, 且回声幅度为水印信号幅度的 20%。在 MP3 压缩攻击中, 压缩码率分别选取 128 kbps 和 96 kbps; 同样地, 在 AAC 压缩攻击中, 也分别采用 128 kbps 和 96 kbps 的码率。对于重采样攻击, 首先将水印音频的采样频率从 44.1 kHz 降采样至 22.05 kHz, 然后再通过升采样恢复至原始的 44.1 kHz。对于低通滤波攻击, 截止频率设定为 8 kHz, 而高通滤波攻击的截止频率设定为 0.5 kHz。

为了评估本方法的鲁棒性, 将本文提出的方法与文献中四种先进的音频水印算法进行了比较。为了便于描述, 我们将本章方法命名为三角调制法, 其它方法依据参考文献序号来命名。如表 3.1 所示, 三角调制法与文献[21]和[59]中的方法一样, 对常见的信号处理攻击表现出较强的鲁棒性, 并获得了较低的 BER。然而, 文献[58]提出的方法无法有效抵抗高通滤波和回声攻击, 这是因为该方法使用的是二级小波分解的近似系数作为嵌入特征, 当高通滤波去除低频分量时, 这些特征会受到明显影响。类似地, 回声攻击引入的延迟回声也会改变这些特征, 从而导致水印提取失败。另外, 文献[42]的方法对大部分攻击缺乏鲁棒性, 其基于直方图的水印嵌入策略对嵌入速率的变化较为敏感。当嵌入速率提高时, 直方图的每个组距宽度会减小, 导致每个组距内特征数量降低, 从而削弱了水印的鲁棒性。

表 3.1 在嵌入率为 30 bps 下, 不同方法对常见攻击的平均 BER 和平均 NCC 值
Table 3.1 The mean BER and mean NCC of different methods against
common attacks at an embedding rate of 30 bps

普通攻击		文献[21]		文献[59]		文献[58]		文献[42]		三角调制	
		BER	NCC	BER	NCC	BER	NCC	BER	NCC	BER	NCC
无攻击		0.00	1	0.08	0.9991	0.00	1	2.96	0.9687	0.00	1
噪声 (dB)	30	0.01	0.9999	0.09	0.9991	0.19	0.998	49.39	0.4877	0.39	0.996
	20	0.03	0.9997	0.09	0.9991	1.96	0.9793	50.31	0.4775	3.70	0.9641
量化		0.00	1	0.08	0.9991	0.10	0.9989	47.69	0.4882	0.27	0.9971
幅度 缩放	0.8	0.00	1	0.08	0.9991	0.00	1	2.53	0.9729	0.00	1
	1.2	0.00	1	0.08	0.9991	0.00	1	2.43	0.9744	0.00	1
回声		1.23	0.9871	1.58	0.9833	21.49	0.7743	49.78	0.4817	2.91	0.9687
MP3 (kbps)	128	0.00	1	0.05	0.9994	0.03	0.9997	39.57	0.5872	0.00	1
	96	0.00	1	0.17	0.9982	0.08	0.9992	46.86	0.5134	0.01	0.9999
AAC (kbps)	128	0.00	1	0.08	0.9991	0.00	1	45.34	0.5262	0.00	1
	96	0.00	1	0.08	0.9991	0.00	1	48.75	0.4908	0.00	1
重采样		0.00	1	0.16	0.9983	0.16	0.9983	41.97	0.5609	0.00	1
低通滤波		0.00	1	0.32	0.9966	1.17	0.9876	48.47	0.4936	0.01	0.9999
高通滤波		0.00	1	1.59	0.9832	42.46	0.5615	49.89	0.4859	0.08	0.9991

针对去同步攻击, 具体设置如下攻击参数: 对于抖动攻击, 每 100 个水印信号样本中删除一个样本; 对于 TSM 攻击, 时间缩放系数取值范围为[0.8, 1.2]; 对于 PSM 攻击, 在保持信号时长不变的前提下, 音高缩放系数同样设定在[0.8, 1.2]范围内。

表 3.2 在嵌入率为 30 bps 下, 不同方法对去同步攻击的平均 BER 和平均 NCC
Table 3.2 The mean BER and mean NCC of different methods against desynchronization attacks
at an embedding rate of 30 bps

去同步攻击		文献[21]		文献[59]		文献[58]		文献[42]		三角调制	
		BER	NCC	BER	NCC	BER	NCC	BER	NCC	BER	NCC
抖动(1%)		49.80	0.4991	49.27	0.4925	28.55	0.6986	3.97	0.958	6.30	0.933
TSM	80%	49.67	0.5023	49.39	0.491	39.63	0.5841	49.74	0.4866	6.74	0.9344
TSM	90%	50.09	0.498	49.47	0.4925	35.60	0.625	50.18	0.4822	4.12	0.9604
TSM	110%	49.95	0.5015	49.69	0.4901	35.61	0.6236	50.24	0.4814	5.02	0.9531
TSM	120%	49.79	0.4986	50.41	0.4806	37.76	0.5989	49.70	0.4838	6.30	0.9401
PSM	80%	50.04	0.4939	50.13	0.4852	39.72	0.5824	50.21	0.4741	16.83	0.8287
PSM	90%	50.31	0.4918	49.62	0.4946	36.14	0.6178	50.44	0.4809	8.05	0.9181
PSM	110%	49.21	0.5031	49.55	0.4899	36.30	0.6186	50.63	0.4749	9.99	0.9009
PSM	120%	50.33	0.4937	50.09	0.4918	39.23	0.5843	49.86	0.4807	15.36	0.8417

表 3.2 展示了不同方法在去同步攻击下的实验结果。从结果中可以看出, 本文提出的方法在抵御去同步攻击方面具有明显优势, 所有情况下的 BER 均低于 17%, 这与第 3.2.5 节的理论分析是一致的。相比之下, 文献[21]和[59]中的方法对去同步攻击完全失效, 平均 BER 接近或达到 50%, 说明其在去同步攻击下无

法成功提取水印信息。文献[58]和[42]的方法仅在抖动攻击下表现良好,但在 TSM 和 PSM 攻击下则完全失效。与之相比,本文所提出的三角调制方法在各类去同步攻击下均表现出优越的鲁棒性。例如,在抖动攻击下,BER 仅为 6.3%,显著低于多数对比方法;在 TSM 攻击下,无论是 TSM(80%、90%)还是 TSM(110%、120%),BER 均稳定控制在 6.3%~6.74%之间,且 NCC 始终高于 0.93,说明水印信息不仅成功提取,且提取精度较高。在 PSM 攻击下,尽管整体误码率略有上升,但仍远低于其他方法,BER 在 8.05%~16.83%范围内,仍能保持较强的稳定性。这说明提出的方法在时域结构发生拉伸或压缩时仍能有效保持水印帧间的能量特征关系。总体而言,去同步攻击破坏了水印嵌入位置的稳定性,显著增加了水印信息准确提取的难度。这些已有方法未能选择合适的特征或技术有效应对这一问题,而本文提出的方法则体现出了更为优越的鲁棒性。

3.3.4 消融实验与分析

为了研究缓冲窗口长度对抵御去同步攻击性能的影响,通过改变缓冲窗口的长度,并对嵌入水印后的音频信号施加不同强度的去同步攻击来测量 BER。图 3.7 展示了不同 ξ 值下的平均 SNR 和平均 ODG。从图中可以看出,随着缓冲窗口长度的增加,SNR 和 ODG 均呈下降趋势,这是合理的,因为较大的长度意味着为了在潜在攻击下成功嵌入和提取水印,需要修改更广泛范围的频域系数。总体而言,嵌入水印后音频的感知质量仍能维持在令人满意的水平。

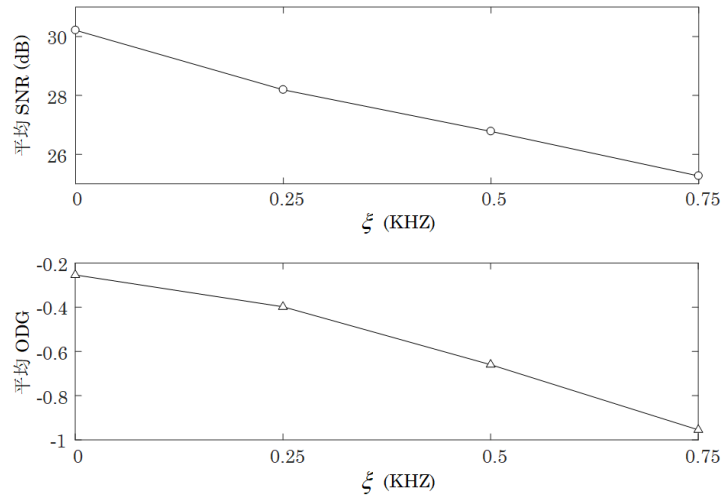


图 3.7 不同 ξ 值下的平均 SNR (dB) 和平均 ODG 值

Figure 3.7 Average SNR (dB) and ODG values under different buffer window lengths(ξ)

另一方面,如图 3.8 所示,BER 随着缓冲窗口长度的增加而逐渐下降,尤其在面对 PSM 攻击时表现得更为明显。这一现象可以归因于第 3.2.5 节所述的 TSM 和 PSM 攻击可建模为沿频率轴的拉伸或压缩操作。这些操作会导致某些未嵌入水印的系数在提取过程中被误识为水印信息。通过应用窗口操作,有效地对这些非水印系数进行了缩放,从而减少了它们对特征提取过程的干扰。

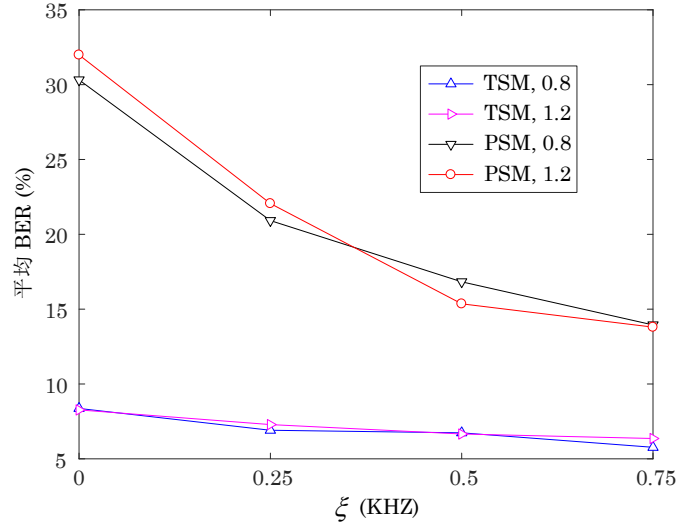


图 3.8 不同 ξ 值和不同攻击下的平均 BER 值
Figure 3.8 Average BER under different ξ values and various attacks

3.4 本章小结

本章研究了抵御 TSM 和 PSM 攻击的鲁棒音频水印算法。首先分析了两类攻击对音频信号的影响,指出现有方法的不足,并提出一种结合三角调制与缓冲补偿的方案。该方法在帧划分与 DCT 变换后,利用三段能量特征的三角关系嵌入水印,并通过缓冲补偿降低去同步攻击引起的提取误差。理论分析表明,该方法能有效抵抗 TSM、PSM 及常见信号处理攻击。实验部分通过 SNR 与 ODG 指标评估了算法的隐蔽性和有效性,结果显示在保持较高嵌入容量的同时具备良好的感知透明性。与四种先进方法的对比进一步验证了本章方法在常见信号处理与去同步攻击下的优势,平均误码率低于 17%,体现了较强的鲁棒性。总体而言,该方法兼顾了不可感知性与鲁棒性,并验证了理论分析的有效性。

第四章 抵抗去同步攻击的 双段式能量调制音频水印算法

第三章所提出的基于三角调制的鲁棒音频水印方案主要针对常规的信号处理攻击、TSM 攻击和 PSM 攻击，但在应对裁剪攻击方面仍存在一定局限性。为进一步提升对 TSM 和 PSM 攻击的鲁棒性，并增强对裁剪攻击的抵抗能力，在本章中提出了一种可同时抵抗常规信号处理攻击与去同步攻击的音频水印方案。该方案首先根据 TSM 和 PSM 攻击的特性设计了稳健的特征；随后，针对裁剪攻击的影响，构建了对称结构的同步码嵌入与检测机制；最后，通过对比实验验证了所提方案的有效性。

4.1 引言

在前一章节中，提出的基于三角调制策略的鲁棒音频水印方法，有效提升了在 TSM 与 PSM 攻击下的性能。然而，该方法在面对裁剪攻击时仍存在一定局限性。由于裁剪操作会直接破坏音频的帧结构，造成水印帧的丢失或位置偏移，进而打破了水印系统的同步关系，导致无法准确提取水印信息。因此，提升音频水印算法在裁剪攻击下的同步恢复能力，成为进一步优化系统性能的关键问题。

针对这一问题，本文提出一种改进式的拼凑算法与对称同步码检测机制的鲁棒音频水印方法。该方法首先在帧内提取两个不相邻子段的中频能量关系进行水印嵌入。同时设计对称同步码嵌入结构，通过帧间冗余嵌入以增强同步定位能力。此外，为进一步平衡鲁棒性与不可感知性，本文在该框架下引入线性递减的缓冲补偿机制。通过对嵌入区域边界频谱系数的平滑处理，有效减弱了嵌入造成的能量突变。相比于第三章的缓冲补偿机制，缓解了音频因嵌入水印后感知的恶化。

本章将围绕该方法的特征提取、水印嵌入、同步检测与提取机制进行详细阐述，并通过多组实验从不同攻击类型下的鲁棒性和隐蔽性两个维度进行性能评估，验证所提算法在抵御去同步攻击方面的有效性与实用价值。

4.2 基于双段式能量调制的水印嵌入方案

为了提升水印算法在 TSM 和 PSM 攻击下的鲁棒性,本文在基于三角调制策略的方法基础上,提出了一种基于能量调制的拼凑水印算法。同时,结合优化后的缓冲补偿机制,显著提升了水印嵌入后的听觉感知效果。针对最为棘手的裁剪攻击,设计了一套同步码的嵌入与检测机制。整体框架如图 4.1 所示。

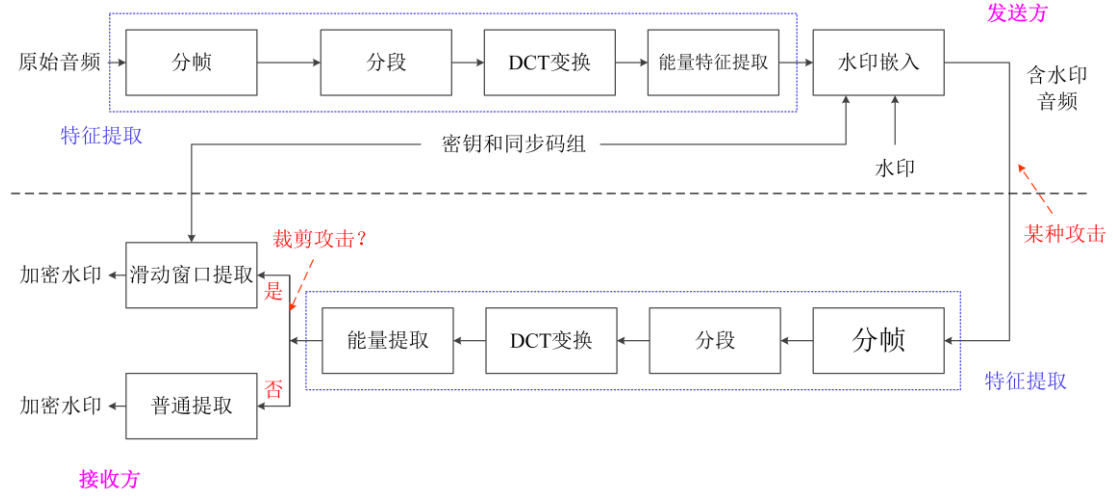


图 4.1 音频水印方案整体框图

Figure 4.1 Overall framework of audio watermarking scheme

4.2.1 特征提取

特征的提取过程与第 3.2.1 节类似,首先对原始音频进行分帧处理,然后将每一帧划分为两个不相邻的子段。对每个子段分别进行 DCT 变换,提取其中频系数的能量作为特征用于水印嵌入。因为 TSM 和 PSM 攻击可近似视为线性的缩放操作。因此,若采用两段能量之间的关系作为嵌入特征,在 TSM 和 PSM 的影响下,该能量关系基本保持不变,其原理如下所示:

假设原始信号为 $f(t)$, 其中 $t \in [t_1, t_2]$ 表示其定义在有限区间内的时间变量。DCT 变换后可以表示为 $F(k)$, 其中 $k \in [k_1, k_2]$ 表示离散频率索引。为简化分析,采用傅里叶变换的缩放性质,对 TSM 和 PSM 攻击近似建模为 $f'(t) = f(at)$, 其中 a 为 TSM 引起的线性缩放因子。根据傅里叶变换的缩放特性,可得攻击后的频域表达为:

$$F'(k) = \frac{1}{|a|} F\left(\frac{k}{a}\right) \quad (3.1)$$

其中, $k \in [ak_1, ak_2]$ 。因此, 频域能量特征为:

$$\begin{aligned} E'_{i,j} &= \int_{k=ak_1}^{ak_2} F'^2(k) dk = \int_{k=ak_1}^{ak_2} \frac{1}{|a|^2} F\left(\frac{k}{a}\right)^2 dk \\ &= \frac{1}{a} \int_{k=ak_1}^{ak_2} F\left(\frac{k}{a}\right)^2 d\frac{k}{a} = \frac{1}{a} E_{i,j} \end{aligned} \quad (3.2)$$

由于使用的是拼凑算法, 假设嵌入水印的两段能量满足 $E_{i,1} > E_{i,2}$, 则经过 TSM 攻击后其大小关系仍满足: $\frac{1}{a} E_{i,1} > \frac{1}{a} E_{i,2}$ 。注意, 该理论是用的模拟信号进行分析, 离散信号分析类似。

4.2.2 音频水印的嵌入

所提出方法是基于拼凑方法的, 根据文献[41]的启发, 本章方法通过修改每帧信号内相邻两段能量的倍数关系进而嵌入水印信息。不失一般性, 总体的修改规则可以建模为:

$$\begin{cases} E'_{i1} \geq \alpha E'_{i2}, W(i) = 0 \\ E'_{i2} \geq \alpha E'_{i1}, W(i) = 1 \end{cases} \quad (3.3)$$

其中 α 为所提出方法的嵌入强度因子, E'_{i1} , E'_{i2} 分别表示嵌入水印后的两段的能量, W 表示水印加密后的版本。嵌入强度因子 α 有助于在水印提取阶段创建错误缓冲区。 α 太小会导致鲁棒性不足, 太大则会影响音频信号的感知质量。上述平均能量的公式是在频率系数修改的基础上完成的, 详细的嵌入策略具体如下:

(1) 当 $W(i)=0$ 时, 通过段内的能量关系嵌入水印信息。如果相邻两段原始的能量关系满足公式 (4.3), 不做任何改变。否则用缩放系数 β 和 γ 来调整段内的能量关系来满足公式 (4.3), 即:

$$E'_{i,j} = \begin{cases} \beta^2 E_{i,1} \\ \gamma^2 E_{i,2} \end{cases} \quad (3.4)$$

为了最大限度的减小在嵌入水印过程中引入的听觉损失，本文对 β 和 γ 的数值进行了优化，其中 β 和 γ 的表示如下：

$$\begin{cases} \beta = \frac{\alpha E_{i2} + E_{i2} \sqrt{\alpha E_{i1} E_{i2}}}{\alpha E_{i2} + E_{i1} E_{i2}} \\ \gamma = \frac{\sqrt{\alpha E_{i1} E_{i2}} + E_{i1} E_{i2}}{\alpha E_{i2} + E_{i1} E_{i2}} \end{cases} \quad (3.5)$$

其中该优化策略将在 4.2.4 节进行说明。

(2) 当 $W(i)=1$ 时，类似于公式 (4.4)，依然用优化后的 β 和 γ 来修改频率系数，此时 β 和 γ 表示如下：

$$\begin{cases} \beta = \frac{E_{i2} + E_{i2} \sqrt{\alpha E_{i1} E_{i2}}}{E_{i2} + \alpha E_{i1} E_{i2}} \\ \gamma = \frac{\sqrt{\alpha E_{i1} E_{i2}} + \alpha E_{i1} E_{i2}}{E_{i2} + \alpha E_{i1} E_{i2}} \end{cases} \quad (3.6)$$

由于 β 和 γ 修改的频率范围为 $[f_{low}, f_{high}]$ ，若嵌入强度因子 α 设置过大，将导致该频段内嵌入水印的系数与其外部未嵌水印区域的系数存在明显差异。某些攻击可能会使特征提取过程涉及到该指定范围以外的频率系数，从而对所提取的特征造成严重干扰。为此，在水印嵌入区域与非嵌入区域之间设置缓冲区，以实现两者之间的平滑过渡，从而提升系统在遭受攻击时的鲁棒性。不同于 3.2.2 节提出的缓冲补偿机制，本节方法对其做了一定的改进。提高了其听觉感知性。具体操作如下所述：

$$\begin{cases} X'_{i,j}(k) = \left[1 + \left(\sqrt{\frac{E'_{i,j}}{E_{i,j}}} - 1 \right) \cdot \frac{k - (f_{low} - \xi)}{\xi} \right] \cdot X_{i,j}(k), & k \in [f_{low} - \xi, f_{high}] \\ X'_{i,j}(k) = \sqrt{\frac{E'_{i,j}}{E_{i,j}}} X_{i,j}(k), & k \in [f_{low}, f_{high}] \\ X'_{i,j}(k) = \left[1 + \left(\sqrt{\frac{E'_{i,j}}{E_{i,j}}} - 1 \right) \cdot \left(1 - \frac{k - f_{high}}{\xi} \right) \right] \cdot X_{i,j}(k), & k \in [f_{high}, f_{high} + \xi] \end{cases} \quad (3.7)$$

同 3.2.2 节一样，频率范围 $[f_{low}, f_{high}]$ 内的系数仍严格遵循既定的嵌入规则，在水印提取阶段也仅使用该范围内的系数进行特征提取。由于三角调制策略能够自适应地选择能量较大的子段并增强其平均能量，同时相应减小其余两个能量较小子

段的能量，因此相比于传统的非自适应拼凑的嵌入方法，三角调制策略在听觉感知方面表现更优。基于此，第 3.2.2 节中采用了在水印区域与非水印区域之间施加相同缩放因子的方法，以平滑两者的系数过渡，如公式 (3.7) 所示。为了进一步提升水印嵌入后的音频隐蔽性，引入了一种线性递减的缓冲补偿策略，即从水印区域向非水印区域之间进行线性过渡，具体形式见公式 (4.7)，该策略一定程度上提升了水印的隐蔽性，其具体实验结果见第 4.3.3 节。

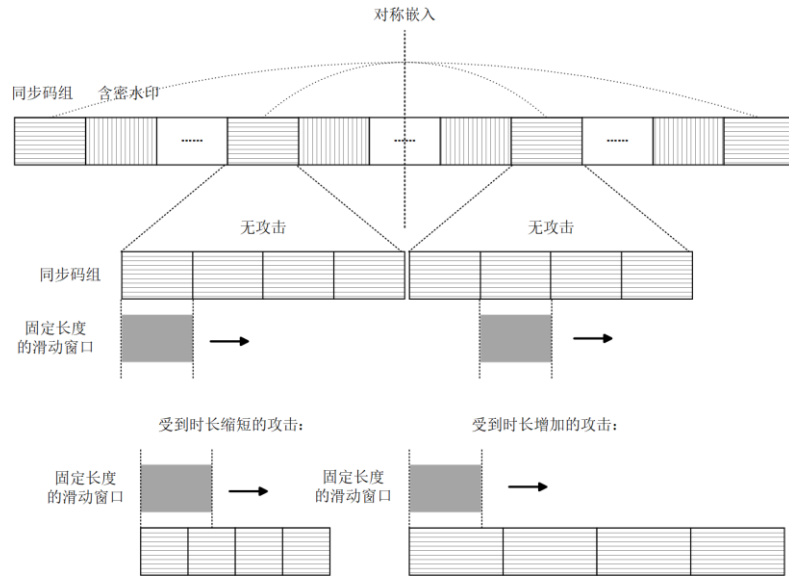


图 4.2 滑动窗口在水印提取过程中的示意图

Figure 4.2 Illustration of the sliding window in the watermark extraction process

基于平均分段的水印策略在面对裁剪攻击时表现出显著的脆弱性。如果缺乏帧同步方法，即便是简单的裁剪操作也可能完全破坏嵌入的水印。传统的同步帧检测方法通常通过固定大小的滑动窗口遍历信号。虽然这种方法对裁剪攻击具有一定的抵抗力，但在遭受 TSM 或抖动攻击时，由于嵌入水印后每帧的长度发生变化，用固定长度的信号提取水印会导致嵌入水印的区域无法完整提取，或者遗漏特定的水印区域，从而影响同步帧的准确检测，如图 4.2 所示。针对这些问题，本节提出了一种改进的对称同步码嵌入与检测机制。该机制的核心在于：在同步码嵌入阶段，依据公式 (4.3) 所定义的嵌入策略，将相同的同步码组以对称方式嵌入音频信号。该设计便于在后续检测阶段，通过统计对称位置上同步码相等的数量，从而判断音频是否遭受了裁剪攻击。然后，再根据攻击类型选取不同的水印提取模式，具体提取模式将在 4.2.3 节进行阐述。

4.2.3 音频水印的提取

在检测阶段,首先利用平均分帧的方法来检测对称位置同步码相等的数量来判断是否为裁剪攻击。具体操作如下,令 x'' 为解码器收到的水印。其中 x'' 可能是 x' 被攻击后接收到的信号,接着用 x'' 来提取水印信息,将接收到的信号进行平均分帧处理,将第 i 帧记为 x_i'' 。随后对每一帧进行平均分割成两个子段,并对每段分别进行 DCT 变换,以提取相应的频率系数。然后再计算每一段的能量特征,并依据公式 (4.8) 提取第 i 帧嵌入的水印信息,记为 w_i'' 。

$$w_i'' = \begin{cases} 0, & E_1' \geq E_2' \\ 1, & E_1' < E_2' \end{cases} \quad (3.8)$$

随后,判断 w_i'' 对称位置的同步码是否相等,如果相等的同步码数量低于预设阈值,则判定为可能发生了裁剪攻击。紧接着,利用固定大小的滑动窗口,从音频两端分别检测同步码组的位置,进而提取水印信息。由于嵌入水印的帧长度未发生变化,这一过程能够确保水印的正确提取。反之,如果相等的同步码数量超过阈值,则直接从 w_i'' 对应的位置来提取水印,理论分析将在 4.2.5 节阐述。

4.2.4 参数优化分析

在本节中讨论 4.2.2 节提出的优化策略,即 β 和 γ 的取值,同 3.2.4 节理论一样,其核心思想是在不改变嵌入条件的前提下,使得公式 (4.9) 成立,同时最小化 DCT 系数的失真。为便于分析,且不失一般性,假设当前嵌入的水印比特为 0,水印为 1 的情形可采用同样的分析方法。结合公式 (4.4),可以得到:

$$\frac{E'_{i1}}{E'_{i2}} = \alpha \quad (3.9)$$

化简之后得到:

$$\beta^2 E_{i1} = \alpha(\gamma^2 E_{i2}) \quad (3.10)$$

重写公式 (4.10) 得到 β 和 γ 的关系即:

$$\beta = \gamma \sqrt{\frac{\alpha E_{i2}}{E_{i1}}} \quad (3.11)$$

由 DCT 系数修改引入的失真可以定义为：

$$D_i = \sum_{j \in [1,2]} \sum_{k \in [f_{low}, f_{high}]} \frac{[X'_{i,j}(k) - X_{i,j}(k)]^2}{|X_{i,j}[f_{low}, f_{high}]|} \quad (3.12)$$

需要指出的是，在频率区间 $[f_{low} - \xi, f_{high} + \xi]$ 内， $[f_{low}, f_{high}]$ 区间内的系数占据了绝大多数。并且如公式 (4.7) 所示，在该主频段以外的部分，采用的是逐渐递减的调制策略，即越接近边缘，所施加的改动幅度越小。因此，失真功率在本文中只考虑在 $[f_{low}, f_{high}]$ 区间内的频率系数。根据公式 (4.11) 和 (4.12)，可以得到：

$$\begin{aligned} D_i &= \sum_{j \in [1,2]} \left(\sqrt{E'_{i,j} / E_{i,j}} - 1 \right)^2 E_{i,j} \\ &= (\beta - 1)^2 E_{i,1} + (\gamma - 1)^2 E_{i,2} \\ &= (\gamma \sqrt{\frac{\alpha E_{i2}}{E_{i1}}} - 1)^2 E_{i,1} + (\gamma - 1)^2 E_{i,2} \end{aligned} \quad (3.13)$$

显然，这是一个关于 γ 的函数。为了最大程度的减少 D_i ，令：

$$\frac{dD_i}{d\gamma} = 0 \quad (3.14)$$

可以得到：

$$\gamma = \frac{\sqrt{\alpha E_{i1} E_{i2}} + E_{i1} E_{i2}}{\alpha E_{i2} + E_{i1} E_{i2}} \quad (3.15)$$

同理基于公式 (4.11)，可以得到：

$$\beta = \frac{\alpha E_{i2} + E_{i2} \sqrt{\alpha E_{i1} E_{i2}}}{\alpha E_{i2} + E_{i1} E_{i2}} \quad (3.16)$$

显然，公式 (4.15) 和公式 (4.16) 匹配公式 (4.5)。在本节中同样设计了使用与不使用该优化策略的对比实验。实验中以 SNR 和 ODG 作为评价指标，对不同嵌入强度下水印音频的感知质量进行了系统分析。如图 4.3 所示，选用了涵盖多种风格的 25 段音频样本进行测试。实验结果清晰表明，引入优化机制后，水印音频在 SNR 与 ODG 两项指标上均获得了不同程度的改善。尤其是在高嵌入强度场景下，该优化策略能够显著缓解因嵌入操作引入的能量畸变，从而有效提升整体感知质量。

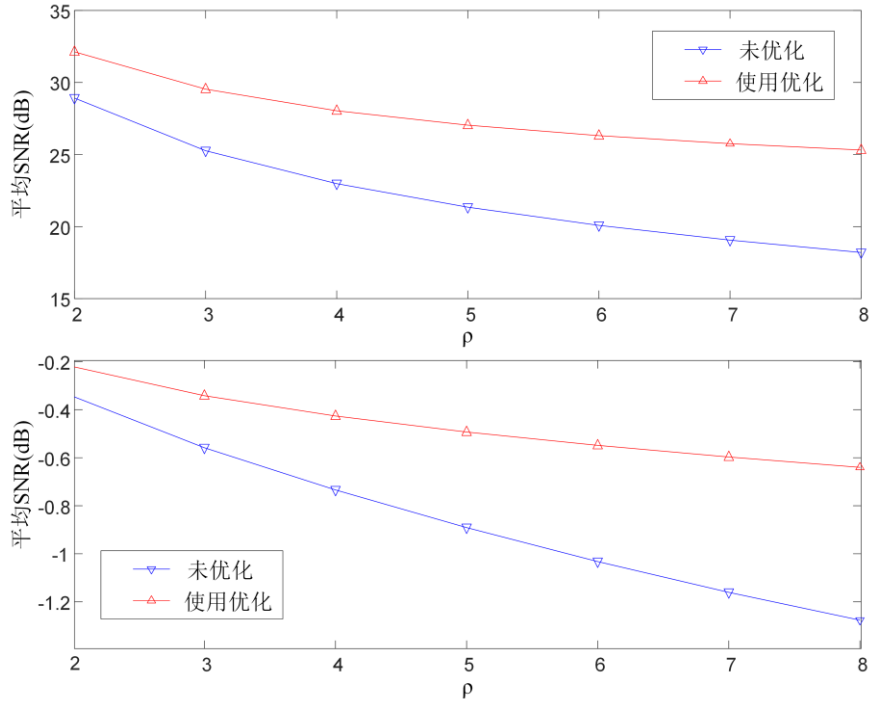


图 4.3 有优化与无优化的平均 SNR 和 ODG 值

Figure 4.3 The mean SNRs and ODGs with and without optimization

4.2.5 鲁棒性分析

本节将从理论层面对本算法在面对常见信号处理攻击及去同步攻击时的鲁棒性进行分析。

当 $w_i' = 0$ 时, 由公式 (4.3) 可知嵌入水印后帧内的关系为 $\frac{E_{i,1}'}{E_{i,2}'} \geq \alpha$, 不失一般

性, 本节以 $\frac{E_{i,1}'}{E_{i,2}'} = \alpha$ 进行鲁棒性的理论分析。经过各种攻击后, 帧内的关系可以

建模为:

$$\frac{E_{i,1}''}{E_{i,2}''} = \frac{E_{i,1}' + \Delta_1}{E_{i,2}' + \Delta_2} \quad (3.17)$$

其中, $\Delta_1 = E_{i,1}'' - E_{i,1}'$, $\Delta_2 = E_{i,2}'' - E_{i,2}'$ 。显然所提出方法本质上对于闭环攻击具有抵抗性, 即 $\Delta_1 = \Delta_2 = 0$ 。低通滤波、MP3 压缩和 AAC 压缩攻击主要影响高频分量。

由于特征嵌入操作在中频范围内, 因而几乎不受此类失真的影响。因此, 可以假设 $\Delta_1 \approx \Delta_2 \approx 0$ 。对于幅度缩放攻击来说, 由于一帧内的系缩放的比例相同, 所以可以得到:

$$\frac{E''_{i,1}}{E''_{i,2}} = \frac{\sigma^2 E'_{i,1}}{\sigma^2 E'_{i,2}} = \frac{E'_{i,1}}{E'_{i,2}} \quad (3.18)$$

其中 σ 为缩放系数，显然本章方法的方法可以抵抗幅度缩放攻击。对于回声，重新采样，量化和噪声攻击来说，可以建模为：

$$\frac{E''_{i,1}}{E''_{i,2}} = \frac{\alpha E'_{i,2} + \Delta_1}{E'_{i,2} + \Delta_2} \quad (3.19)$$

为了保证水印能够正确地提取，根据公式（4.8），我们期望：

$$\frac{E''_{i,1}}{E''_{i,2}} \geq 1 \quad (3.20)$$

化简后得到：

$$E'_{i,2} \geq \frac{\Delta_1 - \Delta_2}{\alpha - 1} \quad (3.21)$$

Mai 等人^[60]认为，噪声瞬时功率谱在每个频率段内以及在足够短的时间段内 近似恒定。基于这一思想，我们认为 $\Delta_1 - \Delta_2 \approx 0$ 。其中， $E'_{i,2}$ 代表着一段内的能量信息，显然大于 $\Delta_1 - \Delta_2$ 。此外，还可以通过调整嵌入强度 α 来创建错误缓冲区，使上述不等式很容易被满足。因此，提出的方法对回声添加，重新采样，量化和噪声攻击具有鲁棒性。

对于 TSM 和 PSM 攻击来说，3.2.5 节提到，TSM 和 PSM 攻击可以看做沿频率轴的拉伸或收缩操作，导致未嵌入水印的系数可能错误地参与了提取过程，而某些嵌入式系数可能已被排除在外。本章提出的改进的缓冲补偿机制平滑了水印区域和非水印区域内的系数，使得 TSM 和 PSM 攻击造成的非水印区域参与的特征计算的影响减小。增大嵌入强度可以使得即使丢弃一些系数依然满足于公式（4.20）。因此，所提出的方法能够抵抗 TSM 和 PSM 攻击。在强度适中的抖动攻击情况下，可以假设 $\Delta_1 - \Delta_2 \approx 0$ 。这表明本方法同样对抖动攻击具有鲁棒性。对于裁剪攻击来说，所提出的对称同步码嵌入和检测的框架有两种特殊情况需要重点考虑：

（1）防止随机性：在裁剪攻击后，从对称位置提取的相等同步码数量应低于阈值。由于裁剪后提取的同步码可视为随机比特。为了减少随机性，通过增加对称同步码的数量来降低对称位置同步码总数的随机波动。

(2) 防止误判：对于可能改变帧长的攻击（如 TSM 和抖动攻击），必须确保从对称位置提取的相等同步码数量高于阈值，以避免误判为裁剪攻击并选择错误的检测方式。这依赖于本章的嵌入方法对这些攻击的鲁棒性。如 4.3.2 节的表 4.1 和表 4.2 所示，该方法在此类攻击下的准确率均超过 80%。因此，设计合理的阈值能够有效抵御裁剪攻击以及 TSM 和抖动攻击。

4.3 实验结果与分析

本节将展示并分析所提音频水印方法的实验结果。首先介绍关键参数的设置依据。随后，从不可感知性和鲁棒性两个方面对所提方法进行实验验证与性能评估，以全面反映其在实际应用中的有效性与稳定性。最后，通过消融实验进一步探讨核心模块的作用及其对整体性能的影响，从而验证方法设计的合理性。

4.3.1 水印隐蔽性测试

本方案选取 3.3.1 节所列的 25 段不同类型和相同长度的音乐片段作为测试音频，音频格式为 WAV，采样率为 44.1 kHz，量化精度为 16 位。在相同的嵌入强度下进行对比实验。为评估音频水印的不可感知性，采用 SNR 与 ODG 作为评估标准，以综合衡量感知质量。

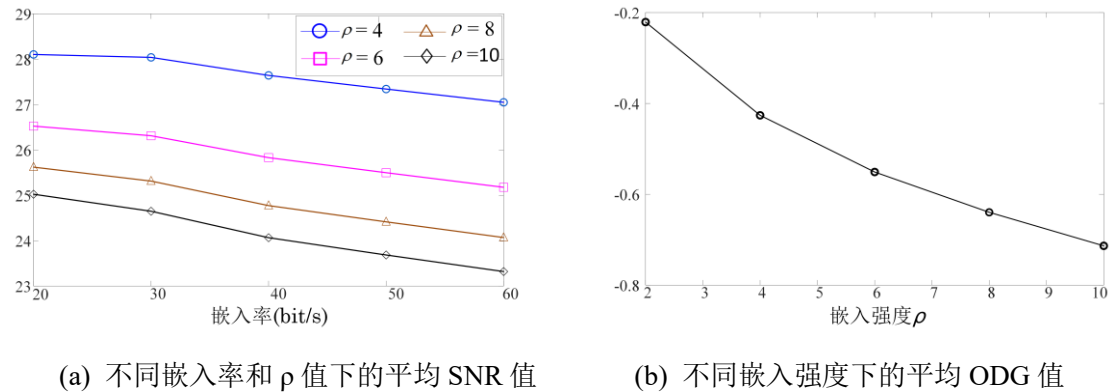


图 4.4 不同嵌入率与嵌入强度对音频质量的影响

Figure 4.4 Impact of embedding rate and strength on audio quality

为评估该方法在不同嵌入强度 ρ 下对应的 SNR，本节同样考察了该方法在多个嵌入强度下的性能表现。如图 4.4(a)所示，该方法在嵌入强度逐渐增加时，信噪比逐渐下降，但整体信噪比依然处于可接受范围。尤其在嵌入速率为 60 bps

的情况下，SNR 能够保持在高水平，表明其在保证容量的同时具备良好的信号保真度。

图 4.4(b)展示了在固定嵌入速率为 30 bps 的条件下，不同嵌入强度对应的 ODG 结果。从结果来看，随着嵌入强度的逐步增加，感知质量略有下降，ODG 值整体呈现出缓慢降低的趋势。然而，在所有测试配置下，ODG 值均未低于 -1。考虑到感知质量与鲁棒性之间存在一定的权衡关系，本方法在设计时重点平衡了两者。经实验验证，默认选取 30 bps 的嵌入速率和一组经验性嵌入强度参数，能够在保障鲁棒性的同时有效抑制感知失真。此外，虽然音频样本本身在风格、节奏及频谱结构上存在较大差异，该参数配置在大多数样本上仍表现出一致且优良的性能表现。因此，建议在实际应用中以此为默认参数，必要时可针对具体音频进行微调，以进一步优化嵌入效果。

对比图 3.6 和图 4.4 可知，该方法在感知质量方面略逊于第三章所提方案，但从后续第 4.3.2 节的鲁棒性实验结果可以看出，适度牺牲听觉感知性以提升鲁棒性是合理且值得的权衡。

4.3.2 水印鲁棒性测试

为了确保评估的公平性，下述实验沿用第 3.3.1 节中的攻击参数配置，将本章提出的方法与当前最先进的抵御去同步攻击方法[40]、方法[61]和三角调制方法进行对比。实验从常见的信号处理攻击和去同步攻击两个维度，分别评估两种方法在鲁棒性方面的表现。首先，在不嵌入同步码的前提下进行常见信号处理攻击实验，旨在验证第 4.2.5 节中所述误判情况的发生概率极低，从而进一步证明所提同步检测机制的可靠性。

（1）常见的信号处理攻击

基于第 2.4.1 节所述的 10 种常见信号处理攻击，表 4.1 展示了方法[40]方法[61]、三角调制方法与本章方法在 30 bps 嵌入速率下的鲁棒性测试结果。本实验选取了 25 段音频样本，在每种攻击下分别计算平均 BER 值和 NCC 的值，并将其作为评估各方法性能的综合指标。

表 4.1 在嵌入率为 30 bps 下, 不同方法对常见攻击的平均 BER 和平均 NCC 值
Table 4.1 The mean BER and mean NCC of different methods against common attacks at an embedding rate of 30 bps

普通攻击		文献[40]		文献[61]		三角调制		本文方法	
		BER	NCC	BER	NCC	BER	NCC	BER	NCC
无攻击		0.00	1.0000	0.00	1.0000	0.00	1.0000	0.00	1.0000
噪声 (dB)	30	0.11	0.9989	0.19	0.9980	0.39	0.9960	0.13	0.9987
	20	1.82	0.9807	1.96	0.9793	3.70	0.9641	1.45	0.9850
量化		0.11	0.9988	0.10	0.9989	0.27	0.9971	0.13	0.9987
幅度缩 放	0.8	0.00	1.0000	0.00	1.0000	0.00	1.0000	0.00	1.0000
	1.2	0.00	1.0000	0.00	1.0000	0.00	1.0000	0.00	1.0000
回声		1.24	0.9870	21.49	0.7743	2.91	0.9687	4.01	0.9600
MP3 (kbps)	128	0.00	1.0000	0.03	0.7743	0.00	1.0000	0.00	1.0000
	96	0.00	1.0000	0.08	0.9992	0.01	0.9999	0.00	1.0000
AAC (kbps)	128	0.00	1.0000	0.00	1.0000	0.00	1.0000	0.00	1.0000
	96	0.00	1.0000	0.00	1.0000	0.00	1.0000	0.00	1.0000
重采样		0.00	1.0000	0.16	0.9983	0.00	1.0000	0.00	1.0000
低通滤波		0.00	1.0000	1.17	0.9876	0.01	0.9999	0.00	1.0000
高通滤波		0.00	1.0000	42.46	0.5615	0.08	0.9991	0.00	1.0000

从表 4.1 中可以看出, 本文方法与文献[40]中的方法一样, 对常见的信号处理攻击表现出较强的鲁棒性, 并获得了较低的 BER。文献[61]的方法对噪声、低通滤波和重采样攻击的抵抗力较弱, 这是因为该方法基于 FDLN 残差特征, 因此对噪声和重采样十分敏感。此外, 该方法的特征提取自频带系数的前 75%, 包含了大量高频信息, 容易受到低通滤波的干扰。我们可以注意到, 与第三章中基于三角调制的方法相比, 可以明显看出本方法在鲁棒性方面有了显著提升, 这主要得益于本章中提出的改进分段策略。为确保音频水印算法的隐蔽性, 实验中选择的是中频范围内的系数作为水印嵌入的区域, 但这种隐蔽性约束了可用的嵌入区域。第三章所使用的三段式分段策略中, 由于每帧被划分为三段, 使得每段所包含的频域系数较少, 遭受攻击时所提取的特征易受干扰, 从而影响鲁棒性。相比之下, 采用两段式的分段方式能够在一定程度上缓解该问题, 提高嵌入特征的稳定性, 从而增强系统的抗攻击能力。

(2) 抖动, TSM 和 PSM 攻击

表 4.2 给出了在 30 bps 嵌入速率下, 针对三种去同步攻击, 对方法[40]方法[61]、三角调制方法与本章所提出方法进行鲁棒性评估的结果。从表中可以看出, 文献[40]与本章方法在抖动和 TSM 攻击下表现相近, 均取得了较低的 BER 值, 显著优于文献[61]所提出的方法。尽管文献[40]对抖动与 TSM 攻击具有一定鲁

棒性，但在面对 PSM 攻击时效果明显不足；而文献[61]尽管在 PSM 和 TSM 攻击下均具备一定抵抗能力，其误码率仍超过 20%，主要归因于其采用了较高的嵌入速率（30 bps）。相比之下，本章所提出的方法在不同强度的 TSM 和 PSM 攻击下均能保持较强鲁棒性，BER 始终低于 10%，充分体现了方法的鲁棒性。

表 4.2 在嵌入率为 30 bps 下，不同方法对去同步攻击的平均 BER 和平均 NCC 值
Table 4.2 The mean BER and mean NCC of different methods against desynchronization attacks at an embedding rate of 30 bps

去同步攻击		文献[40]		文献[61]		三角调制		本文方法	
		BER	NCC	BER	NCC	BER	NCC	BER	NCC
抖动(1%)		3.05	0.9674	31.04	0.5814	6.3	0.933	10.55	0.9022
TSM	80%	0.18	0.9981	23.30	0.7105	6.74	0.9344	1.24	0.9871
TSM	90%	0.12	0.9987	21.25	0.7395	4.12	0.9604	0.68	0.9929
TSM	110%	0.19	0.9980	20.36	0.7522	5.02	0.9531	1.93	0.9799
TSM	120%	0.17	0.9982	23.38	0.7079	6.30	0.9401	7.40	0.9228
PSM	80%	34.33	0.6113	31.19	0.5892	16.83	0.8287	7.03	0.9281
PSM	90%	49.72	0.4317	25.57	0.6771	8.05	0.9181	2.17	0.9776
PSM	110%	53.56	0.3840	21.01	0.7429	9.99	0.9009	3.71	0.9618
PSM	120%	35.69	0.6010	25.85	0.6715	15.36	0.8417	12.31	0.8721

（3）裁剪攻击

为验证本章所提出的对称同步码嵌入与检测策略的有效性，下述实验选取了不同比例的裁剪攻击（10%、15%、20%、25%、30%）对上述三种方法与本章方法进行鲁棒性对比评估。考虑到同步码组的加入对嵌入容量的影响，实验将

表 4.3 在 10 bps 嵌入速率下，不同方法针对裁剪攻击的平均 BER 与平均 NCC
Table 4.3 The mean BER and mean NCC of different methods against cropping attacks at an embedding rate of 10 bps

攻击类型		文献[40]		文献[61]		三角调制		本文方法	
		BER	NCC	BER	NCC	BER	NCC	BER	NCC
裁剪	10%	41.18	0.6509	51.68	0.3299	50.08	0.4644	0.00	1.0000
	15%	46.90	0.5388	52.22	0.3094	50.15	0.4651	0.00	1.0000
	20%	47.08	0.5368	49.12	0.3591	49.38	0.4763	0.00	1.0000
	25%	45.20	0.5548	50.54	0.3348	49.57	0.4791	0.00	1.0000
	30%	46.68	0.5376	50.98	0.3131	49.98	0.4710	0.00	1.0000

嵌入速率统一设置为 10 bps。从表 4.3 可观察到，文献[40]与文献[61]在面对裁剪攻击时均出现完全失败的情况。这是由于裁剪操作造成音频片段丢失，进而引发水印嵌入位置的偏移或丢失，严重影响水印的正确提取。而本章提出的方法通过引入对称同步码组，不仅能够有效检测裁剪攻击的发生，还能准确定位水印嵌入

位置。实验结果表明,即便在高达 30%的裁剪强度下,水印仍可被准确提取,充分验证了该方法在抵御裁剪攻击方面的鲁棒性。

4.3.3 消融实验

为了研究缓冲补偿的窗口长度对去同步攻击性能的影响已在 3.3.4 节进行了系统的分析,此节不再赘述。接下来将研究 3.3.4 节的缓冲补偿和本章提出的改进的缓冲补偿机制在隐蔽性和鲁棒性方面的区别。

表 4.4 $\xi = 250\text{Hz}$ 下缓冲补偿和改进的缓冲补偿的 SNR 和 ODG,并在 TSM 和 PSM 攻击下的 BER

Table 4.4 SNR, ODG, and BER under TSM and PSM Attacks for Buffer Compensation (BC) and Improved Buffer Compensation (IBC) at $\xi = 250\text{Hz}$

名称	SNR	ODG	TSM(80%)	TSM(120%)	PSM(80%)	PSM(120%)
BC	24.18	-0.7944	1.24	7.41	7.03	12.31
IBC	25.31	-0.6401	1.03	5.71	5.58	8.76

表 4.5 $\xi = 500\text{Hz}$ 下缓冲补偿和改进的缓冲补偿的 SNR 和 ODG,并在 TSM 和 PSM 攻击下的 BER

Table 4.5 SNR, ODG, and BER under TSM and PSM Attacks for Buffer Compensation (BC) and Improved Buffer Compensation (IBC) at $\xi = 500\text{Hz}$

名称	SNR	ODG	TSM(80%)	TSM(120%)	PSM(80%)	PSM(120%)
BC	22.55	-1.6464	1.09	4.25	3.43	4.36
IBC	24.62	-0.7982	1.1	5.76	5.42	8.19

如表 4.4 和表 4.5 所示,在参数 ξ 等于 250Hz 和 500Hz 的设置下,对采用 BC 和 IBC 后嵌入水印的音频进行评估,分别比较其平均 SNR 与平均 ODG,并对嵌入水印的音频信号施加多种类型的攻击以评估其鲁棒性。从表 4.4 和表 4.5 可见,尽管第 3.3.4 节所提出的原始缓冲补偿机制在 SNR 上可达到 20 dB 以上,但其 ODG 值已远低于-1,音质劣化明显,易被人耳感知。相比之下,改进后的缓冲补偿机制不仅保持了 20 dB 以上的高 SNR,同时 ODG 值也优于-1,听感显著改善,具有更好的音频感知质量。此外,从表 4.4 和表 4.5 中可以观察到,在 TSM 攻击下,两种缓冲补偿机制的 BER 差异不大;而在 PSM 攻击下,尽管存在一

定差别, BER 的差距仍控制在 4% 以内。综上实验结果表明, 改进后的缓冲补偿机制在显著提升音频感知质量的同时, 依然保持了良好的鲁棒性。

4.4 本章小结

本章针对裁剪攻击导致的同步失效问题, 提出了一种基于能量调制与对称同步码检测机制的鲁棒音频水印算法。该方法通过在帧内选取两个不相邻子段的中频能量关系嵌入水印信息, 在增强水印嵌入稳定性的同时, 有效降低了对音频感知质量的影响。为了提升系统在裁剪攻击下的同步恢复能力, 设计了结构对称的同步码冗余嵌入机制, 并引入滑动窗口检测策略, 实现了水印区域的快速定位与鲁棒提取。此外, 算法还引入线性递减缓冲补偿机制, 进一步平滑了水印区域与非水印区域之间的频谱边界, 显著提升了听觉感知质量。实验结果表明, 所提算法不仅在常规信号处理攻击下表现稳定, 在面对裁剪、TSM、PSM 等典型去同步攻击时仍具备优异的鲁棒性, 验证了本章方法在实际复杂环境中的实用价值与技术优势。

第五章 总结与展望

5.1 全文总结

本论文围绕音频水印技术在面对去同步攻击（TSM、PSM 与裁剪攻击）时的鲁棒性问题展开了深入研究。现有技术在应对此类攻击时，常因同步信息被破坏而导致水印提取失败，尤其裁剪攻击因直接造成信息丢失而被视为最具挑战性的难题之一。为解决这些问题，本文结合变换域拼凑策略与帧内能量关系调制的思想，设计并实现了两种具有高鲁棒性与良好不可感知性的音频水印算法，并通过大量实验验证了所提方法在多种攻击场景下的有效性和优越性。

首先，针对现有音频水印方法在 TSM 与 PSM 攻击下鲁棒性不足的问题，本文在第三章中提出了一种基于三角能量调制与缓冲补偿机制的鲁棒音频水印方法。该方法将每个音频帧划分为三个不相交的子段，并提取其 DCT 域中频段的平均能量作为稳定特征。通过构建这三段能量之间的三角几何关系来嵌入水印：当三段能量可构成等边三角形时表示比特“1”，不能构成三角形时表示比特“0”。为提升隐蔽性，设计了最小化失真功率的优化策略。更重要的是，为应对 TSM/PSM 攻击导致的频率线性缩放，算法引入了缓冲补偿机制，通过平滑嵌入区域与非嵌入区域的频谱边界，有效降低了特征提取误差。实验结果表明，该方法在抵抗常规信号处理攻击、TSM 和 PSM 攻击时均表现出优越的鲁棒性和听觉透明性。

其次，为进一步解决鲁棒性与不可感知性的平衡问题，并重点增强对裁剪攻击的抵抗能力，本文在第四章提出了一种结合双段式帧内能量调制与对称同步码检测机制的改进算法。该算法将每帧划分为两个子段，通过调制两者间的能量关系嵌入水印，提升了特征的稳定性。其核心创新在于设计了对称同步码的嵌入与检测机制：通过在音频信号的对称位置嵌入相同的同步码组，提取时首先比对同步码的完整性，从而准确判断音频是否遭受裁剪攻击。若检测到裁剪，则启用固定长度的滑动窗口进行水印定位与提取；反之，则按常规模式提取。此外，该算法还引入了线性递减的缓冲补偿策略，进一步缓解了补偿机制造成的听觉感知质

量下降。实验证明,该同步码检测机制的有效性,即使在高达 30% 的随机裁剪攻击下,水印信息仍可被零误码率地准确恢复,成功解决了裁剪攻击下的同步失效难题。

综上所述,本文所提出的两种算法均能兼顾水印的不可感知性与鲁棒性,在当前主流攻击模型下表现出良好的综合性能,为鲁棒音频水印技术的发展提供了新的设计思路与实现路径。

5.2 未来研究展望

尽管本文在抵御去同步攻击的鲁棒音频水印方面取得了一定研究成果,但仍存在一些有待改进之处,未来研究可从以下几个方面进一步拓展和完善:

(1) 鲁棒性与容量的权衡优化:目前方案在保障鲁棒性的同时,嵌入速率仍受到同步机制和水印结构的限制。后续研究可尝试结合压缩感知、冗余编码等方法,进一步提升水印嵌入容量的同时保持良好的不可感知性与鲁棒性。

(2) 复杂攻击组合环境下的适应能力提升:本文主要针对单一攻击或典型去同步攻击进行测试,而实际应用中音频信号常常面临多种攻击叠加的复杂环境,未来可重点研究多重攻击情境下的协同防御机制与鲁棒特征提取策略。

(3) 智能化水印提取与识别机制:目前水印提取主要基于固定规则与阈值判断,后续可尝试引入机器学习、深度神经网络等智能技术,实现对不同攻击模式下水印信息的自适应判别与恢复,提高提取的鲁棒性与准确率。

(4) 立体声音频环境下的适应性研究:目前研究主要针对单通道(单声道)音频信号进行水印嵌入与提取,然而在实际应用中,立体声音频被广泛使用,其通道间的差异性以及混合方式对水印的稳健性和提取准确性带来新的挑战。未来可针对立体声结构,设计适应多通道特性的水印嵌入策略与同步机制,以提升算法在复杂声场下的稳定性与鲁棒性。

通过以上研究的持续推进,有望进一步拓展鲁棒音频水印技术的应用,推动其在数字版权保护、音频内容追踪及智能媒体管理等领域的广泛应用。

参考文献

- [1] Shrivass S, Goyal J, Maurya V. The Evolution and Impact of Audiobooks in the Digital Age[M]. New York: Routledge, 2024.
- [2] Liu Z, Huang J, Sun X, et al. A security watermark scheme used for digital speech forensics[J]. Multimedia Tools and Applications, 2017, 76(7): 9297-9317.
- [3] Shahriar M R, Sangjin C, Sangbock C, et al. A High-capacity Audio Watermarking Scheme in the Time Domain Based on Multiple Embedding[J]. IETE Technical Review, 2013, 30(4): 286-294.
- [4] Sudler H. Effectiveness of anti-piracy technology: Finding appropriate solutions for evolving online piracy[J]. Business Horizons, 2013, 56(2): 149-157.
- [5] Zhang Y. Digital Watermarking Technology: A Review[C]//Proceedings of the 2009 ETP International Conference on Future Computer and Communication. Wuhan, China, 2009: 250-252.
- [6] Boney L, Tewfik A H, Hamdy K N. Digital watermarks for audio signals[C]//Proceedings of the 8th European Signal Processing Conference (EUSIPCO-96). Trieste, Italy: Edizioni LINT, 1996: 1-4.
- [7] Yuan X C, Pun C M, Philip Chen C L. Robust Mel-Frequency Cepstral coefficients feature detection and dual-tree complex wavelet transform for digital audio watermarking[J]. Information Sciences, 2015, 298: 159-179.
- [8] Chernock R, Gómez-Barquero D, Whitaker J, et al. ATSC 3.0 Next Generation Digital TV Standard—An Overview and Preview of the Issue[J]. IEEE Transactions on Broadcasting, 2016, 62(1): 154-158.
- [9] Li B, Chen J, Xu Y, et al. DRAW: Dual-Decoder-Based Robust Audio Watermarking Against Desynchronization and Replay Attacks[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 6529-6544.
- [10] Cvejic N, Seppanen T. Increasing robustness of LSB audio steganography using a novel embedding method[C]//Proceedings of the 2004 International Conference on Information Technology: Coding and Computing (ITCC 2004). Las Vegas, NV, USA: IEEE Computer Society, 2004: 533-537.

- [11] Xiong Y, Ming Z X. Covert Communication Audio Watermarking Algorithm Based on LSB[C]//Proceedings of the 2006 International Conference on Communication Technology (ICCT 2006). Guilin, China: IEEE, 2006: 1-4.
- [12] Cho J W, Park H J, Huh Y, Chung H Y, Jung H Y. Echo Watermarking in Sub-band Domain[C]//Kalker T, Cox I, Ro Y M, eds. Digital Watermarking, IWDW 2003. Berlin, Heidelberg: Springer, 2004: 399-412.
- [13] Hyoung Joong K, Yong Hee C. A novel echo-hiding scheme with backward and forward kernels[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(8): 885-889.
- [14] Byeong-Seob K, Nishimura R, Suzuki Y. Time-spread echo method for digital audio watermarking[J]. IEEE Transactions on Multimedia, 2005, 7(2): 212-221.
- [15] Xiang Y, Peng D, Natgunanathan I, et al. Effective Pseudonoise Sequence and Decoding Function for Imperceptibility and Robustness Enhancement in Time-Spread Echo-Based Audio Watermarking[J]. IEEE Transactions on Multimedia, 2011, 13(1): 2-13.
- [16] Kirovski D, Malvar H S. Spread-spectrum watermarking of audio signals[J]. IEEE Transactions on Signal Processing, 2003, 51(4): 1020-1033.
- [17] Attari A A, Shirazi A A B. Robust and Transparent Audio Watermarking based on Spread Spectrum in Wavelet Domain[C]//Proceedings of the 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT 2019). Amman, Jordan: IEEE, 2019: 366-370.
- [18] Xiang S, Huang J. Histogram-Based Audio Watermarking Against Time-Scale Modification and Cropping Attacks[J]. IEEE Transactions on Multimedia, 2007, 9(7): 1357-1372.
- [19] Xiang S, Yang L, Wang Y. Robust and Reversible Audio Watermarking by Modifying Statistical Features in Time Domain[J]. Advances in Multimedia, 2017, 2017(1): 8492672.
- [20] Hu H-T, Hsu L-Y. Robust, transparent and high-capacity audio watermarking in DCT domain[J]. Signal Processing, 2015, 109: 226-235.
- [21] Saadi S, Merrad A, Benziane A. Novel secured scheme for blind audio/speech norm-space watermarking by Arnold algorithm[J]. Signal Processing, 2019, 154: 74-86.

- [22] Fallahpour M, Megías D. Audio Watermarking Based on Fibonacci Numbers[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(8): 1273-1282.
- [23] Wang X, Wang P, Zhang P, et al. A norm-space, adaptive, and blind audio watermarking algorithm by discrete wavelet transform[J]. Signal Processing, 2013, 93(4): 913-922.
- [24] Karajeh H, Khatib T, Rajab L, et al. A robust digital audio watermarking scheme based on DWT and Schur decomposition[J]. Multimedia Tools and Applications, 2019, 78(13): 18395-18418.
- [25] Jiang W, Huang X, Quan Y. Audio watermarking algorithm against synchronization attacks using global characteristics and adaptive frame division[J]. Signal Processing, 2019, 162: 153-160.
- [26] Wei L, Xiangyang X, Peizhong L. Localized audio watermarking technique robust against time-scale modification[J]. IEEE Transactions on Multimedia, 2006, 8(1): 60-69.
- [27] Wang S, Yuan W, Zhang Z, et al. Synchronous Multi-Bit Audio Watermarking Based on Phase Shifting[C]//Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021). Toronto, Canada: IEEE, 2021: 2700-2704.
- [28] Bender W, Gruhl D, Morimoto N, et al. Techniques for data hiding[J]. IBM Systems Journal, 1996, 35(3.4): 313-336.
- [29] Su Z, Zhang G, Yue F, et al. SNR-Constrained Heuristics for Optimizing the Scaling Parameter of Robust Audio Watermarking[J]. IEEE Transactions on Multimedia, 2018, 20(10): 2631-2644.
- [30] 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, 2001, 47(4): 1423-1443.
- [31] Erfani Y, Pichevar R, Rouat J. Audio Watermarking Using Spikegram and a Two-Dictionary Approach[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(4): 840-852.
- [32] Khaldi K, Boudraa A O. Audio Watermarking Via EMD[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(3): 675-680.

- [33] Hwang M J, Lee J, Lee M, et al. SVD-Based Adaptive QIM Watermarking on Stereo Audio Signals[J]. IEEE Transactions on Multimedia, 2018, 20(1): 45-54.
- [34] Zhao X, Guo Y, Liu J, et al. Quantization Index Modulation audio watermarking system using a psychoacoustic model[C]//Proceedings of the 2011 8th International Conference on Information, Communications & Signal Processing (ICICS 2011). Singapore: IEEE, 2011: 1-4.
- [35] Shaoquan W, Jiwu H, Daren H, et al. Efficiently self-synchronized audio watermarking for assured audio data transmission[J]. IEEE Transactions on Broadcasting, 2005, 51(1): 69-76.
- [36] Jiang W, Huang X, Quan Y. Audio watermarking algorithm against synchronization attacks using global characteristics and adaptive frame division[J]. Signal Process, 2019, 162(C): 153–160.
- [37] Natgunanathan I, Xiang Y, Rong Y, et al. Robust Patchwork-Based Embedding and Decoding Scheme for Digital Audio Watermarking[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(8): 2232-2239.
- [38] In-Kwon Y, Hyoung-Joong K. Modified Patchwork Algorithm: a novel audio watermarking scheme[C]//Proceedings of the 2001 International Conference on Information Technology: Coding and Computing (ITCC 2001). IEEE Computer Society Press, 2001: 237-242.
- [39] Xiang Y, Natgunanathan I, Guo S, et al. Patchwork-Based Audio Watermarking Method Robust to De-synchronization Attacks[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(9): 1413-1423.
- [40] Natgunanathan I, Xiang Y, Hua G, et al. Patchwork-Based Multilayer Audio Watermarking[J]. IEEE/ACM Trans Audio, Speech and Lang Proc, 2017, 25(11): 2176–2187.
- [41] Zhao J, Zong T, Xiang Y, et al. Desynchronization Attacks Resilient Watermarking Method Based on Frequency Singular Value Coefficient Modification[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 2282-2295.
- [42] Li J, Xiang S. Audio-lossless robust watermarking against desynchronization attacks[J]. Signal Processing, 2022, 198: 108561.

- [43] Zhang G, Zheng L, Su Z, et al. M-Sequences and Sliding Window Based Audio Watermarking Robust Against Large-Scale Cropping Attacks[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 1182-1195.
- [44] 陈雪松, 李昊天, 贾瑞成, 等. 一种基于混沌加密的 DWT 数字音频水印算法[J]. 计算机与现代化, 2014, (08): 46 – 49, 74.
- [45] 朱媛媛, 张小红, 崔智勇. 基于 BCH 纠错编码技术的小波域音频数字水印研究[J]. 江西理工大学学报, 2009, 30(03): 45-48, 73.
- [46] Van Schyndel RG, Tirkel AZ, Osborne CF. A Digital Watermark[C]//Proceedings of the 1994 IEEE International Conference on Image Processing (ICIP 1994). Austin, TX, USA: IEEE, 1994: 86–90.
- [47] 杨恒欢, 常学义, 荆锐, 等. 一种新颖的双声道数字音频无源水印算法[J]. 上海第二工业大学学报, 2011, 28(1): 65-69.
- [48] Gulbis M, Müller E, Steinebach M. Content-Based Authentication Watermarking with Improved Audio Content Feature Extraction[C]//Proceedings of the 2008 Fourth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2008). Harbin, China: IEEE Computer Society, 2008: 620–623.
- [49] Lei B Y, Soon I Y, Li Z. Blind and robust audio watermarking scheme based on SVD–DCT[J]. Signal Processing, 2011, 91(8): 1973-1984.
- [50] Depovere G, Kalker T, Haitsma J, Maes M, De Strycker L, Termont P, Vandewege J, Hudson A, et al. The VIVA Project: Digital Watermarking for Broadcast Monitoring[C]//Proceedings of the 1999 IEEE International Conference on Image Processing (ICIP-99). Kobe, Japan: IEEE, 1999: 202-205.
- [51] Liu L, Guan T, Zhang Z. Broadcast monitoring protocol based on secure watermark embedding[J]. Computers & Electrical Engineering, 2013, 39(7): 2299-2305.
- [52] Satyanarayana B V V, Sudhir B, Kranthi Kumar G, Srinivasarao B N. Audio Watermarking Implementation – Covert Communication[J]. International Journal of Emerging Communication Technologies, 2011, 2(Spl-1): 48.
- [53] Shelke R D, Nemade M U, et al. Audio watermarking techniques for copyright protection: A review[C]//Proceedings of the 2016 International Conference on

- Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC 2016). Jalgaon, India: IEEE, 2016: 634–640.
- [54] Hunt B R, Li T Y, Kennedy J A, Nusse H E, eds. The Theory of Chaotic Attractors[M]. New York: Springer, 2004.
- [55] 曹军梅. 一种基于LWT和Logistic混沌映射的自适应音频水印算法[J]. 智能计算机与应用, 2017, 7(03): 65-67.
- [56] Lin Y, Abdulla WH, Ma Y, et al. Audio Watermarking Detection Resistant to Time and Pitch Scale Modification[C]//Proceedings of the 2007 IEEE International Conference on Signal Processing and Communications (ICSPC 2007). Honolulu, USA: IEEE, 2007: 209–215.
- [57] Chen S, Malik A, Zhang X, et al. A Fast Method for Robust Video Watermarking Based on Zernike Moments[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(12): 7342-7353.
- [58] Li J-f, Wang H-X, Wu T, et al. Norm ratio-based audio watermarking scheme in DWT domain[J]. Multimedia Tools and Applications, 2018, 77(12): 14481-14497.
- [59] Liu X, Li X, Shi C, et al. A novel SVD-based adaptive robust audio watermarking algorithm[J]. Multimedia Tools and Applications, 2024, 83(27): 69443-69465.
- [60] Mai V K, Pastor D, Aïssa-El-Bey A, et al. Robust Estimation of Non-Stationary Noise Power Spectrum for Speech Enhancement[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(4): 670-682.
- [61] Liu Z, Huang Y, Huang J. Patchwork-Based Audio Watermarking Robust Against De-Synchronization and Recapturing Attacks[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(5): 1171-1180.
- [62] Liu C, Zhang J, Fang H, et al. DeAR: A Deep-Learning-Based Audio Re-recording Resilient Watermarking[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(11): 13201–13209.
- [63] Liu W Z, Li Y, Lin D D, et al. GROOT: Generating Robust Watermark for Diffusion-Model-Based Audio Synthesis[J]. Proceedings of the 32nd ACM International Conference on Multimedia, 2024, 47(10): 1234–1245.

攻读硕士学位期间取得的研究成果

一、论文

- [1] Zhu W, Zhou Y, Wu D, Zhao G, Dong Z, Ye J, Wu H. Desynchronization Resilient Audio Watermarking Based on Adaptive Energy Modulation[J]. *Mathematics*, 2025, 13(17): 2736. (SCI 收录, WOS:001569904700001)

二、专利

- [1] 朱伟南, 吴汉舟. 一种抵御去同步攻击的自媒体音频数据版权保护装置: 审
批编号 2025-1248, 上海大学, 2025. (已提交)

致 谢

岁月不居，时节如流。至此，我的硕士生涯也将画上圆满的句号。从最初的懵懂无知，到如今顺利完成毕业论文，这一路充满了探索的艰辛与收获的喜悦。这段旅程不仅是一场求知的修行，更是一幅由师长的教诲、家人的关爱和朋友的陪伴共同勾勒的画卷。在此落笔之时，我满怀感激与敬意，谨向所有关心和帮助过我的人致以最诚挚的谢意。

首先，我要由衷感谢我的导师吴汉舟老师。您如明灯般，用睿智的学术指导和宽广的视野为我指引方向；亦如园丁般，以严谨的治学态度和细致的耐心雕琢着我的学术成长。在研究过程中，无论是思路的启发，还是方法的改进，您的点拨总让我受益匪浅。没有您的辛勤付出与无私帮助，就没有这篇论文的完成，更没有我今日的收获与成长。

同时，我要感谢我的课题组成员和实验室伙伴们，尤其是吴德阳师兄、陈诗怡师姐、赵葛剑师兄、何承桦、杨之光、林丽娜、史景辉、黄超越师弟。在实验数据采集、技术难题攻克以及论文讨论的过程中，你们的热心协作与真诚建议为我提供了莫大的助力。正因有你们的陪伴，这段充满挑战的研究生生涯多了一份温暖与力量。

感谢本科及研究生期间结识的挚友们。无论是闲暇时的谈笑，还是低谷时的安慰，你们的陪伴让我在学术之外感受到生活的美好。正是这些点滴，构成了我人生中不可或缺的亮色。

最深的感激，留给我的家人。无论我远在求学的他乡，还是因课题挑战而倍感压力，你们始终是最坚实的后盾。奶奶的叮嘱、母亲的关怀、父亲的教诲，给予了我无尽的勇气与支持，让我一次次突破自我、迎接挑战。你们的理解和鼓励，是我成长路上不可或缺的力量。这份学位不仅属于我，更属于一直守护和陪伴我的家人。

无论遇到怎样的困难与挑战，我都会倾尽所有热情，在自我救赎中奋力前行。我始终清楚自己想要什么，也明白自己是谁，我就是我，颜色各异的烟火。希望

未来的我，依旧能够保持这颗少年心，向往自由，常怀热情，勇敢做自己想做的事。

回头看，轻舟已过万重山；

向前看，长路漫漫亦璀璨。

“既而，别之。朝花夕拾言笑安晏，万物皆流，唯情旦旦。”

朱伟南

上海大学宝山校区

2025 年 09 月 26 日