

学校代码 : 10699  
分 类 号 :  
密 级 :  
学 号 : 2018200480

题目 增强现实中  
双耳渲染技术研究

作者 陈洁茹

学 科 专 业 信号与信息处理

指 导 教 师 张雯 陈景东

申请学位日期 2021 年 3 月



西北工业大学  
硕士学位论文  
(学位研究生)

题目：增强现实中  
双耳渲染技术研究

作者：陈洁茹

学科专业：信号与信息处理

指导教师：张雯 陈景东

2021年3月



**Title: Binaural Rendering Technology In Augmented Reality**

**By  
Jieru Chen**

**Under the Supervision of Professor  
Wen Zhang , Jingdong Chen**

A Dissertation Submitted to  
Northwestern Polytechnical University

In partial fulfillment of the requirement  
For the degree of  
Doctor of Signal and Information Processing

Xi'an P. R. China  
March 2021



## 摘要

声学场景的双耳渲染旨在给听众提供身临其境的体验，是虚拟现实、增强现实、沉浸式多媒体产品、虚拟声学等领域的重要研究课题。最近，基于球谐分解的双耳渲染算法得到了广泛关注。该算法对声场和头相关传递函数进行球谐函数展开，并且直接在球谐域进行渲染，避免了虚拟扬声器数目和位置对听感的影响。但是此类算法存在声场和头相关传递函数不匹配的问题，例如通常录音设备的尺寸小于人头尺寸，声场分解阶次（通常小于 5 阶）远小于头相关传递函数的分解阶次（34 阶）。这些不匹配问题导致定位线索的损伤、空间感的降低和音色的改变。

本论文主要针对以上问题，从录制声场和头相关传递函数两方面出发，对原始算法加以改进。主要的研究内容包括以下四个方面：

1. 深入研究了基于球谐分解的双耳渲染算法及其存在的问题。对声场和头相关传递函数的球谐分解及系数获取方法进行研究，针对空心球和刚性球展开讨论，给出了正确获取声场系数所需的麦克风数目，以及两种球阵所获取的声场系数的区别。
2. 研究了头相关传递函数的预处理方法，并通过实验验证了该方法的有效性。该方法采用与频率相关的相位对准，保留低频相位信息的同时对高频的相位加以修正。和直接使用最小二乘方法求解头相关传递函数的球谐分量相比，该预处理方法一方面可以准确地表示头相关传递函数的幅度谱，另一方面有效降低了头相关传递函数的分解阶次。
3. 提出了一种声场扩阶理论和一种新的双耳渲染算法。该理论对麦克风阵列的采集声场进行分析，将入射声场分解为直达波和混响场的叠加。根据直达波入射方向进行空间加窗处理，可以实现声场分量的阶次提升，同时扩大控制区域的半径。并且本研究将声场扩阶理论、头相关传递函数预处理和基于球谐分解的双耳渲染算法相结合，实现了基于声场扩阶的双耳渲染算法。
4. 在消声室和混响环境下，从双耳时间差、双耳声级差和双耳听觉互相关系数这三个评价指标出发，将本文所提出的算法与对标算法进行对比和分析，验证本算法的有效性。实验结果表明，在不同的混响情况下、声源位于不同方向时，本文所提出的双耳渲染算法均能达到较好的效果，优于现有算法。并且对空心球和刚性球在不同双耳渲染算法下的性能进行对比，验证了刚性球的优越性。

**关键字：**双耳渲染、头相关传递函数、空间音频、球谐分解、声场扩阶

本文研究得到 XX 基金(编号:XXXXXXX)资助。



## Abstract

Binaural rendering of acoustic scenes aims to provide the audience with an immersive experience. It is an important research topic in the fields of virtual reality, augmented reality, immersive multimedia products, and virtual acoustics. Recently, binaural rendering algorithms based on spherical harmonic decomposition have received much attention. This algorithm performs spherical harmonic expansion of sound field and head-related transfer function( HRTF ), and renders directly in spherical harmonic domain, avoiding the influence of the number and position of virtual speakers on hearing sense. However, there are some mismatches between the sound field and the head-related transfer function in this kind of algorithm. For example, the size of the recording equipment is usually smaller than the head size, and the decomposition order of the sound field (usually less than 5 order) is much smaller than the decomposition order of the head-related transfer function ( 34 order). These mismatches lead to impairments of localization cues, loss of spaciousness, and changes in timbre.

Aiming at the above problems, this paper improves the original algorithm from two aspects, sound field and head-related transfer function. The main research contents include the following four aspects:

1. The binaural rendering algorithm based on spherical harmonic decomposition and its problems are studied in detail. The spherical harmonic decomposition and coefficient calculation method of sound field and HRTF are studied, and the number of microphones which is needed to acquire sound field coefficient correctly. The difference of sound field coefficient obtained by open and rigid spherical array are discussed.

2. The preprocessing method of head-related transfer function is studied and the effectiveness of this method is verified by experiments. In this method, the frequency-dependent phase alignment is adopted to retain the phase information of low frequency and modify the phase information of high frequency. Compared with the least square method, this method can not only accurately represent the magnitude spectrum of the head-correlation transfer function, but also effectively reduce the decomposition order of the head-correlation transfer function.

3. A new theory for enlarging the order of sound field and a new binaural rendering algorithm are proposed. In this theory, the incident sound field is decomposed into the superposition of direct component and diffuse field. According to the direction of direct component, spatial masking can improve the order of the sound field and enlarge the radius of the control area.

In addition, a new binaural rendering algorithm is realised by combining this theory , the frequency-dependent phase alignment preprocessing method of the head-correlation transfer function and the binaural rendering algorithm based on spherical harmonic decomposition.

4. In anechoic chamber and reverberation environment, the algorithm proposed in this paper is compared and analyzed from three evaluation indexes , interaural time difference , interaural level difference and interaural cross-correlation. The experimental results show that the binaural rendering algorithm proposed in this paper can achieve better results under different reverberation environments and sound source directions, which is superior to the existing algorithms. The performance of open and rigid spherical array under different binaural rendering algorithms is compared to verify the superiority of rigid ball.

**Keywords :** Binaural rendering, head-related transfer function, spatial audio, spherical harmonic decomposition, sound field amplification

# 目录

摘要 .....	I
Abstract .....	III
目录 .....	VI
第一章 绪论 .....	1
1.1 研究背景 .....	1
1.2 国内外研究现状 .....	2
1.2.1 双耳录音 .....	3
1.2.2 双耳合成 .....	5
1.3 研究内容与论文结构 .....	9
1.3.1 研究内容及创新点 .....	9
1.3.2 论文结构安排 .....	9
第二章 双耳渲染算法 .....	11
2.1 虚拟听觉重放技术 .....	11
2.1.1 头相关传递函数 HRTF .....	11
2.1.2 基于 HRIR 和 BRIR 的虚拟听觉方法 .....	12
2.2 基于 Ambisonics 的虚拟扬声器方法 .....	13
2.2.1 Ambisonics 编码 .....	14
2.2.2 Ambisonics 解码 .....	17
2.2.3 虚拟扬声器方法 .....	18
2.3 基于球谐分解的双耳渲染算法 .....	19
2.4 本章小节 .....	21
第三章 基于 HRTF 球谐分解的双耳渲染算法 .....	23
3.1 球形麦克风阵列的声场录制及球谐系数估计 .....	23
3.1.1 空心球的球谐系数估计 .....	24
3.1.2 刚性球的球谐系数估计 .....	25
3.2 HRTF 的球谐分解 .....	27
3.2.1 HRTF 的获取 .....	27
3.2.2 HRTF 的球谐系数估计 .....	29
3.3 HRTF 预处理 .....	31
3.3.1 预处理方法 .....	31

---

3.3.2 实验结果与分析 .....	33
3.4 本章小结 .....	34
第四章 基于声场扩阶的双耳渲染算法 .....	35
4.1 信号模型 .....	35
4.2 声场扩阶原理 .....	37
4.2.1 有界线性算子 .....	37
4.2.2 空域加窗 .....	38
4.2.3 空域加窗后的声场阶次 .....	40
4.3 空域窗函数 .....	41
4.4 整体算法框架 .....	43
4.5 本章小结 .....	45
第五章 实验结果及分析 .....	47
5.1 评价指标 .....	47
5.1.1 双耳时间差 .....	47
5.1.2 双耳声级差 .....	50
5.1.3 双耳相关性 .....	51
5.2 实验结果及分析 .....	51
5.2.1 消声室环境下各算法性能对比 .....	52
5.2.2 空心球与刚性球的性能对比 .....	54
5.2.3 混响环境下各算法性能对比 .....	58
5.3 本章小结 .....	61
第六章 总结与展望 .....	63
6.1 全文总结 .....	63
6.2 工作展望 .....	64
参考文献 .....	65
致谢 .....	71
攻读硕士学位期间发表的学术论文和参加科研情况 .....	73

# 第一章 绪论

## 1.1 研究背景

如今，虚拟现实（Virtual Reality, VR）和增强现实（Augmented Reality, AR）等沉浸式技术正在快速发展，一定程度上改变了个人、企业与数字世界的互动方式，并给人们带来体验全新世界的可能。增强现实真实场景的基础上，叠加计算机模拟产生的虚拟物体、场景，呈现给用户更具沉浸感的新体验。一个完整的身临其境的体验需要将视觉、听觉、触觉、嗅觉、味觉等多种感官融合到一个场景中。目前，增强现实技术大部分集中在视觉方面，例如谷歌眼镜。但是相比于视觉，声音能够更好地带来身临其境的体验，因为人可以感知任意方向的声学信息，而无需转向信号源<sup>[1]</sup>。

世界上所有声音的本质都是三维（3D）的，包含空间信息，然而在过去的大部分时间里，人类对3D声音的研究和使用集中在建筑声学和音乐创作方面。自19世纪以来，更先进的技术允许出现更具沉浸式的声音，目标可能是重现一个完整的声学环境，尽可能接近真实世界，也可能是创造一种新体验，对真实世界进行增强<sup>[2]</sup>。我们可以捕捉、处理、存储和传输来自自然环境的声音，并通过使用扬声器或耳机进行回放以代替或增强自然聆听环境，这将有助于听者的听觉体验。空间音频技术，特别是3D音频，可以再现身临其境的声音体验，已经成为信号与信息处理和人工智能领域的研究热点。从某种意义上来说，这个领域的迅速发展试图重新建立声音和空间之间的联系，而这种联系在某种程度上被早期的录音和复制的技术限制所切断<sup>[1]</sup>。

空间音频重建技术可以分为双耳音频重建技术和声场重建技术，其区别主要在于是否考虑人的影响。其中，声场重建技术从单通道到立体声，再到多通道环绕声，最后到全息声。其最终目标是通过扬声器阵列在某一空间区域内再现自然的全景虚拟声场，尽可能保留原有声音的空间特性。双耳音频重建技术，即双耳渲染技术，利用和模仿人耳听觉系统对空间音频的感知，通过耳机或扬声器，针对听者的双耳模拟再现虚拟的听觉环境<sup>[2]</sup>。这就产生了一种“你在那里”的第一人称视角，与声场重建技术中“他们在这里”的扬声器系统形成对比，体现了双耳音频技术的优越性<sup>[1]</sup>。

本课题聚焦于双耳渲染技术，考虑人耳听觉系统，通过给听者提供空间听觉信息，再现具有高度真实感的三维听觉场景。几十年来，这项技术一直受到科学界众多学者的广泛关注，在虚拟声学、建筑声学、语音通信、信息系统、媒体社交、教育和游戏娱乐等领域有着广泛的应用。

当双耳渲染技术足够成熟时，其应用包括：

- 提高语言通信的可懂度。目前绝大部分的语言通信采用的是单通路的信号传输系统，不能实现目标语言声源与其他声源在空间上分离，因而语言可懂度会变差。如果引入双耳渲染技术，对原始场景进行重现保留声源的空间信息，就可以提高语言通信的质量；
- 引导视觉定位。声音信息的空间化可以引导视觉定位，甚至可以在不借助视觉的情况下寻找目标，从而减少寻找目标以及采取应对措施的时间，提高安全性。主要用途是民用或军用救援搜索、旅游或博物馆向导、盲人的导向和信息系统；
- 减轻视觉负担。当目标的位置超出视觉的范围（如后方的目标），或同时有多个视觉目标时，可以结合双耳渲染技术，将部分目标信息转换成声音的空间信息重放出来，减轻视觉的负担；
- 改变我们聆听音乐、与音乐互动、共存的方式，可以重新定义我们娱乐、交流和合作的方式，改善生活质量。

双耳渲染技术是给听众提供空间音频的一种有效的方式。此外，虚拟听觉环境的音质和能否带来有效的真实感，对听者来说也变得越来越重要。到目前为止，双耳渲染技术得到了快速发展，但仍有许多挑战和待解决的问题。

## 1.2 国内外研究现状

双耳渲染技术的实现有两种方式：基于耳机的双耳渲染和基于扬声器的双耳渲染。其中，通过耳机将双耳信号直接传输至听者耳膜是最自然和有效的方法。基于耳机的双耳渲染有很多优点：

- 具有完美的通道分离特性。耳机将左声道信号直接发送到左耳，右声道信号直接发送到右耳，因此到达每只耳朵的信号是独立呈现的，可以防止任何串音信号到达非预期的耳朵；
- 较小的噪声干扰。除非严格控制，否则任何听力环境都含有背景噪声，这些噪声会叠加在重建的双耳信号上，造成听觉干扰，导致听觉上的声像失真。当使用耳机时，听众与声学环境之间的硬件隔离会减少环境噪声带来的干扰；
- 可用于移动或手持式设备。相对来说，耳机硬件结构简单，重放功耗较小，所需的数据传输量小，且可以得到相对好的音质，比较适合移动和手持式播放设备的应用；
- 基于耳机的重现不受听众位置或方向的影响。除非对头部进行跟踪并使用额外的处理来补偿听者的位置，否则所呈现的信号将与听者的位置或方向无关，并可以使听者始终处于最佳位置。

对于一些应用场景，也可以通过扬声器实现双耳渲染，例如在电影音频、家庭影院系统、车载音箱中。需要注意的是，在使用扬声器进行双耳声信号的回放时，需要进行

串音消除。未来随着个性化听觉感知的发展，这些市场也可能面临来自基于耳机的双耳渲染的日益激烈的竞争。因此本文更多关注基于耳机的双耳渲染技术，接下来对目前的研究现状进行介绍。

### 1.2.1 双耳录音

人类只有两只耳朵来感知三维空间中的声音，人耳听觉系统可以通过双耳接收到的声信号来完成环境感知、声源定位等任务。因此，可以很直观地联想到使用两个分别放置在人头或者假人头模型左、右耳处的高保真麦克风的录制信号来模拟人耳接收到的声信号，这个过程称为双耳录音（binaural recording）<sup>[1][2]</sup>。麦克风录制得到的双耳信号经过放大、传输等过程后，使用一对耳机进行回放，在听者双耳处呈现与原声场一致的主要空间信息，产生逼真的聆听体验。

假人头模型是人类头部的物理重现，具有与成人头部相似的结构特征，例如头部大小、形状、耳朵、耳廓的位置。在某些情况下，该模型还包括肩膀和躯干。在信号被麦克风采集之前，人体结构对声波的反射和散射导致的空间线索会被叠加，因此所获取的双通道采集信号包含了真人听者的空间听觉线索。假人头模型的作用是捕捉听者两耳处的声信号，从这个角度来看，假人头模型与双通道麦克风类似，是双通道立体声录音的一种特定方法。此外，假人头模型也应用于声学研究和测量，例如头相关传递函数的测量。

1931 年，Fletcher 在贝尔实验室利用一个绰号奥斯卡的假人开发了一种双耳录音设备，并出现在 1933 年的芝加哥世界博览会上，第一次记录了使用假人头模型进行双耳录音的演示过程<sup>[1]</sup>。由于 1.4 英寸的麦克风太大，无法放置在耳道，只好将麦克风安装在假人头模型耳朵正前方的颧骨上，两只耳朵的声音被实时传输给戴着耳机的听众，实验中超过三分之一的受试者更喜欢双耳<sup>[3]</sup>。现在有许多假人头模型和双耳声音捕捉设备，在耳朵的位置，或者在耳朵的入口处，或者在耳道的某个地方，都配备了麦克风。可以分为三种类型：假人头，带肩或躯干的假人头和双耳麦克风，如图 1-1 所示。

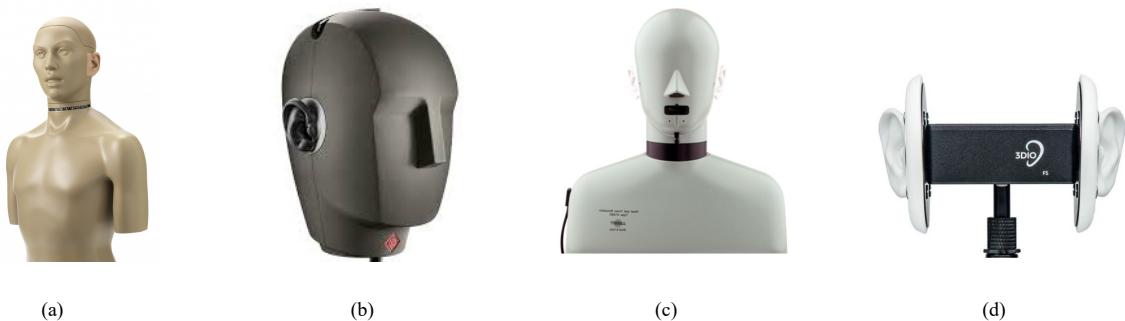


图 1-1 双耳录音设备，(a) KEMAR，(b) Neumann KU-100，(c) HATS，(d) 3Dio

一些广泛使用的双耳录音设备包括仿真人头 KEMAR<sup>[4]</sup>、Neumann KU-100<sup>[5]</sup>、头部和躯干模拟器 HATS<sup>[6]</sup> 和 3 Dio 自由空间双耳麦克风<sup>[7]</sup>。其中，KEMAR 已被广泛用于头相关传递函数 HRTF (头相关脉冲响应 HRIR) 和双耳房间脉冲响应 BRIR 的测量，例如麻省理工学院<sup>[8]</sup>、华南理工大学<sup>[9]</sup> 以及柏林工业大学<sup>[10]</sup>。Neumann KU-100 仅包含了人类头部的结构特征，这种设计便于携带，可以捕获头部和耳廓带来的空间信息，但由于躯干的缺失，肩部所对应的空间线索没有包含在录制的双耳信号中，这一缺陷可能会降低俯仰角线索的强度，从而限制了声音的空间捕捉范围。HATS 的结构与 KEMAR 类似，但其对人体的物理重现精度稍差。3 Dio 自由空间双耳麦克风使用两个仿真人耳进行双耳录音，其间距与人头尺寸相似，但并未考虑头部和躯干对声波的反射和散射。

尽管许多人认为双耳录音是捕捉三维声音和空间线索的最令人信服的方法，但它们的一个限制是，采集的声像代表一个固定的视角。换句话说，一旦声场景被采集，听者就不能转动头部与环境进行互动。相比于静态聆听模式，涉及头部跟踪的交互式系统具有显著的优势，不仅可以创建更加真实和身临其境的聆听环境，还可以提高定位精度，减少前/后和上下混淆<sup>[11][12]</sup>。目前已有许多可用的技术来跟踪听众的位置，例如惯性传感器（加速度计、陀螺仪）、声学跟踪、光学跟踪和磁跟踪。

运动跟踪双耳(Motion-tracked Binaural, MTB)是由Algazi Duda 和 Thompson 于2004年提出的一种方法<sup>[1][13]</sup>，作为一种新的双耳声音捕捉和再现方法，在双耳录音的基础上进行推广，保留了动态头部运动线索，允许听者移动头部，获得以听者为中心的声像。基于使用假人头模型进行声音捕捉的概念，MTB 使用分布在一个接近人头的球体表面上的麦克风获取双耳信号。实际系统中使用一个圆形麦克风阵列，直径设定为人体平均头部直径，如图 1-2 (a) 所示。在回放过程中跟踪听者头部在水平面上的运动，使用头部跟踪器来确定最接近听者耳朵位置的麦克风，并且集成有信号插值模块，可以更有效地获取双耳信号。如果听者耳朵所在位置碰巧与麦克风位置吻合，来自该麦克风的信号直接发送至耳机；如果耳朵在两个麦克风之间，信号经过插值后发送至耳机。

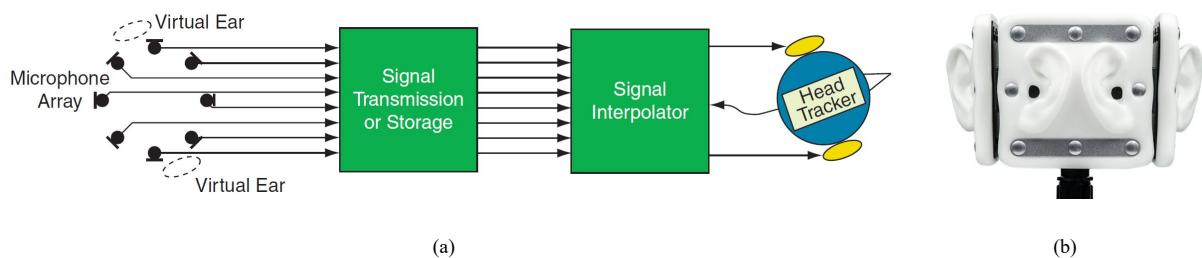


图 1-2 多自由度双耳录音设备：(a) MTB<sup>[13]</sup>；(b) Omni

但是，MTB 方法并非没有限制。由于麦克风没有配备人造外耳或耳廓，所以信号缺乏与听者仰角定位相关的谱线索。如图 1-2 (b) 所示，Omni 双耳麦克风与 MTB 类似，

是由 3Dio 发展而来。其包含沿正方形放置的四对耳廓，每一边均配备了一个左耳耳廓和右耳耳廓，允许从四个不同的角度捕获双耳信号，即可以同时捕获头部朝向四个方向的双耳信号。类似于 MTB，在重放过程中，可以根据听者的方向调整输出双耳信号。

双耳录音的限制包括以下两点：一是其设备尺寸较大，便携性差。二是仿真人头录音要在个体差异和良好的空间感知之间折中。每个人身体都有独特的尺寸和形状特征，因此直接使用仿真人头录音代替人耳可能会造成空间混淆，特别是在仰角定位方面。这个问题也被称为 HRTF 的个性化差异，是 HRTF 的特征之一。

### 1.2.2 双耳合成

一般来说，双耳信号可以分为两个主要成分，一是声音从声源传播到听者所在的位置，二是听者头部等生理结构对声波的影响。“双耳”一词专指进入听者耳朵的双声道声音经过了一系列旨在模仿人类定位的线索的综合滤波。这些线索可以被自然地叠加（例如双耳录音），也可以通过信号处理的方式获得（例如双耳合成）。自从 1989 年 Wrightman 和 Kistler 采用信号处理的方法在耳机重放中模拟出三维空间虚拟源<sup>[14][15]</sup>以来，虚拟听觉重放的研究得到了很大的发展，成为国际上空间声的研究热点，并在许多领域得到了广泛的应用<sup>[16][17][18]</sup>，例如在电脑游戏和军事训练系统中遇到的虚拟声学环境。本节将对现有的双耳合成方法进行介绍。

头相关传递函数 HRTF 在双耳合成中至关重要，是双耳空间研究及虚拟听觉应用的核心问题。其描述了听者的头部、躯干和耳廓等生理结构对入射声波的综合滤波，包含了双耳声压的各种空间线索，尤其是人耳听觉系统对声源进行精确定位的关键线索，例如 Rayleigh 双因素理论中的双耳时间差和双耳声级差，耳廓所带来的单耳因素——谱因素。HRTF 的时域表示为头相关脉冲响应 HRIR。

简单的双耳合成是通过对声源的单声道信号卷积左右耳的 HRIR 来实现。如上所述，一对左右耳 HRIR 包含了听者感知声源方位所必要的空间线索。当我们将包含在 HRIR 中的线索叠加在一个声源信号上，可以创建一个位于听者耳道入口处的双耳信号，由此产生了声音似乎来自 HRIR 所对应方位的一种错觉。

用于双耳合成的 HRIR 可以是通用的、单独测量的、从数据库中选择的和/或定制的，其中许多 HRIR 数据库是在无回声环境中测量的。也可以捕捉扬声器、室内声学和听者 HRIR 的组合特性，从而得到双耳房间脉冲响应 BRIR。同样，通过卷积将左右耳的 BRIR 和扬声器播放信号相结合，即可通过耳机重现一个在房间内的虚拟声源，使感知到的虚拟声像等效于房间里的真实扬声器。

在渲染过程中，当声源位置相对于听者发生动态变化时，要想给听者产生连续空间位置变化的主观印象，而不是突然的声像变化，必须经常更新虚拟声源的位置并使用相应的 HRIR 和 BRIR 滤波器。为了在空间上产生均匀的运动状态，HRIR 和 BRIR 必须

在两个或多个离散测量的位置之间进行插值，以获取各个位置的 HRTF。目前已经有多 种方法和技术应用于时域和频域的插值，例如简单局部线性插值、主成分分析（PCA）、 Karhunen Loeve 展开、小波域表示和球谐函数分解。

上述信号处理方法需要已知声源位置和相应的驱动信号，常用于 VR 产生虚拟声学 环境。对于捕捉的自然声场，经常存在多个声源的叠加，此时需要先对采集信号进行 方位估计和声源分离。该方法的运算复杂度较高，且估计误差会降低双耳信号的质量。 有一些先进的方法提出对整个声场进行统一处理，避免了方位估计和对多个声源的单 独处理，降低复杂声场景下的运算复杂度，同时提升空间音频的效果。其中，Ambisonics 技术已经在物理声场重建领域占据重要地位，现有的部分双耳渲染技术在其基础上 进行<sup>[19][20][21][22]</sup>。

1973 年，Michael Gerzon 首次提出 Ambisonics<sup>[23]</sup> 技术，使用球谐基函数对来自某 采样点周围各个方向的声场进行编码和重建。Ambisonic 技术最大的优点是其编码与播 放系统无关，因此在重建过程中非常灵活。最初，Michael Gerzon 和 Peter Craven 发明了 著名的声场麦克风，使用四个麦克风实现入射声场的一阶近似，包括指向 x、y、z 轴的 八字形麦克风和一个全向型麦克风进行空间三维录音，分别采集声场的前后、左右、上 下和全向信息，即 X、Y、Z、W 分量，这四个正交的分量信号被称为 B 格式音频信号。

由于一阶的 Ambisonics 系统使用低阶截断的球谐分解，所获取的声场空间信息不 足，不能达到声场重建中完美的沉浸式体验效果，因此高阶 Ambisonics 引起了广泛关 注。更高阶的 Ambisonics 系统（HOA）使用额外的麦克风，在数学上进行更高阶次的 球谐波展开来实现声场的更精确逼近，重现效果较为理想，最佳听音区和可实现的音 频带宽得以提高。目前在技术上比较成熟的 Ambisonics 话筒有 CoreSound 的 TetraMic、 TSL 的声场麦克风 SPS200，以及可进行高阶 Ambisonics 录制的 Eigenmike 球形话筒<sup>[24]</sup>， 如图 1-3 所示。



图 1-3 Ambisonics 麦克风

双耳渲染技术对 Ambisonics 方法进行移植，和虚拟扬声器方法相结合，获取双耳信 号。基于耳机的虚拟扬声器方法的目的是仿真听者在含有一个或多个扬声器的场景中的 听感。对所有虚拟扬声器的声源信号卷积相对应方位的左右耳 HRIRs，并进行叠加，以 获取综合的双耳信号。虚拟扬声器方法不仅可以和 Ambisonics 技术进行结合，还可以

与 5.1、7.1、10.2、22.2 等多通道环绕声相结合，创建虚拟听觉。

但是基于 Ambisonics 的虚拟扬声器方法会带来一些问题：一是不同的虚拟扬声器数目和位置会对所获取的双耳信号带来影响，二是扬声器所在位置的声像会不可避免地被加重。

近年来，球形麦克风阵列在声场录制及重放领域中日益普及。球面阵列的使用自然导致在球面谐波（Spherical Harmonics, SH）域中进行三维声场的处理和分析。1998 年，Evans 等人<sup>[25]</sup>首次提出了用球谐函数来表示 HRTF 的方法，接着一些新的双耳渲染工作在球谐域进行<sup>[26][27][28][29]</sup>。其核心思想是不仅将声场在球谐域进行表示，同时将 HRTF 在球谐域展开，并直接在球谐域中进行渲染。

在球谐域渲染双耳信号的优点在于能够直接使用现有的球谐域算法来处理声场或相应的 HRTFs，例如波束形成、定位等。在球谐域实现听者追踪时，只需要对原始声场乘以一个旋转因子即可实现对声场的转向<sup>[30]</sup>，无需传统算法中的 HRTF 匹配和插值。相比于基于 Ambisonics 的虚拟扬声器方法来说，该方法无需解码到虚拟扬声器这一步骤，避免了虚拟扬声器数目和位置对听感的影响。

为了精确地使用球谐函数表示声场和 HRTF，通常需要使用高阶展开。然而实际系统中由于麦克风数量有限，导致声场的阶次较低，一般实际系统中小于 5 阶<sup>[28]</sup>。而对于现有的 HRTF 数据库来说，其球面采样点较多，可以实现较高的分解阶次。Zhang 等人<sup>[31]</sup>对 HRTF 的空间维度进行了研究，得出的结论是，要想对高达 15 kHz 的 HRTF 进行精确的物理表征，大约需要 34 阶。

声场和头相关传递函数 HRTF 之间存在不匹配性，最简单直接的方法是对 HRTF 进行低阶截断，与声场阶次保持一致。但是现有的低阶双耳渲染算法在提供高保真空间声方面仍受到很多限制。HRTF 的直接低阶截断会导致高频能量损失，带来一定的误差，导致空间分辨率受限，直接影响对声源方位的感知、外化 (externalization)、感知声源宽度和音色等<sup>[27][32][33][34]</sup>。

现有的一些改进方法主要集中在 HRTF 的预处理方面，例如，通过使用时间对准或相位对准<sup>[25]</sup>，最小相位表示<sup>[33]</sup>，或方向均衡<sup>[35]</sup>等。文献 [36] 对现有的方法归纳总结，主要分为单纯基于复谱的方法和复合方法，其中复合方法分别处理 HRTF 的绝对值和相位。包括相位上的频域解卷积、空域解卷积、时间对准，幅度上的取对数谱，以及对复频谱进行平滑，并且可以相互结合，如表 1-1 所示。

表 1-1 HRTF 预处理方法

方法	预处理对象	预处理方法
$H_a$	复频谱	时间对准
$H_{sm}$	复频谱	平滑
$ H $	幅度谱	无
$ H _{dB}$	幅度谱	取对数
$ H_{sm} $	幅度谱	平滑
$ H_{sm} _{dB}$	幅度谱	平滑, 取对数
$\angle_f H$	相位	频域解卷绕
$\angle_s H$	相位	空域解卷绕
$\angle_f H_{sm}$	相位	平滑, 频域解卷绕

基于双耳对准的 HRTF 预处理方法（时间对准、相位对准）已被证明是一种有效降低 HRTF 球谐分解阶次的稳健方法，有很多工作基于此进行。

2018 年，Zaunschirm 等人的另一项研究<sup>[28]</sup>也提出了改善球谐域双耳渲染算法中的 HRTF 阶次问题。通过去除高频 HRTF 的线性相位成分，实现了频率相关的时间对准的 HRTF 预处理方法，并在再现信号上添加扩散场约束。2019 年，Zamir Ben-Hur 等人<sup>[37]</sup>将相位对准预处理和新的稀疏采样方式相结合，提出了一种有效的 HRTF 表示方法。该方法能够在相对较少的方向上测量 HRTF，并通过精确的插值重建高空间分辨率的 HRTF。结果表明，该方法仅需要 100 个测量方向的稀疏 HRTF 采样，不到全采样所需测量次数的 10%。

2020 年，Zamir Ben-Hur 将对 HRTF 引入相位对准这种有效的表示融入到整个双耳渲染过程中<sup>[38]</sup>，提出了一个新的概念——双耳 Ambisonics 信号。该信号表示左右耳所在位置的声场 Ambisonics 表示，可以直接录制某场景下两个耳朵处的声场，例如使用两个分别位于左右耳处的 Eigenmike 或声场麦克风，也可以由录制声场计算得到左右耳位置出的声场分量，计算原理与 HRTF 进行双耳对准的逆过程相似。

除此之外，还有一些 HRTF 预处理方法通过对 HRTF 的稀疏采样降低球谐分解的阶次，也有部分方法主要集中在对双耳信号的频谱补偿方面。

2014 年，Bernschütz 等人对 HRTF 数据库进行空间重采样<sup>[26]</sup>，根据所需的阶次设定相对应的低阶复合网格，有效的采样网格可以减少音色失真。采样方式包括等角高斯采样和列别捷夫采样，若 HRTF 数据库不包含所选取的采样点，则需要引入插值的方法获取。

2017 年，Zamir Ben-Hur 等人提出了一种将低 SH 阶双耳信号的频谱校正为高 SH 阶频谱的均衡滤波器<sup>[34]</sup>，使原始双耳信号的频谱分布更为均匀。该方法以刚性球为模

型，通过计算高阶扩散场与低阶截断扩散场的比值获取均衡滤波器，从而减小阶次降低对 HRTF 所带来的影响。所提出的滤波方法能够有效地恢复截断后的双耳信号的音色成分，同时保留甚至改善了信号的一些空间特性。结果显示，当使用均衡滤波器时，再现信号的空间感知能力显著改善，双耳声音的音色得以校正。另外，虚拟源被认为完全外化。

2019 年 Christoph Hold 等人提议通过补偿滤波器和球谐域渐变来校正音色<sup>[39]</sup>。之前的研究直接对 HRTF 进行低阶截断，这相当于一个矩形窗。该方法在 HRTF 在截断过程中进行加窗，降低直接使用矩形窗进行阶次截断所引入的旁瓣，并且和均衡滤波器相结合。结果表明，所提出的方法可以减少音色损失，从而改善音频的感知质量。

还有一部分工作对现有的 HRTF 预处理方法进行比较和进一步验证，例如 Brinkmann 等人在 2018 年和 Lubeck 等人在 2019、2020 年的几项研究<sup>[36][40][41]</sup>验证了各种 HRTF 预处理方法对双耳信号在感知方面的改善。

## 1.3 研究内容与论文结构

### 1.3.1 研究内容及创新点

本论文聚焦于球谐域的双耳渲染技术。尽管这一领域已经取得了重大进展，但基于低阶球谐的高质量双耳渲染算法仍然是一个值得探索的问题。

对于球谐域双耳渲染算法来说，其包含两个同等重要的部分，声场和 HRTF。二者之间存在阶次和尺寸等不匹配问题，直接基于低阶 HRTF 球谐分解的双耳渲染算法会对双耳信号带来众多感知方面的损失。实验研究表明，当双耳信号以更高的球谐波阶次进行重构时，对声源的感知会更清晰和更外化，使用更高的球谐波阶构建双耳表征将导致外部的、更窄的、更远的声像，以及更平衡的音色。

近年来许多学者对低阶球谐分解的双耳渲染算法加以改进，但这些方法都是从 HRTF 角度出发。本文不仅从 HRTF 预处理的角度出发，并且在声场层面对现有的双耳渲染算法加以改进，创新性提出了一种新的声场扩阶理论，并且将声场扩阶理论、头相关传递函数预处理和基于球谐分解的双耳渲染算法相结合，实现了基于声场扩阶的双耳渲染算法。该声场扩阶理论对麦克风阵列的采集声场进行分析，将入射声场分解为直达波和混响场的叠加。根据直达波入射方向进行空间加窗处理，可以实现声场分量的阶次提升，同时扩大控制区域的半径。

### 1.3.2 论文结构安排

论文共分为六个章节，各章节的主要内容如下：

本章（第一章）为绪论。首先对增强现实下双耳渲染技术的研究背景及可预见的应用进行了简单的介绍，接下来对现有的双耳渲染实现方法加以详细阐述，最后说明了本文的研究内容、创新点与论文结构安排。

第二章为双耳渲染算法。本章对三种典型的双耳渲染算法的原理进行详细介绍，其中包含本文所关注的基于球谐分解的双耳渲染算法，并给出了双耳渲染中至关重要的头相关传递函数和声场的球谐分解等基础知识，为后续工作奠定理论基础。

第三章为基于 HRTF 球谐分解的双耳渲染算法。本章首先详细介绍了声场和头相关传递函数 HRTF 的球谐分量求解方法，并对本文所采用的 HRTF 预处理方法进行介绍，最后通过实验验证了预处理算法的有效性。

第四章为基于声场扩阶的双耳渲染算法。本章首先对信号模型和球谐域 MUSIC 定位算法进行介绍，其次从有界线性算子出发，对通过空域加窗实现声场阶次提升的方法进行详细推导，最后将声场扩阶和 HRTF 预处理方法相结合，给出总体算法框架。

第五章为实验结果及分析。本章首先对评价指标及其计算方法加以介绍，其次在消声室和混响环境下，将对标算法和本文所提出的算法进行对比和分析，验证本算法的优越性。

第六章为全文总结。对论文的研究内容进行归纳和总结。

## 第二章 双耳渲染算法

本章对三种典型的双耳渲染算法的原理进行详细介绍，主要包括虚拟听觉重放技术、基于 Ambisonics 的虚拟扬声器方法和基于球谐分解的双耳渲染算法，其中基于球谐分解的双耳渲染算法是本文的研究重点。并给出了双耳渲染中至关重要的头相关传递函数、球谐函数的定义及性质，以及球傅里叶变换及其逆变换等基础知识，为后续工作奠定理论基础。

### 2.1 虚拟听觉重放技术

人类是通过双耳进行声音感知的，除了可以感知声音的响度、音调和音色，还可以对声源的方位和距离进行判断。声源发出的声音经过头部、躯干、耳廓等散射和反射后到达双耳，因此双耳信号中包含了大多数声源定位的因素，例如双耳时间差和双耳声级差。虚拟听觉方法是指采用信号处理的方法来模拟声波经由头部、躯干、耳廓等人体生理结构的散射和反射的过程。声波从声源到双耳的传输过程可以看作一综合滤波器，该滤波器可以通过头相关脉冲响应（HRIR）和头相关传递函数（HRTF）来描述。

#### 2.1.1 头相关传递函数 HRTF

头相关传递函数（Head-related Transfer Function , HRTF）定义为自由场情况下从声源到双耳的频域传输函数，表达了生理结构对声波的综合滤波效果<sup>[18]</sup>。如式 (2-1) 所示，HRTF 是声源到头中心距离  $r$ 、声源的方位角  $\theta$ 、仰角  $\phi$ 、频率  $f$  和人头半径  $a$  的函数。当声源位于远场时，HRTF 与距离无关，一般略去参数  $r$ 。

$$\begin{aligned} H_L &= H_L(r, \theta, \phi, f, a) = \frac{P_L(r, \theta, \phi, f, a)}{P_0(r, f)}, \\ H_R &= H_R(r, \theta, \phi, f, a) = \frac{P_R(r, \theta, \phi, f, a)}{P_0(r, f)}, \end{aligned} \quad (2-1)$$

式中： $P_0$  是头移开后点声源在头中心位置处的频域复数声压， $P_L$ 、 $P_R$  分别为左右耳的频域复数声压。每个人的生理结构和尺寸不同，式 (2-1) 中参数  $a$  的引入就是为了体现 HRTF 的个性化特征，一般不涉及个性化 HRTF 时略去参数  $a$ 。

严格来说， $P_L$ 、 $P_R$  应定义为左、右耳鼓膜处的电压。但实际测量表明，至少在 14 kHz 以下的频率，整个耳道可以近似为一维传输。所以耳道入口、鼓膜，甚至耳道内任一点的声压，都可以作为 HRTF 定义中的左、右耳声压  $P_L$ 、 $P_R$ ，它们都包含了有关声源空间

位置的信息，并且不同测量点的声压之间可以互相转换。将耳道声压的测量点称为参考点，耳道内任意参考点的声压统称为双耳声压。

在时域上，HRTF 表示为头相关脉冲响应（Head-related Impulse Response , HRIR），HRIR 和 HRTF 是一对傅里叶变换，HRIR 表示声源到双耳的脉冲响应，是声源位置  $r$ 、 $\theta$ 、 $\phi$  和时间  $t$  的函数。HRIR 和 HRTF 包含了众多主要的声源定位因素，但并不是包含全部的定位因素，至少没有包括头部微小转动所产生的动态因素。

### 2.1.2 基于 HRIR 和 BRIR 的虚拟听觉方法

为了在双耳渲染中产生空间音频效果，我们可以通过将单声道信号与虚拟声源位置对应的 HRIR 卷积来生成虚拟音效。该技术中双耳声信号和相应的空间信息都是通过信号处理的方法人为产生或虚拟出来的，其中不同的声源信号和方位都是已知的。

在自由场情况下，位置  $(r, \theta, \phi)$  的声源在双耳处产生的声压可以通过相应角度的头相关传递函数  $H_L(r, \theta, \phi, f)$  和  $H_R(r, \theta, \phi, f)$  描述。通过实验测量或理论计算得到头相关传递函数后，将频域信号  $E_0(f)$  代入式 (2-2) 即可计算出双耳声压。

$$\begin{aligned} B_L(f) &= E_0(f)H_L(r, \theta, \phi, f) \\ B_R(f) &= E_0(f)H_R(r, \theta, \phi, f). \end{aligned} \quad (2-2)$$

然而人耳听音的环境大多都是多声源和混响环境，实际系统中通常需要引入多个声源的叠加，并且需要考虑房间的声学特性，如早期反射和晚期混响。使用 HRIR 进行双耳渲染时，需要先对采集信号进行方位估计和声源分离，再对其进行渲染，如图 2-1 所示。此时计算量取决于听觉场景的复杂度，场景中声源数目越多，计算量越大。

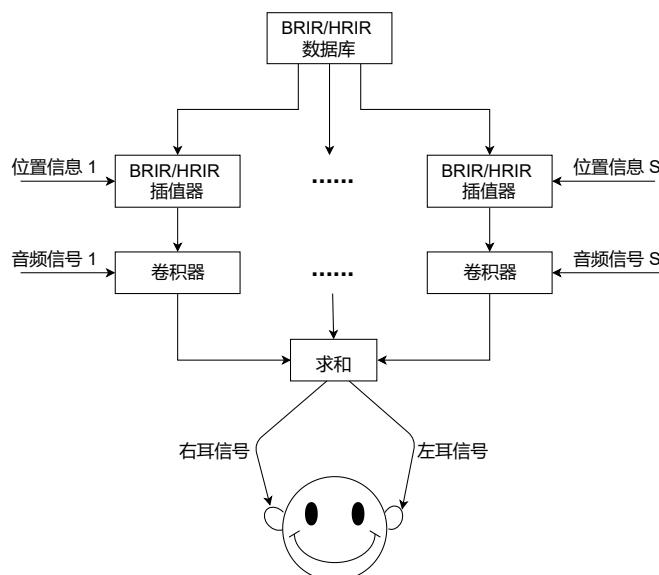


图 2-1 基于 HRIR 和 BRIR 的方法

一个概念上简单的例子是通过双耳房间冲激响应（Binaural Room Impulse Response, BRIR）进行渲染。在一个经过声学优化的混响环境下的听音室中，使用放置在某一固定位置的高性能扬声器播放声源信号，并通过放置在理想收听位置的假人头模型，测量扬声器到麦克风的冲激响应即可获取 BRIR。BRIR 描述了声音从房间里的一点传到听者耳朵的特征，需单独考虑直达波、早期反射、晚期混响，能够准确地捕捉房间的特征，直接对整个环境进行渲染。使用 BRIR 进行渲染是一种有效的方式来给听者提供身临其境的三维空间音频场景。

假设  $S$  个信号源的空间位置  $(r_s, \theta_s, \phi_s)$  和激励信号  $E_s(f)$  已知，通过卷积对应的双耳房间冲激响应获取双耳信号：

$$\begin{aligned} B_L(f) &= \sum_{s=1}^S E_s(f) BRIR_L(r_s, \theta_s, \phi_s, f) \\ B_R(f) &= \sum_{s=1}^S E_s(f) BRIR_R(r_s, \theta_s, \phi_s, f), \end{aligned} \quad (2-3)$$

其中， $BRIR_L(r_s, \theta_s, \phi_s, f)$  和  $BRIR_R(r_s, \theta_s, \phi_s, f)$  分别声源方位为  $(r_s, \theta_s, \phi_s)$  对应的左耳和右耳的房间冲激响应。

基于 HRIR 和 BRIR 的方法在使用耳机进行回放的过程中，需要根据听者和声源的相对位置选择对应的 HRIR/BRIR 数据（当声源位置未包含在数据库时还需进行插值），再对于每个声源，将源信号与其对应的 HRIR/BRIR 进行卷积的结果求和并反馈给耳机。当听者处于运动状态时需要快速产生符合真实听感的声像轨迹，除非使用高速、低延时的算法，否则 HRTF 插值和卷积的组合过程可能导致对头部移动的响应产生不可接受的延时。

## 2.2 基于 Ambisonics 的虚拟扬声器方法

对于许多应用，我们希望能够直接对捕获的整个声场进行渲染，而不需要事先知道源的数量或位置，或声学环境的结构。基于 Ambisonics 的虚拟扬声器方法就是其中一种。

首先对 Ambisonics 技术进行简单介绍，该技术是一种基于空间声场球谐函数分解和重构的声音回放技术，其核心思想是使用某一场景中的麦克风阵列（例如，声场麦克风，球形麦克风阵列）捕获来自不同方向的声压波，并通过位于听众周围的扬声器阵列进行重现<sup>[13]</sup>。

Ambisonics 技术包括编码和解码两大模块，编码是对声场采集信号进行分解，求得各阶球谐分量；解码则是根据扬声器的位置对各阶分量进行组合得到扬声器激励信号，实现声场重建。Ambisonics 系统最大的优点在于其编码和解码模块是互相独立的，编码

时无需事先知道扬声器的摆放位置，解码时根据扬声器位置选择不同的解码矩阵即可，并且扬声器阵列的摆放具有较高的灵活性，便于实现。

Ambisonics 系统的编解码是在以下假设下进行的：(1) 录制采样的声源为空间某方向入射的平面波；(2) 回放过程中，扬声器发出的声波相对听音位置也可视为平面波。以上假设在声源和扬声器均位于远场时成立<sup>[42]</sup>。

### 2.2.1 Ambisonics 编码

对于 Ambisonics 来说，其编码时通过完备正交集球谐函数进行的，通过使用球傅里叶变换对声场进行球谐分解，从而获取声场的球谐系数，接下来将详细介绍。

球谐函数有复数和实数两种等价定义形式<sup>[43][44][45]</sup>，其中  $n$  阶  $m$  次复数形式的归一化球谐函数定义如下：

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos \theta) e^{im\phi}, \quad (2-4)$$

其中， $n = 0, 1, 2, \dots$ ,  $m = 0, \pm 1, \pm 2, \dots, \pm n$ ,  $i = \sqrt{-1}$ ,  $(\cdot)!$  表示阶乘， $P_n^m(\cdot)$  表示关联勒让德函数。 $\sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}}$  是归一化系数，其值由关联勒让德函数决定。 $\theta$  表示俯仰角，定义为方向矢量与  $z$  正轴的夹角， $\phi$  表示水平面的方位角，且  $0^\circ \leq \theta \leq 180^\circ$ ,  $0^\circ \leq \phi \leq 360^\circ$ ，角度定义如图 2-2 所示。

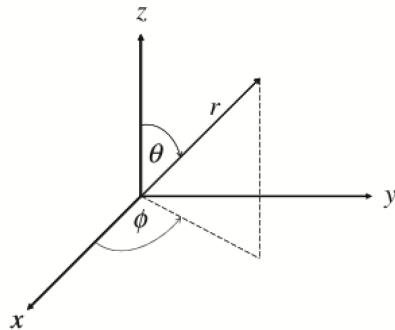


图 2-2 球谐函数角度定义图

球谐函数有许多特性，其中最重要的是正交完备性，将在本文中多次应用。不同阶的  $Y_n^m(\theta, \phi)$  满足以下正交关系式：

$$\int_0^{2\pi} \int_0^\pi [Y_n^m(\theta, \phi)]^* Y_{n'}^{m'}(\theta, \phi) \sin \theta d\theta d\phi = \delta_{nn'} \delta_{mm'}, \quad (2-5)$$

式中， $\delta_{nn'}$ 、 $\delta_{mm'}$  为  $\delta$  函数：

$$\delta_{nn'} = \begin{cases} 1, & n = n', \\ 0, & n \neq n'. \end{cases} \quad (2-6)$$

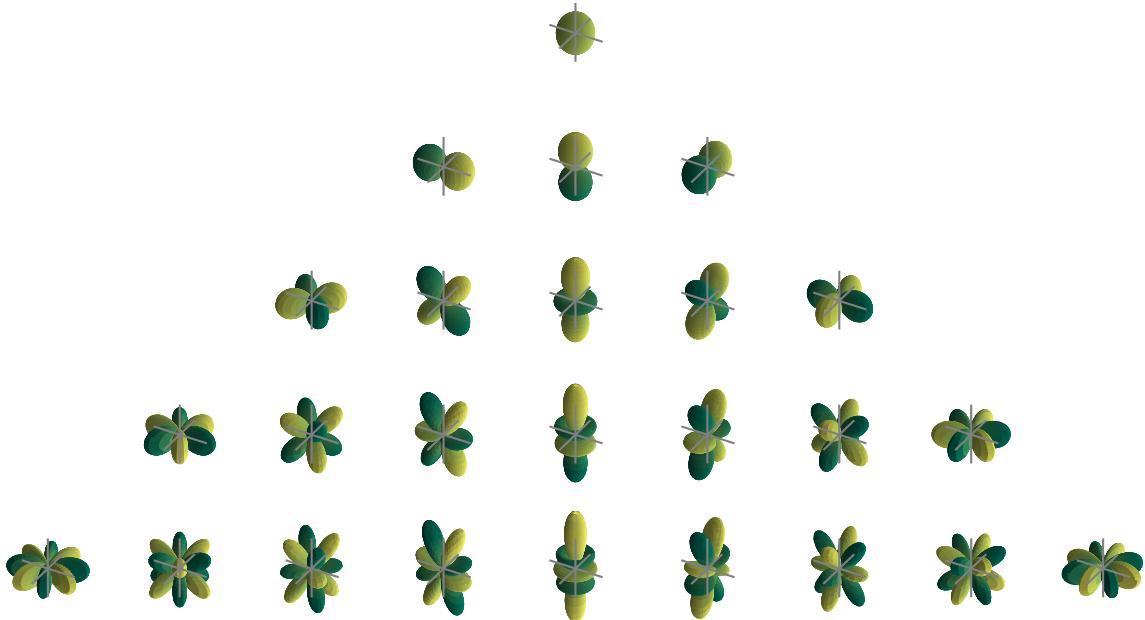


图 2-3 球谐函数三维图

图 2-3 为球谐函数的三维图，其中从上到下分别对应  $n = 0$  到  $n = 4$ ，从左到右分别对应  $m = -n$  到  $m = n$ ，当  $m$  取负值时绘制  $Y_n^m$  的虚部， $m$  取实值时绘制  $Y_n^m$  的虚部。图中黄色部分表示正值，绿色部分表示负值。

正是因为球谐函数的正交完备性，在很多应用中采用球谐函数作为基底函数，以保证声场的正交分解。如式 (2-7) 所示，声场中任意一点的声压可以看作不同阶球谐函数的叠加：

$$S(r, \theta, \phi, f) = \sum_{n=0}^{\infty} \sum_{m=-n}^n S_n^m(r, f) Y_n^m(\theta, \phi), \quad (2-7)$$

其中，球谐系数  $S_n^m(r, f)$  可通过式 (2-8) 得到。

$$S_n^m(r, f) = \int_0^{2\pi} \int_0^\pi S(r, \theta, \phi, f) [Y_n^m(\theta, \phi)]^* \sin \theta d\theta d\phi. \quad (2-8)$$

式 (2-8) 称为  $S(r, \theta, \phi, f)$  的球傅里叶变换 (Spherical Fourier Transform, SFT)，式 (2-7) 称为  $S_n^m(r, f)$  的反球傅里叶变换 (Inverse Spherical Fourier Transform, ISFT)，二者构成逆变换。

如式 (2-7) 可知，声场可以看作是无穷个球谐函数的线性组合，然而在实际系统中这是不可实现的，不能实现声场的无限阶球谐分解，需要对原始声场的分解做截断处理，即使用有限阶的球谐函数来对声场进行近似：

$$\hat{S}(r, \theta, \phi, f) = \sum_{n=0}^N \sum_{m=-n}^n S_n^m(r, f) Y_n^m(\theta, \phi), \quad (2-9)$$

其中,  $\hat{S}(r, \theta, \phi, f)$  是  $S(r, \theta, \phi, f)$  的近似值,  $N$  是球谐函数阶数的最大值。

这种截断处理必然会影响声场的精度, 定义归一化截断误差如下<sup>[46]</sup>:

$$\varepsilon_N(kr) = \frac{\int_0^{2\pi} \int_0^\pi |S(r, \theta, \phi, f) - \hat{S}(r, \theta, \phi, f)|^2 \sin \theta d\theta d\phi}{\int_0^{2\pi} \int_0^\pi |S(r, \theta, \phi, f)|^2 \sin \theta d\theta d\phi}, \quad (2-10)$$

式中, 波数  $k = 2\pi f/c$ , 声速  $c = 343 \text{ m/s}$ 。

对于内部域问题, 即所有声源均位于所关注的区域外, 内部声压场可以表示为<sup>[43]</sup>:

$$S(r, \theta, \phi, f) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_n^m(f) j_n(kr) Y_n^m(\theta, \phi), \quad (2-11)$$

其中,  $A_n^m(f) = S_n^m(r, f)/j_n(kr)$  是与位置无关的球谐系数,  $j_n(\cdot)$  为球贝塞尔函数。

自由场环境下, 对于来波方位为  $(\theta_s, \phi_s)$  的单位幅度平面波来说, 式 (2-11) 及其截断表示为:

$$S(r, \theta, \phi, f) = \sum_{n=0}^{\infty} \sum_{m=-n}^n 4\pi i^n j_n(kr) [Y_n^m(\theta_s, \phi_s)]^* Y_n^m(\theta, \phi). \quad (2-12)$$

$$\hat{S}(r, \theta, \phi, f) = \sum_{n=0}^N \sum_{m=-n}^n 4\pi i^n j_n(kr) [Y_n^m(\theta_s, \phi_s)]^* Y_n^m(\theta, \phi). \quad (2-13)$$

将式 (2-12) 和式 (2-13) 代入到式 (2-10), 归一化截断误差可表示为:

$$\varepsilon_N(kr) = 1 - \sum_{n=0}^N (2n+1)(j_n(kr))^2. \quad (2-14)$$

关于  $kr$  的归一化截断误差  $\varepsilon_N(kr)$  与截断阶数  $N$  之间的关系如图 2-4 所示。

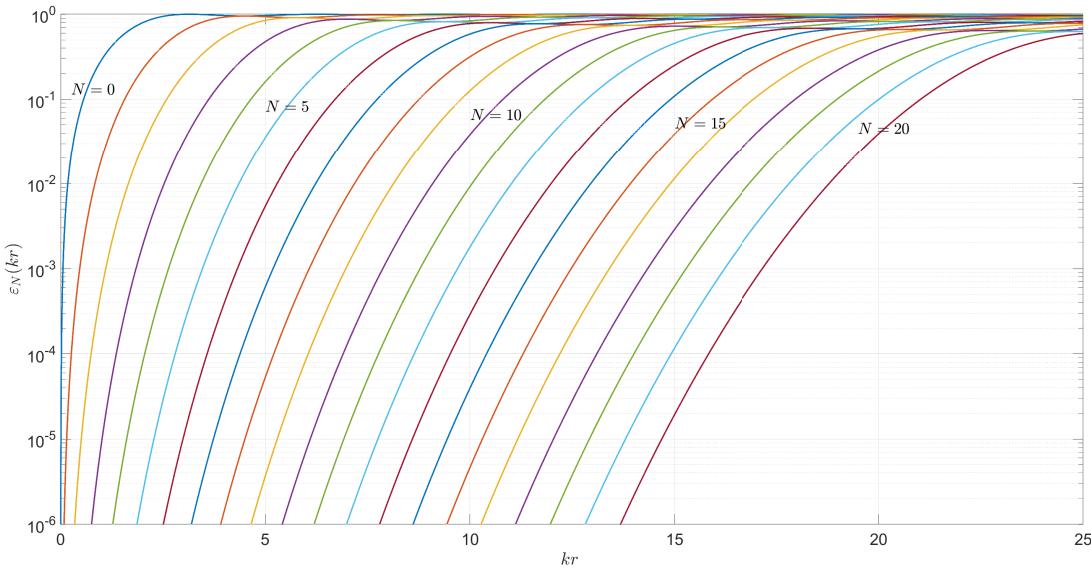


图 2-4 不同阶数  $N$  时归一化截断误差  $\varepsilon_N(kr)$  与  $kr$  之间的关系

在实际应用中, 对于给定的频率  $k$ 、区域半径  $r$  以及可接受的误差, 可以从图 2-4 查询所需要的分解阶次  $N$ 。以 4% 的误差为例<sup>[46]</sup> (对于大多数实际应用已经足够), 可以

看出频率  $k$ 、半径  $r$  和最低截断阶次  $N_0$  之间存在关系:

$$N_0 = \lceil kr_0 \rceil, \quad (2-15)$$

式中  $\lceil \cdot \rceil$  表示向上取整。

在半径小于  $r_0$  的重建区域内，重建声场可近似表示原始声场；在半径大于  $r_0$  的重建区域内，重建声场与原始声场差异较大，重建效果不准确。

## 2.2.2 Ambisonics 解码

Ambisonics 解码过程需要满足平面波声场的假设条件，扬声器均匀分布在球上，且只辐射平面波，实现扬声器阵列中心处的声场重建。采用扬声器阵列重现声场的关键在于求取各个扬声器的激励信号，求解方法为使扬声器阵列产生的声场与原始平面波声场相匹配。原则上，通过使用 Ambisonics 分量信号的线性组合来导出扬声器信号，其中每个信号取决于扬声器相对于重建声场中心的实际位置。接下来从理论上对扬声器激励信号的求解过程进行推导。

由于假设扬声器只辐射平面波，参考式 (2-13)，使用  $L$  个位于  $(\theta_l, \phi_l)$  的扬声器进行声场重建时，所产生的线性叠加声场表示为：

$$\hat{P}(r, \theta, \phi, f) = \sum_{l=1}^L w_l \sum_{n=0}^N \sum_{m=-n}^n 4\pi i^n j_n(kr) [Y_n^m(\theta_l, \phi_l)]^* Y_n^m(\theta, \phi), \quad (2-16)$$

其中， $w_l$  为第  $l$  个扬声器的激励信号。

结合式 (2-13) 和式 (2-16)，建立扬声器阵列产生的总声场与原始平面波声场之间的等式：

$$\sum_{l=1}^L w_l [Y_n^m(\theta_l, \phi_l)]^* = [Y_n^m(\theta_s, \phi_s)]^*. \quad (2-17)$$

将式 (2-17) 写为矩阵形式：

$$\Psi \mathbf{w} = \mathbf{d}, \quad (2-18)$$

其中， $\square$  是大小为  $(N + 1)^2 \times L$  的矩阵，其第  $v$  行  $l$  列元素  $\Psi_{v,l} = Y_n^m(\theta_l, \phi_l)$ ，其中  $v = n^2 + n + m + 1$ ， $\mathbf{w}$  是包含扬声器激励信号的矢量，

$$\mathbf{w} = [w_1, w_2, \dots, w_L]^T, \quad (2-19)$$

$$\mathbf{d} = [Y_0^0(\theta_s, \phi_s), \dots, Y_n^m(\theta_s, \phi_s), \dots, Y_N^N(\theta_s, \phi_s)]^T. \quad (2-20)$$

由于扬声器阵列的配置方式是一定的，因此只有  $\mathbf{w}$  是未知量。根据球谐系数的数量  $(N + 1)^2$  和扬声器的数量  $L$  之间的大小关系，对  $\mathbf{w}$  的求解分情况讨论<sup>[46, 47]</sup>：

- 当  $L > (N + 1)^2$  时, 矩阵  $\Psi$  是一个欠定矩阵, 式 (2-18) 可能没有解, 也可能有无穷多个解。 $\mathbf{w}$  通常可以通过最小二乘问题求解:

$$\min_{\mathbf{w}} \|\mathbf{w}\|^2 \quad \text{subject to } \Psi \mathbf{w} = \mathbf{d} \quad (2-21)$$

其中,  $\|\cdot\|$  表示向量的 2 范数。

- 当  $L = (N + 1)^2$  时, 矩阵  $\square$  是一个非奇异方阵, 式 (2-18) 有唯一解, 采用矩阵求逆的方法求解  $\mathbf{w}$ :

$$\mathbf{w}_{\text{opt}} = \Psi^{-1} \mathbf{d}. \quad (2-22)$$

- 当  $L < (N + 1)^2$  时, 矩阵  $\Psi$  是一个超定矩阵, 式 (2-18) 没有准确解。 $\mathbf{w}$  通常可以通过下式求解:

$$\min_{\mathbf{w}} \|\Psi \mathbf{w} - \mathbf{d}\|^2 \quad (2-23)$$

并且在实际系统中, 精确再现原始声场所需的扬声器数量为  $L \geq (N + 1)^2$ , 即保证扬声器的数目大于需要约束的模态数<sup>[46]</sup>。

对于一阶 Ambisonics 来说, 只包含零阶分量  $W$  和一阶分量  $X$ 、 $Y$ 、 $Z$ , 其扬声器激励信号也可以通过以下方式进行求解。假设扬声器阵列均匀布放, 且扬声器数目  $L \geq 4$ , 则扬声器激励信号为:

$$w_l = \bar{w}_l W + \bar{x}_l X + \bar{y}_l Y + \bar{z}_l Z, \quad (2-24)$$

其中,

$$\begin{aligned} \bar{w}_l &= \sqrt{2}/L \\ \bar{x}_l &= 3 \cos \phi_l \cos \theta_l / L \\ \bar{y}_l &= 3 \sin \phi_l \cos \theta_l / L \\ \bar{z}_l &= 3 \sin \theta_l / L \end{aligned} \quad (2-25)$$

### 2.2.3 虚拟扬声器方法

在多通路环绕声系统中, 为了产生空间中某一方向的虚拟源, 并不一定需要在此方向布放一个扬声器, 而是可以在空间中适当地布置若干个扬声器, 通过调整各个扬声器的激励信号来产生不同方位的虚拟源。借鉴多通路环绕声系统, 在双耳合成过程中, 可以分别合成多个不同方位的虚拟扬声器, 通过改变各个虚拟扬声器的激励信号, 即可模拟出多通道声重放中不同虚拟源产生的双耳声信号。虽然这种双耳信号可能只是理想情况的低阶近似, 但利用多声源综合定位的心理学原理, 可在耳机重放中得到相应空间虚拟源的听觉感知<sup>[48][49]</sup>。这就是双耳合成中虚拟扬声器 (virtual loudspeaker) 方法的基本思路。

由 2.2.2 节可以得到声场重现所需的虚拟扬声器摆放方式和对应的扬声器激励信号，为了让 Ambisonics 系统可以用于耳机实现，我们需要对虚拟扬声器信号加以 HRTF 渲染，该方法可以消除听音空间对扬声器回放产生的影响。

假设整个空间中有  $L$  个虚拟声源（虚拟扬声器），其位置为  $(\theta_l, \phi_l)$ ，激励信号为  $w_l(f)$ ，双耳接收到的信号是多个声源共同作用的结果：

$$\begin{aligned} B_L(f) &= \sum_{l=1}^L H_L(\theta_l, \phi_l, f) w_l(f) \\ B_R(f) &= \sum_{n=l}^L H_R(\theta_l, \phi_l, f) w_l(f), \end{aligned} \quad (2-26)$$

其中， $H_L(\theta_l, \phi_l, f)$  和  $H_R(\theta_l, \phi_l, f)$  是声源位于  $(\theta_l, \phi_l)$  处的左、右耳 HRTF 数据。最后对  $B_L(f)$  和  $B_R(f)$  求反傅里叶变换即可得到时域双耳信号。

但是基于 Ambisonics 的虚拟扬声器方法会带来一些问题：一是不同的虚拟扬声器数目和位置会对所获取的双耳信号带来影响，二是扬声器所在位置的声像会不可避免地被加重。并且当扬声器所处方位在现有的 HRTF 数据库中没有对应的数据，此时需要对原始 HRTF 进行插值。

### 2.3 基于球谐分解的双耳渲染算法

近年来，基于球谐分解的双耳渲染算法受到人们的广泛关注。其核心思想是不仅对声场进行球谐分解，同时对 HRTF 进行球谐分解，直接在球谐域上进行渲染。可以避免了虚拟扬声器引入对双耳信号带来的影响。

对声场使用球谐函数进行表示：

$$S(r, \theta, \phi, f) = \sum_{n=0}^{N_s} \sum_{m=-n}^n S_n^m(r, f) Y_n^m(\theta, \phi), \quad (2-27)$$

其中， $S_n^m(r, f)$  是声学场景的球谐系数，无需声源信号和方位的先验信息。 $N_s$  为声场的球谐分解阶次，由录制设备决定。

HRTF 一般在一个固定的球面上进行多点采样，当球面半径大于 1 米时，可以认为是远场 HRTF，可以使用球谐函数进行分解：

$$\begin{aligned} H_L(\theta, \phi, f) &= \sum_{n=0}^{N_h} \sum_{m=-n}^n \beta_{nm}^L(f) Y_n^m(\theta, \phi) \\ H_R(\theta, \phi, f) &= \sum_{n=0}^{N_h} \sum_{m=-n}^n \beta_{nm}^R(f) Y_n^m(\theta, \phi), \end{aligned} \quad (2-28)$$

其中， $N_h$  是 HRTF 进行球谐分解的阶数。

由于声源可能具有一定体积，例如面源、体源，并且房间混响会产生反射，因此假设声场中声源是连续分布的，则左耳的信号可以表示为<sup>[50][51][52]</sup>：

$$\begin{aligned}
 B_L(f) &= \int_0^{2\pi} \int_0^\pi S(r, \theta, \phi, f) [H_L(\theta, \phi, f)]^* \sin \theta d\theta d\phi \\
 &= \int_0^{2\pi} \int_0^\pi \sum_{n=0}^{N_s} \sum_{m=-n}^n S_n^m(r, f) Y_n^m(\theta, \phi) \left[ \sum_{n'=0}^{N_h} \sum_{m'=-n'}^{n'} \beta_{n'm'}^L(f) Y_{n'}^{m'}(\theta, \phi) \right]^* \sin \theta d\theta d\phi \\
 &= \sum_{n=0}^{N_s} \sum_{m=-n}^n \sum_{n'=0}^{N_h} \sum_{m'=-n'}^{n'} S_n^m(r, f) [\beta_{n'm'}^L(f)]^* \int_0^{2\pi} \int_0^\pi Y_n^m(\theta, \phi) [Y_{n'}^{m'}(\theta, \phi)]^* \sin \theta d\theta d\phi \\
 &= \sum_{n=0}^N \sum_{m=-n}^n S_n^m(r, f) [\beta_{nm}^L(f)]^*, \tag{2-29}
 \end{aligned}$$

其中， $N = \min \{N_s, N_h\}$ ，最后一个等式的推导借助了球谐函数的正交性，如式(2-5)所示。

同理，右耳信号可以表示为：

$$B_R(f) = \sum_{n=0}^N \sum_{m=-n}^n S_n^m(r, f) [\beta_{nm}^R(f)]^* \tag{2-30}$$

由 $N = \min \{N_s, N_h\}$ 可知，在球谐域上进行双耳渲染时的球谐阶次由声场和HRTF进行球谐分解的阶次共同决定。HRTF数据库的球面采样点较为密集，可以实现高阶球谐系数的获取，并且研究表明，要想对高达15 kHz的HRTF进行精确的物理表征，大约需要 $N_h = 34$ 的阶次。声场方面，要想准确获取 $N$ 阶声场，至少需要 $(N + 1)^2$ 个采样点，其详细推导见第3.1节，而实际系统中，对 $N_h = 34$ 对应的1225个通道的采集、传输和处理很难实现，一般来说录制设备的通道数较少，所能准确获取的声场系数的阶次较低，基本满足 $N_s < 5$ <sup>[28]</sup>。二者之间的阶次存在不匹配问题，HRTF虽然可以获取高阶球谐分量，但也只能使用低阶进行表示。HRTF的直接低阶截断会导致高频能量损失，带来一定的误差。除此之外，还存在其他问题，例如录音设备的尺寸小于人头尺寸，录音设备的采样率与HRTF数据库采样率不一致等。这些不匹配问题将会导致所获取的双耳信号的空间分辨率受限、定位线索（双耳时间差ITD、双耳声级差ILD）的损伤、空间感的下降等，直接影响声源方位的感知、外化(externalization)、感知声源宽度和音色等<sup>[27][32][33][34]</sup>。

本文主要针对以上问题，对基于球谐分解的双耳渲染算法进行研究，对头相关传递函数HRTF的预处理加以研究，并且从另外一个方面——声场出发，提出改进方法。

## 2.4 本章小节

本章主要介绍了几种常见的双耳渲染技术及其原理。首先对基于 HRIR 和 BRIR 的虚拟听觉的方法进行了简单介绍，其中头相关传递函数 HRTF 是所有双耳渲染算法中的核心数据。接下来对基于 Ambisonics 的虚拟扬声器方法进行了详细推导，其核心思想是对声场的球谐分解，再利用各阶谐波分量求解重建声场所需的扬声器设置和对应的激励信号，最后结合虚拟扬声器获取双耳信号，其中对声场进行正交球谐分解是本文的重要理论基础。最后对基于球谐分解的双耳渲染技术进行了介绍和理论推导，并对其所存在的问题进行了简单描述，本文将在此算法框架下进行，针对其所存在的问题，提出改进算法。



## 第三章 基于 HRTF 球谐分解的双耳渲染算法

对 HRTF 进行球谐分解并在球谐域上直接渲染的双耳渲染算法近年来得到了广泛关注，本章将在 2.3 节的基础上对其进行详细介绍，其包含两个同等重要的部分：声场和 HRTF。在声场获取方面，麦克风阵列尤其是球形麦克风阵列可以更好地采集声场的特征，具有高质量、高保真的录制效果，是多通道录制与重构、空间音频技术的重要工具。首先对空心球和刚性球进行讨论和对比，分别给出其球谐系数的求解方法。接下来对 HRTF 的获取和其球谐分量的求解方法进行介绍。并且在此基础上，对 HRTF 的预处理方法的原理进行详细介绍，该方法可以有效降低 HRTF 的球谐分解阶次。最后通过实验验证了预处理方法的有效性。

### 3.1 球形麦克风阵列的声场录制及球谐系数估计

声场录制的关键在于如何准确获得待录制声场的高阶球谐系数，使用球形麦克风阵列可以实现此目标。球形麦克风阵列是通过将麦克风放置在三维空间中并记录麦克风所在位置的声压信号来实现的。影响球阵可以准确重现声场的因素包括频率、阵列半径、阵元数目和球面采样方式等<sup>[53]</sup>。录制声场球谐系数的估计依赖于麦克风阵列的采样配置，在一些特殊的采样方式<sup>[45]</sup>例如等角采样、高斯采样和均匀采样（如图 3-1 所示）的情况下，可以使用求和来代替积分计算录制声压的球谐系数，能够方便地计算函数的球傅里叶变换。在实际系统中很难要求所有球阵满足特殊的采样方式，因此需要给出在任意采样配置下使用球傅里叶变换来计算球谐系数的方法。

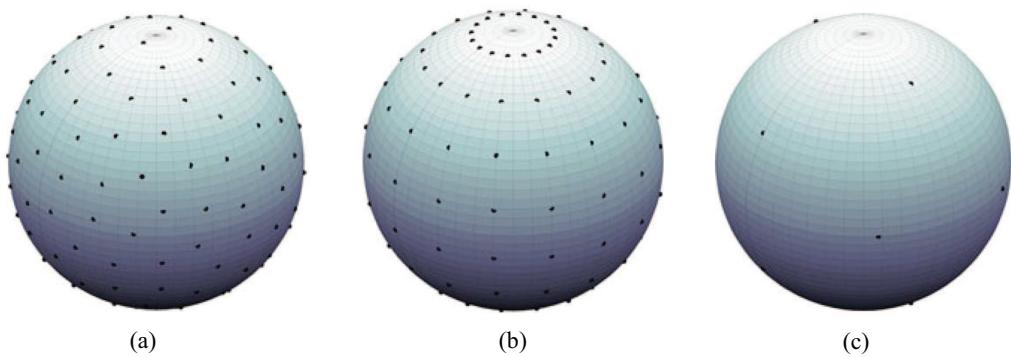


图 3-1 球面采样：(a) 等角采样；(b) 高斯采样；(c) 均匀采样。

使用半径为  $R$  的球形麦克风阵列采集来自  $(\theta_s, \phi_s)$  的单位幅度平面波时，麦克风  $(\theta_q, \phi_q)$  连续分布在球面上。此处假设声源均位于球形麦克风阵列外，麦克风观测

声压可分为空心球和刚性球两种情况进行讨论<sup>[43, 45]</sup>。

### 3.1.1 空心球的球谐系数估计

对于空心球来说，忽略单个麦克风和支架对声波的影响，可以认为其录制的声场即为真实环境下的声场，直接进行球谐系数的估计。

对于特殊的采样方式，球面面积分可以通过  $Q$  个球面采样  $(\theta_q, \phi_q)$  的加权求和来逼近，如式(3-1)所示：

$$\int_0^{2\pi} \int_0^\pi g(\theta, \phi) \sin \theta d\theta d\phi \approx \sum_{q=1}^Q \alpha_q g(\theta_q, \phi_q), \quad (3-1)$$

其中， $\alpha_q$  是采样权重，式(3-2)、(3-3) 和 (3-4) 分别为等角采样、高斯采样和均匀采样下的权重表达式：

$$\alpha_q = \frac{2\pi}{(N+1)^2} \sin(\theta_q) \sum_{q'=0}^N \frac{1}{2q'+1} \sin([2q'+1]\theta_q). \quad (3-2)$$

$$\alpha_q = \frac{\pi}{N+1} \frac{2(1-\cos^2 \theta_q)}{(N+2)^2 [P_{N+2}(\cos \theta_q)]^2}. \quad (3-3)$$

$$\alpha_q = \frac{4\pi}{Q}. \quad (3-4)$$

则通过球傅里叶变换获取球谐系数的式(2-8)可以表示为：

$$S_n^m(r, f) = \sum_{q=1}^Q \alpha_q S(r, \theta_q, \phi_q, f) [Y_n^m(\theta_q, \phi_q)]^*. \quad (3-5)$$

此时，球谐系数可以直接采用上式进行求解。

在某些情况下，从这些或其他预定义的采样配置中选择采样集可能是不可行的，例如，由于麦克风定位的机械限制。因此，在任意给定采样集时，如何计算球面球谐系数在实际应用中具有重要的价值。考虑在单位球面上进行采样，根据球傅里叶变换，第  $q$  个麦克风的接收信号可以写为：

$$S(r, \theta_q, \phi_q, f) = \sum_{n=0}^{N_s} \sum_{m=-n}^n S_n^m(r, f) Y_n^m(\theta_q, \phi_q), \quad (3-6)$$

其矩阵形式为：

$$\mathbf{S} = \mathbf{Y} \mathbf{S}_{nm}, \quad (3-7)$$

其中， $\mathbf{S}$  和  $\mathbf{S}_{nm}$  是长度为  $Q$  和  $(N_s + 1)^2$  的列向量， $\mathbf{Y}$  是大小为  $Q \times (N_s + 1)^2$  的矩阵，

$$\mathbf{S} = [S(r, \theta_1, \phi_1, f), S(r, \theta_2, \phi_2, f), \dots, S(r, \theta_Q, \phi_Q, f)]^T, \quad (3-8)$$

$$\mathbf{S}_{nm} = [S_0^0, \dots, S_n^m, \dots, S_{N_s}^{N_s}]^T, \quad (3-9)$$

$$\mathbf{Y} = \begin{bmatrix} Y_0^0(\theta_1, \phi_1) & \cdots & Y_n^m(\theta_1, \phi_1) & \cdots & Y_{N_s}^{N_s}(\theta_1, \phi_1) \\ Y_0^0(\theta_2, \phi_2) & \cdots & Y_n^m(\theta_2, \phi_2) & \cdots & Y_{N_s}^{N_s}(\theta_2, \phi_2) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_0^0(\theta_Q, \phi_Q) & \cdots & Y_n^m(\theta_Q, \phi_Q) & \cdots & Y_{N_s}^{N_s}(\theta_Q, \phi_Q) \end{bmatrix}. \quad (3-10)$$

当  $Q = (N + 1)^2$  时, 式(3-7)的解可以通过矩阵  $\mathbf{Y}$  的逆求解(要求球形麦克风阵列的采样位置满足矩阵  $\mathbf{Y}$  可逆的条件):

$$\mathbf{S}_{nm} = \mathbf{Y}^{-1} \mathbf{S}, \quad (3-11)$$

在许多情况下都使用过采样, 即  $Q > (N + 1)^2$ , 此时式(3-7)超定, 其最小二乘解由矩阵的伪逆给出::

$$\mathbf{S}_{nm} = \mathbf{Y}^\dagger \mathbf{S} = (\mathbf{Y}^H \mathbf{Y})^{-1} \mathbf{Y}^H \mathbf{S}. \quad (3-12)$$

当  $Q < (N + 1)^2$  时, 采样点数不足, 式(3-7)可能不能提供正确的解。

### 3.1.2 刚性球的球谐系数估计

和空心球相比, 刚性球录制的总声压场, 也就是实验中的被测声压, 可以分为入射声场和散射声场两个部分。从实验的角度来说, 入射声场是针对球体不存在时测量的声压场, 而散射声场是球体存在时声压场的变化量。对于刚性球来说, 其表面的径向速度为 0, 即球体表面处入射场和散射场的总径向速度为 0, 满足 Neumann 边界条件<sup>[43]</sup>。如果散射场为零, 则认为该球体是完全可透射的。

大多数散射问题的研究都是从平面波着手的, 对于来波方向为  $(\theta_s, \phi_s)$  的单幅度平面波来说, 麦克风  $(r, \theta_q, \phi_q)$  处观测到的入射声压(直达声)可以用球贝塞尔函数表示:

$$S_{\text{dir}}(r, \theta_q, \phi_q, f) = \sum_{n=0}^{\infty} 4\pi i^n j_n(kr) \sum_{m=-n}^n [Y_n^m(\theta_s, \phi_s)]^* Y_n^m(\theta_q, \phi_q). \quad (3-13)$$

上式即为空心球情况下麦克风的接收信号。

因为散射声场代表了入射声波在球面的散射而发出的球面声波, 因此散射声压可以利用表示去波的第一类球汉克尔函数展开:

$$S_{\text{scat}}(r, \theta_q, \phi_q, f) = - \sum_{n=0}^{\infty} 4\pi i^n \frac{j'_n(kr)}{(h_n^{(1)})'(kr)} h_n^{(1)}(kr) \sum_{m=-n}^n [Y_n^m(\theta_s, \phi_s)]^* Y_n^m(\theta_q, \phi_q), \quad (3-14)$$

其中,  $j'_n(kr)$  和  $(h_n^{(1)})'(kr)$  分别为  $n$  阶球贝塞尔函数  $j_n(kr)$  和  $n$  阶第一类球汉克尔函数  $h_n^{(1)}(kr)$  的导数。波动方程在频域上的表示称为亥姆霍兹方程, 球坐标系下亥姆霍兹方程中只与距离相关的项所构成的方程被称为球贝塞尔方程, 球贝塞尔函数和球汉克尔函数共同组成了球贝塞尔方程的解。

对于刚性球来说，麦克风的录制信号中除了直达声外，还包含因球面散射形成的散射声，则接收到的总声压可表示为：

$$\begin{aligned} S(r, \theta_q, \phi_q, f) &= S_{\text{dir}}(r, \theta_q, \phi_q, f) + S_{\text{scat}}(r, \theta_q, \phi_q, f) \\ &= \sum_{n=0}^{\infty} 4\pi i^n \left( j_n(kr) - \frac{j'_n(kr)}{(h_n^{(1)})'(kr)} h_n^{(1)}(kr) \right) \sum_{m=-n}^n [Y_n^m(\theta_s, \phi_s)]^* Y_n^m(\theta_q, \phi_q), \end{aligned} \quad (3-15)$$

整合公式 (3-13) 和式 (3-15)，可以得到采用不同麦克风阵列时，第  $q$  个麦克风接收信号的表达式：

$$S(r, \theta_q, \phi_q, f) = \sum_{n=0}^{\infty} b_n(kr) \sum_{m=-n}^n [Y_n^m(\theta_s, \phi_s)]^* Y_n^m(\theta_q, \phi_q), \quad (3-16)$$

其中， $b_n$  表示模态强度：

$$b_n(kr) = \begin{cases} 4\pi i^n j_n(kr) & \text{空心球} \\ 4\pi i^n \left( j_n(kr) - \frac{j'_n(kr)}{(h_n^{(1)})'(kr)} h_n^{(1)}(kr) \right) & \text{刚性球.} \end{cases} \quad (3-17)$$

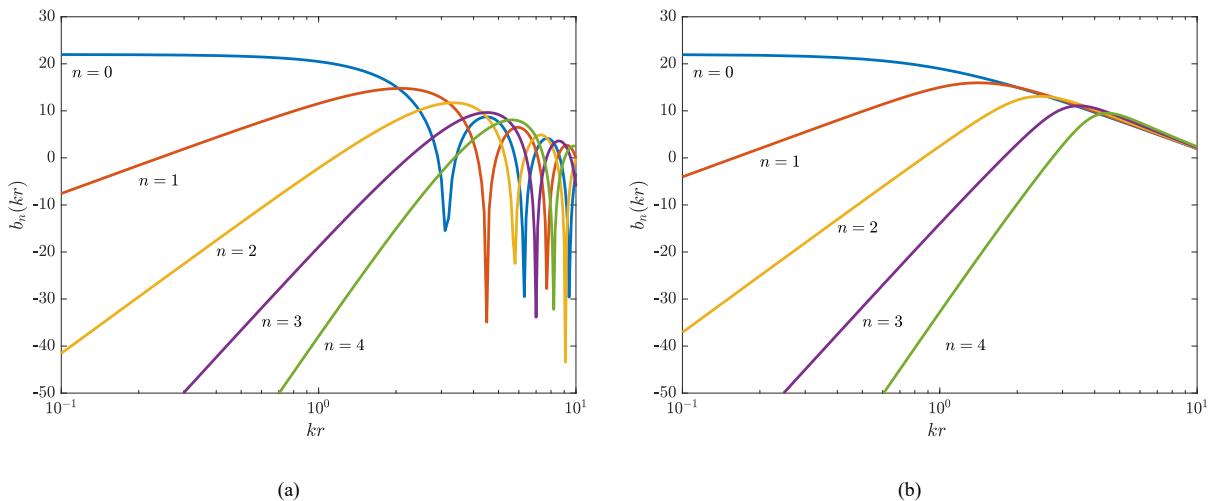


图 3-2 模态强度  $b_n$ : (a) 空心球模态强度; (b) 刚性球模态强度

图 3-2 为模态强度  $b_n$  随  $kr$  的变化趋势图。(a) 中曲线有很深的凹陷，即在某些频率处模态强度为 0。而 (b) 中刚性球可以避免空心球所带来的凹陷（零点）问题，这是因为刚性球麦克风接收到的不仅包含直达声还有声波入射到球体产生的散射声。关于零点问题带来的影响将在第 4.1 节中关于球谐域 MUSIC 定位算法和第 5.2.2 节中的实验结果进行展现。

并且由于散射声场的存在，在高阶次和低频率的情况下，刚性球模态强度的幅度稍大于空心球，因此刚球对传感器噪声和其他误差的影响的鲁棒性有些许提升<sup>[45]</sup>。另

外，实际情况下由于麦克风数目较多，所以做成刚性球更合理，目前常用的刚球为 32 通路 Eigenmike<sup>[24]</sup>。

## 3.2 HRTF 的球谐分解

本节将对头相关传递函数 HRTF 的三种获取方式以及本文所采用的两种数据库加以介绍，并且给出了如何从给定 HRTF 数据库中求解 HRTF 球谐系数的方法。

### 3.2.1 HRTF 的获取

HRTF 和 HRIR 是双耳渲染算法中合成双耳信号的重要数据基础，其可以通过计算、定制与测量三种方法获取<sup>[18]</sup>。

原则上，声波经由头部、耳廓、躯干等生理结构的综合滤波过程可以通过求解加以边界条件约束的波动方程来获得，即边界元法（boundary element method, BEM）<sup>[54]</sup>。但实际上边界表面的不规则性会产生及其复杂的现象，以足够的准确度实现边界表面（尤其是耳廓）的测量具有一定的挑战，并且计算量较大。还有一些分析解决方案适用于非常简单的几何形状，在简化情况下，忽略耳廓和躯干的影响，将头部近似为刚性球体。在此基础上，雪人模型<sup>[55]</sup>将头部和躯干简化为两个不同半径的球，并采用多重散射或多级展开的方法进行求解，数学方法非常复杂。目前已知的仿真 HRTF 数据库为 MESH2HRTF<sup>[56]</sup>，其采用边界元法，频率范围为 100Hz 到 22kHz，以 100Hz 为间隔，球面采样点数为 4334 个，采样方式为列别捷夫采样。

HRTF 的定制主要应用于获取个性化 HRTF，假定 HRTF 和一定的生理参数之间存在高的相关性。主要包括生理参数匹配法、频率标度变换法和生理参数线性回归法，近几年也出现了基于机器学习的个性化 HRTF 生成模型。总体上，定制 HRTF 较测量和计算来说比较简单，且可以得到一定的效果，但其准确性有待进一步改进<sup>[18]</sup>。

现有的技术中，实验测量是获得 HRTF 最重要且最准确的手段<sup>[18]</sup>，一个扬声器在特定的位置产生一个脉冲或其他宽带信号，然后通过记录放置在人工头或真人受试者耳道内的两个麦克风的输出信号来测量。根据 HRTF 的定义，理想情况下的测量应该在自由场（消声室）的环境下进行，但实际上也经常采用非消声室的测量方法，为了避免环境反射声的影响，会加以时间窗处理。

双耳麦克风的尺寸很小可以放置在耳道入口处，并使用泡沫环保持在适当的位置，这被称为封闭耳道法，进入耳道的信号被麦克风捕获并存储为双通道录音。封闭耳道法经常用于 HRTF 的测量<sup>[1]</sup>。

目前国际上已有许多课题组对人工头和真人受试者的远场 HRTF/HRIR 进行了测量，建立了相应的数据库<sup>[18]</sup>，大多数数据库都是在远场条件下进行测量，但也有部分

是在近场进行。现有的较为常用的远场 HRTF 数据库包括 MIT 数据库<sup>[8]</sup> 和 CIPIC 数据库<sup>[57]</sup>，其对应的数据已在网上公开。

除此之外，Benjamin Bernschutz 使用 Neumann KU-100 建立了一个远场 HRTF 数据库<sup>[58]</sup>，测量环境为消声室，共采用了 4 种球面采样方式，包括在以  $1^\circ$  为角度间隔的圆周上均匀采样，2354、2702 点球面列别捷夫采样和以  $2^\circ$  为间隔的球面高斯采样（共计 16020 个采样点），此数据库将在第 3.3.2 节用于体现 HRTF 预处理的结果展示。

本文主要采用 CIPIC 数据库进行双耳渲染，其来自加州大学戴维斯分校 CIPIC 交互实验室。在非消声室环境下，对 43 名真人受试者（27 男，16 女）进行测量，是目前真人 HRTF 数据库中空间分辨率较高，同时受试者人数较多的一个，同时给出了真人受试者的生理尺寸<sup>[18]</sup>。其测量环境为非消声室环境，测量方法为封闭耳道法，HRIR 的长度为 200 点，采样频率为 44.1 kHz。其测量时采用双耳极坐标系统，如图 3-3 (a) 所示，其中， $(\Theta, \Phi) = (0^\circ, 0^\circ)$  为正前方， $(\Theta, \Phi) = (90^\circ, 0^\circ)$  为正右方。

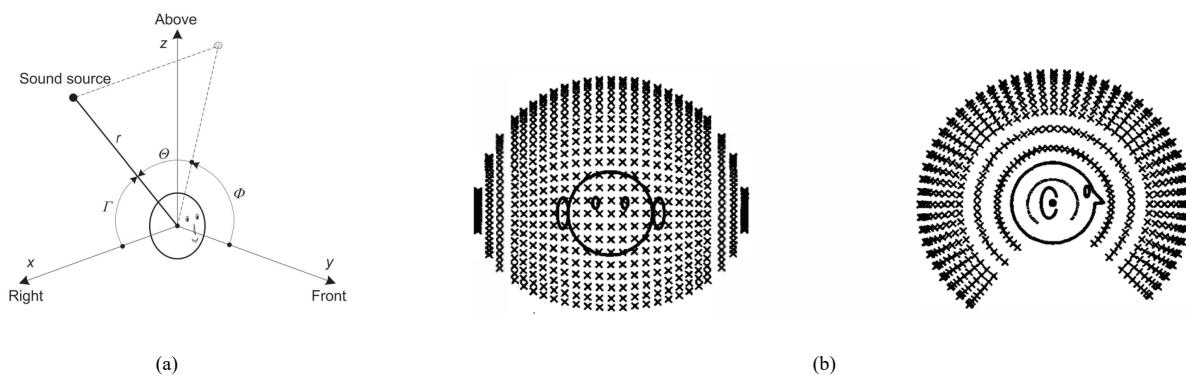


图 3-3 CIPIC 数据库角度说明：(a) 双耳极坐标系<sup>[18]</sup>，(b) 声源位置分布

如图 3-3 (b) 所示，CIPIC 数据库共包括 1250 个测量位置，50 个俯仰角为  $\Theta = -45^\circ + 5.625^\circ \times (0 : 1 : 49)$ ，每个俯仰角均对应 25 个方位角，且非均匀分布，其设定为：

$$[-80 \ -65 \ -55 \ -45 : 5 : 45 \ 55 \ 65 \ 80]$$

以 003 号受试者为例，当声源位于水平面且方位角为  $80^\circ$  时，即声源位于右耳方向附近，左右耳的 ITD 和 ILD 取接近最大值，对应的左、右耳的 HRIR 和 HRTF 如图 3-4 所示。方位角为  $0^\circ$  时，即声源位于正前方，如图 3-4 所示。

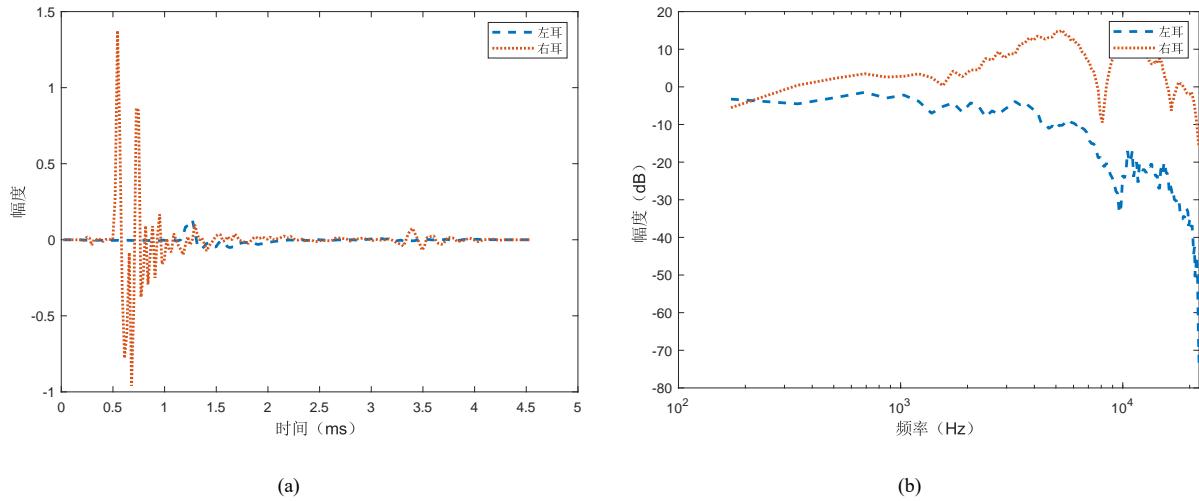


图 3-4 CIPIC 数据库中声源位于水平面且方向为  $80^\circ$  时的左右耳 HRIR 和 HRTF 数据:(a) HRIR, (b) HRTF

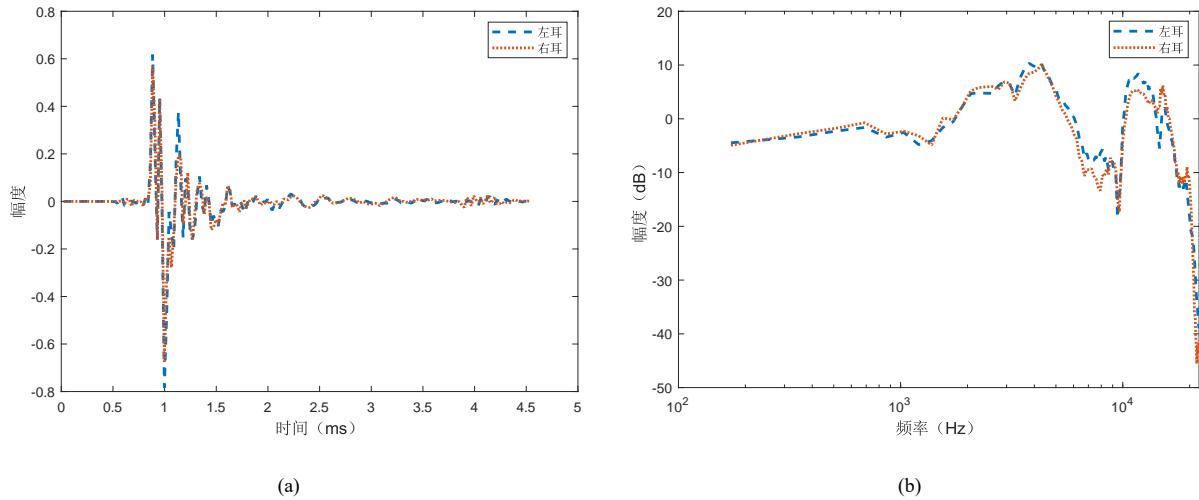


图 3-5 CIPIC 数据库中声源位于水平面且方向为  $0^\circ$  时的左右耳 HRIR 和 HRTF 数据:(a) HRIR, (b) HRTF

### 3.2.2 HRTF 的球谐系数估计

由于 CIPIC 数据库采用双耳极坐标系,与球谐函数的角度定义不一致,所以在 HRTF 球谐分解之前,需要对 CIPIC 数据库的角度进行转换。在转换过程中也可以借助 CIPIC 数据库的 SOFA 格式数据,在 SOFA 格式中,角度定义与图 2-2 类似,仅在俯仰角  $\theta$  的定义上不同,其俯仰角为方向矢量与水平面的夹角。

同时为了方便分析,本文对方位角重新定义。图 2-2 中的俯仰角  $\theta$  为方向矢量与  $z$  正

轴的夹角,  $\theta = 0^\circ, 90^\circ, 180^\circ$  分别表示正上方、水平面和正下方, 其中水平面是指由原点指向正前方和正右(或左)方的两矢量决定的平面。方位角  $\phi$  为方向矢量在水平面的投影与中垂面的夹角, 在水平面上,  $\phi = 0^\circ, 90^\circ, 180^\circ, 270^\circ$  分别表示正前、正左、正后和正右方。调整后, 俯仰角不变, 在水平面上, 方位角  $\phi = 0^\circ, 90^\circ, 180^\circ, 270^\circ$  分别表示正前、正右、正后和正左方。详见表格 3-1。

表 3-1 转换前后的角度定义

	SOFA 格式角度定义	球谐函数角度定义	最终角度定义
俯仰角 $\theta$	矢量与水平面的夹角 $-90^\circ \leq \theta \leq 90^\circ$	矢量与 $x$ 正轴的夹角 $0 \leq \theta \leq 180^\circ$	矢量与 $x$ 正轴的夹角 $0 \leq \theta \leq 180^\circ$
正上方	$\theta = 90^\circ$	$\theta = 0^\circ$	$\theta = 0^\circ$
水平面	$\theta = 0^\circ$	$\theta = 90^\circ$	$\theta = 90^\circ$
正下方	$\theta = -90^\circ$	$\theta = 180^\circ$	$\theta = 180^\circ$
正前方	$\phi = 0^\circ$	$\phi = 0^\circ$	$\phi = 0^\circ$
正右方	$\phi = 270^\circ$	$\phi = 270^\circ$	$\phi = 90^\circ$
正后方	$\phi = 180^\circ$	$\phi = 180^\circ$	$\phi = 180^\circ$
正左方	$\phi = 90^\circ$	$\phi = 90^\circ$	$\phi = 270^\circ$

从 SOFA 格式角度定义到最终角度定义的转换公式为:

$$\begin{aligned}\phi &= \text{mod} (360^\circ - \phi_{\text{SOFA}}, 360^\circ) \\ \theta &= 90^\circ - \theta_{\text{SOFA}},\end{aligned}\quad (3-18)$$

其中,  $\text{mod}$  表示取余。

接下来对进行角度转换后的 HRTF 数据进行球谐分解。由 2.3 节可知,  $P$  个方位的 HRTF 可以使用球谐函数进行分解:

$$H_{\text{L,R}}(\theta_p, \phi_p, f) = \sum_{n=0}^{N_h} \sum_{m=-n}^n \beta_{nm}^{L,R}(f) Y_n^m(\theta_p, \phi_p) \quad (3-19)$$

球谐系数的求解方式  $\beta_{nm}^{L,R}(f)$  的方法与式 (3-7) 类似:

$$\mathbf{H}_{\text{L,R}} = \mathbf{Y} \boldsymbol{\beta}_{nm}^{\text{L,R}}, \quad (3-20)$$

其中,  $\mathbf{Y}$  是大小为  $P \times (N_h + 1)^2$  的矩阵,  $\mathbf{H}_{\text{L,R}}$  和  $\boldsymbol{\beta}_{nm}^{\text{L,R}}$  是长度为  $P$  和  $(N_h + 1)^2$  的列向量,

$$\mathbf{Y} = \begin{bmatrix} Y_0^0(\theta_1, \phi_1) & \cdots & Y_n^m(\theta_1, \phi_1) & \cdots & Y_{N_h}^{N_h}(\theta_1, \phi_1) \\ Y_0^0(\theta_2, \phi_2) & \cdots & Y_n^m(\theta_2, \phi_2) & \cdots & Y_{N_h}^{N_h}(\theta_2, \phi_2) \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ Y_0^0(\theta_P, \phi_P) & \cdots & Y_n^m(\theta_P, \phi_P) & \cdots & Y_{N_h}^{N_h}(\theta_P, \phi_P) \end{bmatrix}, \quad (3-21)$$

$$\mathbf{H}_{\text{L,R}} = [H_{\text{L,R}}(\theta_1, \phi_1, f), H_{\text{L,R}}(\theta_2, \phi_2, f), \dots, H_{\text{L,R}}(\theta_P, \phi_P, f)]^T, \quad (3-22)$$

$$\boldsymbol{\beta}_{nm}^{\text{L,R}} = [\beta_{0,0}^{\text{L,R}}, \dots, \beta_{n,m}^{\text{L,R}}, \dots, \beta_{N_h,N_h}^{\text{L,R}}]^T. \quad (3-23)$$

对于 HRTF 数据库来说，球面采样一般较为密集，假设采样点总数满足  $P > (N_h + 1)^2$ ，因此可以采用式 (3-24) 进行球谐分量的求解。

$$\boldsymbol{\beta}_{nm}^{\text{L,R}} = \mathbf{Y}^\dagger \mathbf{H}_{\text{L,R}} = (\mathbf{Y}^H \mathbf{Y})^{-1} \mathbf{Y}^H \mathbf{H}_{\text{L,R}}. \quad (3-24)$$

球谐分量  $\boldsymbol{\beta}_{nm}^{\text{L,R}}$  也可以用于重建任意方位的 HRTF，又称为 HRTF 插值。任意方向  $(\theta, \phi)$  的 HRTF 的表达式为：

$$\hat{H}_{\text{L,R}}(\theta, \phi, f) = \mathbf{y} \boldsymbol{\beta}_{nm}^{\text{L,R}}. \quad (3-25)$$

其中， $\mathbf{y} = [Y_0^0(\theta, \phi), \dots, Y_n^m(\theta, \phi), \dots, Y_{N_h}^{N_h}(\theta, \phi)]$  为重建方位  $(\theta, \phi)$  对应的球谐函数向量。

### 3.3 HRTF 预处理

从 2.3 节可知，直接使用低阶截断的 HRTF 分量进行双耳渲染会带来许多问题，因此需要引入 HRTF 预处理方法进行改进。一些 HRTF 预处理方法通过寻找对 HRTF 更为紧凑的表示，使大部分能量集中在球谐分解的低阶分量上。本节将对 HRTF 的预处理方法之一——双耳对准加以介绍，并且通过实验验证了该方法的有效性。

#### 3.3.1 预处理方法

为了降低 HRTF 的低阶球谐表示所引起的误差，建议引入预处理来降低 HRTF 的有效球谐阶数。对于式 (3-19) 中对 HRTF 的球谐分解，其高阶球谐分量主要是由高频处快速的相位变化导致的。并且研究表明，去除或者改变高频处的双耳相位差 (interaural phase difference, IPD) 在感知上无关紧要<sup>[33]</sup>。在此基础上，对 HRTF 进行相位修正的预处理方法被用于 HRTF 低阶球谐分解的扩展工作。例如，对高频应用时间对准，即通过减去高频部分的线性相位部分来实现 IPD 均衡<sup>[28][36]</sup>，可以降低 HRTF 数据库的空间复杂度，从而降低所需的球谐分解阶次。

知觉研究表明，听觉系统在判断声源方位时所使用的声学信息是有选择性的，在自由场情况下，人耳听觉系统对单一声源进行定位时，主要定位因素双耳时间差 ITD 和双耳声级差 ILD 随着频率的作用范围如下所示：

- 在频率小于 1.5 kHz 的情况下，双耳时间差 ITD 是方向定位的主要因素；
- 在频率为 1.5 kHz 到 4 kHz 的情况下，双耳时间差 ITD 和双耳声级差 ILD 对方向定位共同起作用；

- 在频率大于 4 kHz 到 5 kHz 的情况下，双耳声级差 ILD 是方向定位的主要因素。

基于高频处 HRTF 的幅度部分对定位的感知起决定作用这一事实，在本研究中，采用了与之前相同的处理方法，对高频相位加以修正。主要关注对高频处 HRTF 幅度的近似，而相位从临界频率开始加以改变。

加以相位对准的头相关传递函数  $H_{L,R}^a(\theta, \phi, f)$  可表示为：

$$H_{L,R}^a(\theta, \phi, f) = \begin{cases} H_{L,R}(\theta, \phi, f), & f < f_c \\ H_{L,R}(\theta, \phi, f)e^{i\zeta}, & f \geq f_c, \end{cases} \quad (3-26)$$

其中， $f_c$  为临界频率，一般选取 1.5 kHz。对于给定的频率， $\zeta$  的表达式为：

$$\begin{aligned} \zeta &= \angle([H_{L,R}(\theta, \phi, f)]^* H_{L,R}(\theta, \phi, f_c)) \\ &= \angle(H_{L,R}(\theta, \phi, f_c)) - \angle(H_{L,R}(\theta, \phi, f)), \end{aligned} \quad (3-27)$$

其中， $\angle(\cdot)$  表示取一个复数的相位信息。

上述公式给出了直接从 HRTF 数据中获取相位差的方法，也可以根据声源方位和人头尺寸等参数获取，如下所示：

$$\zeta = 2\pi(f_c - f)\tau_p^{L,R}, \quad (3-28)$$

其中， $\tau_p^{L,R}$  为声源位于  $(\theta_p, \phi_p)$  位置处对应的时间偏移，即头中心和耳朵之间的时间差。

假设为远场条件，此时声源信号可以看作平面波，因此  $\tau_p^{L,R}$  的计算公式为：

$$\tau_p^r = \cos(\theta_p) \sin(\phi_p) a/c, \quad \tau_p^l = -\tau_p^r \quad (3-29)$$

使用式 (3-28) 通过时间差计算相位差从而对 HRTF 进行对准的方法一般被称为时间对准。时间对准的 HRTF 和相位对准的 HRTF 方法的原理基本一致，核心思想是将原有的 HRTF 中使用自由场下头中心处的声压对双耳声压进行归一化转换为使用双耳所在位置的声压对双耳声压进行归一化，此处方法已被证明是一种有效降低 HRTF SH 阶次的稳健方法，有很多工作基于此进行 [28][37][38]。

和直接使用最小二乘方法求解头相关传递函数的球谐分量相比，该预处理方法一方面准确地表示头相关传递函数的幅度谱，并有效降低了头相关传递函数的分解阶次。

请注意，与其他以前建议的有效的 HRTF 表示方法不同，该相位校正过程是可逆的，这意味着从原始 HRTF 到经过对准的 HRTF 再到原始 HRTF 不会丢失任何信息。

### 3.3.2 实验结果与分析

为了衡量 HRTF 预处理算法对球谐分解所带来的影响，定义 HRTF 球谐分解分量在  $n$  阶下的能量  $E_{L,R}(f, n)$  和所有阶次下的累积能量  $E_{L,R}^C(f, n)$  为：

$$E_{L,R}(f, n) = \sum_{m=-n}^n |\beta_{nm}^{L,R}(f)|^2, \quad (3-30)$$

$$E_{L,R}^C(f, n) = \sum_{n=0}^{N_h} \sum_{m=-n}^n |\beta_{nm}^{L,R}(f)|^2. \quad (3-31)$$

为了避免测试环境对预处理前后阶次变化的影响，本节采用消声室环境下使用 Neumann KU-100 在 2702 点球面列别捷夫采样下的远场数据库 [58]。在计算过程中，对能量进行归一化，并对归一化能量取对数  $10\lg(E_{L,R}(f, n))$ 。预处理前后的球谐分量的能量分布如图 3-6 所示，图中黑色实线表示有效球谐分解阶次，即前  $N$  阶球谐分量包含了 HRTF 99% 的能量。

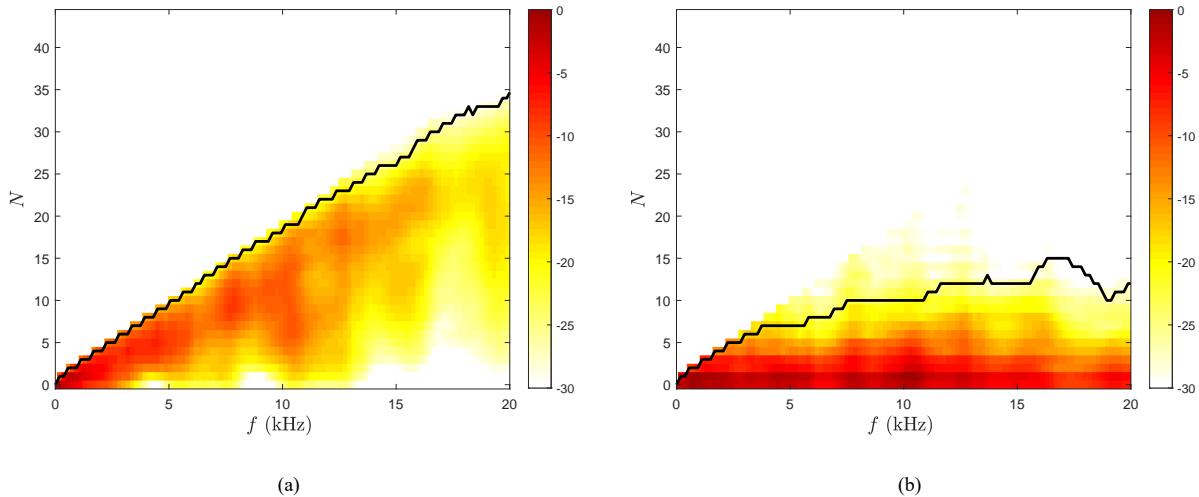


图 3-6 预处理前后 HRTF 的能量分布：(a) 预处理前，(b) 预处理后

从图 3-6 可以看出：

- (1) 未引入预处理时，HRTF 球谐分解所需阶次较大，并且随着频率的增加，球谐分解阶次逐渐增加；
- (2) 对头相关传递函数加以预处理后，HRTF 球谐分解所需阶次较小，并且随着频率的增加，球谐分解阶次趋于稳定；
- (3) 与频率相关的双耳对准预处理算法可以有效降低 HRTF 的分解阶次，由原来的 34 阶降低至 15 阶。

### 3.4 本章小结

本章对基于球谐分解的双耳渲染进行了详细的理论推导，并针对其存在的不匹配问题，对 HRTF 进行预处理。首先，对声场的球谐系数估计进行推导，并且针对空心球和刚性球两种情况给出了不同的计算方法。其次对 HRTF 数据库、HRTF 的球谐系数估计以及对 HRTF 数据库进行球谐分解的前提——角度转换进行了详细介绍。并引入 HRTF 预处理方法，以降低 HRTF 的球谐分解阶次和提升双耳感知。实验结果表明，与频率相关的双耳对准预处理算法可以有效降低 HRTF 的分解阶次。

## 第四章 基于声场扩阶的双耳渲染算法

由于实际情况的限制，基于球谐分解的双耳渲染算法在声场和 HRTF 之间存在不匹配问题，第三章中对 HRTF 进行相位对准预处理，以降低 HRTF 的球谐分解阶次。声场在算法中和 HRTF 有着同等重要的地位，但是目前的算法均集中于 HRTF 预处理。本章将从声场角度出发，提出一种新的声场扩阶算法，以提升声场的球谐分解阶次。首先对信号模型进行介绍，其次对声场扩阶算法的原理进行详细介绍，即通过对原始录制声场进行空域加窗，以实现对录制声场分解阶次的提升，接下来对本文采用的空域窗函数加以介绍。并且将声场扩阶方法与 HRTF 预处理方法相结合，给出整体算法框图，共同解决目前双耳渲染中存在的问题。

### 4.1 信号模型

考虑位于远场的一个声源信号和  $Q$  个麦克风，将麦克风在某个时频点上的接收信号建模为：

$$\mathbf{S} = \mathbf{S}_p + \mathbf{S}_d, \quad (4-1)$$

其中，第一项  $\mathbf{S}_p$  表示平面波，与声源的方位相关。其中，声源方位可以是直达波，也可以是早期反射，描述为在该时频点上能量较大的信号。第二项  $\mathbf{S}_d$  表示扩散声场 (diffuse field)，由来自许多方向的不相关的平面波组成。第一项和第二项分别表示声源分量和环境分量。

该模型假设来源于语音信号的稀疏性。对于语音信号来说，常采用短时傅里叶变换 (Short Time Fourier Transform, STFT)，其在时频域上具有稀疏性，语音能量不仅集中在少数时间帧上，更集中在少量的时频单元上。

其中，第一项  $\mathbf{S}_p$  对应的声源方位可以通过许多方法进行获取，估计声源的到达方向 (direction of arrival, DOA) 一直是人们关注的领域。基于波束形成、最大似然以及子空间方法等各种 DOA 估计方法之前已被广泛研究，并且在球形麦克风阵列上加以扩展 [59]，接下来将简单介绍。

MUSIC 方法是非常经典的子空间 DOA 方法，其算法简单、高空间分辨率高，是目前比较流行的 DOA 估计算法之一，基于球形麦克风阵列的 MUSIC 方法可以在空域和球谐域上进行，本文主要关注基于球阵的球谐域 MUSIC 算法。对球形麦克风阵列的采集信号进行 SH 变换转换到球谐域的好处是，与传统的空域 MUSIC 算法相比，球谐域 MUSIC 算法通过解耦去除球谐域中与频率相关的分量，其阵列流形矩阵与频率无关，

独立于阵列设置，且计算简单<sup>[59]</sup>。

假设有  $J$  个幅度为  $s_j(f)$ , 来波方向为  $(\theta_j, \phi_j)$  的平面波, 由公式(3-16)可得, 第  $q$  个麦克风的接收信号和其球谐分解系数分别为:

$$S(r, \theta_q, \phi_q, f) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \sum_{j=1}^J s_j(f) b_n(kr) Y_n^m(\theta_q, \phi_q) [Y_n^m(\theta_j, \phi_j)]^*, \quad (4-2)$$

和

$$S_n^m(r, f) = b_n(kr) \sum_{j=1}^J s_j(f) [Y_n^m(\theta_j, \phi_j)]^*. \quad (4-3)$$

定义  $s_n^m(f) = S_n^m(r, f)/b_n(kr)$ , 式(4-3)可写为:

$$s_n^m(f) = \sum_{j=1}^J s_j(f) [Y_n^m(\theta_j, \phi_j)]^*, \quad (4-4)$$

其矩阵形式为:

$$\mathbf{s}_n^m = \mathbf{Y}^H \mathbf{s}, \quad (4-5)$$

其中,  $\mathbf{s}_n^m$  是一个长度为  $(N_s + 1)^2$  的向量:

$$\mathbf{s}_n^m = [s_0^0(f), s_1^{-1}(f), s_1^0(f), s_1^1(f), \dots, s_{N_s}^{N_s}(f)]^T, \quad (4-6)$$

$\mathbf{s}$  是一个长度为  $J$  的平面波幅度向量:

$$\mathbf{s} = [s_1(f), s_2(f), \dots, s_J(f)]^T, \quad (4-7)$$

$\mathbf{Y}$  是一个  $J \times (N_s + 1)^2$  的球谐域流形矩阵, 其中第  $j$  行为声源位于  $(\theta_j, \phi_j)$  方位处的球谐域流行向量, 表示为:

$$\mathbf{y}(\theta_j, \phi_j) = [Y_0^0(\theta_j, \phi_j), \dots, Y_n^m(\theta_j, \phi_j), Y_{N_s}^{N_s}(\theta_j, \phi_j)] \quad (4-8)$$

由式(4-8)可知, 球谐域 MUSIC 算法的阵列流形矩阵与频率无关, 且独立于阵列设置, 即与麦克风所在位置无关。

相关谱矩阵可以由下式得到:

$$\hat{\mathbf{S}} = \mathbf{s}_n^m [\mathbf{s}_n^m]^H \quad (4-9)$$

则 MUSIC 谱函数可以表示为:

$$P_{MU}(\theta, \phi) = \frac{1}{\mathbf{y}(\theta, \phi) \mathbf{E}_{N_s} \mathbf{E}_{N_s}^H \mathbf{y}^H(\theta, \phi)}, \quad (4-10)$$

其中, 矩阵  $\mathbf{E}_{N_s}$  是噪声子空间, 每一列对应一个噪声特征向量, 是矩阵  $\hat{\mathbf{S}}$  最小的  $(N_s + 1)^2 - J$  个特征值对应的特征向量, 这些特征向量可以通过对矩阵  $\hat{\mathbf{S}}$  进行特征值分解获

得。令  $(\theta, \phi)$  在空间中变化，搜索  $P_{MU}(\theta, \phi)$  的  $J$  个峰值，对应的角度即为平面波来波方向的估计值。

在本文中，如式(4-1)所示，假设每一个时频点只存在一个平面波，此时  $J = 1$ ，噪声子空间中特征向量的数目为  $(N_s + 1)^2 - 1$ 。

在此对球贝塞尔函数的零点问题（如图3-2所示）给定位算法带来的影响加以讨论。如上述定义  $s_n^m(f) = S_n^m(r, f)/b_n(kr)$  所示，使用球谐域 MUSIC 算法进行定位时需要对所获取的球谐分量  $S_n^m(r, f)$  除以  $b_n(kr)$ 。对于单个空心球来说，即需要除以  $j_n(kr)$ ，这将导致  $s_n^m(f)$  为一极大值，带来很大的定位误差。对于刚性球来说，则可以很好地避免这个问题。

## 4.2 声场扩阶原理

本节将对所提出的声场扩阶原理进行详细介绍，首先给出了有界线性算子及其在球谐域的表示形式，以此为基础对空域加窗算子在球谐域上的表示进行了详细推导，得到加窗后信号的球谐系数表达式，并且对空域加窗带来的声场扩阶现象加以分析。

### 4.2.1 有界线性算子

对于一个在  $L^2(\mathbb{S}^2)$  空间的有界线性算子  $\mathcal{B}$ <sup>[60]</sup>，将  $S(\theta, \phi)$  映射到  $G(\theta, \phi)$ （简便起见，麦克风位置  $r$  和频率  $f$  省略不写），记作：

$$G(\theta, \phi) = (\mathcal{B}S)(\theta, \phi) \quad (4-11)$$

也可以简记为  $G = \mathcal{B}S$ 。在大多数情况下， $\mathcal{B}$  用于表示一个通用的有界操作符，或者在特定有界运算符的表示中用作统称。

在可分离的希尔伯特空间，对于一个给定的完备正交序列  $\{\psi\}_{n=1}^\infty$ ，任意一个有界线性算子  $\mathcal{B}$  可以使用一个无限矩阵  $\mathbf{B}^{(\psi)}$  等价表示。矩阵在第  $n$  行  $m$  列的元素  $b_{n,m}^{(\psi)}$  表示沿着  $\psi_m$  方向的输入与经过该有界线性算子  $\mathcal{B}$  后沿着  $\psi_n$  方向的输出之间的关系。

由于  $L^2(\mathbb{S}^2)$  空间是可分离的，并且球谐函数是一个完备正交序列。因此，可以结合球谐分解得到有界线性算子在球谐域的表示形式。映射前后的信号可表示为：

$$S(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n S_n^m Y_n^m(\theta, \phi). \quad (4-12)$$

$$G(\theta, \phi) = \sum_{p=0}^{\infty} \sum_{q=-p}^p G_p^q Y_p^q(\theta, \phi). \quad (4-13)$$

我们几乎只用球谐函数作为完备正交序列，因此可以去掉对基函数上标  $\psi$  的依赖性，简单地把  $\mathbf{B}^{(\psi)}$  写成  $\mathbf{B}$  的形式。

此时, 算子  $\mathcal{B}$  可以用无限矩阵  $\mathbf{B}$  表示, 其矩阵元素  $b_{p,n}^{q,m}$  表示的是以  $Y_n^m$  为输入, 经过算子  $\mathcal{B}$  后输出信号在  $Y_p^q$  方向的分量, 表示为:

$$b_{p,n}^{q,m} = \langle \mathcal{B}Y_n^m, Y_p^q \rangle, \quad (4-14)$$

其中,  $\langle \cdot \rangle$  表示内积, 式 (4-14) 可进一步表示为:

$$b_{p,n}^{q,m} = \int_0^{2\pi} \int_0^\pi B(\theta, \phi) Y_n^m(\theta, \phi) [Y_p^q(\theta, \phi)]^* \sin \theta d\theta d\phi \quad (4-15)$$

在式 (4-14) 中, 一对  $m$  和  $n$  对应矩阵  $\mathbf{B}$  的一列, 即第  $n^2 + n + m + 1$  列, 一对  $p$  和  $q$  对应矩阵  $\mathbf{B}$  的一行, 即第  $p^2 + p + q + 1$  行, 算子矩阵  $\mathbf{B}$  如下所示:

$$\mathbf{B} = \begin{bmatrix} b_{0,0} & b_{0,1} & b_{0,2} & \dots \\ b_{1,0} & b_{1,1} & b_{1,2} & \dots \\ b_{2,0} & b_{2,1} & b_{2,2} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \left[ \begin{array}{c|ccccc} b_{0,0}^{0,0} & b_{0,1}^{0,-1} & b_{0,1}^{0,0} & b_{0,1}^{0,1} & b_{0,2}^{0,-2} & \dots \\ \hline b_{1,0}^{-1,0} & b_{1,1}^{-1,-1} & b_{1,1}^{-1,0} & b_{1,1}^{-1,1} & b_{1,2}^{-1,-2} & \dots \\ b_{1,0}^{0,0} & b_{1,1}^{0,-1} & b_{1,1}^{0,0} & b_{1,1}^{0,1} & b_{1,2}^{0,-2} & \dots \\ \hline b_{1,0}^{1,0} & b_{1,1}^{1,-1} & b_{1,1}^{1,0} & b_{1,1}^{1,1} & b_{1,2}^{1,-2} & \dots \\ \hline b_{2,0}^{-2,0} & b_{2,1}^{-2,-1} & b_{2,1}^{-2,0} & b_{2,1}^{-2,1} & b_{2,2}^{-2,-2} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \right]. \quad (4-16)$$

因此有界算子  $\mathcal{B}$  将  $S \in L^2(\mathbb{S}^2)$  的球谐系数  $S_n^m = \langle S, Y_n^m \rangle$  映射到  $G \in L^2(\mathbb{S}^2)$  的球谐系数  $G_p^q = \langle G, Y_p^q \rangle = \langle \mathcal{B}S, Y_p^q \rangle$ , 则  $G_p^q$  和  $G(\theta, \phi)$  可以表示为:

$$G_p^q = \sum_{n=0}^{\infty} \sum_{m=-n}^n S_n^m b_{p,n}^{q,m} \quad (4-17)$$

$$G(\theta, \phi) = \sum_{p=0}^{\infty} \sum_{q=-p}^p \sum_{n=0}^{\infty} \sum_{m=-n}^n S_n^m b_{p,n}^{q,m} Y_p^q(\theta, \phi) \quad (4-18)$$

对于复合算子  $\mathcal{P} = \mathcal{B} \cdot \mathcal{D}$ , 其算子矩阵可以简单地由各个算子的算子矩阵给出, 其算子矩阵元素可表示为:

$$p_{p,s}^{q,t} = \sum_{n=0}^{\infty} \sum_{m=-n}^m b_{p,n}^{q,m} d_{n,s}^{m,t}, \quad (4-19)$$

其中,  $p_{p,s}^{q,t}$ 、 $b_{p,n}^{q,m}$  和  $d_{n,s}^{m,t}$  分别为  $\mathcal{P}$ 、 $\mathcal{B}$  和  $\mathcal{D}$  的算子矩阵元素。

#### 4.2.2 空域加窗

在空间域使用窗函数对一个信号进行掩蔽, 即将感兴趣的信号  $S(\theta, \phi)$  与设计合理的窗函数  $h(\theta, \phi)$  在空间域上逐点相乘。将空域加窗对应的空间掩蔽算子记为  $\mathcal{B}_h$ , 则加窗后信号在空间域的表达式为:

$$(\mathcal{B}_h S)(\theta, \phi) = h(\theta, \phi) S(\theta, \phi), \quad (4-20)$$

其中,  $h(\theta, \phi)$  为空域窗函数。

空间掩蔽算子  $\mathcal{B}_h$  对应的算子矩阵元素为:

$$b_{p,n}^{q,m} = \int_0^{2\pi} \int_0^\pi h(\theta, \phi) Y_n^m(\theta, \phi) [Y_p^q(\theta, \phi)]^* \sin \theta d\theta d\phi. \quad (4-21)$$

与式 (4-12) 和 (4-13) 类似, 对窗函数  $h(\theta, \phi)$  进行球谐分解:

$$h(\theta, \phi) = \sum_{s=0}^{\infty} \sum_{t=-s}^s h_s^t Y_s^t(\theta, \phi) \quad (4-22)$$

其中,  $h_s^t$  为窗函数的球谐系数。

因此式 (4-17) 可以表示为:

$$\begin{aligned} G_p^q &= \sum_{n=0}^{\infty} \sum_{m=-n}^n S_n^m b_{p,n}^{q,m} \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n S_n^m \int_0^{2\pi} \int_0^\pi h(\theta, \phi) Y_n^m(\theta, \phi) [Y_p^q(\theta, \phi)]^* \sin \theta d\theta d\phi \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n S_n^m \int_0^{2\pi} \int_0^\pi \sum_{s=0}^{\infty} \sum_{t=-s}^s h_s^t Y_s^t(\theta, \phi) Y_n^m(\theta, \phi) [Y_p^q(\theta, \phi)]^* \sin \theta d\theta d\phi \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n S_n^m \sum_{s=0}^{\infty} \sum_{t=-s}^s h_s^t \int_0^{2\pi} \int_0^\pi Y_s^t(\theta, \phi) Y_n^m(\theta, \phi) [Y_p^q(\theta, \phi)]^* \sin \theta d\theta d\phi \end{aligned} \quad (4-23)$$

为上式的积分定义一个简记法:

$$y(s, t; n, m; p, q) \triangleq \int_0^{2\pi} \int_0^\pi Y_s^t(\theta, \phi) Y_n^m(\theta, \phi) [Y_p^q(\theta, \phi)]^* \sin \theta d\theta d\phi, \quad (4-24)$$

此时, 空间掩蔽算子  $\mathcal{B}_h$  对应的算子矩阵元素为:

$$b_{p,n}^{q,m} = \sum_{s=0}^{\infty} \sum_{t=-s}^s h_s^t y(s, t; n, m; p, q). \quad (4-25)$$

加窗后信号的球谐分解系数可以简洁表示为:

$$G_p^q = \sum_{n=0}^{\infty} \sum_{m=-n}^n S_n^m \sum_{s=0}^{\infty} \sum_{t=-s}^s h_s^t y(s, t; n, m; p, q). \quad (4-26)$$

在量子力学中经常出现包含三个球谐函数的积分, 可以用 Wigner  $3j$  符号表示<sup>[61]</sup>, 式 (4-24) 可表示为:

$$y(s, t; n, m; p, q) = (-1)^q \sqrt{\frac{(2s+1)(2n+1)(2p+1)}{4\pi}} \begin{pmatrix} s & n & p \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} s & n & p \\ t & m & -q \end{pmatrix}, \quad (4-27)$$

其中, 后两项不是矩阵, 而是 Wigner  $3j$  符号, 将在下文对该符号的定义和计算公式进行简单介绍。

Wigner 3j 符号是 Wigner 在文献 [61] 中定义的 3j 符号, 定义为:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \triangleq \frac{(-1)^{j_1-j_2-m_3}}{\sqrt{2j_3+1}} s_{j_3,m_1,m_2}^{j_1,j_2} \delta_{m_1+m_2+m_3,0}, \quad (4-28)$$

其中  $s_{j_3,m_1,m_2}^{j_1,j_2}$  的表达式为:

$$\begin{aligned} s_{j_3,m_1,m_2}^{j_1,j_2} = & \frac{\sqrt{(j_3+j_1-j_2)!(j_3-j_1+j_2)!(j_1+j_2-j_3)!(j_3-m_3)!(j_3+m_3)!}}{\sqrt{(j_1+j_2+j_3+1)!(j_1-m_1)!(j_1+m_1)!(j_2-m_2)!(j_2+m_2)!}} \\ & \times \sum_t \frac{(-1)^{t+j_2+m_2} \sqrt{2j_3+1} (j_3+j_2+m_1-t)!(j_1-m_1+t)!}{(j_3-j_1+j_2-t)!(j_3-m_3-t)!(t)!(j_1-j_2+m_3+t)!} \end{aligned}$$

在关于  $t$  的求和项中要求所有阶乘内的参数都是非负的。

必须同时满足以下条件, Wigner 3j 符号才不是零值:

$$\begin{aligned} m_1 + m_2 + m_3 &= 0 \\ |j_2 - j_3| &\leq j_1 \leq j_2 + j_3 \\ |j_1 - j_3| &\leq j_2 \leq j_1 + j_3 \\ |j_1 - j_2| &\leq j_3 \leq j_1 + j_2 \\ |m_k| &\leq j_k, k = 1, 2, 3 \end{aligned} \quad (4-29)$$

定义和计算 Wigner 3j 符号的方法还有很多, 其中最广泛使用的是 Racah 公式 [62]:

$$\begin{aligned} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = & (-1)^{j_1-j_2-m_3} \sqrt{\Delta(j_1, j_2, j_3)} \\ & \times \sqrt{(j_1-m_1)!(j_1+m_1)!(j_2-m_2)!(j_2+m_2)!(j_3-m_3)!(j_3+m_3)!} \\ & \times \sum_t \frac{(-1)^t}{(t)!(j_1+j_2-j_3-t)!(j_1-m_1-t)!(j_2+m_2-t)!(j_3-j_2+m_1+t)!(j_3-j_1-m_2+t)!} \end{aligned} \quad (4-30)$$

其中  $\Delta(j_1, j_2, j_3)$  的表达式为:

$$\Delta(j_1, j_2, j_3) = \frac{(j_1+j_2-j_3)!(j_1-j_2+j_3)!(j_2-j_1+j_3)!}{(j_1+j_2+j_3+1)!}$$

#### 4.2.3 空域加窗后的声场阶次

在实际情况中, 需要对式 (4-12)(4-13)(4-22) 进行截断处理, 假设加窗前信号、加窗后信号和窗函数球谐分解的最高阶次分别为  $N_s$ 、 $N_g$  和  $N_{h'}$ , 由式 (4-31) 可知  $N_g$  是由  $N_s$  和  $N_{h'}$  共同决定的。

$$G_p^q = \sum_{n=0}^{N_s} \sum_{m=-n}^n S_n^m \sum_{s=0}^{N_{h'}} \sum_{t=-s}^s h_s^t y(s, t; n, m; p, q), \quad 0 \leq p \leq N_g. \quad (4-31)$$

对于轴对称问题，即窗函数沿水平角  $\phi$  为常数的特殊情况， $N_s$ 、 $N_g$  和  $N_{h'}$  有显示关系式。此时  $h(\theta, \phi) = h(\theta)$ ，窗函数的球谐系数为：

$$\begin{aligned} h_s^t &= \int_0^{2\pi} \int_0^\pi h(\theta) [Y_s^t(\theta, \phi)]^* \sin \theta d\theta d\phi \\ &= 2\pi \delta_{t,0} \sqrt{\frac{2s+1}{4\pi}} \int_0^\pi h(\theta) P_s^t(\cos \theta) \sin \theta d\theta \\ &= \sqrt{\frac{4\pi}{2s+1}} h_s \delta_{t,0}, \end{aligned} \quad (4-32)$$

其中， $h_s$  只与  $s$  有关，这个性质是通过求解  $\phi$  的积分得到的，将原始的二维积分简化为一维，关联勒让德函数退化为勒让德函数。在这种情况下， $h_s$  可表示为：

$$\begin{aligned} h_s &= \frac{2s+1}{2} \int_0^\pi h(\theta) P_s(\cos \theta) \sin \theta d\theta \\ h(\theta) &= \sum_{s=0}^{\infty} h_s P_s(\cos \theta), \end{aligned} \quad (4-33)$$

其中， $P_s(\cdot)$  为  $s$  阶勒让德函数。

结合式 (4-32) 和式 (4-29) 可知，只有当  $m = q$  且  $|p - s| \leq n \leq |p + s|$  时， $y(s, t; n, m; p, q)$  才为非零值。此时，式 (4-31) 可表示为：

$$\begin{aligned} G_p^q &= (-1)^q \sqrt{\frac{(2p+1)}{4\pi}} \sum_{s=0}^{N_{h'}} \sum_{n=|p-s|}^{p+s} h_s^0 S_n^q \sqrt{(2s+1)(2n+1)} \\ &\quad \times \begin{pmatrix} s & n & p \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} s & n & p \\ 0 & q & -q \end{pmatrix} \end{aligned} \quad (4-34)$$

由式 (4-34) 可知，对于  $p \leq N_{h'}$ ，加窗后信号的球谐系数  $G_p^q$  是原始信号球谐系数  $S_n^q$  在阶次  $0 \leq n \leq p + N_{h'}$  上的加权和，对于  $p \geq N_{h'}$ ，加窗后信号分量  $G_p^q$  是原始信号分量  $S_n^q$  在阶次  $|p - N_{h'}| \leq n \leq p + N_{h'}$  上的加权和。因此，加窗后信号的球谐系数是原始信号的频谱平滑，平滑的程度取决于窗的带宽和频谱响应的形状。并且，加窗后信号的球谐分解阶次最高可达到  $N_g = N_s + N_{h'}$ 。

### 4.3 空域窗函数

如上节所示，通过对原始信号进行空域加窗可以实现声场扩阶，本节将对采用的窗函数加以介绍。

本文中采用 Von Mises Fisher 函数<sup>[63][64][65]</sup>作为三维空间的窗函数，其可以认为是多元高斯函数在球坐标系下的近似。在  $p-1$  维球体  $\mathbb{S}^{p-1}$  上，Von Mises Fisher 函数的

表达式为：

$$h(\Omega; \mu, \kappa) = C_p(\kappa) \exp^{\kappa \mu^T \Omega} \sin \theta \quad (4-35)$$

其中，主瓣方向（对称轴方向）向量  $\mu$  是一个单位向量，即  $\|\mu\| = 1$ ， $(\cdot)^T$  表示转置， $C_p$  是归一化系数：

$$C_p(\kappa) = \frac{\kappa^d}{2\pi^{d+1} I_d(\kappa)},$$

其中， $I_d(\kappa)$  是  $d$  阶第一类修正贝塞尔函数， $d = p/2 - 1$ ， $\kappa$  是集中因子，取值范围为  $\kappa \geq 0$ 。

集中因子  $\kappa$  是衡量 Von Mises Fisher 函数在主瓣方向附近聚集程度（或是方向分散度）的参数， $\kappa$  值越大，函数越集中于主瓣方向。当  $\kappa \rightarrow \infty$  时，Von Mises Fisher 函数成为一个  $\delta$  函数， $\kappa = 0$  时，函数各向同性。因此可以认为集中因子  $\kappa$  控制着 Von Mises Fisher 函数的宽度。

在笛卡尔坐标系下， $\mu = [\sin \theta_o \cos \phi_o, \sin \theta_o \sin \phi_o, \cos \theta_o]^T$ ，其中  $\theta_o$  和  $\phi_o$  分别表示主瓣方向的俯仰角和水平角， $\Omega = [\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta]^T$ 。此时  $\mu^T$  和  $\Omega$  的内积可以表示为  $\mu^T \Omega = \sin \theta_o \sin \theta \cos(\phi - \phi_o) + \cos \theta_o \cos \theta$ 。

将  $\mu^T \Omega$  带入式 (4-35)，并且假设  $p = 3$ ，此时为二维球体  $\mathbb{S}^2$ ，即普通球面，也就是欧几里得空间  $\mathbb{R}^3$ ，Von Mises Fisher 函数的表达式为：

$$h(\theta, \phi | \theta_o, \phi_o, \kappa) = C_p(\kappa) \exp^{\kappa[\sin \theta_o \sin \theta \cos(\phi - \phi_o) + \cos \theta_o \cos \theta]} \sin \theta. \quad (4-36)$$

式 (4-36) 构成了 Von Mises Fisher 函数的一般形式。假设  $(\theta_0, \phi_0) = (0^\circ, 0^\circ)$ ，此时  $\mu = [0, 0, 1]^T$ ，窗函数的空间分布随着  $\kappa$  的变化如图 4-1 所示，其中红色的点表示在某个固定的集中因子  $\kappa$  下窗函数幅值较大的位置。

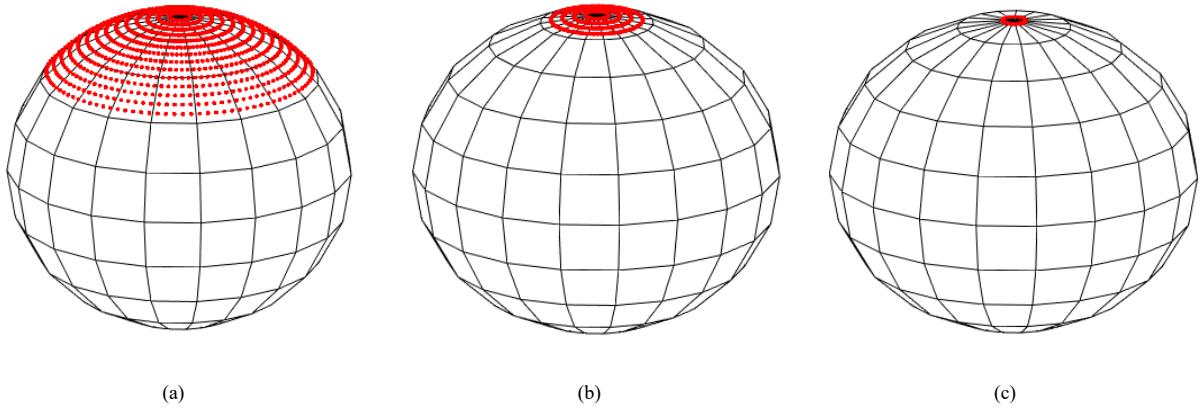


图 4-1 二维球体  $\mathbb{S}^2$  窗函数：(a)  $\kappa = 10$ ，(b)  $\kappa = 100$ ，(c)  $\kappa = 600$

对于  $p = 2$  的一维球体  $\mathbb{S}^1$ ，Von Mises Fisher 函数的表达式为：

$$h(\phi) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\phi - \phi_0)}, \quad |\phi - \phi_0| \leq \pi \quad (4-37)$$

假设主瓣方向  $\phi_0 = 0^\circ$ , 窗函数在半个空间内的分布随集中因子  $\kappa$  的变化如图 4-2 所示。综合图 4-1 和 4-2 可以看出, 集中因子  $\kappa$  越大, 窗函数的宽度越窄, 函数越集中于主瓣方向, 且主瓣方向的幅值越大。

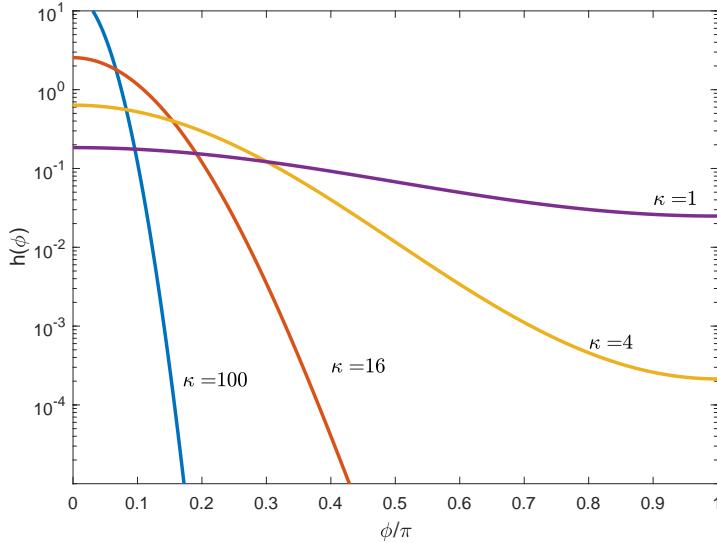


图 4-2 一维球体  $S^1$  窗函数

本文中, 在对原始声场进行空域加窗时, 关于窗函数的选取原则如下:

- (1) 窗函数的主瓣方向: 选取该时频点上平面波的来波方向, 即球谐域 MUSIC 算法的结果;
- (2) 集中因子  $\kappa$ : 需要根据频率进行选择。设定目标半径  $R_T$ , 对于某一频率  $f$ , 根据  $N_T = kR_T = 2\pi f R_T$  计算目标阶次, 选择合适的集中因子  $\kappa$  使窗函数的球谐系数在前  $N_T$  阶的累积能量达到总能量的 90% (或 99% 等其他接近于 1 的数值)。

接下来对窗函数的球谐系数获取方法进行简单介绍。在上述选取原则的指导下, 已经可以得到空域窗函数的具体表达式, 因此可以在空间上任意位置取采样。相比于麦克风阵列的球谐系数获取, 窗函数球谐系数的获取不受空间采样点数的限制, 因此可以直接采用式 (3-24) 的方式进行求解。

#### 4.4 整体算法框架

基于球谐分解的双耳渲染算法直接在球谐域进行处理, 如前所述, 实际系统中录制声场的阶次受硬件设备的限制, 因此和 HRTF 数据之间存在不匹配问题, HRTF 的直接低阶截断会对双耳信号带来损伤。本文所提出的算法分别从声场和 HRTF 两方面出发, 共同解决二者之间的不匹配问题。接下来将对本文的内容进行总结, 包括第 2.3 节、第三章和第四章。

首先, 第 2.3 节中对基于球谐分解的双耳渲染算法框架进行了详细介绍, 给出了如何利用声场和 HRTF 的球谐系数获取双耳信号的方法。第三章首先介绍了声场和 HRTF 球谐系数的获取方法, 接下来引入了相位对准的 HRTF 预处理方法, 并通过实验验证了该方法可以有效降低 HRTF 的球谐分解阶次。第四章创新性提出通过空域加窗的方法来提升录制声场的阶次。首先对信号模型及球谐域 MUSIC 定位算法进行了介绍, 接下来对空域加窗可以实现声场阶次提升的原理进行推导, 并对所采用的窗函数加以介绍。

以上三部分是本算法的理论基础, 接下来给出引入声场扩阶和 HRTF 预处理的基于球谐分解的双耳渲染算法流程, 如图 4-3 所示。其中, STFT 和 ISTFT 分别为短时傅里叶变换及其逆变换, SFT 为球傅里叶变换。

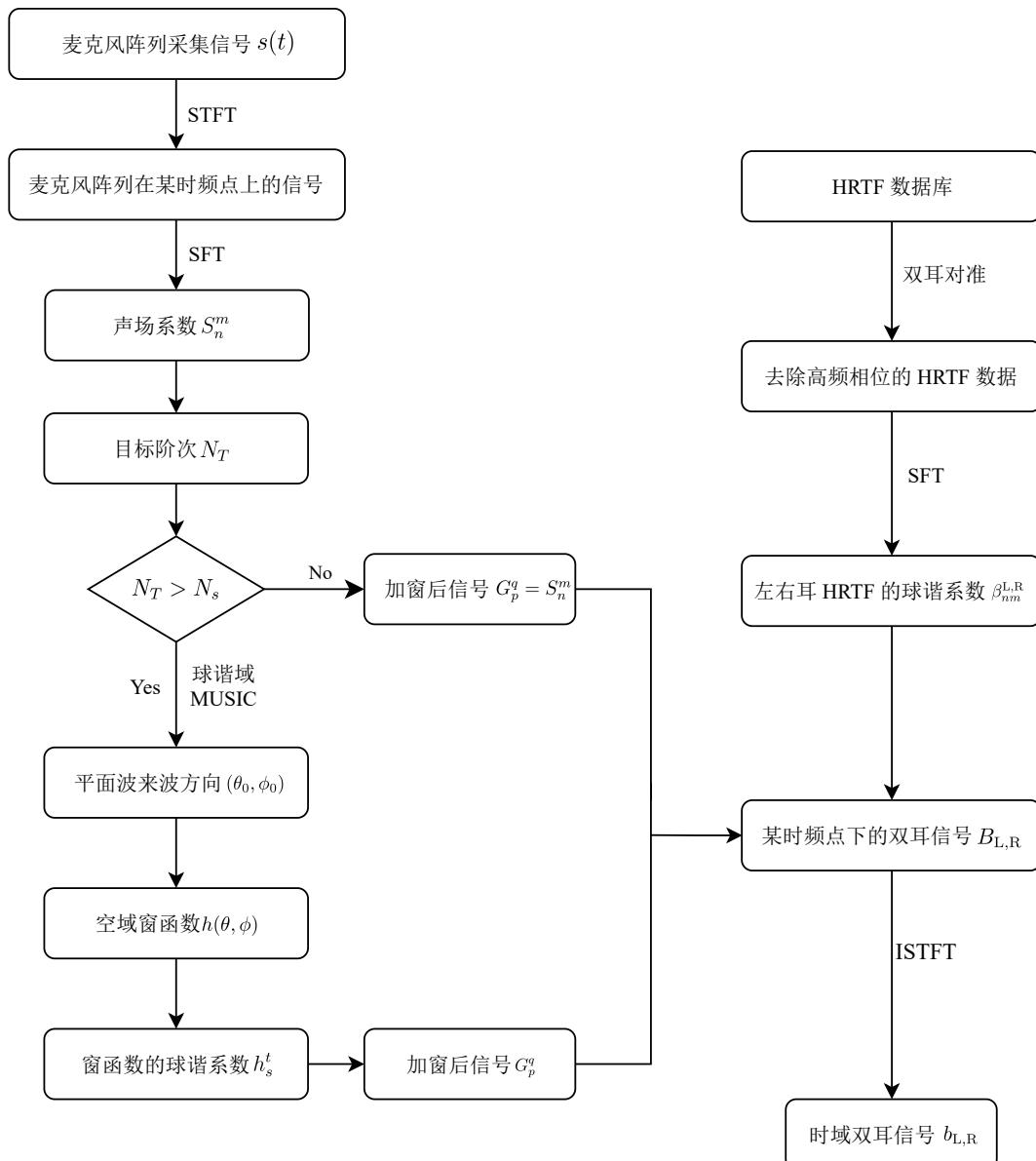


图 4-3 引入声场扩阶和 HRTF 预处理的基于球谐分解的双耳渲染算法流程图

## 4.5 本章小结

本章对同时引入声场扩阶和 HRTF 预处理的基于球谐分解的双耳渲染算法进行了详细介绍。首先对信号模型进行描述，并给出了球谐域 MUSIC 定位算法以获取单时频点上的平面波来波方向。其次从有界线性算子出发，在球谐域上对加窗后信号的球谐系数表达式进行详细推导，并通过分析发现空域加窗可以实现声场阶次的提升。接下来对采用的窗函数及其选取原则和球谐系数求解方法加以介绍。最后将前三节的声场扩阶和第三章的 HRTF 预处理相结合，给出了总体算法框架。



## 第五章 实验结果及分析

随着越来越多的技术进入消费者市场，有必要评估沉浸式音频体验的质量，以量化和优化技术，提高听众体验的真实性。质量评价的目的是测量听者对虚拟听觉环境的感知。在评价双耳沉浸式环境的质量时，我们需要考虑各种物理特征，包括声源定位精度、感知声源宽度（Apparent Source Width, ASW）、音质质量和损伤以及室内声学和环境相关属性<sup>[65]</sup>。

本章首先对评价指标及其计算方法进行详细介绍，主要包括声源定位精度和感知声源宽度两方面。然后从实验的角度出发，对本文所提出的双耳渲染算法的性能进行评估。

### 5.1 评价指标

声源定位是听众通过确定方位角、仰角和距离来定位声源的能力。在声源定位精度方面，主要采用的指标为双耳时间差 ITD 和双耳声级差 ILD。瑞利在 1907 年将双耳时间差 ITD 和双耳声级差 ILD 引入作为声源定位因素，称为瑞利双因素理论。该理论指出人耳听觉系统可以利用低频的 ITD 信息和高频的 ILD 信息决定声源方向。

感知声源宽度是对声源或声场在感知听觉宽度这一属性上的度量。这一主观属性与一种客观测量相关，最明显的是双耳听觉互相关系数。因此在空间感的感知方面，主要采用的指标为双耳听觉互相关系数 IACC。大量的主客观实验数据显示，双耳听觉互相关系数与声重放的空间主观感知有很强的相关性，可以用来衡量感知声源宽度。双耳听觉互相关系数定义为双耳归一化互相关函数的最大值，与空间感主观评测呈现负相关关系，即 IACC 值越小，双耳信号相关性越差，对应的声场越趋向于扩散场，因此感知声源更宽。

接下来将对三个指标的计算方法进行详细介绍。

#### 5.1.1 双耳时间差

双耳时间差（Interaural Time Difference, ITD）是对声源方向定位的一个重要因素，其主要作用于低频段范围（小于 1.5 kHz），其表示的是声波从声源到双耳传输的时间差。当声源位于中垂面（原点指向正前方和正上方的两个矢量构成的平面）时，声波从声源到达双耳的距离相等，传输所需的时间也相等，此时 ITD 为 0。当声源偏离中垂面时，声波到达左、右耳的距离不同，对应的时间不同，此时双耳时间差不再为 0。

以水平面为例，如果不考虑头部对时间差的影响，可以将双耳近似为空间中相距  $2a$  的两点，如图 5-1 中左图所示。则对于来波方向为  $\phi$  的远场平面波，其到达双耳的距离差为  $2a \sin \phi$ ，双耳时间差为：

$$\text{ITD}(\phi) = \frac{2a}{c} \sin \phi \quad (5-1)$$

式中： $c$  为声速。 $\text{ITD} > 0$  表示右耳比左耳先到达； $\text{ITD} < 0$  正好相反。

如果考虑头部对时间差的影响，可以用刚球模型来近似人头，球面上相对的两点分别表示左右耳，如图 5-1 中右图所示。该模型考虑了声波在头部弯曲表面  $a\phi$  的传输时间，此时式 (5-1) 可写为：

$$\text{ITD}(\phi) = \frac{a}{c} (\sin \phi + \phi), \quad 0 \leq \phi \leq \frac{\pi}{2} \quad (5-2)$$

上式给出了部分入射角的 ITD 计算公式，其他入射角的公式可以通过对称性得到。式 (5-2) 被称为 Woodworth 公式。

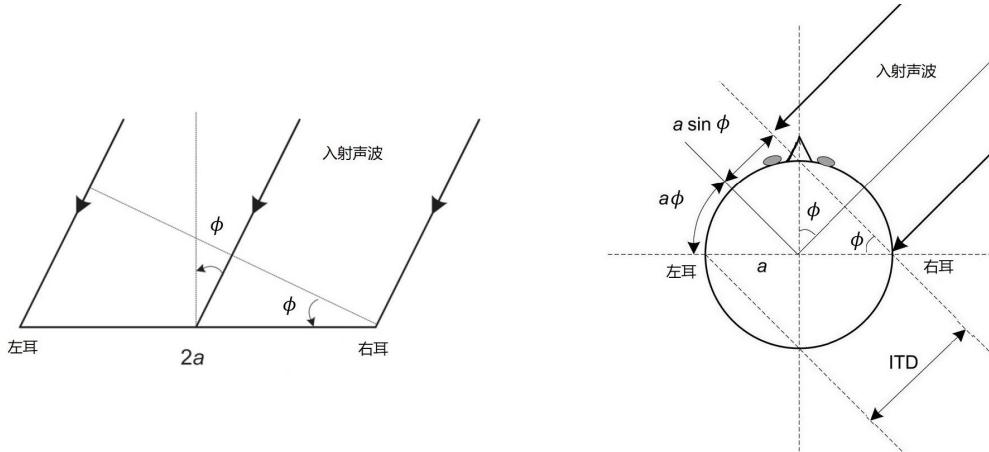


图 5-1 ITD 计算示意图<sup>[18]</sup>

当信号从正前方附近入射时， $\sin \phi \approx \phi$ ，式 (5-1) 和 (5-2) 结果近似相等。但对于其他角度，式 (5-1) 和 (5-2) 结果有一定差别。并且式 (5-1) 和 (5-2) 只适用于求解已知声源方位角时的理论 ITD，对于一个未知声源方位的双耳信号，需要使用下述的三种方法进行求解。

### (1) 时域互相关法：

假设声源在左右耳产生的声压为  $b_L(t)$  和  $b_R(t)$ ，两路信号的互相关函数为：

$$\psi_{LR}(\tau) = \frac{\int_{-\infty}^{\infty} b_L(t + \tau) b_R(t) dt}{\sqrt{\int_{-\infty}^{\infty} b_L^2(t) dt \int_{-\infty}^{\infty} b_R^2(t) dt}} \quad (5-3)$$

如果两路信号的时域表示为  $b_L(t) = g_L s(t + t_L) + n_L(t)$ ， $b_R(t) = g_R s(t + t_R) + n_R(t)$ ， $g_L$  与  $g_R$  分别反映声源信号  $s(t)$  到达左右耳处的幅度衰减函数， $t_L$  和  $t_R$  分别表示声源到

达左右耳处所需的时间。由于  $n_L(t)$  和  $n_R(t)$  为加性随机噪声，与声源信号不相关，且互相独立，因此：

$$\begin{aligned}\psi_{LR}(\tau) &= g_L g_R r_{ss}(\tau - t_L + t_R) + g_L r_{sn_R}(\tau - t_L) + g_R r_{sn_L}(\tau + t_R) + r_{n_L n_R} \\ &= g_L g_R r_{ss}(\tau - t_L + t_R),\end{aligned}\quad (5-4)$$

其中， $r_{ss}$  和  $r_{n_L n_R}$  分别表示信号的自相关函数和噪声的互相关函数， $r_{sn_L}$  和  $r_{sn_R}$  表示信号与噪声的互相关函数。

由互相关函数的性质可知，当  $\tau = t_L - t_R$  时，互相关函数  $\psi_{LR}(\tau)$  取最大值。所以双耳信号  $b_L(t)$  和  $b_R(t)$  之间的时延估计值为：

$$\hat{\tau}_{LR} = \arg \max_{\tau} \psi_{LR}(\tau), \quad (5-5)$$

式中， $\tau \in [-\tau_{\max}, \tau_{\max}]$ ， $\tau_{\max}$  在双耳信号计算中一般取 1 ms。在麦克风阵列中， $\tau_{\max}$  与阵元间距、声速和采样频率有关。

## (2) 频域相干法<sup>[66]</sup>:

对左右耳信号  $b_L(t)$  和  $b_R(t)$  进行短时傅里叶变换(STFT)转换到频域，表示为  $B_L(\mathcal{T}, f)$  和  $B_R(\mathcal{T}, f)$ ，其中  $\mathcal{T}$  为时间帧， $f$  为频率。

两路信号之间的频域相干函数定义为：

$$\Gamma_{B_L B_R}(\mathcal{T}, f) = \frac{P_{B_L B_R}(\mathcal{T}, f)}{\sqrt{P_{B_L B_L}(\mathcal{T}, f) P_{B_R B_R}(\mathcal{T}, f)}}, \quad (5-6)$$

其中， $P_{B_L B_L}(\mathcal{T}, f)$  和  $P_{B_R B_R}(\mathcal{T}, f)$  分别为  $B_L(\mathcal{T}, f)$  和  $B_R(\mathcal{T}, f)$  的自功率谱， $P_{B_L B_R}(\mathcal{T}, f)$  为  $B_L(\mathcal{T}, f)$  和  $B_R(\mathcal{T}, f)$  的互功率谱，计算公式为：

$$\begin{aligned}P_{B_i B_i}(\mathcal{T}, f) &= \alpha P_{B_i B_i}(\mathcal{T} - 1, f) + (1 - \alpha) |B_i(\mathcal{T}, f)|^2, \quad i = L, R \\ P_{B_L B_R}(\mathcal{T}, f) &= \alpha P_{B_L B_R}(\mathcal{T} - 1, f) + (1 - \alpha) B_L(\mathcal{T}, f) B_R^*(\mathcal{T}, f)\end{aligned}\quad (5-7)$$

其中  $\alpha$  为相邻帧之间的平滑因子。

理想情况下单个声源在两个麦克风处的相干函数为：

$$\Gamma(f) = e^{j2\pi f \tau} \quad (5-8)$$

当频率固定时，理想相干函数仅与时间差  $\tau$  有关。

此时，可以使用双耳信号的相干函数与多个  $\tau$  构成的理想相干函数库进行匹配，二者相关系数最高时所对应的  $\tau$  即为双耳信号的 ITD。也可以计算双耳信号相干函数的相位  $\psi$ ，此时某个频点下的 ITD 为：

$$\text{ITD} = \frac{\psi}{2\pi f} \quad (5-9)$$

### (3) 改进的频域相干法:

在频域相干法中，使用关心频段内的所有频率进行 ITD 的计算，本方法对其改善，加入频率有效性判断。在每个频率计算一个频域相干函数，利用频率相干函数的幅度值大小判断该频率的有效性。当某个频率的相干函数幅度高于频段内幅度均值时，则认为该频率有效，最后只采用相关性大的频率进行 ITD 的求解。此种方法可以选取更为可靠的频率进行计算，降低了干扰频率带来的影响，从而进一步提高 ITD 计算结果的准确性和可靠性。

上文对三种计算 ITD 的方法进行了详细介绍，需要注意的是 ITD 是在低频段对声源定位的主要因素，所以在计算时需要对信号加以处理。对于时域计算方法，需要对双耳信号进行截止频率为 1.5 kHz 的低通滤波；对于频域计算方法，只需要对 1.5 kHz 以下的频率进行计算。

## 5.1.2 双耳声级差

双耳声级差 (Interaural Level Difference, ILD) 是声源方向定位的另一个重要因素，其主要作用于中高频，这是由于头部对声波有一定的反射和散射作用，高频时其作用更加明显。当声源偏离中垂面时，与声源同侧耳处的声压有所提升，异侧耳的声压有所衰减。因此，双耳声级差是声源方向和频率的函数。在单声源情况下，ILD 在低频时很小，且基本不随声源方向的改变而改变。随着频率的增加，ILD 逐渐增加，表现出与声源方向及频率的复杂关系<sup>[18]</sup>。心理声学表明，大约在 1.5 kHz 以上，ILD 才开始作为一个有效的方向定位因素。

对于一个双耳信号，ILD 在某个频率  $f$  处的计算方式为：

$$\text{ILD}(f) = 20 \lg \left| \frac{B_R(f)}{B_L(f)} \right| \quad (5-10)$$

式中， $B_R(f)$  和  $B_L(f)$  分别是左耳信号和右耳信号的频域声压。 $\text{ILD} > 0$  表示右耳比左耳先到达， $\text{ILD} < 0$  相反。

有时也需要计算一定频率范围  $f_L \leq f \leq f_H$  内的平均双耳声级差，通过分别计算左右耳信号在某个频段内的总能量来获取 ILD：

$$\text{ILD} = 10 \lg \left( \frac{\int_{f_L}^{f_H} |B_L(f)|^2 df}{\int_{f_L}^{f_H} |B_R(f)|^2 df} \right) \quad (5-11)$$

在上式中选择不同的频率积分范围，即可得到该频带内的平均 ILD，例如，1/3 倍频程带。

在实际计算过程中，需要对式 (5-11) 离散化求解：

$$\text{ILD} = 10 \lg \left( \frac{\sum_{f_L}^{f_H} |B_L(f)|^2}{\sum_{f_L}^{f_H} |B_R(f)|^2} \right) \quad (5-12)$$

与 ITD 的计算相似，ILD 是声源在中高频段定位的主要因素，因此在计算过程中应该选取 1.5 kHz 以上的频率。由于语音的常用采样率为 16 kHz，因此本文最终选取 1.5 ~ 8 kHz 频段进行平均 ILD 的计算。

### 5.1.3 双耳相关性

在室内声学中，空间感（Spaciousness）是总体空间印象的一个重要部分，是衡量厅堂音质的重要主观评价指标。大量的研究表明，厅堂的早期反射对于空间感至关重要，特别是 1.5 kHz 以下的低频成分。Amdo 的研究表明厅堂的空间感与分立反射声的空间方向分布有关，并提出以双耳听觉互相关系数<sup>[18]</sup>（Interaural Cross-Correlation, IACC）作为衡量空间感的一个客观指标。

IACC 是双耳信号空间感评测的重要指标，同时可以用于预测感知声源宽度。其测量两个耳朵信号的相似度，通过计算双耳信号的归一化互相关函数的最大值而来，互相关分析用于表明在给定的一段时间内两个时域波形的相似程度。归一化互相关函数定义如式(5-3)所示，通常最大化互相关是在  $-1 \text{ ms} \leq \tau \leq 1 \text{ ms}$  范围内计算  $\psi_{LR}(\tau)$  的最大值：

$$\text{IACC} = \max_{-1 \text{ ms} < \tau < 1 \text{ ms}} \psi_{LR}(\tau) \quad (5-13)$$

由定义可知，所有的  $0 \leq \text{IACC} \leq 1$ 。IACC 为 1 表示两个信号完全相关，当信号具有较低的相似度或存在随机相关时，IACC 趋近于 0。

有时还会引入半波整流和低通滤波以模拟在高频时听觉系统中发生的包络提取，这种改进的早期 IACC 称为  $\text{IACC}_E$ ，更适合预测 ASW。与 ITD 类似，本文在计算 IACC 前对信号进行截止频率为 1.5 kHz 的低通滤波。

最小可觉差（Just Noticeable Difference, JND）也被称为差别阈值线，体现人们对改变或者误差的容忍度，是测量两种事物差异的数量单位。如果误差  $\Delta$  和 JND 之比小于 1，即  $\Delta$  小于 JND，可以近似认为二者在感知上无差别。

对于 IACC 来说，其最小可觉差<sup>[36][67]</sup>可以通过下式计算：

$$\text{JND}(x) = \max (0.557 - 0.379x - 0.178x^2, 0.007) \quad (5-14)$$

其中， $x$  表示参考双耳信号的 IACC。

为了避免在相除过程中 JND 作为分母取值为 0，对最小值加以约束，即式(5-14)中的限制  $\text{JND}(0.99) = 0.007$ 。

## 5.2 实验结果及分析

在本节中，将使用第 5.1 节所介绍的评价指标对本文所提出的方法进行评估。接下来对参与评估的算法和信号进行简单介绍。

首先,选取了一个参考信号,其声场和 HRTF 的球谐分解均为 34 阶,并且麦克风阵列半径为 0.1 米(接近人头尺寸),参见式(2-29)和式(2-30)获取双耳信号,记作 Original。参与比较的算法包括以下三个:

(1) 直接基于低阶球谐分解的双耳渲染算法。声场的球谐分解为  $N_s$  阶, HRTF 直接进行低阶截断,且球谐分解的阶次与声场保持一致,记作 LS;

(2) 引入 HRTF 预处理的双耳渲染算法。声场的球谐分解为  $N_s$  阶,对 HRTF 进行相位对准预处理,使其能量更加集中在低频,HRTF 的球谐分解阶次仍为  $N_s$  阶,记作 HRTF Pre;

(3) 同时引入声场扩阶和 HRTF 预处理的双耳渲染算法。将声场的球谐分解阶次从  $N_s$  提升为  $N_T$  阶,并对 HRTF 进行相位对准预处理,且球谐分解阶次为  $N_T$  阶,记作 EO+HRTF Pre,其中 EO 表示扩阶(Enlarge Order)。

### 5.2.1 消声室环境下各算法性能对比

仿真环境为消声室环境,麦克风阵列采用 Eigenmike 球阵,其包含 32 个麦克风,半径为 0.042 米,对应的球谐函数分解阶次  $N_s = 4$  阶。麦克风阵列和声源位于同一水平面,即俯仰角为  $\theta = 90^\circ$ ,且相距 2 米。声源以 Eigenmike 为圆心均匀运动,即声源相对麦克风阵列的方位角  $\phi$  均匀变化。方位角  $\phi$  的取值范围从  $0^\circ$ (即正前方)沿顺时针方向到  $90^\circ$ (即正右方),角度间隔为  $10^\circ$ ,其它方位的结论可以较为容易地通过对称来获取。声源信号包括两类,白噪声信号和语音信号。

消声室情况下,主要采用双耳时间差 ITD 和双耳声级差 ILD 这两个评价指标,并且通过计算和对比三种待评估算法(LS、HRTF Pre 和 EO+HRTF Pre)对应的双耳信号与参考双耳信号 Original 在 ITD 和 ILD 上的误差  $\Delta_{ITD}$  和  $\Delta_{ILD}$  来进行评估,如式(5-15)所示。误差越小,则认为该算法越优。

$$\begin{aligned}\Delta_{ITD} &= | ITD_{case} - ITD_{Original} | \\ \Delta_{ILD} &= | ILD_{case} - ILD_{Original} | \end{aligned}\quad (5-15)$$

其中, case 表示 LS、HRTF Pre 和 EO+HRTF Pre。

不同声源类型(白噪声、语音)、不同方位角对双耳信号的影响如图 5-2 所示,图(a)和图(b)为声源为白噪声信号,位于不同方位角时三种算法对应的双耳信号的  $\Delta_{ITD}$  和  $\Delta_{ILD}$  结果,图(c)和图(d)为声源为语音信号,位于不同方位角时三种算法对应的双耳信号的  $\Delta_{ITD}$  和  $\Delta_{ILD}$  结果。

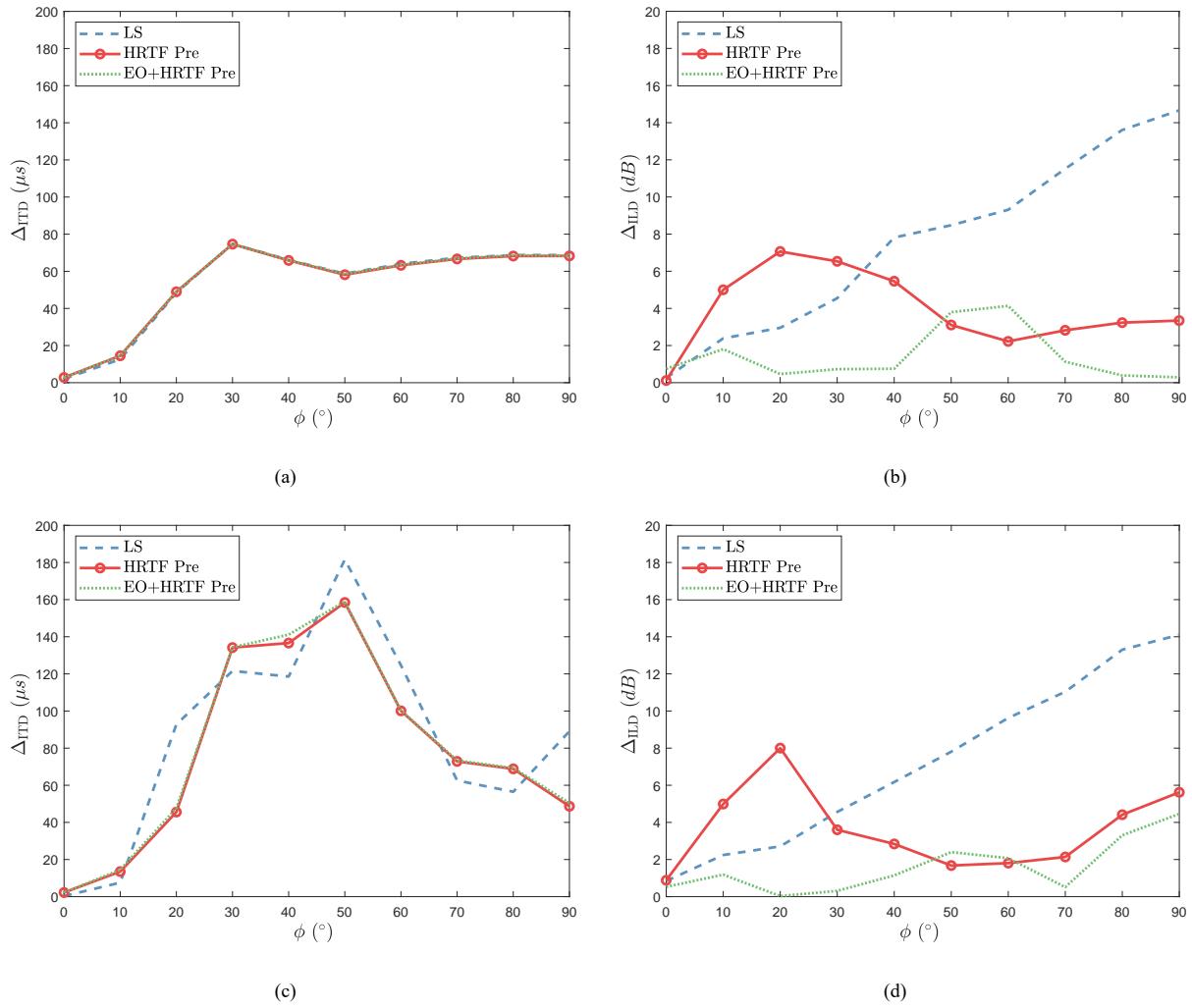


图 5-2 消声室环境下两种声源信号在不同方位角处的实验结果:白噪声:(a) ITD;(b) ILD;语音: (c) ITD; (d) ILD

从图 5-2 可以看出:

(1) 无论是白噪声信号还是语音信号,LS、HRTF Pre 和 EO + HRTF Pre 算法的  $\Delta_{ITD}$  结果基本一致,这是因为 HRTF Pre 算法进行相位对准的起始频率为 1.5 kHz, EO + HRTF Pre 算法在引入声场扩阶的过程中,当  $N_T > N_s$  时才开始进行处理,根据  $N_T = kR_T = 2\pi f R_T / c$ ,  $R_T = 0.1$  米和原有声场阶次  $N_s = 4$  阶可知,进行声场扩阶的起始频率为 2.7 kHz。而评价指标 ITD 的关注频率范围为 1.5 kHz 以下,此频段内 HRTF Pre 和 EO + HRTF Pre 算法并未对双耳信号进行任何处理,因此三种算法的  $\Delta_{ITD}$  结果没有区别。

(2) 无论是白噪声信号还是语音信号, EO + HRTF Pre 算法的  $\Delta_{ILD}$  结果最小,并且在任意方位角下均可达到较小的误差。LS 算法与 Original 的差距最大,且随着方位角的变化,误差一直增大。

(3) 相较于 LS 算法, HRTF Pre 算法在大多数方位角下的  $\Delta_{ILD}$  有所降低,更加接

近 Originial 算法，但是在部分角度 ( $10^\circ$ 、 $20^\circ$ ) HRTF Pre 算法的误差大于 LS 算法，进一步分析发现 HRTF Pre 算法在这些角度出现 ILD 的正负错误，可能会带来少量的声源定位错误，而改进方法 EO + HRTF Pre 并未出现此问题。

(4) 相比于白噪声信号，语音信号的 ITD 和 ILD 结果有少许变化，这是因为语音信号的能量主要集中在低频，并且在各频点上的能量分布不均，而白噪声在整个频谱上均匀分布，避免了计算波动。但是两种声源类型对应的综合结论一致。

上图展示了 1.5 kHz 到 8 kHz 频段内平均 ILD 的结果，接下来以方位角  $\phi = 40^\circ$  为例，对三种算法所获取的双耳信号与参考信号在 8 kHz 以下频段内单个频点处的  $\Delta_{ILD}$  结果进行展示，如图 5-3 所示。

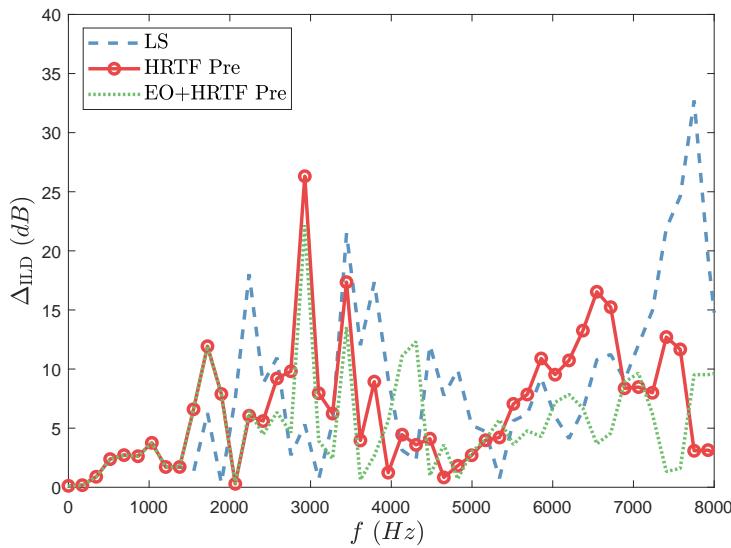


图 5-3 消声室情况下方位角  $\phi = 40^\circ$  时单个频点处的  $\Delta_{ILD}$  结果

从图 5-3 可以看出：

(1) 综合整个频段来看，EO + HRTF Pre 算法的  $\Delta_{ILD}$  结果最小，尤其是在 5 kHz 以上的高频，现有的研究表明，ILD 是高频定位因素，1.5 kHz 以上 ILD 才开始对定位起作用，频率大于 4 kHz 时，ILD 才是方向定位的主要因素。HRTF Pre 算法次之，LS 算法的  $\Delta_{ILD}$  最大，与图 5-2 中 1.5 kHz 到 8 kHz 频段内平均  $\Delta_{ILD}$  的结论一致。

(2) 相比于 LS 算法，HRTF Pre 和 EO + HRTF Pre 算法的 ILD 开始在 1.5 kHz 有所调整。相比于 HRTF Pre 算法，EO + HRTF Pre 算法的 ILD 开始在 2.7 kHz 有所调整，与图 5-2 的讨论中频率相关的结论一致。

### 5.2.2 空心球与刚性球的性能对比

仿真环境为消声室环境，麦克风阵列采用含有 32 个麦克风的空心球阵，对应的球谐函数分解阶次  $N_s = 4$  阶。空心球阵的半径从 0.01 米到 0.1 米均匀取值，间隔为 0.01 米。

麦克风阵列和声源位于同一水平面，俯仰角为  $\theta = 90^\circ$ ，且相距 2 米，声源位于麦克风阵列的正右方，即方位角  $\phi = 90^\circ$ 。声源信号包括两类，白噪声信号和语音信号。

不同声源类型（白噪声、语音）、不同麦克风阵列半径  $R$  对双耳信号的影响如图 5-4 所示，其中图 (a) 和图 (b) 是声源为白噪声信号，采用不同麦克风阵列半径时三种算法对应的双耳信号与参考双耳信号之间的  $\Delta_{ITD}$  和  $\Delta_{ILD}$  结果，图 (c) 和图 (d) 是声源为语音信号，采用不同麦克风阵列半径时三种算法的  $\Delta_{ITD}$  和  $\Delta_{ILD}$  结果。

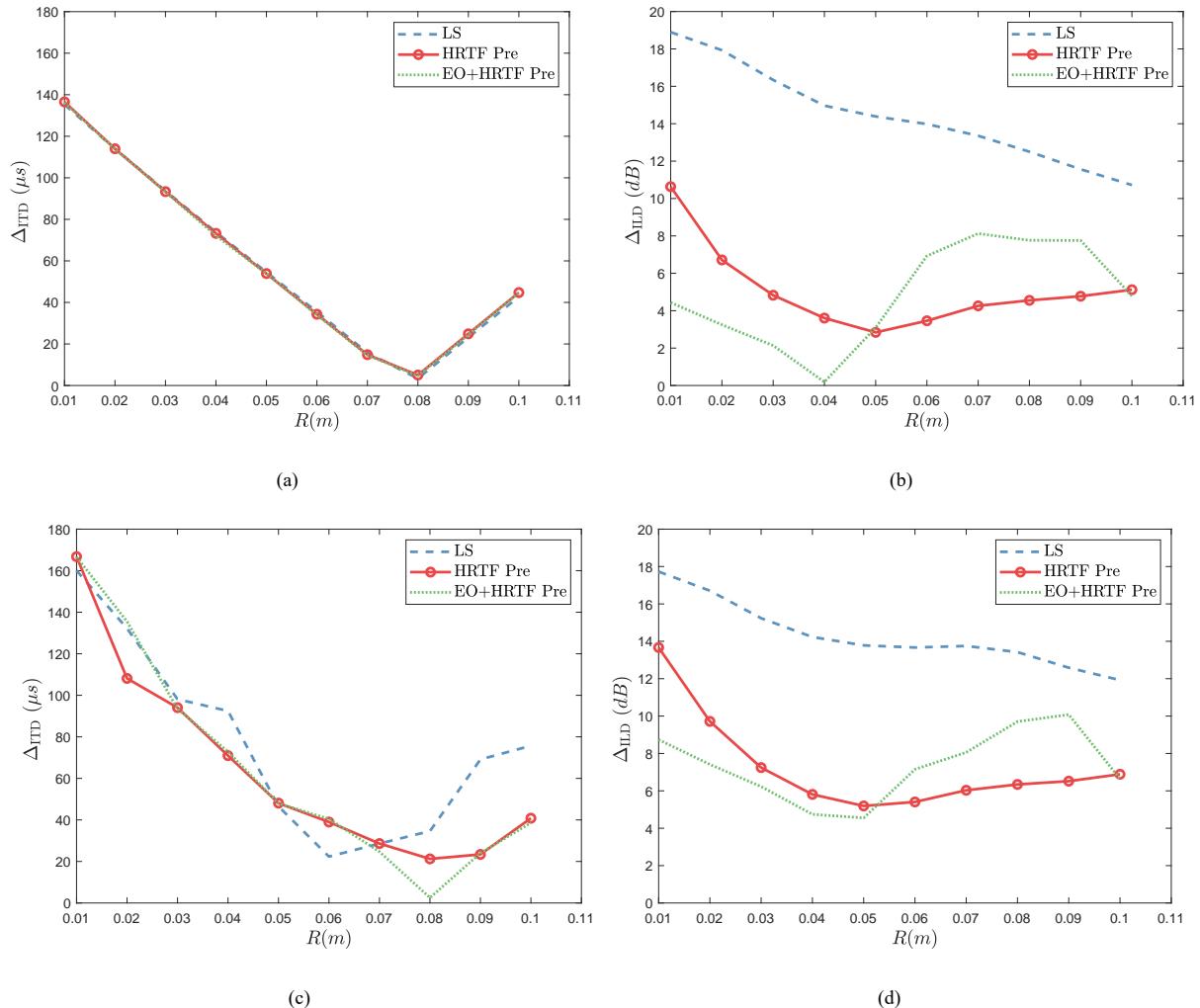


图 5-4 空心球情况下两种声源信号在不同麦克风阵列半径时的实验结果：白噪声：  
(a) ITD；(b) ILD；语音：(c) ITD；(d) ILD

从图 5-4 可以看出：

- (1) 无论是白噪声信号还是语音信号，LS、HRTF Pre 和 EO + HRTF Pre 算法的  $\Delta_{ITD}$  结果完全一致，与图 5-2 的讨论一致。并且当麦克风阵列半径接近人头尺寸时， $\Delta_{ITD}$  达到较小的值。
- (2) 无论是白噪声信号还是语音信号，HRTF Pre 和 EO + HRTF Pre 算法的  $\Delta_{ILD}$  结果

远小于 LS 算法, 且 EO + HRTF Pre 算法更适合半径较小的空心球阵(小于等于 0.05 m), HRTF Pre 算法更适合尺寸较大的麦克风阵列。

(3) 当麦克风阵列半径  $R$  大于等于 0.05 m 时, 声场扩阶算法开始不起作用, 甚至对 ILD 有一定的损伤,  $\Delta_{ILD}$  大于 HRTF Pre 算法。这是因为空心球阵情况下会存在球贝塞尔函数零点问题, 零点的引入会对声场扩阶算法中的球谐域 MUSIC 定位算法有影响, 导致错误的声场扩阶。

(4) 半径  $R = 0.1$  米时, HRTF Pre 和 EO + HRTF Pre 算法的  $\Delta_{ITD}$  和  $\Delta_{ILD}$  结果完全一致, 这是因为 EO + HRTF Pre 算法的目标球体半径  $R_T = 0.1$ , 此时两种算法等效。

对球贝塞尔函数的零点进行进一步分析, 图 5-5 给出了 EO + HRTF Pre 算法在不同麦克风阵列半径下三种算法所获取的双耳信号与参考信号在 8 kHz 以下频段内单个频点处的 ILD 结果, 以及 4 阶球贝塞尔函数  $j_4(kR)$  在不同半径  $R$  下随频率  $f$  的变化情况。

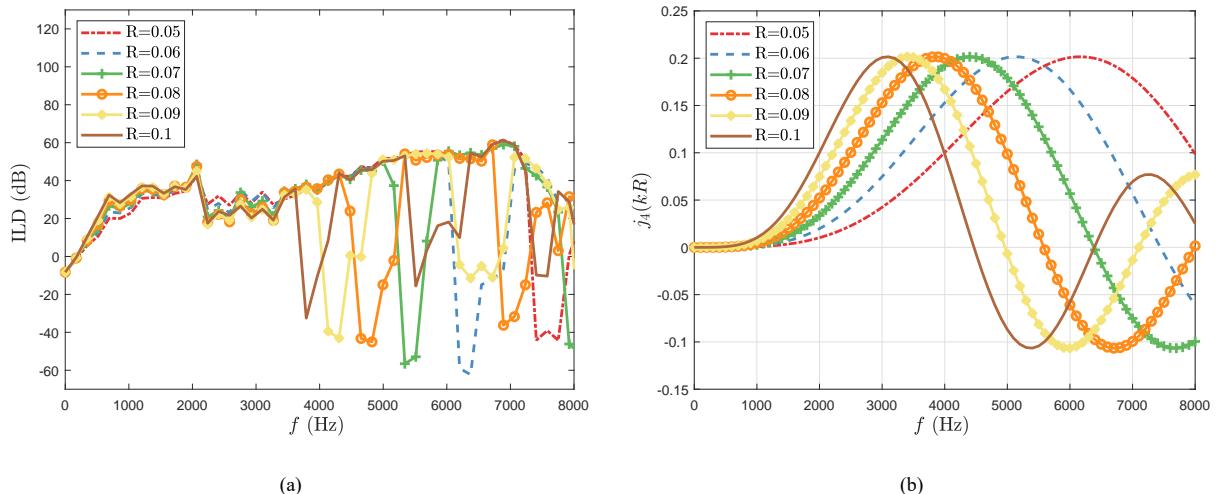


图 5-5 消声室情况下不同麦克风阵列半径在各个频点的 ILD

从图 5-5 可以看出:

(1) 本实验中声源的方位角  $\phi = 90^\circ$ , ILD 应该为正值, 而图 (a) 中部分频率的 ILD 出现负值, 且其绝对值较大, 严重影响了平均 ILD 结果, 因此导致了图 5-4 (b) 和 (d) 中的 ILD 损伤现象。

(2) 对比图 (a) 和图 (b) 可以看出, 固定麦克风半径  $R$ , ILD 出现错误的频率位于该半径下球贝塞尔函数的零点处或零点附近, 这是因为零点的存在导致球谐域 MUSIC 定位算法的错误, 从而导致 ILD 错误。

(3) 图 (b) 中, 对于 4 阶球贝塞尔函数来说,  $R \geq 0.05$  时, 零点开始出现在 8 kHz 以下的频段, 且半径  $R$  越大, 开始出现零点的频率越小, 8 kHz 以下的频段零点出现的次数越多。这也导致了半径越大, EO + HRTF Pre 算法的 ILD 结果与 Original 算法的差距

越大。

在第 3.1 节中对空心球和刚性球进行了详细讨论，并且已经从原理上证明，由于散射声场的存在，刚性球可以避免球贝塞尔函数的零点问题，从而解决空心球带来的一些问题，更加适用于实际系统。

接下来采用刚性球阵进行仿真实验，其余实验设置均保持不变，并且只关注 EO + HRTF Pre 算法出现 ILD 损伤的情况，即麦克风阵列半径从 0.05 米到 0.1 米均匀取值，间隔为 0.01 米。图 5-6 给出了刚性球情况下，不同声源类型和不同麦克风阵列半径下三种算法对应的双耳信号与参考双耳信号之间的  $\Delta_{ITD}$  和  $\Delta_{ILD}$  结果，图 5-7 给出了 EO + HRTF Pre 算法在采用不同麦克风阵列半径的刚性球情况下，三种算法所获取的双耳信号与参考信号在 8 kHz 以下频段内单个频点处的 ILD 结果。

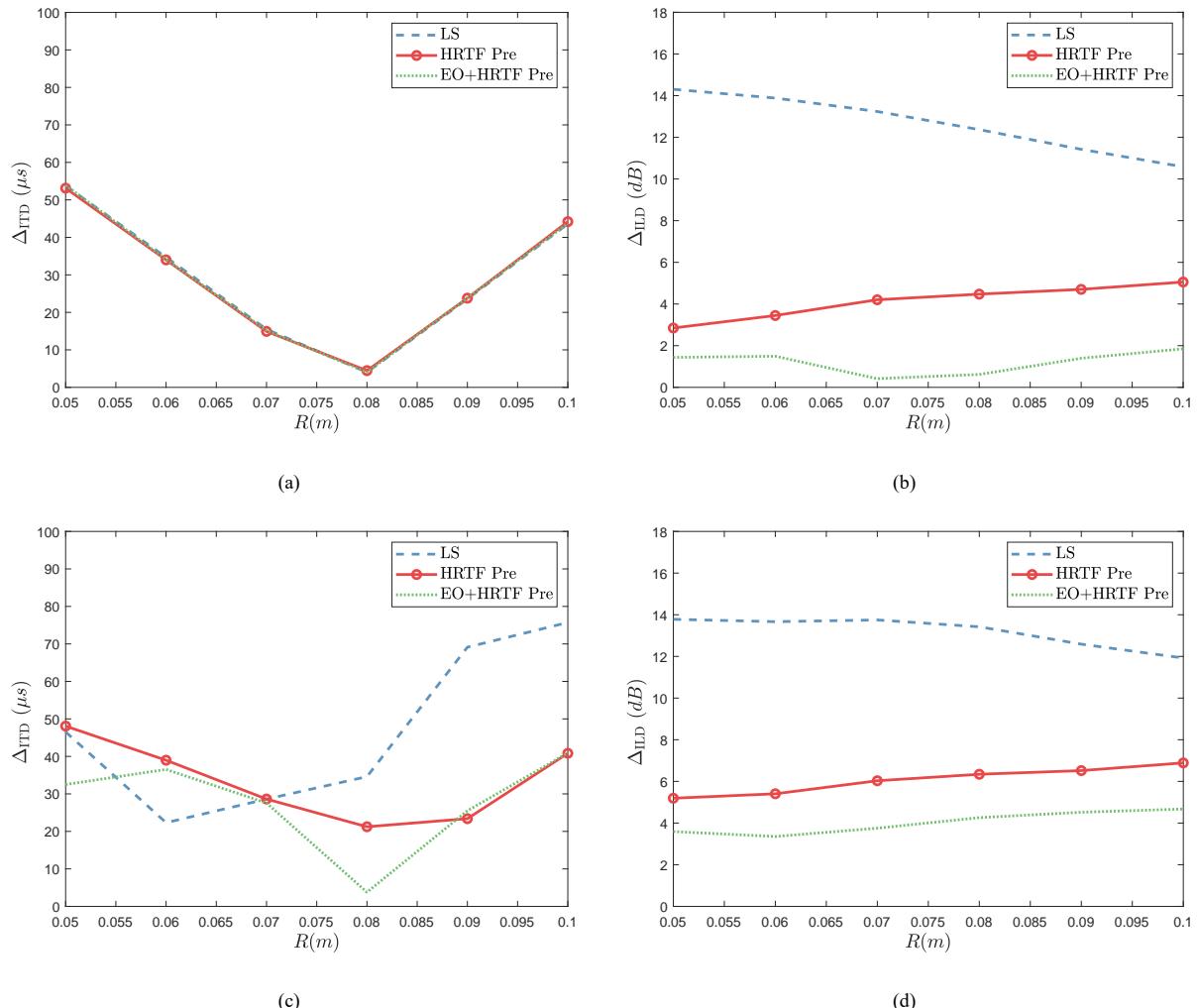


图 5-6 刚性球情况下两种声源信号在不同麦克风阵列半径时的实验结果：白噪声：  
(a) ITD；(b) ILD；语音：(c) ITD；(d) ILD

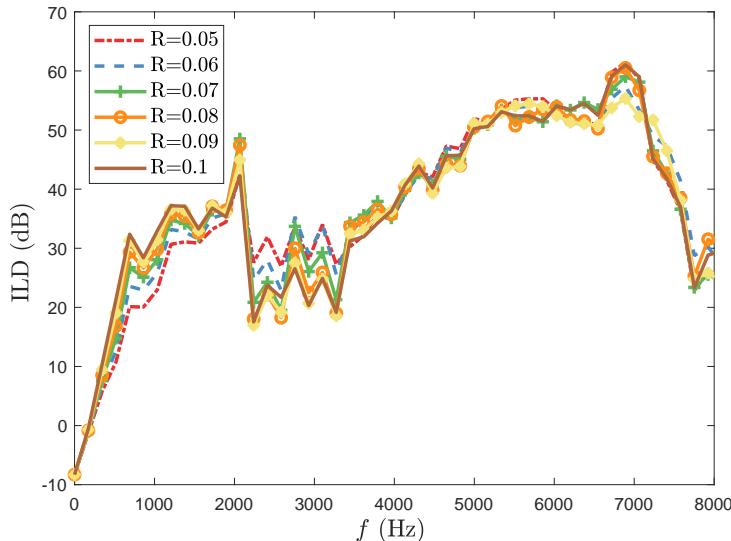


图 5-7 刚性球情况下采用不同麦克风阵列半径的单频点 ILD 结果

从图 5-6 和 5-7 可以看出：

- (1) 无论是白噪声信号还是语音信号，刚性球情况下 EO + HRTF Pre 算法在各个半径的  $\Delta_{\text{ILD}}$  结果最小，HRTF Pre 算法次之，LS 算法的误差最大，且远大于 HRTF Pre 和 EO + HRTF Pre 算法。
- (2) 刚性球的使用避免了 EO + HRTF Pre 算法在空心球情况中较大的麦克风阵列半径下的 ILD 损伤现象，因此 EO + HRTF Pre 算法在刚性球情况下适用于研究范围内的任意半径。
- (3) 对比图 5-5(a) 和 图 5-7 可以看出，刚性球可以避免球贝塞尔函数的零点给 EO + HRTF Pre 算法带来的 ILD 正负错误，体现了刚性球的优越性。

### 5.2.3 混响环境下各算法性能对比

本实验的仿真环境为混响环境，声源在混响室的传播可以用镜像模型方法（Image Source Method, ISM）产生<sup>[68]</sup>。镜像模型方法通常以矩形结构房间为基本模型，可通过调节墙壁反射系数  $\beta$ 、房间尺寸大小、声源位置及麦克风接收位置等参数来获得房间冲激响应，目前已经广泛应用于各种仿真实验中。其基本思想是用虚拟镜像来等效地替代声源在墙壁发生的反射，进而确定反射声的传播路径，一旦确定房间设置（尺寸、反射系数）及声源与麦克风的位置，既可计算所有的镜像声源坐标和对应的能量衰减系数。

反射系数的取值范围是 0 到 1 之间，本文选取的反射系数为 0、0.3、0.5 和 0.7，分别对应无混响、低混响、中等混响和强混响的声学环境。混响的强弱也可以用混响时间  $T_{60}$  这一指标来描述，其表示的是在扩散声场中当声源停止后从初始的声压级衰减 60 dB 时所对应的时间，描述了室内混响声的衰减快慢。常用的两个计算公式为艾润公式和赛宾公

式<sup>[18]</sup>:

$$\begin{aligned} T_{60} &= 0.161 \frac{V}{-S \ln(1 - \alpha)} \\ T_{60} &= 0.161 \frac{V}{S\alpha} \end{aligned} \quad (5-16)$$

其中,  $V$  为房间的体积,  $S$  为总的吸声面积,  $\alpha$  为平均吸声系数,  $\alpha = 1 - \beta^2$ ,  $\beta$  为反射系数。

混响环境下 IACC 是一个重要评价指标, 因此本实验中同时使用双耳时间差 ITD、双耳声级差 ILD 和双耳互相关系数 IACC 这三个评价指标对各种算法进行分析。和  $\Delta_{ITD}$ 、 $\Delta_{ILD}$  类似, 引入  $\Delta_{IACC}$ :

$$\Delta_{IACC} = | IACC_{case} - IACC_{Original} | \quad (5-17)$$

其中, case 表示 LS、HRTF Pre 和 EO+HRTF Pre。

本实验选取尺寸为  $5 \text{ m} \times 7 \text{ m} \times 3 \text{ m}$  的房间, 反射系数为 0、0.3、0.5 和 0.7。麦克风阵列采用 Eigenmike 球阵, 其包含 32 个麦克风, 对应的球谐函数分解阶次  $N_s = 4$  阶, 半径为 0.042 米。Eigenmike 位于房间中心, 与声源处于  $z = 1.5$  米的同一水平面上, 且相距 2 米。声源位于麦克风阵列的正右方, 即方位角  $\phi = 90^\circ$ 。声源信号包括两类, 白噪声信号和语音信号。

不同声源类型(白噪声、语音)、不同反射系数对双耳信号的影响如表 5-1、5-2 和 5-3 所示, 分别为三种算法对应的双耳信号与参考双耳信号之间的  $\Delta_{ITD}$ 、 $\Delta_{ILD}$  和  $\Delta_{IACC}$  结果。表 5-3 中的 JND 为参考双耳信号 Original 的 IACC 对应的最小可觉差, 参见式(5-14)。

表 5-1 两种声源信号在不同反射系数下的  $\Delta_{ITD}$  ( $\mu\text{s}$ )

反射系数	白噪声信号				语音信号			
	0	0.3	0.5	0.7	0	0.3	0.5	0.7
LS	67	26	12	7	163	12	6	4
HRTF Pre	66	27	14	8	123	34	12	5
EO+HRTF Pre	66	27	14	7	124	33	12	5

从表 5-1 可以看出: 无论是白噪声信号还是语音信号, 三种算法在各个反射系数下的  $\Delta_{ITD}$  基本一致, 语音信号情况下三种算法之间的误差相对较大, 这是因为语音信号在各频点上的能量分布不均导致的计算波动。整体结论与之前的实验一致, HRTF Pre 和 EO+HRTF Pre 算法并未对 LS 算法的 ITD 加以改变。

表 5-2 两种声源信号在不同反射系数下的  $\Delta_{ILD}$  (dB)

反射系数	白噪声信号				语音信号			
	0	0.3	0.5	0.7	0	0.3	0.5	0.7
LS	14.8	9.3	6.4	3.6	14.3	6.1	3.2	3.4
HRTF Pre	3.4	3.9	3.6	1.9	5.8	2.1	1.9	0.7
EO+HRTF Pre	0.5	0.6	0.7	0.8	4.6	0.4	2.1	1.9

表 5-3 两种声源信号在不同反射系数下的  $\Delta_{IACC}$ 

反射系数	白噪声信号				语音信号			
	0	0.3	0.5	0.7	0	0.3	0.5	0.7
LS	0.2866	0.0775	0.0172	0.0017	0.1066	0.0176	0.0040	0.0005
HRTF Pre	0.1897	0.0319	0.0076	0.0095	0.0589	0.0018	0.0008	0.0011
EO+HRTF Pre	0.1866	0.0467	0.0028	0.0071	0.0947	0.0153	0.0031	0.0001
JND	0.2847	0.0955	0.0260	0.0070	0.1049	0.0220	0.0085	0.0070

从表 5-2 可以看出：

(1) 无论是白噪声信号还是语音信号，相较于 LS 算法，HRTF Pre 算法和 EO+HRTF Pre 算法的  $\Delta_{ILD}$  明显降低。

(2) 对于白噪声信号，EO+HRTF Pre 算法的  $\Delta_{ILD}$  最小，小于 1 dB，HRTF Pre 次之，LS 算法的误差最大。对于语音信号，不同混响情况下，HRTF Pre 和 EO+HRTF Pre 算法的差异不大。ILD 是一个高频定位因素，因此相对来说白噪声下的结果更为可靠，即 EO+HRTF Pre 算法最优。

从表 5-3 可以看出：

(1) 无论是白噪声信号还是语音信号，在大多数混响环境下（反射系数为 0.7 除外），相较于 LS 算法，HRTF Pre 和 EO + HRTF Pre 算法的  $\Delta_{IACC}$  有所下降。

(2) 反射系数为 0.7 时，对于白噪声信号，HRTF Pre 和 EO + HRTF Pre 算法的  $\Delta_{IACC}$  与 JND 基本一致，且大于 LS 算法。对于语音信号，三种算法的  $\Delta_{IACC}$  均远小于 JND。但是无论是白噪声信号还是语音信号，其  $\Delta_{IACC}$  的数值都很小， $10^{-3}$  量级，可以认为三种算法在空间感上的差异性很小。

综上所述，HRTF Pre 和 EO+HRTF Pre 算法均对 LS 算法加以提升，EO+HRTF Pre 算法对 HRTF Pre 算法在大多数情况下加以提升，部分情况下保持不变。并且本文所提出的算法 EO+HRTF Pre 在任意混响条件下均可达到好的效果。

### 5.3 本章小结

本章主要对所提出的双耳渲染算法进行评估。首先介绍了双耳信号的三个评价指标及其计算方法，包括用于评估声源定位精度的双耳时间差 ITD 和双耳声级差 ILD，以及用于评估空间感知的双耳听觉互相关系数 IACC。

实验方面，从消声室和混响室两种环境对算法进行评估。消声室情况下首先探索了不同声源方位和声源类型对双耳信号的影响，结果表明，两种声源信号下，本文所提出的双耳渲染算法在声源位于任意方位角的情况下均可达到较小的误差。接下来对使用不同阵列半径的空心球和刚性球进行双耳渲染的性能加以对比，详细阐述空心球存在的零点问题及其对定位算法带来的影响，证明了刚性球的优越性。结果表明，空心球只适用于半径较小的球体，以避免球贝塞尔零点出现在关注频段内，而刚性球在任意半径下均可达到最小的误差。最后在不同的混响环境和不同声源类型下，对三种双耳渲染算法进行对比分析，结果表明，本文所提出的双耳渲染算法在多种混响情况下均能达到较好的效果。



## 第六章 总结与展望

### 6.1 全文总结

近年来，增强现实中的双耳渲染技术一直受到科学界众多学者的关注，在虚拟声学、建筑声学、语音通信、信息系统、媒体社交、教育和游戏娱乐等领域有着广泛的应用。本论文主要研究了基于球谐分解的双耳渲染技术，从录制声场和头相关传递函数两方面出发，对原始算法加以改进，进一步提升双耳信号的感知性能，并且使用球形麦克风阵列对所提出的算法进行评估。

本文的研究工作主要包括以下几个方面：

1. 对典型的双耳渲染算法进行了概述和总结，对本文所关注的基于球谐分解的双耳渲染算法及其存在的问题进行深入研究，并对声场的正交球谐分解等基础理论加以详细介绍，为后续工作奠定理论基础。
2. 研究了基于球谐分解的双耳渲染算法中声场和头相关传递函数的球谐分解及系数获取方法。在声场方面，分析了空心球和刚性球的特点，给出了正确获取声场系数所需的麦克风数目，以及两种阵列所获取的声场系数的区别，并说明了刚性球的优越性。在头相关传递函数方面，首先对其获取方式进行概述，并对本文所采用的两种数据库加以详细介绍，最后给出了球谐系数的获取方法。
3. 研究了头相关传递函数的预处理方法，并通过实验证明了该方法的有效性。本文采用与频率相关的相位对准预处理方法，基于高频处头相关传递函数的幅度部分对定位的感知起决定作用这一事实，对高频处的相位加以修正。该预处理方法一方面准确地表示头相关传递函数的幅度谱，并有效降低了头相关传递函数的分解阶次。最后在现有的头相关传递函数数据库上进行仿真实验，结果表明该预处理算法可以将头相关传递函数的分解阶次由原来的 34 阶降低至 15 阶。
4. 提出了一种新的声场扩阶理论，并且将声场扩阶理论、头相关传递函数预处理和基于球谐分解的双耳渲染算法相结合，实现了基于声场扩阶的双耳渲染算法，给出总体算法框架。该声场扩阶理论对麦克风阵列的采集声场进行分析，将入射声场分解为直达波和混响场的叠加。根据直达波入射方向进行空间加窗处理，可以实现声场分量的阶次提升，同时扩大控制区域的半径。除此之外，引入球谐域 MUSIC 定位算法以获取声源的入射方向，对采用的空域窗函数进行介绍并给出了相应的选取原则和球谐系数的获取方法。
5. 通过实验将本文所提出的算法与对标算法进行对比和分析，验证了本算法的有

效性。在消声室和混响环境下，从双耳时间差、双耳声级差和双耳听觉互相关系数这三个评价指标出发，对双耳信号的声源定位精度和空间感知进行评估。实验结果表明，在不同的混响情况下、声源位于不同方向、采用不同声源类型时，本文所提出的双耳渲染算法均能达到较好的效果，优于现有算法。并且在消声室环境下，对空心球和刚性球在不同双耳渲染算法下的性能进行对比，分析空心球存在的零点问题及其危害，验证了刚性球的优越性。

## 6.2 工作展望

本文以基于球谐分解的双耳渲染算法为基础，在声场和头相关传递函数两方面对原始算法加以改进，提出了一种新的声场扩阶理论，并提出一种新的双耳渲染算法，进一步提升了双耳信号的感知性能。本文虽然取得了一定的成果，但是由于时间有限，仍有部分内容没有深入研究，主要集中在以下两个方面：

1. 目前本文所提出的双耳渲染算法还未在实际系统上进行，可以借助已有的实验设备，建立一个演示系统，并在此基础上进行更广范围的听音评测。
2. 未来随着个性化听觉感知的发展，双耳渲染技术与个性化头相关传递函数的结合很有必要，可以将本文采用的头相关传递函数数据库更换为听者特有的头相关传递函数。

## 参考文献

- [1] Roginska A, Geluso P. Immersive sound: The art and science of binaural and multi-channel audio[M]. Taylor & Francis. 2017.
- [2] Zhang W et al. Surround by Sound: A Review of Spatial Audio Recording and Reproduction[J]. Applied ences. 2017,7(5).
- [3] Fletcher H. An acoustic illusion telephonically achieved[J]. Bell Labs Rec. 1933.
- [4] Kemar. <https://www.grasacoustics.com/industries/audiology/kemar>.
- [5] Neumann, dummy head ku-100. [http://www.neumann.com/?lang=en&id=current\\_microphones&cid=ku100\\_description](http://www.neumann.com/?lang=en&id=current_microphones&cid=ku100_description).
- [6] Head and torso simulator hats. <http://www.bksv.com/Products/transducers/ear-simulators/head-and-torso/hats-type-4128c?tab=overview>.
- [7] 3dio free space binaural microphone. [http://3diosound.com/index.php?main\\_page=product\\_info&cPath=33&products\\_id=45](http://3diosound.com/index.php?main_page=product_info&cPath=33&products_id=45).
- [8] Gardner B. HRTF Measurements of a KEMAR Dummy-Head Microphone[J]. MIT Media Lab Perceptual Computing-Technical Report 280. 1994, 280:1-7.
- [9] Xie B et al. Report on Research Projects on Head-Related Transfer Functions and Virtual Auditory Displays in China[J]. Journal of the Audio Engineering Society. 2013, 61(5):314-326.
- [10] Wierstorf H et al. A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances[C]. In 130th AES Convention. 2011.
- [11] Brungart D et al. The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources[C]. In Icad: the Meeting of the International Conference on Auditory Display. 2004.
- [12] Begault DR, Wenzel EM, Anderson MR. Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source[J]. Journal of the Audio Engineering Society Audio Engineering Society. 2001, 49(10):904-916.

- [13] Algazi VR, Duda RO. Headphone-Based Spatial Sound[J]. IEEE Signal Processing Magazine. 2011, 28(1):33-42.
- [14] Wightman FL, Kistler DJ. Headphone simulation of free-field listening. I: stimulus synthesis[J]. The Journal of the Acoustical Society of America. 1989, 85(2):858–867.
- [15] Wightman FL, Kistler DJ. Headphone simulation of free-field listening. II: Psychophysical validation[J]. The Journal of the Acoustical Society of America. 1989, 85(2):868–878.
- [16] Begault DR. 3-D Sound for Virtual Reality and Multimedia[M]. Academic Press Professional, Inc. 1994.
- [17] Wightman F, Kistler D. Measurement and validation of human HRTFs for use in hearing research[J]. Acta acustica united with Acustica. 2005, 91(3):429–439.
- [18] 谢菠荪. 头相关传输函数与虚拟听觉 [M]. 国防工业出版社. 2008.
- [19] Davis LS et al. High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues[C]. In Audio Engineering Society Convention 119. Audio Engineering Society2005.
- [20] Zotter F, Frank M. All-Round Ambisonic Panning and Decoding[J]. Journal of the Audio Engineering Society. 2012, 60(10):807-820.
- [21] Noisternig M et al. 3D binaural sound reproduction using a virtual ambisonic approach[C]. In Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003. VECIMS '03. 2003 IEEE International Symposium on. 2003.
- [22] Jieru C, Shanshan D, Wen Z. Binaural Rendering based on Linear Differential Microphone Array and Ambisonic Reproduction[J]. Journal of Fudan University(Natural Science). 2019.
- [23] Gerzon MA. Periphony: With-height sound reproduction[J]. Journal of the Audio Engineering Society. 1973, 21(1):2–10.
- [24] 张莹, 沈希辰. 从 Ambisonics 到双耳听觉——VR 电影声音制作方案研究 [J]. 复旦学报 (自然科学版). 2017, 56(002):215-221.
- [25] Evans MJ, Agnus JAS, Tew AI. Analyzing head-related transfer function measurements using surface spherical harmonics[J]. The Journal of the Acoustical Society of America. 1998, 104(4):2400-2411.

- [26] Bernschütz B et al. Binaural Reproduction of Plane Waves With Reduced Modal Order[J]. *Acta Acustica united with Acustica*. 2014, 100(5):972-983(12).
- [27] Sheaffer J, Villevial S, Rafaely B. Rendering binaural room impulse responses from spherical microphone array recordings using timbre correction[J]. 1014279/depositonce-4103. 2014.
- [28] Zaunschirm et al. Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint[J]. *Journal of the Acoustical Society of America*. 2018.
- [29] Ahrens J, Andersson C. Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre[J]. *The Journal of the Acoustical Society of America*. 2019, 145(4):2783-2794.
- [30] Rafaely B, Kleider M. Spherical Microphone Array Beam Steering Using Wigner-D Weighting[J]. *IEEE Signal Processing Letters*. 2008, 15(1):417-420.
- [31] Zhang W et al. Insights into head-related transfer function: Spatial dimensionality and continuous representation[J]. *Journal of the Acoustical Society of America*. 2016, 127(4):2347.
- [32] Avni A et al. Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution.[J]. *Journal of the Acoustical Society of America*. 2013, 133(5):2711.
- [33] Romigh GD et al. Efficient Real Spherical Harmonic Representation of Head-Related Transfer Functions[J]. *IEEE Journal of Selected Topics in Signal Processing*. 2015, 9(5):921-930.
- [34] Ben-Hur Z et al. Spectral equalization in binaural signals represented by order-truncated spherical harmonics[J]. *Journal of the Acoustical Society of America*. 2017, 141(6):4087.
- [35] Poerschmann C, Arend JM, Brinkmann F. Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2019, 27(6):1060-1071.
- [36] Brinkmann F, Weinzierl S. Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition[C]. In *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society2018.

- [37] Ben-Hur Z et al. Efficient Representation and Sparse Sampling of Head-Related Transfer Functions Using Phase-Correction Based on Ear Alignment[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2019, PP(99):1-1.
- [38] Ben-Hur Z et al. Binaural Reproduction Based on Bilateral Ambisonics and Ear-Aligned HRTFs[J]. Audio Signal Processing for Binaural Reproduction with Improved Spatial Perception. 2020, :69.
- [39] Hold C et al. Improving Binaural Ambisonics Decoding by Spherical Harmonics Domain Tapering and Coloration Compensation[C]. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019.
- [40] Lübeck T. Perceptual Evaluation of Mitigation Approaches of Errors due to Spatial Undersampling in Binaural Renderings of Spherical Microphone Array Data[J]. 2019.
- [41] Lübeck T et al. Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data[J]. Journal of the Audio Engineering Society. 2020, 68(6):428-440.
- [42] 朱天一, 卢晶. 基于双耳时间差的客观指标对 Ambisonics 回放系统的性能分析 [J]. 南京大学学报: 自然科学版. 2017, (06):1153-1160.
- [43] Williams EG, Mann JA. Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography</i>[M]vol. 108. 2000.
- [44] Kennedy RA, Sadeghi P. Hilbert Space Methods in Signal Processing[J]. Cambridge Uk. 2013, 10.1017/CBO9780511844515:xv-xviii.
- [45] Rafaely B. Fundamentals of Spherical Array Processing[M]vol. 16 of *Springer Topics in Signal Processing*. Cham: Springer International Publishing. 2019.
- [46] Ward DB, Abhayapala TD. Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers[J]. IEEE Transactions on Speech & Audio Processing. 2001, 9(6):p.697-707.
- [47] Gene H. Golub CFVL. Matrix Computations[M]. Johns Hopkins studies in the mathematical sciences. Johns Hopkins University Press3rd edn. 1996.
- [48] Jôt JM, Wardle S, Larcher V. Approaches to binaural synthesis[C]. In Audio Engineering Society Convention 105. Audio Engineering Society1998.

- [49] Jot JM, Larcher V, Pernaux JM. A comparative study of 3-D audio encoding and rendering techniques[C]. In Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction. Audio Engineering Society1999.
- [50] Benesty J, Chen J, Huang Y. Microphone array signal processing[J]. Journal of the Acoustical Society of America. 2008, 125(6):4097-4098.
- [51] Benesty J, Chen J. Study and Design of Differential Microphone Arrays[J]. Springer Topics in Signal Processing. 20136.
- [52] Benesty J, Chen J, Pan C. Fundamentals of Differential Beamforming[J]. 201610.1007/978-981-10-1046-0.
- [53] Rafaely B. Analysis and design of spherical microphone arrays[J]. IEEE Transactions on Speech and Audio Processing. 2005, 13(1):135-143.
- [54] Katz BF. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation[J]. The Journal of the Acoustical Society of America. 2001, 110(5):2440–2448.
- [55] Algazi VR et al. Approximating the head-related transfer function using simple geometric models of the head and torso[J]. The Journal of the Acoustical Society of America. 2002, 112(5):2053–2064.
- [56] Ziegelwanger H, Kreuzer W, Majdak P. MESH2HRTF: AN OPEN-SOURCE SOFTWARE PACKAGE FOR THE NUMERICAL CALCULATION OF HEAD-RELATED TRANSFER FUNCTIONS[C]. In The 22nd International Congress on Sound and Vibration. 2015.
- [57] Algazi VR et al. The CIPIC HRTF database[C]. In IEEE Workshop on Applications of Signal Processing to Audio & Acoustics. 2001.
- [58] Bernschütz B. A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100[C]. In DAGA Fortschritte der Akustik. 2013.
- [59] Sun H et al. Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing[J]. Journal of the Acoustical Society of America. 2012, 131(4):2828-2840.
- [60] Kennedy RA, Sadeghi P. Hilbert Space Methods in Signal Processing[J]. Cambridge Uk. 2013, 10.1017/CBO9780511844515:xv-xviii.

- [61] Wigner, E. P. Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra[J]. American Journal of Physics. 1960, 28(4):408.
- [62] Messiah A, [by] Albert Messiah; translated from the French by J. Potter. Quantum mechanics. V. 2[M]. North-Holland. 1961.
- [63] Mammasis K, Stewart RW, Thompson JS. Spatial Fading Correlation model using mixtures of Von Mises Fisher distributions[J]. IEEE Transactions on Wireless Communications. 2009, 8(4):2046-2055.
- [64] Seon KI. Smoothing of an All-sky Survey Map with a Fisher-von Mises Function[J]. Journal- Korean Physical Society. 200748(3).
- [65] Rumsey F. Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm[J]. Journal of the Audio Engineering Society. 2002, 50(9):651-666.
- [66] 方义、陈友元、牟宏宇、冯海泓. 基于双耳相干函数的鲁棒时延差估计与混响抑制算法 [C]. In 第十四届全国人机语音通讯学术会议.
- [67] Klockgether S, van de Par S. Just noticeable differences of spatial cues in echoic and anechoic acoustical environments[J]. The Journal of the Acoustical Society of America. 2016, 140(4):EL352–EL357.
- [68] Allen JB. Image method for efficiently simulating small-room acoustics[J]. The Journal of the Acoustical Society of America. 1979.

## 致谢

时光飞逝，转眼间已经要说再见，在西工大度过的七年时光历历在目，有着太多的不舍。学习和成长的过程，离不开老师的悉心教导和同学的关心帮助，在此我要向大家表示我最真诚的感谢。

首先要特别感谢我的两位指导老师：张雯教授和陈景东教授。从论文选题到研究工作的开展以及论文的撰写，都是在两位老师的指导下进行的。两位老师认真负责、精益求精的工作态度、渊博的知识储备、平易近人的处事方法深深影响了我，永远是我科研和生活的人生榜样。从本科毕业设计到现在，这三年多的时间里，我从刚开始的科研小白到现在有一定的理解和成果，都离不开两位老师的耐心指导，一次次的讨论让我对自己的研究方向和工作内容有了清晰的认知，使我的学习能力与动手能力得到了提升。老师们不仅在科研上给予帮助，还非常关心我们的生活情况和就业进展。

感谢教研室的师兄师姐和师弟师妹们，他们在整个研究生生涯中给予了我很多帮助，一起度过了很多快乐的时光。在探讨交流过程中，我吸收了很多宝贵的经验，为科研工作奠定基础。其中特别感谢郗经纬同学，在遇到问题时共同讨论、互帮互助，还有同级的陈奕彤和李梦真，在整个论文的撰写过程中相互鼓励，正是这种合作精神，给我们的研究生生涯画上了一个完美的句号。感谢张俊卿师兄在我遇到不懂的问题时给予分析和建议，对论文提出修改意见，同时感谢帮我审阅论文的师妹尤新雅、胡轩琦、师弟王佳乐，感谢你们提出的宝贵意见。感谢我硕士期间的宝藏舍友们，每天回到宿舍都会有一个开心轻松的氛围，互相关心，相信我们七年的友谊在之后的时间里可以得到更好的发展。

在此，还要感谢我的家人和男朋友对我无微不至的关心与理解。你们是我的坚强后盾，对我的任何决定都给予支持，让我可以毫无顾虑地勇往直前。在我遇到困难和压力时，你们的陪伴和鼓励使我能够保持乐观的心态，做出更好的成绩。

同时，我还要感谢我的班主任李道江老师和就业指导中心这个大家庭，在我的生涯规划和求职过程中给予指导和关心。最后还要向本领域进行相关研究的学者专家，以及各位审阅评议论文的老师表示由衷的感谢！



## 攻读硕士学位期间发表的学术论文和参加科研情况

### 发表学术论文

- [1] Chen J, Dou S, Zhang W. Binaural rendering based on linear differential microphone array and ambisonic reproduction[J]. Journal of Fudan University(Natural Science),2019, 58(3):370–377.

### 发表专利

- [1] 专利名称：一种混合型主动降噪方法与系统（已受理，第一发明人）。
- [2] 专利名称：一种主动降噪耳机的滤波器设计方法（已受理）。



## 西北工业大学 学位论文知识产权声明书

本人完全了解学校有关保护知识产权的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属于西北工业大学。学校有权保留并向国家有关部门或机构送交论文的复印件和电子版。本人允许论文被查阅和借阅。学校可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。同时本人保证，毕业后结合学位论文研究课题再撰写的文章一律注明作者单位为西北工业大学。

保密论文待解密后适用本声明。

学位论文作者签名 : \_\_\_\_\_

指导教师签名 : \_\_\_\_\_

年   月   日

年   月   日

## 西北工业大学

### 学位论文原创性声明

秉承学校严谨的学风和优良的科学道德，本人郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容和致谢的地方外，本论文不包含任何其他个人或集体已经公开发表或撰写过的研究成果，不包含本人或其他已申请学位或其他用途使用过的成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式表明。

本人学位论文与资料若有不实，愿意承担一切相关的法律责任。

学位论文作者签名 : \_\_\_\_\_

年   月   日