



华南理工大学
South China University of Technology

博士学位论文

高阶 Ambisonics 可听化重放技术

优化及其感知评估

作 者 姓 名 郭振宇
学 科 专 业 建筑技术科学
指 导 教 师 赵越喆 教授
所 在 学 院 建筑学院
论 文 提 交 日 期 2025 年 7 月

On the optimization and perception evaluation of higher-order Ambisonics based auralization reproduction system

A Dissertation Submitted for the Degree of Doctor of Philosophy

Candidate: Zhenyu Guo

Supervisor: Prof. Yuezhe Zhao

South China University of Technology

Guangzhou, China

分类号： TU1

学校代号： 10561

学 号： 202010100886

华南理工大学博士学位论文

高阶 Ambisonics 可听化重放技术优化 及其感知评估

作者姓名： 郭振宇

指导教师姓名： 赵越喆 教授

申请学位级别： 工学博士

学科专业名称： 建筑技术科学

研究方向： 建筑声学

论文提交日期： 2025 年 7 月 15 日

论文答辩日期： 2025 年 9 月 8 日

学位授予单位： 华南理工大学

学位授予日期： 年 月 日

答辩委员会成员：

主 席： 吴硕贤

委 员： 张东旭 张昌佳 孙海涛 赵越喆

华南理工大学

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。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标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：郭振宇

日期：2025年9月10日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华南理工大学。学校有权保存并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅（除在保密期内的保密论文外）；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。本人电子文档的内容和纸质论文的内容相一致。

本学位论文属于：

保密（校保密委员会审定为涉密学位论文时间：____年____月____日），于____年____月____日解密后适用本授权书。

不保密，同意在校园网上发布，供校内师生和与学校有共享协议的单位浏览；同意将本人学位论文编入有关数据库进行检索，传播学位论文的全部或部分内容。

（请在以上相应方框内打“√”）

作者签名：郭振宇

日期：2025年9月10日

指导教师签名：赵颖伟

日期：2025年9月10日

作者联系电话：18565363720

电子邮箱：zhenyu.guo.cn@outlook.com

联系地址（含邮编）：广东省广州市华南理工大学励吾科技楼，510640

摘要

可听化技术旨在通过数字信号处理等手段在虚拟环境下重现真实空间的声环境，使听众在虚拟环境下产生与目标空间中一致的听觉感知。可听化是建筑声学设计、听觉研究以及虚拟现实等领域中的核心技术之一。实现可听化需要对声源进行建模，获取室外或房间内声源在接收点处产生的脉冲响应，然后对信号进行合成与重放。空间声重放技术是实现可听化的核心环节之一，其中，高阶 Ambisonics 技术凭借其物理声场重构能力，在实现高精度、沉浸式的空间声重放上具有巨大潜力。然而，基于 Ambisonics 重放的可听化系统在实际应用中面临两个主要问题：（1）在技术层面，扬声器规模与重放精度难以兼顾，且算法要求在空间上均匀布置大量扬声器，这在实际应用中难以实现；（2）在感知层面，重放虚拟声环境下的主观感知效果尚不清晰，特别是缺少对高层次感知维度的研究，包括言语清晰度、音色、方向感知、空间感等方面。这两个问题互相关联，首先需要对听众在虚拟重放环境下的感知情况进行评估，然后，依据感知结果对系统进行针对性的改进或简化。因此，本研究围绕高阶 Ambisonics 声场重建与可听化系统的改进优化与重放系统下的感知评估问题，开展了以下几个方面的研究：

第一，建立了一套多通路扬声器阵列高阶 Ambisonics 可听化重放平台。系统实现了完整的信号处理链路，包括 Ambisonics 空间声重放、室内声场空间脉冲响应合成（结合几何声学仿真软件）、多通道均衡校准及重放，以及多声源场景合成、声场分析、声场拾拾等模块。该系统是国内规模最大的高阶 Ambisonics 重放系统之一，为后续研究提供了硬件与软件基础。

第二，探讨了 Ambisonics 系统重放阶数对声场重放误差以及主观感知的影响。首先通过仿真模拟分析了 Ambisonics 重放阶数在客观层面对重放误差的影响，包括声场重构误差大小、双耳接收信号误差以及能量扩散问题。然后，开展主观听音实验，针对重放阶数对言语接受阈值的影响开展研究。实验在不同阶数的 Ambisonics 重放下分别测量了听众的言语接受阈值。研究发现，言语接受阈值随阶数升高而降低，9 阶重放下的结果已接近单扬声器参考水平的结果。低阶系统（1 阶、3 阶）重放下存在严重的多通道串扰，导致频谱失真，显著提高了言语接受阈值，清晰度较低。该研究明确了高阶重放系统（如 9 阶）对保证虚拟环境下语音信号可靠感知的重要性，也为系统的阶数选择提供了数据参考。

第三，提出了一种针对分层扬声器阵列的 Ambisonics 重放算法。该算法针对高阶 Ambisonics 重放要求扬声器阵列在空间上均匀分布这一难点，算法结合二维 Ambisonics 与矢量调幅技术，降低了系统对扬声器数量与空间均匀性的要求，尽可能提高重放阶数，并提高了水平面区域的重放精度。使用头相关传输函数与声场辐射函数分别进行双耳接收信号与声场的模拟，客观模拟验证了分层 Ambisonics 算法相比三维 Ambisonics，合成的虚拟声源更接近理想目标声源。

第四，开展主观听音实验，将分层 Ambisonics 算法与现有的多种混合阶 Ambisonics 算法以及 AllRAD 算法进行对比。主观绝对定位实验与声音保真度对比评估实验表明，本文所提出的算法在水平面和 30°仰角平面等具有实际扬声器阵列的平面上的定位误差显著低于现有混合阶 Ambisonics 及 AllRAD 算法，声音保真度更优。

第五，开展实验研究反射声重放的仰角范围及重放精度对空间感的影响。实验首先对虚拟厅堂声环境进行可听化重放，然后通过对比评分实验，比较了不同反射声重放仰角范围以及能量扩散程度（阶数、算法）对空间感（声源宽度、包围感）的影响。实验的主要发现有：二维重放相比三维重放会显著降低空间感，但听觉对反射的具体仰角不敏感，不同反射声重放的仰角范围对空间感影响较小；早期与后期反射的仰角信息均对听众包围感有显著贡献；早期反射的能量扩散对空间感知影响较大，特别是在反射相对复杂的室内声环境中。该研究探索了反射声仰角信息对听觉感知的影响，为可听化重放算法优化提供了新的可能性，此外，实验还验证了本文所提出的重放算法在空间感知上的可靠性。该实验有助于进一步结合听觉感知的特性，探索简化可听化系统的可能性，

本研究建立了完整的高阶 Ambisonics 可听化系统，能够在多个建筑声学实际应用场景中进行使用。本文深入研究了高阶 Ambisonics 可听化系统在不同虚拟声场景下的感知规律，本文所提出的分层重放算法有效解决了非均匀阵列下的重放难题。本文的研究成果为高阶 Ambisonics 在建筑声学设计中的应用提供了理论依据、基础数据和技术参考，对推动高阶 Ambisonics 可听化技术的发展具有重要意义。

关键词：Ambisonics；可听化；声场重建；声源定位；空间感

ABSTRACT

Auralization technology is devoted to reproducing the sound environment of real spaces in a virtual venue through digital signal processing, enabling listeners to achieve consistent auditory perception under the real space. It has been well recognized as one of the vital technologies in architectural acoustic design, auditory research, and virtual reality experiences. Implementing auralization requires modeling sound sources, acquiring the impulse response of outdoor or indoor sound sources at the receiver, and then synthesizing and replaying the signals. Spatial sound reproduction is one of the key components of auralization, and higher-order Ambisonics (HOA) technology has demonstrated enormous potential in achieving high-precision and immersive spatial sound reproduction due to its physical sound field reconstruction capability. However, Ambisonics-based auralization reproduction face two main challenges in practices: (1) On the technical aspect, there is a trade-off between loudspeaker quantity and reproduction accuracy, and the algorithm requires spatial-uniformly distributed loudspeaker array, which is difficult to achieve in reality; (2) On the perceptual aspect, the subjective perception of reproduced virtual acoustic environments remains unknown, especially regarding the higher-level perceptual dimensions, including speech intelligibility, timbre, direction perception, spaciousness, and etc. These two issues are interrelated, that is, the perception situation under virtual reproduced environment needs to be determined. After that, the system can be optimized or simplified with targets. Therefore, the thesis conducts the following research, focusing on the optimization and perception evaluation of HOA-based sound field reconstruction and auralization:

First, a HOA auralization reproduction platform with multiple-channel loudspeaker array was established. The system implements a complete chain of signal processing, including Ambisonics spatial sound reproduction, indoor sound field spatial impulse response synthesis (combined with geometric acoustic simulation software), multi-channel equalization and playback, as well as multi-source scene synthesis, sound field analysis and spatial sound field record modules. This system is one of the largest HOA reproduction systems in China, providing hardware and software foundations for subsequent research.

Second, the effects of reproduction order on reproduction error and subjective perception were explored. In objective aspect, the effect of Ambisonics reproduction order on reproduction error was first inspected by simulation, including sound-field reconstruction error, binaural received signals, and energy spread issue. After that, subjective listening experiments were

conducted to investigate the impact of Ambisonics reproduction order on the speech reception threshold (SRT). SRTs of listeners were measured under different orders of Ambisonics reproduction conditions. The results showed that SRTs decreased with increasing reproduction order, and results of 9th-order reproduction approached the results of the reference level (single loudspeaker playback). Low-order systems (1st and 3rd orders) reproduction suffered from spectral distortion due to multi-channel crosstalk, significantly increasing SRT values, which indicates degraded speech intelligibility. Additionally, low-order systems were also affected by the spatial release from masking effect caused by energy spread, which enhanced speech intelligibility in noisy environments. This study highlights the importance of high-order reproduction systems (e.g., 9th order) to ensure the reliable perception of speech signal under virtual environments and provides data references for system order selection.

Third, a new Ambisonics reproduction algorithm for layered loudspeaker arrays was proposed. The algorithm focuses on addressing the difficulties that Ambisonics reproduction requires spatial-uniformly distributed loudspeaker array. It combines 2D Ambisonics with vector amplitude panning technology, reducing the system requirements for loudspeaker quantity and spatial uniformity and increasing the order as much as possible, hence improving reproduction accuracy in the horizontal plane. By simulating the binaural signals and sound field with head-related transfer functions and sound radiation functions, it was verified that compared to the 3-D Ambisonics, the virtual sources synthesized by multi-layer Ambisonics were more like the target ideal sources.

Fourth, subjective listening experiments were conducted to compare this algorithm with existing mixed-order Ambisonics algorithms and the AllRAD algorithm. Subjective localization and sound fidelity comparison evaluation experiments showed that the algorithm proposed in the thesis had significantly lower localization errors than existing mixed-order Ambisonics and AllRAD algorithms in the plane with real loudspeaker array including horizontal plane and 30° elevation plane, the sound fidelity was also superior to others.

Fifth, experiments were conducted to investigate the impact of the elevation range and reproduction accuracy of reflected sounds on spaciousness. The experiment first reproduced a virtual concert hall environment with auralization, subjective perception evaluation was then conducted to explore how the elevation range of reflected sound reproduction and the degree of energy spread (order, algorithm) affect spaciousness (source width, envelopment). Key findings of experiments included: 2D reproduction significantly reduced spaciousness compared to 3D reproduction, but auditory was insensitive to the specific elevation of reflections, with little impact of different elevation ranges on spatial awareness; height

information of both early and late reflections significantly contributed to listener envelopment; energy spread of early reflections had a greater impact on spaciousness, especially in complex indoor acoustic environments. This study explores the influence of reflected sound elevation information on auditory perception, offering new possibilities for optimization of auralization algorithms. Additionally, the experiment verified the reliability of the proposed multi-layer Ambisonics reproduction algorithm in spaciousness. The experiment is dedicated to further explore the possibility of simplifying auralization systems by considering the features of auditory perception.

This study establishes a comprehensive HOA auralization system that can be applied to various architectural acoustic applications scenarios. The thesis deeply investigates perceptual patterns in different virtual sound environments, and the proposed multi-layer Ambisonics reproduction algorithm effectively addresses the reproduction challenges of non-uniform arrays. The research findings of this thesis provide theoretical foundations, experimental data, and technical reference for applications of HOA in architectural acoustic design, soundscape research, and virtual reality, contributing to the advancement of HOA auralization technology.

Keywords: Ambisonics; auralization; sound field reconstruction; sound source localization; spaciousness

目录

摘要	I
ABSTRACT	III
目录	VI
图清单	X
表清单	XII
数学符号列表	XIII
缩略词注释	XIV
第一章 绪论	1
1.1 研究背景	1
1.2 空间感知	2
1.2.1 听觉系统	2
1.2.2 空间听觉	4
1.2.3 空间印象感知	6
1.3 空间声重放	7
1.3.1 坐标系	7
1.3.2 声音	8
1.3.3 双耳重放	9
1.3.4 多通路环绕声	10
1.3.5 Ambisonics	11
1.3.6 波场合成	12
1.4 研究现状	13
1.4.1 Ambisonics 重放算法改进	13
1.4.2 Ambisonics 声重放的主观评价	14
1.4.3 Ambisonics 系统在建筑声学中的应用	15
1.4.4 现有研究的不足之处	16
1.5 研究目的与研究内容	17
1.6 研究思路与方法	18
第二章 高阶 Ambisonics 可听化系统的原理与设计	20

2.1 引言	20
2.2 高阶 Ambisonics 的基本原理	20
2.2.1 球谐函数	20
2.2.2 声场的球谐函数分解	22
2.2.3 Ambisonics 编码与解码	23
2.2.4 Ambisonics 解码优化	27
2.2.5 混合阶 Ambisonics	29
2.2.6 Ambisonics 信号捡拾	33
2.3 高阶 Ambisonics 可听化系统设计	36
2.3.1 系统框架	36
2.3.2 硬件系统设计及调试	37
2.3.3 可听化重放实现	46
2.4 本章小结	48
第三章 高阶 Ambisonics 阶数对声场重放误差的影响及分析	50
3.1 引言	50
3.2 Ambisonics 重放阶数对声场重放的误差影响的模拟分析	51
3.3 实验背景	55
3.3.1 重放阶数对感知的影响	55
3.3.2 语言清晰度评价方法	55
3.4 重放阶数对言语接受阈值影响的主观实验	57
3.4.1 实验目的	57
3.4.2 实验对象	57
3.4.3 实验方法	57
3.5 实验结果	61
3.5.1 实验结果统计分析	61
3.5.2 实验结果讨论	65
3.6 本章小结	66
第四章 分层扬声器阵列下的重放算法改进	67
4.1 引言	67
4.2 理论基础	68

4.2.1 算法目标	68
4.2.2 理论基础	69
4.3 分层 Ambisonics 空间声重放算法	71
4.4 算法客观验证	73
4.4.1 HRTF 模拟	73
4.4.2 声场模拟	75
4.4.3 结论	76
4.5 本章小结	77
第五章 分层 Ambisonics 与混合阶 Ambisonics 重放的主观感知对比实验	78
5.1 引言	78
5.2 研究背景	79
5.2.1 混合阶 Ambisonics 研究现状	79
5.2.2 空间声重放系统主观评价方法	80
5.3 混合阶及分层 Ambisonics 的定位与声音保真度主观评估实验	82
5.3.1 实验目的	82
5.3.2 实验对象与刺激信号	82
5.3.3 定位实验	85
5.3.4 声音保真度评估实验	88
5.4 实验结果与讨论	90
5.4.1 定位实验结果与讨论	90
5.4.2 声音保真度评估实验结果与讨论	96
5.5 本章小结	98
第六章 高阶 Ambisonics 可听化下虚拟厅堂的空间感知研究	100
6.1 引言	100
6.2 研究背景	101
6.2.1 反射声仰角对空间感的影响	101
6.2.2 重放阶数对空间感的影响	101
6.3 高阶 Ambisonics 可听化重放的空间感评估实验	102
6.3.1 实验目的	102
6.3.2 实验对象与刺激信号	103

6.3.3 实验 1—反射声重放仰角范围对空间感影响	109
6.3.4 实验 2—早期反射重放方式对空间感影响的评估实验	111
6.4 实验结果与讨论	112
6.4.1 实验 1 结果	112
6.4.2 实验 2 结果	113
6.4.3 结果讨论	115
6.5 本章小结	116
结论与展望	117
研究结论	117
创新点	118
展望	119
参考文献	120
附录 1 N3D 到 N2D 补偿系数 α 的计算过程	129
攻读博士学位期间取得的研究成果	130
致 谢	132

图清单

图 1-1 听觉系统截面图, 修改自 WikiJournal of Medicine ^[12]	3
图 1-2 双耳定位线索: ITD、ILD 与谱线索 (第一、第二频谱峰值 P1、P2; 第一、第二频谱谷值 N1、N2)	5
图 1-3 球坐标系.....	7
图 1-4 论文结构图.....	19
图 2-1 0 至 3 阶球谐函数及角谐波函数的图像.....	22
图 2-2 不同优化下 3 阶 Ambisonics 在水平面上的虚拟指向性.....	28
图 2-3 0 至 3 阶的 Zonal、Tesseral 及 Sectoral 球谐函数.....	29
图 2-4 2 阶 3-D 与 4H2P 混合阶 Ambisonics 下的正前方扬声器指向性.....	31
图 2-5 球形传声器阵列拾信号处理流程.....	34
图 2-6 高阶 Ambisonics 可听化系统框架.....	37
图 2-7 硬件系统框架图.....	38
图 2-8 脉冲响应测量原理图.....	40
图 2-9 传声器指向性测量示意图.....	42
图 2-10 传声器在不同方向上的频率响应以及对应的均衡滤波器.....	43
图 2-11 通道 1 扬声器的脉冲响应及其频谱.....	44
图 2-12 均衡前与均衡后扬声器的脉冲响应及其频谱.....	45
图 2-13 室内声场可听化信号处理流程.....	47
图 3-1 不同重放阶数 Ambisonics 重构($15^\circ, 0^\circ$)方向虚拟声源声场的结果及其误差	52
图 3-2 正前方目标声源以及不同阶数 Ambisonics 重放合成声源在左耳处的 HRTF	53
图 3-3 不同阶数 Ambisonics 重放在空间中的理论能量分布.....	54
图 3-4 言语接受阈值测量实验示意图及过程照片.....	58
图 3-5 SRT 测量过程示例	60
图 3-6 言语接受阈值实验结果.....	62
图 3-7 言语接受阈值实验数据正态性检验的 Q-Q 图.....	63
图 3-8 SRT-SD 的线性拟合结果	64

图 4-1 不同方向上虚拟声源和目标声源在水平面（1000 Hz）的合成声场与误差.....	70
图 4-2 分层 Ambisonics 重放算法概念图与流程图.....	71
图 4-3 不同声源方向下，目标声源、分层 Ambisonics 与 3-D Ambisonics 在左耳处的 HRTF.....	74
图 4-4 分层与 3-D Ambisonics 在水平面（1000 Hz）的合成声场与合成误差.....	76
图 5-1 不同混合阶 Ambisonics 的球谐函数选择策略.....	80
图 5-2 SOFE 系统照片、定位实验过程照片以及声音保真度评估实验 GUI 界面 ...	83
图 5-3 三种混合阶 Ambisonics 重放在 SOFE 系统下的解码矩阵的条件数	85
图 5-4 定位与保真度评估实验流程图与听音位置示意图.....	87
图 5-5 头中定位率与前后混淆率条形图.....	91
图 5-6 定位实验中不同重放方法的反馈方向分布.....	92
图 5-7 中心听音情况下水平面声源的方位角绝对定位误差.....	93
图 5-8 中心听音情况下中垂面和侧垂面声源的仰角定位绝对定位误差.....	94
图 5-9 中心听音情况下不同仰角声源的方位角和仰角绝对定位误差.....	95
图 5-10 水平面声源在两个听音位置下的绝对方位角与仰角定位误差.....	96
图 5-11 声音保真度评估实验结果.....	97
图 6-1 干信号时域波形图和长时平均谱.....	103
图 6-2 实验房间三维模型透视图.....	104
图 6-3 实验 1 中所有条件重放虚拟声场的 T_{20} 与 C_{80}	107
图 6-4 实验 2 中所有条件重放虚拟声场的 T_{20} 与 C_{80}	108
图 6-5 实验流程图、MUSHRA 评分操作界面以及实验照片	110
图 6-6 实验 1 实验结果箱型图.....	112
图 6-7 实验 2 实验结果箱型图.....	114

表清单

表 2-1 Ambisonics 声场重构模型组合.....	25
表 2-2 不同阶数下混合阶 Ambisonics 解码矩阵的条件数	32
表 2-3 球阵列拾声场辐射函数补偿滤波器系数.....	35
表 2-4 192 通道扬声器阵列各扬声器空间方位.....	39
表 3-1 言语接受阈值测量的实验条件列表.....	59
表 3-2 拉丁方矩阵平衡的实验次序.....	59
表 3-3 言语接受阈值测量实验的 rm ANOVA 统计结果	64
表 5-1 SOFE 系统扬声器阵列各扬声器空间方位	84
表 5-2 定位实验虚拟声源方向.....	86
表 5-3 声音保真度评估实验虚拟声源方向.....	89
表 6-1 音乐厅模型主要界面材料吸声系数与散射系数.....	105
表 6-2 实验中所有刺激的反射编码方法.....	106

数学符号列表

符号	含义
Ω	极坐标系下的坐标
θ	方位角
ϕ	仰角
r	距离
P	频域声压
j	复数单位
k	波数
\vec{r}	笛卡尔坐标系下的位置矢量
f	频率
t	时间
c	声速
J_n	第一类贝塞尔函数
j_n	第一类球贝塞尔函数
h_n	第一类球汉克尔函数
$h_n^{(2)}$	第二类球汉克尔函数
Y_n^m	球谐函数
P_n^m	缔合勒让德多项式（无 Condon-Shortley phase 项）
P_n	勒让德多项式
δ_{nn}	冲击函数
Φ_n^m	角谐波函数
N	Ambisonics 阶数
\mathbb{S}^2	二维单位球面
\mathbb{N}	自然数集

缩略词注释

缩略词	英文全称	中文释义
AI	Articulation Index	清晰度指数
ARIR	Ambisonics Room Impulse Response	Ambisonics 房间脉冲响应
ASW	Apparent Source Width	视在声源宽度
BRIR	Binaural Room Impulse Response	双耳房间脉冲响应
DRR	Direct-to-Reverberant Ratio	直达混响比
FIR	Finite Impulse Response	有限冲激响应
FOA	First-order Ambisonics	一阶 Ambisonics
HOA	Higher-order Ambisonics	高阶 Ambisonics
GUI	Graphical User Interface	图形用户界面
HRIR	Head-Related Impulse Response	头相关脉冲响应
HRTF	Head-Related Transfer Function	头相关传输函数
IACC	Interaural Cross-correlation Coefficient	双耳互相关系数
ILD	Interaural Level Difference	双耳声级差
IPD	Interaural Phase Difference	双耳相位差
ITD	Interaural Time Difference	双耳时间差
LEV	Listener Envelopment	听者包围感
LTAS	Long-Term Average Spectrum	长时平均频谱
LTI	Linear time invariant	线性时不变
MLS	Maximum Length Sequence	最大长度序列
MMD	Mode-Match Decoding	模态匹配解码
MUSHRA	Multiple Stimuli with Hidden Reference and Anchor	多刺激隐藏参考锚点法
NFC	Near-field Compensation	近场补偿
RIR	Room Impulse Response	房间脉冲响应
rm ANOVA	Repeated Measures Analysis of Variance	重复测量方差分析
SHT	Spherical Harmonics Transform	球谐函数变换

SII	Speech Intelligibility Index	言语可懂度指数
SNR	Signal-to-Noise Ratio	信噪比
SD	Spectral distortion	谱失真
SRM	Spatial Release from Masking	空间掩蔽释放
SRT	Speech Reception Threshold	言语接受阈值
STI	Speech Transmission Index	语音传输指数
VBAP	Vector Base Amplitude Panning	矢量幅度调幅
WFS	Wave Field Synthesis	波场合成

第一章 绪论

1.1 研究背景

可听化（Auralization）是一项使用仿真、测量或合成的数据产生可听信号的技术^[1]，旨在通过信号控制手段在虚拟环境下重现真实空间的声环境，使听众在重构的虚拟环境下产生与目标空间一致的听觉事件。可听化技术在建筑声学设计、听觉研究及声景设计、虚拟现实体验等领域具有重要的地位。特别是在建筑声学设计中，通过可听化技术，设计者能够在音乐厅、剧院等场馆建成或修缮前，对未建成场馆的声环境进行主观试听，进而预判与展示其室内声环境的质量。尽管现有建筑声学相关研究已经建立了一系列评估声场环境质量的客观指标，如混响时间，清晰度等，主观试听仍是建筑声学设计中的重要参考依据。

实现可听化主要包括四个核心步骤：房间模型的建立、脉冲响应仿真与测量、多通道处理信号处理及卷积运算、声音重放^[2]。其中，房间模型及其脉冲响应可以使用计算机软件进行三维建模，然后通过声学模拟软件对房间脉冲响应（Room impulse response, RIR）进行仿真计算获取，也可以在实地或按一定比例缩小的缩尺模型中进行测量得到^[3]。通过仿真或测量得到 RIR 后，使用空间声重放技术对 RIR 进行多通路信号处理，卷积任意目标干信号后馈给对应重放设备进行播放，进而实现可听化。

根据最近的一项大规模调研，目前超过 40% 的建筑声学设计项目采用耳机或立体声双扬声器作为可听化重放终端^[4]，在算法上使用对应的双耳空间声重放对 RIR 进行编码处理。双耳重放的算法复杂度与重放终端的成本相对较低，并且已有大量的心理声学实验证明这种方式可以让听众获得相当准确的空间定位效果^[5]。但受限于耳机平台，双耳重放存在头中定位^[6]、前后混淆、声音外化效果差、难以支持头部转动、无法多人同时体验、以及无法提供给特殊人群（如助听器、人工耳蜗佩戴者）体验等缺点。

近年来，随着物理声场重构理论与数字信号处理技术的发展，基于扬声器阵列的声场重构系统得到了广泛的关注。此类方法可以使用一定数量的扬声器，在特定区域内直接重建任意复杂声场。扬声器重放克服了上述双耳重放的一系列缺陷。传统的多通路扬声器重放技术包括立体声技术、5.1 通道环绕声技术等，受限于扬声器布置与算法限制，它们难以实现复杂声场的可靠重放。物理声场重构技术能够实现更精确的重放，但物理声场重放通常需要数量更多的扬声器单元，其在算法复杂度、计算量上也显著高于其他方法。

在物理声场重放技术中，Ambisonics 技术能够使用相对较少数量的扬声器在一定区域实现对物理声场的重构，具有在实际系统进行实现与应用的可能性。Ambisonics 重放的主要优势之一是可以实现对三维空间中任意方向声源的声场重构，在实际建筑声环境中，声源与反射声可能来自于空间中的任意方向，这一特性对实现建筑声学环境的可听化具有重要意义。然而，实现三维重放需要大量位于水平面以外的扬声器，这在实际系统中往往是难以实现的。此外，由于重放扬声器数量限制，重构的声场存在一定的误差，这种误差不仅会对简单自由场虚拟声源的感知造成影响，还可能会降低复杂重放声环境下的感知可靠性。

声信号的听觉感知存在较大的冗余性，对于声信号中的特定细节，听觉不一定能够进行识别或区分。例如，对于部分高仰角的声源方向感知，听觉的分辨率仅为 10° 左右^[7,8]。对于语音信号，听觉即使在语音信号存在较大的噪声、混响干扰以及频谱失真的情况下，仍能进行有效的识别^[9]。实际的声重放系统往往需要结合听觉特性，对特定误差进行针对性的降低，以此提高算法的效率与可实现性。

综上所述，基于 Ambisonics 的可听化重放系统在建筑声学研究与应用中具有巨大的潜力，但在实际系统实现及应用时，需要考虑扬声器规模与重放精度的取舍。评估重放虚拟声环境下的听觉感知情况，对指导实际系统构建与算法的改进具有重要意义。

1.2 空间感知

人类听觉系统在长期的演化中逐渐形成了对声音的定位能力。由于声波的波长远比光波更长，声音更容易发生衍射现象，相比光线，声音不容易被障碍物或人体自身的结构所遮挡。准确的声源定位能力，有助于有效地寻找目标，规避未知危险。

空间听觉能够对声源的位置，包括方位与距离两个层面进行感知，还能够对目标声源与周围环境交互产生的复杂声场形成一定的空间感知印象。人类的双耳空间听觉包含以下几方面的作用：（1）声源定位与分离；（2）听觉场景分析；（3）视觉-听觉多模态整合。双耳空间听觉是帮助人类在复杂的日常声学环境下获取信息的核心机制之一，包含了低层级生理处理与高层级的神经认知整合功能^[10]。

1.2.1 听觉系统

人耳的听觉系统是感知声音的生理基础，听觉系统包括外耳、中耳与内耳三个部分（如图 1-1 所示）^[11]。（1）外耳：外耳由耳廓与耳道组成，声波首先经过耳廓与头部的反射与干涉，叠加后进入耳道内，以近似一维传输的方式在耳道内传播至中耳。

(2) 中耳：中耳包括耳膜及其后的听骨链，声波在耳道内传输至鼓膜，激发鼓膜振动。鼓膜的振动带动中耳的听骨链（包含三块听小骨）振动，振动通过卵圆窗传播至内耳中的耳蜗。这里的听骨链起到了阻抗匹配的作用，把空气（低阻抗环境）中的振动转变为了淋巴液（高阻抗的环境）中的振动。（3）内耳：内耳主要包含耳蜗，是将声音从振动信号转化为电生理信号的生理机构，淋巴液的声振动引起耳蜗中基底膜不同区域（对应不同频率）的振动，其上的听觉内毛细胞按照振动的幅度发放电刺激脉冲，传输至高层神经系统，经过大脑处理后形成了听觉感知。

在空间定位和声重放系统的研究中，通常把中耳及中耳以后的部分看成一个“黑箱”，假设为：如果在耳道口或是鼓膜处激发特定的声场（刺激），听觉系统会根据刺激响应出一个相应的听觉事件。对于听力不正常的人群，其内耳系统与正常人的内耳响应不完全一样，对于同样的激励，正常人与听损者响应的听觉事件通常不一致。本文中的所有声重放研究，均局限在外耳接收的物理信号层面，所有感知实验均针对正常听力者展开，不考虑中耳和内耳对声信号的处理。

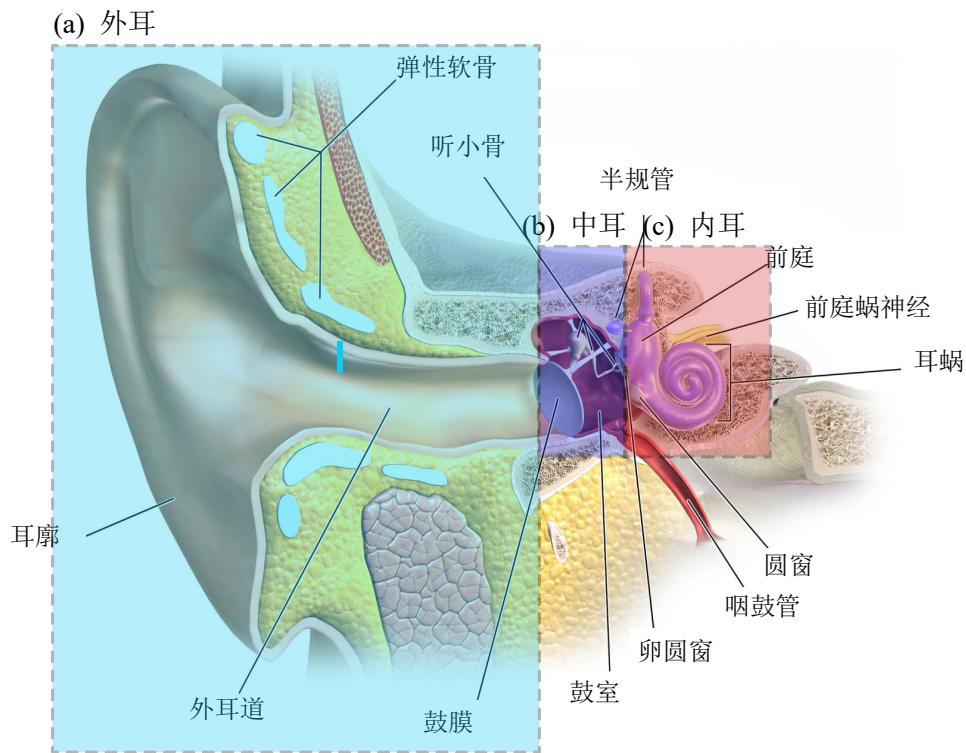


图 1-1 听觉系统截面图，修改自 WikiJournal of Medicine^[12]

Fig. 1-1 The cross-sectional picture of the auditory system, modified from WikiJournal of Medicine

1.2.2 空间听觉

人类听觉系统对声音的感知主要包含响度、音调、音色以及空间感四个属性^[13]。空间听觉是听觉系统的一个重要功能，它能够对声源的方向以及距离进行感知。对于水平面上的声源方向，在自由场单个声源的理想条件下，听觉定位的误差低于 3°^[7]。现在一般认为，在自由场下，听觉空间定位主要依赖以下四个物理线索：双耳时间差（Interaural time difference, ITD）线索、双耳声级差（Interaural level difference, ILD）、谱线索以及动态线索^[14]（ITD, ILD 及谱线索见图 1-2）：

(1) ITD 线索：不同方向的声音抵达两只耳朵的距离差不同，对应的到达时间不同，声音抵达左耳与右耳的时间之比称为 ITD。对于持续性信号，听觉依靠双耳信号的相位差（interaural phase difference, IPD）得到 ITD 信息。声音的相位以 2π 为周期，当声音的波长小于头部尺寸时，相位差可能超过 2π （由于周期性回到 0），导致 ITD 线索失效，因此 ITD 线索是一个低频线索，主要在 1.5 kHz 以下频率范围起作用。

(2) ILD 线索：由于头部对声音的遮挡作用，声音传至两耳的强度差别会随声源方向产生变化，这一定位线索称为 ILD，定义为左耳与右耳声压之比的分贝数。当声音的波长大于头部时，由于衍射作用，低频声音受人头遮挡的影响较小，因此 ILD 是一个中高频定位因素（大于 1.5 kHz）。同时，ILD 数值随频率变化波动较大，是一个宽频带的线索，心理声学实验表明窄带信号的 ILD 线索在空间定位中不可靠。

(3) 谱线索：ITD 和 ILD 线索只能用于区分信号的左右入射角度，从图 1-2 可以看到前后对称位置的 ITD 和 ILD 基本一致，仅通过 ITD 和 ILD 无法区别信号的前后与上下的方向差异，这一问题被称为混乱锥问题。但在实际中，人的听觉系统可以区分前后声音，也能够感受信号大致的高度（仰角）。现有研究认为，耳廓与躯干对声音的反射，对声音的高频部分有梳状滤波效应，这种高频谱线索（见图 1-2 (c)）是前后、垂直声源定位的重要线索。特别是 4 kHz 左右至 16 kHz 左右的前两个频谱峰谷 P1、P2、N1 与 N2，被认为是听觉系统判断声源仰角的核心线索^[15,16]。由于谱线索主要位于声音的高频部分，其具体数值受耳廓形状、躯体尺寸的影响较大，并且每个人的耳廓各有不同，因此谱线索是一种高度个性化的高频定位因素。

(3) 动态线索：上述三个线索即使在声源与头部相对位置不改变时，也可以提供一定的定位依据，被称为静态线索。研究表明，头部转动时，ITD 与 ILD 随着头部转到动态发生变化，这一动态因素也对声源方向定位有重要作用。当引入动态线索后，静态 ITD、ILD 无法区分前后及上下的混乱锥问题得以解决，定位准确度具有较大提升^[14]。

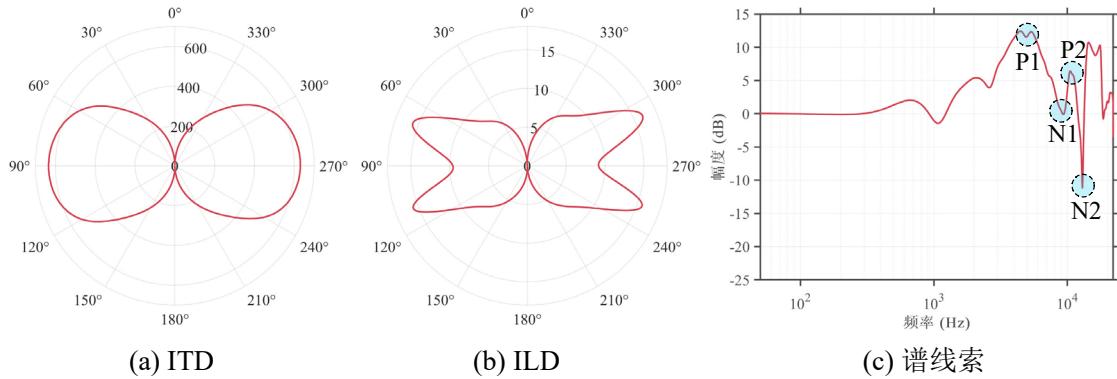


图 1-2 双耳定位线索：ITD、ILD 与谱线索（第一、第二频谱峰值 P1、P2；第一、第二频谱谷值 N1、N2）

Fig. 1-2 Binaural localization cues: ITD, ILD, and spectral cues (the first and second spectral peaks P1 and P2; the first and second spectral notches N1 and N2)

听觉系统还能感知声源的距离，相比方向定位，声源距离感知的精度更低，且受到众多因素（其中也包括非声学因素）的影响，听觉距离定位长期没有得到足够的关注^[17,18]。声源距离定位的物理线索可以分为三类，单耳线索、双耳线索以及动态线索：

(1) 单耳线索：单耳距离定位线索包括声音强度、直达混响比（Direct-to-reverberant ratio, DRR）以及谱线索。其中声音强度线索（声源强度随距离 r 按 $1/r$ 比例衰减）是自由场中最重要的距离定位线索^[19]，但声音强度同时与声源本身强度有关，因此声音强度是一个相对定位线索。在具有反射声的混响环境中，由于反射的干扰，声音强度随距离的变化减弱，但直达声与反射声的能量比例随距离变化，DRR 成为判断声源距离的重要因素^[20]。距离定位中的谱线索主要指，当声源非常接近头部时，由于声波波前为曲面，随着声源距离变近，低频部分相对于高频部分的能量增强更大^[21]，此外，头部对声音的反射作用由于衍射，只在相对高频的部分起作用^[22]。上述二者的共同作用，使得声源的高低频能量比随声源距离改变发生变化，这种变化能够作为声源距离的感知线索。

(2) 双耳线索：当声源距离改变时，声源抵达双耳的时间差几乎不随距离发生变化，ITD 对距离定位几乎没有贡献。而当声源离头部较近时，头部的阴影作用在不同声源距离下影响大小不同，导致 ILD 随距离变化。具体来说，与声源同侧的接收耳归一化声压随距离增大而减小，异侧耳的归一化声压在距离变大时，头部遮挡作用减弱，ILD 随着距离增大而增大^[23]。需要注意的是这里的声压为归一化声压，即双耳接收生压与头中心处声压（移开头部测量）的比值。双耳线索导致侧向声源的距离定位往往

要优于前方和后方，但不同研究中的实验对这种侧向距离定位准确性的提升还没有达成一个统一的结果^[24]。

(3) 动态线索：动态线索包括声源高速移动中产生的多普勒频移（Doppler frequency shift）现象^[25]；不同声源距离下，距离发生改变时强度的变化率改变产生的 acoustic tau 效应^[26]；以及动态视差效应^[27]，即在不同距离上，声源运动所产生的双耳线索变化不一样，这种随距离动态变化而产生的物理线索对距离定位有一定的贡献。

1.2.3 空间印象感知

相比理想的自由场环境，实际房间内的声学环境存在大量反射声，此时听觉系统不能对单个声源进行定位，但听觉系统能够在声源空间感知的基础上，对房间内的整体反射声环境形成更复杂听觉印象感知^[28]。对于重要的声学场所，如音乐厅或歌剧院等，房间的空间听觉效果是重要的考虑因素之一。Beranek 等人认为，音乐厅内最重要的感知因素包括：活跃度（liveness）、清晰度（clarity）、亲切感（intimacy）、空间感（spaciousness）、温暖感（warmth）和响度（loudness）^[29]。其中，受到声源方向影响感知维度主要是空间感，空间感是音乐厅主观感知最重要的维度之一^[30,31]。此类听觉印象具有很强的主观性，各个研究中对空间听觉印象的描述维度也不完全一致，对于空间听觉主观印象感知的研究仍处于比较初步的阶段。

目前的研究普遍认为，空间感主要包含两个属性：视在声源宽度（Apparent source width, ASW）以及听众包围感（Listener envelopment, LEV）^[32]。ASW 指声源在听感上比其实际呈现的更宽，LEV 指听众被声音环绕的感觉。这两种属性都被认为与侧向反射能量有关，其中早期反射对 ASW 感知的贡献更为显著，而 LEV 主要受后期反射的影响。ISO 3382-1 建议分别使用早期侧向能量比 J_{LF} 和后期侧向声能 L_J 来评估声源宽度 ASW 和包围感 LEV^[33]。其中， J_{LF} 的定义为：

$$J_{LF} = \frac{\int_{0.005}^{0.080} p_L^2(t) dt}{\int_0^{0.080} p^2(t) dt} \quad (1-1)$$

L_J 的定义为：

$$L_J = 10 \lg \left[\frac{\int_{0.080}^{\infty} p_L^2(t) dt}{\int_0^{\infty} p_{10}^2(t) dt} \right] \quad (1-2)$$

其中, p_L 和 p 分别表示在房间中使用八字形传声器和无指向传声器测量的声压; p_{10} 表示在消声室中距离声源 10 米处测量的声压。此外, 也有研究表明双耳间互相关系数 (Interaural cross-correlation coefficient, IACC) 与 ASW 和 LEV 具有显著的相关性^[34]。

1.3 空间声重放

1.3.1 坐标系

参考坐标系是空间声研究的基础, 本文后续内容中, 若如无特殊说明, 参考坐标系均为球坐标, 球坐标参考系如图 1-3 所示, 该坐标系定义与大部分空间声相关研究中采用的球坐标系定义相同。

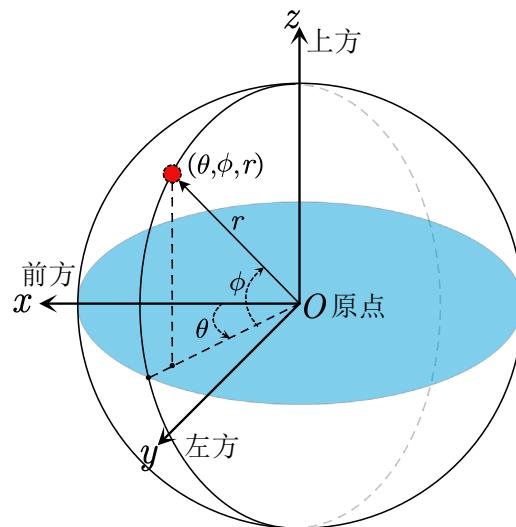


图 1-3 球坐标系

Fig. 1-3 The spherical coordinate system

球坐标系上任意一点的位置用坐标 (θ, ϕ, r) 表示, 其中: θ 为方位角, 范围为 $\theta \in [0^\circ, 360^\circ]$, 0° 方位角指向正前方, 90° 方位角指向左侧; ϕ 为仰角, 范围为 $\phi \in [-90^\circ, 90^\circ]$, 0° 为水平面, 90° 仰角为正上方; r 表示空间点到坐标系原点的距离, $r \geq 0$ 。本文中, 如无特殊说明, 听者头部朝向 x 轴正方向, 左耳方向为 y 轴正方向。该球坐标系下的坐标与笛卡尔坐标系下的坐标 (x, y, z) 转换公式如下:

$$\begin{cases} x = r \cos(\theta) \cos(\phi) \\ y = r \sin(\theta) \cos(\phi) \\ z = r \sin(\phi) \end{cases} \quad (1-3)$$

1.3.2 声音

声音通常有两个不同含义，一是物理上的定义，描述物理振动以波的形式在媒质中的传播，二是该物理振动通过人类听觉系统所激发的听觉事件^[35]。声音的这种物理与心理的双重属性，是空间声重放的实现与优化的基础。

声音由物体振动引起的空气密度变化而产生，是一种压力波，用声压 p 进行描述，单位为 Pascal (Pa)。声压的定义为媒质压力相对于媒质静态压力的改变量，由于声压的变化幅度非常大，通常采用对数计量的声压级 (sound pressure level, SPL) 描述声压大小，计算方法为：

$$SPL = 20 \log_{10} \left(\frac{p}{p_0} \right) \quad (1-4)$$

其中 p_0 为参考声压级，对应人耳在空气中对 1 kHz 纯音的最小可察觉阈值，该数值在标准室温空气的条件下为 2×10^{-5} Pa。

声源在空间中任意处产生的声压是时间和空间的函数，可以通过求解波动方程获得。在最简单的情况下，声波不受到任何界面的反射与吸收，这种条件称为自由场 (free field) 条件。自由场中位于 Ω_s 处的单频无体积球形声源 (点声源) 在空间中任意位置 Ω 处激发的声场为：

$$P_s(\Omega_s, \Omega, k) = \frac{e^{-jk|\vec{r}_s - \vec{r}|}}{4\pi|\vec{r}_s - \vec{r}|} \quad (1-5)$$

其中， \vec{r}_s, \vec{r} 分别为声源，接受点处的位置矢量 (笛卡尔坐标系)； $k = 2\pi f/c$, $\omega = 2\pi f$ 分别为波数与角频率， c 为声速， f 为频率。

当声源距离接收点较远，即 $|\vec{r}_s| \gg |\vec{r}|$ 时，式(1-5)可以用平面波辐射函数近似：

$$P_s(\Omega_s, \Omega, k) = e^{-j\vec{k}(\vec{r}_s - \vec{r})} \quad (1-6)$$

其中， \vec{k} 为波数矢量， $\vec{k} = k\vec{r}_s$ ，可以看到，平面波模型中，声压强度不随距离发生改变，且其相位只受接收点与声源点的法向距离影响，相比平面波更简单。本文后续的研究中主要采用平面波模型作为假设模型。

1.3.3 双耳重放

随着个人耳机的市场发展与普及，双耳空间声重放凭借其低成本与高精度的特点，成为目前最广泛使用的技术之一。双耳空间声重放的基本原理为，对于自由场中位于任意方向 $\Omega_s(\theta_s, \phi_s, r_s)$ 的理想点声源 S ，该声源到耳道口的传输过程可以用头相关传输函数（Head-related transfer function, HRTF）进行描述：

$$\begin{cases} H_L(\Omega_s, f, a) = \frac{P_L(\Omega_s, f, a)}{P_0(r, f)} \\ H_R(\Omega_s, f, a) = \frac{P_R(\Omega_s, f, a)}{P_0(r, f)} \end{cases} \quad (1-7)$$

其中， P_L, P_R, P_0 分别为该声源传输到左耳、右耳与头中心（移开头部时），在频率 f 处激发的频域声压。 H_L, H_R 描述了声源从不同位置传输到左右耳接收位置，受到人体结构影响后的响应。 a 为描述人体结构尺寸的参数集。若欲重放位于方向 Ω_s 的干信号 $x(t)$ ，则耳机的左右通路馈给信号 $y(t)$ 为：

$$y_\alpha(t) = h_\alpha(t) * x(t), \alpha = L, R \quad (1-8)$$

其中， $h(t)$ 为 HRTF 的等效时域形式，即头相关脉冲响应（Head-related impulse response, HRIR），通过傅里叶逆变换计算得到。

HRTF 通常采用封闭耳道法进行测量，即将接收点设置在封闭的耳道口处完成不同方向的 HRTF 测量^[36]。实际使用 HRTF 进行双耳重放或计算双耳分析时，为了获得更好的重放效果或更准确的结果，需要根据实际接受位置，并结合不同的耳机传输特性进行均衡处理。由于声音在耳道内的传输可以近似为一维传输，可以通过简单的阻抗变化进行匹配，此外，还需要使用耳机传输函数（Headphone transfer function, HPTF）对耳机发声单元本身的频率响应进行均衡^[37]，上述均衡过程的频域补偿系数 G 为^[15]：

$$G = \frac{1}{M \cdot HPTF} \cdot \frac{Z_{earcanal} + Z_{headphone}}{Z_{earcanal} + Z_{radiation}} \quad (1-9)$$

其中, M , $Z_{earcanal}$, $Z_{headphone}$, $Z_{radiation}$ 分别为传声器的频率响应 (通常是 1), 耳道阻抗, 耳机阻抗与耳道处的自由场辐射阻抗, $HPTF$ 是在封闭耳道口测量得到的耳机频率响应, 按上述流程均衡后的结果可以近似认为是鼓膜处的接收信号。

可以看到, 一旦获得了按空间分布的 HRTF, 即可使用耳机还原不同方向声音传输到双耳的过程, 实现双耳空间声重放。但 HRTF 与人体的生理结构有关, 因人而异, 实际使用中难以获取每个个体对应的 HRTF。使用非个性化的 HRTF 进行双耳重放会导致定位性能的下降, 容易产生前后混淆、头中心定位等问题。如何有效地获取高精度的 HRTF 数据是双耳重放的核心, 本文的前期研究提出了一种基于生理结构匹配的算法进行 HRTF 的个性化定制^[38]。此外, HRTF 直接反映了不同位置声源在双耳接收点产生的信号, 因此也往往被用于空间声重放系统以及相关听觉研究的模拟分析之中。

1.3.4 多通路环绕声

双耳信号通常只能通过耳机进行重放, 在实际使用中存在一定的限制。使用扬声器作为播放设备实现空间声重放, 需要选用针对扬声器设计的多通路重放算法。基于心理声学原理设计的多通路环绕声重放系统是最早发展起来的, 也是目前应用最广泛的技术之一, 包括但不限于双通路环绕声、5.1 通路环绕声、以及更大型的 22.2 通路环绕声系统等等。此类系统往往使用“基于通路”的控制方法, 通过预先调节好的多通路信号增益与延时, 馈给对应的多通路扬声器, 产生虚拟声像。

双通路立体声 (stereo) 采用布置在听者前方扇形区域的一对扬声器, 坐标分别为 $(-\theta_0, 0^\circ)$ 与 $(\theta_0, 0^\circ)$, 重放方位角位于 $-\theta_0 \sim \theta_0$ 内的虚拟声源。当欲重放声源位于两扬声器位置时, 对应位置上的扬声器直接发声即可重建目标方向的空间声源; 当欲重放声源位置位于两扬声器之间时, 可以通过调整两通路信号幅度, 让听众感觉到虚拟声源位于两扬声器之间的某个位置, 这种调整信号幅度以实现不同虚拟源方向重放的方法称为调幅 (panning) 方法。

在双通路立体声的基础上, 将多通路扬声器扩展为全空间分布, 每三个扬声器构成一个重放扬声器对, 重放界内的虚拟声像, 进而实现整个三维空间的虚拟声源重放。这种空间声重放系统称为矢量幅度调整 (Vector base amplitude panning, VBAP) 重放系统。对于一组 (三个) 重放扬声器, 若扬声器 l 位于 $\Omega_l(\theta_l, \phi_l, r_l)$, 单位方向矢量为 \vec{r}_l (笛卡尔坐标), $\vec{r}_l = [\cos(\theta_s)\cos(\phi_s), \sin(\theta_s)\cos(\phi_s), \sin(\phi_s)]^T$, T 为矩阵转置符号。

欲重放虚拟源的方向矢量为 \vec{r}_s ，通过调整三个扬声器对应的增益 b_l ，实现对目标位置声源的空间重放，各通道增益通过矩阵求逆得到^[39]：

$$\begin{cases} \vec{r}_s = \sum_{l=1}^3 b_l \vec{r}_l \\ [b_1, b_2, b_3]^T = [\vec{r}_1, \vec{r}_2, \vec{r}_3]^{-1} \vec{r}_s \end{cases} \quad (1-10)$$

上式中⁻¹为矩阵求逆运算，为保持重放能量不随方向变化，通常还需对增益进行归一化处理，即 $\hat{b}_l = \frac{b_l}{||(b_1, b_2, b_3)||}$ 。

传统多通路环绕声重放的优点在于，算法简单容易实现，且扬声器的数量以及摆放位置相对灵活。其缺点在于，算法的所依赖的心理声学基础均为速度矢量定位原理，仅能够准确还原低频的 ITD 线索^[40]，定位准确度较低。算法均假设受试者位于预设的单个位置，当受试者偏离中心时信号偏离假设，重放的有效收听区域（甜区）较小。

1.3.5 Ambisonics

Ambisonics 是一类基于物理声场重构的空间声重放系统，区别于多通路环绕声系统，物理声场重构系统的目标为：通过调整多个扬声器的信号，直接在一定区域内重建目标声源产生的声场。此类重放系统通常具有更准确的空间定位效果与更大的甜区，但声场重放往往需要大量的扬声器。

Ambisonics 系统的基本原理为，将由点声源或平面波声源构成的空间声场，使用其在球坐标系统下的本征函数（球谐函数）进行分解，然后使用多个扬声器合成的重构声场逼近分解后的球谐域声场（详细原理见 2.2 节）。Gerzon 最早提出了一阶 Ambisonics（First-order Ambisonics，FOA），用以重放声源的高度信息（也被称为 Periphery）^[41]。FOA 系统仅使用四个扬声器进行全空间的虚拟声源重放，声场重构精度有限，定位效果一般，FOA 系统更多地被认为是 M/S 立体声拾音技术的一种三维扩展。Daniel 等人^[42]以及 Jot 等人^[43]的研究进一步完善了 Ambisonics 系统的理论，将 FOA 扩展至更高的阶数，此类系统称为高阶 Ambisonics 系统（Higher-order Ambisonics，HOA）。HOA 重放使用更多的扬声器，具有更高的声场重构范围与精度，同时也具有更高的算法复杂度，近年来 HOA 得到了广泛的关注。

在 1.3.3 与 1.3.4 节所述的双耳重放与多通路环绕声重放中，馈给信号在设计之初就需要考虑重放方法与扬声器的布置，同一组信号不能适配不同的重放系统。Ambisonics 的一个优势是分离了信号的编码与解码过程，首先使用球谐函数对声场进行编码，重放时再根据扬声器阵列的布置进行解码，生成多通道信号，极大地增加了空间声重放系统的灵活性。通过使用 HRTF 对扬声器进行虚拟化，也能实现使用耳机进行双耳 Ambisonics 空间声重放。Ambisonics 系统的主要缺陷是，低阶系统的重放精度较低，高阶系统对扬声器数量的要求较高，且系统要求扬声器在空间上均匀分布，这使得实际场景下，难以构建具有较好听音效果的 Ambisonics 系统。

Ambisonics 重放相比传统多通路声重放的另一个优势是，既可以预设声源的方向，对目标声信号进行特定方向的 Ambisonics 虚拟合成，也能够对已有的声场进行三维空间声录制，或是测量该环境的三维声场信息。相比传统信号拾拾，Ambisonics 信号拾拾在后期信号处理方面较为复杂，但其配合 Ambisonics 重放系统可以获得较好的空间重放效果，在一些场景下可以避免传统信号拾拾时，需要在多点布置传声器的问题，同时可以克服双耳录制、M/S 录制等三维空间声录制方法中，录制的信号只能使用对应重放装置进行重放的局限性^[44]，具有广阔的应用前景。此方法可以用于录制音乐厅的空间脉冲响应，用于主观听音评价^[45]。由于 Ambisonics 拾拾能够以一种紧凑可移动的设备拾拾空间声场，在声景（soundscape）学的研究中也受到大量的关注^[46-48]。

1.3.6 波场合成

波场合成（Wave field synthesis, WFS）是另一种声场重构技术，基于惠更斯菲涅尔（Huygens-Fresnel）理论实现对物理声场的重构。声源激发的声场可以由波阵面上的连续次级声源完全重构，次级声源的馈给信号为目标初级声源在次级声源处激发的声场。根据基尔霍夫-亥姆霍兹方程，闭合曲面 D 包含的闭合空间内的声场可以由曲面 D 上的声压与梯度完全决定^[49]：

$$P(\Omega, f) = - \iint_D \left[\frac{\partial P(\Omega_l, f)}{\partial \vec{n}_l} G(\Omega, \Omega_l, f) - P(\Omega_l, f) \frac{\partial G(\Omega, \Omega_l, f)}{\partial \vec{n}_l} \right] dD(l) \quad (1-11)$$

其中， l 代表曲面 D 上的点， \vec{n}_l 为 l 处的单位内法向矢量， $G(\Omega, \Omega_l, f)$ 为次级声源格林函数。上式表示，如果能在闭合曲面 D 上布置均匀连续的压力传感器（即无指向传声器，积分内第二项）和速度传感器（积分内第一项）拾拾原始声场，在重放时，分别把上述信号馈给曲面对应位置处的均匀连续无指向次级声源与偶极子次级声源，就能

完全重构曲面内的声场。在实际的声音拾拾或重放时，传声器与扬声器只能离散布置，这会导致空间混叠现象，实际的有效重放频率与次级声源的间隔大小 Δl 相关，当声源垂直次级声源曲面入射时，频率上限 $f_{max} = c/2\Delta l$ 。当欲重放上限频率为3 kHz的信号时，次级声源间隔不能超过5.6 cm，远小于一般的扬声器单元尺寸，这在实际重放中往往难以实现。

相比 Ambisonics 重放，WFS 的优势是具有更大的听音区域，但需要的扬声器数量非常多，有效重放频率较低。当需要重放在三维空间中分布的声源时，所需次级声源数量为二维重放的平方倍，这在现实应用中几乎不可能实现，因此，目前对 WFS 系统的研究主要集中在水平面空间声重放^[50]。由于室内声场中反射声可能来自于空间中的任意方向，包括水平面方位角信息与垂直方向仰角信息，WFS 系统的在重放维度上的缺陷，使其目前暂时不适用于室内可听化重放，因此，本研究采用 Ambisonics 系统作为主要的声重放手段。

1.4 研究现状

1.4.1 Ambisonics 重放算法改进

如 1.3.5 节所述，Ambisonics 重放系统的两个主要问题是：（1）低阶系统在重放时的有效频率较低，定位性能差；（2）系统要求扬声器的空间上均匀布置，这在实际系统中难以实现。目前对 Ambisonics 系统的优化改进方向主要集中在上述两个方面。

低阶 Ambisonics 重放特定方向声源时，周围扬声器也具有较大的馈给能量，特别是有大量的反相能量，容易造成声像不稳定。Daniel 最早提出最大化能量定位矢量的 $\text{max-}r_E$ 优化及避免反向信号的 *in-phase* 优化^[42]，以提高低阶系统的听感及定位性能（见 2.2.4 节）。龚惠哲在 $\text{max-}r_E$ 优化的基础上，使用优化算法进一步调整各扬声器增益，使重放的响度、幅度、速度矢量、能量矢量的综合指标达到最优值^[51]。在准确重构的空间范围外，传声器阵列对重构声场的测量显示， $\text{max-}r_E$ 重放的重构声场误差相比普通 Ambisonics 重放更低^[52]，但在有效重放区域内， $\text{max-}r_E$ 和 *in-phase* 都会降低声场的重构精度^[53]。

另一方面，当扬声器在空间上分布不均匀时，Ambisonics 重放容易出现“病态解码”的问题，部分扬声器会出现较大的增益，极大地降低重放质量（见 2.2.3 节）。Zotter 等人提出结合 VBAP 和 Ambisonics 的 AllRAD（All-Round Ambisonics Decoding）算法以降低 Ambisonics 重放对扬声器空间均匀分布的要求^[54,55]，但该算法同时也引入

了 VBAP 算法的缺陷，一定程度上破坏了 Ambisonics 的物理重构特性，可能会降低定位性能。另一种 Energy-preserving Ambisonics decoding（EPAD）算法通过使用奇异值分解（Singular value decomposition, SVD）的方法去除解码矩阵的奇异值，提高解码矩阵的稳定性，降低了重放对扬声器阵列均匀性的要求，特别是，通过重新构造解码矩阵（相当于改变基函数），在半球形阵列上也能实现稳定的解码^[56]。但该方法破坏了 Ambisonics 解码的物理基础，在均匀阵列中，方向性精度略低于模式匹配法，且其实际听音的效果尚不明确。此外，还有混合阶 Ambisonics 算法，此类算法在三维重放的基础上选择性引入更多高阶谐波，能够在保持三维空间重放精度的前提下，提高水平面的空间分辨率^[57]。

1.4.2 Ambisonics 声重放的主观评价

空间声重放系统的实际重放效果与重建声场的准确度相关，但听觉感知是非线性的，二者通常并不完全一致。例如，Ambisonics 系统的实际有效听音甜区往往要大于理论的有效区域^[58]。为了验证与优化空间声重放系统的效果，需要通过主观听音实验对重放算法效果进行评估。对声重放系统的主观评价通常包含两个层次：（1）反映听觉能力的评价，声源定位准确度、音色等；（2）高层感知印象评价，如声源宽度、沉浸感、真实感、与偏好度等等。

1.4.2.1 声源定位

Ambisonics 系统的有效重放频率上限和甜区大小与其阶数直接相关，随着重放阶数的提高，听众的定位准确度也会提高。FOA 系统的定位准确度较低，定位实验表明 FOA 系统的定位准确性甚至低于 VBAP 系统^[59]。江建亮对比了不同阶数双耳 Ambisonics 重放系统下的定位准确度，发现在不考虑动态头动因素的情况下，5 阶 Ambisonics 已经可以获得接近双耳重放的定位效果，在有头动因素的情况下，则需要高至 18 阶的系统才能获得与双耳重放接近的定位效果^[60]。Frank 等人使用扬声器进行 1、3 与 5 阶 Ambisonics 重放，定位实验表明，随着阶数上升，定位准确度与前后混淆概率均有明显改善^[61]。在对更高阶 Ambisonics 系统的测试中，Barrett 发现 7 阶与 9 阶 Ambisonics 的重放差异几乎不能察觉，但该实验不是一个严格的定位实验^[62]。

Ambisonics 重放中，准确重放的频率范围随着听众至听音中心的距离增大而减小。Stitt 等人的主观定位实验发现，在使用 FOA 进行重放时，听众可能会感受到与实际重放位置相反的虚拟声源^[63]。对于非中心区域听音，虚拟声源容易偏向离听众最近的扬

声器的方向^[61,64]。也有研究发现，尽管理论上 Ambisonics 的准确重放区域较小，但在实际听音实验中，受试者可以在较大的范围内感受到大致正确的虚拟声源方向^[58]。对于不同的优化解码方案，Frank 等人对比了水平面不同重放阶数下， $\max-r_E$ 和 *in-phase* 优化对 Ambisonics 重放系统定位性能的影响，发现即使在 5 阶系统下， $\max-r_E$ 相比无优化的 Ambisonics 仍有一定的定位准确度提升，但 *in-phase* 反而造成定位准确度下降。

1.4.2.2 声音失真与空间印象

Ambisonics 重构声场的误差可能会造成失真问题，提高 Ambisonics 的阶数或靠近听音中心，可以在一定程度上降低重放的声音失真^[65,66]。在一定阶数的 Ambisonics 重放中，有研究表明，使用多于阶数对应的最低扬声器数量的扬声器单元，会引入更强的梳状滤波效应，造成更大的信号频谱失真，进而可能会增大重放信号的失真程度^[67,68]。

对于更复杂的听觉感知印象，例如，声源宽度、空间包围感、混响感及声景真实感等等，这些听觉感知属性同样受 Ambisonics 系统的重放准确度影响。Thery 等人对比了双耳可听化与 Ambisonics 可听化重放信号的空间感，结果表明 Ambisonics 在声源宽度和包围感上与双耳重放存在显著的差别。Bargum 等人的实验对比了 Ambisonics 录制的室内可听化信号与模拟完成的可听化重放信号，实验表明录制的信号混响感更强，虚拟的信号的感知声源宽度远大于录制的信号^[69]。本文的初期研究也显示，相比传统双耳重放，一阶 Ambisonics 捡拾与重放在声景感知上的真实感与方向感更差^[70]。

1.4.3 Ambisonics 系统在建筑声学中的应用

目前，国内外已有一些课题组建立了基于 Ambisonics 物理场重放的可听化系统。Neal 建立了基于虚声源模拟和 3 阶 Ambisonics 重放的可听化系统，并探索了虚拟声场环境下的空间感感知问题^[71]；Favrot 与 Buchholz 等人提出基于扬声器的房间可听化（Loudspeaker-Based Room Auralization，LoRA）系统，该系统借助 CATT 或 Odeon 等几何声学软件对房间声场进行模拟，然后对反射进行空间声重放，通过客观测量检验了重放声场的房间声学参数^[72,73]。国内的龚惠哲也采用类似的方法建立了 3 阶 Ambisonics 重放系统，通过使用 Odeon 软件计算得到的三阶 Ambisonics 信号，实现室内声场可听化^[51]。

上述系统的声场模拟均使用离线计算的方式，无法根据需要实时更新，近年来，实时可听化技术备受关注。Noisternig 等人基于 EVERTims 实时波束追踪软件，实现了

音乐厅声场的实时可听化，并使用 3 阶 Ambisonics 编码进行重放^[74]；Schröder 开发的 Virtual Acoustics (VA) 可听化框架，通过高速的虚声源法仿真与声线追踪算法，实现了实时声场模拟，同时该框架支持对模拟声场进行 Ambisonics 编码重放^[75]；Seeber 等人研发的模拟开放场环境重放 (Simulated Open-Field Environment, SOFE) 系统，使用虚声源法实现对任意形状房间的实时模拟，并支持双耳重放、最近邻扬声器映射、VBAP 和 Ambisonics 重放等多种音频渲染方法，同时该系统还整合了 CAVE 视频显示和运动跟踪功能，实现视觉-听觉虚拟场景重现^[76]。

近年来，Ambisonics 重放系统也被应用于声景声环境的研究之中，其核心价值在于通过传声器阵列实现三维声场的精准拾拾与重构，为户外声景的数字化记录、分析与虚拟再现提供了方便的技术手段^[47]。相比人工头静态的双耳录音，Ambisonics 直接对声场进行拾拾与重放，为实现交互性视听声景环境和装置提供了可能性^[77]。Davis 等人使用 1 阶 Ambisonics 录音，在半消声室通过 8 通道扬声器阵列重放合成声景，验证了其对真实声景的还原能力^[78]。类似地，Hong 等人进一步对比 1 阶 Ambisonics 捡拾重放的声景与真实的声景的感知，详细比较了不同 Ambisonics 重放方式下的感知差异^[46]。更进一步，他们的另一项结合使用全景摄像机实现了三维视觉-听觉重放^[79]，使用虚拟声景研究水声对交通噪声烦恼度的掩蔽作用。

1.4.4 现有研究的不足之处

综上所述，国内外课题组已建立了一定数量的 Ambisonics 重放系统并针对 Ambisonics 声重放下的一系列具体问题开展了一定的研究，现有研究在以下几个方面仍存在一些不足：

(1) 高阶 Ambisonics 重放阶数对声场重构误差及主观感知的影响：对于高阶 Ambisonics 的重放阶数对声场重构误差与主观感知影响的探讨，现有研究主要局限于研究重构误差对自由场噪声信号的定位性能及声音失真两个维度的影响。针对重放阶数对具体建筑声学应用场景下的高层感知，如语音信号感知问题，鲜有研究进行关注。仅有的相关研究中，只探讨了低阶系统下的语音感知问题，没有对高阶系统进行探索，同时缺少对重构误差的分析，无法指导实际重放系统的简化设计。

(2) 三维重放的算法改进：Ambisonics 系统将虚拟声场重放维度从二维空间重放扩展到三维空间重放，但三维 Ambisonics 重放需要大量均匀的非水平面扬声器，实际的扬声器阵列往往在空间中是分层非均匀分布的。目前已有的重放算法主要改进与优

化的目标是提高低阶系统下虚拟声源定位的稳定性，较少关注如何提高此类空间分层分布的扬声器阵列的重放效率。

(3) 已有的 Ambisonics 优化重放算法的对比与评估：针对非均匀扬声器的 Ambisonics 重放问题，已有研究提出了一些优化算法，如混合阶 Ambisonics 与 AllRAD Ambisonics 重放，不同的算法具有不同的假设基础与优化目标，但目前尚未有研究开展主观实验对它们进行系统地对比和验证，这些算法的具体性能表现尚处于未知状态。

(4) 室内声场反射声仰角信息的感知机理缺失：目前，使用 Ambisonics 系统实现室内声场可听化的相关研究主要聚焦在系统实现和算法优化上。高阶 Ambisonics 三维重放对扬声器数量与分布均有严苛要求，但人类听觉对于声源的仰角信息相比水平信息更不敏感，因此在重放时对高度信息进行一定程度的简化是可能的。但目前的研究较少关注在虚拟室内声场中，反射声仰角信息及其重放精度对感知的影响。

1.5 研究目的与研究内容

本文针对高阶 Ambisonics 可听化重放的优化改进与重放感知评估问题开展研究。研究目的为，在高阶 Ambisonics 可听化系统的基础上，通过主观感知评估实验，探究影响虚拟听觉场景真实性的因素，包括重放准确性以及听觉对具体重放内容的敏感性。结合主观感知的结果，对重放系统进行算法层面的优化，并开展听音实验对改进算法进行评估，在感知层面提高虚拟重放的精度，降低系统对实际硬件的需求。研究旨在实现高阶 Ambisonics 系统的改进与优化，最终推动其在建筑声学设计与研究中的落地应用。为了实现这一目标，本文开展多项相关研究：

(1) 建立了一套大型的高阶 Ambisonics 可听化重放系统，该系统实现了完整的 Ambisonics 信号处理流程，包括二维与三维空间声编解码以及声场景可听化合成，能够对多种声环境实现高阶重放。此外，还对系统进行了细致的测量与校准，以实现可靠的可听化重放。该系统是目前国内最大型的高阶 Ambisonics 可听化重放系统，是后续实验的研究平台，并作为重放系统优化的基准。

(2) 基于该系统，设计并开展客观模拟分析与主观听音实验，探讨重放阶数对声场重放误差的影响，以及重放误差与主观感知的关联。针对言语接受阈值这一测试任务，测量并确定了实现可靠的言语主观感知所需要的重放阶数。

(3) 为进一步提高重放精度, 进而保证虚拟声环境下的可靠感知结果。针对分层扬声器阵列的分布特点, 提出了一种分层 Ambisonics 重放算法, 该算法结合二维 Ambisonics 与 VBAP 技术, 降低了三维重放对扬声器阵列的要求。

(4) 进一步开展主观听音实验, 将分层 Ambisonics 重放算法与现有的混合阶 Ambisonics 与 AllRAD 算法在定位性能与声音保真度两个感知维度上进行了比较和评估, 确定了不同算法的具体性能表现, 明确了本文提出的分层 Ambisonics 算法的实际效果。

(5) 为结合听觉特性进一步简化系统, 使用该系统重放音乐厅室内声环境, 开展主观听音实验研究了重放反射声仰角范围、重放阶数及方法对虚拟声环境的空间感的影响。量化了仰角范围与能量扩散对听觉感知的影响, 明仰角方向信息与空间分辨率的冗余度与优化阈值。

1.6 研究思路与方法

本论文围绕高阶 Ambisonics 可听化技术优化及其感知评估问题开展研究, 论文结构与研究思路如图 1-4 所示, 具体而言, 本文的研究思路与方法包括:

本文首先在第一章介绍了研究的基本背景, 介绍了可听化重放系统这一研究课题下的基本概念与术语。初步地对该研究领域的发展情况进行了梳理, 总结了现有研究的不足, 针对这些研究缺陷, 提出本文的研究目的、思路与方法。然后, 在第二章详细阐明了高阶 Ambisonics 空间声重放的原理以及相关信号处理方法。在这一理论及算法的基础上建立了一套多通道高阶 Ambisonics 可听化系统, 对硬件进行了详细的校准, 结合 Odeon 仿真实现了自由场与室内声场的高精度重放。后续章节围绕该重放系统对一系列关键问题开展细致研究。进一步地, 对 Ambisonics 重放的关键重放参数——重放阶数, 在第三章进行了探讨。在客观层面, 分析了重放阶数对重放的误差影响, 在主观层面, 探讨了重放阶数对语音感知的影响。开展实验测量不同阶数下的言语接受阈值, 并分析阶数对重构误差的影响与主观感知结果的关联性。针对高阶重放在实际中难以实现的问题, 第四章针对分层扬声器阵列特点, 提出结合二维 Ambisonics 与 VBAP 的分层重放算法, 以降低三维重放对扬声器数量与空间均匀性要求, 并开展仿真模拟从客观层面探讨算法的有效性。进一步地, 在第五章开展主观听音实验, 将该算法与已有的几类混合阶 Ambisonics 及 AllRAD 算法进行横向对比, 评估算法合成的简单自由场虚拟声源的定位性能与声音失真情况。第六章进一步使用该系统重放虚拟厅堂的声环境, 探讨反射仰角信息重放与高分辨率重放对可听化重放的必要性, 开展两

个主观实验分别探讨重放反射的仰角范围与重放反射的能量扩散对空间感感知的影响。论文的最后归纳了本文的主要结论与发现，总结了全文的主要创新点，并对后续研究进行了展望。

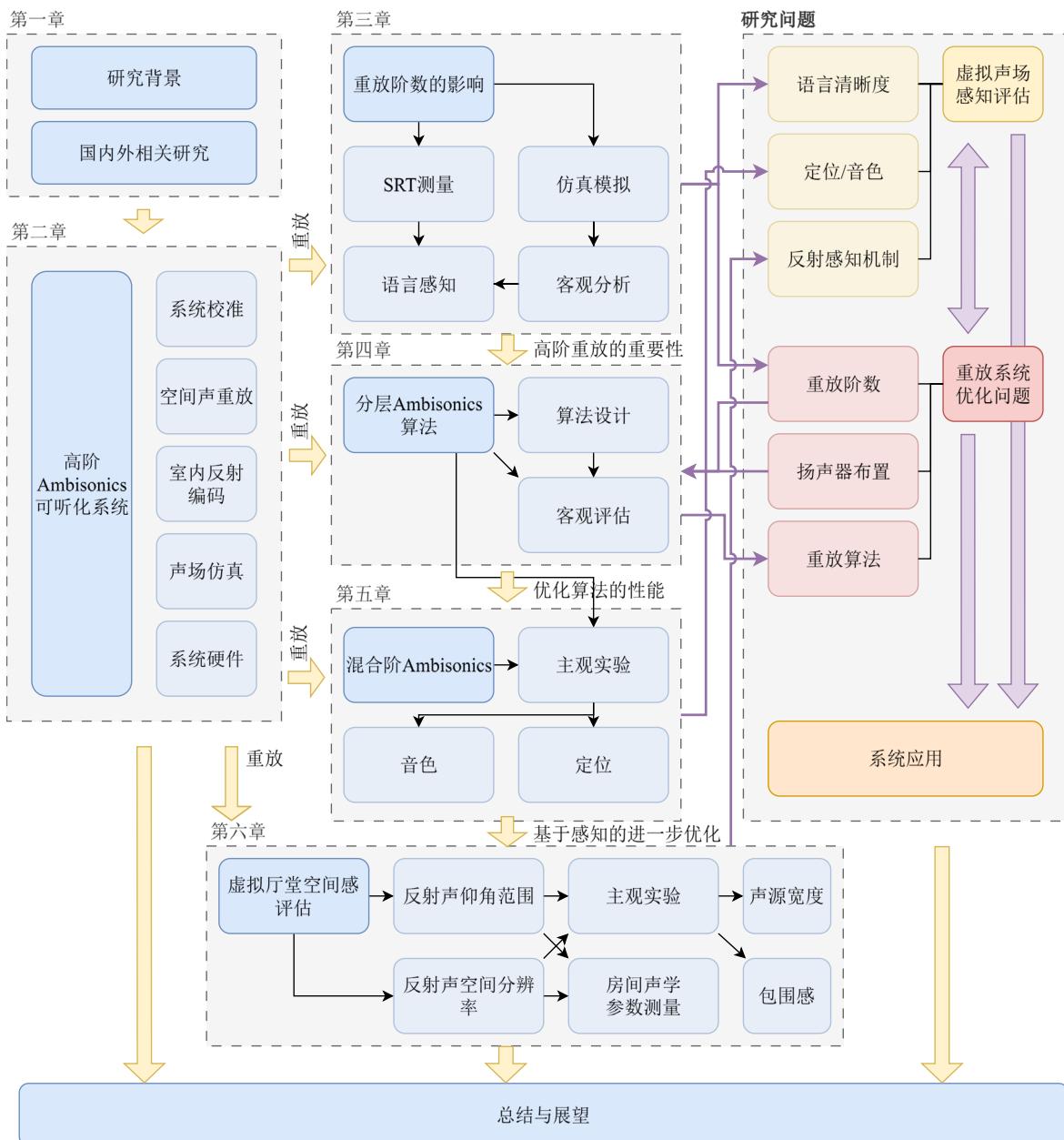


图 1-4 论文结构图

Fig. 1-4 The block diagram of the thesis

第二章 高阶 Ambisonics 可听化系统的原理与设计

2.1 引言

高阶 Ambisonics 重放是一种物理声场重构技术，基于声场球谐函数分解理论，将声场进行分解，然后使用大量扬声器重建目标声场，将目标声场的分解系数与重构声场的分解系数进行逐级匹配。基于高阶 Ambisonics 重放技术，可以实现空间任意位置声源的重放，进一步结合可听化技术，通过仿真或测量得到房间反射信息，对复杂室内声场中的所有声源与反射声进行空间编码，进而实现复杂声环境的可听化重放。根据重放阶数，Ambisonics 系统可以分为一阶 Ambisonics 重放（FOA）与高阶 Ambisonics 重放（HOA）两类；根据重放声场的空间维度，可以分为水平面二维 Ambisonics（2-D Ambisonics）与三维 Ambisonics（3-D Ambisonics）。此外，还有在 3-D Ambisonics 基础上增强水平面分辨率的混合阶 Ambisonics^[68]，考虑近场曲面波前传播特性的近场补偿 Ambisonics（Near-field compensation Ambisonics，NFC-Ambisonics）^[80]等等。

本章对 Ambisonics 的重放、优化以及拾音等信号处理进行了全面的论述，基于该理论，首先建立了一个多通道 Ambisonics 可听化空间声重放系统，用于实现高精度的自由场声源、室内混响声环境等空间的重放。该系统包含了信号生成、仿真、空间化处理等步骤，实现了全链路的可听化 Ambisonics 高阶重放，是后续研究的基础。

本章 2.2 节首先对高阶 Ambisonics 的基本理论进行介绍，包括声场分解理论与 Ambisonics 重放以及拾音理论。2.3 节基于 Ambisonics 球谐函数理论，介绍了高阶 Ambisonics 可听化系统硬件与软件部分的实现与校准。最后，2.4 节对本章内容进行了总结。

2.2 高阶 Ambisonics 的基本原理

2.2.1 球谐函数

对于随时间变化的声音信号，可以使用傅里叶变换将信号分解为不同频率正弦与余弦信号的组合，实现对信号的分析与频域处理。类似地，对于随空间变化的声信号，可以使用球谐波变换将信号分解为不同阶数球谐函数的组合，进而实现对声场信号的分析与后期处理，然后重新组合这些谐波函数分量恢复空间声场信号。

根据声波波动方程，任意声场可以表示为内部声源向外辐射的声场（Exterior field）和外部声源向内辐射的声场（Interior field）的组合叠加。亥姆霍兹方程在球坐标系下

的本征函数为球谐函数，可以使用球谐函数进行分解展开，在球坐标系下，坐标 $\Omega(\theta, \phi, r)$ 处接收点的频域声压 P 可以表示为^[81]:

$$P(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n B_n^m(k) j_n(kr) Y_n^m(\theta, \phi) + \sum_{n=0}^{\infty} \sum_{m=-n}^n A_n^m(k) h_n(kr) Y_n^m(\theta, \phi) \quad (2-1)$$

上式第一项和第二项分别代表外部声源和内部声源的辐射场，对于声重放系统，声源通常位于声场外部，因此仅考虑外部声源的辐射场，即 $A_n^m = 0$ ； j_n, h_n, k 分别为第一类球贝塞尔函数，第一类球汉克尔函数和波数； Y_n^m 是球谐函数，定义如下所示：

$$\begin{cases} Y_n^m(\theta, \phi) = N_n^{|m|} P_n^{|m|}(\sin\phi) \frac{1}{\sqrt{2\pi}} \begin{cases} \sqrt{2} \sin(|m|\theta), m < 0 \\ 1, m = 0 \\ \sqrt{2} \cos(m\theta), m > 0 \end{cases} \\ N_n^{|m|} = \sqrt{\frac{2n+1}{2} \frac{(n-|m|)!}{(n+|m|)!}} \\ P_n^m(x) = \frac{(1-x^2)^{m/2}}{2^n n!} \frac{d^{n+m}}{dx^{n+m}} (x^2 - 1)^n \end{cases} \quad (2-2)$$

$m, n \rightarrow \mathbb{N}, m \in [-n, n], n \in [0, \infty)$

其中， P_n^m 为不包含 Condon-Shortley phase 因子 $(-1)^m$ 的缩合勒让德多项式 (Associated Legendre polynomials)。 $N_n^{|m|}$ 为球谐函数的归一化系数，上述球谐函数被称为 N3D 归一化形式的球谐函数^[82]，具有正交归一化性质：

$$\begin{cases} \int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} Y_n^m Y_{n'}^{m'*} d\Omega = \delta_{nn'} \delta_{mm'} \\ \int |Y_n^m|^2 d\Omega = 1 \end{cases} \quad (2-3)$$

$\Omega \rightarrow \mathbb{S}^2$

上式中 δ_{ij} 为冲击响应函数，仅当 $i = j$ 时， $\delta = 1$ ，其他情况下 $\delta = 0$ 。0 到 3 阶球谐函数图像如图 2-1 所示，由图像可以发现，不同阶数的球谐函数具有不同的指向性，阶数越

高，指向性越尖锐。此外，注意到式(2-2)定义的 N3D 归一化系数 $N_n^{|m|}$ 包含 $\sqrt{2n+1}$ 因子，该因子数值随着阶数 n 提高而增大，使得高阶球谐函数项具有更高的幅值，在实际的重放系统的信号处理过程中，有可能会导致信号幅值超过系统所允许的上限。实际使用中可以移除 $\sqrt{2n+1}$ 项，这种归一化方式的球谐函数被称为 SN3D 球谐函数^[82]：

$$N_n^{|m|}_{SN3D} = \sqrt{\frac{1}{2} \frac{(n - |m|)!}{(n + |m|)!}} \quad (2-4)$$

此时球谐函数不具有归一化性质，需要在具体使用时补回 $\sqrt{2n+1}$ ，为便于叙述，如无特殊说明，本文中的球谐函数均采用 N3D 归一化的形式。

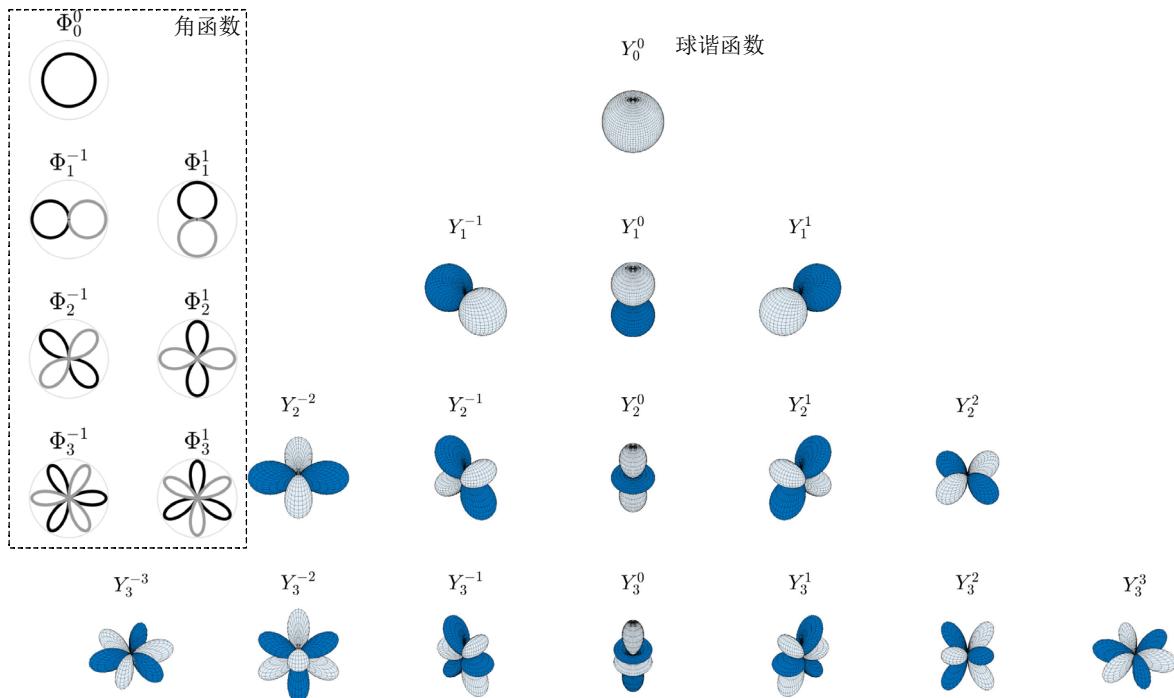


图 2-1 0 至 3 阶球谐函数及角谐波函数的图像

Fig. 2-1 The graphs of spherical harmonics and angular harmonics from order 0 to 3

2.2.2 声场的球谐函数分解

由于球谐函数具有正交性，可对式(2-1)进行逆变化，乘上球谐函数的共轭项，并在球面进行积分，可以得到：

$$B_n^m = \frac{1}{j_n(kr)} \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} P(r, \theta, \phi, k) Y_n^m(\theta, \phi)^* \cos(\phi) d\theta d\phi \quad (2-5)$$

其中， Y_n^{m*} 为 Y_n^m 的共轭对，本文中采用实数形式的球谐函数，球谐函数定义已包含共轭对函数，因此 $Y_n^{m*} = Y_n^m$ 。上式与式(2-1)为声场的球谐变化 (Spherical harmonics transfer, SHT) [83]，这两个公式说明，空间声场可以由一组系数 B_n^m 完全表示。

分别考虑两种形式的常见声波的波场：平面波与球面波，其函数分别如式(1-6)与式(1-5)所示，带入式(2-5)后得到^[84]，对于平面波：

$$\begin{cases} P_s(\Omega, \Omega_s, k) = e^{(-j\vec{k} \cdot \vec{r})} = 4\pi \sum_{n=0}^N \sum_{m=-n}^n j^n j_n(kr) Y_n^m(\Omega_s) Y_n^m(\Omega) \\ B_n^m = 4\pi j^n Y_n^m(\Omega_s) \end{cases} \quad (2-6)$$

对于球面波：

$$\begin{cases} P_s(\Omega, \Omega_s, k) = \frac{e^{-jk|\vec{r}_s - \vec{r}|}}{4\pi |\vec{r}_s - \vec{r}|} \\ = \sum_{n=0}^N \sum_{m=-n}^n j j_n(kr) k h_n^{(2)}(kr_s) Y_n^m(\Omega_s) Y_n^m(\Omega) \\ B_n^m = -j k h_n^{(2)}(kr_s) Y_n^m(\Omega_s) \end{cases} \quad (2-7)$$

其中， j 为虚数单位， $h_n^{(2)}$ 为第二类球汉克尔函数， $h_n^{(2)*} = h_n^{(2)}$ ^[85]。

式(2-6)中，声场信息可以由 B_n^m 完全表征，并且 B_n^m 中仅有 $Y_n^m(\Omega_s)$ 与声源方向有关，理想情况下，信号应该用无穷阶球谐函数进行描述 ($N = +\infty$)，但在实际中受限于实际硬件条件与信号处理算力限制，只能使用有限项 $n \in [0, N]$ ，这种阶数的截断会导致空间混叠，带来误差。

2.2.3 Ambisonics 编码与解码

物理场重构系统的目标是使用多个扬声器调节合成声场，使合成声场与目标声源所辐射出的目标声场一致，实现声场重构。若扬声器 i 位于 $\Omega_i = [\theta_i, \phi_i, r_i]$ ，其信号为 A_i ，阵列包含 L 个扬声器。假设扬声器均辐射平面波，根据式(2-6)，合成声场 P_v 为：

$$P_v(\Omega, \Omega_1, \dots, \Omega_L, k) = \sum_{l=1}^L A_l \sum_{n=0}^N \sum_{m=-n}^n 4\pi j^n j_n(kr) Y_n^m(\Omega_i) Y_n^m(\Omega) \quad (2-8)$$

若合成声场与目标声场一致，即 $P_v = P_s$ ，化简后有：

$$\sum_{l=1}^L A_l Y_n^m(\Omega_l) = Sig \cdot Y_n^m(\Omega_s), m \in [-n, n], n \in [0, N] \quad (2-9)$$

其中, Sig 代表声源的实际信号内容, 上式用矩阵形式表示:

$$\mathbf{Y}_{3D} [A_1, A_2, \dots, A_L]^T = Sig \cdot [Y_0^0(\Omega_s), Y_1^{-1}(\Omega_s), \dots, Y_N^N(\Omega_s)]^T$$

$$\mathbf{Y}_{3D} = \begin{bmatrix} Y_0^0(\Omega_1) & Y_0^0(\Omega_2) & \cdots & Y_0^0(\Omega_L) \\ Y_1^{-1}(\Omega_1) & Y_1^{-1}(\Omega_2) & \cdots & Y_1^{-1}(\Omega_L) \\ \vdots & \vdots & \vdots & \vdots \\ Y_N^N(\Omega_1) & Y_N^N(\Omega_2) & \cdots & Y_N^N(\Omega_L) \end{bmatrix} \quad (2-10)$$

在上式中, T 表示矩阵转置, 上式为一个线性方程组, 可通过矩阵求逆的方式计算得到各个扬声器的馈给信号 $[A_1, A_2, \dots, A_L]^T$:

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_L \end{bmatrix} = \underbrace{\mathbf{D}_{3D} Sig \cdot \begin{bmatrix} Y_0^0(\Omega_s) \\ Y_1^{-1}(\Omega_s) \\ \vdots \\ Y_N^N(\Omega_s) \end{bmatrix}}_{\text{decode}} \quad (2-11)$$

$$\mathbf{D}_{3D} = \mathbf{Y}_{3D}^{-1}$$

对于位于方位 Ω_s 的任意以平面波形式辐射的次级声源信号 Sig , Ambisonics 编码信号 \tilde{Y}_n^m 为 $Sig \cdot Y_n^m(\Omega_s)$, 各通道 Ambisonics 编码信号与声场表征系数的关系为:

$$\tilde{Y}_n^m = Sig \cdot Y_n^m = \frac{B_m^n}{4\pi j^n} \quad (2-12)$$

每个扬声器通路的重放信号实际上是各个 Ambisonics 编码信号的线性叠加组合, 叠加系数可以用合成矩阵 \mathbf{D}_{3D} 表示, 这一过程称为 Ambisonics 信号的解码过程, \mathbf{D}_{3D} 称为解码矩阵。实际求解过程中, \mathbf{Y}_{3D} 不一定存在稳定的逆矩阵 \mathbf{Y}_{3D}^{-1} , 根据其矩阵维度, 存在以下三种情况^[81]:

(1) 若 $(N+1)^2 < L$, \mathbf{Y}_{3D} 为欠定矩阵, 具有无穷多解, 通常使用求伪逆(最小能量约束)的方法进行近似求解, 即 $\mathbf{Y}_{3D}^{-1} = pinv(\mathbf{Y}_{3D}) = \mathbf{Y}_{3D}^H \{ \mathbf{Y}_{3D} \mathbf{Y}_{3D}^T \}^{-1}$, H 表示矩阵的共轭转置运算;

(2) 若 $(N+1)^2=L$, 且次级声源均匀分布, \mathbf{Y}_{3D}^{-1} 存在, 若次级声源不是均匀分布, 通常也采用求伪逆的方法, 存在一定的误差;

(3) 若 $(N+1)^2>L$, 此时 \mathbf{Y}_{3D} 为超定矩阵, 此时没有精确解。

上述解码过程被称为模态匹配解码方法 (mode-match decode, MMD), 实际重放中, 若扬声器布置在空间上不均匀, 矩阵求逆时容易产生病态的解码矩阵^[40], 导致个别通道在重放特定方向声源时增益过大, 即系统不稳定, 引起可察觉的音色改变, 甚至超出实际电声系统的扩声性能。为了评估不同扬声器布置下的解码稳定性, 可以使用矩阵条件数 (condition number) κ 评估矩阵的稳定性^[66]。矩阵条件数是衡量矩阵对输入误差敏感程度的一个指标, 对于任意矩阵 \mathbf{D} , 其条件数 $\kappa(\mathbf{D})$ 的定义为:

$$\kappa(\mathbf{D}) = ||\mathbf{D}|| \cdot ||\mathbf{D}^{-1}|| = \frac{\sigma(\mathbf{D})_{max}}{\sigma(\mathbf{D})_{min}} \quad (2-13)$$

其中, σ 是矩阵的奇异值, 可以通过对矩阵进行奇异值分解计算得到。

上述过程假设目标声源与扬声器均发出平面波, 当声源离场点 (接收点) 较近时, 声源不能近似为平面波声源。根据目标声源与重放扬声器距离接收点的距离, 可以分别使用点声源、平面波两种模型表示, 共有四种可能的组合模型, 声场合成表达式(2-9)可以根据式(2-6)与式(2-7)进行组合, 结果如表 2-3 所示。由于虚拟目标声源通常比扬声器距离接收点更远, 该表中没有给出目标声源采用球面波, 扬声器采用平面波的情况。

表 2-1 Ambisonics 声场重构模型组合

Tab. 2-1 The combinations of Ambisonics sound field reconstruction models

目标声源	扬声器	重声场构等式
平面波	平面波	$\sum_{l=1}^L A_l Y_n^m(\Omega_l) = Sig \cdot Y_n^m(\Omega_s)$
平面波	球面波	$\sum_{l=1}^L A_l Y_n^m(\Omega_l) = Sig \cdot \frac{-4\pi j^n}{jkh_n^{(2)}(kr_l)} Y_n^m(\Omega_s)$
球面波	球面波	$\sum_{l=1}^L A_l Y_n^m(\Omega_l) = Sig \cdot \frac{h_n^{(2)}(kr_s)}{h_n^{(2)}(kr_l)} Y_n^m(\Omega_s)$

当采用球面波模型进行 Ambisonics 重构时, 可以看到扬声器增益与波数 (对应频率) 有关, 这种 Ambisonics 被称为近场补偿 Ambisonics (NFC-Ambisonics)。NFC-

Ambisonics（包括目标源为平面波和点源两种情况）除了引入了一个与阶数、频率相关的系数外，没有改变编码-解码方法本身，该系数可以用滤波矩阵 $\mathbf{H}(k)$ 进行描述，则式(2-11)可以改写为：

$$\begin{bmatrix} A_1(k) \\ A_2(k) \\ \vdots \\ A_L(k) \end{bmatrix} = \underbrace{\mathbf{D}_{3D} \mathbf{H}(k) \text{Sig}}_{decode} \cdot \underbrace{\begin{bmatrix} Y_0^0(\Omega_s) \\ Y_1^{-1}(\Omega_s) \\ \vdots \\ Y_N^N(\Omega_s) \end{bmatrix}}_{encode}$$

$$\mathbf{H}(k) = diag\{h_0(k), h_1(k), h_2(k), \dots, h_N(k)\}$$
(2-14)

其中， $diag$ 表示对角矩阵，对于目标声源为平面波，扬声器为球面波的情况，各个频率上的 $h_n(k) = \frac{-4\pi j^n}{jkh_n^{(2)}(kr_l)}$ ，对于二者均为球面波的情况， $h_n(k) = \frac{h_n^{(2)}(kr_s)}{h_n^{(2)}(kr_l)}$ ，该滤波器的作用是对球面波的曲面波前在不同距离与频率上的差异进行补偿^[86]。

上述过程均是针对三维空间声场的传播与重构，若声源和接收点均位于同一水平面，可以使用二维声场替代三维声场，简化重放过程。类似地，同样只考虑外部声源辐射场，式(2-1)在二维的情况下为^[87]：

$$P(r, \theta, k) = \sum_{n=0}^{\infty} \sum_{m=-1,1} B_n(k) J_n(kr) \Phi_n^m(\theta_s)$$
(2-15)

其中， J_n 为贝塞尔函数。二维声场的本征函数为角度谐波函数 Φ_n^m ，其定义如下式所示，函数图像见图 2-1：

$$\Phi_n^m(\theta) = \frac{1}{\sqrt{2\pi}} \begin{cases} \sqrt{2} \sin(n\theta), & n \neq 0, m = -1 \\ 1, & n = 0 \\ \sqrt{2} \cos(n\theta), & n \neq 0, m = 1 \end{cases}$$
(2-16)

角谐波函数同样满足正交归一化性质。与球谐函数类似，角谐波函数随着阶数提高，具有更尖锐的指向性。对比式(2-2)，可以发现角谐波函数包含在球谐函数中，是球谐函数中与方位角 θ 有关的因子项。对于 N 阶二维 Ambisonics 重放，角谐波的数量为 $2N + 1$ 个，即最少需要 $2N + 1$ 个扬声器进行重放。因此，二维重放所需的扬声器数量远小于三维重放所需的数量。

经过与式(2-9)、式(2-10)与式(2-11)一样的目标声场、重构声场匹配推导过程，二维 Ambisonics 编解码流程可以总结为：

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_L \end{bmatrix} = \underbrace{\mathbf{D}_{2D} \text{Sig}}_{decode} \cdot \underbrace{\begin{bmatrix} \Phi_0^0(\theta_s) \\ \Phi_1^{-1}(\theta_s) \\ \vdots \\ \Phi_N^1(\theta_s) \end{bmatrix}}_{encode} \quad (2-17)$$

$$\mathbf{D}_{2D} = \begin{bmatrix} \Phi_0^0(\theta_1) & \Phi_0^0(\theta_2) & \cdots & \Phi_0^0(\theta_L) \\ \Phi_1^{-1}(\theta_1) & \Phi_1^{-1}(\theta_2) & \cdots & \Phi_1^{-1}(\theta_L) \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_N^1(\theta_1) & \Phi_N^1(\theta_2) & \cdots & \Phi_N^1(\theta_L) \end{bmatrix}^{-1}$$

2.2.4 Ambisonics 解码优化

上一节介绍了 Ambisonics 的模态匹配法，该方法在理论上能够精确重构物理声场，但由于实际使用的 Ambisonics 信号阶数有限，实际重放中能够准确重放的频率和空间范围有限。在扬声器均匀分布的情况下，三维重放的重放阶数与准确重放距离、重放频率的经验公式如下所示^[88]：

$$N = \lceil kr \rceil \quad (2-18)$$

其中， $\lceil \cdot \rceil$ 表示向上取整。阶数的截断会引起空间混叠，当阶数较低时，这种失真会严重影响听音体验，因此，实际的 Ambisonics 解码中往往需要根据实际心理声学实验的听音结果对解码进行优化设计^[89]。Gerzon 基于 Makita 波矢量方向虚拟源定位理论^[90]提出的速度定位矢量 \vec{r}_V 与能量定位矢量 \vec{r}_E 分析方法被广泛接受^[91]。该理论提出两个指标作为合成虚拟声源的方向参考，速度定位矢量 \vec{r}_V 与能量定位矢量 \vec{r}_E 的定义分别为：

$$\vec{r}_V = \frac{\sum_{l=0}^L A_l \vec{r}_l}{\sum_{l=0}^N A_l} \quad (2-19)$$

$$\vec{r}_E = \frac{\sum_{l=0}^L A_l^2 \vec{r}_l}{\sum_{l=0}^N A_l^2} \quad (2-20)$$

其中， A_l , \vec{r}_l , L 分别为第 l 个扬声器的增益，位置矢量和总扬声器个数。一般认为， \vec{r}_v 反映了低频定位线索 ITD 的准确度（与真实声源相比）， \vec{r}_E 反映了中高频定位线索 ILD 的准确度，优化的目标是使 $||\vec{r}_v||=1$ 与 $||\vec{r}_E||=1$ 。由于声场重构理论中，对目标区域声场进行了声压匹配，Ambisonics 系统的 $||\vec{r}_v||$ 通常接近 1，但 $||\vec{r}_E||$ 在扬声器阵列分布不均的情况下，波动较大。Daniel 等人提出采用最大化 $||\vec{r}_E||$ 的 max- r_E 优化方法，使重放更加稳定^[42]。在原有的 Ambisonics 解码下，其优化方式如下所示：

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_L \end{bmatrix} = \underbrace{\mathbf{D}_{3D} \mathbf{A} \mathbf{Sig}}_{decode} \cdot \underbrace{\begin{bmatrix} Y_0^0(\Omega_s) \\ Y_1^{-1}(\Omega_s) \\ \vdots \\ Y_N^N(\Omega_s) \end{bmatrix}}_{encode}$$

$$\mathbf{A} = diag\{a_n\}$$

$$a_n = P_n \left(\cos \left(\frac{137.9^\circ}{N+1.51} \right) \right)$$
(2-21)

$diag$ 表示对角矩阵，上式的效果为对各阶球谐函数谐波乘上一个“窗函数”，从而削减旁瓣。

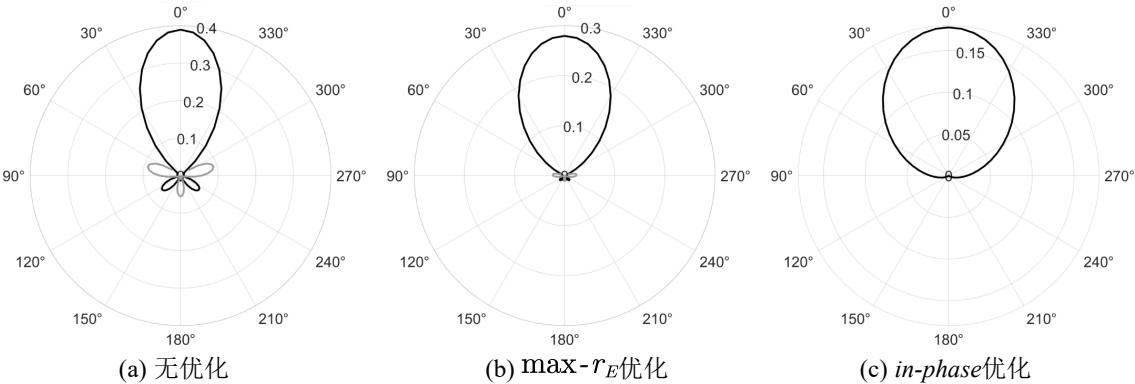


图 2-2 不同优化下 3 阶 Ambisonics 在水平面上的虚拟指向性

Fig. 2-2 The virtual directivity on the horizontal plane of 3rd-order Ambisonics using different decoding optimizations

使用 $max-r_E$ 优化后，重放扬声器虚拟指向性（即重放不同方向虚拟声源时，扬声器的增益随方向的变化）的旁瓣被极大的压制了（见图 2-2），然而，主瓣的宽度也同时增加，这意味着重构声场的空间分辨率可能会下降。类似地，还有 *in-phase* 优化方法，

类似地，通过使用不同的谐波分量权重系数改变窗函数形状，使得虚拟指向性没有负值，优化系数计算方法为：

$$a_n = \frac{N! (N+1)!}{(N+n+1)! (N-n)!} \quad (2-22)$$

2.2.5 混合阶 Ambisonics

混合阶 Ambisonics 对空间 Ambisonics 的基函数进行选择性扩展，提高水平面的球谐函数指向性。目前，混合阶 Ambisonics 有多种基函数的扩展策略^[92]。球谐函数按照其定义式(2-2)，可以被分为三种类型。谐波分类如图 2-3 所示，包括：Zonal 谐波 $\{Y_n^m | n \geq 0, m=0\}$ 、Sectoral 谐波 $\{Y_n^m | n \geq 0, m=\pm n\}$ 以及 Tesseral 谐波 $\{Y_n^m | n \geq 0, 0 < |m| < n\}$ 。图中仅标出 $m \geq 0$ 的部分球谐函数，其余对称部分定义相同。根据式(2-2)， Y_n^m 中与仰角 ϕ 相关的系数为连带勒让德多项式 $P_n^{|m|}(\sin \phi)$ ，对于 Sectoral 谐波，该项为 $P_n^n(\sin \phi)$ ，可以展开为^[93]：

$$P_n^n(\sin \phi) = (-1)^n (2n-1)!! \cos^n \phi \quad (2-23)$$

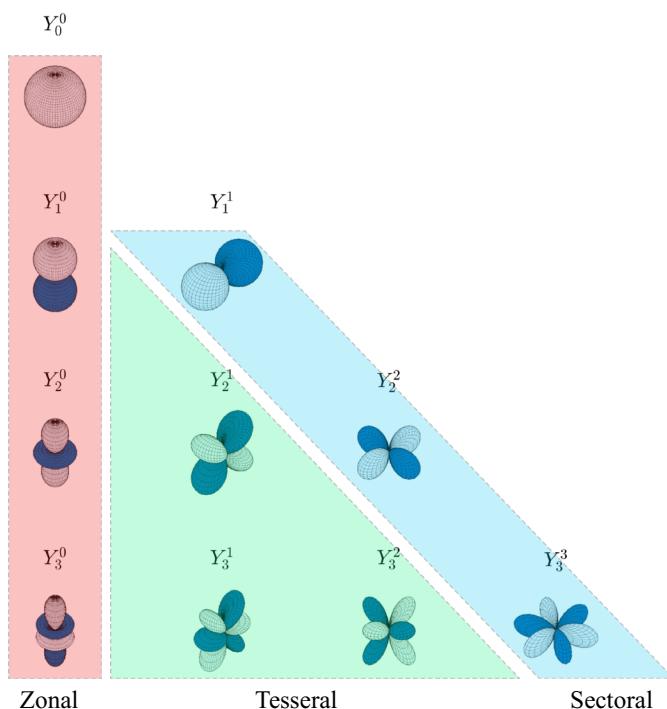


图 2-3 0 至 3 阶的 Zonal、Tesseral 及 Sectoral 球谐函数

Fig. 2-3 Zonal, Tesseral and Sectoral spherical harmonic functions from order 0 to 3

P_n^n 中与仰角相关的系数是 $\cos^n \phi$, 随着仰角数值 $|\phi|$ 增加, P_n^n 的数值大小快速降低, 对应球谐函数 Y_n^n 能量降低, 这说明 Sectoral 谐波能量主要集中在水平面。可以在保留一定阶数 3-D Ambisonics 的全部谐波的基础上, 保留更高阶数的 Sectoral 谐波, 由于这部分谐波仅在水平面及附近仰角存在较大能量, 因此, 即使当水平面以外的扬声器数量较少时, 额外增加的谐波不会过度降低解码矩阵的稳定性。

若原有 3-D Ambisonics 阶数为 M_{3D} , 扩展 Sectoral 谐波阶数至 M_{2D} ($M_{2D} > M_{3D}$), 对于位于 Ω_s 处的声源, 其 Ambisonics 编码信号为^[57]:

$$\begin{aligned}\tilde{Y} = \text{Sig} \cdot [Y_0^0(\Omega_s), Y_1^{-1}(\Omega_s), \dots, Y_{M_{3D}}^N(\Omega_s), \\ Y_{M_{3D}+1}^{-1}(\Omega_s), Y_{M_{3D}+1}^1(\Omega_s), \dots, Y_{M_{2D}}^{-1}(\Omega_s), Y_{M_{2D}}^1(\Omega_s)]^T\end{aligned}\quad (2-24)$$

对应地, 式(2-11)中的 3-D Ambisonics 解码矩阵 \mathbf{D}_{3D} 可以改写为混合阶 Ambisonics 的解码矩阵 \mathbf{D}_{mix} , 如下所示:

$$\mathbf{Y}_{\text{mix}} = \begin{bmatrix} Y_0^0(\Omega_1) & Y_0^0(\Omega_2) & \cdots & Y_0^0(\Omega_L) \\ Y_1^{-1}(\Omega_1) & Y_1^{-1}(\Omega_2) & \cdots & Y_1^{-1}(\Omega_L) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{M_{3D}}^N(\Omega_1) & Y_{M_{3D}}^N(\Omega_2) & \cdots & Y_{M_{3D}}^N(\Omega_L) \\ Y_{M_{3D}+1}^{-1}(\Omega_1) & Y_{M_{3D}+1}^{-1}(\Omega_2) & \cdots & Y_{M_{3D}+1}^{-1}(\Omega_L) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{M_{2D}}^1(\Omega_1) & Y_{M_{2D}}^1(\Omega_2) & \cdots & Y_{M_{2D}}^1(\Omega_L) \end{bmatrix} \quad (2-25)$$

$$\mathbf{D}_{\text{mix}} = \mathbf{Y}_{\text{mix}}^{-1}$$

当重放扬声器个数 L 满足 $L \geq (M_{3D} + 1)^2 + 2(M_{2D} - M_{3D})$ 时, 解码矩阵可以通过矩阵求逆的方式得到。对比高至 M_{2D} 阶的 3-D Ambisonics 系统, 混合阶 Ambisonics 系统对扬声器数量的要求下降了 $M_{2D}^2 - M_{3D}^2$ 个, 并且这种重算法其允许系统通过在水平面布置更密集的扬声器阵列, 提高水平面虚拟声源的空间分辨率。为便于表达, 3-D Ambisonics 谐波记为 \mathbf{P} , Sectoral 谐波记为 \mathbf{H} , 该重放方式可以记为 $M_{2D} \mathbf{H} M_{3D} \mathbf{P}$ 阶的混合阶 Ambisonics^[94]。

特殊地, 除了对球谐函数进行 Sectoral 谐波的扩展, 有时还需要对已编码的 3-D Ambisonics 信号进行 2-D Ambisonics 重放, 即信号的降维^[95]。此时, 采用类似的方法, 仅保留 Sectoral 谐波部分, 然后进行一定能量匹配后, 按照 2.2.3 节所述的 2-D Ambisonics 解码方式进行重放。对比二维角谐波函数 Φ_n^m 和三维球谐函数 Y_n^m , 容易注

意到 $Y_n^{\pm n}(\theta, \phi) = N_n^n P_n^n(\sin \phi) \Phi_n^{\pm 1}(\theta)$, 经过一定的数学展开变形后, 角谐波和球谐波函数有以下转换关系^[92] (推导过程见附录 1) :

$$\begin{aligned}\Phi_n^{\pm 1}(\theta) &= Y_n^{\pm n}(\theta, 0^\circ) \cdot \alpha \\ \alpha &= \frac{\sqrt{\pi}}{Y_n^n(0^\circ, 0^\circ)} = \frac{1}{N_n^n P_n^n(0)} = \frac{\sqrt{2} \cdot 2^n \cdot 2!}{\sqrt{(2n+1)!}}\end{aligned}\quad (2-26)$$

通过转换系数 α 对 3-D Ambisonics 信号的 Sectoral 谐波部分进行选择性保留以及补偿后, 即可转化为 2-D Ambisonics 信号进行二维重放。

混合阶 Ambisonics 具有空间不对称的特性, 在水平面方位角 θ 上具有更高的分辨率, 以本文中的 192 通路扬声器阵列为例, 考察扬声器的虚拟指向性, 对比 2 阶 3-D Ambisonics 与 4H2P 混合阶 Ambisonics 在 MMD 解码下, 正前方 $\Omega(0^\circ, 0^\circ)$ 扬声器的指向性, 如图 2-4 所示。可以看到, 扬声器在方位角 θ 上具有更窄的主瓣, 在水平面方位角上具有更高的空间分辨率, 但二者主瓣在仰角上的分辨率一致。

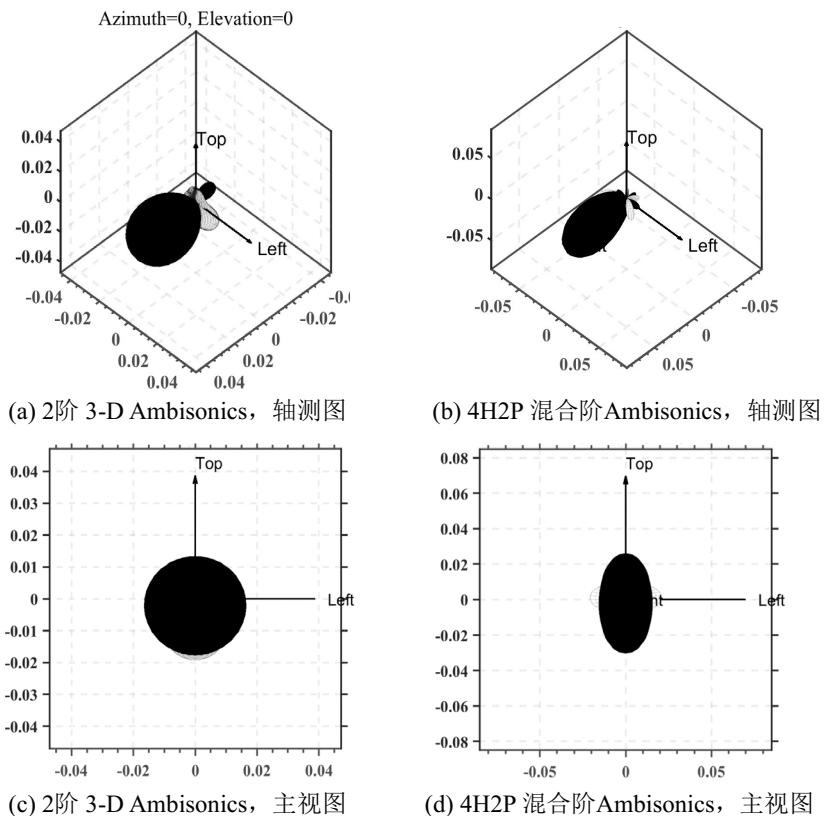


图 2-4 2 阶 3-D 与 4H2P 混合阶 Ambisonics 下的正前方扬声器指向性

Fig. 2-4 Directivity patterns of the front loudspeaker under 2nd-order 3-D Ambisonics and 4H2P mixed-order Ambisonics

对于空间非均匀布置的扬声器阵列，特别是高仰角区域缺少足够扬声器的情况，混合阶 Ambisonics 的解码矩阵具有更低的条件数，可以显著提高重放的稳定性。以本文中的多通路扬声器阵列为例，该阵列在水平面有 36 个扬声器，二维 Ambisonics 的理论阶数上限为 17 阶，表 2-2 对比了该阵列 0 至 6 阶 M_{3D} 时，不同 M_{2D} 下的条件数 κ ，条件数定义见式(2-13)。可以看到，由于该阵列在水平面有大量的扬声器，在固定的 M_{3D} 下，提高 M_{2D} 对解码矩阵的条件数几乎没有影响。

表 2-2 不同阶数下混合阶 Ambisonics 解码矩阵的条件数

Tab. 2-2 The condition numbers of decode matrix of mixed-order Ambisonics under different order

κ	M_{3D}					
M_{2D}	1	2	3	4	5	6
1	1.7					
2	1.7	3.1				
3	1.7	3.1	6.3			
4	1.7	3.1	6.3	16.0		
5	1.1	3.1	6.3	16.0	46.0	
6	1.7	3.1	6.3	16.0	46.0	163.7
7	1.7	3.1	6.3	16.0	46.0	163.7
8	1.7	3.1	6.3	16.0	46.0	163.7
9	1.7	3.1	6.3	16.0	46.0	163.7
10	1.7	3.1	6.3	16.0	46.0	163.7
11	1.7	3.1	6.3	16.0	46.0	163.7
12	1.9	3.2	6.6	16.7	47.9	169.8
13	1.9	3.2	6.6	16.7	47.9	169.8
14	1.9	3.3	6.7	16.8	48.0	170.2
15	1.9	3.3	6.7	16.8	48.0	170.4
16	1.9	3.4	7.0	17.6	50.4	179.0
17	2.0	3.4	7.0	17.6	50.5	179.1

2.2.6 Ambisonics 信号拾拾

在 2.2.2 节所述的 Ambisonics 重放理论中，声源的方向已知时，可以将特定方向点声源或平面波信号编码为 Ambisonics 信号，然后根据扬声器阵列的空间分布进行解码，进而合成虚拟声源的声场。在实际使用中，往往需要直接对已有的声场进行拾拾，此时声源的方向是未知的。如前所述，声场的全部信息由 B_n^m 完全决定，可以使用传声器阵列直接对声环境进行录制，通过信号处理后计算得到声场信息 B_n^m ，进而转化为 Ambisonics 信号。目前，可以通过空间重合传声器或球形传声器阵列对信号进行拾拾，经过一定的信号处理后得到 Ambisonics 编码信号。

空间重合传声器在空间上的同一位置，布置具有各阶球谐函数指向性（见图 2-1）的多个传声器，直接拾拾 Ambisonics 编码信号。具有高阶指向性的传声器在实际中难以实现，目前，空间重合传声器拾拾只能实现 FOA 信号的拾拾。若四个传声器分别指向左前上方向（FLU，方向矢量为 $[1, 1, 1]$ ）、左后下方向（BLD， $[-1, 1, 1]$ ）、右前下方向（FRD， $[1, -1, -1]$ ）以及右后上方向（BRU， $[-1, -1, 1]$ ）四个方向，传声器响应为心形响应， $Response = 1/2 + 1/2\cos(\Delta\alpha)$ ($\Delta\alpha$ 为传声器主轴与声源方向的空间夹角) [96]。假设声源位于 Ω_s ，则 Ambisonics 信号可通过拾拾信号计算得到：

$$\begin{bmatrix} W \\ X \\ Y \\ Z \end{bmatrix} = \frac{\sqrt{3}}{2} \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} FLU \\ FRD \\ BLD \\ BRU \end{bmatrix} \quad (2-27)$$

其中， $[W, X, Y, Z]^T$ 被称为 B-format 制式信号，该信号与 N3D 格式的 FOA 信号关系为：

$$[W, X, Y, Z]^T = \frac{\sqrt{3}}{2\sqrt{\pi}} \left[\tilde{Y}_0^0 / \sqrt{2\pi}, \tilde{Y}_1^{-1}, \tilde{Y}_1^1, \tilde{Y}_1^0 \right]^T \quad (2-28)$$

为了拾拾更高阶的信号，需要使用更多数量的传声器，球形传声器阵列则可以拾拾空间高阶 3-D Ambisonics 信号^[97,98]。球形传声器阵列拾拾是重放的逆过程，各阶 Ambisonics 信号 \tilde{Y}_n^m 可由球阵列的拾拾信号进行线性组合及滤波得到，流程如图 2-5 所示。首先使用阵列上各个通道对信号进行拾拾，然后对每个通道进行加权组合之后，得到各阶声场分解系数。进一步使用各阶对应的辐射函数补偿滤波器对信号进行滤波处理，得到 Ambisonics 编码信号。

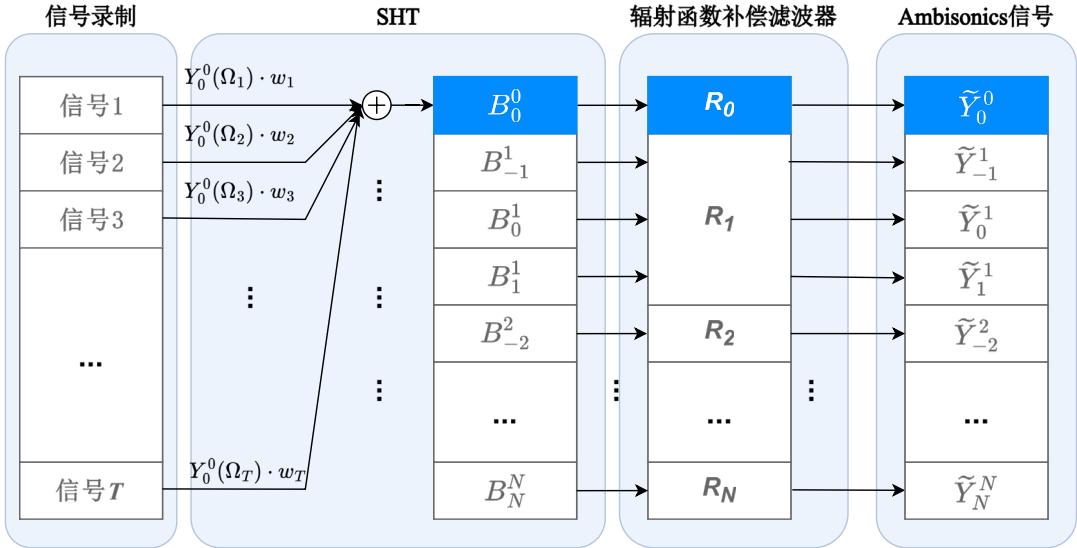


图 2-5 球形传声器阵列检拾信号处理流程

Fig. 2-5 Diagram of signal processing of spherical microphone array

具体而言，若球形传声器阵列的半径为 a （传声器位于阵列表面），对于位于 Ω_s 的平面波声源，阵列上 Ω 位置处的频域声压 P 为声源入射声波与球面反射声波的叠加，可由刚球散射模型计算得到^[99]：

$$P(\Omega, \Omega_s, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n [j_n(ka) - \frac{j_n'(ka)}{h_n^{(2)'}(ka)} h_n^{(2)}(ka)] (2n+1) j^n P_n(\vec{r}_s \cdot \vec{r}) \quad (2-29)$$

其中， \vec{r}_s, \vec{r}, h_n 分别是声源和接收点的单位位置矢量，' 为一阶导数。上式可按式(2-5)与式(2-6)的流程对声场进行球谐函数分解，得到：

$$\begin{cases} P(\Omega, \Omega_s, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n [j_n(kr) - \frac{j_n'(kr)}{h_n^{(2)'}(kr)} h_n^{(2)}(kr)] B_n^m Y_n^m(\Omega) \\ B_n^m = 4\pi j^n Y_n^m(\Omega_s) \end{cases} \quad (2-30)$$

实际检拾中面临的问题是上式的逆问题，阵列上频域声压 P 已知，声场声源数量与其方向 Ω_s 未知，需对 B_n^m 进行求解，根据式(2-5)对频域声压 P 进行 SHT 变换，并用 Wronskian 关系^[100]将 $j_n(ka) - [j_n'(ka)/h_n'(ka)] h_n(ka)$ 转化为 $-j/[(ka)^2 h_n'(ka)]$ ，得到：

$$\begin{aligned}
 B_n^m &= \frac{1}{j_n(ka) - [j'_n(ka)/h'_n(ka)]h_n(ka)} \\
 &\times \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} P(\Omega, k) Y_n^m(\theta, \phi) \cos(\phi) d\theta d\phi \\
 &= (ka)^2 j h_n^{(2)\prime}(ka) \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} P(\Omega, k) Y_n^m(\theta, \phi) \cos(\phi) d\theta d\phi
 \end{aligned} \tag{2-31}$$

式(2-31)中假设球面上的声压处处可知，实际上，传声器在阵列上离散分布，只能取 T 个有限位置，假设传声器 t 位于 Ω_t , $t \in [1, T]$ ，该传声器拾拾的频域声压为 $P_t(\Omega_t, k)$ ，式(2-31)变为离散形式：

$$B_n^m = (ka)^2 j h_n^{(2)\prime}(ka) \sum_{t=1}^T w_t P_t(\Omega_t, k) Y_n^m(\Omega_t) \tag{2-32}$$

其中， w_t 为各个采样点的权重系数，对于传声器均匀分布或近似均匀分布情况， $w_t = 4\pi/T^{[59]}$ 。上述过程用公式总结描述为：

$$\begin{bmatrix} \tilde{Y}_0^0 \\ \tilde{Y}_{-1}^1 \\ \vdots \\ \tilde{Y}_N^N \end{bmatrix} = \frac{4\pi}{T} \text{diag}\{R_n\} \begin{bmatrix} Y_0^0(\Omega_1) & Y_0^0(\Omega_2) & \cdots & Y_0^0(\Omega_T) \\ Y_{-1}^1(\Omega_1) & Y_{-1}^1(\Omega_2) & \cdots & Y_{-1}^1(\Omega_T) \\ \vdots & \vdots & \vdots & \vdots \\ Y_N^N(\Omega_1) & Y_N^N(\Omega_2) & \cdots & Y_N^N(\Omega_T) \end{bmatrix} \begin{bmatrix} P_1(\Omega_1, k) \\ P_2(\Omega_2, k) \\ \vdots \\ P_T(\Omega_T, k) \end{bmatrix} \tag{2-33}$$

其中， R_n 为球形传声器阵列的声场辐射函数补偿滤波器，与球阵列的类型（实心刚性球或空心球）、假设的声源类型（平面波或球面波）有关， R_n 系数如下表所示：

表 2-3 球阵列拾拾声场辐射函数补偿滤波器系数

Tab. 2-3 Coefficients of sound-field radiation compensation filters for spherical microphone array

球阵列类型	声源类型	R_n
实心球	平面波	$\frac{(ka)^2 h_n^{(2)\prime}(ka)}{4\pi j^{(1+n)}}$
实心球	球面波	$-\frac{ka^2 h_n^{(2)\prime}(ka)}{h_n^{(2)}(kr_s)}$

空心球	平面波	$\frac{1}{4\pi j^n j_n(ka)}$
空心球	球面波	$\frac{j}{kj_n(ka) h_n^{(2)}(kr_s)}$

2.3 高阶 Ambisonics 可听化系统设计

基于 2.2 节所述的高阶 Ambisonics 原理，本文首先建立了一个用于声环境可听化的高阶 Ambisonics 重放系统。该系统是后续章节研究的基础，下面分别对系统的总体框架设计、硬件及软件的实现进行介绍。

2.3.1 系统框架

高阶 Ambisonics 可听化声重放系统的设计框架如图 2-6 所示，本文的研究主要关注高阶 Ambisonics 系统的重放问题，图中仅展示了文中所用到的重放相关部分系统，主要包含声场模拟、软件信号处理及重放硬件三部分，涵盖 Ambisonics 信号处理、声场模拟、硬件及声音输出模块，实现了精准、沉浸式的声场可听化。

系统在信号处理上以高阶 Ambisonics 重放为核心，通过预设虚拟声源方向及该方向上的干信号，按照 2.2.3 节所述方法进行 Ambisonics 编码。得到 Ambisonics 编码信号后，可以对 Ambisonics 信号进行变化，包括维度变化（3-D Ambisonics 与 2-D Ambisonics 变换）及声场旋转等后期处理。然后，Ambisonics 信号通过解码后馈给经过均衡处理的扬声器阵列进行多通路播放，实现声场重构，信号处理部分主要在 MATLAB 2024a 平台上编程实现。

声场模拟部分主要通过使用 Odeon 12 几何声学仿真软件完成。Odeon 是建筑声学仿真中常用的软件，该软件结合了虚源法与声线追踪法，分别用于计算早期及后期反射声^[101]。系统根据 Odeon 所计算的早期反射与后期反射信息，对反射声信息进行一定处理后，使用 Ambisonics 信号处理模块进行空间重放。

硬件部分包含扬声器阵列系统与信号传递与转换设备，用于空间可听化信号重放，通过对系统进行多通路脉冲响应测量，设计均衡滤波，降低硬件系统与理论假设模型的偏差，从而提高重放的精度。

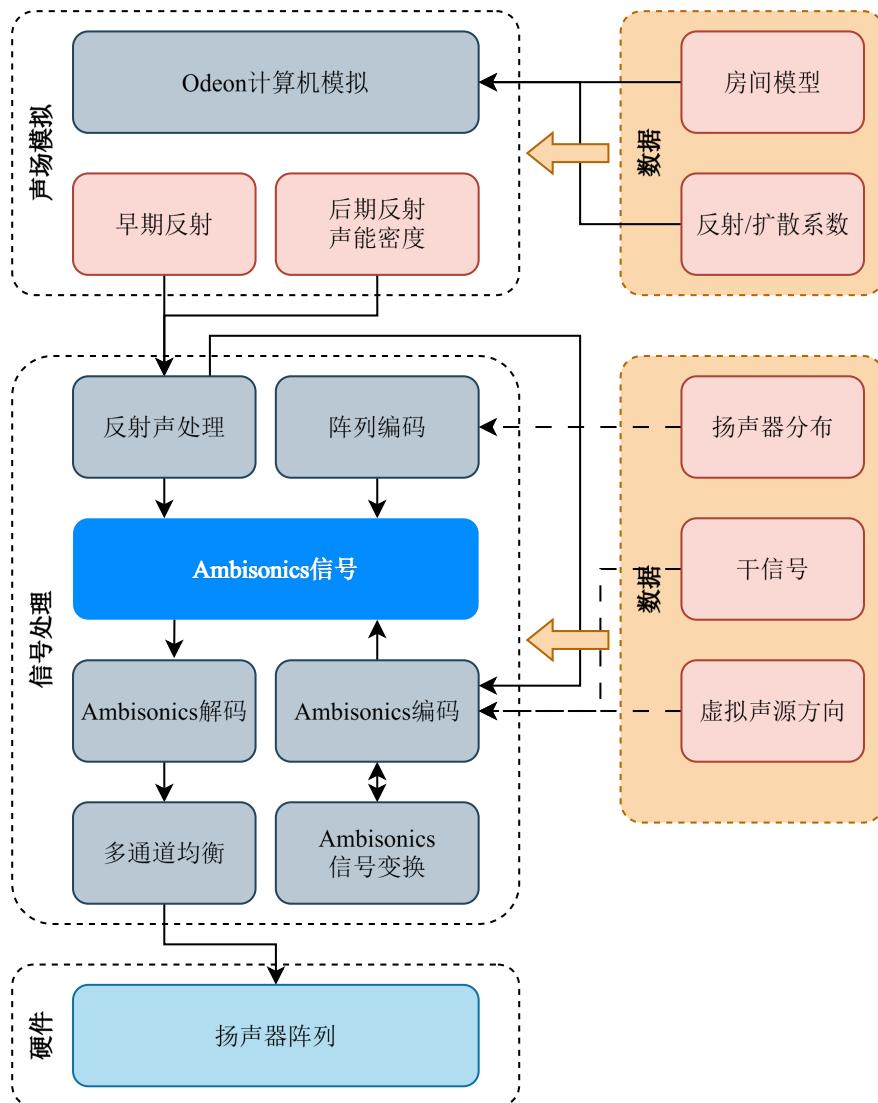


图 2-6 高阶 Ambisonics 可听化系统框架

Fig. 2-6 The Framework of higher-order Ambisonics auralization system

2.3.2 硬件系统设计及调试

2.3.2.1 硬件结构

扬声器阵列包含 192 通道的扬声器，扬声器安装在近似球形的钢结构网架上，其空间分布如表 2-4 所示。扬声器以 15° 为间隔，均匀分布在 -30° 至 90° 共 9 个仰角平面上。扬声器阵列被布置于一个半消声室内（背景噪声为 3.5 dBA），地面铺设了 13 mm 厚的地毯以降低地面反射的干扰，同时阵列支架上包裹了吸音棉以降低支架的反射干扰。

本研究中所用的可听化系统主要包含以下设备：

- (1) RME HDSPE MADI FX PCI-E 声卡 1 台；
- (2) TACSystem VMC-102 监听控制器 2 台；

- (3) Focusrite RedNet D64R MADI-Dante 数字转化接口 3 台;
- (4) Focusrite Red 16Line 数模/模数转换器 12 台;
- (5) Genelec 8010A 主动监听音箱 192 台;
- (6) Cisco SG-300 28 口千兆交换机 2 台;
- (7) Focusrite RedNet AM2 耳机监听单元。

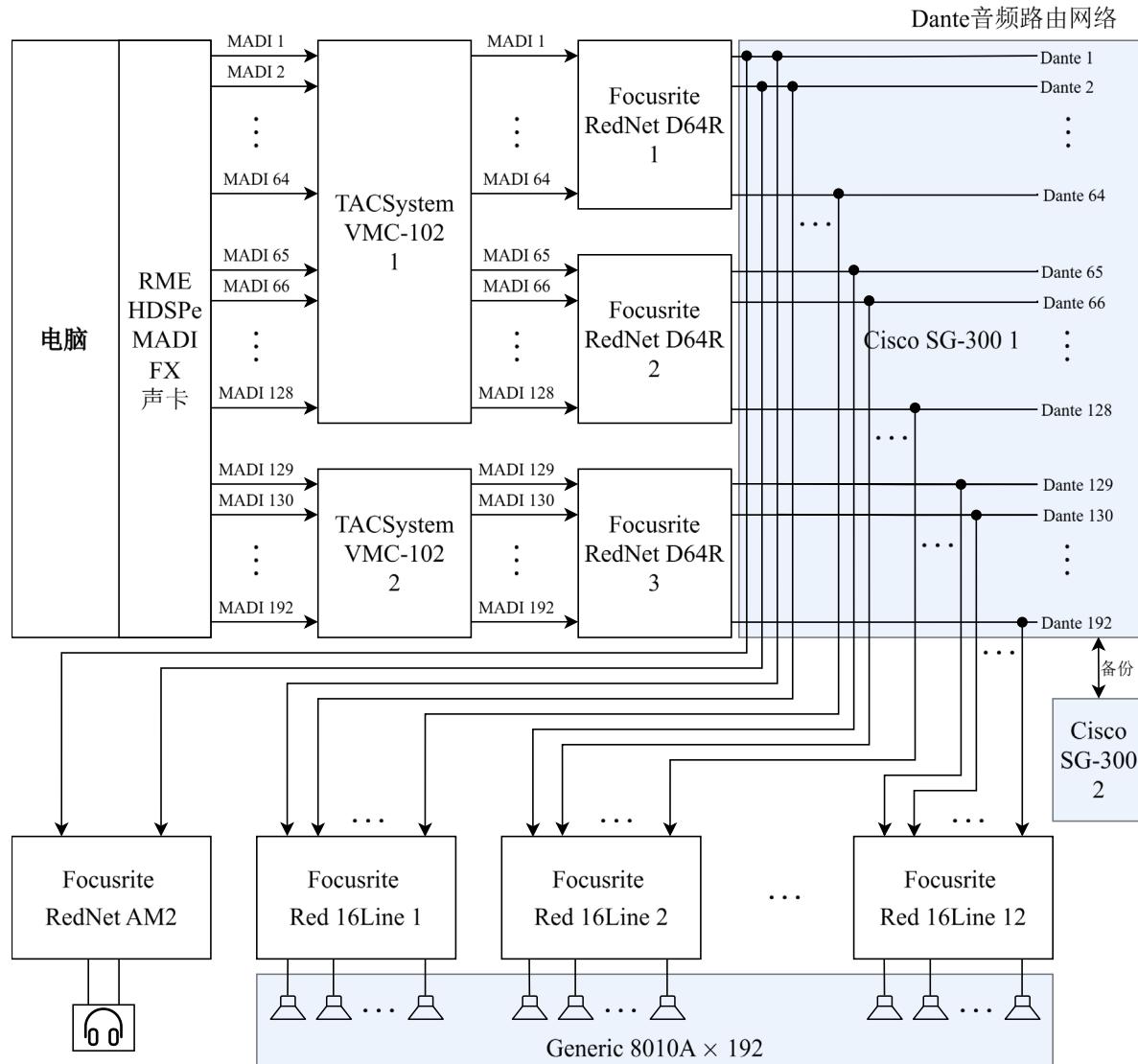


图 2-7 硬件系统框架图

Fig. 2-7 The framework of the hardware system

上述设备的硬件连接如图 2-7 所示。数字信号在电脑上通过系统的软件部分计算生成后，RME 声卡将多通路信号以 MADI 格式传输至 VMC-102。VMC-102 进行增益控制与监控，然后将信号传输至 D64R，转为 Dante 格式的网络音频。D64R 与 12 台

16Line 组成了一个 Dante 音频路由网络，多通路声信号路由至 16Line。16Line 进一步将数字信号转化为模拟信号，分别馈给所有扬声器。

表 2-4 192 通道扬声器阵列各扬声器空间方位

Tab. 2-4 The spatial direction of each loudspeaker in the 192-channel loudspeaker array

仰角 ϕ	方位角 θ	数量
-30°	0° 至 342° 以 18° 等间隔分布	20
-15°	0° 至 345° 以 15° 等间隔分布	24
0°	0° 至 350° 以 10° 等间隔分布	36
15°	0° 至 348.75° 以 11.25° 等间隔分布	32
30°	0° 至 348° 以 12° 等间隔分布	30
45°	0° 至 345° 以 15° 等间隔分布	24
60°	0° 至 337.5° 以 22.5° 等间隔分布	16
75°	0° 至 320° 以 40° 等间隔分布	9
90°	0°	1

2.3.2.2 脉冲响应测量

由于系统包含较多数量的扬声器，每个扬声器在安装时存在一定的位置误差，且扬声器之间本身存在一定的差异。这些误差会导致扬声器在接收点（阵列中心）存在频谱及延时差异，偏离声场重构理论模型。为了测量并修正误差，需要对扬声器进行脉冲响应测量，根据所测量的结果，设计均衡滤波器组对误差进行补偿，保证重放声场与理论声场尽可能一致。

本文中的脉冲响应测量均采用对数正弦扫描频率信号（扫频信号）测量法。扫频信号与最大长度序列（Maximum Length Sequence, MLS）信号测量法是目前使用最广泛的两种脉冲响应测量方法^[102]。相比 MLS 测量，扫频信号测量法通过对估计的脉冲响应进行加窗处理，降低背景噪声的干扰，通常可以获得更高的测量信噪比，对数扫频信号的定义为^[103]：

$$\begin{cases} x(t) = \sin\left[\frac{2\pi f_0 t_1}{a} (e^{at/t_1} - 1)\right] \\ a = \ln(f_1/f_0) \end{cases} \quad (2-34)$$

其中, f_1, f_0, t_1 分别为信号的频率上限、频率下限以及持续时间。

基于扫频信号的脉冲响应测量原理如图 2-8 所示。对于任意的线性时不变 (Linear time invariant, LTI) 系统, 在理想状态下, 系统传输可以由一个脉冲响应 $h(t)$ 进行描述, 系统的输出信号 $y(t)$ 为输入信号 $x(t)$ 与脉冲响应的卷积。在本文中, LTI 系统包括扬声器本身, 扬声器所发出的声波在房间中传输至接收点, 接收点传声器将声信号转化为数字信号的整个过程 (图 2-8 中的蓝色方框部分)。

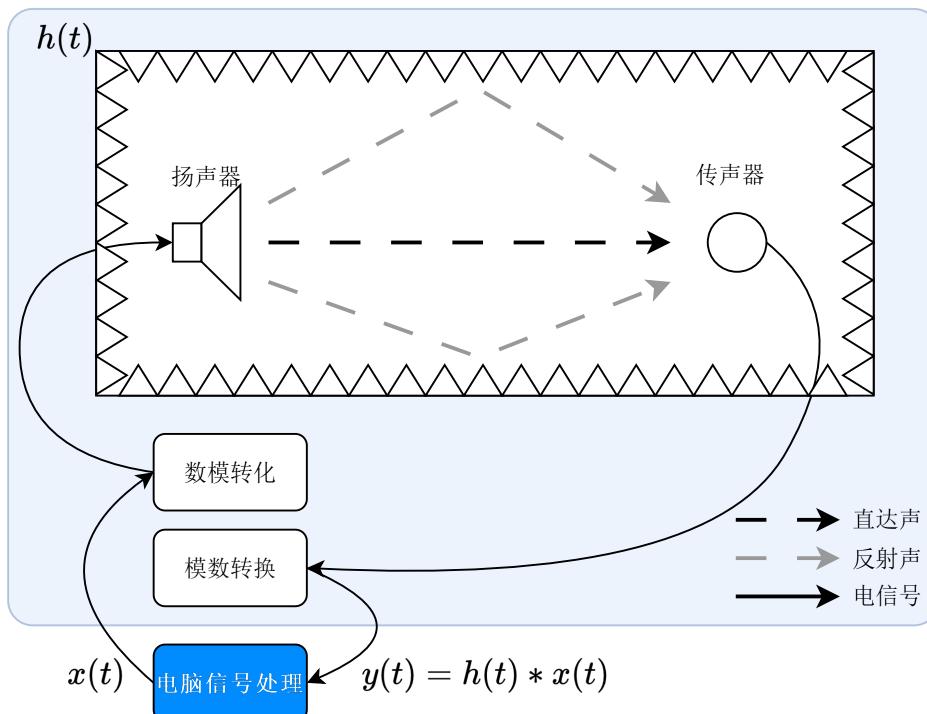


图 2-8 脉冲响应测量原理图

Fig. 2-8 The diagram of impulse response measurement

测量脉冲响应时, 将扫频信号 $x(t)$ 馈给扬声器, 通过传声器测量得到接收信号 $y(t)$ 。时域信号卷积等效为频域信号相乘, 因此可以将上述过程表示为:

$$\begin{cases} y(t) = h(t) * x(t) \\ Y(f) = H(f) \cdot X(f) \end{cases} \quad (2-35)$$

其中, Y, H, X 分别为 y, h, x 对应的频域信号, 通过傅里叶变化计算得到, 则测量得到的系统脉冲响应 $\tilde{h}(t)$ 为:

$$\tilde{h}(t) = \text{IFFT} \left[\frac{Y(f)}{X(f)} \right] \quad (2-36)$$

IFFT 为快速傅里叶逆变换，上述过程被称为解卷积。由于实际测量过程中存在噪声干扰，且实际系统一般不完全符合 LTI 的要求，因此 $\tilde{h}(t)$ 只是系统真实脉冲响应 $h(t)$ 的近似估计。

上述过程中， $X(f)$ 在特定的频率，由于播放设备的实际电声性能限制，或由于实际 LTI 系统本身的特性影响，可能会出现数值陡降（接近 0）的情况，导致进行傅里叶逆变换后的 $\tilde{h}(t)$ 不稳定，出现较大的“振铃”。实际求解的过程中通常采用正则化的方法保持求解的稳定性：

$$\tilde{h}(t) = \text{IFFT} \left[\frac{H^*(f)}{H^2(f) + \varepsilon(f)} \right] \quad (2-37)$$

其中， $H(f) = Y(f)/X(f)$ ， $\varepsilon(f)$ 为正则化系数，在有效频率范围内取接近 0 的数值，在有效频率范围外取相对较大的数值。

2.3.2.3 传声器指向性补偿

本文中，扬声器脉冲响应测量采用 1/2 英寸的自由场电容传声器（Brüel & Kjær Type 4189）。当传声器被放置于声场之中时，传声器本身会对声场的高频部分产生干扰。声源相对传声器主轴沿不同方向入射时，这种干扰的大小会产生变化。自由场传声器在设计时仅针对声源正对振膜（即沿主轴）入射的情况进行了校准处理，补偿了传声器本身对声场的干扰。当声源正对传声器主轴入射时，传声器的频率响应平直。但在本文的扬声器测量中，扬声器被放置于不同方位，而测量传声器主轴只能正对一个方向，为避免传声器本身的频率响应特性对测量结果造成影响，首先需要对传声器的指向性进行测量并设计滤波器进行补偿。

传声器指向性测量实验如图 2-9 所示，测量实验在全消声室开展，本底噪声低于 -12.1 dBA。实验中使用一个 Generic 8010A 扬声器作为测试声源，调整传声器的主轴角度 ϕ ($\phi = 0^\circ$ 时传声器正对声源)，每次调整角度后重新校准传声器中心点的位置，使不同测量角度下中心点位置保持一致。测量声源相对传声器主轴在不同角度入射时的系统脉冲响应。传声器距离声源 350 cm，二者高度均为 140 cm。实验采用对数扫频信号进行脉冲响应测量。

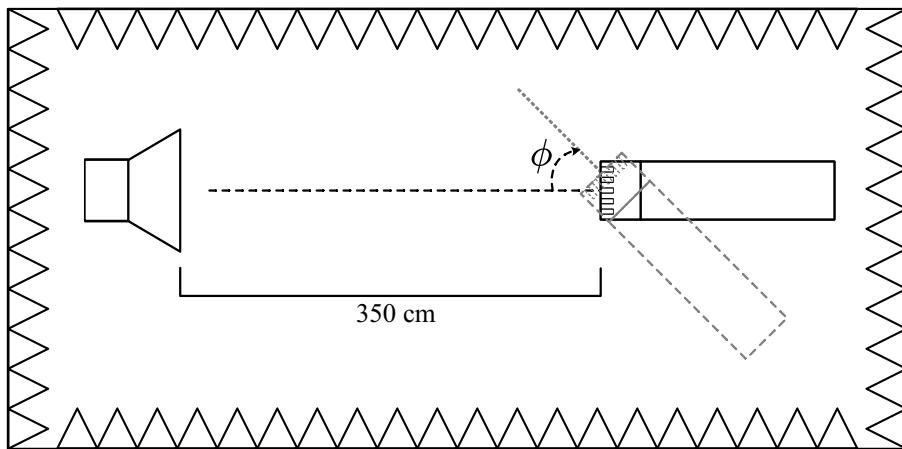


图 2-9 传声器指向性测量示意图

Fig. 2-9 The diagram of microphone directivity measurement

实验中一共测量了 0° 至 180° 之间的 14 个入射角度（以 15° 为步长）下的脉冲响应。然后，所有角度的脉冲响应幅度谱（不考虑相位）均减去 0° 入射情况下脉冲响应的幅度谱。由于本实验中使用的声源低频下限约为 80 Hz，且传声器在低频的响应几乎与入射角度无关，因此低频部分缺失的幅度谱直接使用有效频率部分的数据进行填充。

原始测量的幅度谱与处理后的幅度谱（添加低频部分，并减去 0° 入射情况下脉冲响应的幅度谱），分别如图 2-10 中的 (a) 与 (b) 所示。可以看到，声源与传声器主轴夹角对传声器频率响应的影响位于 1 kHz 以上，在 4 kHz 以上具有较明显的影响，此时频率响应随角度的变化大于 1 dB。

根据该频谱设计逆滤波器，实现不同方向入射的均衡。均衡滤波器为 512 点有限冲激响应（Finite Impulse Response, FIR）线性相位滤波器，频率范围为 0 到 20 kHz，均衡滤波器的时频图如图 2-10 (c) 与 (d) 所示，该滤波器为添加了 256 个采样点延时的零相位滤波器。若声源与传声器主轴的夹角不在所测量的 14 个角度上，则取最近邻的两组均衡滤波器的幅度谱进行插值，然后计算对应方向的均衡滤波器。若目标补偿仰角为 ϕ_s ，相邻两个实测仰角为 ϕ_1 与 ϕ_2 ($\phi_2 > \phi_1$)，则补偿滤波器的频谱 EQ_s 为：

$$EQ_s(f) = \frac{EQ_2(f) - EQ_1(f)}{\phi_2 - \phi_1} \phi_s + EQ_1(f) \quad (2-38)$$

其中, EQ_1 与 EQ_2 分别为 ϕ_1 与 ϕ_2 对应的均衡滤波器。如无特殊说明, 在本文后续的脉冲响应测量均使用了对应方向的均衡滤波器。

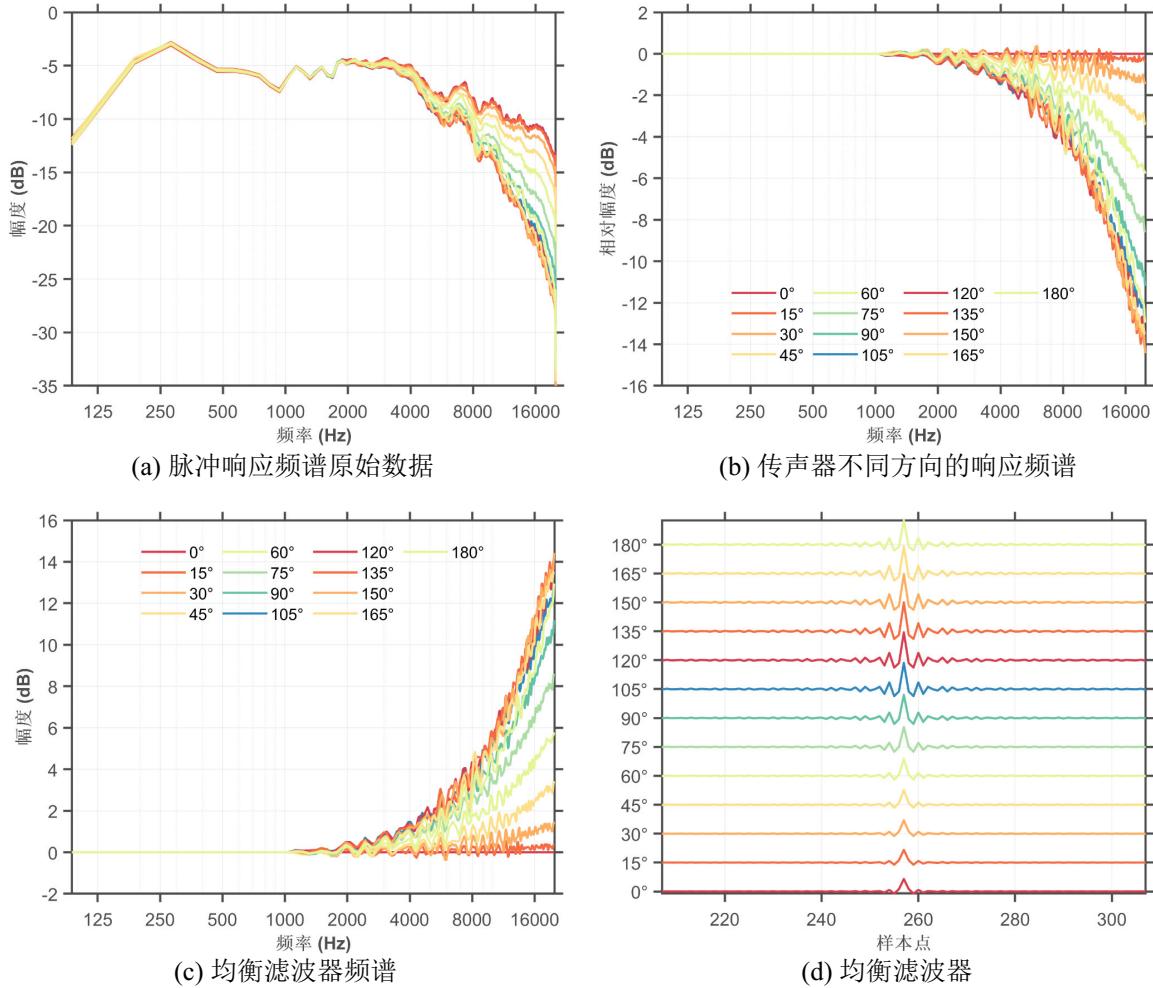


图 2-10 传声器在不同方向上的频率响应以及对应的均衡滤波器

Fig. 2-10 Frequency responses of microphone pointing to different directions and corresponding equalization filters

2.3.2.4 多通道扬声器均衡

基于上述两节的测量结果, 对已架设好的多通路扬声器进行均衡处理, 均衡旨在消除扬声器通路之间的延时、幅度以及相位差异。在扬声器阵列的水平面扬声器圆环上, 沿两条相互垂直的直径各布设一条细绳, 通过两绳的交叉点确定扬声器阵列中心位置, 在测量过程中, 传声器被置于该中心位置。传声器连接至适调放大器 (Brüel & Kjær Type Nexus) 进行信号放大及供电, 模拟信号从适调放大器传输至 16Line, 转化为 Dante 格式的数字信号, 进一步在音频路由网络中传输至 D64R, 转为 MADI 格式音

频信号后回到声卡，进而传输至电脑进行数字信号处理与分析。若无特殊说明，本文后续使用自由场传声器对该系统重放信号的校准与测量的实验均采用此硬件系统与连接方式。

图 2-11 展示了通道 1 扬声器的脉冲响应测量结果，图中绘制了脉冲响应的时域信号与扬声器的频率响应分析结果。

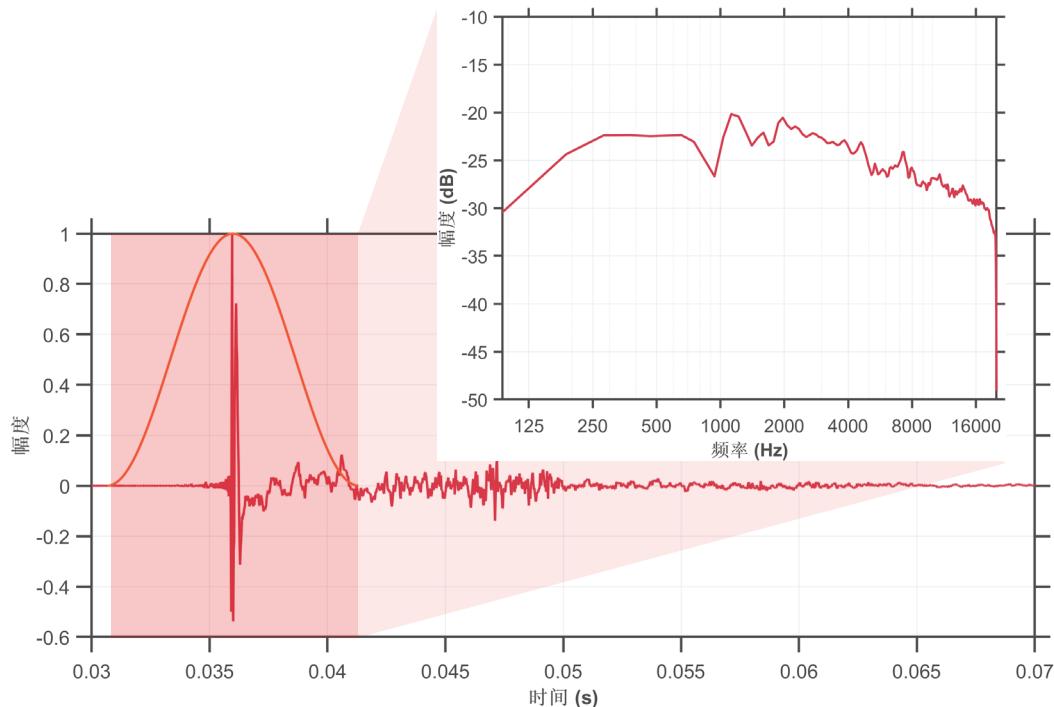


图 2-11 通道 1 扬声器的脉冲响应及其频谱

Fig. 2-11 The impulse response and spectrum of the channel-1 loudspeaker

可以看到，测量脉冲响应中，除直达声外还有一定的反射声，这些反射声来自于地面、其他扬声器以及支架结构。本研究中的均衡只针对扬声器的本身频率响应特性，只对直达声进行处理，因此首先对脉冲响应进行加窗处理（如图 2-11 红线所示），然后计算加窗后直达声的频谱。多通道均衡流程如下所示：

- (1) 首先使用对数正弦扫频信号依次测量所有通道扬声器的脉冲响应；
- (2) 使用峰值检测估计各通道延时，计算所有通道的最低延时。以最低延时处为中心，保留所有通道左右各 512 点的脉冲响应。分别对各通道以直达声为中心进行施加 512 点汉宁窗，将窗函数以外的所有点置零；

- (3) 计算加窗后脉冲响应的复数频谱 $H(f)$, 对频谱求逆, 即 $1/H(f)$, 然后进行傅里叶逆变换。为了避免扬声器低频和高频输出上下限 $H(f)$ 接近 0, 导致其倒数 $1/H(f)$ 过大, 使用 90 至 20000 Hz 的带通滤波器进行幅度限制。对计算得到滤波器再次加窗, 同时增加 256 点延时以保证滤波器的因果性, 得到多通道均衡滤波器组;
- (4) 对通道 1 的均衡滤波器能量进行归一化, 保证信号通过滤波器后总能量不变, 将该归一化系数施加到整个滤波器组, 最后的均衡滤波器为 1024 点的 FIR 滤波器, 其中 512 点用于对扬声器相位、幅度补偿, 额外的 512 点用于容纳滤波器对通道间延时的补偿。

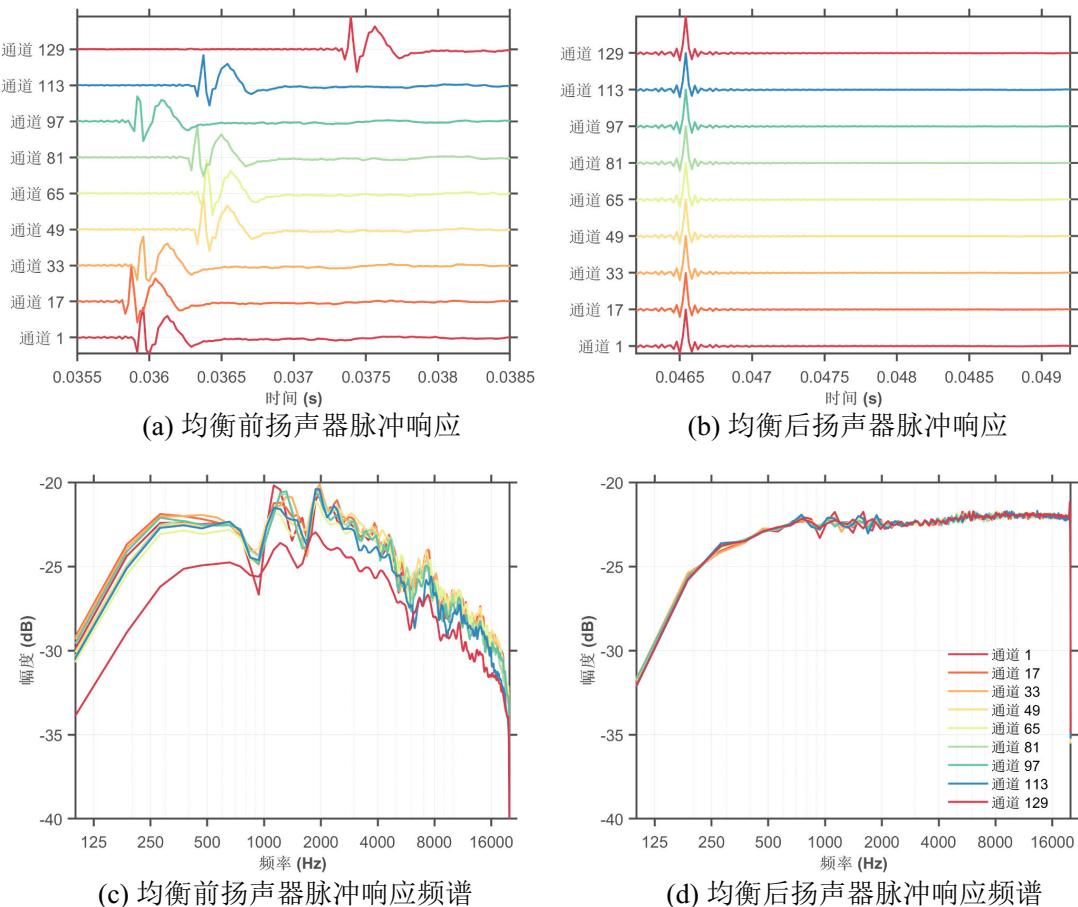


图 2-12 均衡前与均衡后扬声器的脉冲响应及其频谱

Fig. 2-12 Impulse responses and spectral of loudspeakers before and after equalization

上述过程同时对多通道扬声器的幅度谱、相位以及整体延时都进行了均衡处理，将该均衡滤波器组施加到对应通道后，再次测量脉冲响应，得到均衡后的扬声器频谱。为便于展示，图 2-12 绘制了系统中 8 个通道均衡前后的脉冲响应及其频谱。

可以看到，均衡前，8 个通道的扬声器播放信号到中心点的延时存在较大误差，且扬声器频谱不平直，不同通道直接的频谱存在一定差异。且对于每个音箱，其相位没有完全对齐（表现为时域脉冲具有两个峰值），这是由于音箱为二分频设计，具有两个发声单元，发声单元到中心点存在一定的距离差异。经过均衡后，脉冲响应的时域形状接近一个经过带通滤波的冲激响应。通道间的延时，能量大小均被对齐到同一水平，且所有通道的频谱在有效频带内接近均接近平直。在本文的后续内容中，如无特殊说明，均采用该滤波器组对重放进行均衡处理。

2.3.3 可听化重放实现

在 2.3.2 节所述的硬件系统基础上，本文使用 MATLAB 开发高阶 Ambisonics 可听化系统的软件部分，系统架构如图 2-6 所示，系统实现了对预设的声源方向及干信号的空间编码处理，得到 Ambisonics 信号，然后基于扬声器阵列的空间分布，生成用于多通路扬声器阵列的空间可听化信号。软件部分围绕高阶 Ambisonics 编解码算法，主要包含以下几个模块：（1）高阶 Ambisonics 方向信号编码及多通道解码重放；（2）FOA 传声器及球形传声器阵列信号处理；（3）房间反射信号处理。本文研究中主要使用模块（1）与模块（3）。模块（1）与模块（2）的算法处理流程已在 2.2 节进行介绍。本节重点介绍系统的模块（3），即处理房间反射信息实现室内声场可听化的方法。

基于双耳或单通道重放的室内声场可听化，通过对目标空间进行实地测量或仿真（计算机仿真或缩尺模型测量）得到对应的双耳房间脉冲响应（Binaural room impulse response, BRIR）或单通道 RIR，然后卷积任意干信号进行重放即可实现可听化，上述过程只涉及测量或计算空间上单处或两处接收点的脉冲响应。对于 Ambisonics 系统，实现可听化则需要获取 Ambisonics 各阶信号对应的房间脉冲响应，即 Ambisonics 房间脉冲响应（Ambisonics room impulse response, ARIR），在重放时，使用干信号卷积 ARIR 便可以获取包含房间反射信息的 Ambisonics 编码信号，然后对编码信号进行多通道解码实现可听化重放。

和传统的 BRIR 获取方式类似，ARIR 也可以通过实地测量或计算机仿真获取。对于测量，使用 FOA 传声器或是球形传声器阵列，按照 2.3.2.2 节所述的脉冲响应测量方法测量各通道在房间中的脉冲响应，然后按照 2.2.5 节所述的传声器阵列信号处理流程

对多通道脉冲响应进行处理，计算得到 ARIR。也可以通过计算机仿真获取各阶 Ambisonics 脉冲响应，目前部分仿真软件，如 Odeon, CATT-Acoustic 等几何声学仿真软件支持对不超过 3 阶的 ARIR 进行仿真计算。本文所建立的 HOA 系统的重放阶数远高于 3 阶，需要更高阶的 ARIR，因此不能直接使用 Odeon 等软件直接计算得到。Favrot 等人提出对 Odeon 计算的早期反射信息与后期反射声能密度进行后期信号处理，重新编码后得到高阶 ARIR^[72]。本研究参考 Favrot 提出的方法，实现的室内声场可听化重放流程如图 2-13 所示。

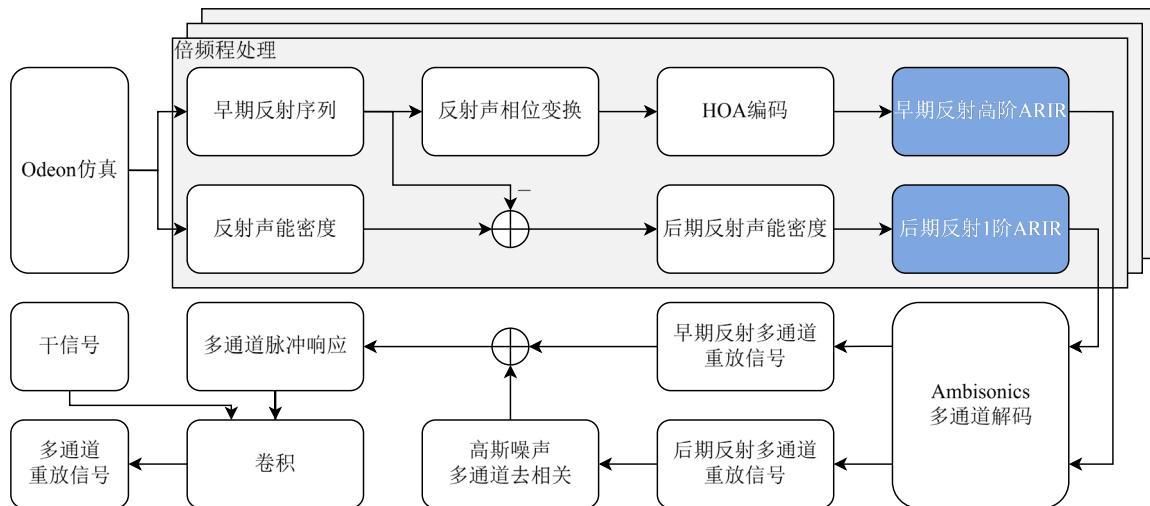


图 2-13 室内声场可听化信号处理流程

Fig. 2-13 Signal processing flow of indoor sound field auralization

将房间模型输入 Odeon 进行几何声学仿真后，计算得到：（1）63 至 8000 Hz 范围内倍频程上共 8 个频率的直达声以及早期反射声的到达时间及相对强度；（2）反射声能随时间变化的统计曲线（沿 $x-y-z$ 三个方向）。前者由虚源法仿真得到，后者由声线追踪计算得到。然后分别对上述两组数据进行处理，分别得到直达声与早期反射声部分的 ARIR 以及后期反射声部分的 ARIR：

（1）对于直达声和早期反射声，根据倍频程能量首先构建直达声及每个反射对应的脉冲响应，对于 63 Hz 以下及 8000 Hz 以上部分的能量，使用 63 Hz 及 8000 Hz 的数据。由于 Odeon 在计算反射时不考虑反射面阻抗的复数部分，即无相位信息，所有脉冲均为线性相位脉冲，计算时随机对部分反射进行相位反转，使整体反射序列具有一定的相位变化，更符合实际情况。然后，

对直达声和每个早期反射的脉冲，根据其方向进行 HOA 编码，得到早期反射的高阶 ARIR；

- (2) 对于后期反射声，首先将反射声能密度减去直达与早期反射声能密度，得到后期反射声能随时间的变化曲线。该数据是反射声在 $x-y-z$ 三个方向的统计量，不包含每个反射的具体方向信息。根据 2.2.3 节所述的 Ambisonics 编码原理，该信号即为后期反射的一阶 ARIR。

然后根据扬声器阵列，对早期反射的高阶 ARIR 和后期反射的一阶 ARIR 分别进行多通道解码。对于直达声和早期反射，直接根据重放扬声器阵列进行 Ambisonics 信号解码。对于后期反射，首先进行解码，解码后的多通路信号仅为反射的包络，不包含相位信息，为避免梳状滤波效应，对每个通道的解码后期反射信号分别乘上随机的高斯噪声，进行去相关处理。最后，合并二者得到完整的多通路空间重放脉冲响应，卷积任意干信号实现室内声环境的可听化重放。

上述过程中，由于 Odeon 在仿真计算后不能获取每个后期反射的具体方向，后期反射的编码仅能根据统计结果实现一阶重放，后期反射的重放分辨率相对较低。但在一般的房间中，后期反射通常被认为近似扩散场分布，根据空间听觉中的优先效应 (Precedence effect)^[104]，听觉难以对此类声音进行定位，因此，后期反射编码的空间分辨率可以适当降低。目前的大部分可听化系统对后期反射的处理均采用类似手段^[76]。

2.4 本章小结

本章首先介绍了高阶 Ambisonics 声场重构的基本原理，包括平面波和点声源的声场 SHT 与 SHT 逆变换、基于球谐函数分解声场的数学基础、Ambisonics 编码解码过程、以及重放阶数与扬声器阵列分布对重放影响的问题，并介绍了矩阵条件数用于评估系统稳定性，为硬件设计提供了理论支撑。同时还介绍 Ambisonics 信号拾拾理论。

基于高阶 Ambisonics 理论，建立了基于高阶 Ambisonics 的可听化系统。硬件系统以 192 通道球形扬声器阵列为核心，阵列布置在半消声室环境以控制反射干扰，利用 MADI 与 Dante 协议构建低延迟数字传输链路。进一步采用对数正弦扫频法测量脉冲响应，对系统进行了全面的校准。首先对传声器的指向性进行了测量，并构建 FIR 滤波器补偿传声器的指向性。进一步对扬声器进行多通道测量，构建均衡滤波器组，实现幅度、相位及延时的全局校准，确保声场重构的频谱一致性。系统软件部分实现了根

据目标声源方向及信号进行空间声重放。在室内声场可听化实现中，使用 Odeon 计算房间早期反射声与后期反射能量，然后分别进行处理合成 ARIR。

该系统具有实现自由场与室内声场的高精度重放的功能，但听众在该系统所重放声场下的具体感知情况，仍然需要通过开展主观实验进行评估。进而确定重放参数对实际重放效果的影响，实现对系统的优化改进。

第三章 高阶 Ambisonics 阶数对声场重放误差的影响及分析

第二章介绍了 Ambisonics 声场重构的基本原理，依此建立了一个高阶 Ambisonics 可听化重放系统，用于建筑声学研究应用。高阶 Ambisonics 重放时，由于阶数截断，声场重构存在空间混叠失真的误差问题，该误差可能会导致重放声环境下感知与实际环境存在差异，但声信号存在较大冗余性，在特定感知层面上，重放误差不一定会带来明显的可感知差异。本章针对高阶 Ambisonics 重放阶数对声场重放的误差与对主观感知的影响进行分析与研究。

3.1 引言

上一章中，基于 Ambisonics 声场重构的基本理论，建立了高阶 Ambisonics 可听化重放平台。根据式(2-1)与式(2-5)，Ambisonics 声场重构的基础在于，三维空间中由任意平面波或球面波声源形成声场的全部信息，可以由单一球面上的声压完全表示。因此，控制单一球面上的声压，即可产生任意目标声场。由于实际系统只能对球面声压在离散且有限数量的采样点上进行控制，根据所能达到的最高阶数，实际重建声场与目标声场相比存在大小不一的误差，随着阶数提高，重构误差降低。对于低阶 Ambisonics 系统重放，低阶球谐函数的指向性相对较弱，因此重放特定方向声源时，目标声源以外的扬声器具有更大的能量，即存在更大的能量扩散。已有研究表明，低阶重放的虚拟声像较为模糊，定位效果差，随着阶数升高，定位性能逐渐提高^[61]。Ambisonics 重放的阶数以及声场重构误差也可能对言语清晰度造成影响，降低重放声环境下的语音感知的真实性。语言清晰度是室内声学设计的重要指标，特别是教室、会议厅等建筑环境。保证听众在可听化虚拟声环境下具有与实际声学环境下具有近似的语音感知结果，是可听化系统的重要目标。

综上所述，本章开展实验研究 Ambisonics 重放阶数对声场重放误差及其对感知的影响。本章的内容安排如下：3.2 节介绍了首先从客观层面分析了重放阶数对声场重构误差的影响，并进行了仿真实验验证。3.3 节介绍了阶数对主观感知影响的相关研究与背景；3.4 节设计并开展了实验，研究 Ambisonics 重放阶数及误差对言语接受阈值这一主观测试指标的影响，3.5 节分析并讨论了实验结果；最后，3.6 节总结了本章的研究。

3.2 Ambisonics 重放阶数对声场重放的误差影响的模拟分析

在 2.2.3 节中提到，使用扬声器阵列进行声场重构时，重构阶数受到扬声器数量的限制，只能取有限阶 N 。重放阶数会影响声场重构的精度，阶数越高，声场重放精度越高。Ambisonics 的理论基础是对声场进行频率-空间的 SHT 变换，声场重构的精度主要体现在重构的空间声场分布与声场中接收点的频率响应准确度两个方面。为研究重放阶数对 Ambisonics 重放性能的影响，首先对本文所建立的多通道扬声器阵列进行模拟，对比 1 阶、3 阶、7 阶与 9 阶重放下的空间声场合成与接收信号的频谱。

首先对理想声源的目标声场与 Ambisonics 重放的合成声场进行模拟。理想声源与重放扬声器均假设为平面波模型，平面波解析表达式如式(1-6)所示。若声源位于 Ω_s ，扬声器阵列各扬声器 l 坐标为 Ω_l ，空间中任意位置处的目标声场 P_s 和由重构声场 P_v 为：

$$\begin{cases} P_s(\Omega_s, \Omega, k) = e^{-j\vec{k}(\vec{r}_s - \vec{r})} \\ P_v(\Omega_1, \dots, \Omega_L, \Omega, k) = \sum_{l=1}^L A_l e^{-j\vec{k}(\vec{r}_l - \vec{r})} \end{cases} \quad (3-1)$$

其中， A_l 为扬声器 l 的合成位于目标声源方向的虚拟源时的增益，计算方法见式(2-11)。

声场合成通过声场合成工具箱（SFS Toolbox）^[105]完成。水平面 $2 \text{ m} \times 2 \text{ m}$ 区域内的目标声场与重构声场在 1000 Hz 时的状态如图 3-1 所示，声源与合成虚拟声源的方向均为 $(15^\circ, 0^\circ)$ ，图中黑色原点为目标声源或扬声器。图中还给出了声场的合成误差 $|P_s - P_v|$ ，蓝色虚线为误差分布的等高线。

可以看到，随着阶数提高，合成声场 P_v 在空间中的分布更接近目标声场 P_s ，准确重构的声场区域随阶数提高逐渐增大，在一阶重放时，准确重放区域较小，有效重构范围的半径仅为 5 cm 左右，当提高阶数至 9 阶时，准确重构的区域半径增加至 50 cm 左右。在准确重构区域以外，空间混叠造成的重放误差与声源方向有关（图 3-1 (d)、(e)、(h)、(i)），在声源方向的路径上，重构误差相对较低，这一结果与 Matthieu 等人对 Ambisonics 合成声场的实测结果类似^[53]。

上述结果展示了 Ambisonics 重放声场在空间中的误差分布。在大多数情况，听众位于阵列中心，通过双耳接收声压信号，进一步对中心听音情况下的双耳接收 HRTF 信号进行分析。使用通过边界元模拟计算得到的 KEMAR 人工头的 HRTF 数据集^[106]，对扬声器重放进行模拟。该数据集在方位角和仰角上为全空间均采样，采样间隔为 1° ，

包含正上方与正下方的采样点。通过 HRTF 模拟得到理想目标声源、1 阶、3 阶、7 阶与 9 阶 Ambisonics 在双耳处产生双耳接收信号频率响应 $HRTF_s$ 与 $HRTF_v$ 如下式所示：

$$HRTF_s = HRTF(\tilde{\Omega}_s)$$

$$HRTF_v = \sum_{l=1}^L A_l \cdot HRTF(\tilde{\Omega}_l) \quad (3-2)$$

其中， $\tilde{\Omega}_l, \tilde{\Omega}_s$ 分别是 HRTF 数据集中，与扬声器方向 Ω_l 和目标声源方向 Ω_s 最接近的方向，最大误差不超过 0.5° 。 A_l 为扬声器 l 的合成位于目标声源方向的虚拟源时的增益，具体计算方法见式(2-11)。

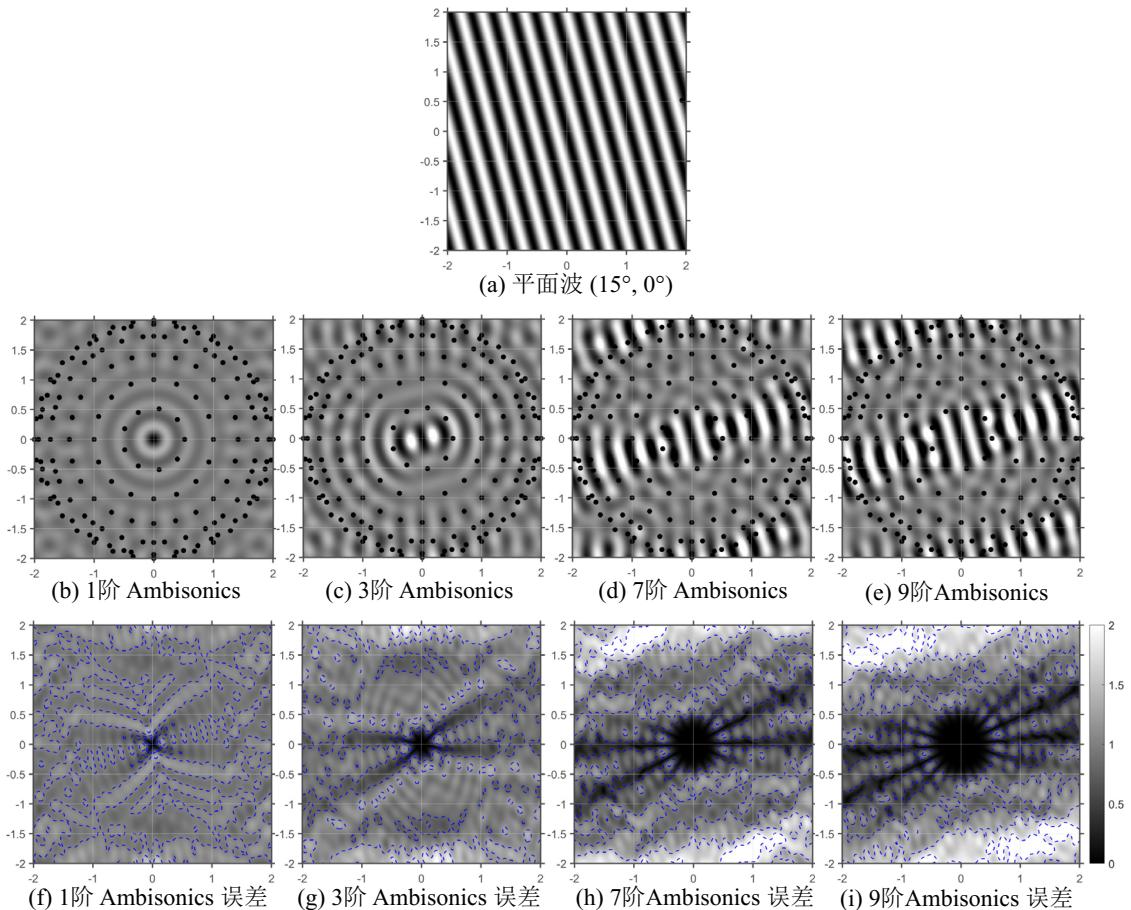


图 3-1 不同重放阶数 Ambisonics 重构($15^\circ, 0^\circ$)方向虚拟声源声场的结果及其误差

Fig. 3-1 Reconstructed sound fields and corresponding errors evoked by virtual sound source at ($15^\circ, 0^\circ$) using Ambisonics with different orders

位于正前方($0^\circ, 0^\circ$)的虚拟声源与合成声源，在左耳处的 HRTF 频谱如图 3-2 所示，图中还计算了重放 HRTF 相比目标 HRTF 的误差，不同频率 f 误差 $error$ 的计算方式为：

$$error(f) = 20 \log_{10} (|HRTF_s(f)| / |HRTF_v(f)|) \quad (3-3)$$

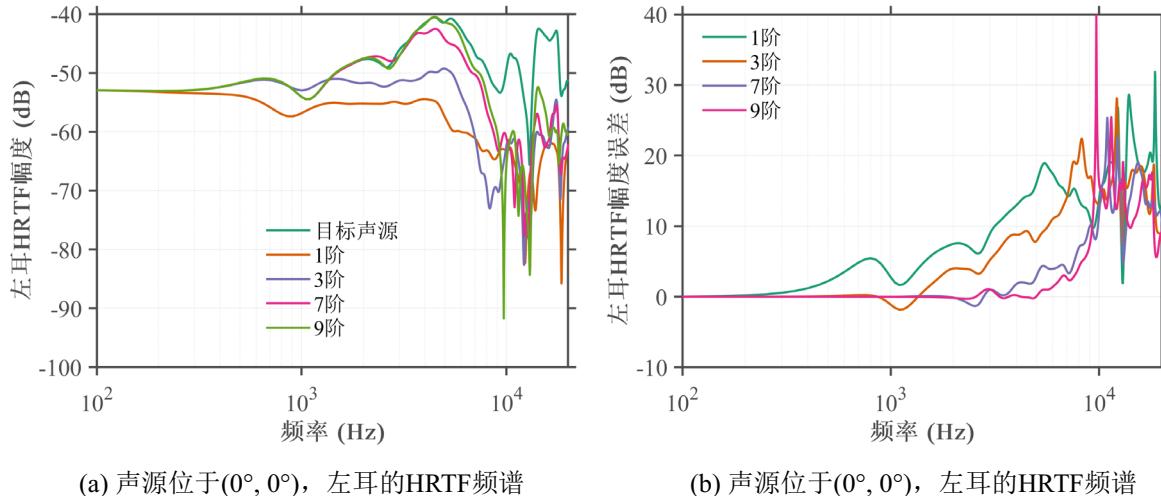


图 3-2 正前方目标声源以及不同阶数 Ambisonics 重放合成声源在左耳处的 HRTF

Fig. 3-2 The left-ear HRTFs of front virtual sources reproduced by different orders Ambisonics and the front target real source

从图中可发现，随着阶数上升，合成声源对应的 $HRTF_v$ 更接近目标声源 $HRTF_s$ 。对于低阶重放，HRTF 在中高频部分与目标声源的 HRTF 差距较大，特别是，在高频部分，低阶重放对应的 HRTF 存在明显的梳状滤波效应，相比目标 HRTF，存在大量峰值与谷点（图 3-2 (b)），这是由于低阶重放时，重放虚拟声源方向周围的扬声器也具有较大的能量，而高阶重放时能量则主要集中在目标方向，即能量扩散较小，多扬声器从不同方向同时播放时，叠加产生的信号是不同方向 HRTF 的合成平均，受到所有方向不同位置高频峰谷的影响。

计算不同阶数与参考信号相比的全频带谱失真（Spectral distortion, SD），SD 计算方法为：

$$SD = \sqrt{\frac{1}{N} \sum_{f=1}^N 20 \log_{10} \frac{|HRTF_s(f)|}{|HRTF_v(f)|}} \quad (3-4)$$

其中， f 为频率采样点， N 为 HRTF 的总频率采样点个数。计算得到 1、3、7、9 阶 Ambisonics 重放条件下的左耳 SD 分别为 15.7、13.2、10.6 与 10.4 dB。由于语音信号主

要能量分布在 8 kHz 以下，同时计算 8 kHz 范围内的 SD (f 限制在 8 kHz 以内的频率采样点)，此时 1、3、7、9 阶 Ambisonics 重放条件下的 SD 为 7.4、5.7、1.7、0.9 dB。

进一步分析不同阶数（1、3、7、9 阶）Ambisonics 重放的能量扩散，如图 3-3 所示，图中目标声源位于正前方($0^\circ, 0^\circ$)，黑色圆点代表为扬声器位置。可以看到，随着阶数提高，能量集中在目标声源附近，距离声源较远的扬声器增益较低，对 HRTF 贡献小，因此频谱更接近理想情况。在一阶 Ambisonics 重放时，能量扩散较大，重放目标声源时几乎整个前半空间都具有较大的能量。

综上所述，提高 Ambisonics 重放阶数，能够增大有效重构声场的大小，降低接收双耳信号的高频失真。同时，重放阶数还影响能量扩散程度，在高阶重放时，能量主要集中在目标声源区域。多通道具有较大能量时，会产生较大的串扰，这些误差可能对主观感知造成不同程度和不同方面的影响。

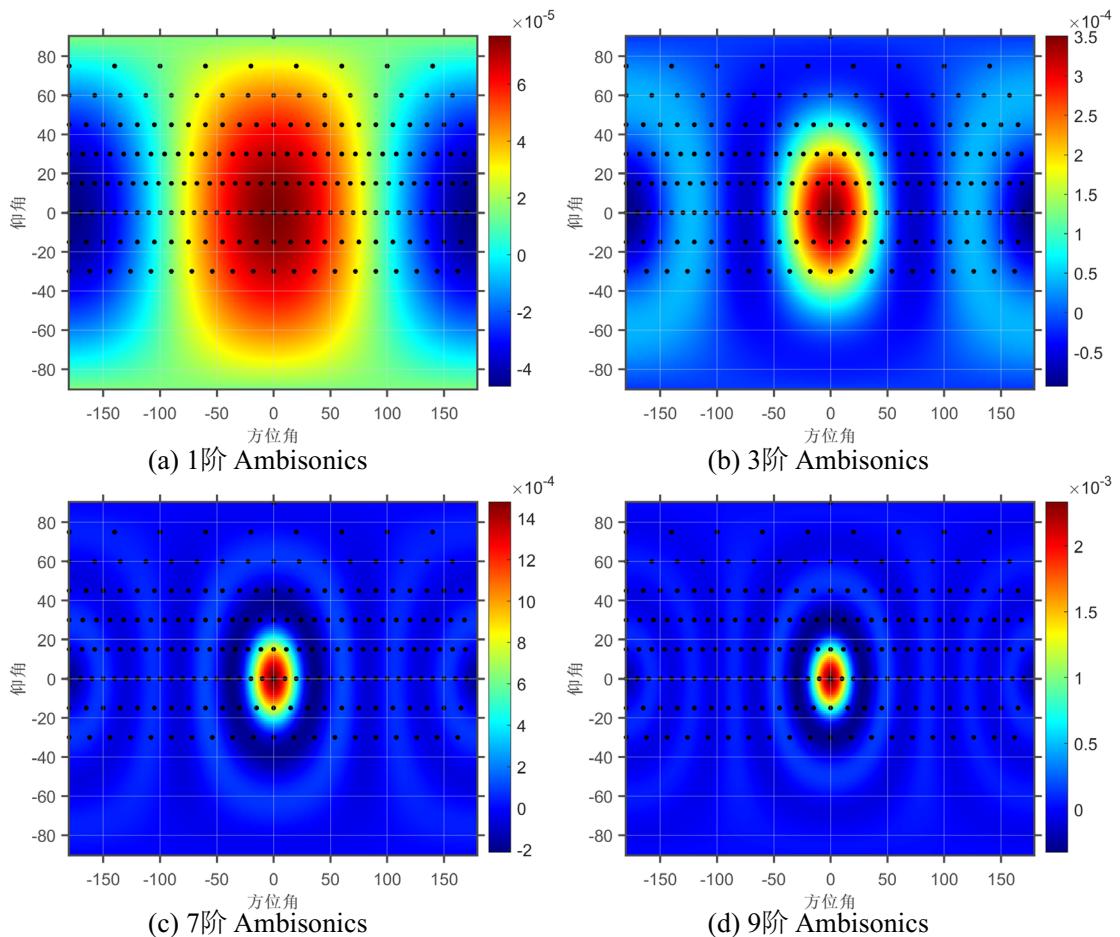


图 3-3 不同阶数 Ambisonics 重放在空间中的理论能量分布

Fig. 3-3 The theoretical energy distribution of different orders of Ambisonics in space

3.3 实验背景

3.3.1 重放阶数对感知的影响

上一节从客观层面的模拟与分析，探讨了 Ambisonics 阶数对重放误差的影响。过去的研究主要探讨了重放阶数对定位的影响，研究发现随着阶数提高，定位准确度有明显改善，需要高至 5 阶的重放才能有与双耳 HRTF 虚拟化重放接近的定位效果^[60]。部分研究探讨了不超过 3 阶重放下信号源对声音失真感知的影响，实验发现，当信号声源频带超过理论重放频率上限时，保真度差异更大，提高阶数可以降低失真^[66]。

除了空间定位与声音失真感知外，重放阶数及其误差还会对高层感知造成影响。由于声信号具有非常大的冗余性^[107]，重放阶数所带来的误差对更高层次的感知维度的影响尚不清楚。现有相关研究中，Au 等人使用一阶 Ambisonics 系统对室内声环境进行拾拾并重放，测量了虚拟混响环境下的带噪环境言语识别率，结果表明听众在虚拟声环境下也存在空间去掩蔽现象，当噪声与语音空间分离时，言语识别率相比不分离的情况下更高^[108]。Ahrens 等人对比了听众在混合阶 Ambisonics 系统（5 阶三维重放，7 阶水平面重放）的条件下，仿真并重放的室内虚拟环境下的语音信号，结果表明，虚拟环境下测量的言语接收阈值（Speech reception threshold, SRT）^[109]与实际声环境下的测试结果仍有一定的差距。此外，Oreinos 等人使用混合阶 Ambisonics 系统（4 阶三维重放，7 阶水平面重放）重构复杂声环境，对助听器降噪后的 SRT 进行评估，类似地，结果同样发现虚拟环境下测得的 SRT 与真实环境下测得的 SRT 有一定的差距^[110]。

上述研究没有直接探讨高阶 Ambisonics 系统重放阶数及误差本身对语音感知的影响。由于语音的冗余度较高，重放阶数可能对语音感知的影响相对较小。Dagan 等人的研究认为虚拟重放声场下的空间掩蔽释放（Spatial masking release, SRM）几乎不受到阶数的影响^[111]。但也有研究指出，多通路扬声器重放存在串扰现象，会对语音信号产生梳状滤波效应，降低言语可懂度。对于 Ambisonics 系统，扬声器的数量与重放阶数会对信号频谱存在不同程度的影响，低阶系统采用过多的扬声器可能会对语音频谱产生较大的破坏，提高阶数可能可以降低这一现象^[67]。上一节的模拟也显示，低阶重放下双耳接收信号的高频部分存在较大的失真，本研究针对这一问题开展实验进行研究。

3.3.2 语言清晰度评价方法

评价语言清晰度的指标主要分为两类，客观评价指标与主观评价指标：

客观指标通过对听觉过程（如听觉带宽、能量掩蔽等）或是信号的物理传输过程（如信号调制、混响等）等特定维度进行建模，根据信号失真的程度与听觉感知的灵敏度，给出具体数值，反映特定系统或条件下听众对语音信号感知的清晰度。客观指标可进一步分为两类，包括侵入式与非侵入式的指标，侵入式指标需要与参考信号（干净的语音）进行对比，非侵入式则无需参考信号^[112]。常用的客观评估指标有：清晰度指数（Articulation index, AI）、言语可懂度指数（Speech intelligibility index, SII），语音传输指数（Speech transmission index, STI）、短时客观可懂度（Short-Time objective intelligibility, STOI）等。客观指标结果相对容易得到，只需测量所需的信号并进行计算即可，但目前的客观指标均只针对部分听觉过程进行模拟，尚没有指标能对完整的听觉感知进行建模，主观听音仍是评价语音信号感知的最可靠判据。

主观指标通常需要开展主观听音实验才能获得，成本较高。主观指标直接反映了特定情况下听众对语音信号的感知情况，通常被认为是最真实可靠的数据。语言清晰度的主要主观测试指标包含：

- 言语识别率，该指标为安静环境或噪声环境下，受试者对句子中词语，或是多个句子的平均识别正确率。该指标直接反映特定环境下的言语识别情况，测试流程和数据分析相对直接，但容易出现天花板或地板效应；
- 言语接受阈（SRT），该指标为一个语音-噪声的信噪比（Signal-to-noise ratio, SNR），在该 SNR 下，受试者对语音的识别率为 50%。通常采用自适应测试或是固定刺激测试两种方法测得。自适应测试根据受试者每个试次言语识别的情况，动态地调整信噪比进行下一个试次，直至得到 50% 识别率对应的信噪比。固定刺激测试则选取多个信噪比条件测量言语识别率，然后通过曲线拟合的方式计算 50% 识别率下对应的信噪比。
- 安静环境言语接收阈（Quiet speech reception threshold, QuietSRT），该指标为一个声压级，在该声压级的语音信号下，受试者的识别正确率 k 为 50%。测试方法与 SRT 类似，但在安静环境下进行，测试时对信号的强度进行调整。

本研究旨在横向对比重放阶数及对应声场重放误差对语音感知的影响，实验在不同阶数的条件下重放自由场虚拟语音信号声源，测量受试者对语音信号的识别情况。由于自由场没有反射干扰，预实验测试结果显示，不同重放阶数条件下受试者的言语识别率都较高，为避免出现天花板效应，本研究使用更稳定的噪声下的言语接受阈 SRT 值作为主观测量指标，对比不同阶数与重放误差下言语接受阈值的区别。

3.4 重放阶数对言语接受阈值影响的主观实验

3.4.1 实验目的

本实验旨在探讨 Ambisonics 多通道重放对言语感知的影响，特别是，具有较大能量扩散的低阶 Ambisonics 重放，可能会降低重放语音信号的清晰度，提高言语接受阈值。本研究的假设是，多扬声器串扰产生的失真以及能量扩散导致的空间掩蔽释放都会对言语接受阈值造成影响。为评估清晰度的下降情况，测量了不同阶次 Ambisonics 重放条件以及单声道播放（参考）条件下的言语接受阈值。实验在自由场声源的条件下展开，噪声与语音信号均不进行室内可听化处理，以衡量 Ambisonics 重放本身对感知的影响。实验针对正常听力受试者，不考虑非中心听音的情况。

3.4.2 实验对象

本实验招募 12 名正常听力的受试者，其中包含 9 名女性，3 名男性。受试者的年龄范围为 21 至 29 岁，平均 24.5 岁，所有受试者均无中耳炎及相关耳科病史。实验开始前，使用自主开发的自动纯音测听软件进行听阈测量，测量前已对测听软件进行了校准^[113]。所有受试者在 500 到 4000 Hz 倍频程的平均气导听阈不超过 20 dB HL。受试者均在得知实验目的与大致实验内容的基础上，签署实验知情同意书后自愿参与实验，受试者根据其参与实验的时间获得了一定的实验酬金。

3.4.3 实验方法

3.4.3.1 实验设备与测试信号

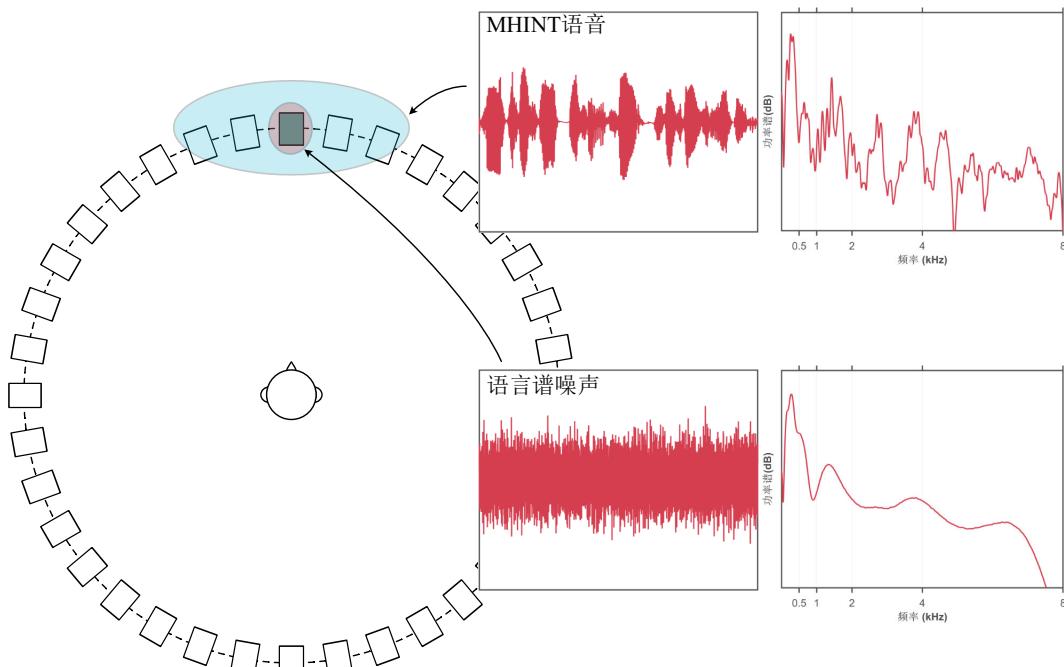
实验使用 2.3 节建立的 192 通路高阶 Ambisonics 重放系统，语音信号与噪声信号均设置在受试者前方 $\Omega_s(0^\circ, 0^\circ)$ ，如图 3-4 (a) 所示。二者在空间位置上重合，以降低空间方向本身对语音感知的影响，重点关注重放阶数及重放误差对 SRT 的影响。

其中，语音信号的信号源为汉语普通话版噪声下言语测试（Mandarin Hearing In Noise Test, MHINT）语料库中的语音信号^[114]，采样率为 48 kHz，通过 Ambisonics 编解码后进行多通道重放。噪声信号为语言谱噪声，通过计算整个 MHINT 语料长时平均谱（Long-term average spectrum, LTAS），使用 LTAS 的幅度谱对白噪声进行滤波，使该噪声具有和整个语料库的相同的平均频谱包络（见图 3-4 (a)）。

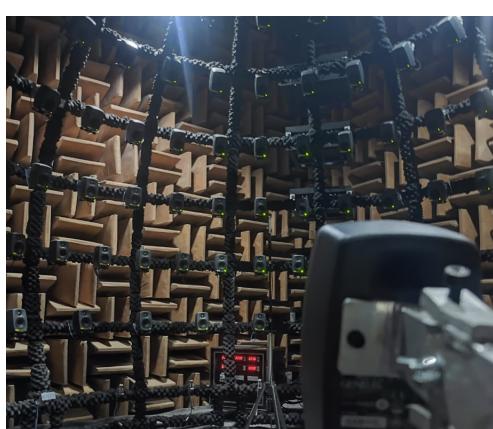
实验中，噪声信号始终由受试者正前方的单一扬声器进行播放，语音信号采用不同阶数 N 的 Ambisonics 进行编解码重放，具有大小不一的能量扩散程度。实验包含 6 测试条件，每个条件重复 2 次，测试条件如表 3-1 所示。所有条件中语音与噪声的预设

空间位置相同。其中，条件 1 为参考条件，语音信号仅从正前方的单个扬声器进行播放，条件 2 为 0 阶 Ambisonics 编码，在该编码条件下，语音信号以相同的强度馈给所有通道进行播放。

使用布置于扬声器阵列中心的自由场传声器（Brüel & Kjær Type 4189）进行声压级校准，确保不同重放条件下的语音信号与噪声信号的声压级保持匹配，测试过程照片见图 3-4 (b)。测试中噪声信号的声压级在实验中保持不变，固定为 65 dBA，语音信号的声压级在实验中随受试者反馈改变，不同条件下的声压级保持在同一水平。



(a) 实验示意图



(b) 声压级校准



(c) 实验过程照片

图 3-4 言语接受阈值测量实验示意图及过程照片

Fig. 3-4 The diagram and photos of the SRT measurement experiment

表 3-1 言语接受阈值测量的实验条件列表

Tab. 3-1 The list of experimental conditions for the SRT measurement

实验条件	语音信号	噪声信号
1	正前方单扬声器	正前方单扬声器
2	0 阶 Ambisonics 重放	正前方单扬声器
3	1 阶 Ambisonics 重放	正前方单扬声器
4	3 阶 Ambisonics 重放	正前方单扬声器
5	7 阶 Ambisonics 重放	正前方单扬声器
6	9 阶 Ambisonics 重放	正前方单扬声器

3.4.3.2 实验流程

正式实验中，每个受试者需要测量 12 次 SRT（6 测试条件 \times 2 重复）。实验前，随机选取两个试次作为训练试次，以便让受试者熟悉测试流程，训练试次的测量结果不纳入最终的实验结果统计。

表 3-2 拉丁方矩阵平衡的实验次序

Tab. 3-2 The experimental sequence balanced with the Latin square matrix

受试者	实验条件											
	1	2	3	4	5	6	7	8	9	10	11	12
S1	1	2	3	4	5	6	7	8	9	10	11	12
S2	2	3	4	5	6	7	8	9	10	11	12	1
S3	3	4	5	6	7	8	9	10	11	12	1	2
S4	4	5	6	7	8	9	10	11	12	1	2	3
S5	5	6	7	8	9	10	11	12	1	2	3	4
S6	6	7	8	9	10	11	12	1	2	3	4	5
S7	7	8	9	10	11	12	1	2	3	4	5	6
S8	8	9	10	11	12	1	2	3	4	5	6	7
S9	9	10	11	12	1	2	3	4	5	6	7	8
S10	10	11	12	1	2	3	4	5	6	7	8	9
S11	11	12	1	2	3	4	5	6	7	8	9	10
S12	12	1	2	3	4	5	6	7	8	9	10	11

SRT 测量过程中，受试者对语音信号的熟悉程度可能会影响实验结果，为避免这种影响，每个试次随机分配一组不重复的句表。MHINT 语料库包含 12 个不重复的句表，每个句表包含 20 段不重复的句子，恰好分配给每个试次。此外，语料库还有 40 个用于训练的句子。为降低实验顺序对测试结果造成显著影响，采用 12×12 的拉丁方矩阵对实验次序进行平衡^[115]。拉丁方矩阵是一个 $n \times n$ 的方阵，其中每一行与每一列都恰好包含 n 个不同的元素。本文实验使用的拉丁方矩阵如表 3-2 所示。可以看到每个实验条件在每个实验次序上出现的次数相同，避免了受试者随着实验进行产生的学习效应对不同实验条件产生有偏的影响。

实验采用自适应测试软件测量 SRT，实验软件界面如图 3-5 所示^[116]，在每次 SRT 测量中，采用一下一上的自适应台阶测试法^[117]。具体流程为：

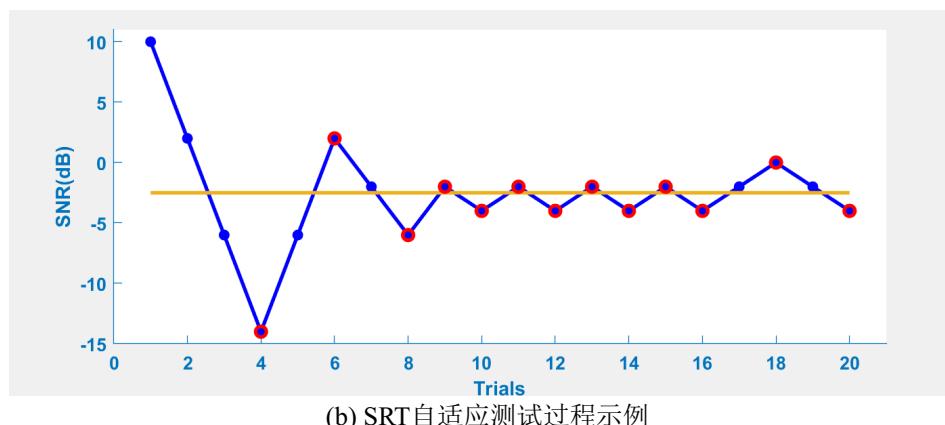
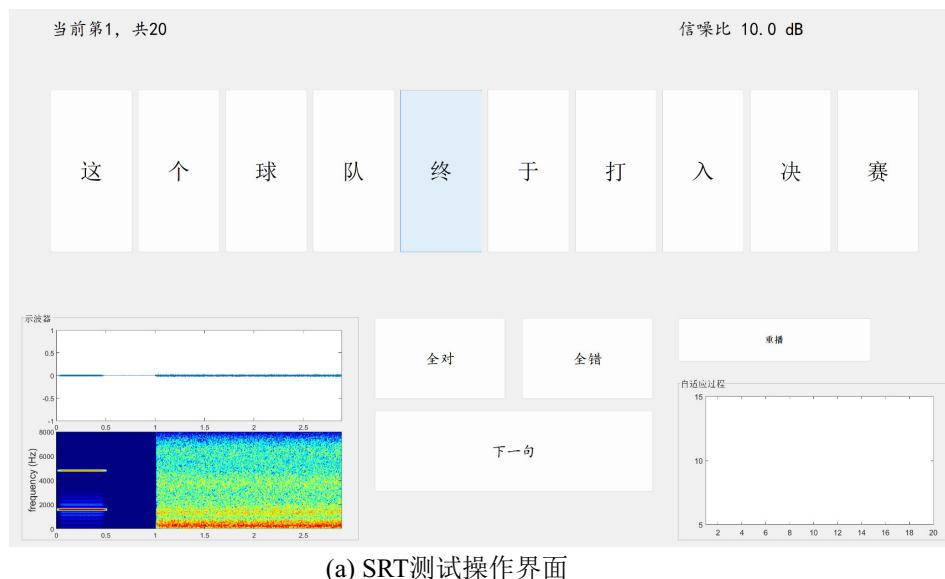


图 3-5 SRT 测量过程示例

Fig. 3-5 An example of the SRT measurement procedure

- (1) 播放带噪声的语音信号，初始 SNR 为 10 dB。受试者根据所听到语音，口头复述听到的内容，实验员根据受试者的反馈在软件界面（图 3-5 (a)）选择受试者答对的字词。MHINT 的每段语音信号包含 10 个字，若受试者答对字数不少于 8 个，则认为受试者给出了正确的反馈，否则认为受试者回答错误。语音信号播放前，有一个短暂的提示音提醒受试者注意聆听；
- (2) 然后，根据受试者回答正确与否，调整语音信号的强度，改变 SNR，若受试者答对，降低一个步长的 SNR，提高难度，否则提高一个步长的 SNR。在本实验中，步长根据实验进程发生改变，第 0 至第 2 个反转点内所有试次的步长是 8 dB，第 3 至第 4 个反转点内试次的步长为 4 dB，第 5 个反转点之后步长为 2 dB。反转点指受试者从反馈错误到反馈正确，或从反馈正确到反馈错误的情况，如图 3-5 (b) 中的红色圆圈所示。
- (3) 当所有句子（20 句）播放完成后，计算最后 8 个句子的平均 SNR 作为 SRT 测试结果。

图 3-5 (b) 展示了某个受试者一次 SRT 的测试情况，实验过程中 SNR 随着受试者的反馈正确与否进行变化，最后趋于稳定，收敛至 50% 言语识别率对应的 SNR。

3.5 实验结果

3.5.1 实验结果统计分析

在正式实验中，总共进行了 144 次 SRT 测试（12 受试者 \times 6 测试条件 \times 2 重复测试），首先将重复测试的结果进行平均，则每个测试条件共有 12 个 SRT 数据样本。实验结果如图 3-6 的箱型图所示。

图中每个箱体对应一个测试条件的统计结果，箱体的上下边界分别为数据的第 75 百分位数 (Q3) 与第 25 百分位数 (Q1)，中间的横线代表中位数 (Q2)。箱体的长度，即 Q3 与 Q1 的差值，称为四分位距 (Interquartile range, IQR)，它反映了数据的离散程度。从箱体上下边界延伸出的线段称为须，须为数据组的最大值和最小值。同时，每组内的所有数据以散点的形式显示在箱图上。

从图 3-6 中可以看到，除 0 阶条件外，随着 Ambisonics 阶数的提升，SRT 数值逐渐降低并趋近于参考条件（单扬声器播放）的结果，这表明语言清晰度随阶数升高而提升。值得注意的是，0 阶条件下的 SRT 明显低于其他条件，且具有更大的方差。从数据变异范围来看，0 阶和 1 阶条件下的 SRT 波动幅度明显高于其他高阶条件（图 3-6 中 0

阶与 1 阶条件的箱图更长）。这一现象可能与低阶 Ambisonics 系统下较小的最佳听音区域大小相关——当受试者位置偏离理想听音区域时，扬声器间串扰效应加剧，导致不同个体的感知差异显著扩大。3 阶、7 阶、9 阶与参考条件的数据变异范围基本类似，标准差分别为 1.0、1.9、0.8 与 0.6 dB。其中，7 阶重放条件存在一个离群数据，导致其标准差相对较高。

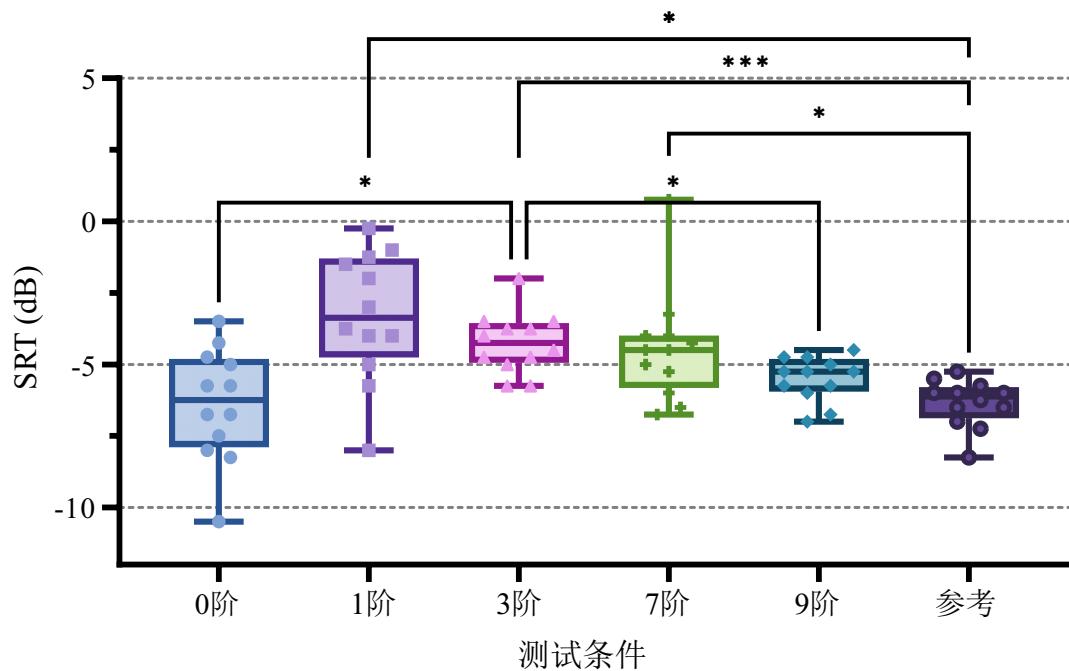


图 3-6 言语接受阈值实验结果

Fig. 3-6 Experimental results of the SRT test

为了对数据的整体分布进行分析，排除个别离群数据的干扰，进一步对数据进行统计分析，首先使用 Anderson-Darling 正态性检验检查各条件下 SRT 是否符合正态分布，检验显示所有条件下 SRT 均符合正态分布，条件 1-6（见表 3-1）对应的 p 值分别为 0.36、0.89、0.69、0.55、0.050 与 0.29，均大于 0.05，数据符合正态性分布，各组数据的 Q-Q (Quantile-Quantile) 图见图 3-7，横轴表示理论分布（如标准正态分布）的分位数，纵轴表示样本数据的分位数，若数据近似服从理论正态分布，则图中数据点应大致分布在一条直线上，从图中可以看到所有条件下的数据基本位于对角线上，说明数据分布基本符合正态分布。

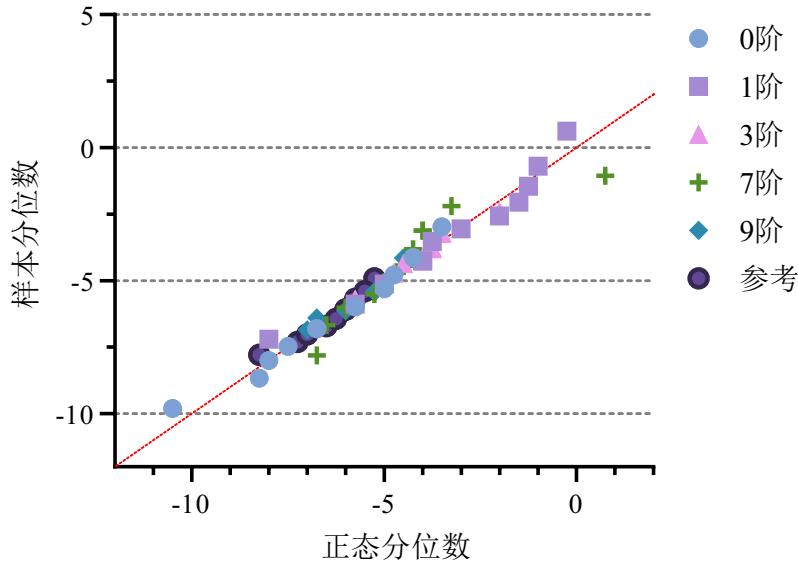


图 3-7 言语接受阈值实验数据正态性检验的 Q-Q 图

Fig. 3-7 Q-Q plot of normality rest for SRT results of the experiment

然后，对符合正态分布的 SRT 数据进行单因素重复测量方差分析（repeated measures ANOVA, rm ANOVA），检验测试条件（重放阶数）是否显著影响 SRT。由于数据不满足球形假设，采用 Geisser-Greenhouse 纠正对 rm ANOVA 进行球度纠正。rm ANOVA 统计结果如表 3-3 所示，结果表明不同实验条件对 SRT 具有显著影响 [$F(2.651, 29.16) = 7.652, p < 0.001$]。进一步通过 Tukey 纠正的两两比较进行事后检验，对比各条件之间的显著性差异，结果（见图 3-6 中的黑色连接线）显示：

- (1) 1 阶、3 阶与 7 阶重放下的 SRT 显著高于参考条件（1 阶 $p < 0.05$ 、3 阶 $p < 0.001$ 、7 阶 $p < 0.05$ ），且 7 阶重放与 3 阶重放无显著区别，这意味着受试者在 1 阶、3 阶和 7 阶 Ambisonics 重放条件下的清晰度显著低于单扬声器播放的参考场景，7 阶重放在单一声源言语感知这一感知情况下与更低阶的重放差距较小，不足以实现可靠的重放感知；
- (2) 9 阶重放的 SRT 与参考条件无统计学显著差异 ($p = 0.07$)，二者平均差为 0.85 dB，这表明高阶（9 阶）系统重放语音信号在清晰度上已接近单声道播放的水平；
- (3) 3 阶重放条件下的 SRT 显著高于 9 阶重放 ($p < 0.05$)，说明从 3 阶到 9 阶的阶数提升对提升言语可懂度具有显著效果。但 1 阶和 9 阶重放并无显著区别，但二者平均 SRT 差值为 2.2 dB，差距较大。这是由于 1 阶重放下 SRT

的波动范围较大，部分受试者在1阶重放下具有更高的清晰度，但这一结果不稳定；

- (4) 尽管0阶重放的SRT均值明显低于其他条件，但在统计上，0阶重放SRT仅显著低于3阶重放($p < 0.05$)。

表 3-3 言语接受阈值测量实验的 rm ANOVA 统计结果

Tab. 3-3 Statistical results of the rm ANOVA for SRT measurement experiment

因素	SS	自由度 (dF)	F	p
测试条件	933.85	5	7.65	< 0.001
个体间效应	32.32	11	1.12	0.31
残差	134.90	55		

结合3.2节计算的客观频谱失真SD，分析SD(全频带与8kHz以下频带)与SRT主观测量平均结果的相关性，结果如图3-8所示。图中横坐标为0、1、3、7、9阶Ambisonics重放的频失真SD，纵坐标为对应条件下的平均SRT结果。可以看到，除0阶Ambisonics条件外，其余重放条件SRT与SD基本符合线性正相关，使用线性拟合对SRT与SD进行拟合，拟合优度 R^2 均较高，为0.79。

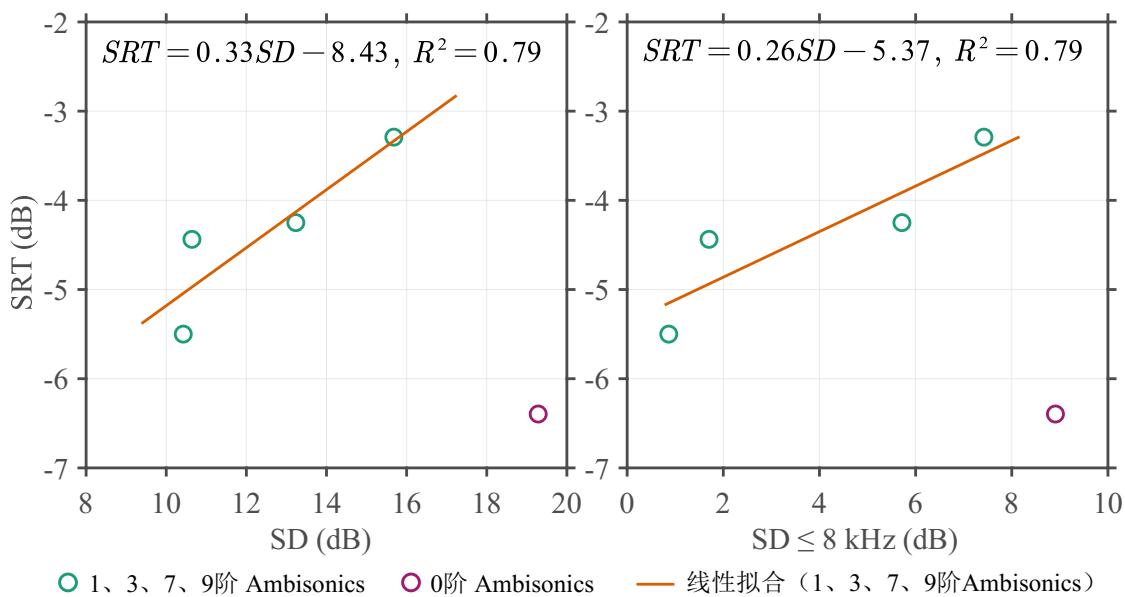


图 3-8 SRT-SD 的线性拟合结果

Fig 3-8 Linear fitting results of SRT to SD

3.5.2 实验结果讨论

听众在低阶 Ambisonics 系统（1 阶、3 阶）下的言语可懂度下降的原因，可能与多扬声器间的信号串扰失真有关，这一结论与立体声系统及 Ambisonics 系统重放下的相关研究结果一致^[67,118]。当多个扬声器同时播放同一个信号时（Ambisonics 重放仅调整不同扬声器的幅度），声波叠加产生的相位干扰会模糊语音信号的时域和频域特征，进而增加听众对语音的识别难度。

值得注意的是，本研究结果与 Ahrens 等人的结论存在差异——他们的研究发现，当噪声与语音在空间位置上重合时，Ambisonics 阶数对 SRT 无显著影响^[119]。这种差异可能源于实验设计上的区别：本实验中噪声固定由前方的单扬声器播放，而语音通过不同阶数的 Ambisonics 系统播放；Ahrens 实验中语音噪声二者均使用 Ambisonics 进行编解码播放。本研究中，语音信号经过 Ambisonics 重放后，存在一定的能量扩散，部分扩散的语音信号和噪声信号会发生一定的空间分离，这种空间分离可能产生一定的空间掩蔽释放（SRM）。具体来说，在低阶条件下，语音能量扩散范围较大，可能通过 SRM 部分抵消噪声的干扰，这或许解释了 0 阶重放条件下 SRT 未随阶数降低而升高的异常趋势，这一结论与 Ahrens 的实验一致，他们的实验发现当语音和噪声分离时，阶数对 SRT 有显著影响，说明能量扩散会影响 SRM。另一方面，1 阶和 3 阶重放相比更高阶的重放，具有更强的 SRM 作用，但这两个条件下的 SRT 更高，清晰度较低，这也侧面印证了多通道串扰对言语接受阈值的影响比 SRM 的影响更强。

本研究通过对比不同阶数 Ambisonics 系统下测量的 SRT 发现：随着阶数提高，重放能量集中在目标区域，使得扬声器多通道串扰降低，能量扩散与重放误差降低，需要高至 9 阶重放，才能使测量的 SRT 接近真实声源参考水平，而低阶系统（如 1 阶、3 阶）因扬声器串扰导致可懂度显著下降。低阶系统的能量扩散特性可能同时引起较为明显的 SRM。在使用多通道 Ambisonics 重放的语音信号清晰度测试感知实验中，应对这种双重作用机制予以特别关注。未来基于 Ambisonics 重放技术开展语音感知实验时，需根据研究目的谨慎选择系统阶数：对于需要精确空间定位的场景，高阶系统更能保证数据有效性；而低阶系统的应用需严格控制受试者位置，避免多通道串扰和有限的听音区域所带来的严重影响。需要注意的是，该实验的结果仅针对单语音-单噪声下的 SRT 感知情况，此时需要高至 9 阶的重放才能使虚拟重放下的感知与参考条件接近，对于其他感知场景，声场重放误差可能具有不同的影响，所需阶数可能不同。

3.6 本章小结

本章围绕高阶 Ambisonics 重放阶数对重放误差与主观感知的影响展开研究，着重研究了阶数对言语接受阈值的影响。首先通过仿真模拟的客观分析手段，分析了阶数对重构声场、双耳接收信号的频谱以及能量扩散三个层面的影响。进一步使用高阶 Ambisonics 重放系统，设计心理声学实验，测量了 12 名受试者在 0 阶至 9 阶 Ambisonics 重放条件及单声道参考条件下的 SRT，分析阶数对言语接受阈值的影响机制。实验结果表明，除 0 阶外，SRT 数值随阶数提升逐渐降低，并趋近单声道水平。其中，1 阶、3 阶与 7 阶重放的 SRT 显著高于参考条件，9 阶重放与参考条件无显著差异，低阶系统数据变异幅度明显更大。低阶系统重放下清晰度下降可能源于多扬声器串扰导致的频谱失真，而 0 阶 SRT 条件下异常较低的 SRT 数值或与低阶系统能量扩散引发的 SRM 效应相关。研究揭示了高阶 Ambisonics（如 9 阶）重放的虚拟声场在可以获得接近真实声源的 SRT 测量结果，低阶系统则因多通道串扰和听音区域限制的问题，对感知结果具有较大的影响。这些实验结果为后续可听化系统设计及语音感知实验的阶数选择提供了理论依据。

第四章 分层扬声器阵列下的重放算法改进

第三章针对 Ambisonics 阶数对重放误差与言语接受阈值的感知影响进行了评估实验，实验结果显示，重放阶数是影响虚拟环境下言语感知可靠性的重要因素，需要高至9阶的重放才能使的虚拟语音声源清晰度与参考声源一致。然而，实现高阶重放需要大量空间均匀分布的扬声器，这在实际系统上难以实现。针对这一问题，本章提出了一种新的重放算法，算法针对空间分层的扬声器阵列结构进行了优化，降低了算法对扬声器数量及空间均匀性的要求。

4.1 引言

在 2.2.3 节介绍的高阶 Ambisonics 编解码原理中提到，三维（3-D）Ambisonics 声场重构，根据所重放阶数 N ，扬声器数量不能低于 N 阶球谐函数的总谐波数量 $(N+1)^2$ 。同时，重放还要求扬声器在空间中均匀分布。扬声器数量要求 $(N+1)^2$ 是阶数 N 的二次项，随着 N 提高迅速增大。另一方面，人类听觉对于声源的仰角方向分辨精度远低于水平面上声源的方位角分辨率，在水平面上正前方，方位角的可辨别阈（Just noticeable difference, JND）可低于 1° ^[120]，而水平面正前方仰角方向的 JND 则约为 3.7° ^[121]。此外，对于声源方位角定位任务，在高仰角的平面上，方位角误差相比低仰角平面上的误差更高^[95]。综上所述，三维 Ambisonics 重放算法对扬声器的要求与实际扬声器阵列（如 NHK 22.2 系统）的设计以及听觉系统的感知特性存在矛盾：（1）安装扬声器时，在水平面搭建数量较多的扬声器相对容易，在其他仰角，特别是在高仰角区域（如 60° 至 90° ），安装难度相对更高；（2）实际扬声器阵列通常是分层布置的（即扬声器仅分布在离散的仰角 ϕ 上），如在水平面、 30° 、 60° 等特定仰角平面布置数圈扬声器阵列；（3）听觉系统对水平面声源的方向最敏感，对其他高度的声源方位分辨能力相对更差，听觉对声源的仰角分辨率也低于方位角分辨率。

对于二维（2-D）高阶 Ambisonics 重放，如 2.2.3 节所述， N 阶二维重放仅要求至少使用 $2N + 1$ 个扬声器，且在水平面布置大量均匀的扬声器相对容易，能够实现高阶二维重放。然而，二维 Ambisonics 仅能重放位于水平面的声源，重放三维声源是 Ambisonics 算法的主要优势之一^[41]，对可听化重放具有重要意义。

如前所述，实际的扬声器阵列更多是分层的空间结构形式，本文建立的高阶 Ambisonics 重放系统的扬声器阵列也是空间分层结构，目前的算法均没有针对这种阵

列进行设计。本章提出一种面向分层扬声器阵列的三维空间声重放方法，在保证高仰角声源定位效果的前提下，通过使用二维重放替代三维重放的方式，尽可能提高实际扬声器区域（包括但不限于水平面）的重放精度。

本章的内容安排如下：4.2 节介绍了算法设计的理论基础，包括高仰角虚拟声源的简化重放方法与垂直方向的虚拟声源合成基础；4.3 节基于上述理论基础，提出了一种新的分层 Ambisonics 空间声重放算法，详细介绍了算法的设计；4.4 节对算法的重放进行了客观模拟验证；最后，4.5 节对本章进行了总结。

4.2 理论基础

4.2.1 算法目标

实际重放系统中，扬声器阵列在空间上通常是分层布置的。分层扬声器阵列是指，扬声器阵列中各扬声器坐标 Ω_l 中的仰角 ϕ_l 分布在较少数量的离散位置上，即 $\phi_l \in \{\phi_1, \phi_2, \dots, \phi_n\}$ ，在每个仰角 ϕ_n 上，布置有数量相对较多且（近似）均匀分布的扬声器。当扬声器仰角仅分布在 $\phi_l = 0^\circ$ 时，系统为水平面重放系统。这种空间分布方式的实际扬声器安装与位置校准中相对容易，目前世界各地建成的大型扬声器阵列大多采用此类结构。例如，奥地利的 IEM CUBE 系统^[56]、丹麦技术大学的 30 通道扬声器阵列^[122]、德国慕尼黑工业大学的 SOFE 系统^[53]、德国亚琛工业大学的 SCaLAr 系统^[123]、国内华南理工大学声学研究所的 29 通道扬声器阵列系统^[68]等等。

使用分层扬声器阵列进行三维 Ambisonics 重放时，即使扬声器阵列通道数量较高，根据实际扬声器阵列仰角分布的稀疏程度，能够重放的三维 Ambisonics 阶数可能相对较低。即使在特定平面（包括但不限于水平面）上布置较多的扬声器，低阶 Ambisonics 重放算法仍具有较大的能量扩散，不能充分利用这些扬声器。

基于上述问题，本章提出了一种针对分层扬声器阵列的三维空间声重放方法，该算法设计旨在实现下述三个主要目标：

- (1) 降低高阶 Ambisonics 对扬声器数量的要求，降低对空间均匀性的要求；
- (2) 在具有实际扬声器的区域，尽量提高声场重放精度；
- (3) 保留一定的垂直虚拟声合成能力。

针对目标（1）与目标（2），算法需要具有可变分辨率的特性。在不同的区域，根据实际扬声器的密度，实现不同的空间分辨率。对于目标（3），算法需要通过主观听

音实验验证其定位效果。下节首先介绍两个基础理论，然后基于这两个基础理论，提出一种分层 Ambisonics 算法。

4.2.2 理论基础

假设欲重放的声源 S 位于方向 $\Omega_s(\theta_s, \phi_s)$ 上，重放扬声器阵列包含 L 个扬声器，其中每个扬声器 l 位于方向 $\Omega_l(\theta_l, \phi_l)$ ，若目标声源与所有扬声器均位于同一平面（包括但不限于水平面），即 $\phi_s = \phi_l = \phi'$ ，根据式(2-6)与式(2-8)，目标声源的声场 P_s 与扬声器阵列的合成声场 P_v 分别为（均假设为平面波，并将球谐函数按照式(2-2)写成完整的展开形式）：

$$P_s(\Omega, \Omega_s, k) = \sum_{n=0}^N \sum_{m=-n}^n 4\pi j^n j_n(kr) (N_n^m)^2 P_n^m(\sin \phi') P_n^m(\sin \phi) \Phi_m(\theta_s) \Phi_m(\theta) \quad (4-1)$$

$$P_v(\Omega, \Omega_1, \dots, \Omega_L, k) = \sum_{l=0}^L A_l \sum_{n=0}^N \sum_{m=-n}^n 4\pi j^n j_n(kr) (N_n^m)^2 P_n^m(\sin \phi') P_n^m(\sin \phi) \Phi_m(\theta_l) \Phi_m(\theta) \quad (4-2)$$

若扬声器阵列的重放声场与目标声场一致，满足 $P_s = P_v$ ，化简得到：

$$\Phi_n^m(\theta_s) = \sum_{l=1}^L A_l \Phi_n^m(\theta_l), m \in \{-1, 1\}, n \in [0, N] \quad (4-3)$$

对比式(2-17)可以发现，此时 A_l 的计算与 2-D Ambisonics 的解码过程一致。上述推导过程说明，对于位于任意仰角 ϕ' 的扬声器阵列，可以使用 2-D Ambisonics 编解码重放位于该仰角 ϕ' 的声源。

使用声场合成工具箱（SFS toolbox）对目标声场和多通路扬声器阵列的重放声场进行模拟^[105]，验证上述结论。模拟两种情况：（1）声源位于 $(15^\circ, 0^\circ)$ ，使用 36 个扬声器对虚拟声源进行重放，扬声器仰角 $\phi_l = 0^\circ$ ，方位角等间隔排列；（2）声源位于 $(15^\circ, 60^\circ)$ ，使用 24 个扬声器对虚拟声源进行重放，扬声器仰角 $\phi_l = 60^\circ$ ，方位角等间隔排列。两个模拟中声源与所有扬声器到原点距离一致，模拟频率均为 1000 Hz。使用最高阶（分别为 17 阶和 11 阶）的 2-D Ambisonics 进行重放，水平面的模拟声场如图 4-1 所示，

黑点表示目标声源和扬声器。对比目标声场和扬声器阵列的重构声场可以发现，2-D Ambisonics 在不同仰角下均能重放该仰角目标声源产生的声场。

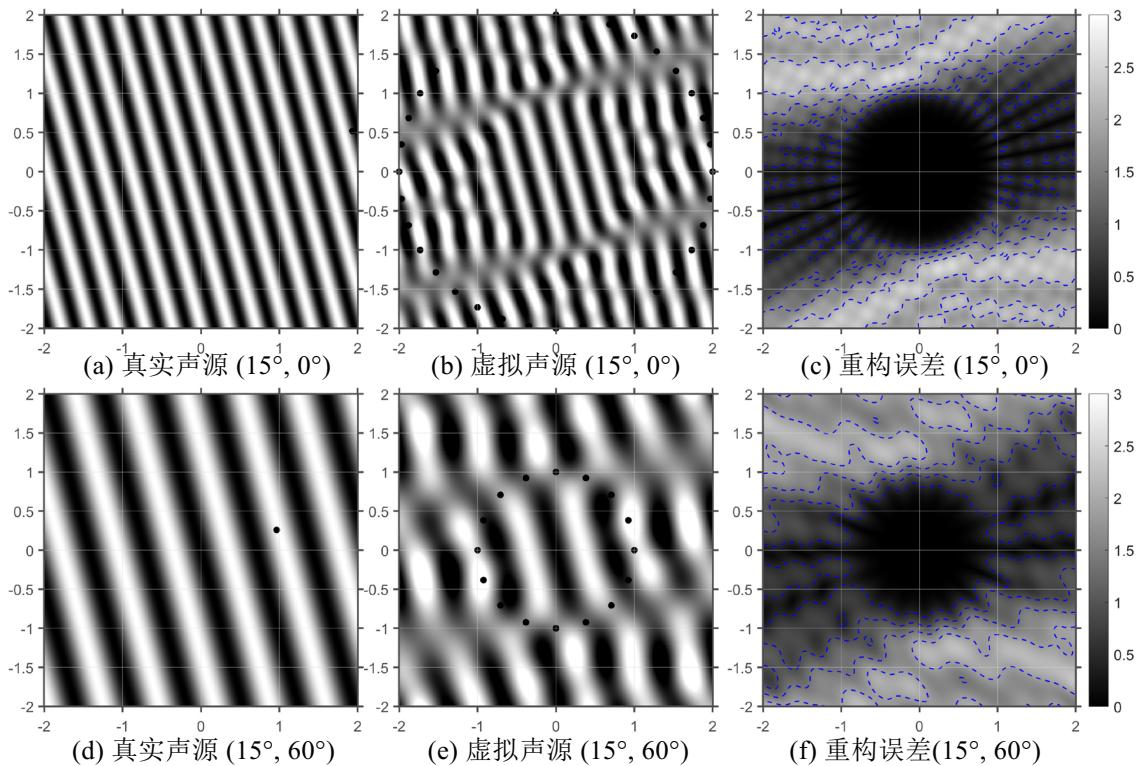


图 4-1 不同方向上虚拟声源和目标声源在水平面（1000 Hz）的合成声场与误差

Fig. 4-1 Synthesized sound fields and errors at the horizontal plane (1000Hz) of virtual sources and target sources in different directions

另一方面，在 Pulkki 等人^[124]和 Wendt 等人^[125]的垂直方向虚拟声源定位研究中，使用位于水平面上下两侧的扬声器进行矢量调幅合成虚拟声源，该虚拟声源可以产生垂直方向的定位效果，但受试者在定位实验中的定位准确度方差较大。Xie 等人进一步探究了使用位于不同仰角的扬声器进行垂直方向的矢量调幅，以合成垂直方向的虚拟声源^[126]。实验发现，位于同一半球空间（同在水平面上方或下方）的扬声器对合成的虚拟声源可产生稳定的垂直定位效果，但使用水平面上下两侧对称分布的扬声器对，无法产生稳定的虚拟声源。在后续研究中，Yi 等人利用中垂面内仰角为-45°、0°和 45°的三个扬声器合成垂直虚声源，扬声器的调幅增益通过一阶 2-D Ambisonics 编解码方法计算，该方法被称为“局部 Ambisonics”^[127]。主观实验表明，该系统不仅能在-45°至 45°仰角范围内生成垂直虚声源，还可以在扬声器边界外产生可以定位的虚拟声源。这一结论在中垂面扬声器对（方位角为 0°）和侧垂面（方位角为 30°）扬声器对的重放

条件下均得到了验证，但实验发现界外虚声源的稳定性相对较差，定位结果的离散度也较大。综上所述，目前的研究认为，使用一对垂直分布的扬声器进行矢量幅度调整产生的虚拟声源，具有一定的垂直定位性能。

4.3 分层 Ambisonics 空间声重放算法

基于上节所述的理论基础，可以得到两个基础性结论：

- A. 二维 Ambisonics 重放不仅可以在水平面合成水平面上的目标声源，还可以在非水平面合成位于该平面上的目标声源；
- B. 使用矢量调幅技术可以在垂直方向进行虚拟声源的合成，且听众能够对这种方式合成的虚拟声源进行垂直方向的定位。

如 2.2.3 节所述，对于三维 Ambisonics 重放， N 阶重放需要 $(N+1)^2$ 个扬声器，对于二维重放， N 阶重放只需要 $2N+1$ 个扬声器。对于包含 M 层的分层扬声器阵列，若在每层上分别使用二维 Ambisonics 重放合成虚拟声源，然后使用这些二维虚拟声源合成三维虚拟声源，则这种重放方式所需扬声器数量不超过 $M(2N+1)$ ，当扬声器在仰角分布上较为稀疏时，所需扬声器数量远低于三维重放。

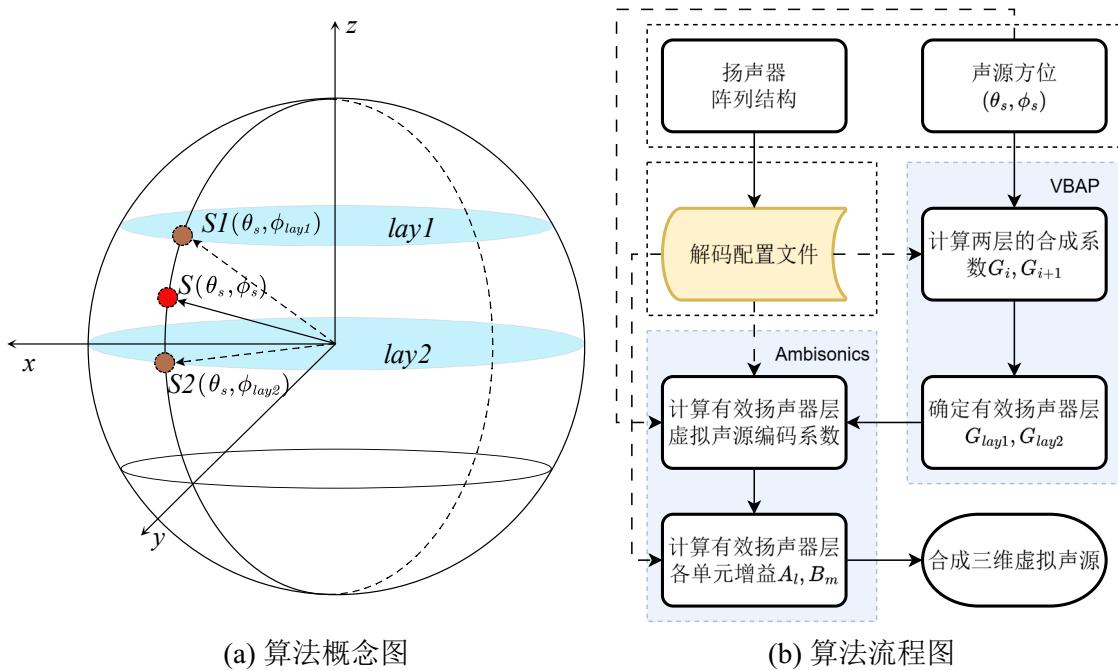


图 4-2 分层 Ambisonics 重放算法概念图与流程图

Fig. 4-2 The concept diagram and flow chart of multiple-layer Ambisonics reproduction algorithm

此外，在每层阵列上，可以自由增加或减少扬声器数量，使用更低阶的二维重放。听觉在高仰角下的定位精度较低，在高仰角下可以使用更少数量的扬声与更低阶的重放。在这种情况下，扬声器数量需求可以进一步降低。这一思路满足了 4.2.1 节算法设计目标的（1）与（2）两点。同时，根据本节开头的结论 B，这一设计思路也满足设计目的（3）。

结合上述思路，本文提出一种结合 Ambisonics 重放与 VBAP 重放的三维空间声重放算法，其概念以及算法流程如图 4-2 所示。算法使用两层物理扬声器阵列 $lay1$ 与 $lay2$ ，通过二维 Ambisonics 分别合成两个位于同一方位角，不同仰角的虚拟源 $S1$ 与 $S2$ ；然后，通过二维 VBAP 的方式调整两个虚拟源的相对强度，合成为目标虚拟源 S 。

对于位于方向 $\Omega_s(\theta_s, \phi_s)$ 的目标声源 S ，使用最接近该声源的上下两层扬声器阵列对目标声源进行重放，其中，上层 $lay1$ 包含 L 个扬声器，每个扬声器 l 的方向为 $\Omega_l(\theta_l, \phi_{lay1})$ ，下层 $lay2$ 包含 M 个扬声器，每个扬声器 m 的方向为 $\Omega_m(\theta_m, \phi_{lay2})$ ，其中， $\phi_{lay2} \leq \phi_s \leq \phi_{lay1}$ ，则目标声源 S 的声场合成公式为：

$$P_v = G_{lay1} \sum_{l=1}^L A_l e^{j k \vec{r}_l \cdot \vec{r}} + G_{lay2} \sum_{m=1}^M B_m e^{j k \vec{r}_m \cdot \vec{r}} \quad (4-4)$$

A_l, B_m 分别为 $lay1$ 和 $lay2$ 上每个扬声器的增益，它们是使用 2-D Ambisonics 合成位于方位角 θ_s 的虚拟声源 $S1$ 与 $S2$ 所对应的扬声器增益，计算方法见式(2-17)，编解码阶数为当前扬声器层上能够进行稳定重放的最高阶，与扬声器数量有关。若扬声器方位角均匀分布，阶数分别为 $\lfloor (L-1)/2 \rfloor$ 与 $\lfloor (M-1)/2 \rfloor$ ，其中， $\lfloor \cdot \rfloor$ 为向下取整运算。

G_{lay1} 和 G_{lay2} 为虚拟声源 $S1$ 和 $S2$ 的增益，通过 2-D VBAP 进行计算^[39,124]：

$$\begin{cases} \begin{bmatrix} G'_{1ay1} \\ G'_{1ay2} \end{bmatrix} = \begin{bmatrix} \cos(\phi_{lay1}), \cos(\phi_{lay2}) \\ \sin(\phi_{lay1}), \sin(\phi_{lay2}) \end{bmatrix}^{-1} \begin{bmatrix} \cos(\phi_s) \\ \sin(\phi_s) \end{bmatrix} \\ \begin{bmatrix} G_{1ay1} \\ G_{1ay2} \end{bmatrix} = \begin{bmatrix} G'_{1ay1} \\ G'_{1ay2} \end{bmatrix} / \| \begin{bmatrix} G'_{1ay1} \\ G'_{1ay2} \end{bmatrix} \| \end{cases} \quad (4-5)$$

可以按照如下方法确定最接近目标声源 S 的扬声器层 $lay1$ 与 $lay2$ 。首先将分层阵列的所有扬声器层的仰角 ϕ_n 按照数值大小从小到大排列，即 $\phi_{n+1} > \phi_n$ 。然后，按照仰

角相邻的原则两两组合，形成 $N-1$ 个扬声器层组合 $\{(\phi_n, \phi_{n+1}) | 1 \leq n \leq N-1\}$ ，对所有 $N-1$ 个扬声器层组合计算调幅系数， n 与 $n+1$ 层扬声器调幅系数为：

$$\begin{bmatrix} G'_n \\ G'_{n+1} \end{bmatrix} = \begin{bmatrix} \cos(\phi_n) & \cos(\phi_{n+1}) \\ \sin(\phi_n) & \sin(\phi_{n+1}) \end{bmatrix}^{-1} \begin{bmatrix} \cos(\phi_s) \\ \sin(\phi_s) \end{bmatrix} \quad (4-6)$$

若 (ϕ_n, ϕ_{n+1}) 仰角层组合的增益满足 $G'_n \geq 0, G'_{n+1} \geq 0$ ，则 n 与 $n+1$ 层为有效扬声器层，分别记为 $lay2$ 与 $lay1$ 。特殊地，若 $lay1$ 或 $lay2$ 只包含一个扬声器（如天花板扬声器或地板扬声器），则该扬声器直接视为 S_1 或 S_2 本身。此外，若扬声器层中各单元不是环状分布，即扬声器到阵列中心的距离不同，应在重放前进行幅度与延时均衡。

该算法的优势在于：当虚拟声源靠近或位于扬声器所在的仰角平面时，仅使用 2-D Ambisonics 进行重放，最大限度地提高了空间重放精度；当声源位于两层扬声器之间时，重放能量扩散仅取决于两层扬声器之间的仰角差值大小，在分层扬声器阵列中，该扩散程度通常小于常规 3-D Ambisonics 算法的能量扩散，且仍具有一定的垂直方向定位性能。可以看到，该重放算法大大降低了扬声器的数量要求，对扬声器空间均匀分布的要求较低，实现了 4.2.1 节所述的设计目标。

该算法的局限性包括：（1）需在已知虚拟声源的实际位置与扬声器阵列的全部信息之后才能进行重放，这削弱了 Ambisonics 信号编解码完全独立的优势，且不能通过传声器阵列对声场进行拾拾；（2）分层 Ambisonics 重放算法无法合成位于最大或最小仰角平面范围以外的虚拟声源，在本研究中，扬声器阵列包含 90° 仰角的天花板扬声器，因此可以重放整个上半空间的虚拟声源，在实际系统建设与信号制作时，需要充分考虑这一限制，将重放仰角限制在扬声器阵列有效仰角范围内，或增设天花板扬声器；（3）算法在垂直方向上采用了 VBAP 进行合成，VBAP 算法不能保证不同方向虚拟声源的能量在较大的区域内保持恒定，特别是，当声源仰角动态变化时，可能会产生可感知的声源强度变化，因此，使用该算法进行合成动态声源时，需要考虑对虚拟声源进行平滑处理，如采用 MDVP（Multiple-direction amplitude panning）等方法^[128]。

4.4 算法客观验证

4.4.1 HRTF 模拟

针对本文所建立的多通道扬声器阵列进行重放模拟验证，分别使用上述分层 Ambisonics 算法与 8 阶 3-D Ambisonics 重放（MMD 解码），对目标方向虚拟声源进行合成，计算各扬声器增益 A_l 。然后，分别计算目标声源与虚拟声源在双耳的 HRTF，

计算方法如3.2节所述。一共对9个方向进行了模拟，模拟的左耳HRTF如图4-3所示。分别将分层Ambisonics、3-D Ambisonics与目标声源的HRTF进行对比，从总体结果来看，两种方法在中频（低于5kHz）都与目标声源的HRTF接近，在高频范围，3-D Ambisonics具有较大的误差。

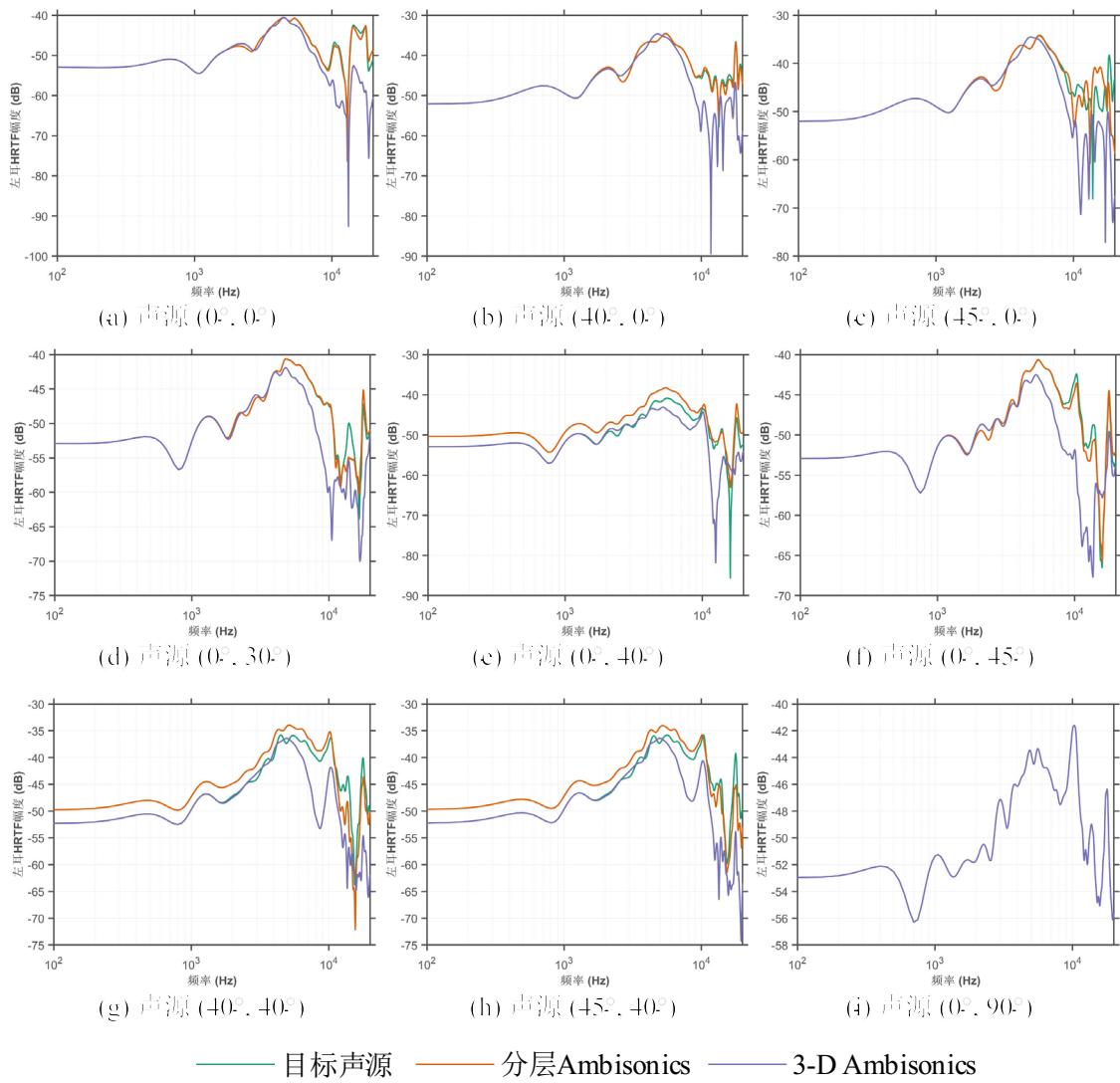


图4-3 不同声源方向下，目标声源、分层Ambisonics与3-D Ambisonics在左耳处的HRTF

Fig. 4-3 The HRTFs of left ear of sources of target, multiple-layer Ambisonics and 3-D Ambisonics, located at different directions

具体分析每个方向，首先对比水平面重放 ($\phi_s = 0^\circ$)，该阵列在水平面具有较高密度的扬声器，对比图4-3(a)、(b)与(c)发现：当重放位置具有实际的扬声器

时，分层 Ambisonics 的频谱几乎与目标声源一致；在方位角为 45° 的情况下，分层 Ambisonics 误差增大，但都比 3-D Ambisonics 更接近目标声源的频谱。

进一步分析更高的重放仰角，对比图 4-3 (d)、(e) 与 (f) 可以发现，当声源位于实际具有扬声器的仰角 (30° 与 45°)，分层 Ambisonics 重构的频谱在全频段都与目标声源频谱较为接近，但当合成声源仰角 (40°) 位置没有实际扬声器阵列时，分层 Ambisonics 在中低频的频谱与目标频谱误差较大，但高频仍然很接近。3-D Ambisonics 的频谱尽管在 40° 时误差也变大，但其中低频与目标频谱重合的较好，主要误差出现在高频。这一区别可能与分层 Ambisonics 在垂直方向上通过 VBAP 合成虚拟源有关，当声源不在实际物理层时，重放误差相比 3-D Ambisonics 特性不同。如 1.2.2 节所述，高频峰谷是垂直方向定位的重要线索，因此，即使对于此方向上的声源，分层 Ambisonics 仍有可能具有较好的垂直定位性能。对比 30° 与 0° 仰角，尽管在两个条件下，分层 Ambisonics 频谱都与目标声源频谱接近，但 30° 仰角的情况误差相对更大，这可能是由于 30° 仰角平面上扬声器密度没有水平面 (0°) 上的扬声器密度高。

最后观察图 4-3 (i) 中重放正上方虚拟声源的结果，可以看到两种重放方式都完美重放了正上方的虚拟声源。

4.4.2 声场模拟

进一步地，使用式(1-6)的平面波解析解，对扬声器阵列在水平面区域合成的声场进行模拟，模拟由 SFS Toolbox^[105]完成。与上一节类似，同样对比分层 Ambisonics 与 8 阶 3-D Ambisonics 的合成结果。选取 1 kHz 作为测试频率，人耳对该频率的声音最敏感，在该频率下，按照式(2-18)计算，8 阶 3-D Ambisonics 的理论有效重放范围约为 0.43 m。模拟了声源位于不同仰角($15^\circ, 0^\circ$)、($15^\circ, 40^\circ$)与($15^\circ, 45^\circ$)的三种情况，不同声源位置、不同重放方法在水平面 $2 \text{ m} \times 2 \text{ m}$ 区域内合成声场的情况。结果如图 4-4 所示，黑色圆点为扬声器，每一组图都展示了合成的声场与相比真实声源的合成误差。可以看到，对于 3-D Ambisonics，其声场合成误差对于不同仰角的虚拟声源几乎一致，有效合成半径约为 0.5 m（图 4-4 (a)、(c) 与 (e)）。对于分层 Ambisonics，其在水平面的有效合成区域最大，超过 1 m。分层 Ambisonics 的声场重构误差与目标虚拟声源的仰角有关，对比图 4-4 (d) 与 (f) 可以发现，该算法在具有实际物理层的仰角重放误差较小。值得注意的是，即使在扬声器物理层以外 (40°)，分层 Ambisonics 重放仍然比 3-D Ambisonics 的重构精度更高。

4.4.3 结论

本节对比了 8 阶 3-D Ambisonics 与分层 Ambisonics 合成虚拟声源的 HRTF 与虚拟声场。主要结论为：

客观模拟的结果符合理论基础与算法设计预期，在具有实际物理扬声器的仰角区域，特别是水平面上，具有较高的声场重放精度。

当合成目标位于实际扬声器物理层外时，分层 Ambisonics 算法误差增大，特别是低频 HRTF 与实际情况差距较大。

相比传统 3-D Ambisonics，分层 Ambisonics 的重构误差随仰角变化，而 3-D Ambisonics 的重放误差在全空间基本类似。

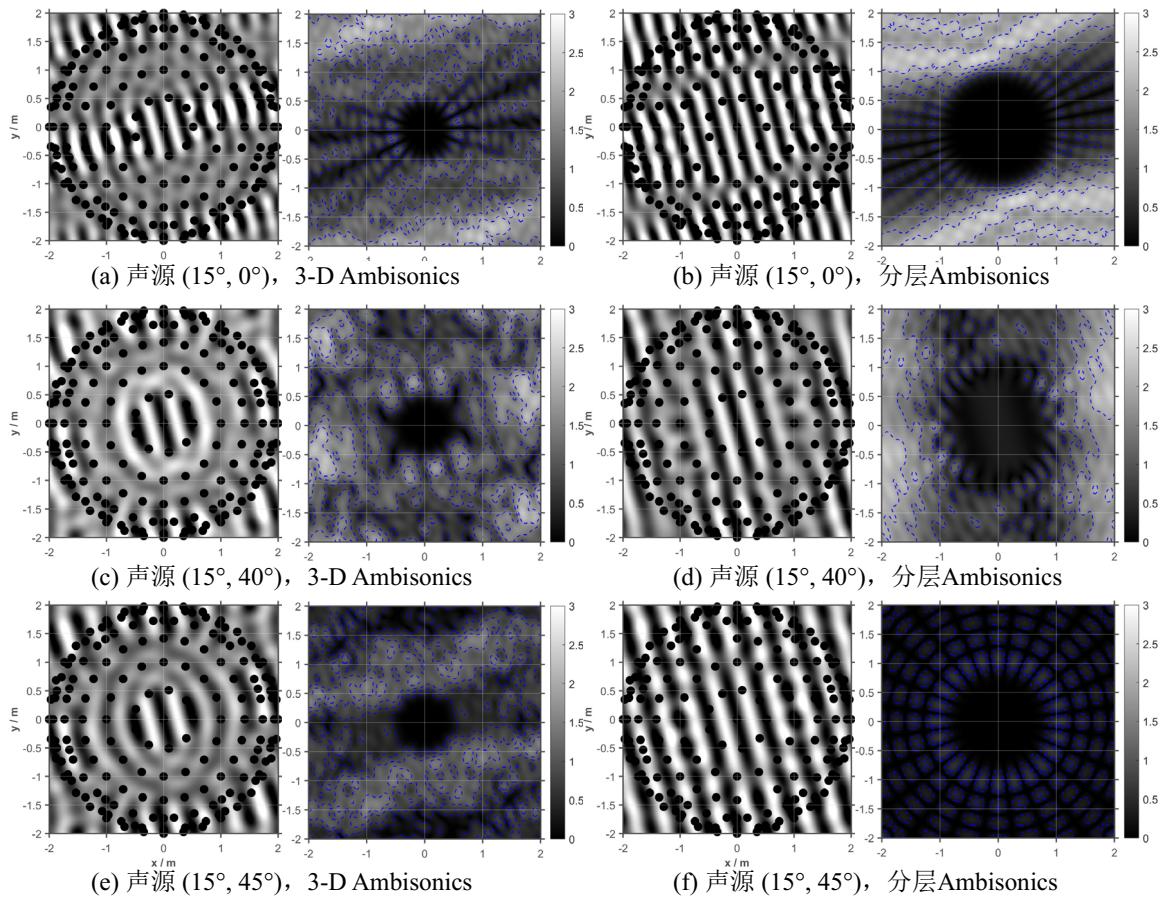


图 4-4 分层与 3-D Ambisonics 在水平面 (1000 Hz) 的合成声场与合成误差

Fig. 4-4 Synthesized sound fields and errors at the horizontal plane (1000 Hz) of multiple-layer Ambisonics and 3-D Ambisonics

4.5 本章小结

本章针对在分层扬声器阵列下进行 3-D Ambisonics 重放的实际困难，包括有效重放阶数低，对扬声器使用不充分等问题，提出了一种结合 2-D Ambisonics 与 VBAP 技术的分层 Ambisonics 重放方法，该重放方法使用两层扬声器阵列分别合成两个位于不同仰角的高精度虚拟声源，再经矢量调幅实现虚拟声源的仰角调节，从而降低了扬声器数量及对空间分布均匀性的需求，提高了具有实际物理扬声器区域的重放精度。

本章首先对分层 Ambisonics 重放的理论基础进行了分析、仿真模拟以及文献调研，然后详细介绍了算法的设计思路与实施流程，并对该算法的优势与缺陷进行了分析。最后，对算法进行了重放模拟，使用 HRTF 模拟了双耳接收的信号，还模拟了水平的声场重构误差，客观分析证实了该算法在目的使用场景下的优势。本章的客观分析仅是初步验证了算法符合设计预期，相比传统 3-D Ambisonics，分层 Ambisonics 在实际物理层有一定的重放精度优势，但其实际的定位性能、声音保真度等方面需要进行进一步的主观实验验证。下一章将开展主观实验，对该算法与相关的 Ambisonics 优化算法进行对比。

第五章分层 Ambisonics 与混合阶 Ambisonics 重放的主观感知对比实验

第四章提出了一种针对分层扬声器阵列的三维空间声重放算法，该算法结合二维 Ambisonics 重放与 VBAP 调幅算法，使用双耳接收信号模拟与声场模拟的方法验证了该算法相比 3-D Ambisonics，在重放准确性上有一定的性能提升，且算法降低了对扬声器的数量与空间均匀性的要求，符合设计预期。但上述分析仅是对合成重放过程的简单模拟，不能直接反应算法的实际听觉感知性能，此外，除了本文所提出的分层 Ambisonics 外，还有一些算法也对非均匀扬声器的重放问题进行算法改进。本章开展主观实验，对比和验证分层 Ambisonics 重放算法与其他相关算法的效果。

5.1 引言

在空间声重放算法的主观感知评估中，定位性能与声音保真度是两个最基础的评估维度：重放算法需要保证听众在听音时能够正确地感知虚拟声源的方向；并且，虚拟声源与真实声源相比，在听感上应具有尽量小的失真。本章针对这两个感知维度，对分层 Ambisonics 进行主观实验效果评估，并将其与相关重放算法进行对比。

除了第四章提出的分层 Ambisonics 算法，已有学者提出多种方法尝试降低 Ambisonics 对扬声器阵列空间均匀分布的要求。Zotter 等人结合 VBAP 以及 Ambisonics 算法，提出 All-Round Ambisonics Decoding (AllRAD) 解码算法降低 Ambisonics 重放对扬声器均匀分布的要求^[54]，该算法在理论上对扬声器空间布局没有严格要求，但其没有提高特定区域，如水平面上的重放精度，在特定区域，该算法可能反而会增大能量扩散程度，降低重放精度。另一种思路是结合二维 Ambisonics 与三维 Ambisonics 重放，在三维球谐函数的基础上扩展主要包含水平面能量的高阶谐波，提高水平面上的重放阶数，这种方法被称为混合阶 Ambisonics^[129]。

尽管不同方法均有其设计目标和对应的优势，但目前尚未有研究开展主观实验对这些算法进行主观听音对比，在设计重放系统时，对重放算法的选择较为困难。因此，本章开展主观定位与保真度评价实验，将分层 Ambisonics 重放算法与已有的混合阶 Ambisonics 算法以及 AllRAD 算法进行对比。

本章的内容安排如下：5.2 节介绍了混合阶 Ambisonics 的基本原理及其研究现状，并对空间声重放算法的主观评估方法进行介绍；5.3 节开展主观定位实验与保真度评估

实验，对分层 Ambisonics 算法、几种混合阶 Ambisonics 算法以及 AllRAD 算法进行主观比较；5.4 节报告了实验结果并进行了讨论；最后，5.5 节对本章进行了总结。

5.2 研究背景

5.2.1 混合阶 Ambisonics 研究现状

除了 2.2.5 节所描述的混合阶 Ambisonics 策略外，近年来一些研究提出了其它的混合阶 Ambisonics 策略。Travis 等人建议保留 $n \in [0, N], m \in [-M, M], M < N$ 的球谐函数谐波^[94]（见图 5-1（b）），它保留了较少的 Sectoral 谐波，更多地保留了在高仰角具有更多能量的谐波，使得水平面和高纬度平面的能量过渡更为平滑。Chang 等人的研究提出一种新的混合阶策略（见图 5-1（c）），保留所有在水平面上能量不为 0 的谐波，以尽可能地提高水平面重构精度^[130]。

与图 5-1（a）中的基础混合阶 Ambisonics 策略（式(2-17)）稍有不同，Grandjean 提出将扬声器阵列分为两部分，空间均匀的部分用于重放 M_{3D} 阶的 3-D Ambisonics，水平面的额外扬声器部分用于直接扩展重放 ($M_{3D} + 1$) 至 M_{2D} 阶的 2-D Ambisonics^[131,132]。Heller 等人提出在基础的混合阶 Ambisonics 上，使用非线性优化算法最大化空间全局的速度矢量和能量矢量^[133]。对于更复杂的扬声器阵列，Trevino 提出将扬声器阵列分为三部分，第一部分为空间均匀的阵列，用于 3-D Ambisonics 重放，第二部分为在水平面具有更多扬声器的三维阵列，采用混合阶 Ambisonics 重放，最后留下一部分扬声器用于补偿声场误差^[134]。此外，混合阶 Ambisonics 理论还能用于非均匀的球形传声器阵列，提高阵列在水平面拾拾的空间分辨率^[135,136]。

上述研究均是针对混合阶 Ambisonics 算法的改进或重新组合，对算法的评估主要依靠声场仿真或是计算特定的客观指标，目前对于混合阶 Ambisonics 系统重放实际效果的主观感知评估较少。Käsbach 等人使用混合阶 Ambisonics 来重现虚拟音乐厅场景，并将空间感知与 3-D Ambisonics 进行了比较，实验证实了相比 3-D Ambisonics，混合阶 Ambisonics 重放的虚拟厅堂空间感更接近参考环境^[122]。Oreinos 和 Buchholz 等人建立了一个混合阶球形传声器阵列，进行声场拾拾并通过混合阶 Ambisonics 系统进行重放，开展实验并比较了虚拟房间和实际声环境下助听器用户的语音可懂度，二者仍存在一定差距^[110]。麦海明对双耳虚拟化的混合阶 Ambisonics 重放的定位性能和音色保真度感知进行了全面的评估，实验证实了，和 3-D Ambisonics 相比，混合阶 Ambisonics 重放

的水平面上的虚拟声源有更大的听音区域与更高的定位精度，重放声源的音色失真也随扬声器数量上升而下降^[68]。

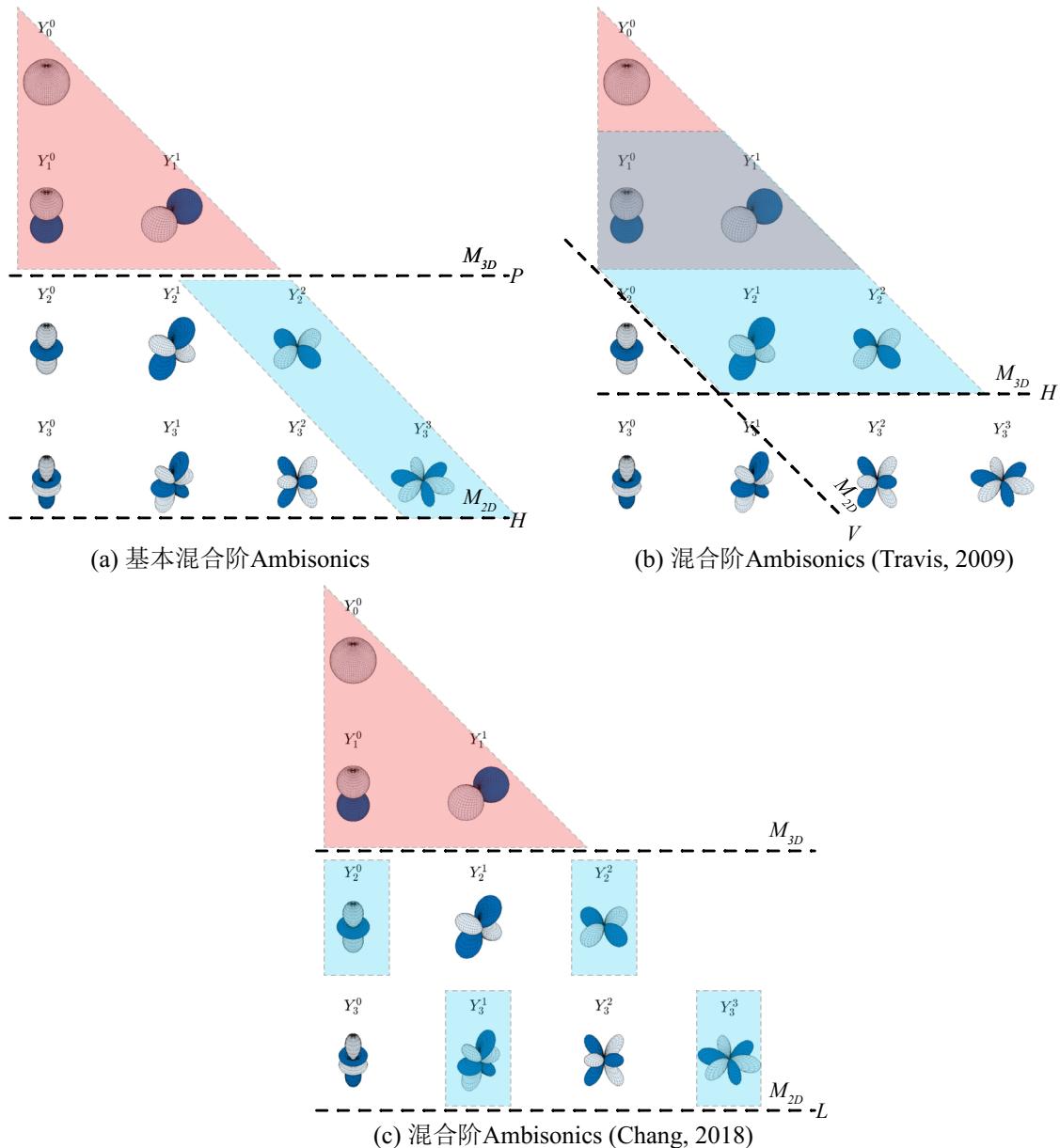


图 5-1 不同混合阶 Ambisonics 的球谐函数选择策略

Fig. 5-1 Spherical harmonic function selection schemes of different mixed-order Ambisonics

5.2.2 空间声重放系统主观评价方法

目前，衡量空间声重放系统性能的金标准是主观听音实验，其中，主要的评价维度包括定位性能和声音失真。

定位实验中常用的两种范式为绝对定位与相对定位。其中，绝对定位是最广泛接受的方法，通过在不同位置播放虚拟声源，要求受试者在没有参考的条件下反馈他们听到的声源方向^[137]；而相对定位范式通常要求受试者将虚拟声源调整至真实声源的参考位置^[124]，此时受试者能够对两个信号的方向进行比较。此外，还有通过判断声源相对位置关系，进而确定虚拟声源的方向分辨角度阈值的测量方法，此类方法通常用于测量最小可辨角度^[121,138]。对于绝对定位实验范式，在收集受试者反馈的方式上，主要包含以自身为中心的指向（Egocentric pointing）和外部中心的指向（Exocentric pointing）^[139]两种类型的方法。前者要求受试者通过手持棍棒（或类似的工具）直接指向感知到的声源位置，这一过程受试者以自身为参考坐标系，对声源相对自身的方向进行定位；后者则要求受试者借助用户图形操作界面（Graphical User Interface, GUI）或三维追踪球等外部接口输入他们感受到的位置^[140]，此时受试者拥有一个外部的方向参考系统。在反馈感知角度的过程中，进一步增加视觉辅助可提升反馈精度^[141]。有研究指出棍棒指向法的准确性普遍优于其他方式，Brungart 等人比较了四种不同的反馈感知角度的方法，发现直接使用棍棒指向的角度误差最低，约为 10°^[142]。

与定位感知不同，主观声音失真评价多通过相对比较实现，常用方法包括成对比较、ABX 比较、多重比较及其变体，如包含隐藏参考与锚点的多刺激对比法（Multiple stimuli with hidden reference and anchor, MUSHRA）。这些方法不仅适用于失真评估，也被广泛用于空间感、语音质量等感知任务的评估^[143]。在数据收集层面，可通过连续量表或离散等级量表让受试者对刺激间的差异排序，或使用 N 项强迫选择测试（N-Alternative forced choice, N-AFC）、ABX 或是成对比较等方法获取二项反馈（正确/错误，高/低）。对于包含多个比较项目的成对比较实验，一般需通过 Thurstone Case V 或 Bradley-Terry 模型将两两比较的分数差异转化为多项条件之间的相对差异大小；二项反馈通常适用于比较差异细微且难以分辨的刺激，否则容易出现天花板效应（所有刺激的回答正确率都接近 100%）。在相关研究案例中，Yang 和 Xie 等人采用 2AFC 法探讨 Ambisonics 阶数、听音位置等对虚拟声源的音色感知的影响^[65]；Nachtigäller 等人^[144]与 Bates 等人^[145]的研究则使用 MUSHRA 评估不同重放条件下的声音差异；Ahrens 和 Andersson 等人使用 AB 比较法对比了不同 Ambisonics 阶数对音色的影响^[146]。此外，问卷调研^[51]和自由访谈等方式^[147]也被用于声音感知评价。

5.3 混合阶及分层 Ambisonics 的定位与声音保真度主观评估实验

5.3.1 实验目的

本章在 4.2.2 节初步论证了分层 Ambisonics 算法的可行性，但算法在实际重放中的效果仍存在诸多未知，尤其是其在主观听觉感知层面的表现，亟需深入探究。另一方面，对于现有的几种混合阶 Ambisonics 重放方法以及另一种结合 VBAP 与 Ambisonics 重放的 AllRAD 算法，尚未有研究对其主观感知性能进行对比。

因此，本研究开展主观听音实验，旨在全面、系统地评估分层 Ambisonics 重放算法的实际效果，并将其与现有的混合阶 Ambisonics 重放算法进行深入对比。如 5.2.2 节所述，对重放算法的主观感知评价主要包含两个维度：定位性能与声音保真度。在定位性能方面，对比不同算法合成的虚拟声源在空间中定位的准确性与稳定性，包括水平面和垂直平面上的定位性能差异。对于保真度表现，评估各算法与真实声源（单扬声器）的声音感知差异。实验主要对比了以下几个变量对两个主观感知维度的影响：不同重放算法；不同的虚拟声源方向；中心听音与非中心听音位置。通过该实验，为本章提出的分层 Ambisonics 重放算法和已有的 Ambisonics 重放方法在非均匀扬声器阵列中的应用提供坚实的主观评价依据，同时为新型算法的设计与优化指明方向。

本实验在德国慕尼黑工业大学的 SOFE (Simulated Open-Field Environment) 扬声器阵列系统下开展^[76]。该扬声器阵列在水平面具有高密度的实际扬声器分布，在空间仰角上分布相对稀疏，该阵列在空间上仅有三个扬声器层与一个天花板扬声器，是一个典型的不均匀分层扬声器阵列，在实际应用中具有可实现性。该系统位于一处全消声室内（见图 5-2 (a)），主要包含一套扬声器阵列声重放系统、一套四台投影仪组成的 CAVE 全景视觉系统，以及一套光学动作捕捉系统（OptiTrack）。

5.3.2 实验对象与刺激信号

5.3.2.1 实验对像

本实验招募 10 名听力正常的受试者，其中包括 5 名女性，5 名男性，受试者的年龄范围为 24 至 35 岁，平均值为 27.4 岁，标准差为 3.41 岁，所有受试者均无中耳炎及相关耳科病史。实验开始前，使用经过校准的手动纯音听力计（Otometrics Astera）和高频测听耳机（HDA 200）对所有受试者进行听力测试，听力测试由一名经过培训的实验员实施，测听流程为改进的 Hughson-Westlake 测听方法^[148]。所有受试者在 500 到

4000 Hz 倍频程的平均气导听阈不超过 20 dB HL。受试者均在得知实验目的与大致实验内容的基础上，签署实验知情同意书后自愿参与实验，并获得了一定的实验酬金。

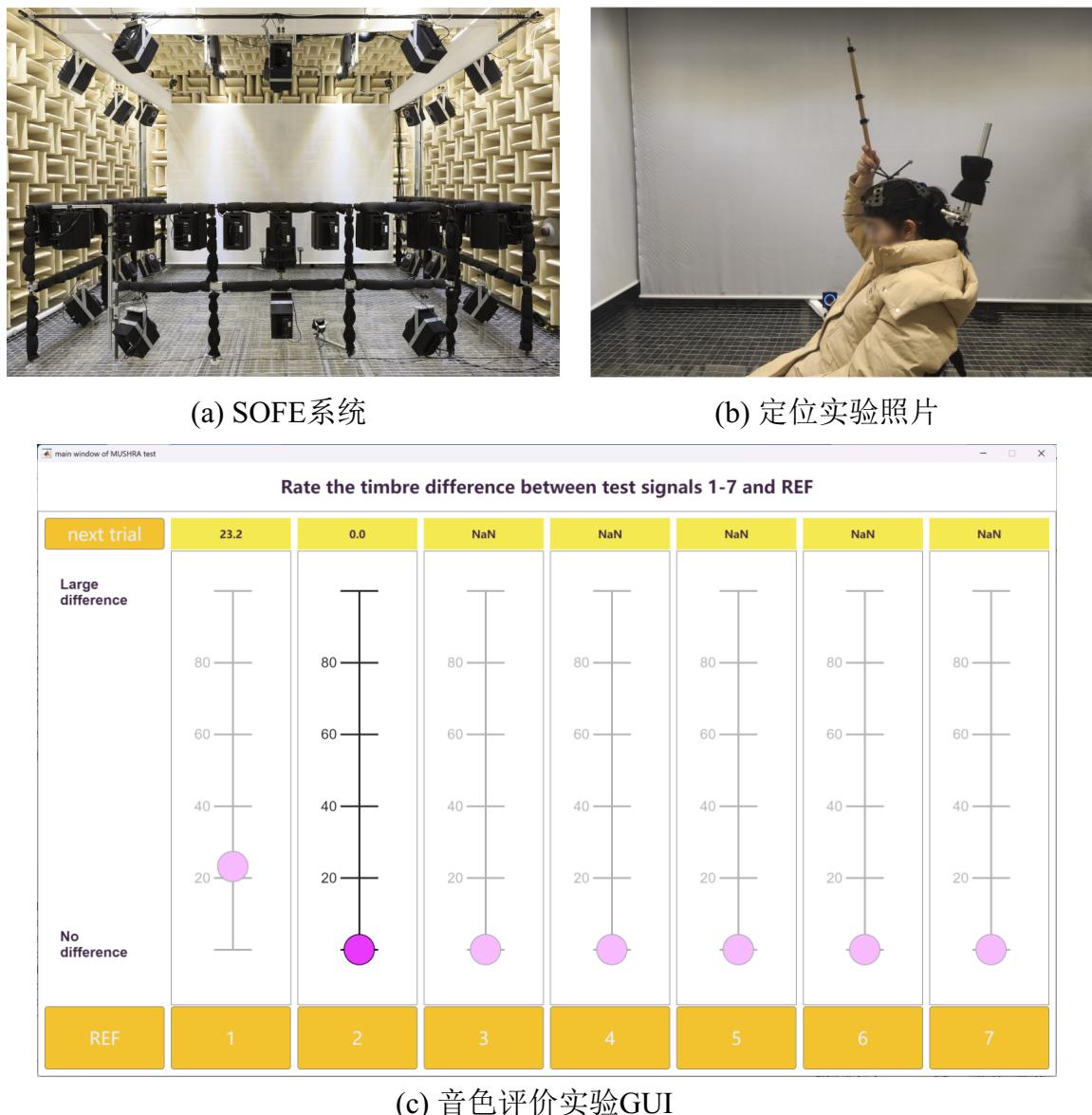


图 5-2 SOFE 系统照片、定位实验过程照片以及声音保真度评估实验 GUI 界面

Fig. 5-2 Photos of the SOFE system, photos of the location experiment, and the GUI of the sound fidelity evaluation experiment

5.3.2.2 实验刺激信号

SOFE 扬声器阵列包含 61 通路的扬声器（Dynaudio BM6A mk II），其空间分布见表 5-1。该阵列在空间上分层分布，包括水平面上均匀分布的 36 个扬声器、-20°仰角层的 12 个扬声器、30°仰角层的 12 个扬声器、以及 1 个天花板扬声器。信号通过 RME

HDSPE FX 声卡以 44.1 kHz 的采样率传输至数模转换器（RME 32DA），将数字信号转换为模拟信号，驱动 61 个扬声器发声。为确保各扬声器输出的一致性，使用 1/2 英寸传声器（MM210，Microtech Gefell）在阵列中心测量扬声器频谱，并采用 1024 点的 FIR 滤波器对扬声器进行校准，在 90 Hz~20 kHz 频率范围内进行延时、幅度和相位均衡，校准流程见 2.3.2 节。在软件层面，高阶 Ambisonics 重放算法实现及对应软件与本文第二章所述的流程一致。

在定位实验和声音保真度评估实验中，采用 500 ms 的宽带粉红噪声作为声源信号，在该信号两端各设置 50 ms 的淡入淡出时间，该信号涵盖了人耳听觉定位所需的全部频率范围。信号时长相对较短，旨在避免受试者过度依赖因头部移动产生的声音动态变化线索进行判断，确保受试者能专注于对静态声源的定位和声音保真度感知。此外，由于不同算法合成的信号声压级可能存在差异，所有信号均通过布置于阵列中心的单通道传声器进行测量与校准，使其在阵列中心处的声压级为 65 dB SPL。

在定位和声音保真度评估实验中对比了五种虚拟重放算法的重放效果，包括本章提出的分层 Ambisonics 算法、三种混合阶 Ambisonics 重放方法（见图 5-1）以及 AllRAD 重放方法^[54]，所有重放方法均在当前系统允许的最高重放阶数下进行测试。。

表 5-1 SOFE 系统扬声器阵列各扬声器空间方位

Tab. 5-1 The spatial direction of each loudspeaker in the SOFE loudspeaker array

仰角 ϕ	方位角 θ	数量
-20°	0° 至 330° 以 30° 等间隔分布	36
0°	0° 至 350° 以 10° 等间隔分布	12
30°	0° 至 330° 以 30° 等间隔分布	12
90°	0°	1

对于分层 Ambisonics 重放，其算法流程按照 4.3 节所述进行，-20°、0° 以及 30° 扬声器层分别使用 17 阶、5 阶以及 5 阶的 2-D Ambisonics 重放。

对于三种混合阶重放方法，根据解码矩阵的条件数 κ ，在保证重放稳定的前提下选择最高的阶数。第一种混合阶 Ambisonics（见图 5-1（a），后文简称混合阶 1）阶数 (M_{2D}, M_{3D}) 为 (17, 3)；第二种混合阶 Ambisonics（见图 5-1（b），后文简称混合阶 2）阶数 (M_{2D}, M_{3D}) 为 (5, 2)；第三种混合阶 Ambisonics（见图 5-1（c），后文简称混合阶

3) 阶数(M_{2D} , M_{3D})为(4, 2)。对于 AllRAD 算法, 采用 6 阶 Ambisonics 重放 (至少要求 49 个实际物理扬声器, 7 阶重放要求 64 个扬声器, 超出此系统的扬声器数量), 虚拟扬声器布置为 12 阶 T-design (对应 86 个虚拟扬声器) [54]。图 5-3 给出了三种混合阶重放的条件数随阶数(M_{2D} , M_{3D})的变化, 其中, 混合阶 2 和 3 在 $M_{3D} \geq 3$ 的情况下条件数无穷大, 因此没有显示在图中。可以看到混合阶 1 能够保留最多的谐波, 混合阶 2 与混合阶 3 由于包含了更多垂直方向具有能量的谐波, 能够稳定解码的谐波数相对较低。

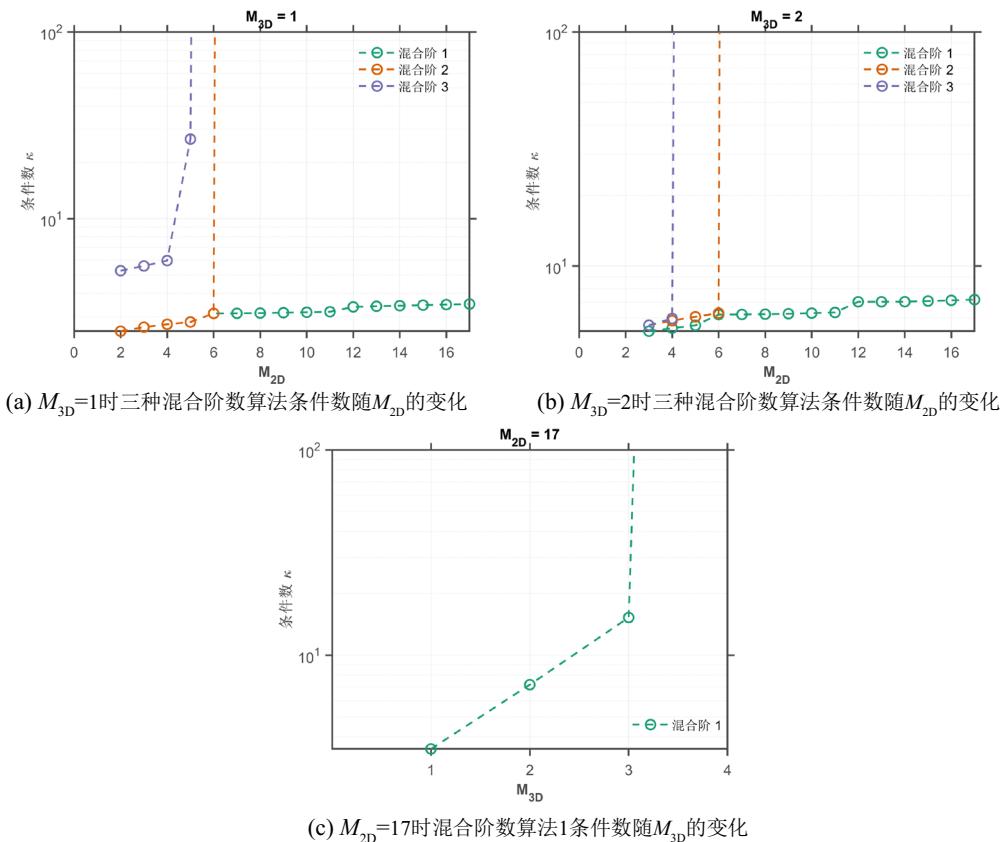


图 5-3 三种混合阶 Ambisonics 重放在 SOFE 系统下的解码矩阵的条件数

Fig. 5-3 The condition numbers of three mixed-order Ambisonics decoding matrices under the SOFE system

5.3.3 定位实验

定位实验采用绝对定位实验范式, 在每个试次中, 受试者在听到信号后使用一根木棍指向感知的声源方位, 受试者确认反馈方向后, 需要按下另一只手中的一个按钮进行确认, 然后, 木棍的顶端和受试者头中心位置被光学动作捕捉系统记录, 系统自动计算受试者反馈方向, 若受试者感觉声源位于头内部而非空间中的某个方向, 则按下另一个按键, 系统记录为头中心定位, 每个试次耗时约 2~3 秒。该实验为三维空间

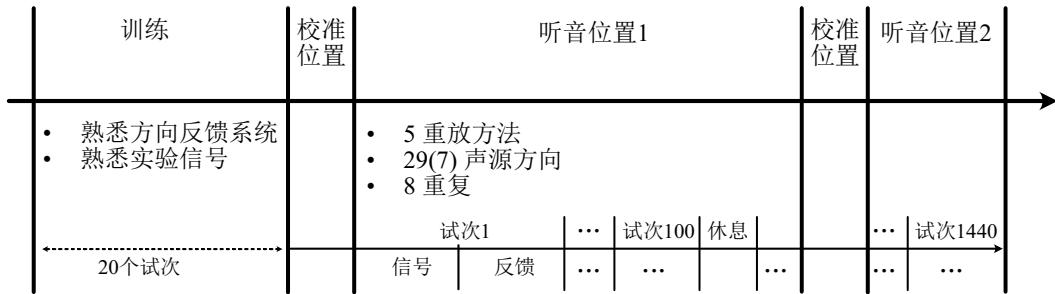
定位实验，即虚拟声源 S 的方位角 θ_s 和仰角 ϕ_s 均随设置的条件进行改变。由于受试者使用木棍指向身体后方的角度相对不便，所有虚拟声源均位于受试者的前半空间，即方位角位于 $-90^\circ \leq \theta_s \leq 90^\circ$ 范围内。

实验对中心听音位置和非中心听音位置两个条件进行了测试：中心听音时，受试者坐在扬声器阵列的中心；非中心听音时，受试者坐在阵列中心右侧 60 cm 处，见图 5-4 (c)。对于中心听音的条件，共设置 29 个虚拟声源方向，分布在 $-15^\circ, 0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ$ 和 90° 共 7 个仰角平面。对于非中心听音的条件，共设置 7 个虚拟声源方向，分布在 $0^\circ, 30^\circ$ 及 60° 共 3 个仰角平面，详细的虚拟声源坐标条件见表 5-2。每个方向重复测试 8 次，因此，每个受试者需要进行 1440 个定位试次 [5 种重放方式 \times (29 + 7) 个声源方向 \times 8 次重复]。其中，中心听音条件包含 1160 个试次，非中心听音条件包含 280 个试次，重放方式和声源方向条件出现次序随机以降低学习效应带来的固定偏差。为避免疲劳影响实验结果，受试者每完成 100 次试验后，根据疲劳情况进行适当休息，每个受试者实验耗时为 2.5 至 3 个小时。

表 5-2 定位实验虚拟声源方向

Tab. 5-2 The directions of virtual sources in the localization experiment

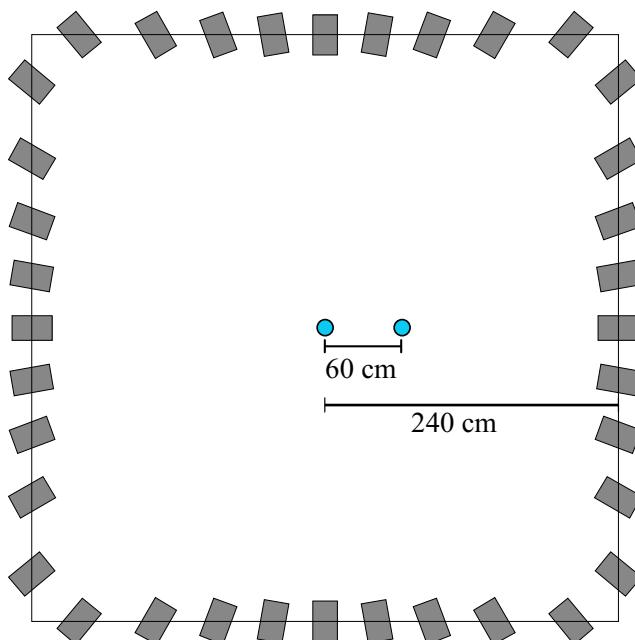
条件	ϕ_s	θ_s	方向数量
中心听音	-15°	$0^\circ, \pm 90^\circ$	3
	0°	$0^\circ, \pm 5^\circ, \pm 45^\circ, \pm 50^\circ, \pm 90^\circ$	9
	15°	$0^\circ, \pm 90^\circ$	3
	30°	$0^\circ, \pm 45^\circ, \pm 90^\circ$	5
	45°	$0^\circ, \pm 45^\circ, \pm 90^\circ$	5
	60°	$0^\circ, \pm 60^\circ$	3
	90°	0°	1
合计			29
非中心听音	0°	$0^\circ, \pm 45^\circ$	3
	30°	$\pm 45^\circ$	2
	60°	$\pm 45^\circ$	2
合计			7



(a) 定位实验流程图



(b) 音色实验流程图



(c) 听音位置

图 5-4 定位与保真度评估实验流程图与听音位置示意图

Fig. 5-4 The flow charts of localization and fidelity evaluation procedures, and diagram of listening positions

定位实验流程如图 5-4 (a) 所示。为降低学习效应对实验的影响，保证实验数据的可靠性，在正式实验前，受试者首先参与培训环节，以熟悉测试信号和方向反馈系

统。培训过程分为两个阶段：在第一阶段中，依次使用左侧($90^\circ, 0^\circ$)、前方($0^\circ, 0^\circ$)、右侧($-90^\circ, 0^\circ$)以及斜上方($0^\circ, 30^\circ$)的扬声器播放测试信号，并告知受试者信号方向，要求其使用木棍指向声源方向，系统记录方向并反馈给受试者，帮助其掌握方向反馈系统的操作。第二阶段模拟正式实验场景，实验流程与正式实验几乎一致，但测试信号直接通过对称方向的扬声器进行播放，不使用虚拟重放进行合成。受试者坐在阵列中心感知信号方向并反馈角度，训练过程中使用位于水平面，方位角 $\theta_s \in \{-90^\circ, -60^\circ, -30^\circ, 0^\circ, 30^\circ, 60^\circ, 90^\circ\}$ 的7个方向的扬声器声源、以及位于中垂面，仰角 $\phi_s \in \{0^\circ, 30^\circ, 90^\circ\}$ 的3个垂直方向的扬声器声源，合计10个方向，每个方向重复测试2次，共20个试次，试次顺序随机。受试者听到信号后需要反馈声源方向，并按下按钮确认，该训练过程中没有反馈。

正式实验中包含中心听音和非中心听音两个测试条件，测试顺序随机，使用拉丁方矩阵对次序进行平衡，两个听音位置测试条件的实验流程相同，仅是试次不同（见表5-2）。以中心听音条件为例，实验前首先通过光学追踪系统对受试者位置进行三维校准，确保受试者头中心位于阵列中心，校准完成后开始实验，受试者头部由一个头枕进行固定。实验中，每位受试者需完成1140个试次，在每个试次中，受试者需仔细聆听刺激信号的方向，然后使用木棍指向感知的声源方向，确认方向后，受试者通过一个手持按钮进行确认，按下按钮后系统记录结果；若受试者感觉声源位于头内部，则按下另一个按钮进行报告。每个试次内，受试者只能听一次信号，不允许重听。此外，为消除扬声器可能带来的视觉提示，在受试者进入消声室前，实验员拉下扬声器前的投影幕布进行遮挡。实验过程照片见图5-2（b）。

5.3.4 声音保真度评估实验

声音保真度评估实验旨在通过主观比较，探究不同虚拟音频重放算法相较于真实声源的声音听感变化，同时对比这种变化在不同虚拟声源位置和听音位置下的差异。实验采用MUSHRA测试范式进行多信号比较评分，MUSHRA测试是一种用于评估音频质量的标准化主观评价方法，受试者在该测试中，需要将多个测试信号与参考信号进行对比并评分，测试信号除了包含待测试的目标信号外，还包含一个隐藏的参考信号与一个或多个锚定信号（已知高质量和低质量的基准信号）。

与定位实验类似，声音保真度评估实验也包含中心听音和非中心听音两个条件，中心听音位置条件下受试者头中心位于阵列中心，非中心听音位置条件下受试者位于

阵列中心右侧 60 cm 处, 见图 5-4 (c)。其中, 中心听音条件下, 设置 6 个声源方向, 包括 0°、30°和 60°三个仰角平面, 非中心听音条件下包含 5 个声源位置。详细的声源位置设置见表 5-3。每个方向重复测试 5 次, 因此, 每个受试者需要进行 55 次 MUSHRA 测试[(6 + 5)个声源方向 × 5 次重复], 其中, 中心听音条件包含 30 个试次, 非中心听音条件包含 25 个试次, 实验耗时约 2 小时。

声音保真度评估实验流程见图 5-4 (b), 在正式实验前, 受试者首先接受训练, 训练环节包含两个项目。第一个项目旨在帮助受试者理解声音差异的概念, 实验员依次播放不同的信号让受试者感受声音强度、方向以及保真度差异。对于声音强度感知, 实验员在正前方扬声器直接播放粉红噪声、增大及减小 5 dB 的粉红噪声副本, 让受试者理解声音强度差异。对于声音方向感知, 实验员使用正前方、左方(10°, 0°)及斜上方(0°, 30°)的扬声器直接播放粉红噪声, 让受试者感受声源方向差异。对于声音保真度差异, 实验员使用正前方的扬声器播放原始粉红噪声、经过 3000 Hz 低通滤波器处理的粉红噪声、以及经过 3000 Hz 高通滤波器处理的粉红噪声, 让受试者感受信号在保真度上的差异。第二个训练项目包含 2 个试次的训练, 旨在让受试者熟悉实验流程和 MUSHRA 评分软件。训练试次和正式实验试次相同, 受试者坐在阵列中心进行听音, 声源位于正前方或正右方, 随机出现, 训练试次的测试结果不纳入最后的统计分析。

表 5-3 声音保真度评估实验虚拟声源方向

Tab. 5-3 The directions of virtual sources in the sound fidelity evaluation experiment

条件	ϕ_s	θ_s	方向数量	
中心听音	0°	0°, -60°, -65°	3	
	30°	-60°, -65°	2	
	60°	-60°	1	
合计				6
非中心听音	0°	-60°, -65°	2	
	30°	-60°, -65°	2	
	60°	-45°	1	
合计				5

正式实验中包含中心听音和非中心听音两个测试条件, 使用拉丁方矩阵对次序进行平衡。在实验前使用光学动捕系统校准头部中心的位置, 然后开始 MUSHRA 测试。

每次 MUSHRA 测试中，受试者需要将 7 个测试信号（包括 5 种不同重放方式、1 个隐藏的参考信号和 1 个隐藏锚点信号）与已知参考信号（单一扬声器）的声音保真度进行比较并评分。测试信号与参考信号均来自同一方向（每次 MUSHRA 测试中的方向随机，方向条件见表 5-3）。锚点刺激使用一阶 Ambisonics 进行合成重放，与参考信号相比，其声音失真较大，作为评分的参考基准。

评分范围为 0 至 100 分，0 分表示测试信号听感与参考信号一样，100 分表示二者差距较大，受试者通过操作平板电脑上显示的 GUI 界面完成所有评分（见图 5-2 (c)）。刺激信号的排序随机生成，受试者可通过按钮自由播放或切换刺激信号。受试者需在选择并收听某个刺激信号后才能对其进行评分，整个过程中保持面向前方，不能转动或移动头部。受试者每进行半小时的评分实验或主动反疲劳后，需要休息一段时间后继续才能继续实验。

5.4 实验结果与讨论

5.4.1 定位实验结果与讨论

受试者对每个虚拟声源反馈有两种反馈结果，包括感知方向 $\Omega_p(\theta_p, \phi_p)$ 或头中定位。对于每个实验条件（听音位置、重放方法、虚拟声源方向），共有 80 个样本值（10 受试者 \times 8 次重复）。

5.4.1.1 头中定位和前后混淆

首先分析受试者的头中心定位与前后混淆率。头中心定位结果由受试者主动反馈得到，前后混淆是指受试者将原本在前半平面的声源定位在后半平面。若条件 $\{(\theta_s, \theta_p) | |\theta_p - \theta_s| \geq |\theta_p + \theta_s|, |\theta_s - 90^\circ| \geq 30^\circ, |\theta_s - 270^\circ| \geq 30^\circ\}$ 成立，即当声源方向距离耳间轴（正左至正右）超过 30° ，且反馈方位角的前后镜像离实际声源的方位角更近时，认为受试者产生了前后混淆^[149]。

统计不同听音位置与重放方法下，所有虚拟声源与所有受试者的整体头中定位率与前后混淆率，结果如图 5-5 所示。

可以看到，头中定位率在整个实验结果中比例较低，中心听音和非中心听音两个条件下的头中定位没有明显的区别。在中心听音和非中心听音的情况下，AllRAD、混合阶 1 和分层 Ambisonics 的头中定位率更低，混合阶 2 与混合阶 3 的头中定位率相对较高，但也在 3% 以内。对于前后混淆率，非中心听音下的混淆率明显高于中心听音下的混淆率。在中心听音的情况下，5 种重放方式的混淆率均较低，低于 1%，其中混合阶

2 和混合阶 3 相对高一些。在非中心听音的情况下，混合阶 2 和混合阶 3 的混淆率显著提高，AllRAD 也有较高的混淆率，混合阶 1 和分层 Ambisonics 混淆率接近。

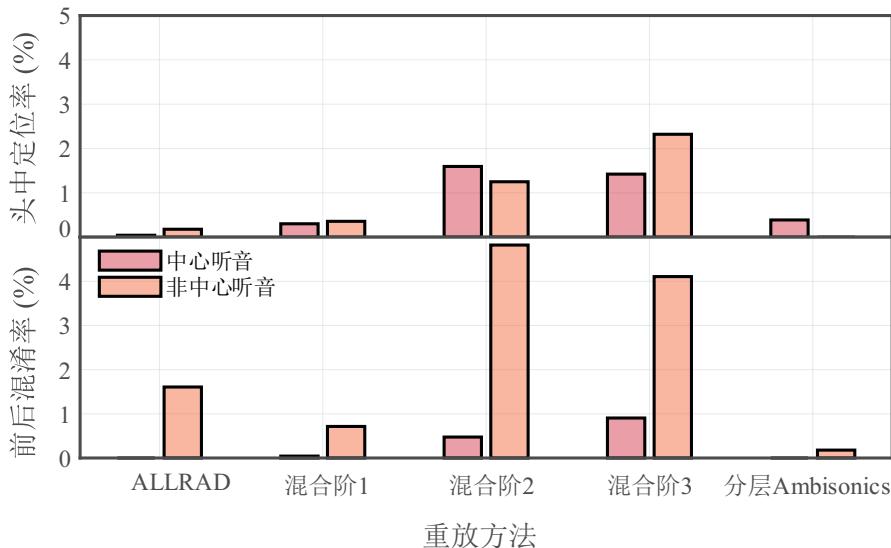


图 5-5 头中定位率与前后混淆率条形图

Fig. 5-5 The bar chart In-head localization rate and front-back confusion rates

5.4.1.2 定位误差分布

在上一节分析的基础上，排除所有头中定位和前后混淆的样本，分析各个条件下受试者的三维定位准确度，统计受试者感知方向 Ω_p 的分布情况^[150]，使用 Kent 和 Fisher 模型进行统计分析。Kent 分布与 Fisher 分布是球面统计学中两种重要的概率分布模型，Fisher 分布用于统计旋转对称模型，即反馈方向的分布成圆形；Kent 分布用于统计非对称模型，反馈方向的分布成椭圆形。结果如图 5-6 所示。图中标出了每个条件下目标声源的方向和受试者的主观反馈方向，若反馈方向空间对称，则用 Fisher 分布画出，否则使用 Kent 分布画出。从图 5-6 中可以看到，在中心听音和非中心听音两个条件下，不同重放方式的声源定位中，仰角定位误差范围均要大于方位角定位误差范围。在图 5-6 (b) 所示的非中心听音的情况下，定位误差相比中心听音的情况均较大。随着虚拟声源仰角增大，定位误差范围也逐渐增大。

5.4.1.3 水平面方位角定位误差

进一步对不同方向虚拟声源的定位进行分析，由于中心听音条件和非中心听音条件下测试的虚拟声源方向不同，需要对二者分别分析。首先分析在中心听音的情况下，

水平面虚拟声源的方位角绝对定位误差，方位角绝对误差定义为 $\Delta\theta = |\theta_s - \theta_p|$ ，其结果如图 5-7 所示，从图中可以看到，整体上混合阶 1 算法和分层 Ambisonics 算法相对其他方法误差较低。

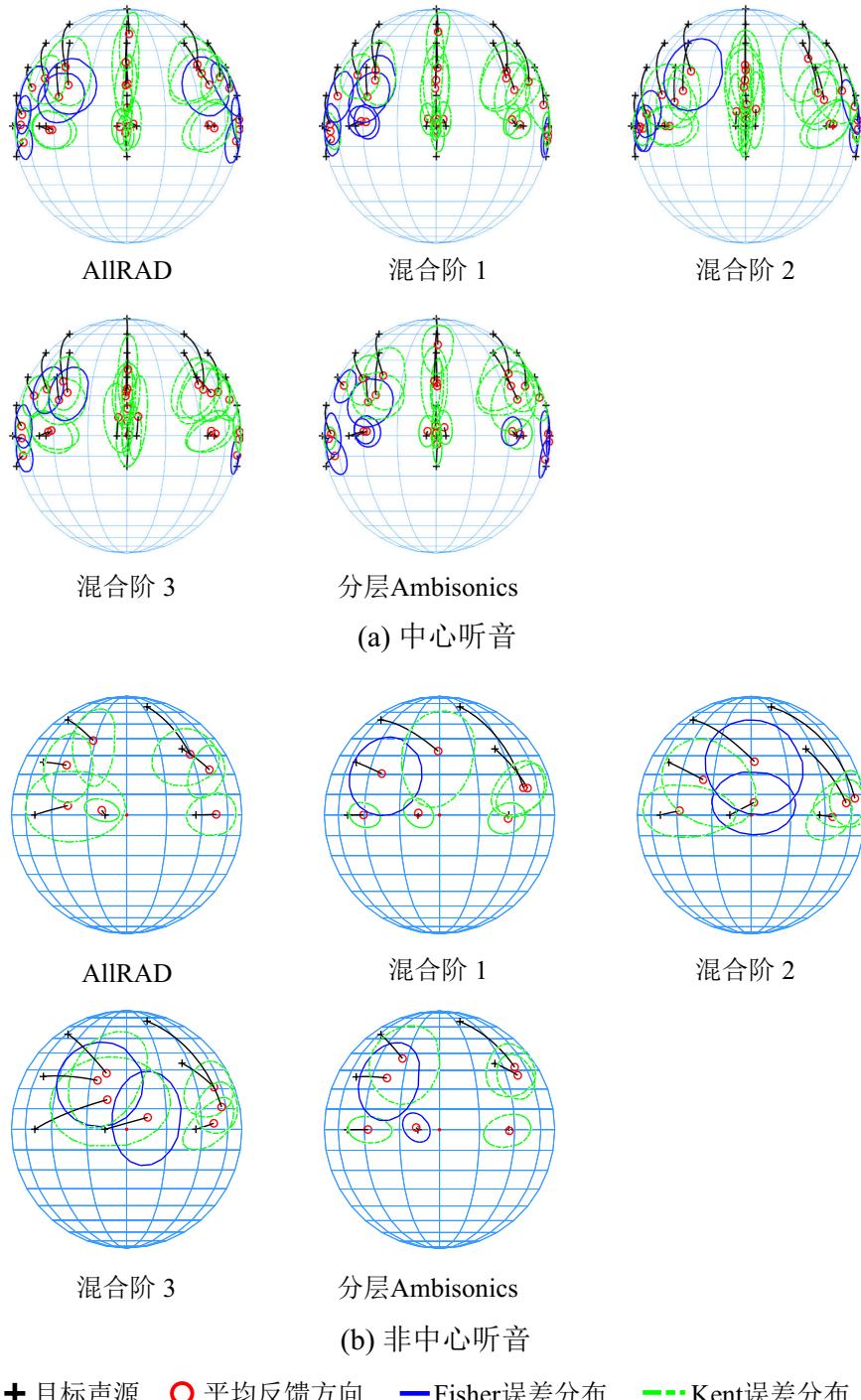


图 5-6 定位实验中不同重放方法的反馈方向分布

Fig. 5-6 The distributions of feedback directions under different reproduction methods in the localization experiment

使用线性混合效应模型对定位误差进行方差分析（R *lme4*），组内因素为虚拟声源方向以及重放方法。结果表明重放方法与声源方向均对方位角定位误差具有显著效应[F(4, 3503) = 28.22, $p < 0.001$; F(4, 3503) = 28.22, $p < 0.001$]，且二者存在交互效应[F(32, 3503) = 2.15, $p < 0.001$]。

进一步使用 Tukey 事后多重检验显示，在特定方向上（见图 5-7 (b)），分层 Ambisonics 算法的方位角定位误差显著低于混合阶 2 算法与混合阶 3 算法（对于混合阶 2，方位角为 270°、310°、315°、45°；对于混合阶 3，方位角为 310°、315°、45°、5°）；混合阶 1 在 310°、315° 以及 45° 方向上定位误差低于混合阶 2，在 310°、31°、45° 和 5° 方向上优于混合阶 3。此外，AllRAD 算法在 310° ($p < 0.05$) 和 315° ($p < 0.001$) 方向的表现略低于分层 Ambisonics 算法，在 315° 方向误差高于混合阶 1 ($p < 0.001$)。

相比中心位置的声源，侧向声源下所有算法的方位角定位误差均显著增大。在水平面上，对比虚拟声源位于真实扬声器方向的情况（方位角为 0°、±50°）与位于真实扬声器以外的情况（方位角为 ±5°、±45°），所有算法的定位误差均没有显著区别。

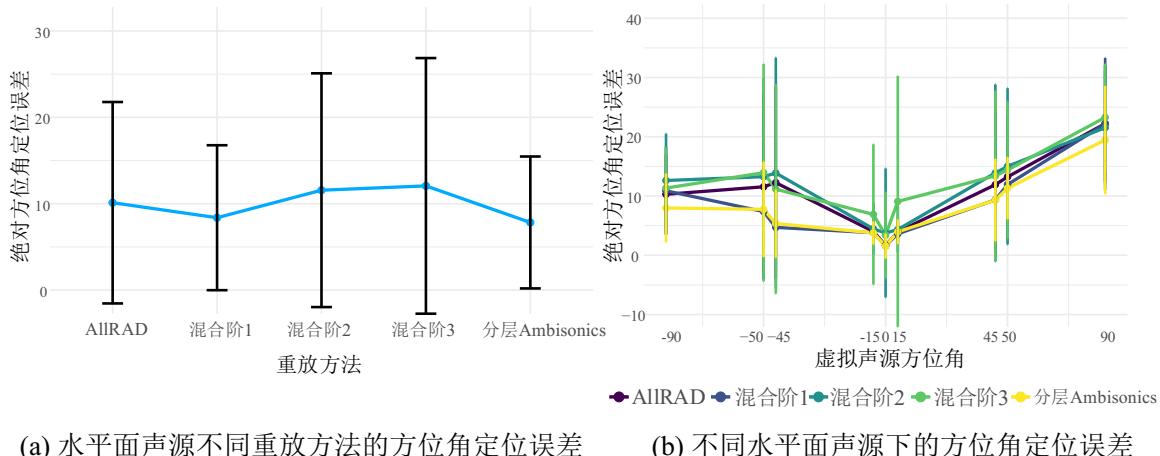


图 5-7 中心听音情况下水平面声源的方位角绝对定位误差

Fig. 5-7 Absolute azimuth localization errors of sound source in horizontal plane under centre listening condition

5.4.1.4 仰角定位误差

类似地，进一步分析绝对仰角定位误差，其定义为 $\Delta\phi = |\phi_s - \phi_p|$ ，结果如图 5-8 (a) - (d) 所示，图 5-8 (a) 显示 AllRAD、混合阶 1 与分层 Ambisonics 算法的仰角定位误差较低。

首先对中心听音情况下中垂面 ($\theta_s = 0^\circ$) 的所有虚拟声源方向使用线性混合效应模型进行分析(见图 5-8 (b) 与 (c))，结果显示重放方法和虚拟声源方向均对绝对仰角定位误差有显著影响 [$F(4, 2686) = 28.36, p < 0.001$; $F(6, 2686) = 324.20, p < 0.001$]，且二者存在交互效应 [$F(24, 2686) = 4.66, p < 0.001$]。对于正前方声源，混合阶 2 与混合阶 3 算法的仰角定位误差显著高于混合阶 1、AllRAD 和混合算法；随着仰角升高，各算法定位误差均增大，其中分层 Ambisonics 算法在 30° ($p < 0.001$) 和 45° ($p = 0.04$) 仰角下的误差显著低于混合阶 2。对于头顶的声源($0^\circ, 90^\circ$)，混合阶 2 和混合阶 3 算法的表现远差于其他算法 ($p < 0.001$)。进一步对比不同算法在中垂面不同声源方向下的定位误差，可以看到，在中垂面上，前方声源的仰角定位误差较大，其中混合阶 2 算法 [$F(4, 3523) = 26.47, p < 0.001$]、混合阶 3 算法 [$F(4, 3523) = 27.60, p < 0.001$] 和 AllRAD [$F(4, 3523) = 2391, p = 0.02$] 算法下，中垂面不同声源方向的仰角定位误差显著不同。

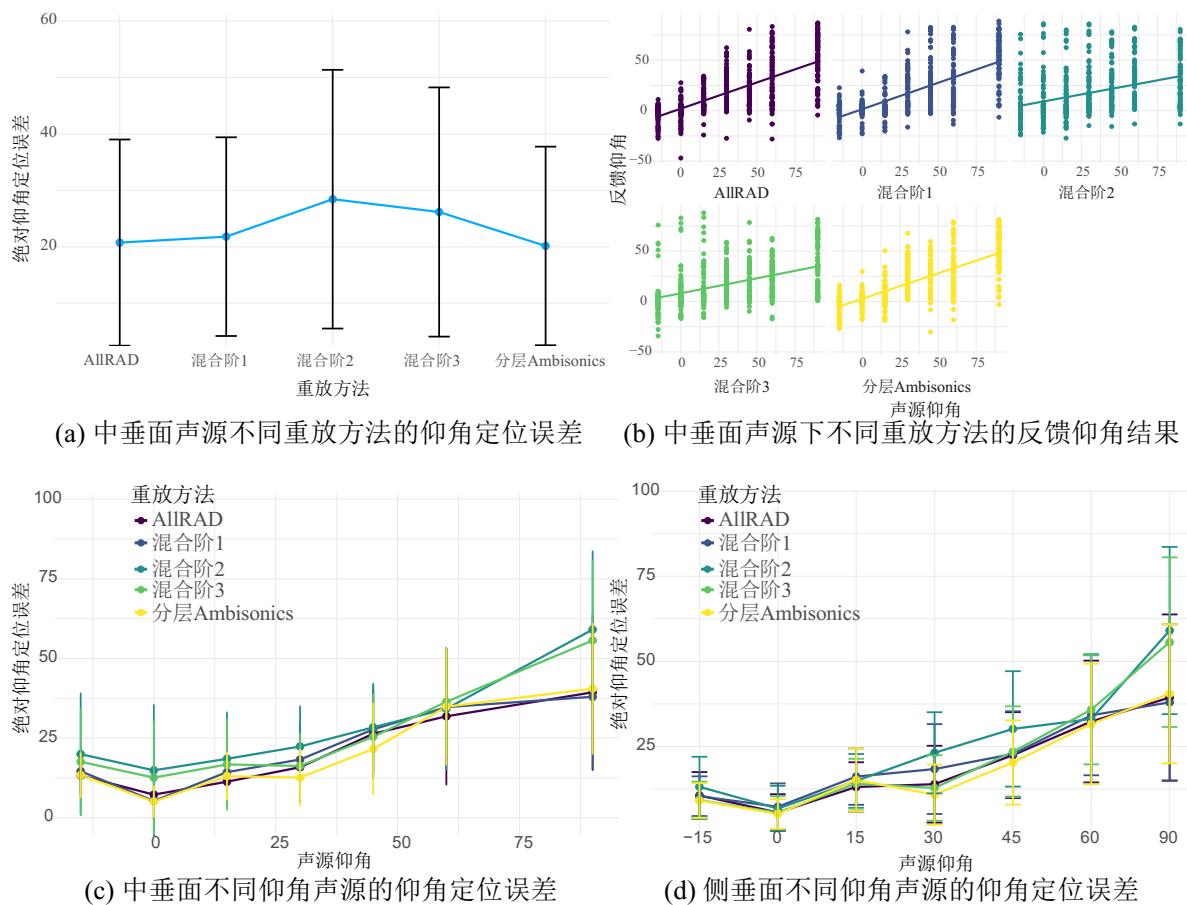


图 5-8 中心听音情况下中垂面和侧垂面声源的仰角定位绝对定位误差

Fig. 5-8 Absolute elevation localization errors of sound source in the median and lateral planes under the centre listening condition

侧垂面（90°和 270°方位角）的结果见图 5-8 (d)，重放方法[F(4, 5133) = 52.61, $p < 0.001$]和仰角[F(5, 5133) = 869.97, $p < 0.001$]均有显著主效应，二者具有显著的交互效应[F(20, 5133) = 12.37, $p < 0.001$]。简单效应分析表明，各算法在高仰角（30°、45°和 60°）及低仰角（-15°）下的定位存在差异。在 30°仰角时，混合阶 1 和混合阶 2 算法的定位误差显著高于其他算法（ $p < 0.001$ ）；45°仰角时，混合阶 2 算法表现最差（ $p < 0.001$ ）；60°仰角时，混合阶 3 算法的误差显著高于分层 Ambisonics 算法（ $p = 0.01$ ）。

5.4.1.5 不同虚拟声源仰角下的垂直定位误差

进一步对不同仰角的全部虚拟声源进行定位误差分析，结果如图 5-9 所示。简单效应分析显示，仅在水平面声源定位时（图 5-9 (a)），算法间的绝对方位角误差存在显著差异；在绝对仰角定位误差方面（图 5-9 (b)），混合阶 2 算法在各虚拟声源仰角下（除 15°仰角）的误差均显著高于其他算法。AllRAD 和混合阶 3 算法的仰角定位误差相对较小，在 15° 和 30° 仰角时，AllRAD 算法显著优于混合阶 1 ($p < 0.05$)；混合阶 1 算法在水平面的定位误差较低，但随仰角升高误差快速增大；分层 Ambisonics 算法在除 15°仰角外的各仰角下，定位误差均最低，尤其是在存在实际扬声器的仰角（0°和 30°）下，该算法相对其他算法误差更低。

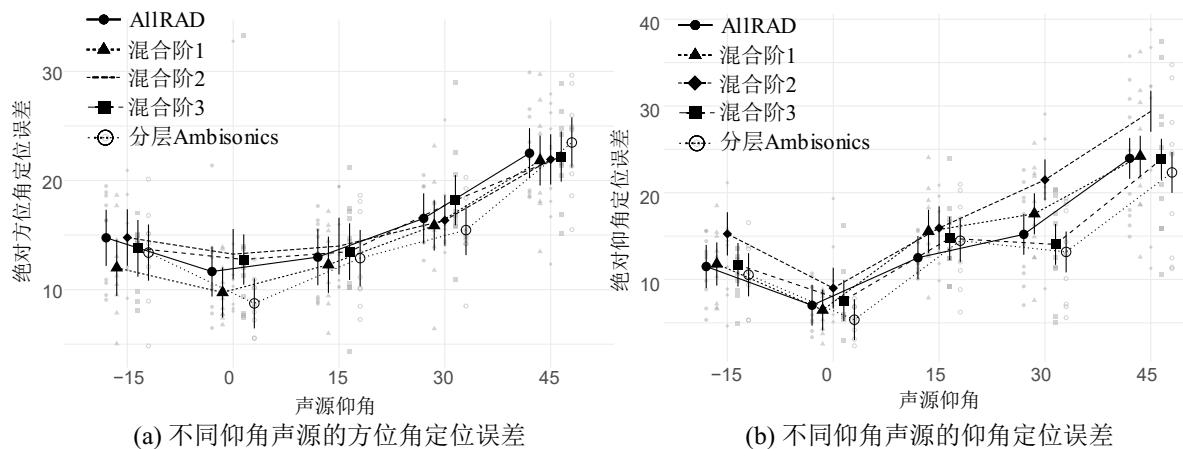


图 5-9 中心听音情况下不同仰角声源的方位角和仰角绝对定位误差

Fig. 5-9 Absolute azimuth and elevation localization errors of sound sources in different elevation planes under centre listening condition

5.4.1.6 非中心听音条件下的误差

在非中心位置听音的场景下，对于水平面上的虚拟声源的方位角与仰角定位结果如图 5-10 所示，作为对比，图中还画出了中心听音条件的结果。

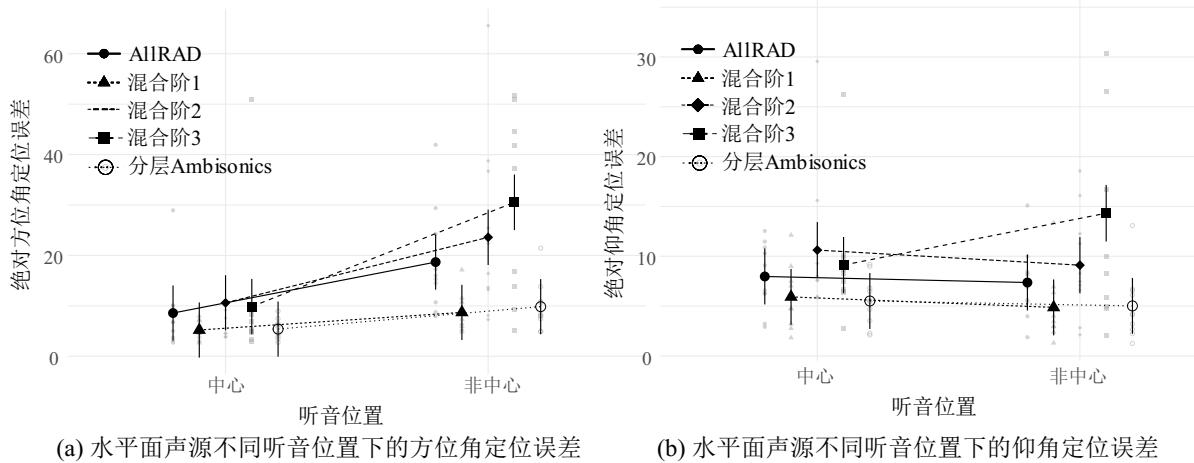


图 5-10 水平面声源在两个听音位置下的绝对方位角与仰角定位误差

Fig. 5-10 The absolute azimuth and elevation localization errors of sound sources in the horizontal plane under centre listening and off-centre listening conditions

在图 5-10 (a) 中, 听音位置对绝对方位角定位误差有显著的主效应 [$F(1, 2307) = 298.33, p < 0.001$] , 图 5-10 (b) 中, 听音位置对绝对仰角定位误差无显著影响 [$F(1, 2307) = 0.56, p = 0.45$]。混合阶 1 和分层 Ambisonics 算法在聆听者偏离中心位置时, 方位角定位误差增加的程度相对较小。在 30°仰角上, 所有算法的方位角定位误差均显著增大, 而 AllRAD ($p = 0.52$) 和分层 Ambisonics 算法 ($p = 0.46$) 在两种听音位置下的仰角定位误差无显著差异; 60°仰角下的定位也是类似的结果。

5.4.2 声音保真度评估实验结果与讨论

声音保真度评估实验要求受试者在不同条件 (声源方向和听音位置) 比较不同重放方法的虚拟声源与参考声源的保真度差异, 每个刺激有 50 个样本点 (10 受试者 \times 5 重复), 每个样本点为 0-100 的评价分数。在中心听音和非中心条件下, 所有声源方向的总体结果如图 5-11 (a) - (b) 的箱型图所示, 箱型图的定义与图 3-6 一致, 可以看到不同重放方法的声音失真评分差异较大。

图 5-11 (a) - (b) 中存在部分离群值, 在中心听音位置条件下, 分层 Ambisonics 与隐藏参考条件分别有 2 个与 5 个离群值 (占比 4% 与 10%) ; 在非中心听音位置条件下, 混合阶 1、分层 Ambisonics 与隐藏参考条件分别有 3 个、1 个与 6 个离群值 (占比 6%、2% 与 12%)。离群值主要出现在评分较低 (失真小) 的重放方式上, 这是由于这些条件与参考信号非常接近, 受试者对失真大小的判断相对较为困难。考虑到这些离

群数据与 MUSHRA 测试范式有关，且不同重放方法中的离群值数量较少，同时隐藏参考条件不是本文的分析重点，后续分析中不排除以上离群值。

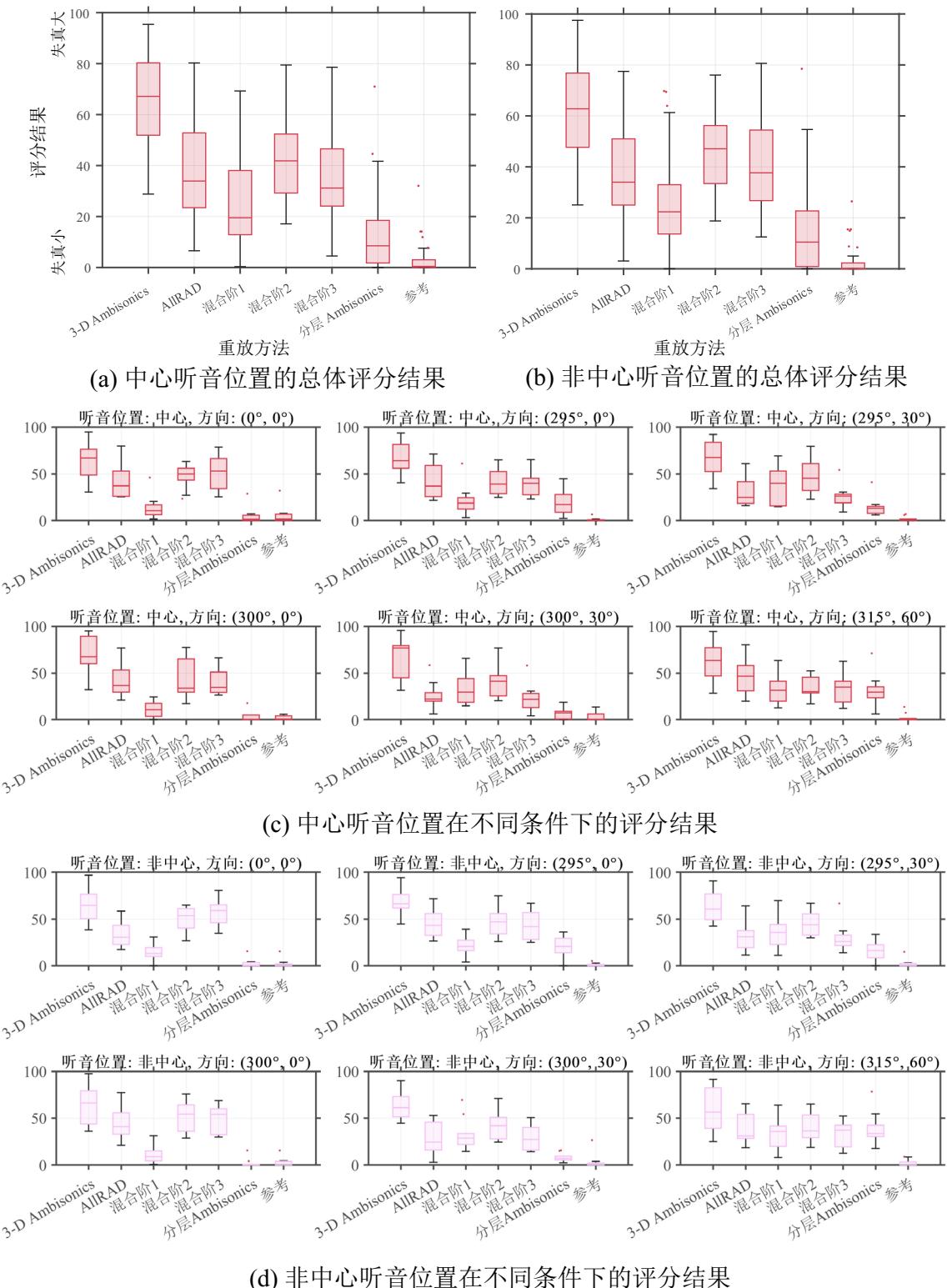


图 5-11 声音保真度评估实验结果

Fig. 5-11 Results of the sound fidelity evaluation experiments

按照 5.4.1 节定位实验结果分析中所使用的统计方法，进一步对结果进行线性混合模型的方差分析，结果显示，重放方法 [$F(6, 747) = 488.81, p < 0.001$] 与声源方向 [$F(5, 747) = 4.80, p < 0.001$] 对声音失真评分结果有影响显著，但听音位置对评分无显著的主效应 [$F(1, 747) = 8.17, p = 0.14$]；重放方法与声源方向存在显著交互效应 [$F(30, 747) = 13.82, p < 0.001$]，重放方法与听音位置存在微弱的交互效应 [$F(6, 747) = 2.11, p = 0.05$]，没有发现显著的三因素交互效应 [$F(30, 747) = 0.48, p = 0.99$]。

分别对每个条件下的结果（图 5-11 (c)）进行统计分析，在中心聆听位置下，对于水平面声源，分层 Ambisonics 算法和混合阶 1 算法的保真度显著优于混合阶 2、混合阶 3 和 AllRAD 算法 ($p < 0.001$)；在 30°仰角平面，分层 Ambisonics 的声音保真度显著优于混合阶 1、混合阶 2 和 AllRAD 算法，且混合阶 1 和混合阶 2 算法在此仰角平面的表现相比其他重放方法明显变差；当仰角升至 60°时，算法间差异缩小，AllRAD 算法的失真显著增大，表现显著差于分层 Ambisonics 和混合阶 1 算法。整体来看，混合阶 1 和 AllRAD 算法在 30°和 60°高仰角时失真更大，分层 Ambisonics 在除 60°仰角外的条件下表现良好，混合阶 3 算法在高仰角下相对表现较好。

在声源方向影响方面（图 5-11 (c) - (d) 中的各子图），对比前方和侧方的声源，各重放方法之间未发现显著的失真大小差异。对比声源位于实际扬声器位置与偏离位置的情况（60°和 65°），分层 Ambisonics 在水平面声源的条件下，在中心 ($p < 0.05$) 和非中心听音的条件 ($p < 0.001$) 均存在显著差异；混合阶 1 算法在声源偏离实际扬声器位置时，声音失真程度增大，但未达显著水平（中心听音条件 $p = 0.269$ ，非中心听音条件 $p = 0.18$ ）。

由于实验设计特性，受试者没有直接比较中心与非中心位置的声音失真，上述结果仅能用于粗略的比较分析。但通过上述分析仍清晰地呈现不同算法在多种条件下的声音保真度表现特征，为深入理解算法性能、优化空间音频技术的声音还原效果提供了重要依据。

5.5 本章小结

本章围绕分层扬声器阵列下的混合阶 Ambisonics 定位、声音保真度感知及重放算法改进展开研究，开展了主观定位和保真度评估实验，对比了分层 Ambisonics 算法与现有的混合阶 Ambisonics 及 AllRAD 算法的性能。主观实验表明，分层 Ambisonics 在中心及非中心听音时，对水平面和 30°仰角平面虚拟声源，方位角定位误差显著低于混

合阶 Ambisonics 及 AllRAD 算法。在非中心听音的情况下，所有算法的定位误差都较大，但分层 Ambisonics 与混合阶 1 算法误差上升的幅度较小。分层 Ambisonics 算法在水平面失真更小，高仰角平面表现稳定。高仰角平面（如 60° ）下各算法虚拟声源听感相比参考条件的失真都较大，算法间的差异缩小，但 AllRAD 算法失真显著增大，而分层 Ambisonics 算法与混合阶 1 表现相对稳定。本研究验证了分层 Ambisonics 算法在非均匀阵列中的实际效果，同时还对比验证了混合阶 Ambisonics 策略在提升水平面感知精度中的价值，弥补了该领域研究的数据缺失，为虚拟现实、建筑声学等领域的声场可听化重放提供了理论与实验依据。

第六章 高阶 Ambisonics 可听化下虚拟厅堂的空间感知研究

相比本文第三章中的自由场语音信号，室内混响环境的特点是具有大量空间分布的反射声源，准确地重放三维反射声源需要大量扬声器。本文的第四章与第五章提出了一种新的算法，降低重放对扬声器数量与均匀性的要求。受限于物理限制，基于算法的优化不能在保证重放精度的前提下，无限地提高重放精度，进一步的优化需要结合听觉在特定声环境下的感知特性。听觉对声源的高度信息相对不敏感，针对这一听觉特性，对室内反射场景的重放实现进一步的优化是可能的。综上所述，本章开展实验研究重放室内反射声的高度信息及空间分辨率对主观感知的影响。

6.1 引言

在 1.3.5 节中，我们提到高阶 Ambisonics 声场重放的两个重要特点是：（1）能够实现在一定区域内的物理声场重构；（2）能够对三维空间的声源进行重放，即能够重放具有高度信息的声源。在第五章内，我们详细探讨了高阶 Ambisonics 在三维重放时对扬声器数量与空间均匀分布的严苛要求，并提出了一种新算法以降低三维重放对扬声器的数量及分布要求，同时，对自由场单个声源的重放进行了主观评价实验。

对具有空间反射的室内声场进行精确的重放是可听化技术研究的核心目标，将基于扬声器的可听化系统从二维重放扩展到三维重放，理论上能提高虚拟声场下听众的感知体验。相比二维重放，三维重放能够保留具有高度信息的高仰角方向反射声，但实现高仰角反射的重放仍然需要投入大量的扬声器成本。因此，实际系统需要在明确高仰角反射对虚拟声场重放感知影响的前提下，权衡重放的精度与系统的可实现性。目前尚未有研究探讨在可听化环境下，高仰角反射如何影响听众的感知。在室内声场感知的维度中，反射的仰角方向不改变混响时间与早期、后期反射能量比例，主要的潜在影响对象是室内声环境下的空间感。此外，提高 Ambisonics 重放阶数也能提高重放空间分辨率，但提高阶数需要更多的重放扬声器，阶数对虚拟重放声场下的空间感知影响也缺少足够的研究。

综上所述，本章使用本文所建立的高阶 Ambisonics 可听化系统，探讨在可听化重放的虚拟室内声环境下：（1）高仰角反射以及（2）空间分辨率对空间感的感知影响。本章的内容安排如下：6.2 节介绍了研究与实验的基本背景与概念，着重介绍了相关研究对反射声高度信息感知以及重放阶数对空间感影响的探索；6.3 节详细描述了主观评估实验的实验设计与信号生成方法，并对不同信号条件下重放声场的客观室内声场指

标进行了测量；随后，6.4 节报告了实验结果并进行了一定的讨论；最后，6.5 节对本章的主要内容与结论进行了总结。

6.2 研究背景

6.2.1 反射声仰角对空间感的影响

Barron 和 Marshall 等人最早系统地研究了反射方向对空间感的影响，他们使用真实扬声器调整早期反射的方向、延时和强度，探究受试者在不同条件下的空间感，实验使用包含直达声与单一反射的简化人工脉冲响应，结果表明反射声仰角对空间感的贡献基本符合 $\sin(\phi)$ 的变化规律，正上方的反射具有最强的空间感^[151]。上述研究的反射声条件较为简单，更进一步地，Furuya 等人使用包含多个反射的更复杂的反射脉冲响应，将早期反射分为两组，分别从水平面扬声器与位于 50°仰角的扬声器发出，调整两组反射的能量比，结果发现增加高仰角反射的能量比可以增强视在声源宽度 ASW 与听众包围感 LEV^[152]。类似地，Evjen 等人通过八个扬声器直接播放反射，研究了高仰角侧向反射对 LEV 的影响。他们的研究发现，在有、无高仰角侧向反射的两个条件下，LEV 感知差异很小，侧向反射的能量占比主导了 LEV 的感知^[153]。上述几个研究均是采用少量扬声器直接播放反射，在扬声器阵列进行虚拟可听化的研究中，Power 等人的研究比较了单声道、立体声、5.1 声道、二维 Ambisonics 和三维 Ambisonics 重放系统下的 LEV 感知，着重对比三维重放和二维重放条件下 LEV 感知。结果表明，在部分声音场景中，具有高仰角扬声器通道的三维重放提升了听众的 LEV 评分^[154]。在一个更复杂的研究中，Dick 使用球形传声器阵列拾声并使用 3 阶 Ambisonics 在扬声器阵列重放室内声场，实验对混合拾声脉冲响应的早期与后期部分进行混合，让受试者对多个具有不同反射方向、延时及能量的混合后的虚拟声场进行空间感评分，根据这些结果建立了一个包含反射方位角、仰角、延时与能量的 LEV 预测指标^[155]。该研究对虚拟声场环境下的空间感进行了重要的探索，但没有直接对比不同反射仰角对空间感的影响。

综上所述，已有研究表明反射仰角可能会影响空间感，但过去的研究尚未在复杂的反射声环境下，特别是在使用高阶 Ambisonics 重放的虚拟复杂环境下，直接对比反射仰角方向对空间感的具体影响。

6.2.2 重放阶数对空间感的影响

重放声场的空间分辨率也可能影响虚拟室内声环境的空间感。Avni 等人探究了 Ambisonics 重放的阶数对听众感知的影响，该研究使用双耳 Ambisonics 技术重放虚拟

音乐厅，受试者反馈更高阶的 Ambisonics 能带来更精准的定位、更明亮的音色以及更强的空间感^[147]。Nowak 与 Klockgether 等人的研究采用了相似的可听化重放方法，但更具体地探究了阶数对 ASW 和 LEV 的影响^[156]。他们发现，总体而言，ASW 和 LEV 均随 Ambisonics 阶数的提升而增强，但不同虚拟房间下的实验结果并不完全一致，该研究同样使用双耳 Ambisonics 重放。Ahrens 与 Andersson 等人对双耳 Ambisonics 重放和双耳拾拾重放技术进行了更直接的 AB 对比实验^[146]。结果表明，对于前方声源，阶数对空间感和音色影响均不明显；而对于侧向声源，需要高达 8 阶重放才能具有与人工头拾拾重放条件接近的重放效果，提高阶数后 ASW 与 LEV 评分有所上升。目前的研究探讨了重放阶数如何影响双耳 Ambisonics 可听化重放下的空间感知，但对于使用扬声器阵列的实际 Ambisonics 重放，尚未有研究开展测试。

此外，本文第四章所提出的重放算法与传统 Ambisonics 重放的空间感的感知区别，也需要进一步探究。

6.3 高阶 Ambisonics 可听化重放的空间感评估实验

6.3.1 实验目的

针对上一节所述的 Ambisonics 可听化研究的不足，本节开展两个主观听音实验，分别探讨在虚拟可听化中，重放时的反射声仰角范围与重放时的 Ambisonics 阶数及算法对空间感感知的影响。室内音质的空间感主要包含声源宽度与包围感（LEV）两个维度，分别对其进行探讨。结合两个实验，总结反射声三维空间信息对空间感主观感知的影响程度与其重要性。

实验 1 探索反射声重放仰角范围对空间感的影响，如前所述，三维重放尽管可能对提升虚拟声环境的空间感有益，但重放三维反射需要大量的扬声器。本实验探讨不同的早期或后期反射声重放仰角范围如何影响包围感的感知，实验对比变量包括：早期反射与后期反射的重放仰角范围，包括不同范围大小，二维与三维编码重放；不同的虚拟房间。本实验的实验假设为，重放反射声仰角范围对 LEV 感知有影响，仰角范围越小，可听化重放的 LEV 越差；

实验 2 探讨早期反射在不同重放方式及重放阶数编码条件下对空间感的影响，旨在确定反射声能量扩散程度对声源宽度与包围感的影响。实验条件包括：早期反射的重放方式，包括不同阶数的 Ambisonics 重放以及分层 Ambisonics 重放；不同的虚拟房间，本实验的实验假设为：早期反射声重放的能量扩散对空间感存在一定影响。

6.3.2 实验对象与刺激信号

6.3.2.1 实验对象

本实验招募 16 名听力正常的受试者，其中包括 8 名女性，8 名男性。受试者以在校大学生为主体，受试者的年龄范围为 20 至 31 岁，平均值为 23.9 岁，标准差为 3.02 岁，所有受试者均无中耳炎及相关耳科病史。实验开始前，使用经过校准的自动测听平台对所有受试者进行听力测试^[113]。所有受试者在 500 到 4000 Hz 倍频程的平均气导听阈不超过 20 dB HL。受试者均在得知实验目的与大致实验内容的基础上，签署实验知情同意书后自愿参与实验，并获得了一定的实验酬金。

6.3.2.2 实验刺激

本实验使用 2.3 节所述的高阶 Ambisonics 可听化重放系统进行声场重放，首先计算多通道的房间脉冲响应（流程见图 2-13），与干信号进行卷积后，馈给扬声器从而实现室内声场景可听化。在两个实验中，干信号均为一段在消声室录制的意大利歌剧，信号包含人声与器乐声^[157]，该录音原始信号长 106 s，选取其中 54 s 至 65 s 约 11 s 部分高潮片段作为测试干信号，并对信号进行淡入淡出处理。该干信号的波形图与长时平均谱如图 6-1 所示。

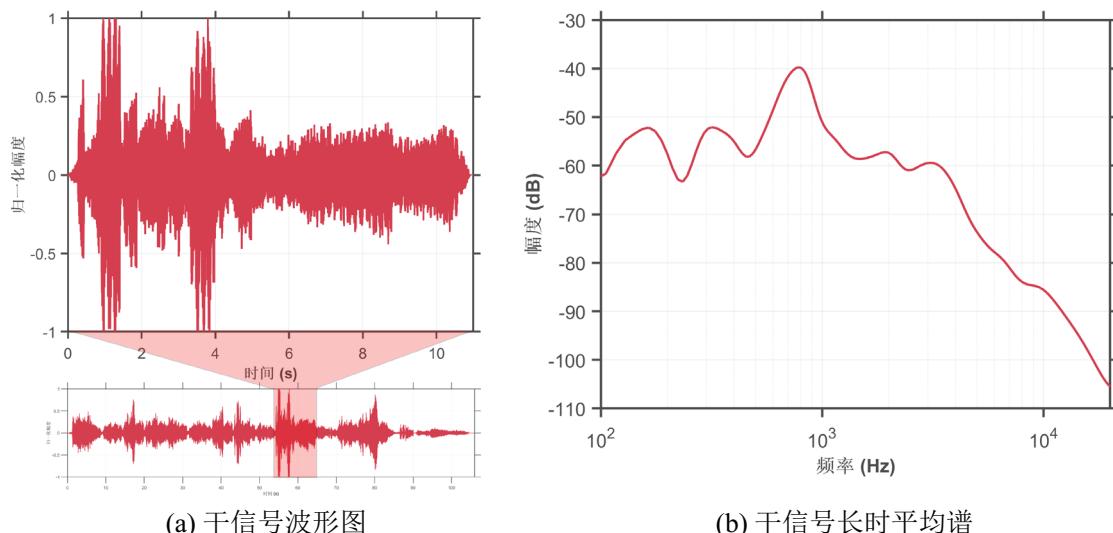


图 6-1 干信号时域波形图和长时平均谱

Fig. 6-1 The time-domain waveform and long-time average spectrum of the dry signal

实验使用两个房间作为可听化对象：房间 1 为一个小型多功能音乐厅；房间 2 为一个矩形房间，分别用于探索实际厅堂与简化房间下的可听化感知效果。两处房间的三维模型如图 6-2 所示。

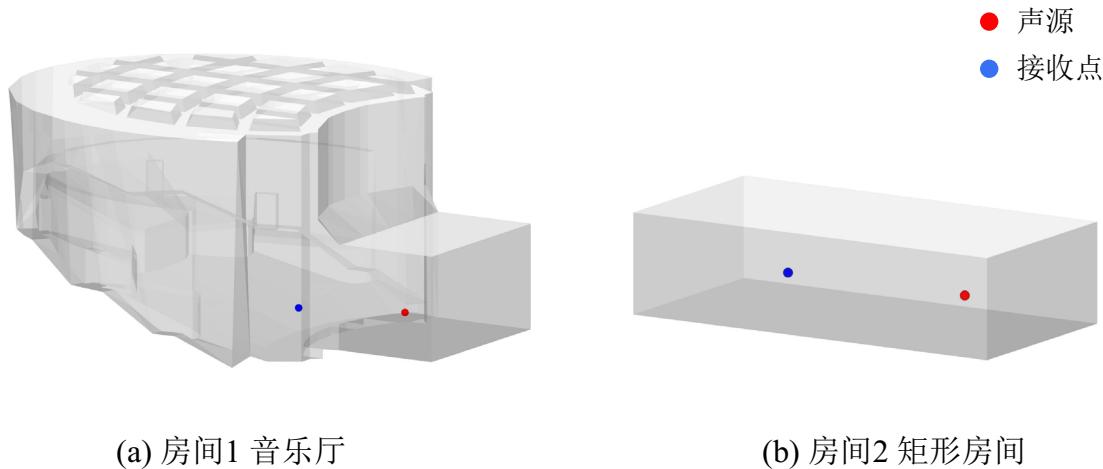


图 6-2 实验房间三维模型透视图

Fig. 6-2 Perspective pictures of 3D room models of the experiment

房间 1 的最大长、宽、高为 $33.7 \times 26.0 \times 15.6$ m，总表面积为 3101.7 m²，包含 500 个坐席。该多功能音乐厅截面近似椭圆形，座位区呈山地葡萄园式布局，由池座及 3 层楼座组成。音乐厅池座最长约为 20 m，池座进深约为 11 m。天花吊顶距 3 层楼座最高处约为 5.7 m，距池座最低处约为 14.7 m。音乐厅体积为 7154.9 m³。房间各界面的吸声系数与扩散系数根据实际材料进行设置，如表 6-1 所示。声源设置在舞台中心处（见图 6-2），距离后墙 5.70 m，高度为 1.50 m，接收点位于前排座位，距离声源点 11.0 m，高度为 1.40 m，接受点正对声源。

房间 2 为简化的矩形房间，长、宽、高为 $20.0 \times 10.0 \times 5.0$ m，总表面积为 700.0 m²，所有界面的各频带吸声系数 α 均设置为 0.1，散射系数为 0.05。声源设置在房间中轴线位置，距离房间中心点 5.0 m，高度为 1.50 m，接收点位于对称位置，高度也为 1.50 m，接收点正对声源。

使用 Odeon 分别对两个房间的房间反射进行计算，早期反射计算阶次设置为 4 次，计算完成后，按照图 2-6 与图 2-13 所示流程分别对早期反射与后期反射进行编码，计算多通道 RIR，实现可听化重放。

表 6-1 音乐厅模型主要界面材料吸声系数与散射系数

Tab. 6-1 Sound absorption and scattering coefficients of materials in the concert hall model

界面	吸声系数 α								散射系数
	63	125	250	500	1000	2000	4000	8000	
墙面	0.15	0.15	0.20	0.10	0.10	0.10	0.10	0.10	0.30
显示屏	0.18	0.10	0.06	0.04	0.03	0.02	0.02	0.02	0.10
楼座护栏	0.15	0.15	0.20	0.10	0.10	0.10	0.10	0.10	0.20
后墙	0.15	0.15	0.20	0.10	0.10	0.10	0.10	0.10	0.10
木墙	0.15	0.15	0.20	0.10	0.10	0.10	0.10	0.10	0.10
走道	0.15	0.15	0.11	0.10	0.07	0.06	0.07	0.07	0.10
天花板	0.15	0.15	0.20	0.10	0.10	0.10	0.10	0.10	0.40
坐席	0.51	0.51	0.64	0.75	0.80	0.82	0.83	0.83	0.70
舞台地面	0.15	0.15	0.11	0.10	0.07	0.06	0.07	0.07	0.10
舞台侧墙	0.15	0.15	0.20	0.10	0.10	0.10	0.10	0.10	0.40
舞台天花板	0.15	0.15	0.20	0.10	0.10	0.10	0.10	0.10	0.40
舞台后墙	0.15	0.15	0.20	0.10	0.10	0.10	0.10	0.10	0.40
地面	0.15	0.15	0.11	0.10	0.07	0.06	0.07	0.07	0.10
门	0.14	0.14	0.10	0.06	0.08	0.10	0.10	0.10	0.10
窗户	0.18	0.18	0.06	0.04	0.03	0.02	0.02	0.02	0.05

在早期反射编码时，将计算得到的反射声的仰角 ϕ_{cal} 压缩到不同的范围 $[\phi_{\min}, \phi_{\max}]$ 内，探讨反射重放仰角范围对感知的影响。同时采用不同的重放方式对早期反射进行编码，探讨早期反射编码方式的空间扩散对空间感的影响。若反射的仰角在设定的范围外，则将该反射的仰角移动至设定范围的边界上，最终反射仰角 ϕ 为：

$$\phi = \begin{cases} \phi_{\max}, & \phi_{cal} > \phi_{\max} \\ \phi_{cal} & \\ \phi_{cal}, & \phi_{cal} < \phi_{\min} \end{cases} \quad (6-1)$$

在早期范围的编码过程中，不同重放仰角方位的条件下，不对方位角进行修改。

实验 1 设置 7 个具有不同反射声仰角范围的刺激，研究反射仰角对空间感的影响，具体包括：（1）早期反射位于-15°至 90°仰角范围，使用 8 阶 3-D Ambisonics 编码，后期反射使用 3-D FOA 编码；（2）早期反射仰角位于-15°至 60°范围，使用 8 阶 3-D HOA 编码，后期反射使用 3-D FOA 编码；（3）早期反射仰角位于-15°至 30°范围，使用 8 阶 3-D HOA 编码，后期反射使用 3-D FOA 编码；（4）早期反射仰角压缩至水平面，使用 8 阶 2-D HOA 编码，后期反射使用 3-D FOA 编码；（5）早期反射位于-15°至 90°仰角范围，使用 8 阶 3-D Ambisonics 编码，后期反射使用 2-D FOA 编码（方法见 2.2.5 节）；（6）早期反射仰角压缩至水平面，使用 8 阶 2-D HOA 编码，后期反射使用 2-D FOA 编码。（7）早期反射位于-15°至 90°仰角范围，使用 8 阶 3-D Ambisonics 编码，后期反射使用 3-D FOA 编码，随后将所有通道的信号直接相加，下混合为单通道信号，由正前方扬声器播放。

表 6-2 实验中所有刺激的反射编码方法

Tab. 6-2 Reflections encoding methods of all stimuli in the experiment

序号	早期反射仰角范围	直达声/早期反射编码	后期反射编码
1	[-15°, 90°]	8 阶 3-D HOA	3-D FOA
2	[-15°, 60°]	8 阶 3-D HOA	3-D FOA
3	[-15°, 30°]	8 阶 3-D HOA	3-D FOA
4	[-0°, 0°]	8 阶 2-D HOA	3-D FOA
5	[-15°, 90°]	8 阶 3-D HOA	2-D FOA
6	[-0°, 0°]	8 阶 2-D HOA	2-D HOA
7	[-15°, 90°]	8 阶 3-D HOA 单通道下混	3-D FOA
8	[-15°, 90°]	最近邻重放	3-D FOA
9	[-15°, 90°]	3 阶 3-D HOA	3-D FOA
10	[-15°, 90°]	3-D FOA	3-D FOA
11	[-15°, 90°]	分层 Ambisonics	3-D FOA

实验 2 设置 5 个不同的刺激，研究反射空间分辨率对空间感的影响：（1）早期反射仰角位于-15°至 90°范围，直达声与早期反射采用最近邻编码（反射声映射到最接近的扬声器上直接播放）；（2）早期反射仰角位于-15°至 90°范围，采用 3-D FOA 编码，后期反射采用 3-D FOA 编码；（3）早期反射仰角位于-15°至 90°范围，采用 3 阶 3-D

HOA 编码，后期反射采用 3-D FOA 编码；（4）早期反射仰角位于-15°至 90°范围，采用 8 阶 3-D HOA 编码，后期反射采用 3-D FOA 编码；（5）早期反射仰角位于-15°至 90°范围，直达声与早期反射采用分层 Ambisonics 重放。

实验中的所有刺激设置如表 6-2 所示，所有信号的重放强度为 68 dBA，使用放置在阵列中心的自由场传声器（Brüel & Kjær Type 4189）进行校准。

进一步对各个条件下可听化重放声场的房间声学指标进行测量，使用布置在阵列中心的自由场传声器（Brüel & Kjær Type 4189）测量重放环境的脉冲响应，测试方法见 2.3.2.2 节，使用扫频信号作为干信号，经过本节中不同条件下的虚拟室内反射编码方式进行编码后重放，传声器记录下经过可听化重放后的扫频信号，解卷积得到虚拟声场的脉冲响应。

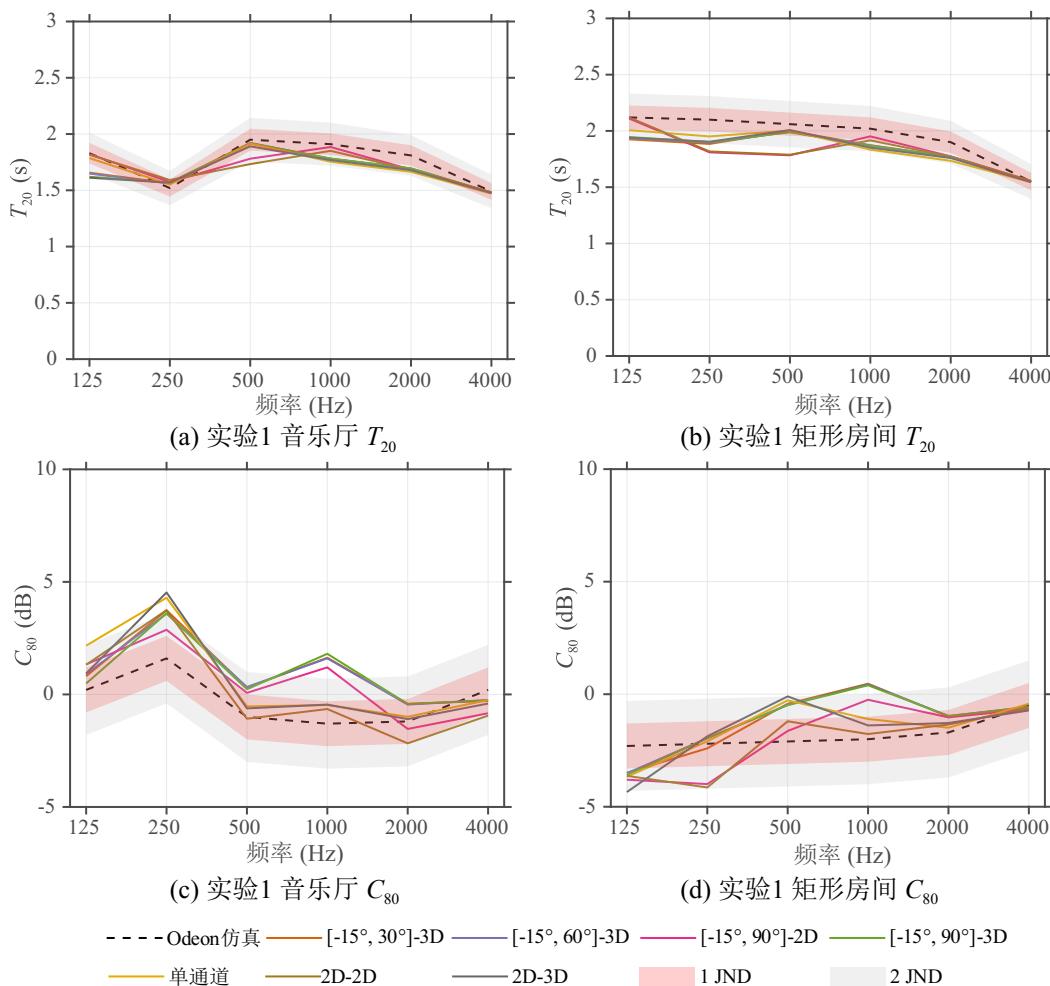


图 6-3 实验 1 中所有条件重放虚拟声场的 T_{20} 与 C_{80}

Fig. 6-3 The T_{20} and C_{80} values of virtual sound fields reproduced by all conditions in the experiment 1

使用 ITA Toolbox 分别计算表征混响时间的 T_{20} 与表征清晰度的 C_{80} ^[158]。 T_{20} 定义为声音能量从初始值衰减 5 dB 与 25 dB 处，拟合曲线外推至 60 dB 衰减所需的时间， C_{80} 脉冲响应 80 ms 前与 80 ms 后全部能量的比值，定义如下所示^[33]：

$$C_{80} = 10 \log_{10} \frac{\int_0^{80 \text{ ms}} p^2(t) dt}{\int_{80 \text{ ms}}^{\infty} p^2(t) dt} \quad (6-2)$$

其中 $p(t)$ 为脉冲响应数值。

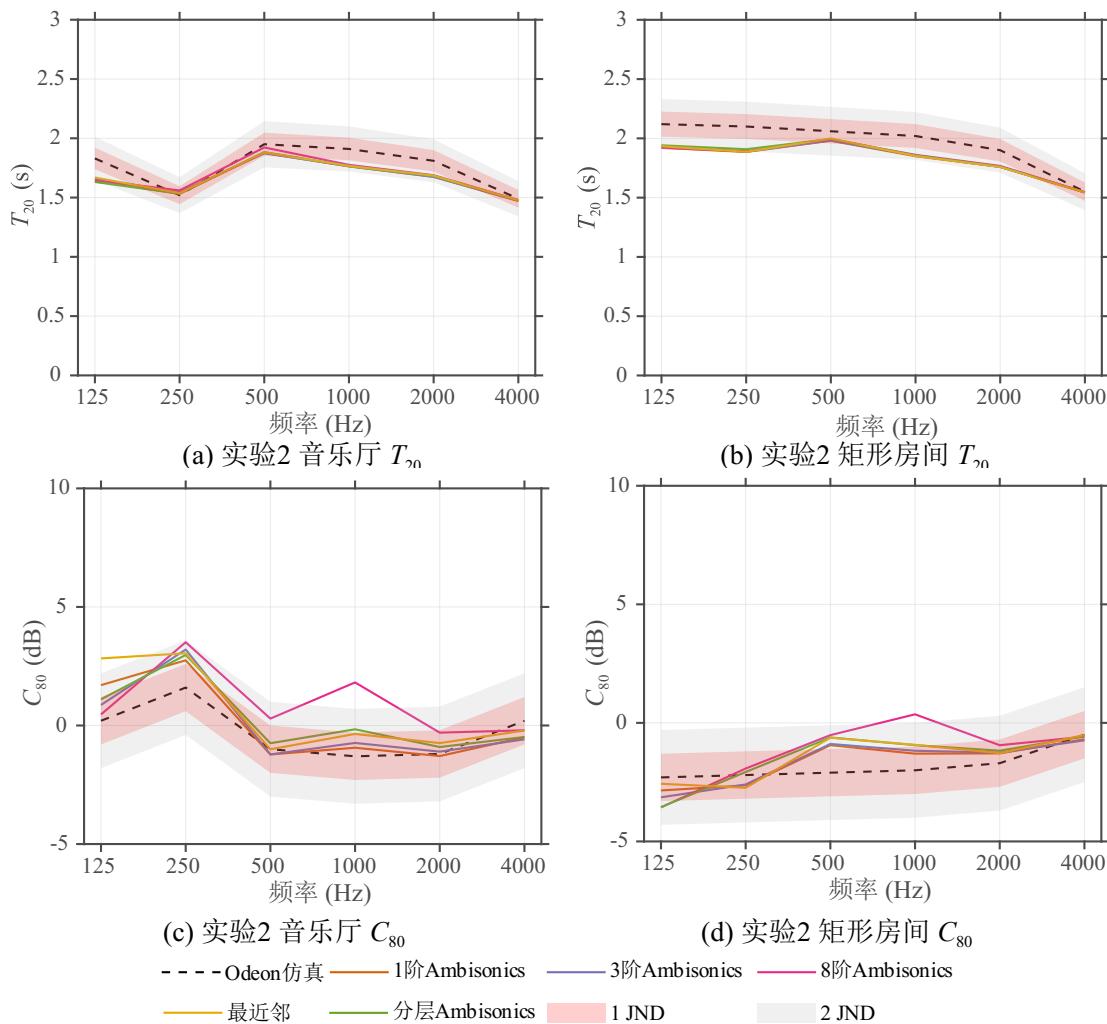


图 6-4 实验 2 中所有条件重放虚拟声场的 T_{20} 与 C_{80}

Fig. 6-4 The T_{20} and C_{80} values of virtual sound fields reproduced by all conditions in the experiment 2

实验 1 中所有虚拟重放条件的声场测量结果如图 6-3 所示，图中还绘制了 Odeon 中计算得到的 T_{20} 与 C_{80} 以及二者对应的一个 JND 与两倍 JND 的范围，作为重放参考范围，其中 T_{20} 的 JND 为当前 T_{20} 的 5%， C_{80} 的 JND 为 1 dB。结果表明，重放声场与 Odeon 仿真的声场客观指标随频率变化趋势基本一致。在 2000 Hz 及以上频率范围内，所有条件的重放声场客观指标与 Odeon 仿真的结果基本一致，误差在一个 JND 以内。在中低频率范围，客观指标与 Odeon 结果的差距范围基本在两个 JND 内。在 1000 Hz，不同条件的重构声场与 Odeon 仿真结果差距相对较大，部分条件误差超过两个 JND。上述误差可能来源于，本文对反射声的编码方式、倍频程滤波方式与 Odeon 不一致，且声场指标计算与工具与 Odeon 存在差异。对比不同重放条件， T_{20} 与 C_{80} 的波动范围基本位于 1 JND 内，但音乐厅虚拟场景 1000 Hz 与矩形房间 250 Hz 的 C_{80} 存在相对较大波动，超过 1 个 JND。

类似地，实验 2 中虚拟重放条件的声场测量结果如图 6-4 所示。测试结果与实验 1 信号的测试结果类似，变化趋势与 Odeon 仿真结果一致，在不低于 2000 Hz 的高频段与仿真结果更接近。相比实验 1，实验 2 中不同条件对 T_{20} 与 C_{80} 影响更小，除 8 阶 Ambisonics 重放在 1000 Hz 处的结果外，其余条件的波动范围均低于一个 JND。

6.3.3 实验 1—反射声重放仰角范围对空间感影响

实验 1 采用与 5.3.4 节中类似的 MUSHRA 测试范式，受试者将不同信号与参考信号进行对比并对感知的 LEV 进行打分。实验包含 2 个虚拟房间条件，每个条件重复 3 次，受试者总计需要进行 6 个试次的实验（2 房间 \times 3 重复），实验试次的顺序随机生成，耗时约为 1.5 至 2 小时。受试者每进行 20 分钟的实验，或受试者主动反馈疲劳后，需休息一次，避免疲劳影响对实验结果造成影响。

实验流程如图 6-5 所示，实验前，受试者首先经过训练环节，该环节包括两个部分：LEV 感知训练与 MUSHRA 训练。在 LEV 感知训练环节，实验员首先口头向受试者解释 LEV 的概念，然后，受试者坐在阵列中心，面向正前方，分别聆听具有不同包围感的信号。这些信号包括：使用正前方单一扬声器播放的干信号、使用 8 阶 2-D Ambisonics 重放的在水平面上均匀分布的 4 个干信号、使用 8 阶 3-D Ambisonics 重放的空间中均匀分布的 4 个干信号、使用 8 阶 3-D Ambisonics 重放的空间中均匀分布的 16 个干信号以及使用 8 阶 3-D Ambisonics 重放的空间中均匀分布的 32 个干信号，上述信号空间包围感依次增强。受试者对上述信号进行有反馈的两两对比，直到能正确区分

所有的信号组合。随后进行 MUSHRA 训练，旨在让受试者熟悉实验信号与 MUSHRA 评分软件操作。该训练要求受试者独立完成一次 MUSHRA 测试，MUSHRA 测试中的刺激与正式实验相同，测试虚拟房间条件随机，该训练的结果不纳入最终的实验结果。



图 6-5 实验流程图、MUSHRA 评分操作界面以及实验照片

Fig. 6-5 Diagram of the experimental produces, MUSHRA GUI, and the photo of the experiment

在正式实验的每次 MUSHRA 测试，受试者需要将 7 个信号（包括 5 个测试信号、1 个隐藏的参考信号和 1 个隐藏的锚点信号）与参考信号进行对比，按照感知 LEV 差异

的大小进行评分。测试信号为分别使用反射编码方式 2~6（见表 6-2）进行可听化重放的刺激，参考信号为使用反射编码方式 1 进行可听化重放的刺激，锚点刺激为使用反射编码方式 7 进行可听化重放的刺激，其空间感与参考信号差距较大。评分范围为 0 至 100 分，0 分表示测试信号的 LEV 与参考信号相同，100 分表示二者 LEV 差距较大，受试者通过操作平板电脑上显示的 GUI 上完成所有评分（见图 6-5（b））。7 个刺激信号的排序随机产生，受试者可通过按钮自由播放或切换刺激信号，或点击底部进度条播放特定时间段的信号。受试者需在选择并聆听刺激信号后才能对其进行评分，整个过程中需保持面向前方，不能转动或移动头部。

6.3.4 实验 2—早期反射重放方式对空间感影响的评估实验

实验 2 的流程如图 6-5 所示，流程与实验 1 相似，但实验 2 中，受试者需要对包围感与声源宽度两个感知维度进行评分。实验包含 2 个虚拟房间条件，每个条件重复 3 次，受试者需要进行 12 个试次的实验（2 房间 \times 3 重复 \times 2 感知维度）。继实验 1 之后，首先进行 LEV 的评分实验，包含 6 个试次，试次的顺序随机生成，随后进行声源宽度感知训练与评分实验，两个感知维度的感知实验信号相同，仅评分目标不同。整个实验耗时约为 1.5 至 2 小时。受试者每进行 20 分钟的实验，或受试者主动反馈疲劳后，需休息一次，避免疲劳影响对实验结果造成影响。

在声源宽度评分实验前，受试者需要经过声源宽度感知训练环节。在该训练中，实验员首先口头向受试者解释声源宽度的概念，然后，受试者坐在阵列中心，面向正前方，分别聆听具有不同声源宽度的信号。这些信号包括：使用正前方单一扬声器播放的虚拟声源、在水平面方位角-30°至 30°范围内按 10°等间隔分布的虚拟声源组、在水平面方位角-60°至 60°范围内按 10°等间隔分布的虚拟声源组、在水平面方位角-90°至 90°范围内按 10°等间隔分布的虚拟声源组。上述虚拟声源均为 8 阶 3-D Ambisonics 重放的干信号，这些信号声源宽度依次增强。受试者对上述信号进行有反馈的两两对比，直到能正确区分所有信号组合。

在实验 2 的 MUSHRA 测试中，受试者需要将 5 个信号（包括 4 个测试信号和 1 个隐藏的参考信号）与参考信号进行对比，按照感知声源宽度或 LEV 的差异的大小进行评分。需要注意的是，实验 2 的 MUSHRA 测试没有锚点信号。测试信号为分别使用反射编码方式 1、9、10 和 11（见表 6-2）进行可听化重放的刺激，参考信号为使用反射编码方式 8 进行可听化重放的信号，具有最小的能量扩散。评分范围为 0 至 100 分，0 分表示测试信号的 ASW 或 LEV 与参考信号相同，100 分表示二者差距大，受试者通过

操作平板电脑上显示的 GUI 完成所有评分。5 个刺激信号的排序随机产生，受试者可通过按钮自由播放或切换刺激信号，或点击底部进度条播放特定时间段的信号。受试者需在选择并聆听刺激信号后才能对其进行评分，整个过程中需保持面向前方，不能转动或移动头部。

6.4 实验结果与讨论

6.4.1 实验 1 结果

实验 1 共获得 672 个评分数据（7 刺激 \times 2 场景 \times 3 重复 \times 16 名受试者），每个条件（虚拟房间-刺激类型）有 48 个数据样本。不同虚拟房间下 7 个刺激类型的评分结果如图 6-6 所示，每个箱线图的箱子中线、底部与顶部边缘显示以下信息：中位数、下四分位数和上四分位数，箱子上下两端的须线为排除离群值的最大值与最小值，小圆点为离群值。从图中可以看到不同重放方法的 LEV 评分差距较大，音乐厅和矩形房间两个条件下评分趋势一致。

进一步对数据使用双因素 rm ANOVA 进行方差分析，检验因子为虚拟房间与刺激类型，并使用 Geisser-Greenhouse 对 rm ANOVA 进行球度校正。本实验中锚点信号为单通道重放，作为包围感最差的参考，不具有对比价值，因此不纳入统计分析。统计结果显示，刺激类型对 LEV 评分存在显著差异 [$F(4.288, 201.5) = 65.35, p < 0.0001$]，该因子对结果的解释度为 36.87%。不同虚拟房间对 LEV 评分没有显著影响 [$F(1.000, 47.00) = 1.164, p = 0.2861$]，且二者不存在交互作用。

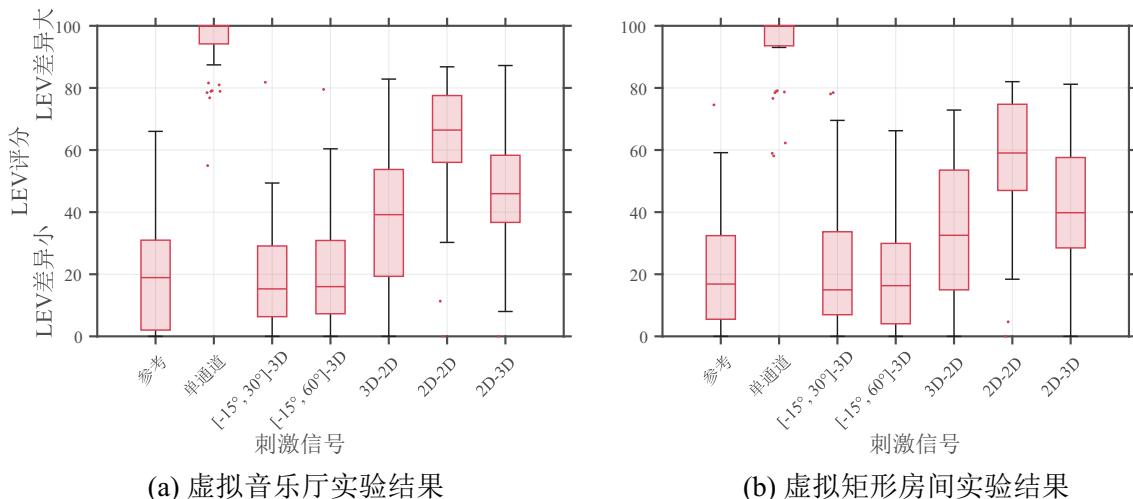


图 6-6 实验 1 实验结果箱型图

Fig. 6-6 The box plots of the experimental results of the experiment 1

进一步对结果进行 Tukey 校正的事后多重比较，由于虚拟房间没有显著效应，且与刺激类型没有交互作用，直接对主效应进行两两对比。首先关注不同重放仰角范围对 LEV 的影响，对比参考信号（参考信号为[-15°, 90°]-3D）、[-15°, 60°]重放仰角范围与[-15°, 30°]重放仰角范围三组信号条件，结果显示它们之间均没有显著差距，这说明限制重放仰角范围对空间感的影响较为有限。

然而，当早期反射被压缩至水平面时（即 2D-3D 条件），LEV 相比参考信号、[-15°, 60°]信号与[-15°, 30°]信号显著变差（ p 值均小于 0.0001）。类似地，当后期反射采用 2D Ambisonics 重放时（3D-2D 条件），LEV 相比参考条件也显著变差（ $p < 0.0001$ ）。从结果上看，无论早期反射或后期反射，当它们采用二维重放时，对 LEV 的影响程度近似，2D-3D 与 3D-2D 没有显著差异（ $p = 0.908$ ）。当二者均采用二维重放时，LEV 最差，显著低于其他所有条件（ p 均小于 0.0001）。

6.4.2 实验 2 结果

实验 2 要求受试者对 LEV 和声源宽度两个感知维度进行打分，每个维度共获得 480 个评分数据（5 刺激 \times 2 场景 \times 3 重复 \times 16 名受试者），每个条件（虚拟房间-刺激类型）有 48 个数据样本。不同虚拟房间在 5 个刺激类型的评分结果如图 6-7 所示，箱型图定义与图 6-6 的定义相同。从图中可以看到不同重放方法的 LEV 与声源宽度评分在两个虚拟厅堂环境表现不同，在音乐厅虚拟声场下，不同早期反射的编码方式会影响声源宽度和 LEV 的感知，在矩形房间下，尽管不同刺激也有差距，但差距相对较小。

使用双因素 rm ANOVA 探讨了不同重放方法与房间类型对空间感感知（LEV 与声源宽度）的影响，结果表明，对于 LEV，刺激类型 [$F(4, 188) = 25.71, p < 0.0001$]、房间类型 [$F(1, 47) = 28.64, p < 0.001$] 及其交互作用 [$F(4, 188) = 13.28, p < 0.0001$] 均对感知结果具有显著影响。其中，刺激类型对 LEV 评分具有更大的影响，解释了 18.25% 的总效应，房间仅解释了 2.21% 的总效应。

进一步对 LEV 的评分进行 Tukey 校正的事后多重比较，由于房间类型和刺激类型存在交互作用，在两个房间下对不同刺激分别进行比较。在虚拟音乐厅下，相比参考信号（最近邻编码），1 阶、3 阶、8 阶 3-D Ambisonics 的 LEV 评分均显著高于参考信号， p 值均小于 0.001。但参考刺激和分层 Ambisonics 重放没有显著的 LEV 区别（ $p = 0.990$ ）。进一步考察重放阶数对 LEV 的影响，随着阶数的上升，LEV 更接近参考信号。1 阶 Ambisonics 的 LEV 差异评分显著高于 8 阶 Ambisonics（ $p < 0.0001$ ）。3 阶 Ambisonics 评分显著高于 8 阶 Ambisonics（ $p < 0.001$ ）。

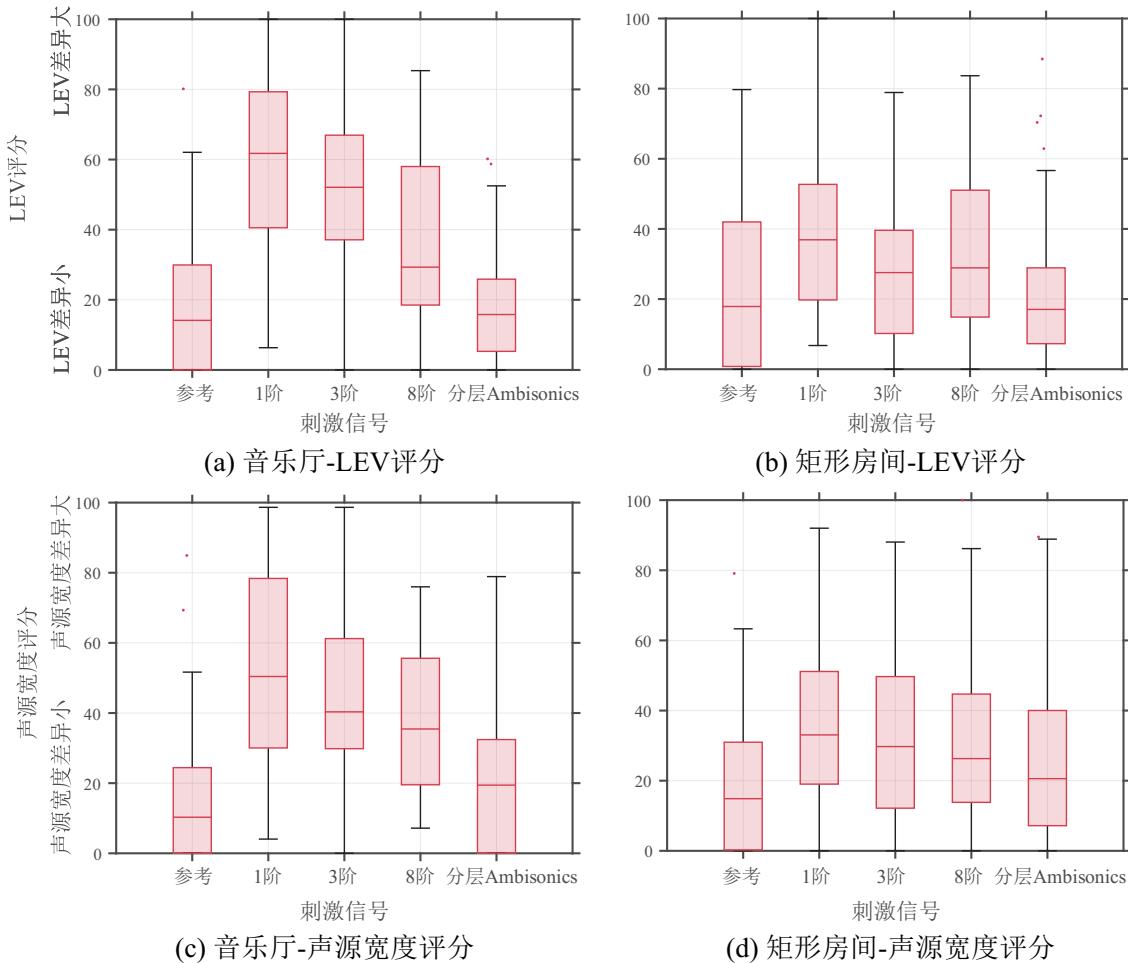


图 6-7 实验 2 实验结果箱型图

Fig. 6-7 The box plots of the experimental results of the experiment 2

在矩形房间条件下，LEV 的差距变化较小，两两对比仅发现 1 阶信号和参考信号与分层 Ambisonics 之间有显著区别 (p 值分别小于 0.0001 和 0.01)。1 阶与 3 阶之间 ($p < 0.05$)，8 阶和参考之间也存在显著差异 ($p < 0.001$)。

声源宽度感知结果与 LEV 类似，刺激类型 $[F(4, 188) = 22.83, p < 0.0001]$ 、房间类型 $[F(1, 47) = 15.27, p < 0.001]$ 及其交互作用 $[F(4, 188) = 4.197, p < 0.05]$ 均对感知结果具有显著影响。其中，刺激类型对声源宽度评分具有更大的影响，解释了 15.36% 的总效应，房间类型仅解释了 1.23% 的总效应。

在虚拟音乐厅下，相比参考信号（最近邻编码），1 阶、3 阶、8 阶 3-D Ambisonics 的声源宽度评分均显著高于参考信号， p 值均小于 0.0001。然而参考刺激和分层 Ambisonics 重放没有显著的声源宽度区别 ($p = 0.816$)。进一步考察重放阶数对声源宽

度的影响，随着阶数的上升，声源宽度更接近参考信号。1 阶 Ambisonics 的声源宽度差异评分显著高于 8 阶 Ambisonics ($p < 0.05$)。尽管 3 阶 Ambisonics 的均值要高于 8 阶 Ambisonics (差值 5.4)，但二者差距不显著 ($p = 0.741$)。

在矩形房间中，声源宽度评分的差距变化较小，两两对比仅发现 1 阶和 8 阶重放与分层 Ambisonics 之间有显著区别 (p 值分别小于 0.001 和 0.05)。

6.4.3 结果讨论

实验 1 结果表明，早期反射声的仰角范围（如压缩至 $[-15^\circ, 60^\circ]$ 或 $[-15^\circ, 30^\circ]$ ）对听众的 LEV 包围感没有显著影响，但当早期反射被压缩至水平面并使用二维重放时，LEV 评分显著下降。这一现象说明：在真实厅堂中，来自天花板、侧墙等垂直方向的反射声对 LEV 有影响，但听觉系统对反射的具体高度并不敏感，对 LEV 感知影响不大。此外，后期反射采用二维重放（3D-2D 条件）时，LEV 评分同样下降，说明后期反射的高度信息对空间感的贡献与早期反射信息相当。这与对 LEV 的通常认识不同，ISO 3382-1 建议的指标认为 LEV 主要与后期反射相关（见 1.2.3 节），但本研究的实验说明了早期反射的高度信息同样会影响 LEV。上述结论几乎不受房间的反射模式影响。

实验 2 发现，随着 Ambisonics 重放阶数从 1 阶提升至 8 阶，声源宽度和 LEV 评分逐渐接近参考信号，且本文提出的分层 Ambisonics 算法在两个厅堂的声源宽度与 LEV 评分均最接近参考信号。这一实验结果表明，早期反射的能量扩散会对声源宽度和 LEV 造成类似的可感知影响，因此重放时应尽量降低虚拟声源的能量扩散。值得注意的是，在矩形房间（简单声学环境）中，不同重放方式的空间感评分差异小于音乐厅场景。这可能是由于简单环境中反射声数量少、方向分布集中，低阶系统即可满足感知需求，而复杂厅堂具有更均匀的反射，能量扩散对其影响更大。

综上，上述实验证实了反射高度信息的必要性与冗余性：受试者难以辨别不同的反射仰角重放范围，但压缩至水平面会导致感知显著退化，这为实际系统设计提供了折中方案，以在保证 LEV 的同时降低硬件成本。另一方面，降低能量扩散对保证可靠的 LEV 与声源宽度感知非常重要，应尽量提高 Ambisonics 重放阶数。实验证实了本文提出的分层 Ambisonics 是一种有效降低能量扩散的重放方法。

需要注意的是，本文的实验结果仅基于两个虚拟房间模型，没有考虑反射声到达时间、时间间隔与重放仰角范围及能量扩散对空间感感知的交互作用。进一步考察房间反射序列对空间感的影响，建立完整的空间感感知模型，是本研究未来的方向。

6.5本章小结

本章围绕反射声仰角范围及 Ambisonics 重放方式对虚拟厅堂空间感感知的影响展开研究。通过设计两组对比实验，系统探讨了不同重放参数设置下 LEV 与声源宽度主观评价指标的变化规律，实验一以探究反射声仰角范围对 LEV 的影响为核心，采用单因素实验设计，将早期反射声的重放仰角范围压缩到不同的区间，包括二维重放（仅水平面）等多个条件。同时，后期反射声也分别设置二维与三维的重放条件，形成多组对比场景。实验二则着重研究早期反射能量扩散对空间感的影响，采用不同阶数 Ambisonics 与不同重放算法对反射进行编码。实验分别在矩形房间（简单声学环境）与音乐厅（复杂声学环境）两种虚拟场景下进行，全面评估重放方式在不同场景下的有效性与感知区别。

研究结果表明，反射声高度信息的保留是保证空间包围感的关键因素，当早期反射被强制压缩至水平面（二维重放）时，LEV 评分显著下降；但听觉对反射的具体仰角方向不敏感，不同仰角范围的重放对 LEV 的影响较小。此外，早期反射与后期反射的高度信息均对 LEV 影响。这一发现强调了重放时保留反射高度信息的重要性，同时为可听化系统硬件简化提供了理论依据。另一方面，早期反射的能量扩散会显著影响声源宽度和 LEV 感知，本文提出的分层 Ambisonics 在可听化环境下取得了与参考信号非常接近的感知评分。此外，研究还揭示了声环境反射复杂度与重放能量扩散的耦合性，在简单声学环境中，低阶系统即可满足感知需求，而复杂厅堂场景则更依赖高阶系统或优化算法。

本章研究成果不仅深化了对空间听觉感知机制的理解，也为建筑声学领域中 Ambisonics 系统的参数优化与工程部署提供了重要参考，同时为虚拟现实场景下的多模态可听化设计指明了方向。

结论与展望

研究结论

在建筑声学和环境声学中，可听化技术用于模拟和重放建筑空间或环境的声场效果，是建筑声学设计与研究中的核心技术之一。本文将一种基于精确物理声场重建的高阶 Ambisonics 声重放技术应用于可听化重放中，建立了目前国内最大型的高阶 Ambisonics 可听化系统，实现了高精度的复杂虚拟声场重放。

本文围绕高阶 Ambisonics 可听化系统重放的改进与感知评估问题，开展了一系列主观听音实验对重放的听觉场景的效果进行了感知评估研究，这些研究的核心目的在于，在保证感知可靠性的基础上，尽可能提高重放精度，降低重放系统的实施难度，推动高阶 Ambisonics 可听化系统的实际落地使用。本文的主要研究结论如下：

(1) 首先，基于声场球谐函数分解理论，搭建了一个大型多通道扬声器阵列的高阶 Ambisonics 系统，系统包含硬件与软件两部分。软件部分实现了完整的 Ambisonics 可听化重放信号处理链路，包括虚拟声源的空间声编解码、空间声场信号的拾拾信号处理算法、反射声空间编码、虚拟声场景合成、多通道重放等部分。硬件部分包含多通道的信号控制与转换，并测量了系统所有通道的脉冲响应，通过 FIR 滤波器对所有通道实现了幅度、相位及延时的全局均衡。软硬件的配合确保了声场重构的准确性。

(2) 探讨 Ambisonics 虚拟重放阶数对声场重构误差与主观感知的影响。首先针对阶数对 Ambisonics 重放误差的影响进行了客观分析，然后，设计并开展了实验，针对言语接受阈值这一主观感知指标，研究不同 Ambisonics 重放阶数下语音信号虚拟声源的感知情况。实验发现，除 0 阶外，SRT 随阶数提升逐渐降低，并接近单声道重放参考水平。低阶系统（1 阶、3 阶）因多扬声器串扰导致言语可懂度下降，而 9 阶系统已接近真实声源水平。实验发现 Ambisonics 重放下，言语接受阈值同时受多通道串扰导致的质量下降与能量扩散导致的 SRM 影响，且串扰导致的频谱失真对清晰度的负面影响更显著。这一实验结果为使用 Ambisonics 系统进行语音清晰度的研究提供了阶数选择及系统规模要求的参考依据，强调了高阶重放的重要性。

(3) 为了尽可能地提高重放阶数，降低 Ambisonics 重放对扬声器硬件的要求，本文提出了一种分层 Ambisonics 重放算法，该算法结合二维 Ambisonics 与 VBAP，通过两层扬声器合成不同仰角的高阶虚拟声源，再经矢量调幅实现仰角调节。算法降低了扬声器数量及均匀性要求，同时提高了水平面等实际扬声器区域的重放精度，通过

HRTF 与平面波声场模拟，在客观指标层面验证了该算法对提高重放精度的有效性。客观模拟发现，在具有实际物理扬声器的平面上，重放声源的准确性和有效重放区域均优于 3-D Ambisonics 系统。但在不具有实际物理扬声器平面的情况下，分层 Ambisonics 算法由于采用了 VBAP，其合成的 HRTF 在低频部分相比目标 HRTF 误差较大，

(4) 进一步开展主观实验对分层 Ambisonics 算法进行验证，实验对比了该算法与相关的算法，包括几种混合阶 Ambisonics 算法与 AllRAD 算法的定位性能与声音保真度差异。定位与失真评估实验表明，本文所提出的新算法在中心及非中心听音时，水平面和 30°仰角平面的方位角定位误差显著低于混合阶 Ambisonics 及 AllRAD 算法，声音保真度更优，在合成高仰角虚拟声源时表现稳定。本研究还填补了当前对几种混合阶 Ambisonics 以及 AllRAD 算法实际性能的基础性验证数据上的缺失，为实际系统的算法选取与设计提供了重要参考。

(5) 在前述研究的基础上，进一步探讨复杂室内声环境可听化的简化可能。由于三维重放需要消耗大量硬件资源，而听觉对高度信息相对不敏感，本文开展实验研究可听化重放时反射声仰角的重放范围以及能量扩散对空间感感知的影响。实验发现早期反射声与后期反射声的高度信息均对 LEV 感知有显著影响，二维重放的 LEV 与参考信 LEV 存在较大差异，但听觉对反射具体仰角不敏感，不同的仰角重放范围对 LEV 的影响较小。此外，随着重放阶数的提高，反射声能量扩散逐渐减小，空间感（声源宽度、LEV）评分都逐渐接近参考信号的评分。此外，实验还发现，相比传统的三维 Ambisonics 算法，本文所提出分层 Ambisonics 算法在本文建立的非均匀扬声器阵列下取得了更好的听音结果，空间感感知与参考条件下的感知结果更接近。实验还发现在复杂厅堂场景中，反射能量扩散对空间感影响更大，简单声学环境的重放或可使用更低阶的重放系统。这些结论不仅为理解听觉在复杂室内声场的感知提供了新视野，还为可听化系统的简化提供了依据。

综上，本研究通过理论创新与实验验证，为高阶 Ambisonics 在建筑声学设计及虚拟现实中的应用奠定了基础。

创新点

综上所述，本文的创新点可归纳为如下三点：

(1) 提出一种针对分层扬声器阵列的分层 Ambisonics 算法并开展主观听音评价实验验证。该算法在保持高精度水平面定位基础上，仍具有较好的垂直方向定位效果，

且声音失真较低。该算法大大降低了三维 Ambisonics 重放对扬声器数量与空间分布均匀性的要求。

(2) 首次全面对比并验证了非均匀扬声器阵列布置下，分层 Ambisonics 算法相比现有的几类 Ambisonics 重放算法在定位性能与声音保真度上的差异。结果表明，分层 Ambisonics 算法相较混合阶 Ambisonics 以及 AllRAD 算法在定位性能与声音保真度上均有较大提升。该项工作补充了该领域研究的数据缺失，为后续相关研究的算法选择与优化设计提供依据。

(3) 首次在高阶 Ambisonics 重放的虚拟室内声环境下，探讨了反射声高度方向信息与能量扩散对空间感的影响。开展主观听音评价实验研究了重放反射声的仰角范围与重放阶数对空间感感知的影响，为系统设计中高仰角通道的简化提供理论依据。

展望

基于以上的研究内容与结论，未来的研究将聚焦于：

(1) 在重放系统方面，目前系统对室内声环境的计算依赖于商业软件，导致后期反射仅能实现一阶重放，未来将开发独立的室内声场几何声学仿真模块，实现更高效率与更高精度的重放。目前，系统尚未建立同步的视觉重放链路，视觉重放与听觉重放是分离的，未来将引入视觉重放装置与软件，实现完整的虚拟视听场景重放。

(2) 在建筑声学应用上，基于已有的高阶 Ambisonics 重放软件，开发易于操作的声场合成工具，将高阶 Ambisonics 系统与建筑设计软件深度整合，开发实时可听化插件，实现室内声学及声景设计的快速可听化场景构建。

(3) 在室内声场的听觉感知问题方面，目前的研究仅探讨了针对空间感的感知评估，可听化重放的目标是使得声场整体印象与真实空间一致，未来将进一步开展实验直接对比虚拟声场与真实声场的整体听觉感知差异。

(4) 对于空间感感知问题，目前的研究仅探讨了反射声仰角范围与分辨率对空间感的影响，未来将进一步扩展主客观实验，纳入反射声序列的完整信息，包括反射延时与强度等信息，综合分析不同条件下的室内空间感感知情况，并结合客观参数构建完整的空间感预测模型。

(5) 目前的所有实验均是基于静态声源基础，进一步的研究将探讨听觉在混响环境下对动态声源的感知精度与分辨率，为可听化系统的动态重放设计与实现提供基础性数据。

参考文献

- [1] Vorländer M. Auralization[M]. Berlin: Springer, 2020.
- [2] 吴硕贤, 赵越皓. 室内声学与环境声学[M]. 广东科技出版社, 2003.
- [3] 孙海涛. 建筑声学缩尺模型测量关键技术研究[D]. 广州: 华南理工大学, 2014.
- [4] Thery D., Boccara V., Katz B.F.G. Auralization uses in acoustical design: A survey study of acoustical consultants[J]. *J Acoust Soc Am*, 2019, 145(6): 3446-3456.
- [5] Wenzel E.M., Arruda M., Kistler D.J., et al. Localization using nonindividualized head-related transfer functions[J]. *J Acoust Soc Am*, 1993, 94(1): 111-123.
- [6] Xie B. Head-related transfer function and virtual auditory display[M]. Plantation: J. Ross Publishing, 2013.
- [7] Mills A.W. On the Minimum Audible Angle[J]. *The Journal of the Acoustical Society of America*, 1958, 30(4): 237-246.
- [8] Middlebrooks J.C. Sound localization[M]. *The Human Auditory System*: Vol. 129. 2015: 99-116.
- [9] Van Schoonhoven J., Rhebergen K.S., Dreschler W.A. Towards measuring the Speech Transmission Index in fluctuating noise: Accuracy and limitations[J]. *The Journal of the Acoustical Society of America*, 2017, 141(2): 818-827.
- [10] Blauert J., Braasch J. *The Technology of Binaural Understanding*[M]. Cham: Springer, 2020.
- [11] Schnupp J., Nelken I., King A. *Auditory neuroscience: Making sense of sound*[M]. MIT press, 2011.
- [12] Blausen.com staff. Medical gallery of Blausen Medical 2014[J]. *WikiJournal of Medicine*, 2014, 1(2).
- [13] Moore B.C. *An Introduction to the Psychology of Hearing*[M]. Brill, 2012.
- [14] 谢菠荪. 动态听觉定位因素及其在空间声重放的作用[J]. *声学学报*, 2024, 49(6): 1131-1151.
- [15] Iida K., Ishii Y. Effects of adding a spectral peak generated by the second pinna resonance to a parametric model of head-related transfer functions on upper median plane sound localization[J]. *Applied Acoustics*, 2018, 129: 239-247.
- [16] Iida K., Itoh M., Itagaki A., et al. Median plane localization using a parametric model of the head-related transfer function based on spectral cues[J]. *Applied Acoustics*, 2007, 68(8): 835-850.
- [17] Xie B., Yu G. Psychoacoustic Principle, Methods, and Problems with Perceived Distance Control in Spatial Audio[J]. *Applied Sciences*, 2021, 11(3): 11242.
- [18] Guo Z., Zhao Y., Wang L., et al. Distance discrimination thresholds of proximal sound sources in a real anechoic environment[J]. *Applied Acoustics*, 2023, 203: 109223.
- [19] Kolarik A.J., Moore B.C., Zahorik P., et al. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss[J]. *Atten Percept Psychophys*, 2016, 78(2): 373-395.
- [20] Kolarik A., Cirstea S., Pardhan S. Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues[J]. *J Acoust Soc Am*, 2013, 134(5): 3395-3398.
- [21] Coleman P.D. Dual role of frequency spectrum in determination of auditory distance[J]. *J Acoust Soc Am*, 1968, 44(2): 631-632.

- [22] Brungart D.S., Rabinowitz W.M. Auditory localization of nearby sources. Head-related transfer functions[J]. *J Acoust Soc Am*, 1999, 106(3): 1465-1479.
- [23] Yu G., Wang L. Effect of Individualized Head-Related Transfer Functions on Distance Perception in Virtual Reproduction for a Nearby Sound Source[J]. *Archives of Acoustics*, 2019: 251-258.
- [24] Arend J.M., Liesefeld H.R., Pörschmann C. On the influence of non-individual binaural cues and the impact of level normalization on auditory distance estimation of nearby sound sources[J]. *Acta Acustica*, 2021, 5: 10.
- [25] Rosenblum L.D., Carello C., Pastore R.E. Relative effectiveness of three stimulus variables for locating a moving sound source[J]. *Perception*, 1987, 16(2): 175-186.
- [26] Ashmead D.H., Davis D.L., Northington A. Contribution of listeners' approaching motion to auditory distance perception[J]. *Journal of Experimental Psychology: Human Perception and Performance*, 1995, 21(2): 239-256.
- [27] Genzel D., Schutte M., Brimijoin W.O., et al. Psychophysical evidence for auditory motion parallax[J]. *Proceedings of the National Academy of Sciences*, 2018, 115(16): 4264-4269.
- [28] Seeber B.U., Clapp S. Auditory Room Learning and Adaptation to Sound Reflections[M]. *The Technology of Binaural Understanding*. Berlin: Springer, 2020.
- [29] Beranek L.L. Concert halls and opera houses: music, acoustics, and architecture[M]. Springer, 2004.
- [30] Ando Y., Raichel D.R. Architectural acoustics: blending sound sources, sound fields, and listeners[M]. *Acoustical Society of America*, 1998.
- [31] Blauert J., Lindemann W. Auditory spaciousness: Some further psychoacoustic analyses[J]. *The Journal of the Acoustical Society of America*, 1986, 80(2): 533-542.
- [32] Soulodre G. Can Reproduced Sound be Evaluated using Measures Designed for Concert Halls[A]. *Proceedings of the Workshop on Spatial Audio and Sensory Evaluation Techniques*[C]. Guildford, UK, 2006.
- [33] ISO 3382-1:2009, Acoustics—Measurement of the reverberation time—Part 1: Performance spaces[S]. Geneva: International Organization for Standardization, 2009.
- [34] Klockgether S., van de Par S. A Model for the Prediction of Room Acoustical Perception Based on the Just Noticeable Differences of Spatial Perception[J]. *Acta Acustica united with Acustica*, 2014, 100(5): 964-971.
- [35] Pulkki V., Karjalainen M. Communication acoustics: an introduction to speech, audio and psychoacoustics[M]. John Wiley & Sons, 2015.
- [36] Li S., Peissig J. Measurement of Head-Related Transfer Functions: A Review[J]. *Applied Sciences*, 2020, 10(14): 5014.
- [37] Paquier M., Koehl V. Discriminability of the placement of supra-aural and circumaural headphones[J]. *Applied Acoustics*, 2015, 93: 130-139.
- [38] Guo Z., Lu Y., Zhou H., et al. Anthropometric-based clustering of pinnae and its application in personalizing HRTFs[J]. *International Journal of Industrial Ergonomics*, 2021, 81: 103076.
- [39] Pulkki V. Virtual sound source positioning using vector base amplitude panning[J]. *Journal of the Audio Engineering Society*, 1997, 45(6): 456-466.
- [40] 谢菠荪. 空间声原理[M]. 北京: 科学出版社, 2019.
- [41] Gerzon M.J. Periphony: With-Height Sound Reproduction[J]. *Journal of the Audio Engineering Society*, 1972, 21: 2-10.

- [42] Daniel J., Rault J.B., Polack J.D. Ambisonics encoding of other audio formats for multiple listening conditions[A]. Audio Engineering Society Convention 105[C]. San Francisco, CA: Audio Engineering Society, 1998.
- [43] Jot J.M., Larcher V., Pernaux J.M. A comparative study of 3-D audio encoding and rendering techniques[A]. Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction[C]. Rovaniemi, Finland: Audio Engineering Society, 1999.
- [44] Lee H. Multichannel 3D Microphone Arrays: A Review[J]. Journal of the Audio Engineering Society, 2021, 69(1/2): 5-26.
- [45] Lokki T. Recording and reproducing concert hall acoustics for subjective evaluation[A]. International Seminar on Virtual Acoustics[C]. Valencia, 2011: 32-37.
- [46] Hong J.Y., Lam B., Ong Z.T., et al. Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications[J]. Building and Environment, 2019, 149: 1-14.
- [47] Hong J., He J., Lam B., et al. Spatial Audio for Soundscape Design: Recording and Reproduction[J]. Applied Sciences, 2017, 7(6): 627.
- [48] Davies W.J., Bruce N.S., Murphy J.E. Soundscape reproduction and synthesis[J]. Acta Acustica united with Acustica, 2014, 100(2): 285-292.
- [49] Fazi F.M. Sound field reproduction[D]. University of Southampton, 2010.
- [50] Wierstorf H. Perceptual assessment of sound field synthesis[D]. Berlin: Technische Universitaet Berlin (Germany), 2014.
- [51] 龚惠哲. Ambisonic 声重发系统的优化改进[D]. 广州: 华南理工大学, 2009.
- [52] Murillo Gomez D.M., Fazi F.M., Shin M. Evaluation of ambisonics decoding methods with experimental measurements[A]. Proceedings of the EAA Joint Symposium on Auralization and Ambisonics[C]. Berlin, 2014.
- [53] Kuntz M., Bischof N.F., Seeber B.U. Sound field synthesis for psychoacoustic research: In situ evaluation of auralized sound pressure levels[J]. The Journal of the Acoustical Society of America, 2023, 154(3): 1882-1895.
- [54] Zotter F., Frank M. All-Round Ambisonic Panning and Decoding[J]. Journal of the Audio Engineering Society, 2012, 60(10): 807-820.
- [55] Zotter F., Frank M. Ambisonic Decoding with Panning-Invariant Loudness on Small Layouts (AllRAD2)[A]. Audio Engineering Society Convention 144[C]. Audio Engineering Society, 2018.
- [56] Zotter F., Pomberger H., Noisternig M. Energy-preserving Ambisonic decoding[J]. Acta Acustica united with Acustica, 2012, 98(1): 37-47.
- [57] Favrot S., Marschall M., Käsbach J., et al. Mixed-order Ambisonics recording and playback for improving horizontal directionality[A]. Audio Engineering Society Convention 131[C]. Audio Engineering Society, 2011.
- [58] Frank M., Zotter F. Exploring the perceptual sweet area in Ambisonics[A]. Audio Engineering Society Convention 142[C]. Berlin, Germany: Audio Engineering Society, 2017.
- [59] Zotter F., Frank M. Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality[M]. Cham: Springer, 2019.
- [60] 江建亮. 双耳 Ambisonics 声重放系统的研究[D]. 广州: 华南理工大学, 2019.
- [61] Frank M., Zotter F., Sontacchi A. Localization experiments using different 2D Ambisonics decoders[A]. Proceedings of 25th Tonmeistertagung-VDT International Convention[C]. Leipzig, 2008.

- [62] Barrett N. . The Perception, Evaluation and Creative Application of High Order Ambisonics in Contemporary Music Practice[R]. 2012.
- [63] Stitt P., Bertet S., van Walstijn M. Off-Centre Localisation Performance of Ambisonics and HOA For Large and Small Loudspeaker Array Radii[J]. *Acta Acustica united with Acustica*, 2014, 100(5): 937-944.
- [64] Bates E., Kearney G., Furlong D., et al. Localization accuracy of advanced spatialisation techniques in small concert halls[J]. *The Journal of the Acoustical Society of America*, 2007, 121(5): 3069-3070.
- [65] Yang L., Bosun X. Subjective evaluation on the timbre of horizontal ambisonics reproduction[A]. 2014 International Conference on Audio, Language and Image Processing[C]. Shanghai: IEEE, 2014: 11-15.
- [66] 刘阳. Ambisonics 声重放系统的稳定性与音色的研究[D]. 广州: 华南理工大学, 2014.
- [67] Solvang A. Spectral Impairment for Two-Dimensional Higher Order Ambisonics[J]. *Journal of the Audio Engineering Society*, 2008, 56: 267-279.
- [68] 麦海明. 混合阶 Ambisonics 声场与感知性能的研究[D]. 广州: 华南理工大学, 2019.
- [69] Bargum A.R., Kandpal D., Kristjansson O.I., et al. Virtual Reconstruction of a the Ambisonic Concert Hall of the Royal Danish Academy of Music[A]. 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)[C]. 2021: 99-102.
- [70] Shi X., Guo Z., Zhao Y. Influence of virtual audio system on psychological restoration effects of soundscapes: Investigating water sounds of a Chinese classical garden[J]. *Applied Acoustics*, 2024, 221: 109991.
- [71] Neal M.T. Investigating the sense of listener envelopment in concert halls using third-order Ambisonic reproduction over a loudspeaker array and a hybrid room acoustics simulation method[D]. Pennsylvania: The Pennsylvania State University, 2015.
- [72] Favrot S., Buchholz J.M. LoRA: A Loudspeaker-Based Room Auralization System[J]. *Acta Acustica united with Acustica*, 2010, 96(2): 364-375.
- [73] Favrot S.E. A loudspeaker-based room auralization system for auditory research[D]. Denmark: Technical University of Denmark, 2010.
- [74] Noisternig M., Katz B.F., Siltanen S., et al. Framework for real-time auralization in architectural acoustics[J]. *Acta Acustica united with Acustica*, 2008, 94(6): 1000-1015.
- [75] Schröder D., Vorlaender M. RAVEN: A real-time framework for the Auralization of interactive virtual environments[A]. Proceedings of Forum Acusticum[C]. Aalborg, Denmark: Acta acustica united with Acustica, 2011: 1541-1546.
- [76] Seeber B.U., Kerber S., Hafer E.R. A system to simulate and reproduce audio-visual environments for spatial hearing research[J]. *Hear Res*, 2010, 260(1-2): 1-10.
- [77] Li H., Lau S.K. A review of audio-visual interaction on soundscape assessment in urban built environments[J]. *Applied Acoustics*, 2020, 166.
- [78] Davies W.J. , Bruce N.S. , Murphy J.E. . Soundscape reproduction and synthesis[J]. *Acta Acustica United with Acustica*, 2014, 100(2): 285-292.
- [79] Hong J.Y., Lam B., Ong Z.T., et al. The effects of spatial separations between water sound and traffic noise sources on soundscape assessment[J]. *Building and Environment*, 2020, 167: 106423.
- [80] Daniel J. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format[A]. Audio Engineering Society

- Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction[C]. Copenhagen, Denmark: Audio Engineering Society, 2003.
- [81] Poletti M.A. Three-dimensional surround sound systems based on spherical harmonics[J]. Journal of the Audio Engineering Society, 2005, 53(11): 1004-1025.
- [82] Nachbar C., Zotter F., Deleflie E., et al. AMBIX - A Suggested Ambisonics Format[A]. 3rd Ambisonics Symposium[C]. Lexington, KY, 2011.
- [83] Rafaely B. Fundamentals of spherical array processing[M]. Cham: Springer, 2015.
- [84] Williams E.G. Fourier acoustics: sound radiation and nearfield acoustical holography[M]. Elsevier, 1999.
- [85] Brychkov Y.A. Handbook of special functions: derivatives, integrals, series and other formulas[M]. CRC press, 2008.
- [86] Daniel J., Moreau S. Further Study of Sound Field Coding with Higher Order Ambisonics[A]. Audio Engineering Society Convention 116[C]. Berlin, Germany: Audio Engineering Society, 2004.
- [87] Wu Y.J., Abhayapala T.D. Soundfield reproduction using theoretical continuous loudspeaker[A]. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing[C]. 2008: 377-380.
- [88] Ward D.B., Abhayapala T.D. Reproduction of a plane-wave sound field using an array of loudspeakers[J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(6): 697-707.
- [89] Frank M. How to make Ambisonics sound good[A]. Forum Acusticum[C]. Krakow, 2014.
- [90] Makita Y. On the directional localization of sound in the stereophonic sound field[J]. EBU review, 1962, 73(2): 1536-1539.
- [91] Gerzon M.A. General metatheory of auditory localisation[A]. Audio Engineering Society Convention 92[C]. Audio Engineering Society, 1992.
- [92] Carpentier T. Reduction of 3D Ambisonic to 2D using plane-wave decomposition[A]. Audio Engineering Society Convention 152[C]. Audio Engineering Society, 2022.
- [93] Thomas M.R.P., Ahrens J., Tashev I. A method for converting between cylindrical and spherical harmonic representations of sound fields[A]. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. Florence, Italy: IEEE, 2014: 4723-4727.
- [94] Travis C. A new mixed-order scheme for Ambisonic signals[A]. Proc. of the Ambisonics Symposium[C]. Graz, Austria, 2009.
- [95] Mironovs M., Lee H. On the Accuracy and Consistency of Sound Localization at Various Azimuth and Elevation Angles[A]. Audio Engineering Society Convention 144[C]. Audio Engineering Society, 2018.
- [96] Talbot-Smith M. Sound Engineer's Pocket Book 2nd Edition[M]. Routledge, 2000.
- [97] J. Meyer, G. Elko. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield[A]. 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing[C]. 2002: II-1781.
- [98] Abhayapala T.D., Ward D.B. Theory and design of high order sound field microphones using spherical microphone array[A]. 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing[C]. 2002: II-1949-II-1952.
- [99] 莫尔斯 P.M., 英格特 K.U., 吕如榆, 等. 理论声学[M]. 科学出版社, 1984.
- [100] Olver F.W.J., Lozier D.W., Boisvert R.F., et al. NIST Handbook of Mathematical Functions[M]. New York: Cambridge University Press, 2010.

- [101] Naylor G.M. ODEON—Another hybrid room acoustical model[J]. Applied Acoustics, 1993, 38(2-4): 131-143.
- [102] Xiang N. Architectural Acoustics Handbook[M]. J. Ross Publishing, 2017.
- [103] Farina A. Advancements in impulse response measurements by sine sweeps[A]. Audio Engineering Society Convention 122[C]. Audio Engineering Society, 2007.
- [104] Litovsky R.Y., Colburn H.S., Yost W.A., et al. The precedence effect[J]. The Journal of the Acoustical Society of America, 1999, 106(4): 1633-1654.
- [105] Wierstorf H., Spors S. Sound field synthesis toolbox[A]. Audio Engineering Society Convention 132[C]. Audio Engineering Society, 2012.
- [106] Rui Y., Yu G., Xie B., et al. Calculation of individualized near-field head-related transfer function database using boundary element method[A]. Audio Engineering Society Convention 134[C]. Audio Engineering Society, 2013.
- [107] Kong F., Zhou H., Zheng N., et al. Sparse representation of speech using an atomic speech model[J]. The Journal of the Acoustical Society of America, 2025, 157(3): 1899-1911.
- [108] Au E., Xiao S., Hui C.T.J., et al. Speech intelligibility in noise with varying spatial acoustics under Ambisonics-based sound reproduction system[J]. Applied Acoustics, 2021, 174: 107707.
- [109] Ahrens A., Marschall M., Dau T. Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments[J]. Hear Res, 2019, 377: 307-317.
- [110] Oreinos C., Buchholz J.M. Evaluation of Loudspeaker-Based Virtual Sound Environments for Testing Directional Hearing Aids[J]. Journal of the American Academy of Audiology, 2016, 27(7): 541-556.
- [111] Dagan G., Shabtai N.R., Rafaely B. Spatial release from masking for binaural reproduction of speech in noise with varying spherical harmonics order[J]. Applied Acoustics, 2019, 156: 258-261.
- [112] Andersen A.H., de Haan J.M., Tan Z.H., et al. Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions[J]. Speech Communication, 2018, 102: 1-13.
- [113] Guo Z., Yu G., Zhou H., et al. Utilizing True Wireless Stereo Earbuds in Automated Pure-Tone Audiometry[J]. Trends Hear, 2021, 25: 23312165211057367.
- [114] Wong L.L.N., Soli S.D., Liu S., et al. Development of the Mandarin Hearing in Noise Test (MHINT) : Ear and Hearing[J]. [2025].
- [115] Torben P. Psychoacoustic Measuring Methods[M]. Technical University of Denmark, 2005.
- [116] 周华莉, 余光正, 孟庆林. 评估人工耳蜗编码策略的心理声学实验平台研究[D]. 华南理工大学. 2021.
- [117] Levitt H. Transformed up-down methods in psychoacoustics[J]. J Acoust Soc Am, 1971, 49(2): 467-477.
- [118] Shirley B., Kendrick P., Churchill C. The effect of stereo crosstalk on intelligibility: comparison of a phantom stereo image and a central loudspeaker source[J]. Journal of the Audio Engineering Society, 2007, 55(10): 852-863.
- [119] Ahrens A., Marschall M., Dau T. The effect of spatial energy spread on sound image size and speech intelligibility[J]. J Acoust Soc Am, 2020, 147(3): 1368.
- [120] Larsen E., Iyer N., Lansing C.R., et al. On the minimum audible difference in direct-to-reverberant energy ratio[J]. J Acoust Soc Am, 2008, 124(1): 450-461.

- [121] Perrott D.R., Saberi K. Minimum audible angle thresholds for sources varying in both elevation and azimuth[J]. *The Journal of the Acoustical Society of America*, 1990, 87(4): 1728-1731.
- [122] Käsbach J. Development and evaluation of a mixed-order Ambisonics playback system[D]. Denmark: Technical University of Denmark, 2010.
- [123] Pausch F., Behler G., Fels J. SCaLAr – A surrounding spherical cap loudspeaker array for flexible generation and evaluation of virtual acoustic environments[J]. *Acta Acustica*, 2020, 4(5).
- [124] Pulkki V. Localization of amplitude-panned virtual sources II: Two-and three-dimensional panning[J]. *Journal of the Audio Engineering Society*, 2001, 49(9): 753-767.
- [125] Wendt F., Frank M., Zotter F. Panning with Height on 2, 3, and 4 Loudspeakers[A]. Proceedings of the 2nd International Conference on Spatial Audio[C]. Erlangen, Germany, 2014.
- [126] Xie B., Mai H., Rao D., et al. Analysis of and experiments on vertical summing localization of multichannel sound reproduction with amplitude panning[J]. *Journal of the Audio Engineering Society*, 2019, 67(6): 382-399.
- [127] Yi K., Xie B. Local Ambisonics panning method for creating a virtual source in the vertical plane of the frontal hemisphere[J]. *Applied Acoustics*, 2020, 165: 107319.
- [128] Pulkki V. Uniform spreading of amplitude panned virtual sources[A]. Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No.99TH8452)[C]. 1999: 187-190.
- [129] Daniel J. Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia[D]. France: University of Paris, 2001.
- [130] Chang J., Marschall M. Periphony-lattice mixed-order ambisonic scheme for spherical microphone arrays[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018, 26(5): 924-936.
- [131] Grandjean P., Berry A., Gauthier P.A. Sound Field Reproduction by Combination of Circular and Spherical Higher-Order Ambisonics: Part II—Hybrid System[J]. *Journal of the Audio Engineering Society*, 2021, 69(3): 166-181.
- [132] Grandjean P., Berry A., Gauthier P.A. Sound Field Reproduction by Combination of Circular and Spherical Higher-Order Ambisonics: Part I—A New 2.5-D Driving Function for Circular Arrays[J]. *Journal of the Audio Engineering Society*, 2021, 69(3): 152-165.
- [133] Heller A., Benjamin E., Lopez-Lezcano F. Optimized Decoders for Mixed-Order Ambisonics[A]. arXiv, 2021.
- [134] Trevino J., Okamoto T., Iwaya Y., et al. High order Ambisonic decoding method for irregular loudspeaker arrays[A]. Proceedings of 20th International Congress on Acoustics[C]. Sydney, Australia, 2010: 23-27.
- [135] Weller T., Buchholz J.M., Oreinos C. Frequency dependent regularization of a mixed-order Ambisonics encoding system using psychoacoustically motivated metrics[A]. Audio Engineering Society Conference: 55th International Conference: Spatial Audio[C]. Audio Engineering Society, 2014.
- [136] Trevino J., Koyama S., Sakamoto S., et al. Mixed-order Ambisonics encoding of cylindrical microphone array signals[J]. *Acoustical science and technology*, 2014, 35(3): 174-177.
- [137] 钟小丽. 双耳听觉定位及其虚拟实现[M]. 广州: 华南理工大学出版社.

- [138] Meng R., Xiang J., Sang J., et al. Investigation of an MAA Test With Virtual Sound Synthesis[J]. *Front Psychol*, 2021, 12: 656052.
- [139] Cui D., Cai Y., Yu G. A Graphical-User-Interface-Based Azimuth-Collection Method in Autonomous Auditory Localization of Real and Virtual Sound Sources[J]. *IEEE Journal of Biomedical and Health Informatics*, 2021, 25(4): 988-996.
- [140] So R.H., Ngan B., Horner A., et al. Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study[J]. *Ergonomics*, 2010, 53(6): 767-781.
- [141] Seeber B. A new method for localization studies[J]. *Acta Acustica united with Acustica*, 2002, 88(3): 446-450.
- [142] Brungart D.S., Rabinowitz W.M., Durlach N.I. Evaluation of response methods for the localization of nearby objects[J]. *Perception & Psychophysics*, 2000, 62(1): 48-65.
- [143] ITU-R BS. 1284-2, General methods for the subjective assessment of sound quality[S]. ITU, Tech. Rep, 2019.
- [144] Nachtigäller P. Perception of coloration changes in a virtual acoustic environment[D]. Munich: Technical University of Munich, 2022.
- [145] Bates E., Dooney S., Gorzel M., et al. Comparing Ambisonic Microphones—Part 2[A]. Audio Engineering Society Convention 142[C]. Audio Engineering Society, 2017.
- [146] 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.
- [147] Avni A., Ahrens J., Geier M., et al. Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution[J]. *The Journal of the Acoustical Society of America*, 2013, 133(5): 2711-2721.
- [148] Carhart R., Jerger J.F. Preferred method for clinical determination of pure-tone thresholds[J]. *Journal of speech and hearing disorders*, 1959, 24(4): 330-345.
- [149] Carlile S., Leong P., Hyams S. The nature and distribution of errors in sound localization by human listeners[J]. *Hearing Research*, 1997, 114(1): 179-196.
- [150] Leong P., Carlile S. Methods for spherical data analysis and visualization[J]. *Journal of Neuroscience Methods*, 1998, 80(2): 191-200.
- [151] Barron M., Marshall A.H. Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure[J]. *Journal of Sound and Vibration*, 1981, 77(2): 211-232.
- [152] Furuya H., Fujimoto K., Takeshima Y., et al. Effect of early reflections from upside on auditory envelopment.[J]. *Journal of the Acoustical Society of Japan (E)*, 1995, 16(2): 97-104.
- [153] Evjen P., Bradley J.S., Norcross S.G. The effect of late reflections from above and behind on listener envelopment[J]. *Applied Acoustics*, 2001, 62(2): 137-153.
- [154] Power P., Davies B., Hirst J. Investigation into the impact of 3D surround systems on envelopment[A]. Audio Engineering Society Convention 137[C]. Audio Engineering Society, 2014.
- [155] Dick D.A. A New Metric to Predict Listener Envelopment Based on Spherical Microphone Array Measurements and Higher Order Ambisonic Reproductions[D]. Pennsylvania: The Pennsylvania State University, 2017.
- [156] Nowak J., Klockgether S. Perception and prediction of apparent source width and listener envelopment in binaural spherical microphone array auralizations[J]. *The Journal of the Acoustical Society of America*, 2017, 142(3): 1634-1645.

- [157] D'Orazio D. Anechoic recordings of Italian opera played by orchestra, choir, and soloists[J]. The Journal of the Acoustical Society of America, 2020, 147(2): EL157-EL163.
- [158] Berzborn M., Bomhardt R., Klein J., et al. The ITA-Toolbox: An open source MATLAB toolbox for acoustic measurements and signal processing[A]. Proceedings of the 43th Annual German Congress on Acoustics, Kiel, Germany[C]. 2017: 6-9.

附录1 N3D 到 N2D 补偿系数 α 的计算过程

对比球谐函数与角谐波函数系数，二者的系数之差为：

$$\alpha = \frac{1}{N_n^n P_n^n(0)} \quad (\text{A-1})$$

其中， $N_n^n = \frac{2n+1}{\sqrt{2(2n+1)!}}$ ， $P_n^n(0)$ 的求解较为复杂，根据缔合勒让德多项式的闭

式解：

$$P_l^m(x) = 2^l \cdot (1-x^2)^{m/2} \cdot \sum_{k=m}^l \frac{k!}{(k-m)!} \cdot x^{k-m} \cdot \binom{l}{k} \binom{\frac{l+k+m}{2}}{l} \quad (\text{A-2})$$

带入 $x = 0, l = m = n$ ，上式存在 0^0 项，可以验证，在本式中的 $0^0 = 1$ 具有实际物理意义。因此，当 $x = 0$ 时，化简得到：

$$P_l^m(0) = \frac{2^l}{(l-m)!} \prod_{i=0}^{l-1} \left(\frac{l+m-1}{2} - i \right) \quad (\text{A-3})$$

代入 $l = m = n$ ，即可得到：

$$\begin{aligned} P_n^n(0) &= 2^n \prod_{i=0}^{n-1} \left(\frac{2n-1}{2} - i \right) \\ &= \frac{2^n}{2^n} \prod_{i=0}^{n-1} [2(n-i)-1] \\ &= \prod_{t=1}^n (2t-1) \\ &= \frac{2^n \cdot n!}{(2n)!} \end{aligned} \quad (\text{A-4})$$

将式(A-4)代入式(A-1)即可得到 $\alpha = \frac{\sqrt{2} \cdot 2^n \cdot 2!}{\sqrt{(2n+1)!}}$ 。

攻读博士学位期间取得的研究成果

一、已发表(包括已接受待发表)的论文, 以及已投稿、或已成文打算投稿、或拟成文投稿的论文情况(只填写与学位论文内容相关的一部分):

序号	作者(全体作者, 按顺序排列)	题 目	发表或投稿刊物名称、级别	发表的卷期、年月、页码	与学位论文哪一部分(章、节)相关	被索引收录情况
1	Guo Zhenyu (第一作者) Lu Yigang Zhou Huali Li Zhelin Fan Youming Yu Guangzheng	Anthropometric-based clustering of pinnae and its application in personalizing HRTFs,” International Journal of Industrial	Journal of Industrial Ergonomics SCI JCR Q2	2021 1月 81 103076	第二章	SCI 收录
2	Guo Zhenyu (第一作者) Yu Guangzheng Zhou Huali Wang Xianren Lu Yigang Meng Qingling	Utilizing True Wireless Stereo Earbuds in Automated Pure-Tone Audiometry	Trends in Hearing SCI JCR Q1	2021 4月 25 23312165211057367	第三章	SCI 收录
3	Guo Zhenyu (第一作者) Zhao Yuezhe Wang Liliang Chu Yijing Yu Guangzheng	Distance discrimination thresholds of proximal sound sources in a real anechoic environment	Applied Acoustics SCI JCR Q1	2023 1月 203 109223	第一章	SCI 收录
4	Shi Xian Guo Zhenyu (共同第一作者) Zhao Yuezhe	Influence of virtual audio system on psychological restoration effects of soundscapes: Investigating water sounds of a Chinese classical garden	Applied Acoustics SCI JCR Q1	2023 11月 221 109991	第一章	SCI 收录

5	Zhou Hengzhi Zhou Huali Guo Zhenyu (共同通信作者) Meng Qinglin	Automated Pure-Tone Audiometry Using True Wireless Stereo Earbuds with Active Noise Control	International journal of audiology SCI JCR Q2	2024 12月 / 1-9	第三章	SCI 收录
6	Guo Zhenyu (第一作者) Zhou Huali Zhao Yuezhe	Evaluating speech intelligibility degradation under different orders of Ambisonics	24th International Congress on Acoustics EI 收录会议论文	2022	第三章	EI 收录
7	Guo Zhenyu (第一作者) Wang Xianren Zhou Huali Lu Yigang Yu Guangzheng Meng Qingling	Automated pure tone audiometry with true wireless stereos earbuds. The Journal of the Acoustical Society of America	The 179th Meeting of the ASA 会议论文	2020	第三章	\
8	Guo Zhenyu (第一作者) Zhao Yuezhe Seeber Bernhard U.	Evaluation of the performance of a multi-layer Ambisonics reproduction method	The Journal of the Acoustical Society of America, SCI JCR Q2	拟投稿	第五章	拟投稿
9	Guo Zhenyu (第一作者) Zhao Yuezhe	On the influence of reproduction order and range of elevation under virtual room environment	Applied Acoustics SCI JCR Q1	拟投稿	第六章	拟投稿

二、与学位内容相关的其它成果(包括专利、著作、获奖项目等)

致谢

五年多的博士生涯即将结束，从我硕士期间开始接触声学研究至今，已过去七年有余。我与声学学科的缘分起源于本科期间偶然接触到的扬声器制作项目，当时，我们非常不成熟地设计了一款扬声器和功放电路，为它能够成功发声感到十分兴奋。现在，我的博士课题最终也在关上扬声器后结束。接近八年的时间过去，现在的我相比当时有何变化？我想，我从“如何能让它工作”这一问题更多地转向了“如何能让它工作得更好”这一问题。刨根问底，关注每一个细节的工作学习态度，是我在研究生生涯中从身边优秀的人身上学到的最重要的内容。

首先要感谢我的导师赵越皓教授，没有赵老师的帮助与指导，本文的多项研究无从开展。赵老师是一个非常严谨的学者，在一次次通宵达旦的实验测量与论文修改中，身体力行地向我示范了如何开展严肃可信的科学的研究。赵老师对研究方方面面、事无巨细的严谨态度，给我留下了非常深刻的印象，这也是我至今努力学习的方向。

感谢我的硕士指导老师余光正教授与孟庆林教授。感谢你们在我博士期间仍持续不断地给我提供科研、生活与人生上的建议与帮助，多次帮助解决了具体实验中遇到的困难。余光正老师在研究上同样非常细致严谨，带我进入了声学的研究大门。孟庆林老师的热情大度，向我展现了学术研究的多元性和包容性。

还要感谢我在德国访学期间的导师 Bernhard U. Seeber 教授，教授在声学信号处理与听觉研究上的深厚学识令我佩服。Seeber 教授治学严谨，富有激情，总是能第一时间针对我遇到的困难，与我进行深入且有益的探讨，在短时间内教会了我掌握了大量的声学技能，收获匪浅。

感谢实验室的孙海涛老师与褚铁景老师，两位老师在我博士期间给予了我莫大的帮助。孙海涛老师经常抽出宝贵的时间帮助我完成声学测量实验，帮助我确定实验细节和思路，褚铁景老师则是不厌其烦地与我探讨理论分析与研究思路。两位老师的热心让我感受到了实验室团队强大的凝聚力。

感谢吴硕贤院士在论文预答辩上的指导，吴院士在文理诗词乐器上无一不通，让我认识到了，优秀的科学家对人生探索的可能性是无限的。

感谢我的同门王亚敏博士生、李曦博士、高建亮博士、潘丽丽博士以及我的同学周华莉博士、朱俊博士生、谭伟博士、王翘楚博士，感谢他们在日常实验中的帮助与研究上的讨论。与你们的互相交流学习避免了我的博士课题成为闭门造车的研究，漫

长的研究生生涯，是你们在最孤独的日子陪我度过难关。感谢德国 AIP 实验室的 Matthieu Kuntz 博士、Payman Azaripasand 博士生、Maryam Bajool 博士生以及张铧月博士生，你们的热心帮助让我在异国他乡也感受到了温情，帮助我度过了孤独的时刻。还要感谢我的师弟师妹们，石宪、黄兆杰、江泽鑫、周恒志、何彦庆，感谢你们的热情帮助，也祝你们早日毕业！

最后，我要感谢我的家人。读博是一个漫长且孤独的旅途，我的家人在这数年间没有给我任何压力，是我能够放心开展研究的最大基础。你们才是本文研究的最大幕后功臣！

本文的研究得到了非常多人的支持与帮助，篇幅有限，不能一一表示感谢，非常感谢曾经参与我的研究，曾经帮助或参与实验的所有人！

声学研究博大精深，知识体系繁杂，本文尚有非常多的不足之处，希望读者能够多多包涵，并从我的博士论文中有所收获！

郭振宇

2025 年 6 月

四、答辩委员会决议书

答辩委员会对学位（毕业）论文的评语（主要内容包括：1. 论文选题与综述；2. 学术水平与创新性；3. 论文综合能力表现；4. 学位（毕业）论文答辩情况；5. 存在的不足之处和建议等）

郭振宇的博士学位论文《高阶 Ambisonics 可听化重放技术优化及其感知评估》具有重要的理论意义和实际应用价值。

论文建立了一套高阶 Ambisonics 可听化重放软硬件平台，针对 Ambisonics 重放阶数对声场重构误差的影响进行客观模拟分析，同时开展主观听音实验量化系统阶数对言语清晰度的影响；提出面向分层扬声器阵列的重放优化算法，并通过主客观实验比较了该算法与现有相关算法在声像定位与声音保真度方面的性能；进而探索了该系统重放室内反射声的空间信息对听觉的感知影响，为室内声环境可听化系统的简化设计提供了理论与数据支持。研究成果可为推动高阶 Ambisonics 可听化系统在建筑声学等领域 的实际应用提供参考。

论文研究方法正确、实验设计合理、结果真实可靠，写作规范、条理清晰。论文作者在答辩过程中表达清楚，回答问题正确。论文表明作者掌握了宽广扎实的基础理论和系统深入的专业知识，具有较强的工程实践与独立从事科学研究的能力。

答辩委员会一致同意通过博士学位论文答辩，并建议授予工学博士学位。

论文答辩日期：2025 年 9 月 8 日 答辩委员会委员 5 人

毕业表决票数：通过（5）票；不通过（0）票

学位表决票数：通过（5）票；不通过（0）票

答辩委员会决议（打“√”）：

毕业：通过（√）；不通过（ ）

学位：通过（√）；不通过（ ）

答辩成员 签名	李研华（主席） 孙伟伟 王海涛	赵惠娟 张文涛	段东旭
答辩秘书 签名	李柳		