

Synthesis Algorithm for Dynamic Binaural Room Impulse Responses

Haokun Hong, Shaobo Wu[†]

Beijing Information Science and Technology University, Beijing, 100010, China

[†]Email: wushaobo@bistu.edu.cn

Abstract

Current methods for constructing spatial reverberant sound fields required extensive measurements of binaural room impulse responses (BRIRs) or relied on interpolation of multiple spatial room impulse responses (SRIRs), which were time-consuming and computationally intensive. To address these limitations, this paper proposed a lightweight dynamic BRIR synthesis algorithm based on monophonic room impulse responses (RIRs). The algorithm separately processed direct sound, early reflections, and late diffuse reverberation by combining the image source method (ISM) and an auto-regressive moving average (ARMA) model, enabling efficient dynamic BRIR synthesis. Experimental results demonstrated that the proposed method achieved a root mean square error (RMSE) of 0.2 ms for early reflection arrival times, showed an average deviation of 3.7% in reverberation time (RT_{∞}) across full frequency bands, and maintained binaural coherence (IC) below 0.05. The results indicate that the proposed synthesis method effectively balances computational efficiency and spatial acoustic accuracy.

Keywords: Binaural Room Impulse Response; Monophonic Room Impulse Response; Dynamic Synthesis; Auto-Regressive Moving Average Model

动态双耳房间脉冲响应的合成算法

洪浩坤, 吴韶波[†]

北京信息科技大学, 北京 100010

摘要: 现在构建空间混响声场的方法需要采集大量的双耳房间脉冲响应数据或者使用多个空间房间脉冲响应插值, 但是这些方法采集时间长且需要占用大量计算资源。针对上述矛盾, 提出了一种基于单耳房间脉冲响应的动态双耳房间脉冲响应的合成算法, 该方法通过直达声、早期反射和晚期弥漫混响分开处理, 结合镜像源与自回归滑动平均模型, 实现了动态双耳房间脉冲响应的高效合成。根据实验对比分析, 早期反射到达时间均方根误差为 0.2ms, 混响时间全频段平均偏差为 3.7%, 全频段双耳相干性小于 0.05。结果表明, 所提出的合成方法有效地平衡了计算效率和空间声学精度。

关键词: 双耳房间脉冲响应; 单耳房间脉冲响应; 动态合成; 自回归滑动平均模型

1 引言

在虚拟/增强现实 (VR/AR) 系统中, 动态空间声场的构建需要平衡声学一致性与计算开销。房间脉冲响应 (Room Impulse Response, RIR) 虽能表征环境混响特性, 但其单声道属性难以支撑精准的 3D 方位感知。为此, 双耳房间脉冲响应 (Binaural Room Impulse Response, BRIR) 通过嵌入头部相关传输函数 (Head-Related Transfer Function, HRTF), 在 RIR 基础上引入包括时域延迟 (ITD) 和声强差 (ILD) 等参数, 实现声源定位精度与空间沉浸感的同步提升。尽管真实声场重建能为用户提供极高的沉浸感, 但多数应用场景仅需满足感知合理性即可实现虚拟声源与真实环境的自然融合^[1]。当前主流的声场构建方法存在一些局限性: (1) 基于几何声学的声场构建方法虽能物理精确地模拟声场, 但其复杂的场景建模流程, 与高昂的计算成本阻碍了实际

应用；(2)基于测量的双耳声场重建依赖专用设备采集 BRIR 或 HRTF^[2]，单点测量耗时过长且无法支持听者动态移动；(3)基于多通道声场合成的方法虽通过 BRIR 插值实现灵活调整^[3]，但由于计算资源通常需与其他高负载模块共享，合成必须保持轻量化。

合成 BRIR 的常用方法是将声场的定向分量和扩散分量分解和合成。为了确定方向分量，这些方法利用空间房间脉冲响应（Spatial Room Impulse Response, SRIR）的方向信息来估计直达声的 DOA 和早期反射。Sanaguano 等人^[4]提出了一种基于变分自编码器（VAE）和长短期记忆神经网络（LSTM）的快速 BRIR 生成方法，在固定声源的房间内布置静态接收点并生成初始 SRIR，通过改进的双线性插值（MBI）进行数据增强，结合位置聚类与解码器实时生成 BRIR。Gari 等人^[5]提出了一种基于空间分解方法（SDM）与波场合成（WFS）结合的声场再现可行性研究，SDM 通过分析紧凑麦克风阵列测量的 SRIR，将其分解为具有方向信息（DOA）的声学事件序列，然后评估了平面波（VNLS）和点源（PS）在两种 WFS 阵列的 BRIR 合成效果。

通过对近期相关的研究发现，现在构建空间混响声场的方法需要对大量的 BRIR 数据合成和插值或者利用多个 SRIR 进行分解合成运算。但是 BRIR 的采集时间长且无法支持听者动态移动，而对多个 SRIR 的分解合成则需要占用大量计算资源。针对上述矛盾，提出一种基于单声道房间脉冲响应的轻量化双耳声场合成框架。该方法通过声场参数解耦与动态外推机制，从单耳 RIR 中提取直达声幅度、TOA 及早期反射，并利用简化镜像源模型（Image Source Method, ISM）推导反射的 DOA。晚期混响分量通过自回归滑动平均（Auto Regressive Moving Average, ARMA）模型合成，结合双耳噪声分段卷积与扩散场 HRTF 滤波，确保耳间相干性。算法进一步支持听者位置实时更新，通过几何声学动态修正传播路径，实现声源距离、方位及混响特性的自适应调整。实验结果表明，合成 BRIR 在关键声学参数上还原性良好。本方法通过单耳 RIR 的低成本采集与轻量化合成，为 VR/AR 动态声场渲染提供了可扩展的解决方案。

2 合成双耳房间脉冲响应的算法

2.1 算法模型

算法的核心创新在于仅使用一个人工合成的单声道房间脉冲响应作为输入，即可生成可用于动态声学模拟的双耳房间脉冲响应，而无需依赖额外的多声道 RIR 进行声场分解。图 1 展示了本算法的基本结构。在处理过程中，算法首先将 RIR 分为早期部分和晚期部分。早期部分包括直达声和早期反射，其中直达声和早期反射的 TOA 可以通过时域分析计算得到，而它们的 DOA 通常无法直接从 RIR 中获得，因此算法使用基于鞋盒形房间的 ISM^[6]进行推导获得。对于晚期混响部分，则通常被认为在时间和空间上均匀分布地到达听者，测量脉冲响应中超过感知混响时间后的部分可以进行简化处理^[7]。所以晚期混响通过自回归滑动平均模型合成，并通过与双耳噪声卷积实现空间扩散特性的重建。此外，本算法支持听者位置实时更新，能够实现虚拟房间内任意方向的声场模拟，相较于传统全声场建模方法显著降低了计算资源需求，增强了算法的可用性。

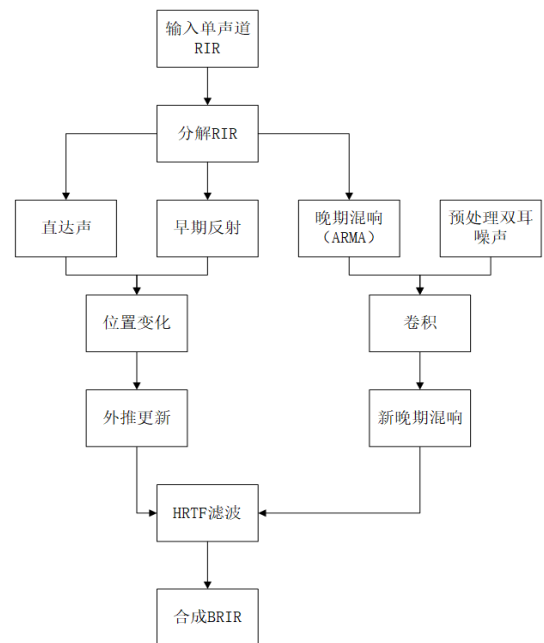


图 1 双耳房间脉冲响应合成的结构图

2.2 早期部分的 TOA

2.2.1 直达声

直达声的到达时间 t_d 通过起始检测阈值确定，该阈值设置为 10 倍上采样后的 RIR 的最大值的-20dB。直

达声的振幅 A_d 计算采用了 t_d 周围的非对称窗口的 RMS。

$$A_d = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (1)$$

其中 N 是非对称窗口内的采样点数量， x_i 是 RIR 信号中的采样值。根据文献^[8]的建议，窗口从范围 t_d 的前 0.5ms 开始，到 t_d 后 1ms 结束。窗口起始点略微提前于 t_d 的设计旨在有效抑制预振铃伪影，而后 1ms 的窗口长度则与直达声能量集中区的时间范围相匹配，确保振幅计算的准确性。

2.2.2 早期反射

对于早期反射，为提取感知显著的早期反射，算法采用一个滑动的矩形窗口（1ms）对 RIR 进行窗口化处理，其中反射的到达时间 T_a 被定义为局部能量达到窗口中能量中位数的 3 倍的时间点。对于每个检测到的反射，在一个 1.5ms 的非对称窗口中，计算振幅的 RMS 以及峰值振幅，以确保准确捕捉感知显著的早期反射。为避免检测到在时间上非常接近的多个反射的情况，算法在检测到反射后 1 ms 内进行冲突检测。若存在另一个反射，则保留振幅较大的反射作为有效反射，而将具有较低振幅的反射排除。

$$T_k = \{k | E_k \geq 3E_m\} \quad (2)$$

$$T_a(i) = \{i | A_{T_k} > A_{T_{k+1}}\} T_k - T_{k+1} \leq 1ms, i = 1, 2, \dots, 10 \quad (3)$$

其中， T_k 为早期反射的初步筛选的时间点， E_k 为在窗口内的第 k 个采样点的局部能量，能量中位数为 E_m ， A_{T_k} 为时间点 T_k 的振幅。

研究表明，选择最响亮的 i 个反射（其中 $i=6\sim10$ ）是感知上重现反射的有效方法^[8]。因此，算法根据反射振幅对所有检测到的反射进行排序，并选择前 10 个最响亮的反射进行空间化与动态再现，如图 2 所示。

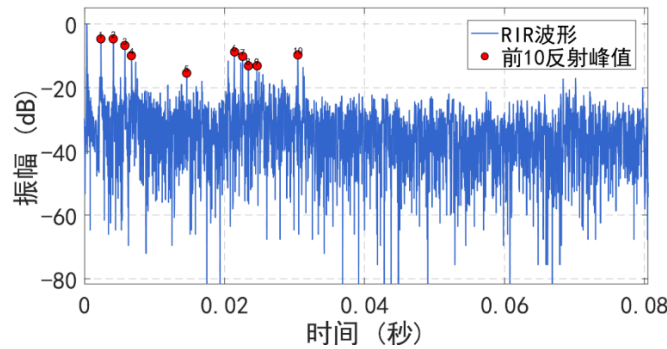


图 2 RIR 的早期反射分解

2.3 早期部分的 DOA

在需要将人工 RIR 生成 BRIR 的 VR/AR 的声学渲染系统中，基于已知的房间几何参数以及声源和接收器位置坐标，可以近似房间的实际几何形状，可通过 ISM 重构直达声与早期反射的到达方向。理论分析与感知评估表明，6~10 个早期反射通常已能有效表征早期声场特性，因此本算法采用二阶镜像反射模型以平衡计算效率与物理精度^[1]。

为了估算所选早期反射的入射方向，利用 1.1 节估算得到的 TOA 与模拟计算的 TOA 进行比较，并将 TOA 相等或 TOA 差异最小的反射进行配对并选择相应的方向。然后，将选定早期反射的 DOA 分配为对应模拟反射的 DOA。

2.4 晚期部分

人工混响生成的 RIR 的后期弥漫混响受限于混响器模型复杂度，导致声场方向性过强或空间感不够，声场不具有高度扩散性。为了解决这个问题，晚期弥漫混响通过 ARMA 模型生成。

晚期部分的弥散混响被视为在时间和空间上均匀到达听者的高斯随机过程。由于听者无法感知到与完全弥散混响之间的显著差异，因此无需对其时间和空间结构进行精确重建。本算法采用了早期反射与弥散混响之间的重叠模型^[9]。

弥散混响 $h(t)$ 参考文章^[10]中的 ARMA 模型进行合成。可以用功率时间曲线(Power Temporal Profile, PTP)来表示晚期部分在每个时间点的功率期望值：

$$\bar{h}(t) = E[|h(t)|^2] = P_0^2 e^{-2t/\tau}, \quad 1t \in T = \begin{cases} 1 & t \in T \\ 0 & t \notin T \end{cases} \quad (4)$$

其中 P_0 是一个与晚期混响总功率相关的定值， τ 与混响时间 RT_{60} 相关，可以描述混响能量的衰减速率：

$$\tau = \frac{RT_{60}f_s}{3\log(10)} \quad (5)$$

其中 RT_{60} 表示混响时间，即声波衰减到初始声音的 60 dB 所需的时间， f_s 是采样率。

由于 $H(f)$ 是广义高斯平稳随机过程，则该过程的自协方差函数 $R(m)$ 和功率谱密度函数 $S(t)$ 为：

$$R(m) = E[H(f)H_i(f-m)^*] \quad (6)$$

$$S(t) = \frac{1}{T} E[|F_T\{H(f)\}|^2] \quad (7)$$

其中 $F_T\{H(f)\}$ 是 $H(f)$ 的离散傅里叶变换， $H_i(f-m)^*$ 表示 $H_i(f-m)$ 复数共轭。

功率谱密度函数与功率时间曲线之间有如下关系：

$$S(t) = T\bar{h}(T-t) \quad (8)$$

根据维纳-辛钦-爱因斯坦定理、诺里斯-艾林公式^[11]和(8)，可以更新自协方差函数为：

$$R(m) = \sigma_r^2 \frac{1-e^{2/\tau}}{1-e^{2(T-t_0+1)/\tau}} \frac{1-e^{(j2\pi m/T+2/\tau)(T-t_0+1)}}{1-e^{j2\pi m/T+2/\tau}} \quad (9)$$

$$\sigma_r^2 = E[|H(f)|^2] = C_r \frac{1-\alpha}{\pi\alpha S} \quad (10)$$

$$\alpha = 1 - e^{-24\ln(10)V/(cSRT_{60})} \quad (11)$$

其中 α 是平均吸收系数， C_r 是一个经验常数为 75, S 是房间的总面积， c 为光速。

根据^[12]的理论框架，晚期混响分量可通过因果可逆的自回归滑动平均模型表征。其频域表达式为：

$$\sum_{i=1}^p \alpha_i L^i H(f) = \sum_{i=1}^q \beta_i L^i N(f) \quad (12)$$

其中 L 为滞后算子， $\sum_{i=1}^p \alpha_i L^i$ 和 $\sum_{i=1}^q \beta_i L^i$ 分别为自回归部分和滑动平均部分的参数， $N(f)$ 为中心化的复高斯噪声，方差为 σ_N^2 。ARMA 的参数 α_i 和 β_i 通过 $R(m)$ 与理论自协方差函数 (ACVF) 的数学对应关系，建立并解非线性方程组算出。

由这个 ARMA 模型可得联合概率分布：

$$p(H(f))_f \propto \prod_{f=0}^{T/2} Z_c\left(\frac{\sum_{i=1}^p \alpha_i L^i H(f)}{\sum_{i=1}^q \beta_i L^i}; 0, \sigma_N^2\right) \quad (13)$$

晚期弥漫混响通过混响时间 RT_{60} 、体积 V 、表面积 S 推导的理论自协方差函数 $R(m)$ 求解，首先设定 AR 阶数 $p = 10$ 与 MA 阶数 $q = 1$ ，基于 Yule-Walker 方程构建非线性方程组解析模型参数 α_i 和 β_i ，随后通过傅里叶反变换将频域响应 $H(f)$ 转换为时域信号，最后得到晚期弥漫混响。

为了合成的 RIR 保持实测 RIR 的频率相关包络以及双耳噪声的双耳特征和相关的声音广阔感。弥散混响通过将合成的晚期部分的小时间段与双耳高斯白噪声的时间段进行分段卷积叠加来合成，即采用 0.67ms (48 kHz 采样率下 32 样本) 短时噪声片段增强空间感知，避免 2.67ms (128 样本) 长片段导致的中性音染。然后通过 HRTF 的弥散声场响应和弥散声场耳间相干性的一致性滤波器进行滤波，确保其在物理精度与听觉感知之间达成平衡。

3 听者位置变化

当听者在 VR/AR 的动态声学场景中里自由移动时，静态 BRIR 因无法实时适配听者位置与朝向的动态变化，将导致声源定位偏差和距离感知失真，严重削弱沉浸式体验的感知真实性。单声道 RIR 和空间参数足以在接收位置对听者的任何头部转向进行动态合成，但是对于听者的位置变化，也需要调整直达声和早期反射的振幅、TOA 和 DOA。为实现听者自由移动时的动态声场重建，本算法基于 ISM 的简单几何模型通过以下机制实现参数自适应调整。每个镜像声源与听者之间的距离由相应的反射峰与直达声峰之间的延迟来确定。另外根据听者位置的变化，重新计算每个反射的距离、入射方向和振幅。

图 3 为听者位置变化的简单示意图，实线表示原始位置下直达声与一阶早期反射的声传播路径，虚线表征位置变化后的路径修正。

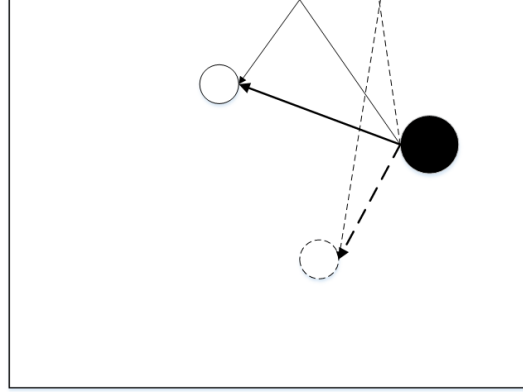


图 3 听者位置变化示意图

假设原始听者位置处测得到达时间 t_i ，到达方向 (ϕ_i, θ_i) ，振幅系数 A_i 。通过声速 c 计算距离 d_i ：

$$d_i = t_i \cdot c \quad (14)$$

映射至球坐标系：

$$\begin{cases} x_i = d_i \sin \theta_i \cos \phi_i \\ y_i = d_i \sin \theta_i \sin \phi_i \\ z_i = d_i \cos \theta_i \end{cases} \quad (15)$$

给定听者位移向量 $\Delta \mathbf{D} = (\Delta x, \Delta y, \Delta z)$ ，更新声源相对坐标：

$$\mathbf{D}'_i = \mathbf{D}_i - \Delta \mathbf{D} = (x_i - \Delta x, y_i - \Delta y, z_i - \Delta z) \quad (16)$$

计算修正后的声学参数：

$$\begin{cases} d'_i = \|\mathbf{D}'_i\| \\ \phi'_i = \arctan 2(y'_i, x'_i) \\ \theta'_i = \arccos\left(\frac{z'_i}{d'_i}\right) \\ t'_i = \frac{d'_i}{c} \\ A'_i = A_i \left(\frac{d_i}{d'_i}\right)^2 \end{cases} \quad (17)$$

其中振幅 A'_i 根据与声源或镜像声源的距离变化，使用平方反比定律进行调整。

4 实验结果及讨论

本文采用基于房间脉冲响应分解的双耳音频重建方法进行仿真验证。算法输入端为采样率 4.8 kHz、32 位深度的纯净 RIR 信号，经 RIR 分解、晚期混响成分合成及个性化 HRTF 滤波处理，最终生成高保真 BRIR 信

号。实验数据来源于 Bacila 等^[13]开发的开放式 360°双耳声场数据库，该数据库在声学特性可控的实验室内采集，通过六自由度头部旋转装置与哈德斯菲尔德声学分析研究工具箱（Huddersfield Acoustic Analysis Research Toolbox, HAART）^[14]的系统级集成，实现了 13 个接收位点的全自动高精度数据采集。各位置测量分布图如图 4 所示。

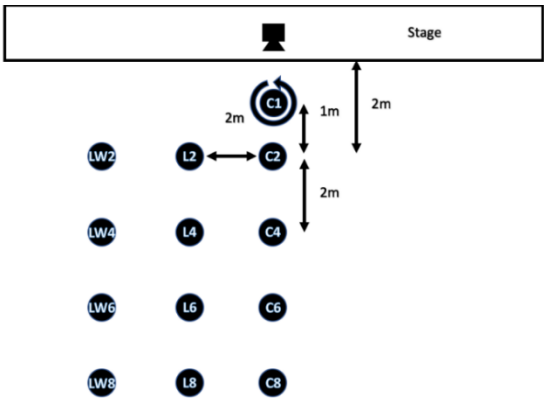


图 4 各位置测量分布图

本文选取 C2 测点的 RIR 样本进行算法验证。房间音质的主观研究表明，由脉冲响应测量得到的几个参数与房间音质的主观感受有密切关系，混响时间 RT_{60} 是描述房间音质的基础指标，同时一些如双耳相干性 IC 的新参数可以描述人们在房间中主观感受的“空间印象”，这些参数结合分析能更全面的描述房间音质。

图 5 展示了实测与合成 BRIR 信号左耳声道的时域波形，左侧为实测 BRIR，右侧为合成 BRIR。

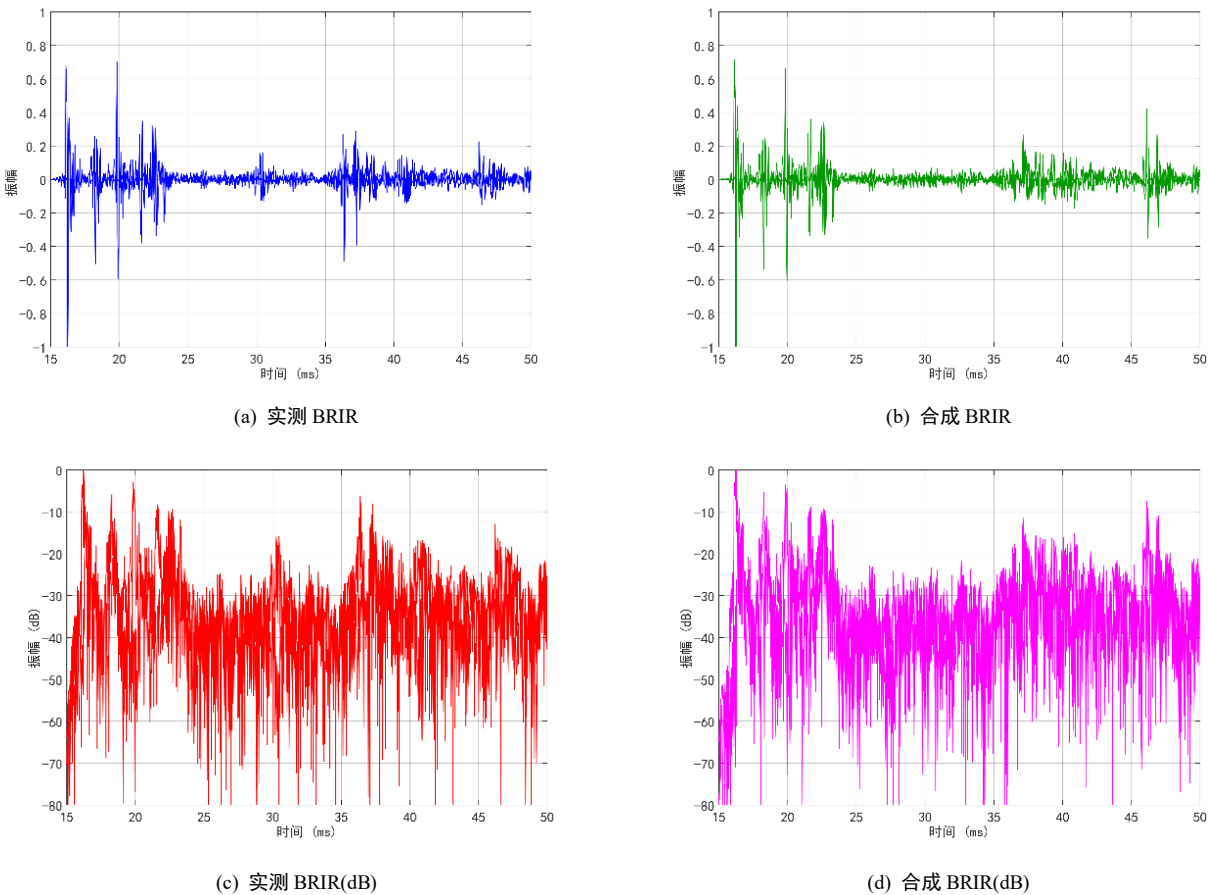


图 5 实测 BRIR 与合成 BRIR (左耳)

从整体上分析，早期反射到达时间均方根误差为 0.2ms。对比分析(a)、(b)，从早期混响成分（15ms~30ms 区间）可见，合成 BRIR 在时延特征与振幅分布方面与实测数据具有良好一致性，表明本算法能有效保留直达声与早期反射的物理传播特性。但在中期混响阶段（30ms~45ms），由于实验策略仅提取前 10 个最响亮的早期反射，导致该区间的反射密度显著低于实测数据。在后期混响阶段（>45ms），合成 BRIR 呈现出了更强的反射，考虑是晚期弥漫混响合成模块叠加产生的效果。此外，进一步分析(c)、(d)，因为晚期弥漫混响部分是通过与双耳白噪声叠加卷积，所以能量结构也保持了相似性。

图 6 展示的是基于 1/3 倍频程的实测与合成 BRIR 分频段 RT_{60} 的对比分析结果，选择中心频带为 250~8k 覆盖了人耳可听范围的主要部分的标准的声学分析频段。

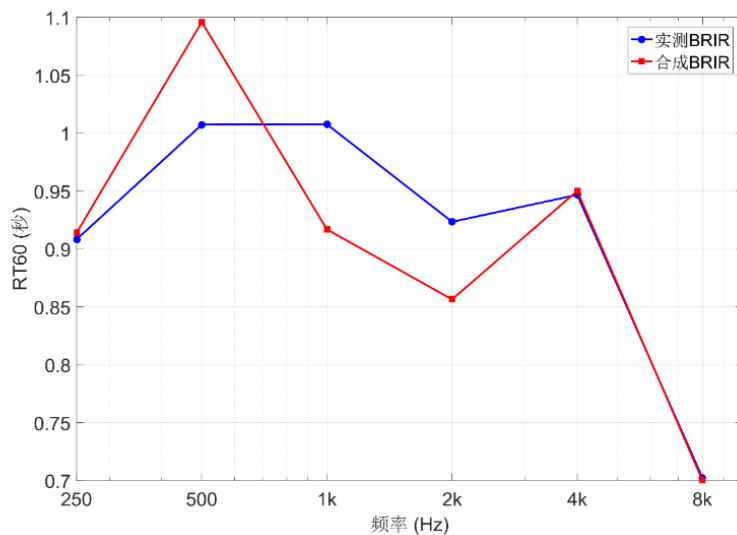


图 6 BRIR 分频段混响时间

图像数据显示在 250~750Hz 的中低频频带上，合成的 BRIR 的平均 RT_{60} 高于实测数据 5.6%；而在 1.5k~4k 的核心语音频带上，合成值较实测值平均降低 6.7%；4k 及以上的高频频带的 RT_{60} 基本相同。从整体上看，每个频段上 RT_{60} 差距均小于 0.1s，全频段平均偏差为 3.7%，表明提出的 RIR 分解合成 BRIR 效果良好。此外，还表明在有效保持了扩散声场衰减的物理本征特性的情况下，所提出的后期弥漫混响合成模块非常适合用于从 RIR 合成 BRIR。

图 7 比较实测 BRIR 和合成 BRIR 分频段的 IC。

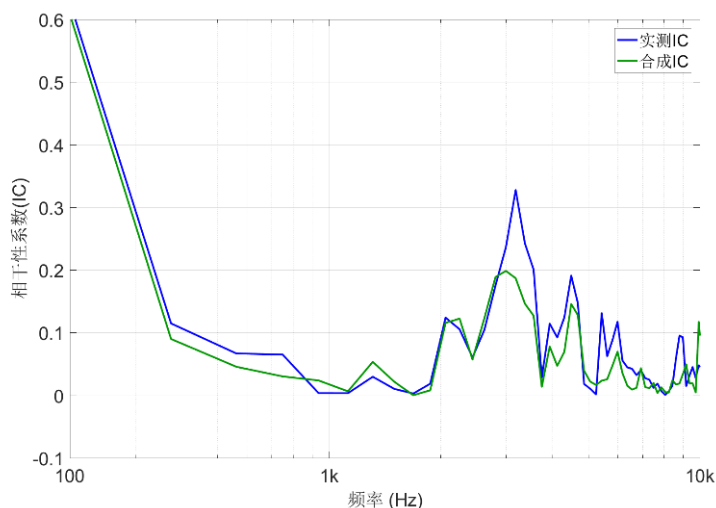


图 7 BRIR 分频段 IC

根据^[15]的算法框架，通过汉明窗(窗口长度 256 样本点，5.33ms)，50%帧重叠率（步长 2.67ms）计算。根据图像显示，在 0~300Hz 的低频频带，IC 具有高度一致性。然而在 3kHz 以上的高频频带，两者出现轻度差异，可能是合成过程中 HRTF 的滤波特性产生了非相干分量。而全频段平均差异 $\Delta IC < 0.05$ 。

RT₆₀以及 IC 是从整体混响特性和整体相似性进行评估，下面从早期主观混响感和集中声场空间特性分析每个测量点位早期衰变时间 EDT 和双耳互相关 IACC，检验算法的性能。所合成的各点位 BRIR 均基于 C2 位置处测量的 BRIR，并通过本章节算法生成，具体数值结果如下表 1。

表 1 各点位 EDT 与 IACC

点位	EDT(左耳) (s)		IACC	
	实测 BRIR	合成 BRIR	实测 BRIR	合成 BRIR
C2	0.82	0.79	0.22	0.17
C4	0.85	0.80	0.35	0.31
C6	0.89	0.85	0.33	0.36
C8	0.92	0.89	0.20	0.26
L2	0.86	0.82	0.16	0.13
L4	0.88	0.86	0.39	0.28
L6	0.91	0.89	0.39	0.40
L8	0.95	0.93	0.22	0.18
LW2	0.88	0.84	0.13	0.13
LW4	0.92	0.86	0.35	0.39
LW6	0.93	0.86	0.38	0.39
LW8	0.97	0.93	0.22	0.18

EDT 作为表征声场早期能量衰减特性的关键参数，其数值直接影响语音清晰度和音乐细节的主观感知密切相关。EDT 的计算是利用 Schroeder 反向积分法得到的能量衰减曲线，选取 0~-10dB 线性区间进行线性拟合后用拟合斜率外推到-60dB 确定衰减时间。表 1 左侧显示的是各点位的 EDT 的实验数据，从整体上看，实测 BRIR 与合成 BRIR 的 EDT 相对偏差约为 5%，只有 C4 和 LW4、LW6 大于 5%，相对偏差的点位出现在 LW6 约为 7.5%，基本处于 ISO 3382-1 标准^[16]中提出的 5%混响 JND 阈值。

IACC 是特定于听觉空间感知的指标，通过限制时间延迟范围，量化声场的集中或扩散特性。IACC 值越接近 1，声场越集中；值越接近 0，声场越扩散。低 IACC 值表明声场扩散性良好，听众感知空间感更强。本文选择 250Hz~8kHz 的倍频程测量 IC，在 $\tau = \pm 1\text{ms}$ 内查找最大绝对值，即为 IACC 的值。表 1 右侧显示的是各点位的 IACC 数值，仅在 L4 点位差值绝对值为 0.11，其他点位的差值绝对值均小于 ISO 3382-1 定义的 0.075 的 JND 阈值。C2 点位处于 JND 范围内，其余点位差异均显著低于阈值。整体而言，测量 BRIR 的 IACC 的值域范围在 0.13~0.39，合成 BRIR 则为 0.13~0.40，表明所提出算法在声场扩散性方面和实测 BRIR 具有高度一致性。

5 结论

本研究提出了一种基于单耳房间脉冲响应的动态双耳房间脉冲响应的合成算法，该方法通过直达声、早期反射和晚期弥漫混响分开处理，结合 ISM 与 ARMA 模型，实现了声场参数的高效提取与动态重建。实验结果表明，合成 BRIR 在关键声学参数上与实测数据较为一致：早期反射到达时间均方根误差为 0.2ms，RT₆₀ 全频段平均偏差为 3.7%，全频段平均差异 $\Delta IC < 0.05$ 。在 12 个测试点位中，EDT 和 IACC 大部分显著低于 ISO 3382-1 定义的 JND 阈值($< 5\%$, < 0.075)，验证了算法在保持声场物理特性与听觉感知一致性方面的有效性。验证了算法在保持声场物理特性与听觉感知一致性方面的有效性。此外，算法通过根据听者位置变化动态更新参数，实现了合成 BRIR 的动态修正，结合公开单声道 RIR 数据库，为低成本、高灵活性的 VR/AR 应用声场构建提供了一种解决方案。

参考文献

- [1] Neidhardt A, Schneiderwind C, Klein F. Perceptual matching of room acoustics for auditory augmented reality in small rooms - literature review and theoretical framework[J]. Trends in Hearing, 2022, 26: 23312165221092919.
- [2] 魏永健,周静雷.基于人体测量参数的个性化 HRTF 建模[J].电工技术,2024,(18):118-122.DOI:10.19768/j.cnki.dgjs.2024.18.033.
- [3] Zhao J, Zheng X, Ritz C, et al. Interpolating the directional room impulse response for dynamic spatial audio reproduction[J]. Applied Sciences, 2022, 12(4): 2061.
- [4] Sanaguano-Moreno D, Lucio-Naranjo J F, Tenenbaum R A, et al. Rapid BRIR generation approach using Variational Auto-Encoders and LSTM neural networks[J]. Applied Acoustics, 2024, 215: 109721.
- [5] Gari S V A. The spatial decomposition method meets wave field synthesis: A feasibility study[J]. Proc. 50. Jahrestagung Für Akustik (DAGA 2024), 2024: 1630-1633.
- [6] Allen J B, Berkley D A. Image method for efficiently simulating small-room acoustics[J]. The Journal of the Acoustical Society of America, 1979, 65(4): 943-950.
- [7] Pörschmann C, Zebisch A. Psychoakustische Untersuchungen zu synthetischem diffusen Nachhall Psychoacoustic Investigations on synthetically created diffuse Reverberation[J]. Proc. of the 27th Tonmeistertagung-VDT Int. Conv, 2012: 539-550.
- [8] Brinkmann F, Gamper H, Raghuvanshi N, et al. Towards encoding perceptually salient early reflections for parametric spatial audio rendering[C]//Audio Engineering Society Convention 148. Audio Engineering Society, 2020.
- [9] Coleman P, Franck A, Jackson P J B, et al. Object-based reverberation for spatial audio[J]. Journal of the Audio Engineering Society, 2017, 65(1/2): 66-77.
- [10] 吴礼福,陶明明,郭业才.频域合成房间频率响应的人工混响方法[J].应用声学,2020,39(02):163-168.
- [11] Speech dereverberation[M]. Springer Science & Business Media, 2010.
- [12] Leglaive S, Badeau R, Richard G. Autoregressive moving average modeling of late reverberation in the frequency domain[C]//2016 24th European Signal Processing Conference (EUSIPCO). IEEE, 2016: 1478-1482.
- [13] Bacila B I, Lee H. 360 binaural room impulse response (BRIR) database for 6dof spatial perception research[C]//Audio Engineering Society Convention 146. Audio Engineering Society, 2019.
- [14] Johnson D, Harker A, Lee H. HAART: a new impulse response toolbox for spatial audio research[J].Journal of The Audio Engineering Society, 2015.
- [15] Menzer F, Faller C. Stereo-to-Binaural Conversion Using Interaural Coherence Matching[C]//AES 128th Convention.2010.DOI:10.1111/j.1463-1326.2011.01431.x.
- [16] GB/T 36075.1-2018,声学 室内声学参量测量 第1部分:观演空间[S].