

非均匀二次声源分布下的 2.5D 高阶 Ambisonics 声场合成算法

董 石¹, 章 强¹, 谢夏风¹, 钟 睿²

(1. 华中师范大学教育信息技术学院, 湖北武汉 430000; 2. 华中师范大学计算机学院, 湖北武汉 430000)

摘 要: 为了提高感兴趣区域音源的声场合成质量, 提出了一种非均匀二次声源分布的声场分析与合成方法. 首先, 将连续二次声源的采样分为两个不同的采样区域, 在总扬声器数目一定的情况下, 增大感兴趣区域的样点数目并减少感兴趣区域外样点数目. 然后, 利用采样混叠矩阵的 Hermitian 特性, 使用瑞利熵来估计不同非均匀采样方案的最大误差, 以指导非均匀采样方案的选择. 最终得到非均匀二次声源驱动信号的频域计算方法. 实验结果表明在 2.5D 圆形区域边界下, 所提非均匀排布方法与典型的均匀排布的最佳模态带宽相比, 对感兴趣区域音源的无混叠半径提高了 12.9% ~ 22.6%, 验证了提出方法的有效性.

关键词: 声信号处理; 非均匀采样; 声场合成; 球谐函数

中图分类号: TN911 **文献标识码:** A **文章编号:** 0372-2112 (2021)01-0014-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20200010

Non-Uniform Secondary Source Distribution for 2.5D Higher Order Ambisonics Sound Field Synthesis

DONG Shi¹, ZHANG Qiang¹, XIE Xia-feng¹, ZHONG Rui²

(1. School of Educational Information Technology, Central China Normal University, Wuhan, Hubei 430000, China;

2. School of Computer, Central China Normal University, Wuhan, Hubei 430000, China)

Abstract: In order to improve the sound field synthesis quality of the sound source in the region of interest, a non-uniform secondary source distribution method for sound field analysis and synthesis is proposed. First, the sampling of the continuous secondary source is divided into two different sampling regions. With a certain number of loudspeakers, the number of samples in the region of interest is increased and the number of samples outside the region of interest is reduced. Then, using the Hermitian characteristic of the sampling aliasing matrix, Rayleigh entropy is used to estimate the maximum error of different non-uniform sampling schemes to guide the selection of non-uniform sampling schemes. Finally, the frequency domain calculation method of the driving signal for the non-uniform secondary source distribution is obtained. The experimental results show that the proposed non-uniform distribution method improves the aliasing-free radius by 12.9% ~ 22.6% for the sound sources from the region of interest under the 2.5D circular boundary condition, compared with the optimal modal bandwidth of the uniform distribution. The effectiveness of the proposed method is verified.

Key words: acoustic signal processing; non-uniform sampling; sound field synthesis; spherical harmonics

1 引言

随着虚拟现实技术的复兴和 3D 影视产业的蓬勃发展, 3D 音频越来越受到业界和学术界的关注. 目前有两类 3D 音频技术受到广泛关注, 一类基于听觉的感知重构, 如头部相关传递函数 (Head-Related Transfer

Function, HRTF); 另一类基于物理的声场合成 (Sound Field Synthesis, SFS), 如波场合成 (Wave Field Synthesis, WFS) 和近场补偿高阶 Ambisonics (Near-Field Compensated Higher Order Ambisonics, NFC-HOA). 此外, 还有一类多通道 3D 音频技术, 如 NHK 22.2 多通道系统. 这类技术源于传统的环境声系统并与其兼容, 可以使

用矢量幅度平移 (Vector Base Amplitude Panning, VBAP)^[1,2]生成可感知的 3D 声效或使用 SFS 算法^[3,4]生成物理 3D 声效. 物理三维声场合成需要非常多的扬声器,数量与重构声场的频率成正比. 但受到成本和环境的限制,扬声器的数量不能任意地增加. 因此,如何在扬声器数量一定的情况下提高声场重建质量成为一个重要课题.

NFC-HOA 的主要目标是重构球形或圆形边界中心区域的声场. NFC-HOA 的一种基本合成方法称为模式匹配,它使用离散的次级声源分布 (Secondary Source Distribution, SSD) 来匹配原始声场的球谐函数系数^[5]. 对于高频声场,扬声器数目一般少于奈奎斯特采样的需求,HOA 扬声器驱动信号的求解是一个超定问题. 如果按照重建误差的 l_2 范数最小化标准,个别扬声器的信号会变得非常大,会分散听众的注意力并导致 3D 听感失真. 在 2005 年, Poletti 提出了一种平滑驱动信号能量的加权函数来缓解此问题^[6]. HOA 要求扬声器位于相同半径的球形表面上,播放环境一般无法满足这种严格要求. 2011 年, Gupta 等将 SSD 的半径视为变量,并利用连带勒让德函数来获得多半径的 HOA 驱动信号,这使得扬声器的灵活布置成为可能^[7]. 从空间频率分析的观点来看, Ahrens 等认为 NFC-HOA 是窄带信号处理方法,其频率被截断以减少由驱动信号采样引起的频率混叠^[8]. 2014 年, Zhang 等将希尔伯特空间映射算子的概念引入声场合成,并将扬声器驱动信号的求解视为空间映射,以支持非球形和非均匀的二次声源分布^[9]. 此外,其他一些方法也被提出以减轻空间混叠并为感兴趣区域重构声场^[10],如局部带宽限制 (Spatial Bandwidth Limitation, SBL) 和频谱加窗.

WFS 利用扬声器阵列重构参考位置的声场,也需要大量扬声器. 2014 年, Koyama 利用人类对声源仰角变化的不敏感性,使得 WFS 扬声器配置得到了简化. 他们使用圆柱阵列扬声器和波场重构滤波器来生成驱动信号. 结果表明,与传统的平面阵列相比,圆柱阵列可以减少中心区域的空间混叠现象^[11]. 2017 年, Firtha 等用稳态相位近似法 (Stationary Phase Approximation, SPA) 来分析任意形状的二次声源分布的无失真重建区域^[12]. 随后在 2018 年, Firtha 等又利用 SPA 用于 WFS 的驱动信号分析,推导出的局部波数关系能确定驱动信号与无失真重建位置的物理关系^[13]. Winter 等还研究了 WFS 的音色失真问题,主观试验证明调整 SBL 参数可以缓解重建声场的音色问题^[14]. Hahn 等引入了时域 WFS 驱动信号计算方法来合成 2.5D 平面声场,与频域的驱动信号计算相比降低了复杂度^[15]. 国内,贾懋坤等学者研究了声场合成中聚焦音源的近场失真问题,通过对称 Sigmoid 函数调节高频声场的频域能量分布,减

少了聚焦音源的合成失真^[16,17]. 李娟等研究了室内声场合成受反射声影响失真的问题,并进一步利用生成反射声增强 3D 声场的距离感^[18,19].

总之, HOA 对于球形区域具有更好的声场质量,但在无混叠区域之外的声场失真则更为明显. 近年, HOA 驱动信号的加权函数被重新用于改善中心区域外部的声场的方向性^[20,21]. Okamoto 等提出了一种角度频谱到球面频谱的转换方法,该方法可以使圆形扬声器阵列重建线性麦克风阵列记录的声场^[22]. Winter 等提出了一种几何模型来预测声场的频率失真,并利用 SPA 进行分析,直接将单个位置的声场失真与 SSD 中的特定位置的二次声源相关联. SPA 可以预测任意凸 SSD 区域内的无失真合成区域^[23]. Kentgens 等提出在能够预先知道主声源方向的情况下,通过在低阶球谐函数系数中保留更多能量来提高合成声场质量^[24]. 另外, Mail 等人还研究了水平扬声器的数量对声场音色的影响^[25]. 如何用更少声道的扬声器阵列或耳机,来进行 HOA 声场合成也受到了关注^[26,27].

可见,在限定数目的二次声源下提高声场质量是一个重要问题,而非均匀二次声源排布是一种可行的方法. 实际上, 3D 音频通常与 3D 视频一起使用,其中大多数声源应来自屏幕方向^[3,28,29]. 针对性地优化来自这些方向声源的声场将是更有价值的. 本文首先在理论上分析了存在频率混叠时均匀采样和非均匀采样的表现,结果表明非均匀采样可以减少来自采样密集区域的声源的频域混叠. 随后利用混叠矩阵的 Hermitian 特性,利用 Rayleigh 熵提出了非均匀采样误差估算方法,构成了非均匀采样的选择标准. 声场合成仿真结果表明,提出的非均匀二次声源排布方法能更好地保存感兴趣区域声源的信号能量,提高声场合成质量.

2 理论模型与方法

2.1 频域混叠分析

根据亥姆霍兹积分方程,声场合成通常需要边界处的声压和粒子速度. NFC-HOA 的简单声源法可以舍去粒子速度,代价是只能重建边界内的声场. 如图 1 所示,内部区域的声场 $f_i(\mathbf{x}, k)$ 可以通过在边界上用 Green 函数 $G(\mathbf{x}, k)$ 与二次声源的驱动信号 $d(\mathbf{x}, k)$ 进行卷积来获得.

$$f_i(\mathbf{x}, k) = \oint_S d(\mathbf{x}', k) G(\mathbf{x} - \mathbf{x}', k) dS \quad (1)$$

其中, \mathbf{x} 为空间位置的矢量形式, $k = 2\pi f/c$ 为波数, S 表示区域边界.

在实际应用中,连续的二次声源无法实现,只能用离散的扬声器代替. 这等效于对驱动信号进行采样,当采样率不足时就会发生频域混叠. 已经证明在 3D 球形

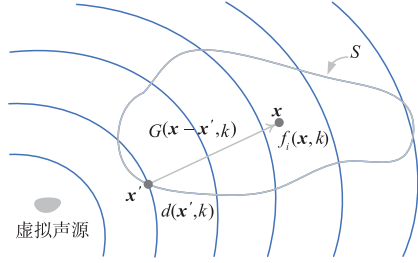


图1 使用简单声源合成法进行声场合成示意图

边界情况下,当波数为 k 且区域半径为 r 时, $L \geq (\lceil kr \rceil + 1)^2$ 可以实现无失真的声场重建. 同样,在 2.5D 圆形边界情况下, $L \geq (2\lceil kr \rceil + 1)$ 也可以实现无失真的声场重建,其中 L 为采样点数也就是扬声器的数目.

可以将连续二次声源的驱动信号 $d(\mathbf{x}, k)$ 视为圆形边界上的 2π 周期信号,因此 HOA 驱动信号的离散化可以表示为周期狄拉克采样函数

$$p(\phi, L) = \sum_{n=-\infty}^{\infty} \sum_{l=0}^{L-1} \delta(\phi - \alpha_l - 2\pi n) - \pi \leq \alpha_0 < \alpha_1 < \dots < \alpha_{L-1} < \pi \quad (2)$$

其中, α_l 是采样间隔, L 是每个周期的采样数. 对于均匀采样,采样间隔 $\alpha_l = 2\pi l/L$. 采样函数 $p(\phi, L)$ 的傅立叶变换可以表示为

$$P(\omega, L) = \sum_{n=-\infty}^{\infty} \delta(\omega - n) \underbrace{\sum_{l=0}^{L-1} e^{-jn\alpha_l}}_{b_n(L)} \quad (3)$$

它表明空间域的采样等效于频域中的频移和加权,其中 $b_n(L)$ 是采样函数的傅立叶系数. 连续二次声源的驱动信号可以表示为傅立叶级数

$$d(\phi, k) = \sum_{n=-N}^N a_n(k) e^{jn\phi} \quad (4)$$

这里 $d(\mathbf{x}, k)$ 用 $d(\phi, k)$ 表示,因为边界上的二次声源的位置 \mathbf{x} 仅由方位角 ϕ 确定. $d(\phi, k)$ 具有有限的带宽 $N = \lceil kr \rceil$, k 是波数, $a_n(k)$ 是与波数相关的傅立叶系数. 因此离散驱动信号 $d^s(\phi, k)$ 是式(4)连续驱动信号与式(2)采样函数的乘积

$$d^s(\phi, k) = p(\phi, L) d(\phi, k) \quad (5)$$

根据傅立叶变换的性质,空间域中的乘积对应于频域中的卷积,因此采样后驱动信号的傅立叶变换可以表示为

$$\begin{aligned} D^s(\omega, k) &= \frac{1}{2\pi} \left(\sum_{n=-\infty}^{\infty} \delta(\omega - n) \sum_{l=0}^{L-1} e^{-jn\alpha_l} \right) * \left(\sum_{n=-N}^N 2\pi \delta(\omega - n) a_n(k) \right) \\ &= \sum_{n=-\infty}^{\infty} \delta(\omega - n) \underbrace{\sum_{n'=-N}^N a_{n'}(k) b_{n-n'}(L)}_{a_n(k)} \end{aligned} \quad (6)$$

进一步,可以用矩阵乘法形式表示

$$\tilde{\mathbf{A}}_k = \mathbf{B}_L \mathbf{A}_k \quad (7)$$

其中, $\tilde{\mathbf{A}}_k = [\tilde{a}_{-N}(k), \tilde{a}_{-N+1}(k), \dots, \tilde{a}_N(k)]^T$ 和 $\mathbf{A}_k = [a_{-N}(k), a_{-N+1}(k), \dots, a_N(k)]^T$ 分别是连续驱动信号 $d(\phi, k)$ 的混叠傅立叶系数和原始傅立叶系数. $\mathbf{B}_L = [b_{m-n}(L); m, n = 0, 1, \dots, 2N]$ 是采样混叠矩阵,因为 $b_{m-n}^* = b_{n-m}$,它同时是一个 Hermitian Toeplitz 矩阵,如下所示

$$\mathbf{B}_L = \begin{bmatrix} b_{0-0}(L) & b_{0-1}(L) & \dots & b_{0-2N}(L) \\ b_{1-0}(L) & b_{1-1}(L) & \dots & b_{1-2N}(L) \\ \vdots & \dots & b_{m-n}(L) & \vdots \\ b_{2N-0}(L) & b_{2N-1}(L) & \dots & b_{2N-2N}(L) \end{bmatrix} \quad (8)$$

其中, $b_{m-n}(L) = \sum_{l=0}^{L-1} e^{-j(m-n)\alpha_l}$. 傅立叶系数可以由混叠 $\tilde{\mathbf{A}}_k$ 复原得到

$$\hat{\mathbf{A}}_k = \mathbf{B}_L^+ \tilde{\mathbf{A}}_k = \mathbf{B}_L^+ \mathbf{B}_L \mathbf{A}_k \quad (9)$$

其中, $\hat{\mathbf{A}}_k$ 是重建的傅立叶系数, \mathbf{B}_L^+ 是 \mathbf{B}_L 的 Moore-Penrose 逆.

对于具有有限带宽的 2.5D HOA 信号,如果采样速率超过奈奎斯特率 ($L \geq 2N + 1$),可以保证无失真重建. 在这种情况下,均匀采样方法的混叠矩阵 \mathbf{B}_L 是单位矩阵;在非均匀采样情况下, \mathbf{B}_L 通常不是单位矩阵. 但根据非均匀采样理论^[30,31], \mathbf{B}_L 仍然是可逆的. 因此,均匀采样和非均匀采样都可以无失真地重建原始声场. 同时,边界上二次声源的驱动信号的归一化重建误差可以表示为

$$\varepsilon = \frac{|\mathbf{A}_k - \hat{\mathbf{A}}_k|^2}{|\mathbf{A}_k|^2} = \frac{\mathbf{A}_k^* (\mathbf{I} - \mathbf{B}_L^+ \mathbf{B}_L)^2 \mathbf{A}_k}{\mathbf{A}_k^* \mathbf{A}_k} \quad (10)$$

其中, $|\cdot|$ 表示取模, $(\cdot)^*$ 表示共轭转置,归一化重建误差的倒数即是信噪比 (Signal to Noise Ratio, SNR).

2.2 低于奈奎斯特率时的优化

本节讨论如果采样点数也即扬声器数目小于奈奎斯特率时如何选择样本. 首先,可以证明当使用奈奎斯特率采样时, L 个均匀分布的二次声源 $\{\alpha_l\}_{l=0}^{L-1}$ 的归一化驱动信号 $\{\mathbf{A}_k^l\}_{l=0}^{L-1}$ 构成矢量空间 \mathbf{A}_k 的正交基.

当以 $L = 2N + 1$ 均匀采样时, $\mathbf{A}_k^* (\mathbf{I} - \mathbf{B}_L^+ \mathbf{B}_L)^2 \mathbf{A}_k = 0$ 对于任何驱动信号 \mathbf{A}_k 都成立. 这意味着所有 \mathbf{A}_k 都包含在矩阵 $(\mathbf{I} - \mathbf{B}_L^+ \mathbf{B}_L)$ 的零空间中. 由于 \mathbf{A}_k^l 也在 $(\mathbf{I} - \mathbf{B}_L^+ \mathbf{B}_L)$ 的零空间中,我们只需要证明 \mathbf{A}_k^l 彼此正交即可. 点声源 $S(r_s, \phi_s, k)$ 的 2.5D 傅立叶表示如下所示,其中 $h_n^{(2)}(\cdot)$ 是第二类球面汉克尔函数^[32]:

$$d_{ps,2.5D}(\phi, k) = \frac{1}{2\pi r_0} \sum_{n=-\infty}^{\infty} \frac{h_{|n|}^{(2)}(kr_s) e^{-jn\phi}}{h_{|n|}^{(2)}(kr_0)} e^{jn\phi} \quad (11)$$

其中,第 l 个二次声源位于 $r_s = r_0$ 和 $\phi_s = \alpha_l$,对于 \mathbf{A}_k^l ,我

们可以得到 $a_n(k) = 1/\sqrt{2N+1}e^{-jn\alpha}$. 因此对于不同的二次声源,我们有

$$\begin{aligned}\langle \mathbf{A}_k^l, \mathbf{A}_k^{l'} \rangle &= \frac{1}{2N+1} \sum_{n=-N}^N e^{-jn(\alpha_l - \alpha_{l'})} \\ &= \frac{1}{2N+1} \sum_{n=-N}^N e^{-jn\frac{2\pi(l-l')}{2N+1}} \\ &= \delta_{ll'}\end{aligned}\quad (12)$$

这证明了 \mathbf{A}_k^l 是相互正交的,并且构成矢量空间 \mathbf{A}_k 的正交基. 因此,低于奈奎斯特速率的采样等同于保留 $\{\mathbf{A}_k^l\}_{l=0}^{L-1}$ 的一部分正交基. 为了确保声场误差最小,应保留具有最大内积 $|\langle \mathbf{A}_k, \mathbf{A}_k^l \rangle|^2$ 的 \mathbf{A}_k^l . 对于点声源 $\mathbf{x} = (r_s, \phi_s)$,我们有

$$\langle \mathbf{A}_k, \mathbf{A}_k^l \rangle = \sum_{n=-\infty}^{\infty} \frac{h_{|n|}^{(2)}(kr_s) e^{-jn\phi_s}}{h_{|n|}^{(2)}(kr_0)} e^{jn\alpha_l}\quad (13)$$

这个函数关系十分复杂,为了继续分析每个 \mathbf{A}_k^l 分量上驱动信号的能量分布,需要使用球面汉克尔函数的渐近性质^[33]:

$$h_n^{(2)}(kr) \sim \begin{cases} j \frac{(2n-1)!!}{(kr)^{2n+1}}, & kr \rightarrow 0 \\ (j)^{n+1} \frac{e^{jkr}}{kr}, & kr \rightarrow \infty \end{cases}\quad (14)$$

从附录中可看出,当 α_l 远离 ϕ_s 时, $|\langle \mathbf{A}_k, \mathbf{A}_k^l \rangle|^2$ 减小. 因此,当 α_l 接近 ϕ_s 时, \mathbf{A}_k^l 保留了更多的驱动信号能量. 总之,如已知声源的方向,则应在声源的方向附近增加二次声源数目. 如图 2 所示,非均匀采样方法在低于奈奎斯特采样率的情况下采样误差比均匀误差更小.

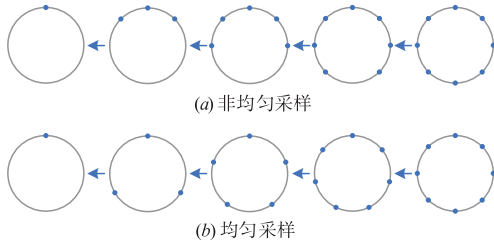


图2 非均匀采样和均匀采样示意图

2.3 非均匀采样方案

受以上讨论的启发,在高频声场的采样率小于奈奎斯特率($L < 2N+1$)的情况下,本节提出了一种双采样率的非均匀采样方法,它增加了靠近声源方向的采样点,以减少驱动信号的混叠误差,如下所示

$$\alpha_l = \begin{cases} \phi_1 + \frac{(\phi_2 - \phi_1)l}{L_1}, & 0 \leq l < L_1 \\ \phi_2 + \frac{(2\pi - \phi_2 + \phi_1)(l - L_1)}{L_2}, & L_1 \leq l < L_1 + L_2 \end{cases}\quad (15)$$

其中, $L_1 + L_2 = L$, 方位角 $\phi_1, \phi_2 \in [-\pi, \pi]$ 且 $\phi_1 < \phi_2$,

它们由 3D 屏幕区域或感兴趣区域确定, r_0 是二次声源半径, r_s 和 ϕ_s 分别是声源距离和方位角,如图 3 所示.

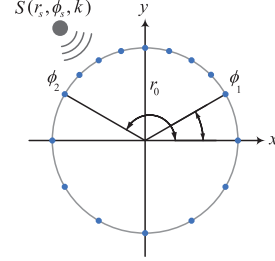


图3 双采样率的非均匀采样方案示意图

为了进一步显示均匀采样与非均匀采样之间的差异,针对所有 $\phi_s \in [-\pi, \pi]$ 方位的点声源我们通过式(10)计算其归一化重建误差. 将式(15)引入式(3)可以得到混叠矩阵 \mathbf{B}_L , 结合式(11)中的点声源傅立叶系数 \mathbf{A}_k , 使用式(10)可以计算出不同方位角的重建误差. 结果如图 4 上图所示, 设置为 ($f = 2000\text{Hz}$, $r_0 = 1.5\text{m}$, $r_s = 2.0\text{m}$, $\phi_1 = -\pi/2$, $\phi_2 = \pi/2$, $N = 33$), 蓝线为均匀采样. 图 4 下图设置为 ($N = 33$, $L = 48$). x 坐标是声源方位角, y 坐标是边界上驱动信号的信噪比. 当 $\phi_s \in [\phi_1, \phi_2]$ 时, 增加 ϕ_1 和 ϕ_2 之间的采样率可以大大减小驱动信号的混叠误差. 具体来说, 尽管只使用了 48 个非均匀样本, 但仍可以获得与 54 个均匀样本相当的信噪比. 但是, 还应注意当声源 $\phi_s \notin [\phi_1, \phi_2]$ 时, 信噪比下降到 42 个均匀采样的水平. 因此, 非均匀采样以这种方式可以在特定方向上优化声场.

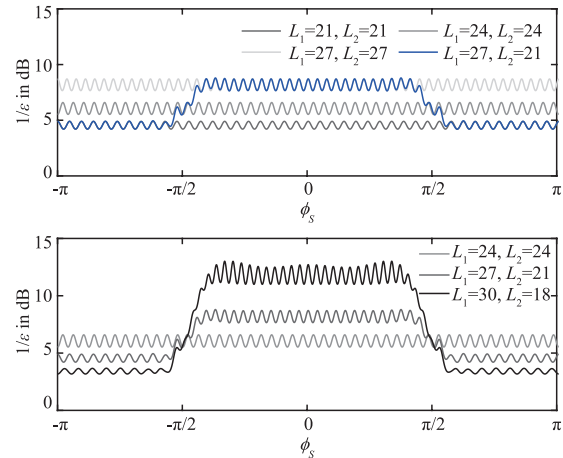


图4 采样信号的信噪比示意图

只要保证 $L_1 + L_2 = L$ 的 L_1 和 L_2 都是可选的, 因此会有很多 L_1 和 L_2 的数值组合. 如上所述, 感兴趣区域的 L_1 增大会导致外部质量下降, 见图 4 下图. 因此整个非均匀采样方案的最大误差应尽可能小, 这可以避免感兴趣区域之外的声音质量过度降低. 由于 $(\mathbf{I} - \mathbf{B}_L^* \mathbf{B}_L)^2$ 也是 Hermitian 矩阵, 因此可以使用 Rayleigh 熵来估计最大误差, 该误差表示为

$$\frac{\mathbf{A}_k^* (\mathbf{I} - \mathbf{B}_L^* \mathbf{B}_L)^2 \mathbf{A}_k}{\mathbf{A}_k^* \mathbf{A}_k} \leq \lambda_{\max} \quad (16)$$

其中, λ_{\max} 是 $(\mathbf{I} - \mathbf{B}_L^* \mathbf{B}_L)^2$ 的最大特征值^[34]. 所以, 我们应该在众多可选择的 L_1 和 L_2 中, 选择具有较小 λ_{\max} 的非均匀采样方案.

2.4 非均匀重构方案

当重构声场时, 需要根据声源的方位角 ϕ_s 设置驱动信号 $d^s(\phi, k)$ 的最大频率 N^s . 如果虚拟音源位于 ϕ_1 和 ϕ_2 之间, N^s 就由 L_1 的采样率决定, 就如同 L_1 对整个边界采样一样; 否则, N^s 就由 L_2 的采样率决定. 可以从奈奎斯特关系式推导出这个最大频率, 其表达式为

$$N^s = \begin{cases} \left\lfloor \frac{L_1 \pi}{\phi_2 - \phi_1} - \frac{1}{2} \right\rfloor, & \phi_s \in [\phi_1, \phi_2) \\ \left\lfloor \frac{L_2 \pi}{2\pi - \phi_2 + \phi_1} - \frac{1}{2} \right\rfloor, & \phi_s \in [\phi_2, 2\pi + \phi_1) \end{cases} \quad (17)$$

二次声源的驱动信号可以根据下式得到

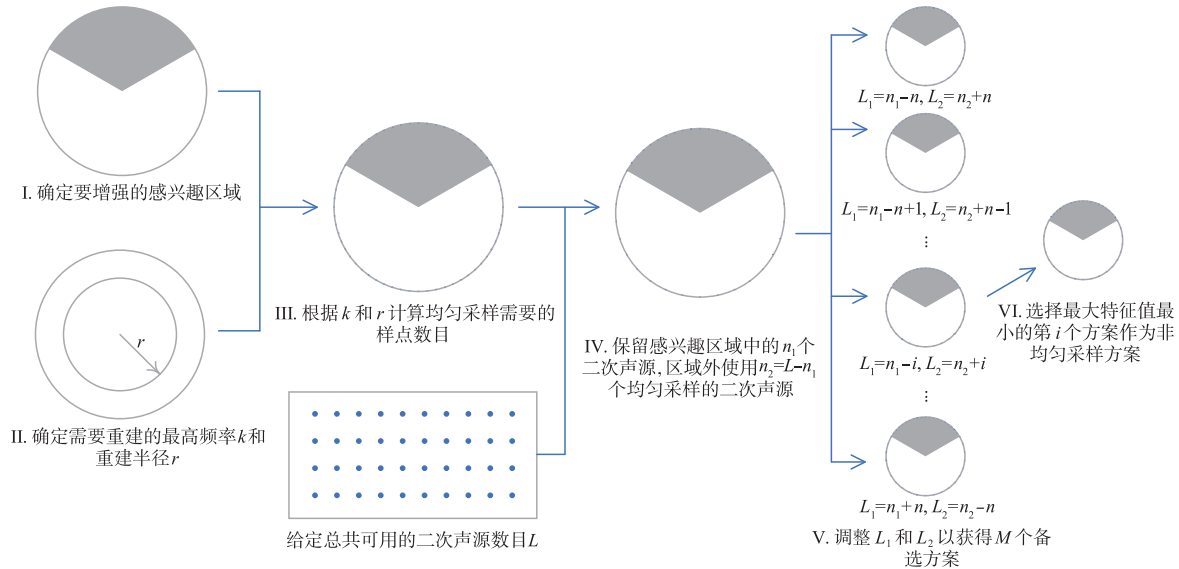


图5 非均匀二次声源分布设计过程示意图

3 结果与讨论

为了验证本文提出的非均匀二次声源分布算法, 本节测试了不同 L_1 和 L_2 配置下的点源合成. 比较算法是均匀采样的最佳模态带宽 (Optimal Modal Bandwidth)^[23]. 其中驱动信号的带宽 $N = \lfloor (L - 1)/2 \rfloor$, 详细信息如表 1 所示. 由于合成声场的无混叠半径只由最高频率决定, 低于最高频率的信号不会影响无混叠半径, 因此这里采用声场合成仿真中对比最高单频信号的方法进行实验. 最高频率采用 $f = 2000\text{Hz}$, 声源方位角为 $\phi_s = \pi/2$, 声源距离为 $r_s = 3\text{m}$. 扬声器阵列半径为

$$d^s(\phi, k) = \sum_{n=-N^s}^{N^s} a_n(k) e^{jn\phi} \quad (18)$$

最后, 非均匀二次声源分布的整个设计过程如下 (见图 5):

步骤 1 确定由 (ϕ_1, ϕ_2) 表示的感兴趣区域. 这个区域不能太大或太小, 否则与均匀采样没有区别.

步骤 2 确定应用所需的声场最大波数 k 和无混叠半径 r .

步骤 3 通过 $n = (2\lceil kr \rceil + 1)$ 来计算均匀分布所需的最小二次声源数目.

步骤 4 给出总共可用的二次声源数目 L 并保留 (ϕ_1, ϕ_2) 中的 n_1 个二次声源, 剩下的 $n_2 = L - n_1$ 个二次声源在 (ϕ_1, ϕ_2) 外部均匀分布.

步骤 5 对 n_1 和 n_2 的数目作微调从而得到 M 个替代方案.

步骤 6 选择一个非均匀分配方案并且它的 $(\mathbf{I} - \mathbf{B}_L^* \mathbf{B}_L)^2$ 具有较小的 λ_{\max} .

$r_0 = 1.5\text{m}$, 所需的无混叠半径为 $r = 0.75\text{m}$, 二次声源的总数为 $L = 48$, 感兴趣的区域是由方位角 $\phi_1 = 0$ 和 $\phi_2 = \pi$ 决定的前半圆区域. 本节使用重建声场混叠信号比 (Reproduced Aliasing to Signal Ratio, RASR) 来表示声场混叠失真的强度, 其定义为频率混叠引起的声场失真和原始声场^[35] 的能量比. $f(\mathbf{x}, k)$ 表示原始声场, $f^s(\mathbf{x}, k)$ 是由离散的二次声源使用驱动信号 $d^s(\phi, k)$ 重构的声场.

$$\text{RASR}(\mathbf{x}, k) = 10 \log_{10} \frac{|f(\mathbf{x}, k) - f^s(\mathbf{x}, k)|^2}{|f(\mathbf{x}, k)|^2} \quad (19)$$

表 1 实验配置

音频	$f = 2000\text{Hz}$
最大波数	$k = 36.6$
阵列半径	$r_0 = 1.5\text{m}$
声源方位角	$\phi_s \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$
声源半径	$r_s = 3.0\text{m}$
感兴趣区域	$(\phi_1, \phi_2) = (0, \pi)$
所需的混叠半径	$r = 0.75\text{m}$
所需的样点数	$n = 58$
(ϕ_1, ϕ_2) 中的保留点数	$n_1 = 29$
比较算法	最佳模态带宽 ($L_1 = 24, L_2 = 24$)
备选方案 1	$(L_1 = 29, L_2 = 19)$
备选方案 2	$(L_1 = 33, L_2 = 15)$

3.1 合成声场评估

图 6 给出了合成声场的波形和相应的 RASR. 从图 6(c) 中可以看出, 均匀分布的二次声源 ($L_1 = 24, L_2 = 24$) 产生了一个 $r = 0.62\text{m}$ 的无混叠区域. 实验设置为 ($f = 2000\text{Hz}, r_0 = 1.5\text{m}, r_s = 3.0\text{m}, \phi_1 = 0, \phi_2 = \pi, N =$

$55, L = 48$), 图 6(c) 是均匀采样的合成声场, 图 6(d) 和 (b) 是非均匀采样 ($L_1 = 29, L_2 = 19$) 的合成声场, 其中图 6(d) 显示音源 $\phi_s = \pi/2$, 图 6(b) 显示音源 $\phi_s = -\pi/2$ (相当于交换 L_1 和 L_2). 图 6(e) 和 (a) 是非均匀采样 ($L_1 = 33, L_2 = 15$) 的合成声场, 其中图 6(e) 显示音源 $\phi_s = \pi/2$, 图 6(a) 显示音源 $\phi_s = -\pi/2$. 图 6(f) ~ (j) 是相应的 RASR. 在图 6(d) 中, 当二次声源增加到 ($L_1 = 29, L_2 = 19$) 后, 无混叠区域半径增加到 $r = 0.70\text{m}$. 这意味着 L_1 值增加 20.8%, 则 r 值增加 12.9%. 图 6(e) 中, 当二次声源增加到 ($L_1 = 33, L_2 = 15$) 时, 无混叠区域半径变为 $r = 0.76\text{m}$, 分别增加了 37.5% 和 22.6% 可见, 增长率与奈奎斯特关系 $L = 2\lceil kr \rceil + 1$ 相当.

图 6(a) 和图 6(b) 展示了感兴趣区域外的声源. 为了进行直接比较, 保持声源方位角为 $\phi_s = \pi/2$, 而交换 L_1 和 L_2 的值, 这等效于声源来自方位角 $\phi_s = -\pi/2$. 在图 6(b) 中, 当二次声源减少到 ($L_1 = 19, L_2 = 29$) 时, 无混叠区域的半径减小到 $r = 0.48$. 进一步将二次声源减少到 ($L_1 = 15, L_2 = 33$) 后, 无混叠区域的半径也随之减小到 $r = 0.37\text{m}$, 所以降低率也满足奈奎斯特关系.

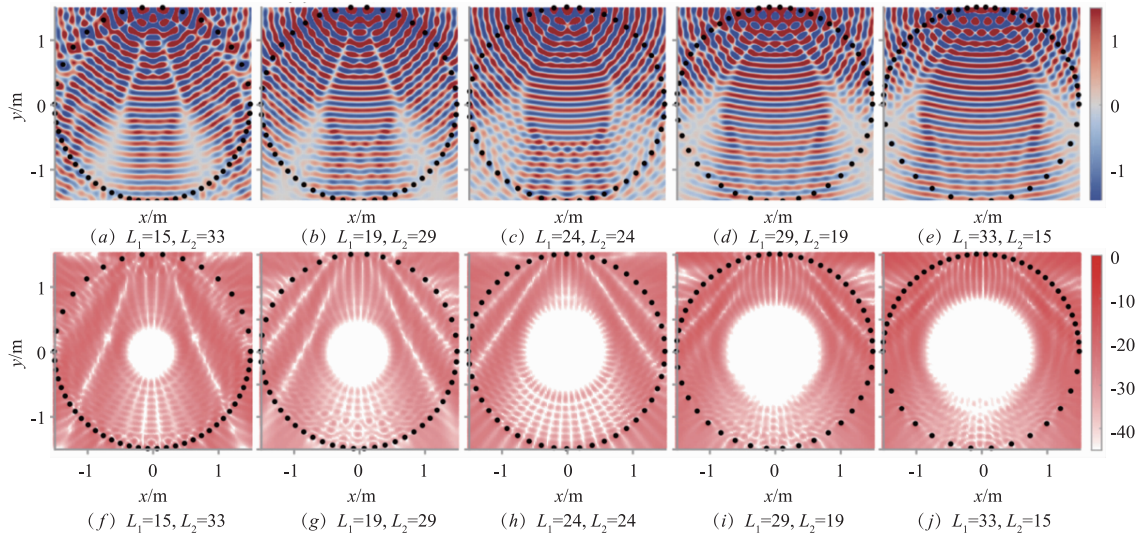


图 6 合成声场评价结果示意图

对于二次声源 ($L_1 = 29, L_2 = 19$) 和 ($L_1 = 33, L_2 = 15$), 最大特征值分别为 $\lambda_{\max} = 2.05$ 和 $\lambda_{\max} = 3.07$. 根据式 (16), $L_1 = 29$ 的最大采样误差小于 $L_1 = 33$, 这与图 6(a) 和 (b) 的结果一致. 当声源来自最差方位角的情况时, 二次声源排布方式 ($L_1 = 29, L_2 = 19$) 具有较大的无混叠区域. 在此情况下, 我们应该选择 ($L_1 = 29, L_2 = 19$). 值得注意的是, ($L_1 = 29, L_2 = 19$) 方案产生的无混叠半径为 0.70m 而不是 0.75m , 这是因为使用的二次声源数量 $L = 48$ 远小于无失真所需的 $n = 58$.

3.2 平均 RASR 评估

表 2 依次给出了水平方位为 $\phi_s = \{\pi/2, \pi/3, \pi/6, 0, -\pi/6, -\pi/3, -\pi/2\}$ 声源的声场合成结果, 其中 r 是无混叠半径, RASR 是阵列内部区域的平均 RASR. 从表中我们可以看出, 与统一采样相比, ($L_1 = 29, L_2 = 19$) 和 ($L_1 = 33, L_2 = 15$) 的前方声源 RASR 都降低了. 此外, ($L_1 = 33, L_2 = 15$) 比 ($L_1 = 29, L_2 = 19$) 的 $[\phi_1, \phi_2]$ 中增加的样本更多并且其平均 RASR 更低. 当声源从 $\phi_s = \pi/2$ 变为 $\phi_s = -\pi/2$ 时, ($L_1 = 33, L_2 = 15$) 的 RASR 性能下降的幅度大于 ($L_1 = 29, L_2 =$

19). 这是因为 ($L_1 = 33, L_2 = 15$) 在 $[\phi_1, \phi_2]$ 之外具有更少的二次声源. 最后, 均匀采样在所有方向上具有几乎相同的 RASR.

可见, 非均匀二次声源分布增强了区域 $[\phi_1, \phi_2]$ 中声源的合成声场质量, 但代价是 $[\phi_1, \phi_2]$ 区域以外声源的合成声场质量会降低. 广义上来说, 圆形排布的 HOA

和 WFS 可以看做两种极端的排布, HOA 是完全均匀采样, WFS 是只在前半圆均匀采样而后半圆完全不采样. 因此在采样点数相同时, WFS 对于前方而来的音源, 能够获得两倍于 HOA 的无混叠重建区域; 但对于后方而来的音源, 则完全丢失了无混叠区域, 这与实验得到的结果是一致的.

表 2 不同方位源的平均 RASR 和无混叠半径

ϕ_s		$\pi/2$	$\pi/3$	$\pi/6$	0	$-\pi/6$	$-\pi/3$	$-\pi/2$
$L_1 = 33$	r/m	0.76	0.72	0.69	0.45	0.43	0.39	0.37
	RASR/dB	-30.44	-29.39	-28.57	-8.45	-7.72	-6.86	-6.51
$L_1 = 29$	r/m	0.70	0.68	0.66	0.53	0.50	0.48	0.48
	RASR/dB	-26.25	-25.63	-25.20	-11.86	-11.58	-11.31	-11.23
$L_1 = 24$	r/m	0.62	0.62	0.62	0.62	0.62	0.62	0.62
	RASR/dB	-19.02	-19.03	-19.03	-19.03	-19.03	-19.03	-19.01

4 结论

在声场合成中, 如何在扬声器数量受限的情况下扩大无混叠听音区域是一个重要问题. 在实际应用中, 声源可能主要是来自某个区域, 例如视频屏幕区域. 在这种情况下, 牺牲区域外音源的声场合成质量以改善感兴趣区域声源的声场质量是一个可行的思路, 非均匀二次声源排布是实现此目标的一种有效方法. 本文研究表明可以将重建声场的边界划分为多个子区域, 并对不同的子区域使用不同的采样率, 增加子区域的采样率将减少该区域声源的采样混叠. 本文提出了一种简单的方法, 将 2.5D 圆形边界的二次声源分为两个子区域, 并讨论了恢复混叠频率和生成驱动信号的方法. 同时建议利用混叠矩阵的 Hermitian 特性, 使用 Rayleigh 熵来估计非均匀采样的最大混叠误差. 实验结果表明, 与均匀采样相比所提方法可用于增强特定方向的声场合成质量, 尤其是在已知大多数声源来自某个感兴趣区域的情况下.

5 附录

因为虚拟点声源在扬声器阵列之外, 这里用 $q = r_0/r_s$ 表示扬声器阵列半径和音源距离的比值, 且有 $q < 1$. 当 kr 较小时, 内积可计算为

$$\begin{aligned}
 \langle \mathbf{A}_k, \mathbf{A}_k^t \rangle &= \sum_{n=-\infty}^{\infty} \frac{j(2|n|-1)!! / (kr_s)^{2|n|+1}}{j(2|n|-1)!! / (kr_0)^{2|n|+1}} e^{-jn\phi_s} e^{jn\alpha_t} \\
 &= \sum_{n=-\infty}^{\infty} q^{2|n|+1} e^{-jn\phi_s} e^{jn\alpha_t} \\
 &= q \sum_{n=0}^{\infty} [q^2 e^{-j(\phi_s - \alpha_t)}]^n + q \sum_{n=0}^{\infty} [q^2 e^{j(\phi_s - \alpha_t)}]^n - q \\
 &= \frac{q(1-q^4)}{(1-2q^2 \cos(\phi_s - \alpha_t) + q^4)} \quad (20)
 \end{aligned}$$

因此, 位于 α_t 的点声源驱动信号能量的偏导可以表示为

$$\begin{aligned}
 \frac{\partial}{\partial \alpha_t} (|\langle \mathbf{A}_k, \mathbf{A}_k^t \rangle|^2) &= \frac{\partial}{\partial \alpha_t} \left(\frac{q^2 (1-q^4)^2}{(1-2q^2 \cos(\phi_s - \alpha_t) + q^4)^2} \right) \\
 &= \frac{4q^4 (1-q^4)^2 \sin(\phi_s - \alpha_t)}{(1-2q^2 \cos(\phi_s - \alpha_t) + q^4)^3} \quad (21)
 \end{aligned}$$

其中, $[1-2q^2 \cos(\phi_s - \alpha_t) + q^4] > 0$. $\phi_s \in [0, 2\pi)$ 时, 只有两个零解 $\phi_s = \alpha_t$ 和 $\phi_s = \alpha_t + \pi$. 当 $\phi_s > \alpha_t$ 时, $|\langle \mathbf{A}_k, \mathbf{A}_k^t \rangle|^2$ 单调增加; 当 $\phi_s < \alpha_t$ 时, $|\langle \mathbf{A}_k, \mathbf{A}_k^t \rangle|^2$ 单调减少.

当 kr 较大时, 内积可以通过以下渐进公式计算

$$\begin{aligned}
 \langle \mathbf{A}_k, \mathbf{A}_k^t \rangle &= \sum_{n=-\infty}^{\infty} \frac{(j)^{|n|+1} e^{jkr_s} / kr_s}{(j)^{|n|+1} e^{jkr_0} / kr_0} e^{-jn\phi_s} e^{jn\alpha_t} \\
 &= q e^{jk(r_s - r_0)} \sum_{n=-\infty}^{\infty} e^{-jn\phi_s} e^{jn\alpha_t} \\
 &= q e^{jk(r_s - r_0)} \delta(\phi_s - \alpha_t) \quad (22)
 \end{aligned}$$

当 $\phi_s \neq \alpha_t$ 时 $|\langle \mathbf{A}_k, \mathbf{A}_k^t \rangle|^2 = 0$.

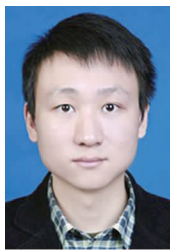
参考文献

- [1] Kim S, Lee Y W, Pulkki V. New 10.2-channel vertical surround system (10.2-VSS); Comparison study of perceived audio quality in various multichannel sound systems with height loudspeakers [A]. Audio Engineering Society Convention 129 [C]. New York, USA: Audio Engineering Society, 2010. Paper Number 8296.
- [2] Ando A. Conversion of multichannel sound signal maintaining physical properties of sound in reproduced sound field [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 19(6): 1467-1475.
- [3] Herre J, Hilpert J, Kuntz A, et al. MPEG-H audio: the new standard for universal spatial/3D audio coding [J]. Journal

- of the Audio Engineering Society, 2014, 62 (12): 821 – 830.
- [4] Politis A, Vilkamo J, Pulkki V, et al. Sector-based parametric sound field reproduction in the spherical harmonic domain[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2015, 9(5): 852 – 866.
- [5] 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.
- [6] Poletti M A. Three-dimensional surround sound systems based on spherical harmonics[J]. *Journal of the Audio Engineering Society*, 2005, 53(11): 1004 – 1025.
- [7] Gupta A, Abhayapala T D. Three-dimensional sound field reproduction using multiple circular loudspeaker arrays [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, 19(5): 1149 – 1159.
- [8] Ahrens J, Spors S. A modal analysis of spatial discretization of spherical loudspeaker distributions used for sound field synthesis [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(9): 2564 – 2574.
- [9] Zhang W, Abhayapala T D. Three dimensional sound field reproduction using multiple circular loudspeaker arrays: functional analysis guided approach[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2014, 22(7): 1184 – 1194.
- [10] Winter F, Ahrens J, Spors S, et al. On analytic methods for 2.5D local sound field synthesis using circular distributions of secondary sources[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2016, 24(5): 914 – 926.
- [11] Koyama S, Furuya K, Hiwasaki Y, et al. Wave field reconstruction filtering in cylindrical harmonic domain for with-height recording and reproduction[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2014, 22(10): 1546 – 1557.
- [12] Firtha G, Fiala P, Schultz F, et al. Improved referencing schemes for 2.5D wave field synthesis driving functions [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2017, 25(5): 1117 – 1127.
- [13] Firtha G, Fiala P, Schultz F, et al. On the general relation of wave field synthesis and spectral division method for linear arrays [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2018, 26(12): 2393 – 2403.
- [14] Winter F, Wierstorf H, Hold C, et al. Colouration in local wave field synthesis [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2018, 26(10): 1913 – 1924.
- [15] Hahn N, Winter F, Spors S, et al. Synthesis of a spatially band-limited plane wave in the time-domain using wave field synthesis[A]. *The 25th European Signal Processing Conference (EUSIPCO)* [C]. Kos, Greece: IEEE, 2017. 673 – 677.
- [16] 周岭松, 鲍长春, 贾懋坤, 步兵. 基于高阶 Ambisonics 的 2.5 维近场声源合成[J]. *电子学报*, 2016, 45(3): 520 – 526.
ZHOU Ling-song, BAO Chang-chun, JIA Mao-shen, BU Bing. 2.5D near-field sound source synthesis using higher order Ambisonics [J]. *Acta Electronica Sinica*, 2016, 45(3): 520 – 526. (in Chinese)
- [17] Jia M, Wang W, Sun J, et al. Sound field reproduction of focused virtual sources with symmetry sigmoid regularization function[J]. *Chinese Journal of Electronics*, 2018, 27(4): 835 – 844.
- [18] 李娟, 李军锋, 颜永红. 波场合成中声像感知距离重建 [J]. *声学学报*, 2013, 38(06): 743 – 748.
LI Juan, LI Jun-feng, YAN Yong-hong. Synthesis of perceived distance based on wave field synthesis [J]. *Acta Acustica*, 2013, 38(06): 743 – 748. (in Chinese)
- [19] 李娟, 付强, 颜永红. 波场合成与波场分析的有源房间补偿方法[J]. *声学学报*, 2014, 39(1): 137 – 144.
LI Juan, FU Qiang, YAN Yong-hong. Active listening room compensation based on wave field synthesis and wave field analysis [J]. *Acta Acustica*, 2014, 39(1): 137 – 144. (in Chinese)
- [20] Kleijn W B, Allen A, Skoglund J, et al. Incoherent idempotent ambisonics rendering [A]. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* [C]. New Paltz, NY, USA: IEEE, 2017. 209 – 213.
- [21] Kleijn W B. Directional emphasis in ambisonics [J]. *IEEE Signal Processing Letters*, 2018, 25(7): 1079 – 1083.
- [22] Okamoto T. 2.5D higher order ambisonics for a sound field described by angular spectrum coefficients [A]. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* [C]. Shanghai, China: IEEE, 2016. 326 – 330.
- [23] Winter F, Schultz F, Firtha G, et al. A geometric model for prediction of spatial aliasing in 2.5D sound field synthesis [J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019, 27(6): 1031 – 1046.
- [24] Kentgens M, Jax P. Space warping based dimensionality reduction of higher order ambisonics signals [A]. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* [C]. Brighton, United Kingdom, United Kingdom: IEEE, 2019. 131 – 135.
- [25] Mai H, Xie B, Jiang J. Influence of the number of loudspeakers on the timbre in mixed-order ambisonics re-

- production[A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C]. Calgary, AB, Canada: IEEE, 2018. 556 – 560.
- [26] Hold C, Gamper H, Pulkki V, et al. Improving binaural ambisonics decoding by spherical harmonics domain tapering and coloration compensation[A]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C]. Brighton, United Kingdom: IEEE, 2019. 261 – 265.
- [27] Menzies D, Fazi F M. Ambisonic decoding for compensated amplitude panning [J]. IEEE Signal Processing Letters, 2019, 26(3): 470 – 474.
- [28] Okamoto T, Cui Z L, Iwaya Y, et al. Implementation of a high-definition 3D audio-visual display based on higher-order ambisonics using a 157-loudspeaker array combined with a 3D projection display[A]. The 2nd IEEE International Conference on Network Infrastructure and Digital Content[C]. Beijing, China: IEEE, 2010. 179 – 183.
- [29] André C R, Corteel E, Embrechts J J, et al. Subjective evaluation of the audiovisual spatial congruence in the case of stereoscopic-3D video and wave field synthesis[J]. International Journal of Human-Computer Studies, 2014, 72(1): 23 – 32.
- [30] Yen J. On nonuniform sampling of bandwidth-limited signals[J]. IRE Transactions on Circuit Theory, 1956, 3(4): 251 – 257.
- [31] Margolis E, Eldar Y C. Nonuniform sampling of periodic bandlimited signals[J]. IEEE Transactions on Signal Processing, 2008, 56(7): 2728 – 2745.
- [32] Ahrens J. Analytic Methods of Sound Field Synthesis[M]. Berlin, Germany: Springer, 2012.
- [33] Earl G Williams. Fourier Acoustics[M]. USA: Academic Press, 1999. 183 – 234.
- [34] Horn R A, Johnson C R. Matrix Analysis[M]. Cambridge, United Kingdom: Cambridge University Press, 2012.
- [35] Rabenstein R, Spors S. Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis[A]. Proceedings of the Audio Engineering Society[C]. USA: AES, 2006. No. 6711.

作者简介



董石 男, 1986 年 12 月出生, 湖北武汉人. 2008 年和 2014 年分别在华中科技大学、武汉大学获工学学士和工学博士学位. 现为华中师范大学教育信息技术学院讲师, 硕士生导师. 主要研究方向为音频信号处理、三维声场技术、信号压缩技术.

E-mail: dongshi@mail.ccnu.edu.cn



章强 男, 1996 年 9 月出生, 安徽铜陵人. 现为华中师范大学教育信息技术学院硕士研究生. 主要研究方向为语音合成、语音信号处理.

E-mail: 751938379@qq.com