空间音频线索的分析探索：可微多球体散射模型

Siminfar Samakoush Galougah、PranavPulijala、Ramani Duraiswami*马里兰大学感知智能与现实实验室，美国马里兰州学院公园

{simin95, pulijala, ramanid}@umd.edu

arXiv:2603.02205v1 [cs.SD] 2 Mar 2026

摘要

开发合成空间听觉系统面临的一项主要挑战，尤其是在水下环境中，是如何准确地对声散射建模。生物体通过利用声音在自身身体上的散射，产生随位置变化的双耳声级差与时间差（ITD/ILD），从而实现三维空间听觉。基于刚性散射的头相关传递函数（HRTF）模型足以描述陆地人类的听觉，但由于水与软组织之间近乎阻抗匹配，因此无法适用于水下环境。受水下动物声学解剖结构的启发，我们提出一种新颖的解析闭式前向模型，用于描述声波从一个包含两个刚性球形散射体的半透明球体上散射的过程。该模型能够准确地将声源方向、频率和材料属性映射到声压场，捕捉分层可穿透结构中的复杂物理现象。尤为重要的是，我们在完全可微的框架中实现了该模型，使其能够与机器学习算法集成，以优化主动定位的代价函数。我们展示了采用物理信息频率加权方案可提升噪声环境下定位的收敛性，并通过扩展卡尔曼滤波器（EKF）和解析计算的雅可比矩阵，实现准确的运动声源跟踪。我们的研究表明，对分层刚性与透明几何结构中的散射进行可微建模，有望为麦克风阵列提供新的基础，使其能够利用基于散射的空间线索，而非传统波束形成，并适用于陆地和水下应用。我们的模型将以开源形式发布。

关键词： 声学散射、可微仿真、多极展开、声源定位、水下声学

1 引言

生物体仅凭两个传感器，便能在复杂环境中进行听觉场景分析。即使存在噪声、干扰声和混响，它们也能实时解析分布在三维空间中的宽带声源方向。实现这一能力的一个基本机制，是入射声场在生物体自身身体上的散射；这一过程会以依赖于位置和频率的方式改变到达双耳的信号。由此产生的双耳声级差与时间差（ILD/ITD）构成空间听觉的主要双耳线索。这些线索与相互连接的神经计算相结合，用于解决声学场景分析中的各种问题，例如定位和跟踪。

在陆生哺乳动物中，身体大部分不透声，空间线索由声波绕过头部时产生的衍射，以及外耳（即耳廓）复杂的几何结构共同塑造。头相关传递函数（HRTF）能够体现这些散射过程的综合作用，其作用已有数十年的研究历史 [1]。研究人员已针对人类听者测量、计算并近似估算头部相关传递函数，并建立了多个数据库。简化的几何模型已

*计算资源由马里兰大学的 Nexus 高性能计算集群提供。我们感谢南丹麦大学（SDU，丹麦）的 Magnus Wahlberg 教授就水下空间听觉提供讨论和背景信息。本研究由美国海军研究办公室通过项目编号 N00014-23-1-2086 资助。数据和代码可按需提供。

在理解基于HRTF的感知背后的物理机制方面发挥了重要作用。刚性球体模型 [2] 描述了代表人头的单个球体周围的衍射，而“雪人”式头部与躯干模型则 [3] 利用多极展开与平移，将这一模型扩展为两个刚性球，并加入躯干反射和仰角线索 [3, 4, 5]。这些模型表明，即使简单的几何结构也能产生丰富的空间信息。然而，在人工系统中将其用于仿生实时定位或跟踪，仍处于起步阶段。

水下的情况从根本上截然不同。海豚和齿鲸等水生哺乳动物尽管没有外耳廓，却能实现精确的空间听觉 [6, 7]。软组织的声阻抗与海水的声阻抗非常接近，因此身体对声音而言大体上是半透明的。因此，空间线索并非来自声音绕过不透明障碍物时产生的衍射，而是源于声音穿过声学特性各异的分层结构并在其中发生散射。解剖学和计算研究表明，这些动物的头部由半透明脂肪组织构成，其中包裹着刚性骨结构，气囊则充当声学反射体 [8, 9, 7, 10, 11, 12, 13, 14]。行为测量已证实，海豚会利用双耳时间差与强度差进行空间感知 [15]。这项研究工作揭示了水下空间感知的一般物理原理：含有内部刚性散射体的半透明介质能够被动地将入射声场塑造成方向相关线索。

为将这一原理抽象出来，不依赖任何特定的生物解剖结构，并建立一个适用于水下和陆地环境的简单框架，我们研究了由典型几何结构产生的散射：一个半透明球体内含两个刚性球形散射体。我们利用多极展开和平移解析求解该问题 [16, 17], ，得到一个闭式前向模型，可根据频率、声源方向、材料属性和几何参数，计算外表面任意一点处的声压场。该模型以 JAX 实现 [18], ，完全可微，因此能够直接根据双耳观测，通过基于梯度的优化估计声源方向。这在两个方面拓展了简化几何HRTF模型的研究传统：外球可透声而非刚性，反映了生物组织在水中的半透明特性；此外，该模型适用于结构与周围介质之间阻抗对比很小的声学环境。

我们通过若干应用展示该模型。首先，我们分析这一几何结构产生的HRTF、ILD和ITD模式，确认其能够生成随方向变化的双耳线索。随后，我们将声源定位表述为一种与感知相联系的计算机制，通过基于梯度的优化来最小化联合ILD-ITD损失。我们评估模型对测量噪声的鲁棒性，并使用扩展卡尔曼滤波器（EKF）跟踪移动声源，同时利用可微模型计算所需的雅可比矩阵。

本文其余部分安排如下。第2 节回顾相关背景。第3.1 节介绍问题表述和解析解。第4 节定义空间线索和声源定位框架。第5节给出仿真结果，第6节总结全文。

2 背景

2.1 经典头相关传递函数建模

头相关传递函数描述了入射声场在到达鼓膜之前，受到听者头部、躯干和耳廓散射而发生变化的情况 [1]。对于位于$\mathbf{r}_s$ 处的声源，以及位于头部上或附近$\mathbf{r}$ 处的接收器，头相关传递函数定义为存在散射体时的声压与自由场中同一点处声压的比值：

$$H(\mathbf{r}, \mathbf{r}_s, \omega) = \frac{p_{\text{total}}(\mathbf{r}, \omega)}{p_{\text{free}}(\mathbf{r}, \omega)}, \tag{1}$$

其中， ω 表示角频率。头相关传递函数是复值的，编码了取决于频率以及声源与接收器相对几何关系的幅度和相位变化。在两只耳朵处进行评估时，头相关传递函数对会产生双耳声级差和耳间时间差，这两者是空间听觉的主要线索。

头相关传递函数的球形建模始于刚性球近似。Duda and Martens [2]表明，单个刚性球能够捕捉人头衍射的一些关键特征，产生

与测量结果相当吻合、适用于方位角定位的频率相关的耳间声级差和耳间时间差。“雪人”头部与躯干模拟器 [4] 将这一模型扩展为两个刚性球，从而捕捉躯干反射并改善仰角线索。沿着这一思路，人们开发了若干简化的几何HRTF模型 [3, 5], 这些模型表明，即使是简单的几何形状，也能产生丰富的双耳行为。

虽然这些刚性球体模型适用于陆地听觉（此时头部对声音基本不透明），但并不适用于水下环境。当散射体的声阻抗接近周围介质的声阻抗时，大部分入射能量会透射而非反射，此时刚性边界条件不再适用。

2.2 多球体与可透射散射模型

当球的声阻抗与周围介质不同、但并非远高于周围介质时，这种球称为半透明 球。声波入射到此类球上时，部分声波会被反射，部分会被透射，二者的比例由阻抗对比决定。在界面处，声压和质点速度的法向分量都必须连续，因此需采用透射条件，而非用于不透明散射体的刚性（诺伊曼）或软（狄利克雷）边界条件。当球内外阻抗完全匹配时，球便完全透明，不会产生散射。

为模拟具有多层不同声学特性的结构，研究人员将单球体框架扩展至可透射的多层几何结构 [19]。对于同心球壳，各界面处的声压和法向速度连续性会使各层的模态系数相互耦合，从而形成可通过传递矩阵求解的系统 [20]。

当内球与外壳不同心时，必须将每个球体产生的散射场转换到其他球体的坐标系中。这通过使用再展开系数 [16] 对波函数展开进行平移来实现。由此产生两类重新展开：奇异到规则（$S|R$）系数将一个球体的散射（出射）场转换为关于另一中心的规则（入射）展开；规则到规则（$R|R$）系数则用于在不同中心之间转换规则场。这些重新展开操作是精确的，并且能够保持场的模态结构。

Gumerov 和 Duraiswami [17] 提出了一个利用多极再展开计算多个刚性球散射的框架，这构成了本文工作的数学基础。他们的方法通过平移算子耦合边界条件，同时求解每个球体的散射场系数。在本文中，我们将这一框架扩展到刚性球体封闭于半透明球体内的情形，此时需要同时满足内球上的刚性边界条件以及外球上的透射条件。

2.3 可微分声学仿真

边界元法（BEM）、有限元法（FEM）和快速多极法（FMM）等数值方法能够以很高的精度计算任意几何结构的头相关传递函数（HRTF） [21]。然而，这些方法的计算成本很高，尤其是在高频条件下，而且难以直接提供输出相对于几何参数或材料参数的梯度。这限制了它们在优化或反问题中的应用；在这些问题中，人们希望调整物理设计以实现所需的声学特性。

自动微分（AD）框架（如 JAX [18] 和 PyTorch）主要为深度学习而开发，为基于物理的优化提供了另一种方法。在此类框架中实现前向模型后，即可通过链式法则，精确且高效地计算任意标量损失函数相对于所有输入参数的梯度，无需使用有限差分或进行伴随推导。这一范式已应用于日益增多的研究中，包括可微物理仿真 [22] 和可微麦克风阵列 [23]。在本研究中，使用 JAX 实现多极散射模型，可使整个正向映射（从几何参数、材料属性和声源方向到传感器位置处的声压）完全可微。这不仅能够通过基于梯度的优化来估计声源方向，实现定位，还可用于灵敏度分析，并原则上用于物理结构的逆向设计。

3 方法论

3.1 问题几何结构与控制方程

我们研究简谐声场在由三个球构成的结构中的散射：一个半透明（可穿透）外球内含两个刚性内球，且所有球均浸没在均匀流体介质中。其几何结构如图 1。

令 S_1 表示半透明外球，其半径为 a_1，中心为 $\mathbf{r}_1'$；令 S_2 和 S_3分别表示两个刚性内球，其半径为 a_2、a_3，中心分别为 $\mathbf{r}_2'$、$\mathbf{r}_3'$。外部介质（S_1之外）的密度为 ρ_o，声速为 c_o；内部介质（S_1之内但在 S_2 和 S_3之外）的密度为 ρ_i，声速为 c_i。相应的波数为 $k_o = \omega/c_o$和 $k_i = \omega/c_i$，其中 ω 为角频率。为简便起见，设 S_1 和 S_2 同心，因此 $\mathbf{r}_2' = \mathbf{r}_1'$，而 S_3 偏离共同中心。

我们引入两个坐标系：$(O_1x_1y_1z_1)$以$\mathbf{r}_1'$ 为中心（它也作为 S_2的坐标系），以及以$\mathbf{r}_3'$为中心的 $(O_3x_3y_3z_3)$。由于 S_1 和 S_2 共用同一中心，它们的坐标系之间的重新展开为恒等变换；只需在 O_1 和 O_3 之间进行平移。

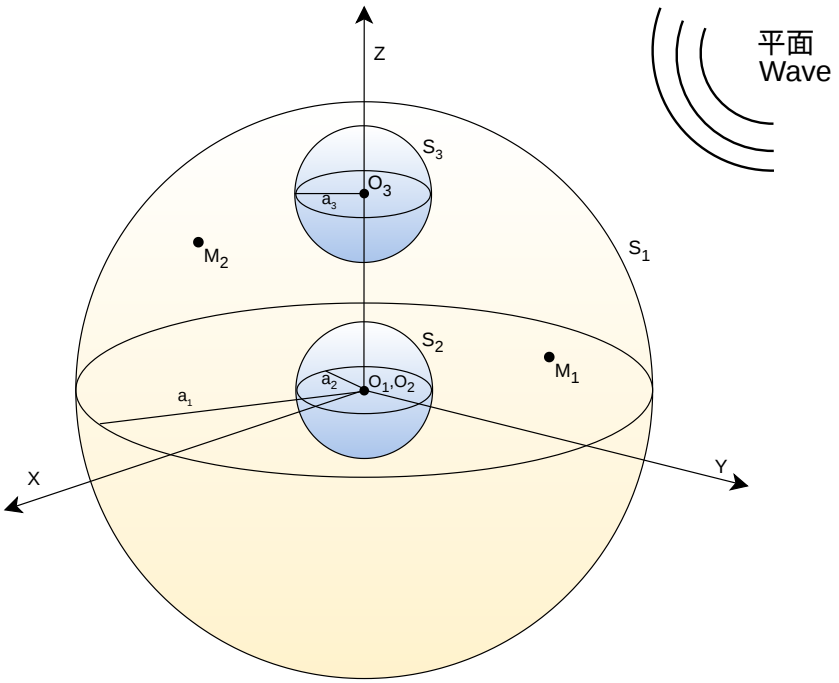


图 1：问题几何结构：半透明球体 S_1 内有两个刚性球 S_2 和 S_3。球 S_1 和 S_2 同心。外部介质的密度为 ρ_o，声速为 c_o；内部介质的密度为 ρ_i，声速为 c_i。

3.1.1 控制方程

各区域中的声压均满足波动方程。下文将讨论声压的傅里叶变换，ψ，其定义为

$$\psi(\mathbf{r},\omega) = \int_{-\infty}^{\infty} e^{i\omega t} p(\mathbf{r},t)\,\mathrm{d}t,$$

它满足亥姆霍兹方程。在外部区域（ S_1之外）：

$$\nabla^2\psi_o + k_o^2\psi_o = 0, \tag{2}$$

以及内部区域（在 S_1 内部、在 S_2 和 S_3 外部）：

$$\nabla^2\psi_{\text{internal}} + k_i^2\psi_{\text{internal}} = 0. \tag{3}$$

3.1.2 边界条件

在刚性球 S_2 和 S_3 的表面上，质点速度的法向分量为零，因此满足诺伊曼条件：

$$|\mathbf{r} - \mathbf{r}_2'| = a_2: \quad \frac{\partial\psi_{\text{internal}}}{\partial n_2} = 0, \tag{4}$$

$$|\mathbf{r} - \mathbf{r}_3'| = a_3: \quad \frac{\partial\psi_{\text{internal}}}{\partial n_3} = 0, \tag{5}$$

其中，法向导数沿外法向$\mathbf{n}_q$ 在表面 S_q处取值，适用于 $q = 2,\ 3$。

在半透明球 S_1 的表面上，法向质点速度和声压都必须在界面两侧保持连续。这些透射条件为：

$$|\mathbf{r} - \mathbf{r}_1'| = a_1: \quad \frac{\partial\psi_{\text{internal}}}{\partial n_1} = \frac{\partial\psi_o}{\partial n_1}, \tag{6}$$

$$\omega_i\rho_i\ \psi_{\text{internal}} = \omega_o\rho_o\ \psi_o, \tag{7}$$

其中 ω_i 和 ω_o 分别是内部和外部的角频率 S_1。这些条件确保法向质点速度的连续性（公式6）以及按介质阻抗缩放的声压连续性（公式 7）。

此外，散射场必须满足无穷远处的 Sommerfeld 辐射条件，以确保其表示出射波。

3.1.3 场分解

外部区域中的总场分解为已知入射场与由 S_1 引起的散射场之和：

$$\psi_o(\mathbf{r}) = \psi_{\text{in}}(\mathbf{r}) + \psi_{1,\text{scat}}(\mathbf{r}). \tag{8}$$

内部区域中的总场由 S_1 内部的透射入射场以及两个刚性球散射的场组成：

$$\psi_{\text{internal}}(\mathbf{r}) = \psi_{1,\text{inside}}(\mathbf{r}) + \psi_{2,\text{scat}}(\mathbf{r}) + \psi_{3,\text{scat}}(\mathbf{r}). \tag{9}$$

3.1.4 多极表示

我们使用球面多极展开表示各个场分量。在球坐标$\mathbf{r} = (r, \theta, \phi)$ 中，正则基函数和奇异基函数定义为

$$R_l^s(\mathbf{r}) = j_l(kr)\, Y_l^s(\theta,\phi), \qquad S_l^s(\mathbf{r}) = h_l(kr)\, Y_l^s(\theta,\phi),$$

其中，j_l 是第一类球贝塞尔函数，h_l 是第一类球汉克尔函数，Y_l^s是次数为 l 、阶数为 s的球谐函数，定义为 [24]

$$Y_l^s(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}}\sqrt{\frac{(l-|s|)!}{(l+|s|)!}}\, P_l^{|s|}(\cos\theta)\, e^{js\phi}, \tag{10}$$

其中， $P_l^{|s|}$ 是缔合勒让德函数。正则基函数 R_l^s 在原点处有限，表示入射场或透射场；而奇异基函数 S_l^s 满足辐射条件，表示散射场（向外传播的场）。

入射场和透射场关于 O_1 按正则基函数展开：

$$\psi_{\text{in}}(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{s=-l}^{l} E_l^s \, R_l^s(\mathbf{r}_1), \tag{11}$$

$$\psi_{1,\text{inside}}(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{s=-l}^{l} A_l^{(1)s} \, R_l^s(\mathbf{r}_1), \tag{12}$$

以及散射场在各自的坐标系中按奇异基函数展开：

$$\psi_{1,\text{scat}}(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{s=-l}^{l} B_l^{(1)s} \, S_l^s(\mathbf{r}_1), \tag{13}$$

$$\psi_{2,\text{scat}}(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{s=-l}^{l} C_l^{(1)s} \, S_l^s(\mathbf{r}_1), \tag{14}$$

$$\psi_{3,\text{scat}}(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{s=-l}^{l} D_l^{(3)s} \, S_l^s(\mathbf{r}_3). \tag{15}$$

系数上的上标表示其定义所在的坐标系。待确定的未知系数为 $A_l^{(1)s}$（S_1 内部的透射场）、$B_l^{(1)s}$（来自 S_1 的散射场）、$C_l^{(1)s}$（来自 S_2 的散射场）和 $D_l^{(3)s}$（来自 S_3 的散射场）。

3.1.5 入射场

对于强度为 Q、位于 $\mathbf{r}_s$ 处的单极点源，入射场为

$$\psi_{\text{in}}(\mathbf{r}) = Q\, G_k(\mathbf{r} - \mathbf{r}_s) = Q\, \frac{\exp(ik_o|\mathbf{r} - \mathbf{r}_s|)}{4\pi|\mathbf{r} - \mathbf{r}_s|},$$

其中 G_k 是自由空间格林函数。围绕 O_1 展开为多极子形式，得到 [16]

$$\psi_{\text{in}}(\mathbf{r}) = Q\, ik_o \sum_{l=0}^{\infty} \sum_{s=-l}^{l} S_l^{-s}(\mathbf{r}_s)\, R_l^s(\mathbf{r}),$$

因此 $E_l^s = Q\, ik_o\, S_l^{-s}(\mathbf{r}_s)$。

对于从方向 (θ_s, ϕ_s) 入射的平面波，入射场为 $\psi_{\text{in}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$，其展开式为

$$\psi_{\text{in}}(\mathbf{r}) = \sum_{l=0}^{\infty} i^l (2l+1)\, j_l(k_o r)\, P_l(\cos\theta),$$

由此得到 Condon–Shortley 相位约定下的系数 $E_l^s = 4\pi\, i^l\, \overline{Y_l^s}(\theta_s, \phi_s)$，其中上划线表示复共轭。本文所述的数值实验中，我们使用平面波入射场。

3.2 解析解

为求解第 3.1 节定义的耦合边值问题，我们将多极展开式代入每个球上的边界条件。由于 S_3 的散射场以坐标系 O_3 表示，而在 O_1 中对 S_1 和 S_2 施加边界条件最为自然，因此我们必须首先使用奇异到规则的平移，将 $\psi_{3,\text{scat}}$ 在 O_1 附近重新展开：

$$\psi_{3,\text{scat}}(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{s=-l}^{l} D_l^{(1)s} \, R_l^s(\mathbf{r}_1), \tag{16}$$

其中

$$D_l^{(1)s} = \sum_{m=|s|}^{\infty} (S|R)_{ml}^{s}(\mathbf{r}'_{31})\, D_m^{(3)s}. \tag{17}$$

这里，$\mathbf{r}'_{tz} = \mathbf{r}'_z - \mathbf{r}'_t$ 表示从球 S_t 的中心到球S_z的中心的平移向量，且 $(S|R)_{ml}^{s}$ 是奇异到正则同轴重展开系数， [16], 用于将一个中心处的外向（奇异）展开转换为另一个中心处的正则展开。

3.2.1 S_1上的边界条件

将多极展开代入速度连续性条件（6），并利用球谐函数的正交性，可得对每个阶 l 和阶数 s:

$$A_l^{(1)s}\, k_i\, j_l'(k_i a_1) + C_l^{(1)s}\, k_i\, h_l'(k_i a_1) + D_l^{(1)s}\, k_i\, j_l'(k_i a_1) - B_l^{(1)s}\, k_o\, h_l'(k_o a_1) = E_l^{s}\, k_o\, j_l'(k_o a_1). \tag{18}$$

类似地，由压力连续性条件（7），定义 $d = \omega_i \rho_i / (\omega_o \rho_o)$ ：

$$d\Big[A_l^{(1)s}\, j_l(k_i a_1) + C_l^{(1)s}\, h_l(k_i a_1) + D_l^{(1)s}\, j_l(k_i a_1)\Big] - B_l^{(1)s}\, h_l(k_o a_1) = E_l^{s}\, j_l(k_o a_1). \tag{19}$$

3.2.2 S_2上的边界条件

由于 S_2 与 S_1同心，所有场均已在共同坐标系 O_1中表示。应用诺伊曼条件（4）可得:

$$A_l^{(1)s}\, k_i\, j_l'(k_i a_2) + C_l^{(1)s}\, k_i\, h_l'(k_i a_2) + D_l^{(1)s}\, k_i\, j_l'(k_i a_2) = 0. \tag{20}$$

3.2.3 S_3上的边界条件

为在 S_3上施加诺伊曼条件，需要将内部场在坐标系O_3中表示。透射场$\psi_{1,\,\mathrm{inside}}$ 与散射场$\psi_{2,\,\mathrm{scat}}$均定义于 O_1附近，并使用适当的平移算子关于 O_3 重新展开:

$$\psi_{\mathrm{internal}}(\mathbf{r}_3) = \psi_{1,\mathrm{inside}}(\mathbf{r}_3) + \psi_{2,\mathrm{scat}}(\mathbf{r}_3) + \psi_{3,\mathrm{scat}}(\mathbf{r}_3) = \sum_{l=0}^{\infty}\sum_{s=-l}^{l}\Big[A_l^{(3)s}\, R_l^{s}(k_i\mathbf{r}_3) + C_l^{(3)s}\, R_l^{s}(k_i\mathbf{r}_3) + D_l^{(3)s}\, S_l^{s}(k_i\mathbf{r}_3)\Big], \tag{21}$$

其中平移系数为

$$A_l^{(3)s} = \sum_{m=|s|}^{\infty} (R|R)_{ml}^{s}(\mathbf{r}'_{13})\, A_m^{(1)s}, \tag{22}$$

$$C_l^{(3)s} = \sum_{m=|s|}^{\infty} (S|R)_{ml}^{s}(\mathbf{r}'_{13})\, C_m^{(1)s}. \tag{23}$$

此处，$(R|R)_{ml}^{s}$ 为正则到正则同轴重展开系数，用于将正则展开从一个中心平移到另一个中心 [16]。注意，$\psi_{2,\,\mathrm{scat}}$是关于 O_1的奇异展开，使用$(S|R)$系数进行平移，从而将其转换为关于 O_3的正则展开。

应用诺伊曼条件（5），在 O_3 坐标系中可得:

$$A_l^{(3)s}\, k_i\, j_l'(k_i a_3) + C_l^{(3)s}\, k_i\, j_l'(k_i a_3) + D_l^{(3)s}\, k_i\, h_l'(k_i a_3) = 0. \tag{24}$$

关于求解上述小节中提到的线性方程组以及求出系数A、 B、 C、和 D 的更多细节见附录A。

从几何结构和材料参数到任意评估点处声压的整个前向模型均以 **JAX** [18], 实现，因此可通过自动微分对所有输入求导。

4 空间线索

给定散射问题的解后，我们可以计算 S_1表面上任意一点的声压。我们在外球上的 M_L 和 M_R 处放置两个传感器，分别代表左侧和右侧接收器。这两处信号随方向变化的差异构成了用于声源定位的双耳空间线索。

4.1 定义与公式

点 M 在 S_1 的表面上的头相关传递函数定义为M 处的外部总声压与来自声源的自由场声压之比:

$$H(M) = \frac{\psi_o(M)}{Q\,G_k(r_s)}, \qquad M \in S_1, \tag{25}$$

其中， Q 为声源强度， $G_k(r_s)$ 是在声源距离处计算得到的自由空间格林函数。头相关传递函数为复数值，其取决于频率以及声源相对于散射体的角位置。

双耳声级差量化了两个传感器接收幅度之间的差异，以分贝表示:

$$\mathrm{ILD}(f) = 20\,\log_{10}\frac{|H(M_R, f)|}{|H(M_L, f)|}. \tag{26}$$

正双耳声级差表示声源更靠近右侧传感器，或受到右侧传感器的遮挡更少；由于几何结构产生的频率相关散射，其幅度和符号会随频率和声源方向而变化。

耳间时间差（ITD）量化了两个传感器之间的到达时间差。我们根据双耳相位差计算频率相关的耳间时间差:

$$\mathrm{ITD}(f) = \frac{\angle H(M_R, f) - \angle H(M_L, f)}{2\pi f}, \tag{27}$$

其中， $\angle(\cdot)$ 表示展开相位。这一定义基于相位，可得到频率相关的耳间时间差，捕捉散射在分层结构中产生的色散效应。在此处的几何结构中，内部介质的声速与外部介质不同，因此这一点尤为重要。

双耳声级差（ILD）和耳间时间差（ITD）共同以互补的方式编码声源方向：当散射体尺寸相对于波长较大时，双耳声级差在较高频率下最具信息量；而耳间时间差则能在较低频率下提供鲁棒的方向信息。第 3.1–3.2 节所述的前向模型将这两种线索计算为声源方向（θ， ϕ）、频率 f以及所有几何结构和材料参数的可微函数。

我们通过一系列数值实验评估所提出的模型。首先考察该几何结构产生的头相关传递函数（HRTF）、双耳声级差（ILD）和耳间时间差（ITD）模式，以确认其呈现符合物理预期的方向依赖性。随后，我们展示基于双耳线索的梯度声源定位，评估模型在不同声源方向网格上的抗测量噪声鲁棒性，评价波束形成性能指标，并展示使用扩展卡尔曼滤波器（EKF）进行运动声源跟踪的效果。

4.2 结果

本节中，外部介质均为水，其密度为 $\rho_o = 1000\ \mathrm{kg/m^3}$，声速为$c_o = 1500\ \mathrm{m/s}$；内部介质的密度为 $\rho_i = 920\ \mathrm{kg/m^3}$，声速为 $c_i = 1420\ \mathrm{m/s}$，这些数值代表生物软组织的特性 [10]。半透明外球的半径为 $a_1 = 0.2\ \mathrm{m}$，两个刚性内球的半径均为 $a_2 = a_3 = 0.1\ \mathrm{m}$。由于内外介质之间的阻抗对比较小（$\rho_i c_i / \rho_o c_o \approx 0.87$），半透明外球几乎完全透明，因此散射主要由两个刚性内球产生。

4.2.1 双刚性球头相关传递函数

图 2展示了本文提出的现有双刚性球模型的头相关传递函数（HRTF） [17], ；在本研究中，该模型表示外部介质为水时，半透明脂肪球形环境中的两个刚性球的头相关传递函数。我们在研究中使用此模型，因为它更准确；而且，两个刚性球模型与实际模型的头相关传递函数在频率和声源角度变化时，其整体结构模式和幅值范围大致相同 ϕ。这是因为，在半透明球体中的两个刚性球模型中，脂肪介质的声学特性与周围水的声学特性相似。

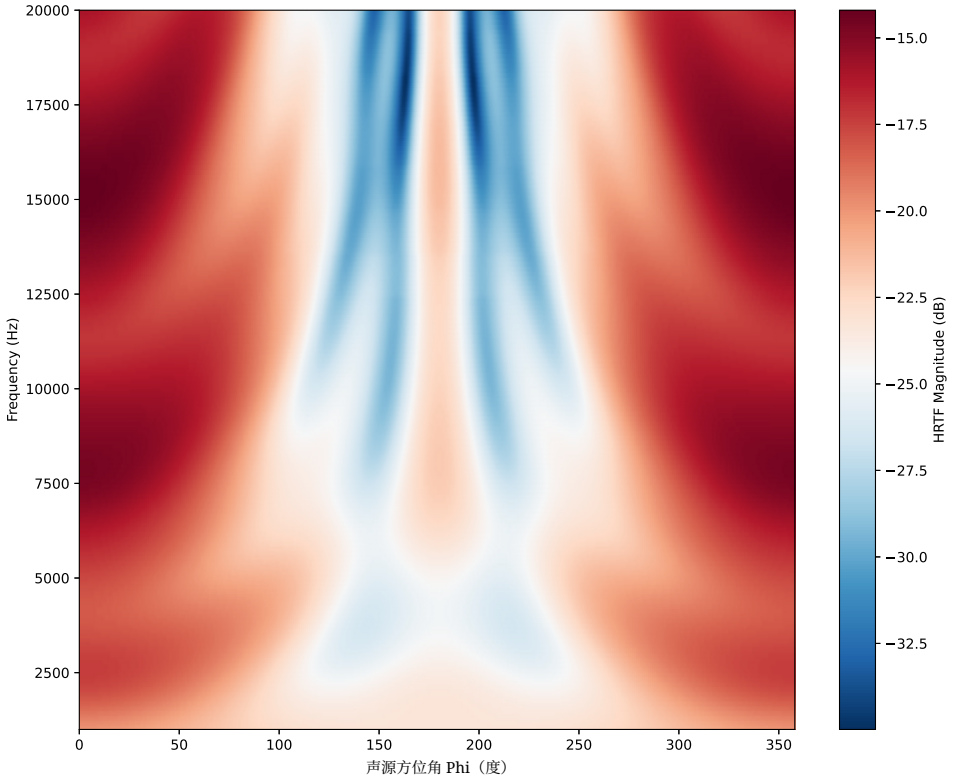
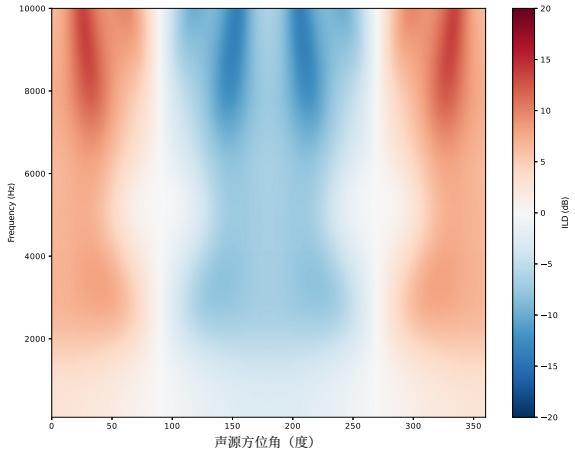


图 2：三球头相关传递函数随声源方位角 Phi（度）和频率（Hz）的变化

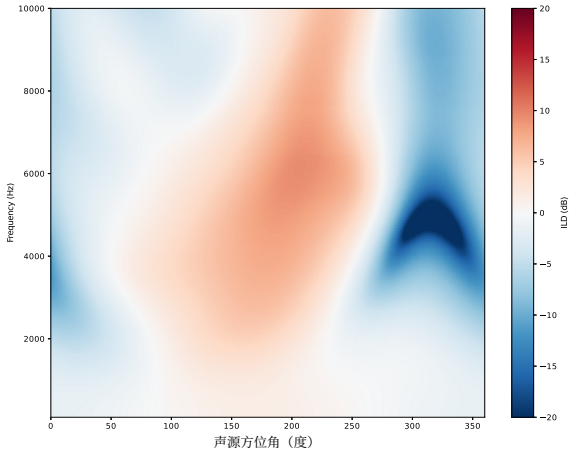
4.2.2 双耳线索

在图 3中，子图 (a) 展示了对称传感器的双耳声级差随声源方位角和频率的变化，子图 (b) 则展示了非对称传感器对应的变化。从图 3中可以观察到一些重要的空间模式。在较低频率（$f \leq 1500$）下，子图 (a) 和 (b) 中的双耳声级差随方位角平滑变化。我们可以观察到，当声源固定在 $\theta = \pi/2$，且方位角 ϕ 从 $0°$ 变化到 $360°$时，双耳声级差在两端达到最大值，并在 $\phi = 180°$ 或左侧传感器的方位角附近达到最小值。这是因为在这些方位角下，声源更靠近右侧传感器，因此接收到的声压更高，双耳声级差也随之增大。

图 4 展示了耳间时间差随声源方位角和频率的变化。子图 (a) 和 (b) 表明，低频时变化近似呈正弦曲线，而高频时变化更为平缓；当声源位于传感器附近时，延迟最大，在中矢面处则最小。我们可以观察到，当声源固定在 $\theta = \pi/2$，而方位角 ϕ 从 $0°$ 变化到 $360°$时，耳间时间差在两端达到最小值，并在 $\phi = 180°$附近达到最大值。这种现象是因为在较小方位角下，声源距离右侧传感器更近，导致传播时间更短，因而耳间时间差更小。对于靠近传感器的水平声源，特别是在对称情况下，较大的耳间时间差

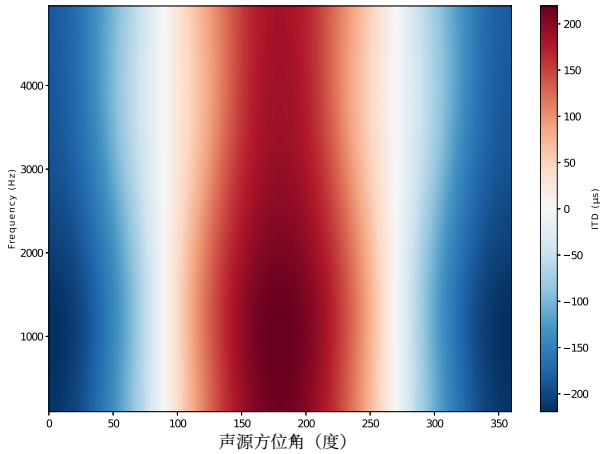


（a）对称麦克风传感器位置的双耳声级差。

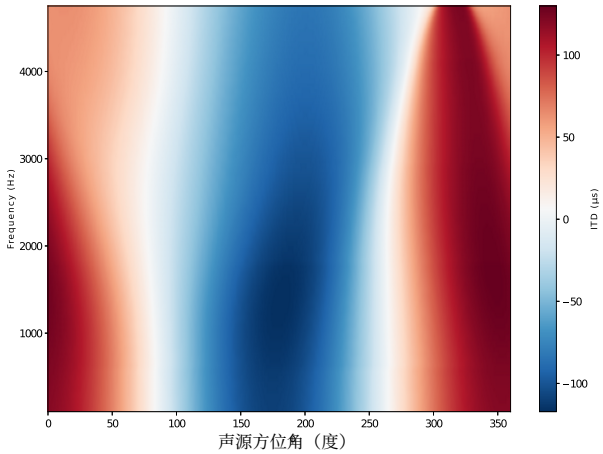


（b）非对称麦克风传感器位置的双耳声级差。

图 3：指向性指数与声源 ϕ （度）及频率（Hz）的关系。



(a) 对称麦克风传感器位置下的耳间时间差。



(b) 非对称麦克风传感器位置下的耳间时间差。

图 4：耳间时间差与声源方位角 ϕ 及频率（Hz）的关系。

与球形散射模型的理论发现一致。这些结果表明，该模型成功捕捉了双耳线索预期的方向依赖性。

可以看出，我们提出的模型生成了符合物理规律的双耳线索，呈现出水下散射所预期的正确方向依赖性、对称特性和频率变化趋势。这些模式表明，该模型捕捉到了海豚提取空间信息的基本机制。头相关传递函数和前向模型所呈现的这些指向性指数和耳间时间差特征，可用于后续章节中开发的声源定位和声源跟踪算法。

5 声源定位

本节使用所提出的散射模型来表述声源定位问题。目标是根据接收到的双耳空间线索估计声源方向。我们将定位视为一个连续优化问题，使预测的双耳声级差与时间差与测量值相匹配。

5.1 定义与公式

我们将声源定位表述为一个连续优化问题。给定在两个传感器处测得的一组观测双耳线索 ILD⋆(*f*) 和 ITD⋆(*f*)，我们要寻找声源方向 (*θ*， *ϕ*)，使预测线索与观测值最为匹配：

$$\min_{\theta,\,\phi}\ \mathcal{L}(\theta,\phi)\quad \text{s.t.}\quad \theta\in[0,\pi],\ \phi\in[0,2\pi), \tag{28}$$

其中，损失函数将一组分析频率$\mathcal{F}$上的双耳声级差与耳间时间差平方误差相结合：

$$\begin{aligned}\mathcal{L}(\theta,\phi)=\sum_{f\in\mathcal{F}}\Big[&\big(\overline{\mathrm{ILD}}(f;\theta,\phi)-\overline{\mathrm{ILD}}^{\star}(f)\big)^2\\&+\big(\overline{\mathrm{ITD}}(f;\theta,\phi)-\overline{\mathrm{ITD}}^{\star}(f)\big)^2\Big].\end{aligned} \tag{29}$$

此处，上划线表示归一化量：由于双耳声级差（单位为dB）与耳间时间差（单位为秒）的量级不同，二者在求和前均被归一化到可比尺度。具体而言，每个线索除以其在该频带内的取值范围，从而使任一项都不会主导损失。

另一种用于衡量我们优化框架性能的指标，是单位球面上估计声源方向与真实声源方向之间的角度误差 [16]，并按如下方式计算

$$\varepsilon=\arccos\left(\frac{4\pi}{3}\,\mathrm{Re}\sum_{s=-1}^{1}Y_1^s(\widehat{\theta},\widehat{\phi})\,\overline{Y_1^s(\theta^*,\phi^*)}\right). \tag{30}$$

$\widehat{\ }$ $\widehat{\ }$其中（*θ*， *ϕ*）为估计方向，（*θ*∗， *ϕ*∗）为真实方向，上划线表示复共轭。

由于前向模型在JAX中实现且完全可微，梯度 $\partial\mathcal{L}/\partial\theta$和 $\partial\mathcal{L}/\partial\phi$ 可通过自动微分精确计算。我们使用Adam优化器 [25], 最小化损失，该优化器以自适应学习率执行梯度下降。该表述将定位视为连续逆问题，而非离散网格搜索，其角度分辨率仅受优化器收敛容差的限制，而不受预定义网格间距的限制。

为避免双耳线索中的前后混淆，我们将左侧和右侧传感器置于球面上略微不对称的位置，从而打破双侧对称；否则，对称声源对会产生相同的线索模式。

5.2 声源定位：单一方向

首先，我们演示单个声源方向的定位过程。真实源位于 (*θ*∗， *ϕ*∗) = (2。13，1。10) 弧度。使用 Adam优化器，以 0。02 的学习率，从初始猜测 (θ_0， ϕ_0) = ($\pi/4$， $\pi/2$)，最小化联合耳间声级差–耳间时间差损失函数 (29。

图 5 展示了损失景观随仰角和方位角变化的情况。对于此配置，联合损失在真实声源方向处呈现一个吸引域。双耳声级差和耳间时间差分量提供了互补信息：双耳声级差损失的等高线相对宽缓，而耳间时间差损失则能提供更精确的定位；两者结合后形成了明确的极小值。

图6展示了收敛行为。角度误差单调下降，降至 1°以下（第43次迭代），并达到约 0.03° 的最终值（图6a）。损失与参数轨迹（图6b）证实了向真实源方向的平滑收敛，最终损失为$\mathcal{L}=1.37\times10^{-3}$。

为评估对初始化的敏感性，我们从另外两个起始点重复进行了优化。从 (θ_0， ϕ_0) = (0.57，0.29) 出发，优化器经过 72 次迭代后收敛，最终角度误差为 0.66° 。从 (θ_0， ϕ_0) = (2.57，1.29) 出发，优化器经过 69 次迭代后收敛至 0.19° 。在所有情况下，优化器都到达了真实声源附近，但收敛速度和最终精度取决于与解的初始距离。

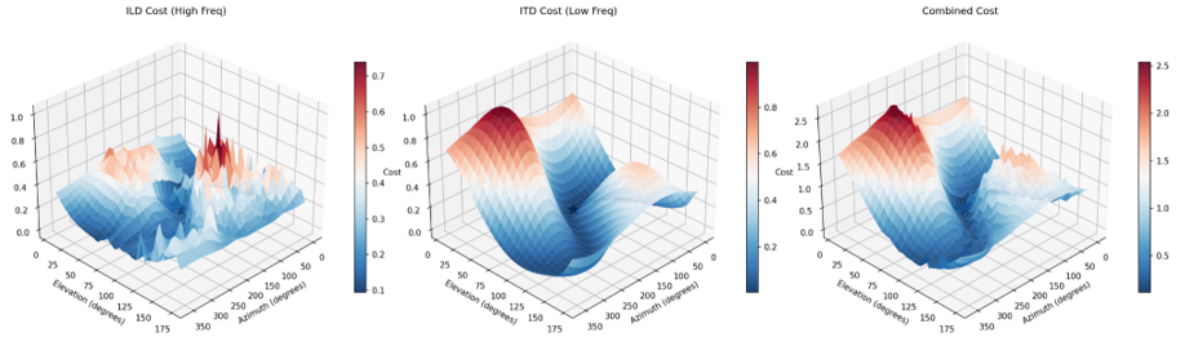


图 5：联合 ILD–ITD 目标函数随声源仰角和方位角变化的损失景观。左：ILD 分量。中：ITD 分量。右：组合损失，显示出单一吸引域。

表 1：每种条件下 32 个方向和 20 次噪声实现的定位性能。

SNR (dB)	平均误差 (度)	中位数误差 (度)	比例 $< 5°$	比例 $< 10°$
20	10.70	9.07	0.666	0.803
10	15.25	12.66	0.391	0.631
0	33.85	26.57	0.097	0.267

5.3 声　　源定位：跨方向与噪声水平的鲁棒性　　ls

为系统评估该定位框架，我们在32个声源方向组成的网格上对其进行了测试（4个仰角 $\times$ 8 方位角：$\theta \in \{60°, 90°, 120°, 150°\}$ 和 $\phi \in \{0°, 45°, \ldots, 315°\}$），并在三个线索级信噪比（信噪比）下进行：20、10和0 dB。在每个方向和信噪比下，测试了20次独立噪声实现，共计 $32 \times 3 \times 20 = 1920$ 次定位试验。

噪声被直接施加到双耳声级差和耳间时间差线索上。对于给定的信噪比，噪声标准差被设置为与干净线索的均方根成比例：

$$\sigma_{\text{cue}} = \frac{\text{RMS}(\text{cue}_{\text{clean}})}{10^{\text{SNR}/20}}. \tag{31}$$

每次试验使用多起点Adam优化器，在球面上分布四个初始化，并保留最终损失最低的结果。优化器最多运行100次迭代，并在连续30步无改进后早停。

表 1 总结了结果。在高信噪比（20 dB）下，平均角误差较小，绝大多数试验都能收敛到真实方向的 $5°$ 以内。随着噪声增加，性能平稳下降：在 10 dB 下，中位误差仍处于适中水平；而在 0 dB 下——此时噪声标准差等于提示信号的均方根——优化器在大多数试验中仍能将声源定位到 $10°$ 以内，但低于该阈值的$5°$ 结果比例有所下降。

这些结果证实，可微前向模型能够在各种声源方向和噪声条件下实现可靠的基于梯度的定位，其性能会随信噪比呈现可预测的变化。

5.4 噪声鲁棒性：详细示例

为说明噪声对单个声源方向的影响，图7 展示了在加性高斯噪声条件下的收敛行为，其中噪声为 $\sigma_{\text{ILD}} = 0.5$ dB，且 $\sigma_{\text{ITD}} = 10$ μs。优化器收敛至最终角度误差 $1.73°$，其中仰角分量误差为 $0.78°$，方位角分量误差为 $1.81°$。收敛轨迹保持平滑，但最终损失（5.80×10^{-2}）高于无噪声情形，反映出测量噪声引入了无法消除的失配。

6 使用匹配滤波器的空间波束形成

尽管本研究的重点是空间线索生成和声源定位，但在实际应用中，麦克风传感系统通常会采用波束形成来提升信号质量。因此，我们使用匹配滤波器评估所提出麦克风阵列的波束形成性能，并采用白噪声增益（WNG）和指向性指数（DI）等经典指标展示其空间滤波特性。这些指标分别衡量波束形成器对传感器噪声的鲁棒性及其空间选择性 [26, 27]。

6.1 定义与公式

假设复 数域中对应于方向 Ω 上声源的导向向量 $= (\theta,\ \phi$ $)$ 定义为

$$\mathbf{h}(f,\Omega) = \begin{bmatrix} H_L(f,\Omega) \\ H_R(f,\Omega) \end{bmatrix}, \tag{32}$$

其中，H_L 和 H_R 分别表示使用所提出的散射模型计算得到的左侧和右侧麦克风传感器处的头相关传递函数。

对于目标方向 Ω_0, ，匹配滤波器以如下权重组合两个通道接收到的信号，

$$\mathbf{w}(f) = \frac{\mathbf{h}_0(f)}{\mathbf{h}_0^H(f)\,\mathbf{h}_0(f)} \tag{33}$$

此归一化确保在目标方向上具有无失真响应，即 $\mathbf{w}^H(f)\mathbf{h}_0(f) = 1$，其中目标方向上的导向矢量为

$$\mathbf{h}_0(f) = \begin{bmatrix} H_L(f,\Omega_0) \\ H_R(f,\Omega_0) \end{bmatrix}. \tag{34}$$

因此，波束形成器的输出为

$$Y(f) = \mathbf{w}^H(f)\mathbf{X}(f), \tag{35}$$

其中，$\mathbf{X}(f)$ 是两个麦克风接收到的信号向量。

白噪声增益 衡量波束形成器对空间不相关白色传感器噪声的敏感度。归一化白噪声增益定义为

$$\mathrm{WNG}(f) = \frac{1}{\mathbf{w}^H(f)\mathbf{w}(f)}. \tag{36}$$

指向性因子 用于评估波束形成器的空间选择性，定义为目标方向上接收到的波束功率与所有方向上的平均波束功率之比：

$$\mathrm{DF}(f) = \frac{\left|\mathbf{w}^H(f)\mathbf{h}(f,\Omega_0)\right|^2}{\frac{1}{4\pi}\int_{4\pi}\left|\mathbf{w}^H(f)\mathbf{h}(f,\Omega)\right|^2 d\Omega}. \tag{37}$$

上式中对所有方向的球面积分通过采用带有角度加权的离散方向网格进行近似。指向性指数（DI）是以分贝表示的指向性因子（DF），其表达式为

$$\mathrm{DI}(f) = 10\log_{10}\left(\mathrm{DF}(f)\right). \tag{38}$$

6.2 结果

使用指向目标方向（$\theta,\ \phi$） $=$ (2.1293，1.0996) 弧度的匹配滤波波束形成器，我们首先验证了所有评估频率下的无失真响应，其中 $|\mathbf{w}^H\mathbf{h}_0| = 1.0$，确认归一化正确。整个频带内的白噪声增益均为正，范围为 3.35 分贝至 5.96 分贝（平均值为 ≈ 5.21 分贝），表明噪声放大表现稳定，且对传感器噪声不过度敏感。

DF 和 DI 是在一个包含 162 个方向的球面网格上计算得出的。DF 的变化范围为 1。86 至 2。41，对应的 DI 范围为 2。70 dB 至 3。81 dB（平均 DI 为 ≈ 3.16 dB）。图 8 展示了白噪声增益和 DI 随频率变化的范围（单位：dB）。

7 跟踪移动声源

本节评估所提出的双麦克风系统的时域波束形成性能。我们通过应用卡尔曼滤波器进行时域滤波，以了解并测量所提出系统的性能。

7.1 定义与公式

为实现对移动声源的连续跟踪，我们采用具有恒定速度状态模型的扩展卡尔曼滤波器 [28, 29, 30]。状态向量定义为

$$\mathbf{x}_t = \begin{bmatrix} \theta_t & \phi_t & \dot{\theta}_t & \dot{\phi}_t \end{bmatrix}^T, \tag{39}$$

其中 θ_t 和 ϕ_t 分别表示声源的仰角和方位角。

在采样间隔为 Δt的条件下考虑恒定角速度，运动模型为

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t, \tag{40}$$

with

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{41}$$

以及过程噪声$\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$。过程噪声模型的更多细节见附录B。

非线性测量模型为

$$\mathbf{z}_t = h(\mathbf{x}_t) + \mathbf{v}_t, \tag{42}$$

其中$\mathbf{z}_t$ 将各频率分箱中测得的双耳声级差和耳间时间差线索堆叠起来，构成一个高维观测向量。 $h(\cdot)$表示基于物理原理的正向声学模型，而$\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$为测量噪声。

在每个时间步，扩展卡尔曼滤波器都会进行预测

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}\hat{\mathbf{x}}_{t-1|t-1}, \tag{43}$$

$$\mathbf{P}_{t|t-1} = \mathbf{F}\mathbf{P}_{t-1|t-1}\mathbf{F}^T + \mathbf{Q}, \tag{44}$$

此处，$\mathbf{P}^{t|t}$ 表示状态估计误差协方差矩阵，用于量化角位置和速度估计的不确定性。

由于基于物理原理的声学测量函数 $h(\mathbf{x})$关于声源角度是非线性的，扩展卡尔曼滤波器会通过一阶泰勒展开，在预测状态 [29]附近对其进行线性化。雅可比矩阵$\mathbf{H}_t$刻画了双耳声级差和耳间时间差线索对微小角度扰动的局部敏感性，从而实现递归不确定性传播。

$$\mathbf{H}_t = \left. \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_{t|t-1}}, \tag{45}$$

雅可比矩阵由可微声学模型计算得到，从而能够对测量函数进行一致的线性化，得到

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^T \left(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^T + \mathbf{R}\right)^{-1}, \tag{46}$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\big(\mathbf{z}_t - h(\hat{\mathbf{x}}_{t|t-1})\big), \tag{47}$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}. \tag{48}$$

扩展卡尔曼滤波方程的详细推导见附录C。

7.2 结果

我们假设声源在 $T = 60$ 个离散时间步内移动。声源轨迹定义为球坐标系中的平滑线性运动，其中仰角 θ 从 110° 变化到 140°，方位角 ϕ 从 50° 变化到 110°。在每个时间步，我们使用所提出的前向模型生成频率相关的耳间声级差和耳间时间差线索。

测量噪声建模为均值为零的高斯噪声，并添加到双耳声级差和耳间时间差观测值中。假设的标准差分别为 $\sigma_{\mathrm{ILD}} = 0.5\mathrm{dB}$ 和 $\sigma_{\mathrm{ITD}} = 10\,\mu\mathrm{s}$，与真实音频感知系统中的线索不确定性一致。

使用扩展卡尔曼滤波器进行声源跟踪，时间步长为 $\Delta t = 1.0$。通过角加速度项引入过程噪声，其中$\sigma_{\theta,\ \mathrm{acc}} = 0.03$ 和$\sigma_{\phi,\ \mathrm{acc}} = 0.03$，从而使轨迹平滑演变但不完全确定（参见附录B）。滤波器初始化于（θ_0，ϕ_0）=(2.2,2.6)，并根据带噪的多频耳间声级差和耳间时间差测量值递归估计声源角度。

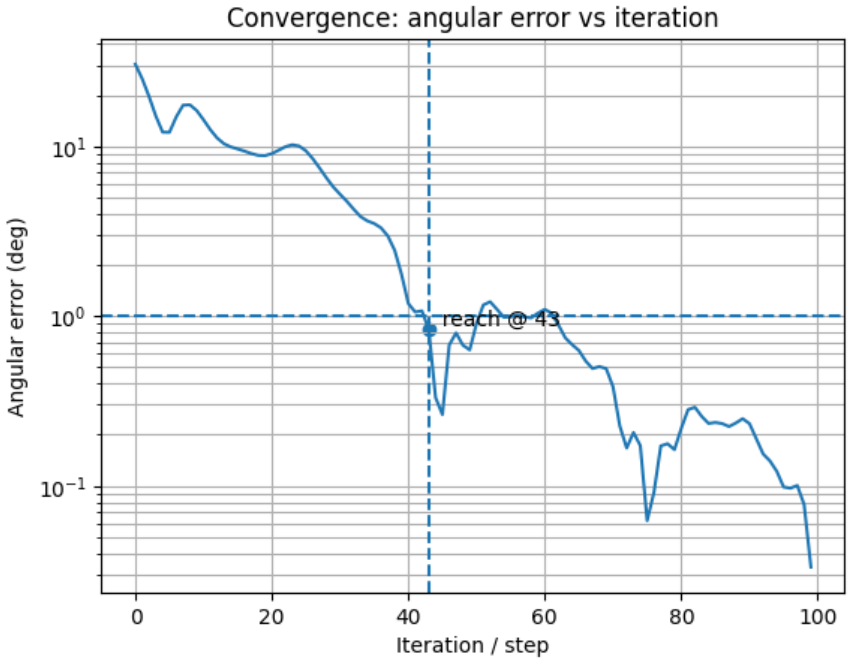
图9 展示了所提出的扩展卡尔曼滤波器的跟踪性能。从不准确的初始点出发经过几个时间步后，随着滤波器输出收敛到真实点，角度误差迅速减小。收敛后，如图10 所示，在轨迹的大部分区间内，跟踪误差保持在 1° 以下；瞬态阶段过后，$\pm 1\sigma$ 置信带收窄，表明估计确定性不断提高。这些结果表明，可微前向模型使扩展卡尔曼滤波器能够对非线性前向模型进行线性化，从而实现准确、稳定的声源跟踪。

8 讨论

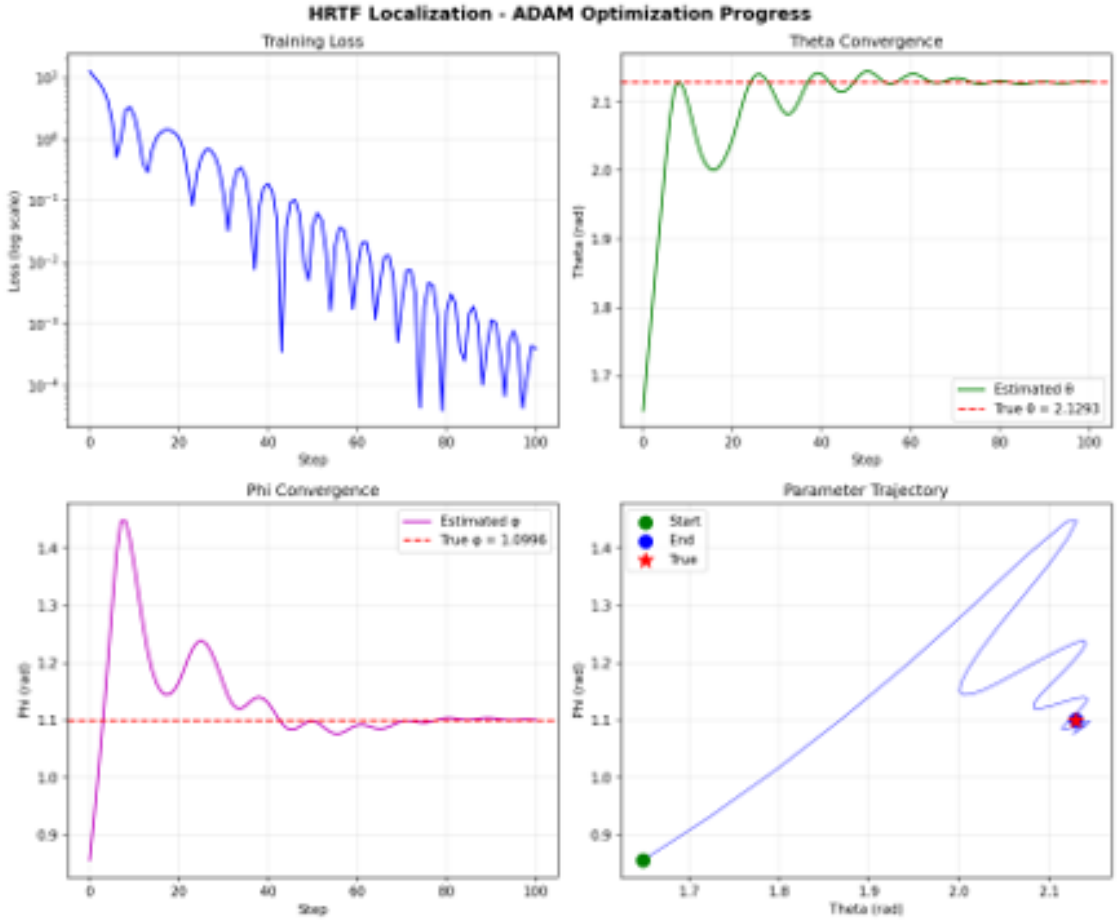
与以往研究不同，我们通过对散射进行建模，设计出一种设备，仅用最少的双传感器阵列即可生成空间线索，从而实现精确定位和稳定跟踪。通过对声散射场建模，该框架能够捕捉在自由场假设下难以产生的双耳空间线索。未来的研究可探索更复杂的运动模型、与真实世界测量的整合，以及在存在混响和多个声源的复杂声学环境中进行评估。虽然所提出的框架采用了最少数量的传感器，但利用更多传感器可增强鲁棒性。此外，研究自适应或基于学习的线索融合策略，或许能够提升系统在严苛条件下的性能。

9 结论

我们研究了受水生哺乳动物内部解剖特征启发、采用多球散射模型的双传感器系统性能。我们使用多极展开法，探究海豚如何生成并利用空间线索进行水下导航和定位。本文提出了一个解析与数值框架，用于确定该多球模型左右传感器处的头相关传递函数。通过求解声散射问题，我们获得了具有一致性的头相关传递函数、耳间时间差和耳间声级差模式，这些特征描述了定位、波束形成和跟踪所用的空间线索。解析结果与数值结果相吻合，验证了所提出模型的准确性和物理真实性。本研究为理解散射系统（包括水下环境中的散射系统）的空间机制，提供了一个在数学和物理层面均可解释的整体基础。

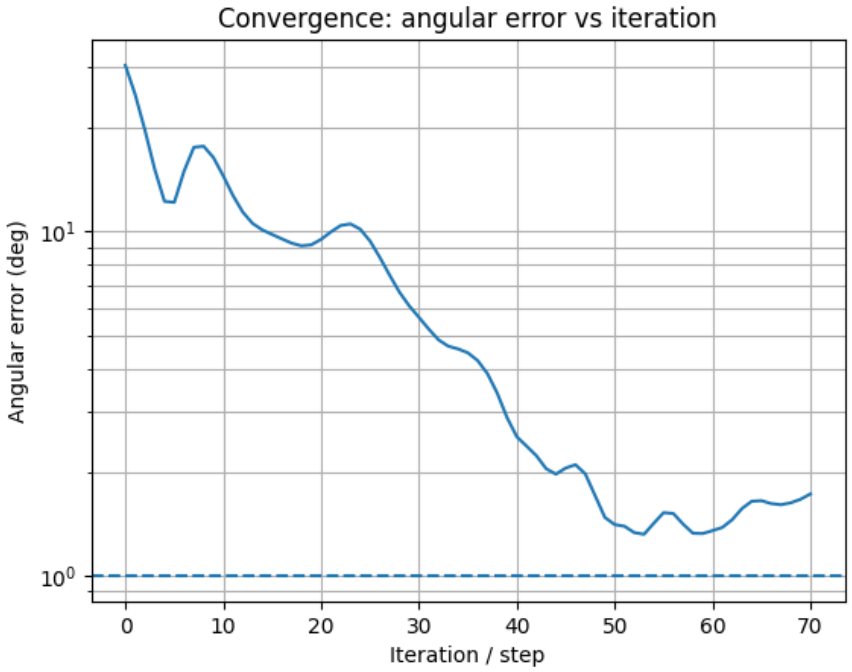


(a) 角度误差与迭代次数的关系

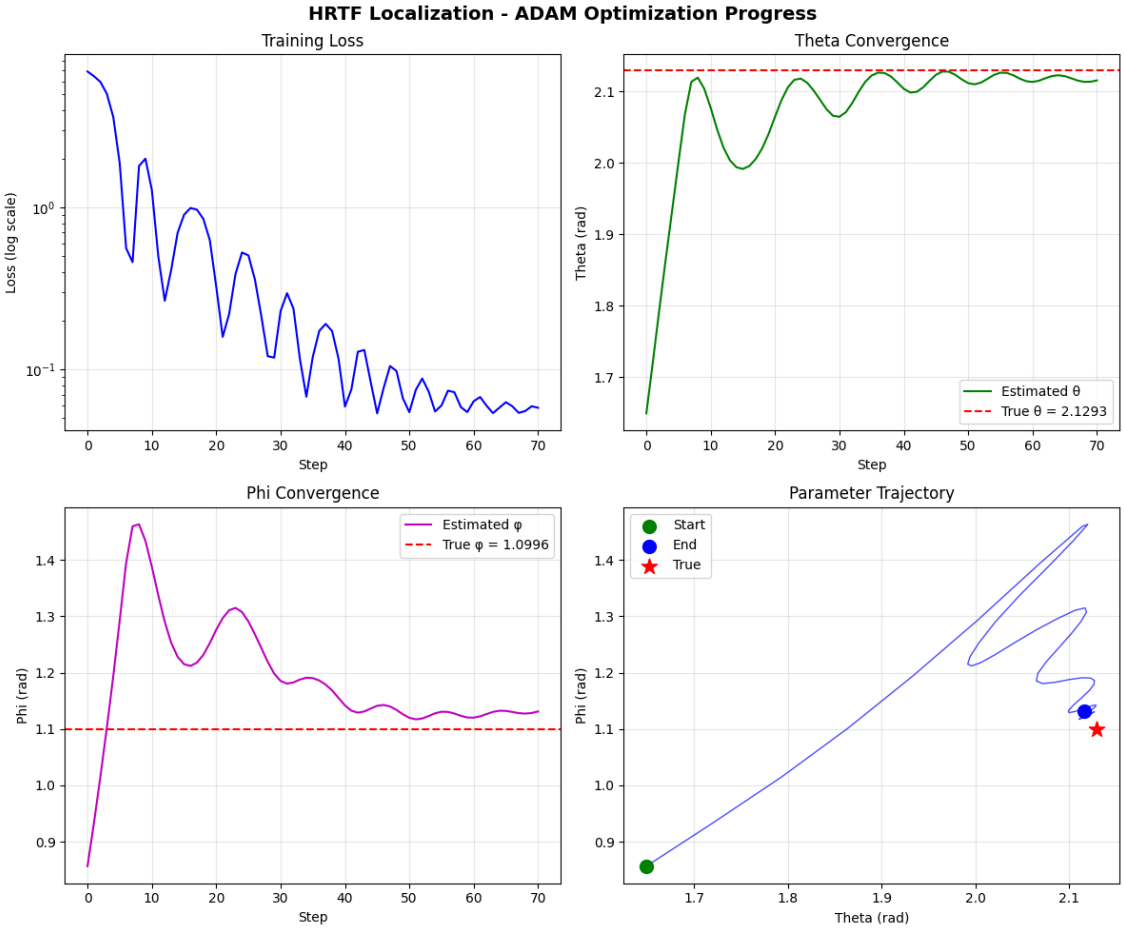


(b) 损失及参数轨迹

图 6：单个声源位于 $(\theta^*, \phi^*) = (2.13,\ 1.10)$ 弧度处时，基于梯度的定位的收敛情况。(a) 角度误差与优化器迭代次数的关系。(b) 损失值和估计的 (θ, ϕ) 与迭代次数的关系。



(a) 角度误差随迭代次数的变化



(b) 损失和参数轨迹

图7：测量噪声下的定位（$\sigma_{\mathrm{ILD}} = 0.5$ dB, $\sigma_{\mathrm{ITD}} = 10\ \mu$s）。（a）角度误差随迭代次数的变化。（b）损失和参数轨迹。

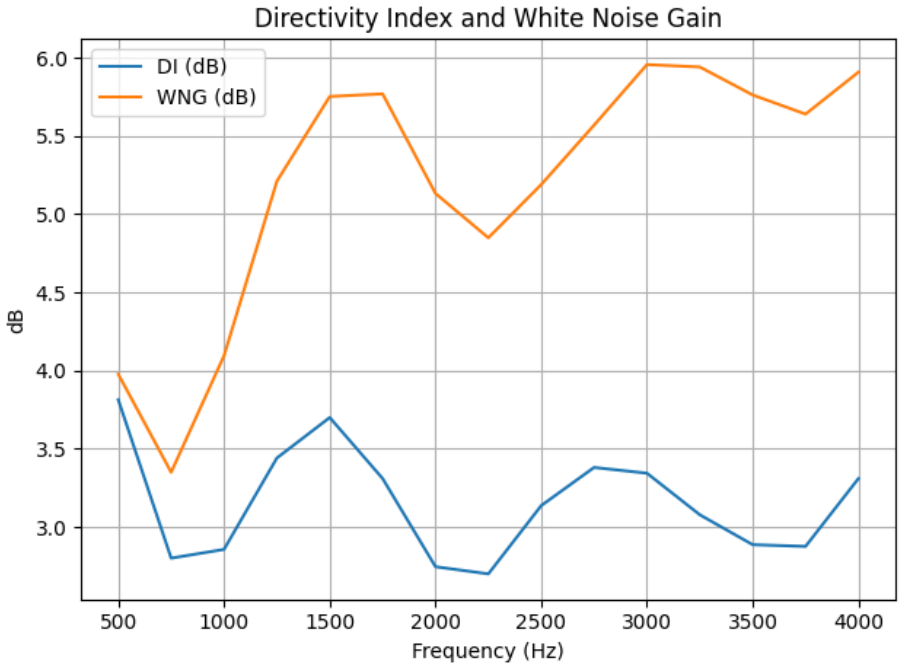


图 8：白噪声增益和指向性指数随频率的变化

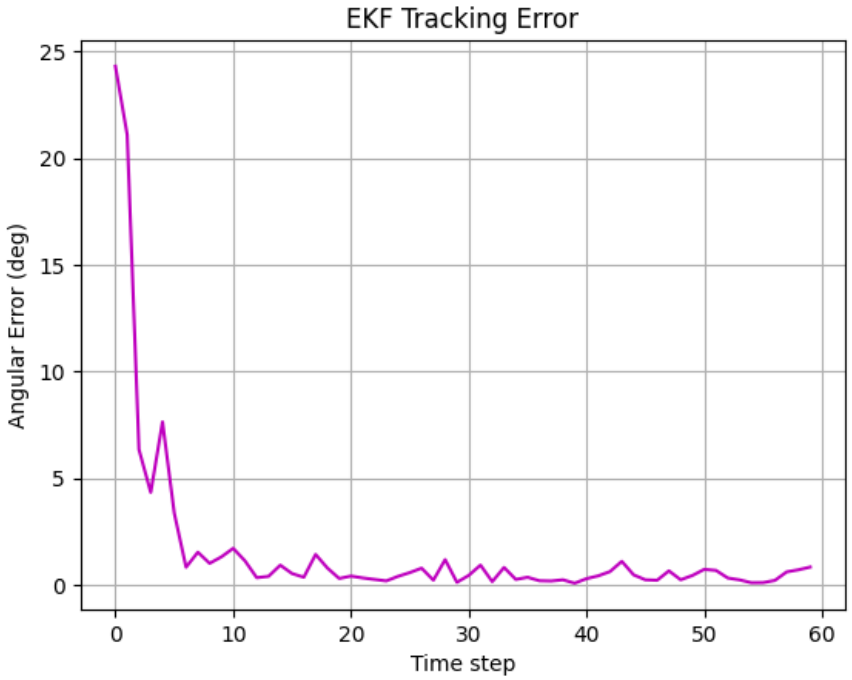


图 9：扩展卡尔曼滤波器的角度跟踪误差随时间的变化。

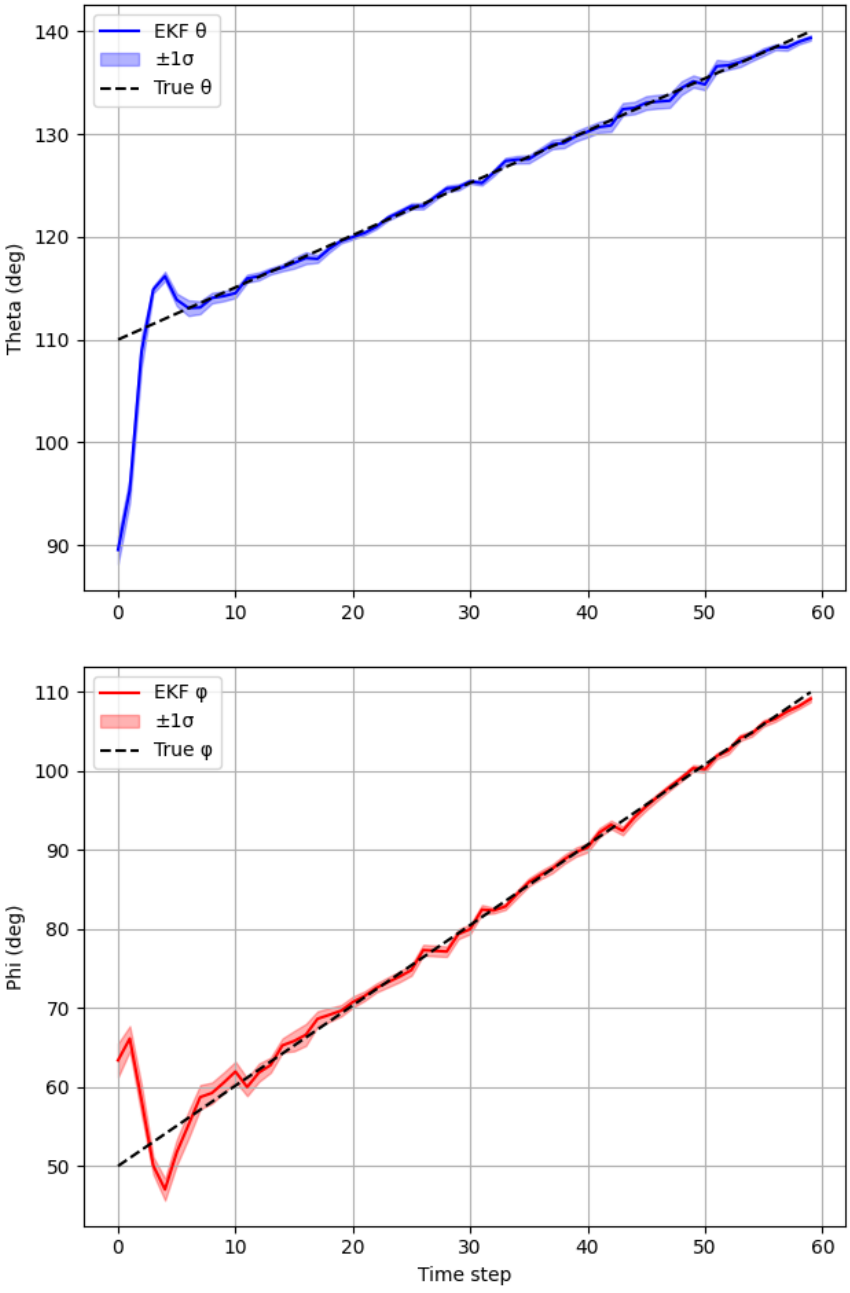


图10：扩展卡尔曼滤波器对仰角（上）和方位角（下）的跟踪结果。

附录

耦合线性系统

方程（18）、（19）、（20）和（24）构成了关于未知系数 $A^{(1)}$、$B^{(1)}$、$C^{(1)}$和 $D^{(3)}$的耦合无限维线性系统。为求数值解，我们将展开截断至阶数$p-1$，保留模态 $l=0$、1、.....、$p-1$。按照 [17],，当 10^{-3}时，$p \geq 3k_i a_1$收敛至。

为将截断系统表示为矩阵-向量形式，我们定义以下量。各阶的平移矩阵 s 为

$$(X|R)^s(\mathbf{t}) = \begin{pmatrix} (X|R)^s_{|s|,|s|} & (X|R)^s_{|s|,|s|+1} & \cdots & (X|R)^s_{|s|,p-1} \\ (X|R)^s_{|s|+1,|s|} & (X|R)^s_{|s|+1,|s|+1} & \cdots & (X|R)^s_{|s|+1,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ (X|R)^s_{p-1,|s|} & (X|R)^s_{p-1,|s|+1} & \cdots & (X|R)^s_{p-1,p-1} \end{pmatrix} \tag{49}$$

其中 $X \in \{S, R\}$、$s=0$、±1、.....、$\pm(p-1)$，且 $\mathbf{t}$ 表示球心之间的平移向量。表示每个球上局部边界关系的对角矩阵为

$$\mathbf{\Lambda}_f^{(q)s} = \operatorname{diag}\left(\Lambda_{|s|}^{(q)f},\ \Lambda_{|s|+1}^{(q)f},\ \ldots,\ \Lambda_{p-1}^{(q)f}\right), \tag{50}$$

$$\mathbf{\Gamma}_f^{(q)s} = \operatorname{diag}\left(\Gamma_{|s|}^{(q)f},\ \Gamma_{|s|+1}^{(q)f},\ \ldots,\ \Gamma_{p-1}^{(q)f}\right), \tag{51}$$

其元素为

$$\Lambda_l^{(q)f} = \frac{j_l'(k_f a_q)}{h_l'(k_f a_q)}, \qquad \Gamma_l^{(q)f} = \frac{j_l(k_f a_q)}{h_l(k_f a_q)}, \tag{52}$$

对于 $l=|s|,\ \ldots,\ p-1$、$q \in \{1,2,3\}$ 和 $f \in \{i,o\}$。这些比值源于将边界条件表示为每个球面上正则基函数与奇异基函数之比。未知系数向量为

$$\mathbf{X}^{(q)s} = \begin{pmatrix} X_{|s|}^{(q)s} \\ \vdots \\ X_{p-1}^{(q)s} \end{pmatrix}, \qquad X \in \{A, B, C, D\}, \tag{53}$$

而源项源于已知的入射场系数以及 S_1 上的透射条件，为

$$\mathbf{E}^{(1)s} = \begin{pmatrix} E_{|s|}^s \dfrac{j_{|s|}(k_o a_1)}{j_{|s|}(k_i a_1)} \\ \vdots \\ E_{p-1}^s \dfrac{j_{p-1}(k_o a_1)}{j_{p-1}(k_i a_1)} \end{pmatrix}, \qquad \mathbf{E}^{(2)s} = \begin{pmatrix} E_{|s|}^s \dfrac{j'_{|s|}(k_o a_1)}{j'_{|s|}(k_i a_1)} \\ \vdots \\ E_{p-1}^s \dfrac{j'_{p-1}(k_o a_1)}{j'_{p-1}(k_i a_1)} \end{pmatrix} \tag{54}$$

其中 $s=0, \pm1, \ldots, \pm(p-1)$。

综合这些定义，每个阶数 s 对应的耦合系统采用分块矩阵形式：

$$\begin{pmatrix} -\mathbf{\Gamma}_o^{(1)s} & \mathbf{I} & \mathbf{\Gamma}_i^{(1)s} & (S|R)^s(\mathbf{r}'_{31}) \\ -\mathbf{\Lambda}_o^{(1)s} & \mathbf{I} & \mathbf{\Lambda}_i^{(1)s} & (S|R)^s(\mathbf{r}'_{31}) \\ \mathbf{0} & \mathbf{I} & \mathbf{\Gamma}_i^{(2)s} & (S|R)^s(\mathbf{r}'_{31}) \\ \mathbf{0} & -(R|R)^s(\mathbf{r}'_{13}) & -(S|R)^s(\mathbf{r}'_{13}) & \mathbf{\Gamma}_i^{(3)s} \end{pmatrix} \begin{pmatrix} \mathbf{A}^{(1)s} \\ \mathbf{B}^{(1)s} \\ \mathbf{C}^{(1)s} \\ \mathbf{D}^{(3)s} \end{pmatrix} = \begin{pmatrix} \mathbf{E}^{(1)s} \\ \mathbf{E}^{(2)s} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \tag{55}$$

每个行块对应一个边界条件：前两行对 S_1 施加透射条件（声压和速度连续性），第三行对 S_2 施加诺伊曼条件，第四行则对 S_3 施加该条件。对于每个阶数 s，该系统有 $4(p-|s|)$ 个方程和 $4(p-|s|)$ 个未知数。

系统（55）对每个阶数 $s = 0$、±1、.....、$\pm(p-1)$ 分别独立求解。我们采用 LU 分解：每个阶数的系统矩阵只需分解一次，随后将分解因子重复用于对应不同频率和入射场配置的多个右端项。每个阶数的计算成本为 $\mathcal{O}\big((p-|s|)^3\big)$，主要来自矩阵分解 [17]。

B 基于角加速度的噪声建模

状态转移模型假设相邻时间步之间的角速度恒定。然而，真实运动声源会因角加速度而产生平滑的速度变化。这种不确定性代表过程噪声，并被建模为角加速度 [29]。

考虑时间上的连续角运动由下式描述

$$\ddot{\theta}(t) = a_\theta(t), \qquad \ddot{\phi}(t) = a_\phi(t), \tag{56}$$

其中 $a_\theta(t)$和 $a_\phi(t)$为零均值高斯白加速度过程，方差分别为$\sigma^2_{\theta,\mathrm{acc}}$ 和$\sigma^2_{\phi,\mathrm{acc}}$。假设各角维度上的加速度相互独立。

在采样间隔 Δt 上离散化，得到随机状态更新：

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t, \tag{57}$$

其中过程噪声协方差矩阵$\mathbf{Q}$ 由当前加速度噪声导出，并按如下方式计算：

$$\mathbf{Q} = \begin{bmatrix} \frac{\Delta t^4}{4}\sigma^2_{\theta,\mathrm{acc}} & 0 & \frac{\Delta t^3}{2}\sigma^2_{\theta,\mathrm{acc}} & 0 \\ 0 & \frac{\Delta t^4}{4}\sigma^2_{\phi,\mathrm{acc}} & 0 & \frac{\Delta t^3}{2}\sigma^2_{\phi,\mathrm{acc}} \\ \frac{\Delta t^3}{2}\sigma^2_{\theta,\mathrm{acc}} & 0 & \Delta t^2\sigma^2_{\theta,\mathrm{acc}} & 0 \\ 0 & \frac{\Delta t^3}{2}\sigma^2_{\phi,\mathrm{acc}} & 0 & \Delta t^2\sigma^2_{\phi,\mathrm{acc}} \end{bmatrix}. \tag{58}$$

加速度方差$\sigma_{\theta,\mathrm{acc}}$ 和$\sigma_{\phi,\mathrm{acc}}$ 控制运动模型的灵活性。较大的值允许更快的轨迹变化，而较小的值则强制更平滑的运动。

C 扩展卡尔曼滤波器更新方程的推导

跟踪移动声源的EKF更新方程推导如下 [29]。测量模型为

$$\mathbf{z}_t = h(\mathbf{x}_t) + \mathbf{v}_t, \tag{59}$$

其中， $h(\cdot)$ 是将声源角度映射到堆叠的耳间声级差和耳间时间差向量的非线性正向头相关传递函数模型，并且

$$\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

是均值为零的高斯测量噪声。

在时间步 t，预测状态和协方差记为

$$\hat{\mathbf{x}}_{t|t-1}, \qquad \mathbf{P}_{t|t-1}.$$

由于 $h(\mathbf{x})$ 关于 (θ, ϕ) 是非线性的，我们在预测状态附近使用一阶泰勒展开对其进行近似：

$$h(\mathbf{x}_t) \approx h(\hat{\mathbf{x}}_{t|t-1}) + \mathbf{H}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}), \tag{60}$$

其中，雅可比矩阵为

$$\mathbf{H}_t = \left.\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{t|t-1}}. \tag{61}$$

定义新息（测量残差）

$$\mathbf{y}_t = \mathbf{z}_t - h(\hat{\mathbf{x}}_{t|t-1}). \tag{62}$$

代入线性化模型可得

$$\mathbf{y}_t \approx \mathbf{H}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \mathbf{v}_t. \tag{63}$$

这对应于一个局部线性测量方程。新息协方差 is

$$\mathbf{S}_t = \mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^T + \mathbf{R}. \tag{64}$$

使后验均方估计误差最小的最优卡尔曼增益为 s

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^T\mathbf{S}_t^{-1} = \mathbf{P}_{t|t-1}\mathbf{H}_t^T\left(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^T + \mathbf{R}\right)^{-1}. \tag{65}$$

后验状态估计更新为

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\left(\mathbf{z}_t - h(\hat{\mathbf{x}}_{t|t-1})\right). \tag{66}$$

后验协方差可写成简化形式

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}. \tag{67}$$

为提高数值稳定性，也可以使用约瑟夫形式：

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)^T + \mathbf{K}_t\mathbf{R}\mathbf{K}_t^T. \tag{68}$$

注意：由于声学测量函数仅取决于状态向量的角度分量（θ， ϕ）

$$\mathbf{x}_t = \begin{bmatrix}\theta_t & \phi_t & \dot{\theta}_t & \dot{\phi}_t\end{bmatrix}^T,$$

雅可比矩阵具有分块结构

$$\mathbf{H}_t = \left[\begin{array}{cccc}\dfrac{\partial h}{\partial \theta} & \dfrac{\partial h}{\partial \phi} & \mathbf{0} & \mathbf{0}\end{array}\right]_{\hat{\mathbf{x}}_{t|t-1}}. \tag{69}$$

作者贡献

Siminfar Samakoush Galougah 建立了分析框架，实现了主要仿真，开展了实验，并主导了文稿的撰写。

Pranav Pulijala 参与了仿真代码的开发与实现、软件调试、部分实验的开展、手稿部分内容的撰写，以及技术讨论和手稿修订。

Ramani Duraiswami 构思了本研究，提供了概念指导，获得了研究经费，并审阅和编辑了手稿。

伦理

本研究未在现实环境中对人类或动物开展实验。

术语表

指向性因子：指向性因子。**指向性指数**：指向性指数。**扩展卡尔曼滤波器**：扩展卡尔曼滤波器。**头相关传递函数**：头相关传递函数。**双耳声级差**：耳间声级差。**耳间时间差**：耳间时间差。**白噪声增益**：白噪声增益。

参考文献

[1] Jens Blauert。《空间听觉：人类声音定位的心理声学》。《空间听觉：人类声音定位的心理声学》。麻省理工学院出版社，1997。[2] Richard O. Duda 和 William L. Martens。球形头部模型响应的距离依赖性。《美国声学学会会刊》，104(5):3048–3058，1998。[3] N. A. Gumerov、R. Duraiswami 和 Zhihui Tang。躯干对头相关传递函数影响的数值研究。*2002 IEEE*国际会议，2002。[4] V Ralph Algazi、Richard O Duda、Ramani Duralswami、Nail A Gumerov 和 Zhihui Tang。利用头部和躯干的简单几何模型近似头相关传递函数。美国声学学会期刊，112(5 Pt 1):2053–2064，2002年11月。[5] N Gumerov 和 R Duraiswami。邻近边界对头相关传递函数影响的建模。*2001 IEEE*国际声学、语音与信号处理会议论文集（目录号*01CH37221*），5:3337–3340，第5卷，2001年5月。[6] Whitlow WL Au。《海豚的声呐》。Springer Science & Business Media，1993。[7] Darlene R. Ketten。鲸类耳朵。载于 Whitlow W. L. Au、Arthur N. Popper 和 Richard R. Fay 编，《鲸和海豚的听觉》，第43–108页。施普林格，纽约，2000。[8] Kenneth S. Norris。齿鲸声学机制的进化。载于 E. T. Drake 编，《进化与环境》，第297–324页。耶鲁大学出版社，1968。[9] Ted W. Cranford、Mats Amundin 和 Kenneth S. Norris。齿鲸鼻部复合体的功能形态学与同源性：对声音产生的启示。形态学杂志，228:223–285，1996。[10] Heather N. Koopman、Suzanne M. Budge、Darlene R. Ketten 和 Sara J. Iverson。齿鲸下颌脂肪体内脂质的地形分布：显著的复杂性与一致性。*IEEE*海洋工程学报，31(1):95–106，2006。

[11] Randall L. Brill。颌部听觉海豚：初步行为与声学证据。载于 Paul E. Nachtigall 和 Patrick W. B. Moore 编，《动物声呐：过程与性能》，第281–287页。普林纽姆出版社，纽约，1988。[12] Ted W. Cranford、Petr Krysl 和 Mats Amundin。通向齿鲸耳朵的新声学通道，以及鼓泡骨复合体的振动分析。《公共科学图书馆·综合》，5(8):e11927，2010。[13] Ted W. Cranford、Petr Krysl 和 John A. Hildebrand。揭示的声学通路：柯氏喙鲸（Ziphius cavirostris）声音传输与接收的模拟。《生物启发与仿生学》，3:016001，2008。[14] J L Aroyan。Delphinus delphis 听觉的三维建模。美国声学学会期刊，110(6):3305–3318，2001年12月。[15] Patrick W. B. Moore、Deborah A. Pawloski 和 Lindsay Dankiewicz。宽吻海豚（Tursiops truncatus）的双耳时间差和强度差阈值。载于 Ronald A. Kastelein、Jeanette A. Thomas 和 Paul E. Nachtigall 编，《水生哺乳动物的感觉系统》，第11–25页。De Spil出版社，伍尔登，1995。[16] Nail A. Gumerov 和 Ramani Duraiswami。《三维亥姆霍兹方程的快速多极子方法》。爱思唯尔，2004。[17] Nail A Gumerov 和 Ramani Duraiswami。利用多极子再展开计算N个球的散射。美国声学学会期刊，112(6):2688–2701，2002年12月。[18] James Bradbury、Roy Frostig、Peter Hawkins、Matthew James Johnson、Chris Leary、Dougal Maclaurin、George Necula、Adam Paszke、Jake VanderPlas、Skye Wanderman-Milne 等。JAX：Python语言+ NumPy程序的可组合变换。2018。[19] L. W. Cai、D. K. Dacol、G. J. Orris、D. C. Calvo 和 M. Nicholas。含电流变层多层球形弹性散射体的声散射。《美国声学学会会刊》，129(1):12–23，2011。[20] Allan D Pierce。《声学：物理原理及应用导论》。施普林格，2019。[21] Yuvi Kahana、Philip A Nelson、Maurice Petyt 和 Sunghoon Choi。仿真人头及外耳传递函数的数值建模。载于音频工程学会会议：第 *16* 届国际会议：空间声音再现。音频工程学会，1999。[22] Deniz Oktay。《通过自动微分连接科学计算与机器学习》。博士论文，普林斯顿大学，2024。[23] Siminfar Samakoush Galougah 和 Ramani Duraiswami。应用自动微分优化差分麦克风阵列设计，2024。[24] Nail A. Gumerov 和 Ramani Duraiswami。利用多极子再展开计算n个球的散射。载于《美国声学学会会刊》，第112卷，第2688–2701页，2002。[25] Diederik P. Kingma 和 Jimmy Ba。Adam：一种随机优化方法。载于 Yoshua Bengio 和 Yann LeCun 编，第三届国际学习表征会议（*ICLR 2015*），美国加利福尼亚州圣迭戈，*2015*年*5*月*7*日至*9*日，会议论文集，2015。[26] Harry L. Van Trees。《最优阵列处理：检测、估计与调制理论第四部分》。Wiley，纽约州纽约，2002。[27] Robert J. Urick。《水下声学原理》。麦格劳-希尔，纽约州纽约，第3版，1983。[28] Rudolf E. Kalman。线性滤波与预测问题的一种新方法。《基础工程学杂志》，82(1):35–45，1960。

[29] Dan Simon。 最优状态估计：卡尔曼、H无穷与非线性方法。Wiley，霍博肯，新泽西州，2006。[30] Yaakov Bar-Shalom、X. Rong Li 和 Thiagalingam Kirubarajan。 估计及其在跟踪与导航中的应用。Wiley，纽约，纽约州，2001。