arXiv:2305.04447v4 [eess.AS] 1 Mar 2024

神经转向器：利用频率和方向上的因果神经场进行新型导向矢量合成

Diego Di Carlo[1,2] *Aditya Arie Nugraha*[1,2] *Mathieu Fontaine*[3,1] *Yoshiaki Bando*[4,1] *Kazuyoshi Yoshii*[2,1]

[1]RIKEN先进智能项目（AIP），RIKEN，日本[2]京都大学信息学研究生院，日本[3]LTCI，T´é´lecom Paris，巴黎综合理工学院，法国[4]日本产业技术综合研究所（AIST），日本

摘要

我们针对如何利用一种称为神经场的深度学习框架，准确地对测量得到的无回声室导向矢量进行插值这一问题展开研究。对于精确的声源分离和定位而言，这项任务在减少资源密集型测量方面发挥着关键作用，而声源分离和定位是语音识别前端不可或缺的环节。传统的插值方法依赖于在固定离散频率集合上，对空间中邻近测量值进行线性加权。受计算机视觉领域神经场在新视角合成方面成功应用的启发，我们提出了神经转向器，这是一种以频率和方向为输入、输出相应导向矢量的连续复值函数。重要的是，它纳入了通道间相位差的信息以及用于强制满足滤波器因果性的正则化项，这对于准确建模导向矢量至关重要。我们使用真实测量导向矢量数据集开展实验，结果表明，我们的无分辨率限制模型能够有效地对这类测量值进行插值。

索引词— 导向向量、神经场、空间音频、插值、表征学习

1. 引言

导向向量是多通道音频信号处理中的基本概念，用于描述消声环境下声源与一组麦克风之间的声学关系。它是语音增强 [1]，声源分离 [2]和定位 [3], 以及声学信道估计 [4]的核心组成部分。因此，准确表征导向向量对于实现稳健的声音分析和真实感渲染至关重要。在大多数实际应用中，导向向量基于代数消声模型计算，或通过测量头部相关传递函数（HRTF）、麦克风阵列的指向性特征以及声学传递函数来估计。

代数导向向量以解析方式编码空间和频率上的直接传播，通常是声波入射角的函数。然而，在实际场景中，这种代数模型会受到多种因素的限制，通常会被测得的通用滤波器取代 [5]。扩展公式还包括指向性和滤波模型，以及声音与接收器的相互作用模型，例如遮挡、衍射和散射。在助听器应用中，导向矢量编码了 HRTF，捕捉用户的

本研究得到 ANR SAROUMANE项目（ANR- 22-CE23-0011）、Hi! Paris MASTER-AI项目、JST PRESTO（编号：JPMJPR20CB）以及 JSPS KAKENHI（编号：JP20H00602、JP21H03572、JP23K16912、JP23K16913）的资助。代码见 https://github.com/Chutlhu/nsteerer。

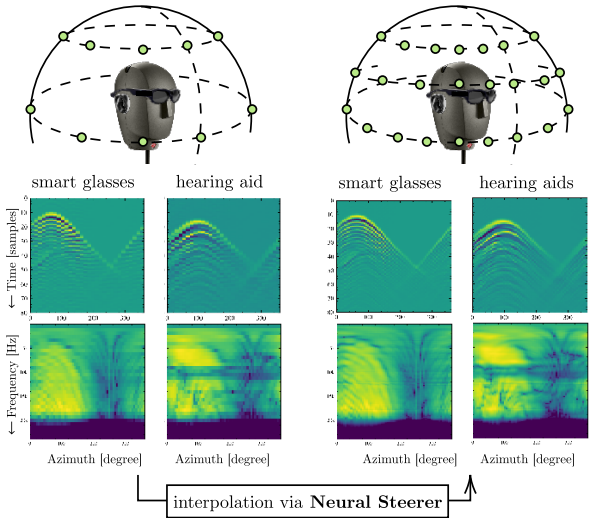

图1：所提出的用于测量所得消声导向矢量插值的神经转向器示意图。

耳廓、头部和躯干的影响。为了对每个空间位置和频率下的所有影响进行建模，可以使用专用模拟器，但这类模拟器在真实度与计算复杂度之间存在显著权衡 [6]。另一种方法是在专用设施中测量导向矢量 [1]。然而，由于成本高、设置复杂，要以高空间分辨率进行测量十分繁琐，甚至可能无法实现。

过去十年间，可靠的数据驱动方法在导向向量和头部相关传递函数插值方面受到广泛关注 [7–9]。如图1所示，这些方法以粗略空间网格上的测量值为基础，再通过插值获得新位置（通常位于更精细网格上）的测量估计值。然而，这类方法有两个主要缺点。首先，它们假设目标量在所需空间位置处是某个基函数的多项式变换，这可能无法准确反映现实情况，并引入伪影。其次，这些方法仅在空间域进行插值，频率（或时间）网格的分辨率保持不变，因此对目标量的建模不够充分。

近期深度学习研究表明，基于坐标的架构能够连续地编码时空信号 [10–12]。这些方法也称为神经场（NFs），可以通过数据拟合损失函数和基于模型的损失函数轻松训练，从而以任意分辨率对目标量进行插值和超分辨率处理。其出色的应用包括从稀疏二维图像合成新视角 [13], 图像超分辨率 [11], 以及人体动画—

化 [14]。与特定分辨率下基于网格的离散测量不同，神经场的内存复杂度随数据复杂度变化，而非随所需分辨率变化。此外，作为过参数化的模块化网络模型，神经场在适当正则化的条件下，能够有效回归适定性不佳的问题。然而在实践中，它们往往会过拟合，其分辨率能力也受网络容量和训练方案的限制 [12]。

有趣的是，物理信息神经网络 （PINNs） [15–17] 是计算物理领域提出的类似模型。在此，利用偏微分方程（PDE）作为正则化项，通过自动微分在任意分辨率下对坐标进行评估。由此得到的神经场很可能是定义于整个输入域上的、符合物理规律的连续函数，也不易对嘈杂或稀疏的输入观测过拟合。然而，这些神经网络需要相对于偏微分方程中使用的输入变量可微，而现有的基于音频的神经场模型返回的是离散频率集合，因此并不具备这一特性。

在本研究中，我们借助神经场领域的最新进展，解决现有离散导向向量插值方法的局限。所提出的方法将导向向量（与头相关传递函数相关）的幅度和相位建模为一个在空间和频率上连续定义的复值场。基于信号处理的理论基础，我们提出了一种新颖的正则化项，用于约束估计出的滤波器满足因果性，从而相比标准方法获得更准确的滤波器。最后，得益于其在空间和时间上的连续形式，所提出的方法可用于物理信息神经网络框架，根据模型输入计算波动方程中的各项。

2. 相关工作

给定频域或时域中的一组导向矢量测量，提高空间分辨率的一种简单方法，是对多个周围位置的已知测量值进行加权平均，以估计新位置的值。这类方法是多项式插值的扩展：将数据投影到特设基函数（例如球谐函数或空间特征函数）上，再对这些基函数的系数进行插值，从而得到所需位置 [7–9] 处的值。这些方法众所周知的局限性包括：训练数据需要近乎均匀采样，而且难以在优化中纳入特设先验信息。此外，性能取决于基函数的选择及其插值算法。

在音频处理中，近年来已有研究提出使用神经场（NF）进行声学脉冲响应（AIR）的估计与插值 [18], 、在任意球坐标下进行头相关传递函数幅度编码 20], [19, [21 23] [20], ，以及双耳渲染——。有趣的是，在估计 AIR 时，会显式纳入早期声音传播的代数模型。然而，这些方法将频率（或时间）视为与神经网络输出维度相对应的离散量，从而限制了模型的适用性。

将频率（或时间）显式设为神经网络的输入变量后，模型便可在 PINN 框架内进行训练，并利用波动方程确保物理一致性。这种方法已用于恢复未见位置处的时域 AIR [24] ，以及针对给定的单个频率频点对复数 HRTF 进行空间超分辨率重建 [25]。尽管取得了令人鼓舞的结果，但估计出的滤波器形状并无保证，例如可能包含反因果分量。此外，值得注意的是，除了 [21,25], ，导向向量的相位分量也被忽略了。这一局限会显著影响空间处理和渲染，因为准确处理离不开相位信息。

3. 所提方法

从理论上计算导向向量入手是合理的。考虑一个消声导向向量，它将位于位置 $\mathbf{s}_j$ 的第 j 个声源到以 $\mathbf{r}$ 为中心的 I 麦克风阵列之间的直接信号传播编码为入射方向（DoA）的函数，该方向由方位角 θ_j 和仰角 φ_j 表示。

根据频域中的远场代数模型，位于$\mathbf{m}_i$ 处的第 i个麦克风在频率 f 下的消声导向向量 [26]，$d_{ij}(f)$表示为

$$d_{ij}(f) = \exp\left(-\frac{\jmath 2\pi f \mathbf{n}_j^\top(\mathbf{m}_i - \mathbf{r})}{c}\right), \tag{1}$$

其中，$\mathbf{n}_j = [\cos\,\theta_j \cos\,\varphi_j, \sin\theta_j \cos\,\varphi_j, \sin\varphi_j]^\top$ 是指向第 j 个声源的单位范数向量，c 是声速，$\jmath$ 表示虚数单位。

由于空气中的信号传播和麦克风指向性 [27]等复杂物理现象引发的各种滤波效应，实际导向向量会偏离理想的消声导向向量。因此，我们将导向向量 $h_{ij}(f)$表示为 $d_{ij}(f)$的修正形式，如下所示：

$$h_{ij}(f) = \underbrace{\exp(-2\pi f \jmath\tau)}_{\triangleq \bar{d}(f)}\, g_j^{\text{air}}(f)\, g_{ij}^{\text{mic}}(\mathbf{m}_i, f)\, d_{ij}(f), \tag{2}$$

其中，$g_j^{\text{air}}(f)$表示与频率相关的空气衰减以及最终与声源相关的特性；$g_{ij}^{\text{mic}}(\mathbf{m}_i, f)$表示麦克风指向性，其相位中心定义在$\mathbf{r}$处，通常取决于麦克风的类型和制造工艺 [28]。最后，全局时延项 $\bar{d}(f)$用于补偿全局恒定时间偏移量 τ，该偏移量常见于真实世界的测量中。

3.1. 神经转向器

导向矢量族可表示为以下函数：

$$\begin{aligned} \mathcal{M}: \quad & \mathbb{S}^2 \quad \times \mathbb{R}_+ \to \mathbb{C} \\ & ((\theta_j, \varphi_j), \quad f) \mapsto h_{ij}, \end{aligned} \tag{3}$$

其中，$\mathbb{S}^2$ 是单位球面上的极坐标集合。为便于阅读，省略了对$\mathbf{m}_i$ 的依赖，因为它们是常数。

我们使用一个神经场 [10–12], 对 $\mathcal{M}$ 进行参数化。该神经场在有限观测集上训练后，便可在任意分辨率下评估任意输入坐标；其超分辨率能力取决于神经网络的归纳偏置，例如架构拓扑和正则项。与 [21], 类似，我们将 τ 和$\mathbf{m}_i$ 建模为模型的自由参数，并在训练期间对其进行优化。随后，可根据输入坐标使用公式（1）以代数方式计算 $\bar{d}$ 和 d_{ij}。因此，神经网络仅预测项 $g_j^{\text{air}}(f)$和 $g_{ij}^{\text{mic}}(\mathbf{m}_i, f)$。这种受物理启发的建模方式很有帮助，因为它施加了归纳偏置并提供了良好的初始化，尽管各个项未必对应实际的物理贡献。

3.2. 网络架构

所提出的架构如图 2 所示。与 [19], 一样，这里使用 SIREN 网络，即具有正弦激活函数的多层感知机（MLP） [11]。该模型记为 $\text{SIREN}_{\text{Phase}}$，以期望的声源到达方向（$\theta_j$、$\varphi_j$）和频率 f为输入，并返回式 (2) 中的分量：$\mathcal{G}_j \triangleq \left\{g_j^{\text{air}}(f)\, \{g_{ij}^{\text{mic}}(\mathbf{m}_i、\ f)\}_{i=1}^{I}\right\}$。

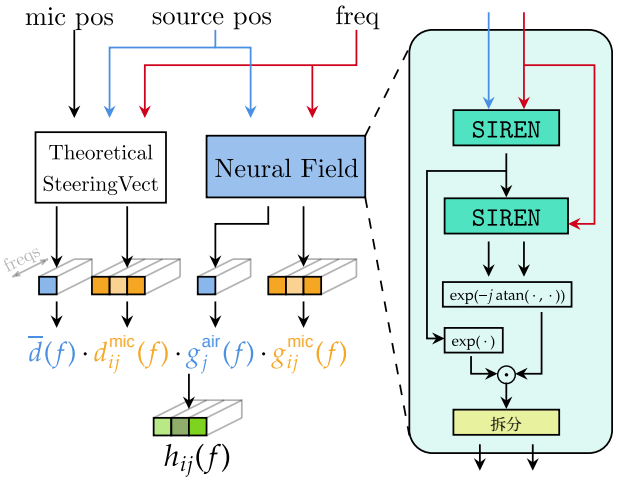

图 2：所提出的神经转向器示意图。 $\odot$ 和 $\cdot$ 均表示逐元素乘法。

设 g^* 是 $\mathcal{G}_j$的一个元素。我们不直接预测复数值的 g^* 或其实部和虚部，而是提出隐式估计其幅度和相位。更准确地说，对于每个 g^*，神经网络都会返回一个三维向量 $\mathbf{g}^* \in \mathbb{R}^3$，并利用其幅度和相位得到$g^*$，如下所示：

$$g^* = \exp\left(\{\mathbf{g}^*\}_1\right) \exp\left(-\jmath 2\pi \arctan\left(\frac{\{\mathbf{g}^*\}_2}{\{\mathbf{g}^*\}_3}\right)\right). \tag{4}$$

这种表示方式处理了幅相格式所遭受的相位缠绕问题；与实部-虚部表示相比，它还提高了稳定性和收敛速度，从而取得了更好的结果。

给定由神经场估计的 $g_j^{\text{air}}(f)$ 和 $g_{ij}^{\text{mic}}(\mathbf{m}_i, f)$，以及通过已知麦克风位置和偏移量分别计算得到的理论导向向量 $d_{ij}^{\text{mic}}(f)$ 和全局时延 $d_j(f)$，导向向量 $h_{ij}(f)$ 随后按照式 (2) 获得。

继承 [29], 中的工作，我们提出了一种替代架构（记为 SIREN$^{\text{Mag}\to\text{Phase}}$），其中相位估计以幅度估计为条件。实际上，该神经网络由两个级联的 SIREN 组成，并进行联合训练：第一个 SIREN 根据输入坐标输出各分量的幅度，第二个 SIREN 则基于坐标和幅度重建相位。

该模型将频率视为连续输入变量。在这种设置下，以下称为连续频率（CF）模型，神经网络输出 $3 \times (I+1)$ 个实数，其中 3 是$\mathbf{g}^*$的维度，用于表示复数值。在离散频率（DF）建模中，神经网络的最后一层经过修改，可针对给定的单声源到达方向输出 $3 \times (I+1)F$，其中 F 是在 $[0, F_s]$ 中按规则采样的频率总数，类似于离散傅里叶变换操作，其中 F_s 是采样频率。

3.3. 训练损失与离网格正则化

训练神经网络的目标是，对于给定的声源到达方向（θ_j，φ_j）和频率 f，返回在频域和时域中相位失真都较低的滤波器。与 [30]，中提出的损失函数类似，我们优化频域中的显式幅度误差和相位误差，以及一个显式的时域损失项 ℓ_2，即

$$\mathcal{L} = \sum_{ij \in \mathcal{B}} \left(\frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} \mathcal{L}_{ijf}^{\text{freq}} + \frac{1}{F} \mathcal{L}_{ij}^{\text{time}} \right) \tag{5}$$

$$\mathcal{L}_{ijf}^{\text{freq}} = \mathcal{L}_{\text{LogMag}}(\hat{h}_{ij}(f), h_{ij}(f)) + \lambda_1 \mathcal{L}_{\text{Phase}}(\hat{h}_{ij}(f), h_{ij}(f))$$

$$\mathcal{L}_{ij}^{\text{time}} = \lambda_2 \left\| \text{iDFT}\left([\hat{h}_{ij}]_F\right) - \text{iDFT}\left([h_{ij}]_F\right) \right\|_2^2,$$

其中，$\mathcal{B}$和 $\mathcal{F}$分别是一个批次中使用的随机训练到达方向和频率集合，且 $|\mathcal{F}| < F$。 $[h_{ij}]_F$表示在 $[0, F_s]$ 中等间隔频率处的 F 个元素的串联。

实际上，正如 [31], 所报告的，我们还发现，在对数幅度谱上使用 ℓ_1 损失，并结合对相位分量的余弦和正弦使用 ℓ_1 损失，效果最佳，分别记为 $\mathcal{L}_{\text{LogMag}}$ 和 $\mathcal{L}_{\text{Phase}}$。此外，我们通过实验证实，与将频率建模为离散输出变量相比，对连续频率（CF）进行建模要困难得多。具体而言，我们发现神经场无法捕捉频率维度上的细节，最终收敛到错误的解。为解决这一问题，我们修改了损失函数 $\mathcal{L}_{\text{LogMag}}$ 和 $\mathcal{L}_{\text{Phase}}$，以考虑频率的序列性质。实际上，这些损失会根据先前频率的累积残差损失 [32]进行反向加权。由于篇幅限制，具体细节不再赘述。

式 (5) 中损失函数的主要局限在于，它只能在训练数据上计算。借鉴 PINNs 中采用的训练策略，我们可以使用物理模型，以无监督的方式在任意分辨率下针对声源位置对目标函数进行正则化 [17, 33]。其中一个主要要求是，对于每个声源位置，我们估计的 $\hat{}$ 导向向量 h_{ij} 及其分量都必须是因果的。反因果分量可能导致伪影和不连贯的导向矢量。为此，我们提出了一种正则化项，它基于一条经典关系：传递函数的虚部必须是其实部的希尔伯特变换 $\mathcal{H}$ [34]:

$$\mathcal{L}_{\text{Causal}} = \left\| \mathcal{H}\left(\Re\left\{[\hat{h}_{ij}]_F\right\}\right) - \Im\left\{[\hat{h}_{ij}]_F\right\} \right\|_2^2, \tag{6}$$

其中 $\Re\{\cdot\}$ 和 $\Im\{\cdot\}$ 分别为实部和虚部。在离散频率情形下，通过在 $\mathbb{R}^3$中随机采样声源位置，可以直接计算该正则化项。在连续频率条件下，则需要对输入频率进行采样。为简化希尔伯特变换和傅里叶变换的计算，我们选择在 $[0, F_s]$中随机采样若干个等间隔分布的频率。

4. 评估

我们的目标是在时域和频域中获得可靠的导向向量表示。因此，我们采用三项指标评估所提出的 SIREN$_{\text{Phase}}$ 和 SIREN$^{\text{Mag}\to\text{Phase}}$ 模型的性能，即均方根误差 (RMSE)和余弦距离，用于衡量时域中估计导向向量与参考导向向量之间的差异；以及对数谱距离 (LSD) [dB]，用于衡量估计头相关传递函数与参考头相关传递函数的幅度差异。我们将模型与 [19] 中使用的原始 SIREN（SIREN*）进行比较，该模型经过修改后可输出幅度和相位；此外，还与一种基于空间特征函数（SCF*）的插值方法 [8]进行比较。

我们的评估使用了 EasyCom 数据集 [1]，其中包含一个 6 通道麦克风阵列的导向矢量，该阵列由安装在头戴式智能眼镜上的 4 个麦克风和位于用户耳道内的 2 个双耳麦克风组成。所有数据均在 F_s =48 千赫兹下，于一个包含 60 个等间距方位角的球面网格上测量，并且

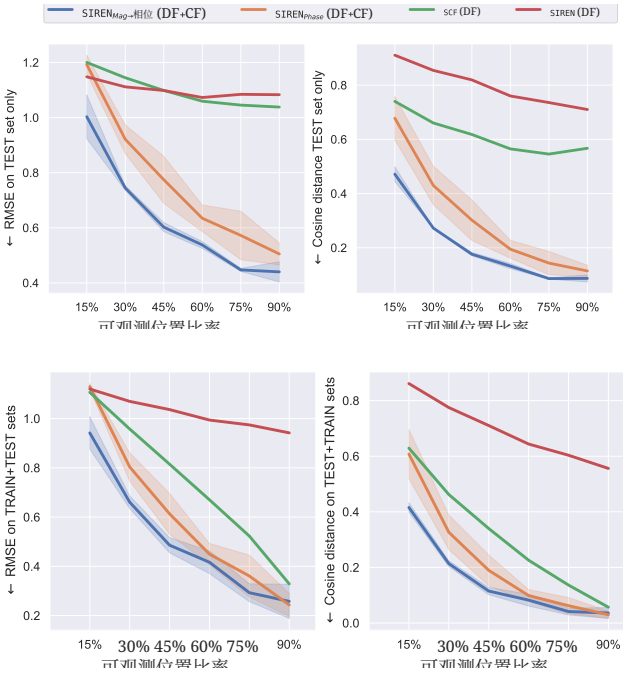

图 3：测试集中的位置（上）和随机数据上的整个球面重建（下）对应的时域 steer- ing 向量的均方根误差（左）和余弦距离（右）。阴影区域表示连续频率模型和离散频率模型的置信区间。

17 个近乎等间距的仰角。除非另有说明，我们考虑 $F = 257$ 正频率。

我们的 SIREN$_{\text{Phase}}$ 和 SIREN$^{\text{Mag}\to\text{Phase}}$ 模型采用了由四个 512 维隐藏层组成的 SIREN 架构。此外，SIREN$^{\text{Mag}\to\text{Phase}}$ 还使用了另一个包含两个 512 维隐藏层的 SIREN。这些模型使用大小为 $|\mathcal{B}| = 18$ 的批次进行训练，学习率初始化为 10^{-3}，并在每个 epoch 以 0.98 的因子进行缩放。我们采用早停机制，通过监控在部分（20%）训练数据上计算的损失来避免过拟合。在所有实验中，我们将 $\lambda_1 = \lambda_2 = 10$ 设置为与不同损失项的尺度相匹配。$\mathcal{L}_{\text{Causal}}$ 使用在单位球内均匀采样的 B 个随机坐标计算得到。基线 SIREN* 使用了应用于我们的 SIREN$_{\text{Phase}}$ 模型的相同参数化方式，但仅直接输出幅度和相位，不进行公式（4）中的变换。

我们首先考虑在规则网格上对导向向量进行插值的任务。在该任务中，训练数据集由按因子2定期下采样的导向向量和位置组成，如 [19]所示。表1报告了在完整频率分辨率下，未见（缺失）数据点的平均均方根误差和余弦距离。使用所提出正则项训练的模型在这两项指标上均优于基线模型。具体而言，通过引入新的正则化技术，并根据幅度对相位估计进行条件约束，模型性能得到了提升。此外，在连续频率建模的背景下，加入离网格正则项 $\mathcal{L}_{\text{Causal}}$ 后，得到的重建结果与通过离散频率建模获得的结果相当（p值为0.0032）。

随后，我们分析了可用点数量（占比）对重建能力的影响，这些点是从原始网格中随机采样得到的。图 3 展示了两个时域目标函数在仅包含未见数据（上图）以及完整测试数据集（下图）上的结果；完整测试数据集包括训练数据点和未见数据点。这些曲线再次印证了前文讨论的结果，表明以幅度估计为条件进行相位估计

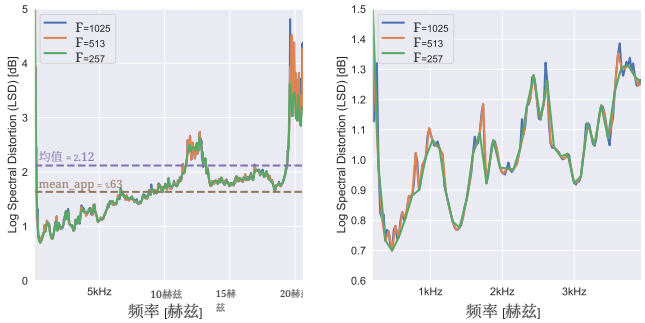

图4：不同分辨率下整个频率范围（左）和选定频率范围（右）的平均对数谱失真。模型使用 $F = 257$ 个频率箱进行训练。

表1：导向向量插值任务中，估计的时域导向向量与参考时域导向向量之间的平均均方根误差和余弦距离（括号内）。数值为 6 个通道的平均值。* 表示基线方法。

频率	模型	$\mathcal{L}_{\text{MSE}}$	$\mathcal{L}_{\text{LogMag}} + \lambda_1\mathcal{L}_{\text{Phase}}$	$+\lambda_2\mathcal{L}_{\text{iDFT}}$	$+\mathcal{L}_{\text{Causal}}$
DF	SCF*	1.42 (0.94)	-	-	-
DF	SIREN*	1.09 (0.78)	-	1.02 (0.70)	**1.01 (0.73)**
DF	SIREN$_{\text{Phase}}$	1.26 (0.77)	0.36 (0.06)	0.34 (0.06)	**0.34 (0.06)**
DF	SIREN$^{\text{Mag}\to\text{Phase}}$	1.22 (0.66)	0.32 (0.05)	0.29 (0.05)	**0.28 (0.05)**
CF	SIREN$_{\text{Phase}}$	1.42 (0.88)	0.63 (0.17)	0.58 (0.17)	**0.33 (0.06)**
CF	SIREN$^{\text{Mag}\to\text{Phase}}$	1.37 (0.80)	0.53 (0.13)	0.45 (0.09)	**0.37 (0.08)**

能取得更好的效果。值得注意的是，可以看到，仅以最小化复数域中的 ℓ_2 范数为训练目标的 SIREN* 模型，在重建已见点和未见点时表现都不理想，即使训练期间已见点占比达到90%。

最后，我们分析了沿频率轴的插值性能。具体而言，我们训练了一个 CF 变体的 SIREN$^{\text{Mag}\to\text{Phase}}$模型，以拟合在 $F = 257$ 个频率下测得的导向矢量，然后在不同频谱分辨率下进行评估。图4所示结果表明，在不同评估网格下，重建误差始终保持较低。不过，需要注意的是，误差会随频率升高而增大，且在边界处更为明显。这一现象可能归因于目标信号的带限特性，也与 [19]中报告的先前发现一致。尽管如此，在经典语音处理应用通常使用的频率范围内，即 $f \in [40, 20000]$ 赫兹，平均对数谱距离约低 0.5 dB。

5. 结论

本文提出了一种新颖的神经场模型，该模型能够基于连续编码，在空间和频率坐标上对导向向量进行连续表示，而输入仅为一组离散测量值。我们的方法特别强调准确捕获目标导向向量的相位分量，同时确保滤波器保持其因果性。实验结果表明，我们的模型在重建真实导向向量测量值方面具有良好的有效性。展望未来，我们认为该模型具有广泛的潜在应用前景。具体而言，导向向量的连续合成有望应用于声源定位和分离任务中的波束形成。此外，当我们将麦克风位置优化为模型参数时，所提出模型的通用性也使其能够应用于麦克风校准。

6. 参考文献

[1] Jacob Donley、Vladimir Tourbabin、Jung-Suk Lee、Mark Broyles、Hao Jiang、Jie Shen、Maja Pantic、Vamsi Krishna Ithapu 和 Ravish Mehra，“EasyCom：支持噪声环境中轻松交流算法的增强现实数据集”，arXiv 电子预印本，2021，arXiv:2107.04174v2。[2] Kouhei Sekiguchi、Yoshiaki Bando、Aditya Arie Nugraha、Kazuyoshi Yoshii 和 Tatsuya Kawahara，“基于指向性感知联合可对角化空间协方差矩阵的快速多通道非负矩阵分解，用于盲源分离”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 28 卷，第 2610–2625 页，2020。[3] Ralph Schmidt，“多发射源定位与信号参数估计”，*IEEE* 天线与传播汇刊，第 34 卷，第 3 期，第 276–280 页，1986。[4] Paolo Annibale、Jason Filos、Patrick A Naylor 和 Rudolf Rabenstein，“温度变化条件下房间几何结构的几何推断”，载于国际控制、通信与信号处理研讨会论文集，2012，第 1–4 页。[5] Sharon Gannot、David Burshtein 和 Ehud Weinstein，“利用波束形成和非平稳性进行信号增强及其在语音中的应用”，*IEEE* 信号处理汇刊，第 49 卷，第 8 期，第 1614–1626 页，2001。[6] Sylvain Argentieri、Patrick Danes 和 Philippe Sou`eres，“机器人学中的声源定位综述：从双耳方法到阵列处理方法”，计算机语音与语言，第 34 卷，第 1 期，第 87–112 页，2015。[7] Ville Pulkki，“使用矢量基振幅声像定位进行虚拟声源定位”，音频工程学会期刊，第 45 卷，第 6 期，第 456–466 页，1997。[8] F´abio P. Freeland、Luiz W. P. Biscainho 和 Paulo S. R. Diniz，“头部相关传递函数（HRTF）的插值：一种多源方法”，载于*EUSIPCO* 论文集，2004。[9] Dmitry N. Zotkin、Ramani Duraiswami 和 Nail A. Gumerov，“使用球谐函数进行正则化头相关传递函数拟合”，载于*IEEE WASPAA* 论文集，2009，第 257–260 页。[10] Matthew Tancik、Pratul Srinivasan、Ben Mildenhall、Sara Fridovich-Keil、Nithin Raghavan、Utkarsh Singhal、Ravi Ra-mamoorthi、Jonathan Barron 和 Ren Ng，“傅里叶特征使神经网络能够学习低维域中的高频函数”，*NeurIPS* 论文集，第 33 卷，第 7537–7547 页，2020。[11] Vincent Sitzmann、Julien Martel、Alexander Bergman、David Lindell 和 Gordon Wetzstein，“具有周期性激活函数的隐式神经表示”，载于*NeurIPS* 论文集，2020，第 33 卷，第 7462–7473 页。[12] Yiheng Xie、Towaki Takikawa、Shunsuke Saito、Or Litany、Shiqin Yan、Numair Khan、Federico Tombari、James Tompkin、Vincent Sitzmann 和 Srinath Sridhar，“视觉计算及其他领域中的神经场”，载于计算机图形论坛，2022，第 41 卷，第 641–676 页。[13] Ben Mildenhall、Pratul P. Srinivasan、Matthew Tancik、Jonathan T. Barron、Ravi Ramamoorthi 和 Ren Ng，“NeRF：将场景表示为神经辐射场以实现视图合成”，载于*ECCV* 论文集，2020，第 405–421 页。[14] Guy Gafni、Justus Thies、Michael Zollhofer 和 Matthias Nießner，“用于单目 4D 面部化身重建的动态神经辐射场”，载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，2021，第 8649–8658 页。[15] Maziar Raissi、Paris Perdikaris 和 George E. Karniadakis，“物理信息神经网络：求解涉及非线性偏微分方程的正问题和逆问题的深度学习框架”，计算物理学杂志，2019。

[16] George Em Karniadakis、Ioannis G Kevrekidis、Lu Lu、Paris Perdikaris、Sifan Wang 和 Liu Yang，《物理信息机器学习》，*Nature Reviews Phys.*，第 3 卷，第 6 期，第 422–440 页，2021 年。[17] Sifan Wang、Yujun Teng 和 Paris Perdikaris，《理解并缓解物理信息神经网络中的梯度流病态问题》，*SIAM J. Sci. Comput.*，第 43 卷，第 5 期，第 A3055–A3081 页，2021 年。[18] Alexander Richard、Peter Dodds 和 Vamsi Krishna Ithapu，《深度脉冲响应：使用深度网络估计并参数化滤波器》，载于*Proc. IEEE ICASSP*，2022 年。[19] You Zhang、Yuxiang Wang 和 Zhiyao Duan，《HRTF 场：将实测 HRTF 幅度表示与神经场统一起来》，载于*Proc. IEEE ICASSP*，2023 年。[20] Jin Woo Lee、Sungho Lee 和 Kyogu Lee，《通过超条件特征的学习仿射变换进行全局 HRTF 插值》，载于*Proc. IEEE ICASSP*，2023 年。[21] Jin Woo Lee 和 Kyogu Lee，《用于双耳语音渲染的神经傅里叶移位》，载于*Proc. IEEE ICASSP*，2023 年。[22] Andrew Luo、Yilun Du、Michael J Tarr、Joshua B Tenenbaum、Antonio Torralba 和 Chuang Gan，《学习神经声场》，载于*Proc. NeurIPS*，2022 年，第 1–13 页。[23] Kun Su、Mingfei Chen 和 Eli Shlizerman，《Inras：音频场景的隐式神经表示》，*NeurIPS*，2022 年。[24] Mirco Pezzoli、Fabio Antonacci 和 Augusto Sarti，《利用物理信息神经网络对房间脉冲响应早期部分进行重建的隐式神经表示》，载于*Proc. Forum Acousticum.*，2023 年。[25] Fei Ma、Thushara D Abhayapala、Prasanga N Samarasinghe 和 Xingyu Chen，《用于头相关传递函数上采样的物理信息神经网络》，*arXiv preprint arXiv:2307.14650*，2023 年。[26] E. Vincent、T. Virtanen 和 S. Gannot，《音频源分离与语音增强》，John Wiley & Sons，2018 年。[27] Robin Scheibler、Eric Bezzam 和 Ivan Dokmani´c，《Pyroomacoustics：用于音频房间仿真和阵列处理算法的 Python 软件包》，载于*Proc. IEEE ICASSP*，2018 年。[28] Prerak Srivastava、Antoine Deleforge、Archontis Politis 和 Emmanuel Vincent，《如何（虚拟地）训练说话人定位器》，载于*INTERSPEECH 2023*，2023 年。[29] Aditya Arie Nugraha、Kouhei Sekiguchi 和 Kazuyoshi Yoshii，《语音复数谱图的深度生成模型》，载于*Proc. IEEE ICASSP*，2019 年，第 905–909 页。[30] Alexander Richard、Dejan Markovic、Israel D. Gebru、Steven Krenn、Gladstone Alexander Butler、Fernando Torre 和 Yaser Sheikh，《从单声道音频合成双耳语音》，载于*Proc. ICLR*，2021 年，第 1–13 页。[31] Jacob Donley 和 Paul Calamia，《DARE-Net：语音去混响与房间脉冲响应估计》，技术报告，斯坦福大学，2022 年。[32] Sifan Wang、Shyam Sankaran 和 Paris Perdikaris，《训练物理信息神经网络，只需尊重因果性》，arXiv 电子预印本，2022 年，arXiv:2203.07404v1。[33] Diego Di Carlo、Dominique Heitz 和 Thomas Corpetti，《利用物理信息神经网络对稀疏瞬时二维速度场进行后处理》，载于*Proc. Int. Symp. Appl. Laser Imag. Tech. FluidMech.*，2022 年。[34] Athanasios Papoulis，《信号分析》，Mcgraw-Hill College，1977 年。