混合声学啸叫抑制：结合卡尔曼滤波器与深度学习的声学啸叫抑制方法

张浩[1], 于萌[1],吴宇中[2],于涛[2],于东[1]

[1]腾讯 AI 实验室，贝尔维尤，华盛顿州，美国[1]腾讯天籁音频实验室，深圳，广东，中国

{aaronhzhang@global., raymondmyu@global., yvzhongwu@, taoyuty@, dyu@global.}tencent.com

arXiv:2305.02583v1 [eess.AS] 4 May 2023

摘要

近年来，深度学习被引入以高效抑制声学啸叫（AHS）。然而，啸叫的循环特性导致离线训练与流式推理之间存在失配，从而限制了增强语音的质量。为解决这一局限，我们提出一种混合方法，将卡尔曼滤波器与自注意力循环神经网络（SARNN）相结合，以发挥二者各自的优势，实现稳健的声学啸叫抑制。在离线训练阶段，我们使用卡尔曼滤波器生成的预处理信号，以及通过教师强制训练策略生成的理想麦克风信号来训练深度神经网络（DNN）。在流式推理阶段，DNN 的参数保持固定，同时将其输出用作参考信号来更新卡尔曼滤波器。在离线和流式推理场景中，使用模拟数据和真实录制数据进行的评估表明，所提出的方法能够高效抑制啸叫，并始终优于基线方法。

索引词：声学啸叫抑制，卡尔曼滤波器，教师强制，深度声学啸叫抑制，混合方法

1. 引言

声学啸叫是一种发生在扩声系统中的现象：扬声器发出的声音被麦克风拾取，并在反馈回路中不断循环放大，最终产生令人不悦的高频啸叫声[1, 2]。这种现象可能出现在各种场景中，例如音乐会、演讲、公共广播系统和助听器。声学啸叫抑制（AHS）是指减少或消除声学啸叫发生的过程。目前已有多种方法被提出，包括通过物理隔离麦克风和扬声器等被动方法，以及增益降低 [3, 4], 、陷波滤波器 [5, 6, 7], 和自适应滤波 [8]等主动方法。在这些方法中，自适应滤波（例如卡尔曼滤波器 [9, 10], ）能够实时动态调整信号，以防止反馈回路的形成，并获得相对较好的语音质量。然而，卡尔曼滤波器容易受到控制参数和干扰的影响，且无法解决放大器和扬声器引入的非线性失真问题。

近年来，深度学习凭借其建模复杂非线性关系的能力，被用于解决声学啸叫抑制相关任务。Chen 等人 [11]提出了一种基于深度学习的啸叫检测方法。文献 [12]介绍了一种基于深度学习的啸叫噪声消除方法。Zheng 等人 [13]利用深度学习解决声反馈系统的边缘稳定性问题，并将该方法命名为 DeepMFC。最近，又有研究提出了一种纯深度学习方法（深度声学啸叫抑制），用于声学啸叫抑制 [14]。然而，基于深度学习的方法容易受到离线训练与流式推理之间失配的影响，导致语音质量表现下降。

尽管声学啸叫抑制（AHS）方法的发展已取得显著进展，但现有方法仍面临诸多挑战，尤其是在抑制性能与信号失真之间的权衡问题。受将传统自适应方法与深度学习相结合以解决声学回声消除问题的成功经验启发 [15, 16, 17, 18], ，本文提出了一种用于声学啸叫抑制的混合方法。

所提出的方法称为混合声学啸叫抑制，它结合了两种解决声学啸叫问题的方法：一种是传统方法，即频域卡尔曼滤波器（FDKF）；另一种是基于自注意力循环神经网络（SARNN）的深度神经网络（DNN）模块 [19]。具体而言，FDKF 与 SARNN 以级联方式结合，FDKF 的预处理输出作为额外输入，用于训练 SARNN 模块。随后，在流式推理过程中使用预训练的 SARNN，并将其输出用作更新 FDKF参数的参考信号。在离线训练期间，混合声学啸叫抑制模型以教师强制方式进行训练 [20, 21] ，该方式假定麦克风信号中只有目标语音会被发送到扬声器。这有助于将递归式啸叫抑制过程转化为语音分离问题，并在流式推理期间展现出有效的啸叫抑制性能。所提出的方法充分利用了传统自适应滤波方法和基于深度学习的方法各自的优势。混合声学啸叫抑制的优势体现在两方面：1）使用经传统方法预处理的信号可为模型训练提供更多信息，并有助于减小离线训练与流式推理之间的失配；2）融入深度学习以进一步增强传统方法的输出，可解决非线性失真造成的泄漏，从而实现稳健的解决方案。

本文其余部分安排如下。第2节介绍声学啸叫问题。第3节介绍所提出的混合声学啸叫抑制方法。第4节和第5节分别介绍实验设置和实验结果。第6节总结全文。

2. 声学啸叫抑制

2.1. 声学啸叫

典型的单通道声学放大系统如图 1(a) 所示。该系统由麦克风和扬声器组成，麦克风拾取目标语音 $s(t)$，随后将其发送至扬声器进行声学放大。扬声器信号 $x(t)$ 播放后到达麦克风，形成声学反馈，记作 $d(t)$:

$$d(t) = NL[x(t)] * h(t) \tag{1}$$

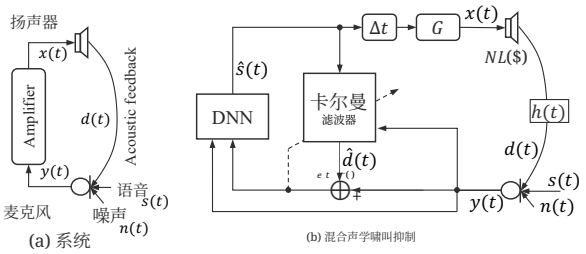

图1: （a）声学放大系统和（b）所提出的混合声学啸叫抑制方法的示意图。

其中， $NL(\cdot)$ 表示扬声器引入的非线性失真， $h(t)$ 表示从扬声器到麦克风的声学路径， $*$ 表示线性卷积。

如果不进行任何处理，播放信号 $d(t)$ 将反复重新进入拾音端，相应的麦克风信号可表示为：

$$y(t) = s(t) + n(t) + NL\left[y(t-\Delta t)\cdot G\right] * h(t) \quad (2)$$

其中， $n(t)$ 表示背景噪声， Δt表示从麦克风到扬声器的系统延迟， G是放大器的增益。 $y(t)$ 与$y(t-\Delta t)$之间的递归关系会导致播放信号被反复放大，并形成反馈回路，产生令人不适的高频啸叫声，这种现象被称为声学啸叫。

值得注意的是，声学啸叫和声学回声是两种不同的现象，尽管对声学回声处理不当可能会导致啸叫。这两种现象的主要区别如下：1. 虽然二者本质上都是播放信号，但啸叫的特征是信号能量以递归方式逐渐累积。2. 导致啸叫的信号与目标信号由同一信号源产生，因此抑制啸叫更加困难。

2.2. 现有的声学啸叫抑制方法

一般而言，声学啸叫抑制的目标是在尽可能保留所需信号的同时，降低或消除啸叫。然而，这一目标并不总是容易实现，因为抑制啸叫往往会造成一定程度的信号失真。基于卡尔曼滤波器的方法在各种信号处理应用中已经取得了长期的成功，并且在某些环境下能够有效抑制声学啸叫 [10]。但是，这类方法依赖于系统的统计模型，而该模型在存在非线性的场景下往往难以准确估计，并会导致明显的泄漏和/或信号失真。另一方面，基于深度学习的方法能够学习复杂的关系，并且在无法通过卡尔曼滤波器对啸叫进行良好建模的环境中，也能有效抑制声学啸叫 [14]。然而，离线训练与流式推理之间的失配会导致不可避免的信号失真。

挑战在于找到抑制性能与特定应用可接受的信号失真之间的平衡。因此，我们考虑结合传统方法与基于深度学习的方法，以实现更高的性能、更强的鲁棒性，并能够充分发挥两种方法各自的优势，从而实现更有效的解决方案。

3. 提出的方法：HybridAHS

为了解决上述声学啸叫抑制方法的缺点并发挥其优势，本文提出了一种称为混合声学啸叫抑制的混合方案。图 1（b）展示了所提出方法的示意图，该方法由两个模块组成：卡尔曼和深度神经网络。

3.1. 问题表述

考虑到啸叫具有递归特性，将声学啸叫抑制方法集成到声学环路中是抑制啸叫的最佳方式。然而，对于基于深度学习的方法而言，这种方式计算开销大且效率低。为应对这一挑战，我们沿用深度声学啸叫抑制 [14] 的方法，并采用教师强制训练策略，在模型训练期间将声学啸叫抑制表述为语音分离问题。所提出的方法基于以下假设：经过适当训练后，混合声学啸叫抑制模型能够抑制干扰，并仅将目标语音传输至扬声器。因此，在模型训练期间，图 1(b) 中的实际输出$\hat{s}(t)$ 可替换为理想目标（教师信号） $s(t)$，而公式 (2) 中递归定义的麦克风信号则转换为目标信号、背景噪声和由 $s(t)$确定的一次性播放信号的混合：

$$y(t) = s(t) + n(t) + h(t) * NL[s(t-\Delta t)\cdot G] \quad (3)$$

因此，AHS 的整体任务在离线训练期间转化为语音分离问题。其目标是利用卡尔曼滤波器输出 $e(t)$ 作为额外输入，从理想麦克风信号中提取目标信号 $s(t)$。理想麦克风信号定义见公式 (3)，且仅用于模型训练，从而联合抑制啸叫和噪声。

3.2. 卡尔曼滤波器

卡尔曼模块利用麦克风信号 $y(t)$ 和增强信号$\hat{s}(t)$ 作为参考（记为 $r(t)$），以获得声学路径 $\hat{}h(t)$ 的估计值及相应的反馈 $\hat{}d(t)$。随后，从麦克风信号中减去估计的反馈，并将所得误差信号 $e(t)$ 用于权重更新。整个过程可视为一个包含两个步骤（预测和更新）的流程，卡尔曼滤波器的权重通过这两个步骤的迭代反馈进行更新。

在预测步骤中，近端信号估计为

$$\mathbf{E}(k) = \mathbf{Y}(k) - \mathbf{R}(k)\hat{\mathbf{H}}(k), \quad (4)$$

其中，$\mathbf{E}$、$\mathbf{Y}$和$\mathbf{R}$ 分别是 $e(t)$、 $y(t)$ 和 $r(t)$ 的短时傅里叶变换（STFT）， k 表示帧索引。$\hat{}\mathbf{H}(k)$ 表示频域估计回声路径。

$\hat{}$回声路径$\mathbf{H}(k)$在更新步骤中进行更新：

$$\hat{\mathbf{H}}(k+1) = A[\hat{\mathbf{H}}(k) + \mathbf{K}(k)\mathbf{E}(k)], \quad (5)$$

其中， A 是转移因子。$\mathbf{K}(k)$表示卡尔曼增益，它是利用根据状态估计误差、观测噪声和过程噪声 [9]计算得到的协方差求得的。

3.3. 输入与特征提取

图2所示的深度神经网络模块接收经过卡尔曼滤波器处理的信号 e 和通过教师强制学习生成的理想麦克风信号 y 作为输入

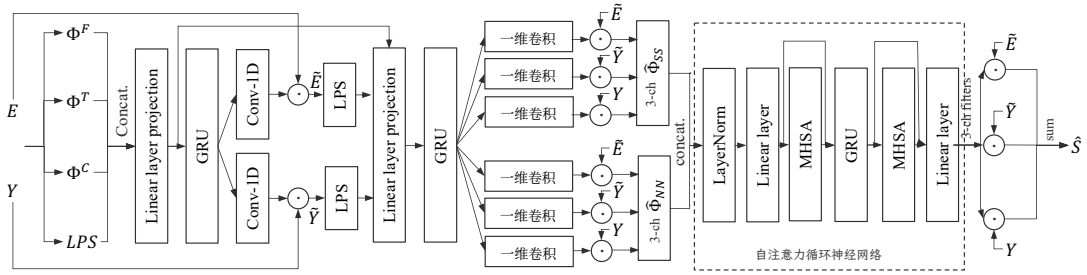

图2： 用于声学啸叫抑制的深度神经网络模块架构。每个“一维卷积”块都会输出一个复数比率滤波器，然后通过深度滤波将其应用于特定信号，记为 ⊙。

用于模型训练。输入信号以 16 千赫采样，并被分割为 32 毫秒的帧，帧移为 16 毫秒。随后对每帧进行 512 点短时傅里叶变换，得到频域输入 Y 和 E。除了归一化的对数功率谱（LPS）外，我们还提取输入信号在各时间帧和频率频点上的相关矩阵，以捕捉信号的时间依赖性和频率依赖性。这些特征有助于区分啸叫与音调成分。我们还计算输入信号的通道协方差（Y 和 E），作为另一项输入特征，以考虑它们之间的互相关。将这些特征拼接后，通过线性层进行特征融合，用于模型训练。有关特征设计的更多细节，请参阅 [19]

3.4. 网络结构

深度神经网络模块采用自注意力循环神经网络（SARNN）实现。该神经网络由三个主要部分组成。第一部分包含一个具有 257 个隐藏单元的门控循环单元（GRU）层和两个一维卷积层。这些层用于估计两个复数滤波器，并通过深度滤波将其应用于输入信号 [22]，以获得中间输出，分别记为 $\tilde{Y}$ 和 $\tilde{E}$。之所以要获得这些中间输出，是因为它们可以用作学习得到的非线性参考信号 [16, 17]，并为啸叫抑制提供更多信息。随后，将这些中间信号的对数功率谱与融合特征拼接，并作为另一 GRU 层的输入。~ ~我们将 Y、Y和 E 视为三通道输入，并针对每个输入通道采用两个一维卷积层，以估计其中的回放/噪声分量和目标语音分量。计算回放/噪声 Φ_{NN} 和目标语音 Φ_{SS} 对应的协方差矩阵，并将其拼接后作为第三部分 SARNN 的输入。SARNN 部分采用两个线性层、两个多头自注意力（MHSA）模块、一个 GRU 和残差连接，来估计三通道增强滤波器。随后，通过多通道深度滤波得到增强信号 $\hat{S}$。最后，使用逆短时傅里叶变换（iSTFT）得到波形 $\hat{s}$。网络结构的详细信息见 [19]。

3.5. 损失函数

我们在时域中使用尺度不变信号失真比（SI-SDR） [23]，并在频域中使用频谱幅度的平均绝对误差（MAE），结合起来对模型进行训练：

$$Loss = -\text{SI-SDR}(\hat{s}, s) + \lambda \text{MAE}(|\hat{S}|, |S|) \quad (6)$$

λ 设置为 10000，以平衡这两个 l 的取值范围 失函数。

4. 实验设置

4.1. 数据准备

使用 AISHELL-2 数据集 [24] 在回放、背景噪声和非线性失真的情况下开展实验。在数据模拟过程中，我们使用镜像法[25]，根据随机生成的房间特性和 0 至 0.6 秒的混响时间（RT60）范围，生成 10,000 组房间脉冲响应（RIR）。每组 RIR 包含近端说话人、扬声器和背景噪声位置对应的 RIR。随机选取一组 RIR，在 $[0.1, 0.3]$ 秒范围内随机生成系统延迟 Δt，并在 $[1, 3.2]$范围内随机选择放大增益，据此生成目标语音 $s(t)$ 及其对应的一次回放信号 $d(t)$。放大器和扬声器引入的非线性失真通过硬削波和 Sigmoid 函数 [26, 27]模拟为饱和型非线性。离线训练所用的麦克风信号由混合信号构成，其中信号与回放比（SPR）从 $[-10, 10]$ dB 中随机选取，信噪比（SNR）则在 -10 dB 至 30 dB 之间。初始阶段采用卡尔曼滤波器对训练信号进行预处理，并获得相应输出。训练集、验证集和测试集分别生成 10k、0.3k 和 0.5k 条信号。测试数据使用与训练和验证数据不同的语句及 RIR 生成。模型训练 60 个 epoch，批量大小为 32。

4.2. 评估指标

所提方法的性能通过两种不同方式进行评估：离线评估和流式推理 [14]。离线评估使用式（3）中生成的信号作为输入，以评估回放衰减性能。采用尺度不变信号失真比和语音质量感知评估（PESQ） [28] 评估回放衰减程度和目标语音质量。数值越高表示性能越好。

在流式推理中，我们将深度学习模块嵌入声学环路，并递归生成增强信号。这种评估方式考虑了泄漏/失真在闭合声学环路中的潜在重新引入，并评估所提方法的实时啸叫抑制性能 [14]。我们展示了递归处理后信号的频谱图，以说明所提方法的有效性。

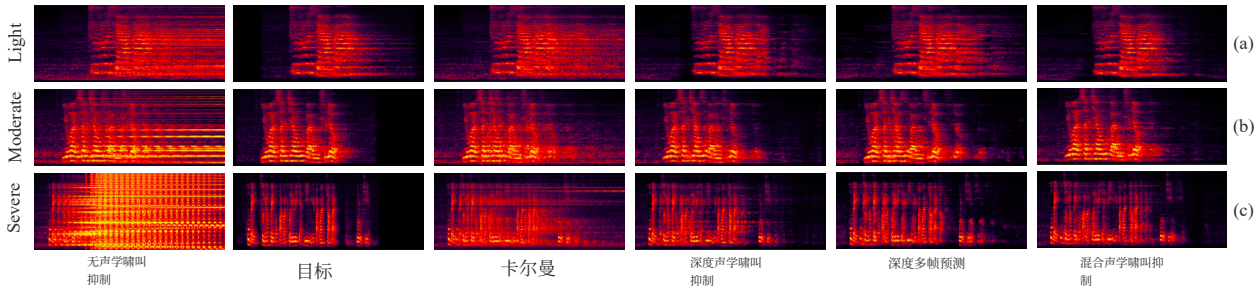

图3： 轻度、 （*b*）中度和（*c*）严重啸叫场景下流式结果的频谱图。每个面板均显示无声学啸叫抑制、目标信号、卡尔曼、*DeepAHS*、深度多帧预测和所提出的混合声学啸叫抑制的信号。

T表 1： 用于回放衰减的模型离线评估 。

	尺度不变信号失真比（dB）			PESQ		
G	1	2	3	1	2	3
未处理	8.59	2.82	-0.66	2.83	2.41	2.18
卡尔曼	7.66	-1.42	-10.56	2.88	2.44	2.07
深度声学啸叫抑制 [14]	16.32	13.32	11.46	3.60	3.41	3.27
深度多帧预测 [13]	14.82	8.60	2.61	3.46	3.13	2.81
混合声学啸叫抑制	**20.16**	**17.11**	**14.43**	**3.76**	**3.60**	**3.43**

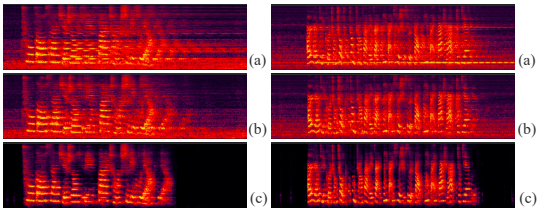

图 4： 使用真实录音和小型模型进行两项测试的频谱图，其中 *(a)* 关闭卡尔曼模块和深度神经网络模块，*(b)* 仅开启卡尔曼模块，以及 *(c)* 同时开启卡尔曼模块和深度神经网络模块（所提出的混合声学啸叫抑制）。

5. 实验结果

5.1. 离线评估

我们首先评估所提方法在回放衰减方面的性能，并将其与近期提出的两种基于深度学习的声学啸叫抑制方法 [13, 14] 以及一种基于卡尔曼滤波器的方法 [10]进行比较。比较结果见表1，其中 G 表示放大增益。请注意，此处展示的卡尔曼结果甚至可能差于“未处理”信号，因为前者以流式方式获得，而后者是使用式（3）所述的教师强制训练策略生成的理想麦克风混合信号。为确保公平比较，我们使用相同的网络和训练数据训练了深度声学啸叫抑制、深度多帧预测和混合声学啸叫抑制。比较结果表明，混合声学啸叫抑制方法始终优于所有基线方法。我们还按照 [14] 的建议，尝试使用延迟麦克风信号作为另一种参考信号。然而，我们发现，将延迟麦克风纳入混合声学啸叫抑制并不能提升性能，因为卡尔曼滤波器的输出已提供充足的参考信息，其效果优于将延迟麦克风作为额外输入所带来的收益。

5.2. 流式推理

本节采用流式推理评估所提出方法的有效性。通过在流式推理过程中逐步增大放大增益 G，生成软啸叫、中等啸叫和严重啸叫三种测试场景。处理后的信号频谱图如图3所示。可以观察到，所有基于深度学习的声学啸叫抑制方法均能成功防止啸叫发生，其中混合声学啸叫抑制方法的整体性能最佳。

5.3. 使用真实录制信号和可部署模型的性能

我们进一步使用真实录音信号和可部署模型评估所提出方法的啸叫抑制性能。为此，我们搭建了一个简单的声学放大系统，并使用真实录音和一个小型可部署模型训练混合声学啸叫抑制方法。该模型是一个长短期记忆网络（LSTM），由一个包含100个单元的隐藏层组成，可训练参数量为0.13 M。为了使模型能够部署到真实设备上，我们将帧长和帧移分别缩短至8 ms和4 ms，并且在模型训练中仅使用对数功率谱特征作为输入。处理结果如图4所示，证明了所提出方法在啸叫抑制方面的有效性，以及其进一步增强卡尔曼滤波器输出的能力。

6. 结论

在本研究中，我们介绍了一种混合声学啸叫抑制方法，将传统卡尔曼滤波与深度学习相结合，以抑制声学啸叫。该方法包括：使用经过卡尔曼滤波预处理的信号，以及采用教师强制训练策略生成的麦克风信号，对 SARNN 进行离线训练。在流式推理过程中，将预训练模型嵌入闭合声学环路中，对输入信号进行递归处理。通过结合卡尔曼滤波与深度学习，与基线技术相比，该方法在离线和流式场景下均能提升非线性声学啸叫抑制的效果和语音质量。未来的工作包括探索设备端实现等实际问题，并扩展该方法以应对多通道场景。

7. 参考文献

[1] R. V. Waterhouse，“混响房间中的啸叫反馈理论”，《美国声学学会杂志》，第37卷，第5期，第921–923页，1965年。[2] T. Van Waterschoot 和 M. Moonen，“声学反馈控制五十年：现状与未来挑战”，*IEEE*学报，第99卷，第2期，第288–327页，2010年。[3] M. R. Schroeder，“通过频移改善声学反馈稳定性”，《美国声学学会杂志》，第36卷，第9期，第1718–1724页，1964年。[4] E. Berdahl 和 D. Harris，“用于抑制声学啸叫的频移”，载于第*13*届数字音频效果国际会议论文集，奥地利格拉茨，第610卷，2010年。[5] W. Loetwassana、R. Punchalard、A. Lorsawatsiri、J. Koseeyaporn 和 P. Wardkein，“音频放大器系统中的自适应啸叫抑制器”，载于*2007*年亚太通信会议。IEEE，2007年，第445–448页。[6] P. Gil-Cacho、T. Van Waterschoot、M. Moonen 和 S. H. Jensen，“用于声学啸叫抑制的正则化自适应陷波滤波器”，载于*2009*年第*17*届欧洲信号处理会议。IEEE，2009年，第2574–2578页。[7] T. v. Waterschoot 和 M. Moonen，“基于陷波滤波器的啸叫抑制中啸叫检测准则的比较评估”，音频工程学会杂志，第58卷，第11期，第923–940页，2010年。[8] H. A. L. Joson、F. Asano、Y. Suzuki 和 T. Sone，“助听器中采用频率压缩的自适应反馈消除”，《美国声学学会杂志》，第94卷，第6期，第3248–3254页，1993年。[9] G. Enzner 和 P. Vary，“免提电话声学回声控制的频域自适应卡尔曼滤波器”，信号处理，第86卷，第6期，第1140–1156页，2006年。[10] F. Albu、L. T. Tran 和 S. Nordholm，“用于自适应反馈消除的混合简化卡尔曼滤波器”，载于*2018*年国际通信会议（*COMM*）。IEEE，2018年，第45–50页。[11] Z. Chen、Y. Hao、Y. Chen、G. Chen 和 L. Ruan，“面向实时通信应用的基于神经网络的啸叫检测方法”，载于*2022*年*IEEE*国际声学、语音与信号处理会议（*ICASSP*）。IEEE，2022年，第206–210页。[12] H. Gan、G. Luo、Y. Luo 和 W. Luo，“利用深度神经网络在时频域中消除啸叫噪声”，载于第六届信息与通信技术国际大会论文集。Springer，2022年，第319–332页。[13] C. Zheng、M. Wang、X. Li 和 B. C. Moore，“针对助听器声反馈系统边缘稳定性问题的深度学习解决方案”，《美国声学学会杂志》，第152卷，第6期，第3616–3634页，2022年。

[14] H. Zhang、M. Yu 和 D. Yu，“深度声学啸叫抑制：一种用于声学啸叫抑制的深度学习方法”，*arXiv* 预印本 *arXiv:2302.09252*，2023。

[15] S. Zhang, Z. Xie, H. J. Fu, T. Xie, Q. Yu 和 L. B. Xie，“采用回声感知损失的多任务深度残差回声抑制”，载于*ICASSP 2022-2022 IEEE* 国际声学、语音与信号处理会议 *(ICASSP)*。IEEE，2022，第 9127–9131 页。

[16] H. 张，S. Kandadai，H. Rao，M. Kim，T. Pruthi，以及 T. Kristjansson，“深度自适应声学回声消除：深度学习与自适应声学回声消除的混合方法”，载于 *ICASSP 2022-2022 IEEE* 国际声学、语音与信号处理会议（*ICASSP*）。IEEE，2022 年，第 756–760 页。

[17] Y. Zhang、M. Yu、H. Zhang、D. Yu 和 D. Wang，“KalmanNet：一种用于声学回声消除的可学习卡尔曼滤波器”，*arXiv*预印本 *arXiv:2301.12363*，2023。

[18] H. Zhang 和 D. Wang，“用于多通道声学回声抑制的神经级联架构”，*IEEE/ACM*《音频、语音与语言处理汇刊》，第 30 卷，2336，2022 2326 页。–。[19] M. Yu、Y. Xu、C. Zhang、S.-X. Zhang 和 D. Yu，“NeuralEcho：一种用于统一声学回声抑制和语音增强的自注意力循环神经网络”，*arXiv*预印本 *arXiv:2205.10401*，2022。[20] R. J. Williams 和 D. Zipser，“用于持续运行的全循环神经网络的学习算法”，神经计算，第 1 卷，第 2 期，第 270–280 页，1989。[21] A. M. Lamb、A. G. Alias Parth Goyal、Y. Zhang、S. Zhang、A. C. Courville 和 Y. Bengio，“教授强制：一种训练循环网络的新算法”，神经信息处理系统进展，第 29 卷，2016。[22] W. Mack 和 E. A. Habets，“深度滤波：使用复数时频滤波器进行信号提取与重建”，*IEEE*信号处理快报，第 27 卷，第 61–65 页，2019。[23] J. Le Roux、S. Wisdom、H. Erdogan 和 J. R. Hershey，“SDR——半生不熟还是恰到好处？”载于*2019 IEEE*国际声学、语音与信号处理会议（*ICASSP*）。IEEE，2019，第 626–630 页。[24] J. Du、X. Na、X. Liu 和 H. Bu，“AISHELL-2：将普通话自动语音识别研究推进至工业规模”，*arXiv*预印本 *arXiv:1808.10583*，2018。[25] J. B. Allen 和 D. A. Berkley，“用于高效模拟小房间声学的镜像法”，美国声学学会杂志，第 65 卷，第 4 期，第 943–950 页，1979。[26] A. Birkett 和 R. Goubran，“用于免提声学回声消除的非线性扬声器补偿”，电子快报，第 32 卷，第 12 期，第 1063–1064 页，1996。[27] H. Zhang 和 D. Wang，“噪声和双讲场景下的声学回声消除深度学习”，*Interspeech 2018*论文集，第 3239–3243 页，2018。[28] A. W. Rix、J. G. Beerends、M. P. Hollier 和 A. P. Hekstra，“语音质量感知评估（PESQ）——一种用于评估电话网络和编解码器语音质量的新方法”，载于*2001 IEEE*国际声学、语音与信号处理会议（*ICASSP*）。，第 2 卷。IEEE，2001，第 749–752 页。