arXiv:2305.06640v1 [eess.AS] 11 May 2023

扬声器振膜位移预测：深度注意力与在线适应

Yuwei Ren、*Matt Zivney*、*Yin Huang*、*Eddie Choy*、*Chirag Patel* 和 *Hao Xu*

高通人工智能研究{ren, matthewz, yinh, echoy, cpatel 和 hxu}@qti.qualcomm.com

摘要

扬声器保护算法利用播放信号的特性，在维持最大响度的同时防止振膜过度位移，这对配备微型扬声器的手机尤为重要。本文提出高效的深度学习解决方案，以准确建模和预测非线性位移；这对传统方法而言颇具挑战。首先，我们搭建实验和预处理流程，将反馈电流和电压采样作为输入，并采用激光器测量位移，作为真实值。其次，我们提出一种 FFTNet 模型，用于探索占主导地位的低频成分及其他未知谐波，并将其与基线 ConvNet 模型进行比较。此外，我们设计了 BN 重新估计来探索在线适应；并基于人工智能模型效率工具包（AIMET[1]）进行 INT8 量化，以进一步降低复杂度。我们在两个扬声器和 3 种典型部署场景中验证了所提出的算法， >99% 残余直流位移小于 0.1 mm，远优于传统方法。

膈肌位移、深度学习、在线适应

1. 引言

扬声器是一种电声换能器，可将功率放大器产生的电信号转换为声音。通常，扬声器的音圈通过悬架连接到安装在固定框架上的振膜。由于存在磁场，流经音圈的电流会产生作用力 f_c，使振膜上下运动 [1], ，其运动表示为振膜偏移量 x_d，且具有特定范围。如果偏移超出该范围，扬声器就会表现出非线性行为，进而产生声音失真，并降低声学回声消除性能 [2]。此外，电流流经扬声器音圈时，部分电能会转化为热能，而不是声音。当扬声器驱动过强时，这种振动会使振膜升温，导致塑料熔化，在边缘附近形成可见的气泡，并进一步造成面板不对称。

Qualcomm AIResearch 是 Qualcomm Technologies, Inc. 的一项计划。[1]AIMET 是高通创新中心, Inc. 的产品。

像活塞一样振动，而这一问题在更小巧、更便携的说话人中会更加突出，例如耳塞式耳机。

通用的解决方案是构建扬声器保护模块（SPL），用于监测电流和电压，并根据监测数据预测振膜的偏移状态。一旦偏移量 x_d 超过阈值，SPL 就会被触发，通过衰减输入功率或修改源信号来降低偏移量。在这一过程中，基于监测数据精确预测 x_d 是一项困难的任务。传统的均衡（EQ）滤波器虽然有效，但采用的是较为保守的解决方案，能够覆盖多种运行因素，例如具有较大动态范围的各种类型音频信号。[1] 对扬声器主动控制进行了全面研究。[3] 通过构建扬声器机电模型的精确逆模型，提出了一种用于封闭式箱体中直接辐射式扬声器非线性失真的均衡器。[2] 对数字扬声器模型进行估计并预测振膜偏移量，随后在偏移域中利用动态范围压缩对其进行控制。然而，这些保守的方法并未充分发挥扬声器的实际性能极限。例如，即使音频信号能量较低且振膜偏移量处于允许范围内，均衡滤波器仍会衰减音频输出，从而降低音频性能和响度。

近期，已有研究提出深度学习（DL）方法，用于对音圈执行器（VCA）的行为建模，并将循环神经网络（RNN）融入多物理场仿真。此外，研究人员还探索了利用深度学习求解微分方程（DE），这些方程可部分表征多种非线性系统的模型，例如振幅预测和 VCA 建模。[5]提出了一种可直接求解常微分方程的物理信息神经网络。[6] 通过对积分核进行参数化，在傅里叶空间中构建了神经算子。然而，目前尚不清楚如何利用深度学习探索扬声器保护所需的振幅特征。此外，深度学习方案高度依赖数据集质量；当不同扬声器单元之间的振膜振幅特征存在较大差异时，数据记录设置会对其产生影响，通常导致过拟合。

面对这些挑战，我们构建了理论模型（第 2.1 节）和一种振膜位移测量方案（第 2.2 节），以采集高质量数据集，并进一步.....

提出一种基于深度学习（DL）的振膜位移预测方法（第3节），以探索有效特征，并在两种典型说话人中进行验证（第4节）。最后，探讨 AIMET [7] [8] 中用于在线适应的 BN 重新估计（第3.2节）和 INT8 量化（第4.2节）。

2. 系统设计与理论模型

2.1. 振幅的理论模型与分析

扬声器可以表示为从电气域到机械域的转换。其电气特性的连续时间模型如 [1]所示：

$$v_c(t) = R_{eb} i_c(t) + \phi_0 \dot{x}_d(t) \quad (1)$$

其中，$v_c(t)$ 是加在音圈两端的电压输入，$i_c(t)$ 是音圈电流，R_{eb} 是阻塞电阻，$\dot{x}_d(t)$ 是振膜位移速度，ϕ_0 是平衡状态$x_d(t)$（即振膜位移）下的换能系数。扬声器的这些机械特性主要由参数 R_{eb} 和 ϕ_0决定，这些参数在很大程度上取决于其几何结构、材料、振膜和箱体。要准确建立数学模型来跟踪其非线性分布和变化并不容易。此外，在式 (1) 中，这种非线性特征可以表示为[$v_c(t)$ $i_c(t)$ $x_d(t)$],，并进一步用于监督式深度学习方案，以学习其特性。

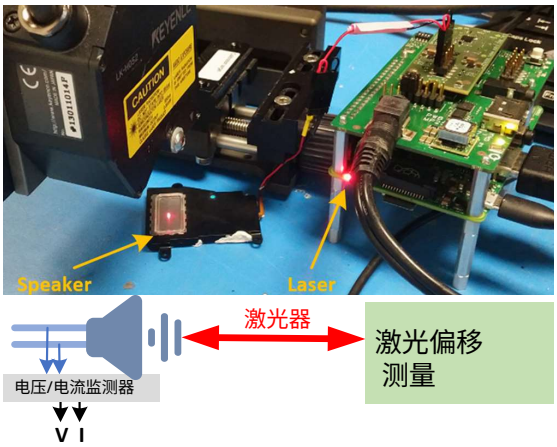


图 1。实验装置与框图。

2.2. 基于激光器测量的实验设置

激光器指向说话人的中心，以跟踪振膜位移。振膜振幅记录为 $x_d(t)$，同时测量相应的实时电流$i_c(t)$和电压 $v_c(t)$，如图1所示。接下来，将公式(1)转换为

$$v_c(t), i_c(t), x_d(t) \Rightarrow f_{R_{eb},\phi_0}(\bullet) \quad (2)$$

其中 $f(\bullet)$ 是用于表征机械特性的函数。例如，在一个电压控制的说话人中，电压是由语音内容编码的源输入，用于驱动振膜位移。而机械响应则嵌入在反馈电流中。一旦某个模型经过优化或学习，能够表征 $f(\bullet)$ 并保持稳定，就可以根据该模型，利用实时记录的电流和电压来预测相应的振膜位移，如式（3）所示，

$$x_d(t) = f_{R_{eb},\phi_0}(v_c(t), i_c(t)) \quad (3)$$

其中，我们的目标是基于记录的数据集学习 $f_{R_{eb},\ \phi_0}(\bullet)$，并利用该模型，在给定实时 $v_c(t)$ 和 $i_c(t)$ 的情况下预测振幅位移。

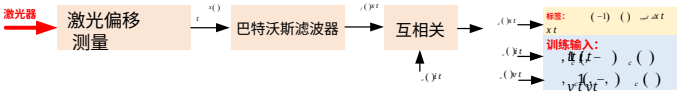


图 2。标签预处理与数据集准备。

激光器测量振膜位移，该位移受噪声和说话人安装位置的影响，导致记录的数据集始终存在噪声。此外，考虑到持续较长时间的大幅位移是损坏振膜的主要原因，我们重点预测直流漂移，并消除噪声的影响。为此，采用一个截止频率为 10 Hz 的二阶滤波器，提取位移的直流分量作为模型输出，如图 2 所示。此外，记录数据时，电流/电压与激光器使用不同的时钟，因此需要进行同步以对齐序列。为此，使用互相关（例如 xcorr）来调整电流与测得位移之间的偏移。

3. 频域深度学习

此前有多项研究 [2] [9], 通过构建数学函数来对位移建模，揭示了与未知频率及若干谐波分量相关的直流漂移。因此，本文采用深度学习（DL）方案提取此类有效的频率特征，并进一步预测直流漂移。

在 DL 中，每个样本定义为 $\{\mathbf{s}\ x_N\}$，其中 $\mathbf{s} = (s_1, s_2, \ldots, s_N) \in \mathbf{R}^{2N}$，映射到缓冲序列中最新的 N 个时间戳，并分别对应电压和电流分量。x_N 是输出标签，对应于序列中 s_N 的时间戳 t。

傅里叶注意力算子层（FAOL）用于提取有效的频率分量。此外，本文提出了 FFT-Net：首先使用一个 conv1d 层增加通道特征，再通过多个 FAOL 层提取特征。经过平均池化对特征进行下采样后，由 FC 输出预测结果，如图 3 所示。

此外，作为对比，还提出了一种典型的基于 ResNet 的 Conv1D 网络（ConvNet），其具有相似的结构：输入和输出格式与 FFTNet 相同，包含 1 个 Conv1D】【。

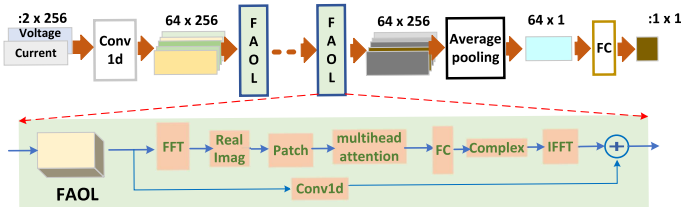


图 3。FFTNet 结构：输入为 256 个采样点，在 48千赫兹采样率下对应 5.3毫秒；复数中的实部/虚部沿通道维度拼接；随后，将通道拆分以恢复复数值。

层、若干 ResNet块、平均池化层和全连接层，用于回归预测的直流漂移，详细内容见第 4 节。

3.1. 傅里叶注意力算子层（FAO） L)

假设神经网络中有 J 个 FAOL，每层的输出为 g_j，其中 $j = 1, 2, 3, \ldots\ldots, J$。对于每一层的输入，我们进行离散傅里叶变换 F，将输入的时域样本转换到频域。受 [10], 中工作的启发，我们使用由 ϑ 参数化的多头自注意力模块在频域中进行学习，然后通过傅里叶逆变换 F^{-1} 恢复时域序列。这个过程定义为傅里叶注意力算子 K（ϑ），表示为

$$K(\vartheta)(g_{j-1}) = F^{-1}(\varphi R_\vartheta(F(g_{j-1}))) + \psi(g_{j-1}) \quad (4)$$

其中，R_ϑ(•) 是多头注意力模块，用于根据给定的数据块学习系数。随后，φ 是权重张量，用于对频域中的各个模态进行线性组合。第 j 层的输出将 F^{-1}输出与经过卷积操作 ψ(•) 加权的初始时域序列相加，这种时域中的跳跃路径旨在恢复被舍弃的有价值的频率成分。此外，多个 FAOL 模块串联起来，形成的深层结构有助于提取谐波特征，并结合使用 ReLU激活和一维卷积操作。

3.2. 模型泛化与在线适应

说话人受到许多未知机械特性的显著影响，例如生产质量，甚至同一说话人的不同单元之间也难以实现对齐。为了实现在线适应，我们提出了一种批归一化（BN）重新估计方法，以适应此类变化，无需任何标注或微调，从而满足边缘设备严格的功耗和计算资源限制。BN 层最初旨在缓解内部协变量偏移问题——这是训练非常深层神经网络时常见的问题，其定义如公式（5）所示，

$$\hat{x}_j = \frac{x_j - E[\mathbf{X}_{\cdot j}]}{\sqrt{Var[\mathbf{X}_{\cdot j}]}}, \quad y_j = \gamma_j \hat{x}_j + \beta_j, \quad (5)$$

其中，x_j 和 y_j 分别表示一个样本中单个神经元响应的输入/输出标量，$\mathbf{X}_{\cdot j}$ 表示一个 BN 层中输入数据的第 j_{th} 列，$\mathbf{X} \in R^{n \times p}$, $j \in \{1 \ldots p\}$。 n表示批量大小，p 表示特征维度。 γ_j和 β_j 为可训练参数。训练完成后，BN 中的此类参数将被冻结，在推理过程中保持不变。

当模型通过 BN 重新估计部署到新设备或新场景中进行推理时，如算法 1 所示，输入数据会在给定窗口内进行缓冲，以计算均值和方差。此外，为了跟踪变化，采用一个单抽头 IIR滤波器，以探索整个空间中的优化。

算法1 推理过程中的BN重新估计 要求：给定缓冲窗口 T 和滤波系数 α确保：在推理过程中启用BN重新估计 while 神经元 j, t次推理前向传播 do 拼接第 t个输入上的神经元响应：$buffer = [buffer, x_j(t)]$在前向传播中计算批归一化输出：$y_j(t) = \gamma_j \frac{x_j(t) - \mu_j}{\sigma_j} + \beta_j$if mod $(t, T) == 0$ then 基于$buffer$计算均值和方差：$\mu_j^b = \mathbb{E}(buffer)$, $\sigma_j^b = \sqrt{Var(buffer)}$ 更新神经元 j的均值和方差：$\mu_j = \alpha\mu_j + (1-\alpha)\ \mu_j^b$, $\sigma_j = \alpha\sigma_j + (1-\alpha)\ \sigma_j^b$重置 $buffer = [\,]$end if end while

4. 实验与结果

在实验中，考虑了 3 种典型场景：正常、加热和直流注入。使用来自两个扬声器（SBS2 和 AAC）的 14 个单元，其中 8 个单元用于训练，4 个单元用于验证，2 个单元用于测试。验证了 3 种方法：**FFTNet** 由 3 个 FAOL 块组成；**ConvNet** 包含 4 个 ResNet块；它们的复杂度如表 1 所示；**DSP**采用有限的运算和较少的参数，以降低内存开销。需要注意的是，DSP 仍利用训练数据集对算法进行微调，类似的 DSP 实现已在 [11] [12] 中讨论。

由于较大的直流漂移会对扬声器振膜造成更严重的损伤，因此优化过程应忽略受噪声影响的零点附近的小损失，并增强超过阈值的较大损失。在训练过程中提出了一种缩放损失函数，如 $L = \frac{1}{S}\sum_{i=1}^{S} \delta\left(x_{out,i}^T - x_{out,i}^P\right)^4$ 所示，其中 S 表示批次大小，$x_{out,i}$ 表示用于预测和真实值的第 i 个样本，而 δ 用于调整与典型 L1 损失相同的点。

4.1. 浮点性能

图4展示了推理性能，预测结果能够跟踪直流位移（DC）的峰值，且残差为0。对于受随机噪声影响而产生的小幅直流位移抖动，例如来自电路或机械振膜噪声的干扰，预测结果难以跟踪其变化。最大损失出现在直流位移骤变时。理想情况下，当输入为零时，例如断电或没有语音输入时，$v_c(t) = i_c(t) = 0$，相应的振膜位移应为零。然而，由于振膜受到机械约束，其位移应逐渐恢复至零，而不会与电信号控制完全同步。尽管这种骤变会导致较大的损失，但由于振膜位移正在减小，因此不会对扬声器造成损坏。

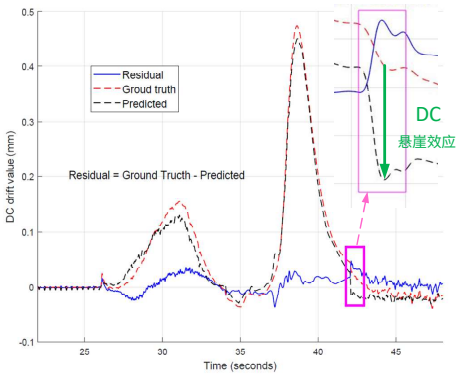


图 4。基于 FFTNet FP32 模型的一段测试序列，其最大 L1 损失为 $= 0.0478$mm。预测的直流漂移序列能够跟踪真实的直流变化。

表 1 展示了在直流注入场景下启用 FP32 的实验结果；这是最具挑战性的边界场景。与 ConvNet 相比，FFTNet 展现出更有希望的性能。但其中涉及的 FFT 处理仍难以部署在真实硬件设备上，例如由 NPU 加速的设备。表 2 验证了用于在线适应的 BN 重新估计：一个仅基于 SBS2 说话人训练的模型，进一步迁移到了 AAC 说话人。通过一个 AAC 单元验证性能，并进一步探究不同滤波器系数 α 的影响。实验表明，这种适应方法可以大幅提升性能。例如，与基线相比，选项 $\alpha = 0.1$ 在最大损失方面可提升 21%，但相应的平均损失略高于其他选项。最优 α 高度依赖于数据/模型权重分布，因此难以确定一个最佳值。

4.2. 模型量化

模型量化是降低模型在边缘设备上部署复杂度的关键。在这里，我们利用 AIMET 进行 INT8 量化。由于量化

[2]不包含 FFT 和 IFFT 处理。

表 1。L1 损失与复杂度结果：验证数据集中采用 DC 注入场景的 SBS2 说话人，ConvNet 与 FFT-Net 的对比。

	均值（mm）	最大值（mm）	浮点运算次数	参数
FFTNet	0.0077	0.2169	0.3M[2]	1.7K
卷积神经网络	0.0091	0.2314	3.2M	19K

表2。ConvNet FP32 模型直接部署在单个 AAC 单元中，定义为基线。仅启用最后一个 BN 层，滤波器设置为 $\alpha = \{0.001, 0.01, 0.05, 0.1\}$。

$\alpha =$	基线	0.1	0.05	0.001	0.0001
均值（毫米）	0.0081	0.0035	0.0024	0.0010	0.0028
最大值（mm）	0.7466	0.4179	0.5849	0.6802	0.7055

由于 FFT 算子的实现不够友好，所提出的 ConvNet 采用了量化处理，而此处不考虑 FFTNet。在实验中，针对 SBS2 和 AAC 分别专门设计了两个独立模型。

如表 3 所示，与 DSP 相比，ConvNet32 的平均损失大幅降低，最大损失接近但略高于目标值（0.1 mm）。此外，经过 8 位量化后，与基线 FP32 相比，INT8 的性能略有下降，最大 L1 损失从 0.1121 mm 增至 0.1298 mm，但仍远优于 DSP 方案。SBS2 说话人也得出了与 AAC 类似的结论，即其平均损失略高于 DSP。

表 3。L1 损失结果：测试数据集中的 3 种场景。

		平均值（mm）	最大值（mm）
AAC	DSP	0.0140	0.2569
	卷积神经网络（ConvNet）FP32	0.0038	0.1121
	ConvNet INT8（卷积神经网络 INT8 量化）	0.0076	0.1298
SBS2	DSP	0.0020	0.2711
	ConvNet FP32	0.0032	0.13171
	ConvNet INT8（卷积神经网络 INT8）	0.0046	0.1408

5. 结论与未来工作

本文提出了一种用于预测扬声器保护中直流漂移的端到端流程。注意力机制用于提取频率特征，其性能优于 ConvNet 和传统 DSP 解决方案。此外，当模型部署到新场景时，可通过 BN 重新估计实现在线适应。最后，我们在 3 种场景中验证了此类模型，并使用 AIMET 将其量化为 8 位。未来有多个研究方向，例如：进一步量化 FFTNet；设计用于监控在线模型适应的机制，例如使用一个黄金序列。

6. 参考文献

[1]Andrew Paddock Bright、Finn Jacobsen、Jean-Dominique Polack 和 Karsten Bo Rasmussen，《扬声器主动控制：实际应用研究》，2002 年。

[2] Chenchi Luo、James McClellan、Milind Borkar 和 Arthur Redfern，《一种基于模型的扬声器振幅保护算法》，载于*2012*年 *IEEE* 国际声学、语音与信号处理会议（*ICASSP*）。IEEE，2012年，第233–236页。

[3] Shreyas Srikanth Payal、V John Mathews、Ajay Iyer、Russ Lambert 和 Jeffrey Hutchings，《扬声器振幅与电流相关非线性的均衡》，载于*2014* 年 *IEEE* 声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2014年，第6697–6701页。[4] Tianjian Lu、Michael Smedegaard 和 Ken Wu，《使用循环神经网络对音圈执行器进行多物理场建模》，*IEEE* 多尺度与多物理场计算技术期刊，第4卷，第12–19页，2019年。[5] Jochen Stiasny、Samuel Chevalier 和 Spyros Chatzivasileiadis，《无数据学习：用于快速时域仿真的物理信息神经网络》，*arXiv*预印本 *arXiv:2106.15987*，2021年。[6] Zongyi Li、Nikola Kovachki、Kamyar Azizzadenesheli、Burigede Liu、Kaushik Bhattacharya、Andrew Stuart 和 Anima Anandkumar，《用于参数化偏微分方程的傅里叶神经算子》，*arXiv*预印本 *arXiv:2010.08895*，2020年。[7]Sangeetha Siddegowda、Marios Fournarakis、Markus Nagel、Tijmen Blankevoort、Chirag Patel 和 Abhijit Khobare，《使用人工智能模型效率工具包（AIMET）进行神经网络量化》，*arXiv*预印本 *arXiv:2201.08442*，2022年。[8] 高通创新中心，《人工智能模型效率工具包（AIMET）》，https://github.com/quic/aimet。[9] Rong Hu、Zheng Yan 和 Jie Su，《扬声器过冲程保护》，2018年6月5日，美国专利9,992,571。

[10]Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、GeorgHeigold、SylvainGelly 等人：《一张图像值 16×16 个词：用于大规模图像识别的 Transformer》，*arXiv* 预印本 *arXiv:2010.11929*，2020 年。

[11] Leung Kin Chiu、David V Anderson 和 Benjamin Hoomes，“音频输出增强算法用于压电式扬声器，”载于*2011*年数字信号处理与信号处理教育会议（*DSP/SPE*）。IEEE，2011年，第317–320页。

[12] Susan Xue、Supriyo Palit、Peter Wei 和 Andrew Du：《SmartPA 扬声器保护算法》，载于德州仪器应用报告。德州仪器公司，2018 年，第 1–14 页。