arXiv:2304.08541v1 [eess.AS] 17 Apr 2023

面向超低功耗端侧关键词检测的模拟滤波器组特征可以做得多小?

Subhajit Ray 哥伦比亚大学 纽约，美国
subhajit.ray@columbia.edu

Xinghua Sun 哥伦比亚大学 纽约，美国
xs2445@columbia.edu

Nolan Tremelling 哥伦比亚大学 纽约，美国
nt2109@columbia.edu

Maria Gordiyenko 巴纳德学院 纽约，美国
mg4366@barnard.edu

Peter Kinget 哥伦比亚大学 纽约，美国
pk171@columbia.edu

摘要

模拟特征提取是一种节能且重新兴起的信号处理范式，可用于在端侧关键词检测系统中实现前端特征提取器。尽管模拟特征提取具有能效优势且再度受到关注，但对于其关键模块——模拟滤波器组的架构参数应设为何值，尚未形成共识，尽管这些参数会显著影响功耗。为推动共识形成并接近功耗的基本限制，我们通过仿真发现，仔细选择架构参数后，典型的最先进模拟滤波器组的功耗可降低 33.6×，而通过后端神经网络进行下游 10 词关键词检测时，准确率仅损失 1.8%。

关键词

关键词检测、端侧、超低功耗、音频特征提取、模拟特征提取、模拟滤波器组

ACM 引用格式:

Subhajit Ray、Xinghua Sun、Nolan Tremelling、Maria Gordiyenko 和 Peter Kinget。2023。《超低功耗端侧关键词检测中的模拟滤波器组特征可以做得多小？》。载于*tinyML*研究研讨会论文集（*tinyML Research Symposium'23*）。ACM，纽约，纽约州，美国，共 7 页。

1 引言

将设备端关键词检测（KWS）的功耗降至超低水平，对于实现泛在智能的愿景至关重要。关键词检测系统通常由前端特征提取器和后端神经网络组成。尽管针对关键词检出的神经网络已经开展了大量软件研究，[4, 12, 16], 硬件研究，尤其是芯片设计研究表明，最先进的关键词检测芯片 [6] 的功耗瓶颈并非后端神经网络，而是前端特征提取器，其功耗约占关键词检测系统总功耗的 70%。因此，低功耗前端特征提取研究十分重要。

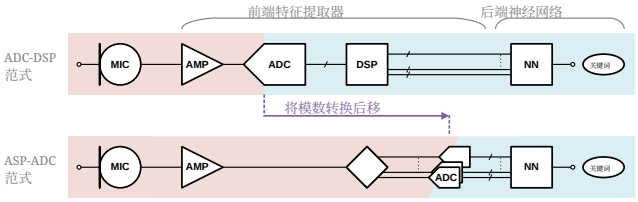


图 1：采用模数转换后进行数字信号处理（模数转换-数字信号处理，ADC-DSP）的传统前端特征提取范式，与采用模拟信号处理后进行模数转换（模拟信号处理-模数转换，ASP-ADC）的非传统但重新兴起且能效更高的范式。本文关注的范式是 ASP-ADC，其中的关键模块是 ASP 内部的模拟滤波器组。

与 [6],采用的传统数字密集型前端特征提取范式相比，功耗效率更高的是较不传统、但正迅速重新兴起的 [2, 5, 7, 11, 15, 20, 22, 23]模拟密集型范式。前者采用奈奎斯特速率模数转换加数字信号处理（模数转换器-数字信号处理），而后者采用模拟信号处理加特征速率模数转换（模拟信号处理-模数转换）[13, 19], ，如图 1 所示。一个典型例子是，基于模拟信号处理-模数转换的最低功耗前端特征提取器 [22] ，其功耗效率 12× 高于基于模数转换器-数字信号处理的最低功耗前端特征提取器（放大器和模数转换器 [1],数字信号处理 [14]）。尽管总体而言，模拟信号处理-模数转换的可编程性较低、可变性较高，但研究基于模拟信号处理-模数转换的前端特征提取仍然十分重要，因为它功耗效率高，而且适用范围广，不仅适用于 tinyML 音频，也适用于 tinyML 视觉 [10]。

基于 ASP-ADC 的前端特征提取器中，关键模块是模拟滤波器组，它位于图1中的“ASP”模块内。尽管滤波器组的架构参数会显著影响功耗，但对于这些参数应设为何值，业界尚未达成共识；表1 展示了最先进的模拟特征提取芯片所采用的参数值选择，揭示了这一点 [7, 11, 20, 22]。这种缺乏共识的现象

允许出于个人用途制作本作品部分或全部内容的数字版或纸质副本，或允许将本文用于课堂教学，且无需付费，但不得为营利或商业利益而复制或分发，且副本首页必须附有本声明和完整引文。必须尊重本文中第三方组件的版权。其他用途请联系所有者/作者。*tinyML Research Symposium'23，2023*年*3*月，美国加利福尼亚州旧金山 © 2023 版权所有者/作者所有。

表1：总结了最先进的模拟特征提取器芯片在模拟滤波器组主要架构参数取值上的选择，结果显示，尽管所有芯片的目标任务都是关键词检出，但各方并未达成共识。

	$N_{\text{滤波器}}$	$f_{\text{最大，滤波器组}}$	$Q_{\text{滤波器}}$
固态电路期刊'22 [11]	31	4kHz	3
固态电路期刊'22 [7]	16	10千赫兹	2
固态电路期刊'21 [22]	16	5kHz	2
IEEE电路与系统技术委员会学报'21[20]	32	8kHz	26
典型值（平均值）	24	7kHz	8

尤其令人惊讶，因为这些芯片的目标任务都是关键词识别。鉴于这种缺乏共识的情况，以及对功耗基本极限的探索，我们提出一个问题：模拟滤波器组可以缩减到多"小"，同时又不会显著降低分类准确率？我们发现，通过谨慎选择其架构参数，可以使功耗降低 33.6×，而关键词识别准确率仅损失 1。8%。

本文其余部分的组织如下：第 2 节介绍模拟特征提取和滤波器组的背景；第 3 节讨论仿真方法；第 4 节报告仿真结果，提出一种微型模拟滤波器组，并将其与典型滤波器组进行比较；第 5 节在更广泛的背景下讨论这些结果；最后，第 6 节总结本文的要点。

2 背景

2.1 模拟特征提取

在用于前端特征提取的 ASP-ADC 范式中，这一过程通常被称为模拟特征提取。图 1 中标记为"ASP"的模块通常由模拟滤波器组和其后的模拟包络检波器组组成。滤波器组由一组带通滤波器构成，其作用是将输入信号分解为各个频率分量。随后，包络检波器组用于检测每个频率分量的包络，即其平均功率的度量。因此，模拟特征提取器的输出是一个特征向量，其中每个元素的值与某个频率子带内随时间变化的短时平均功率成正比。这样，随着时间推移将特征（列）向量横向堆叠起来，便得到输入信号的声谱图，该声谱图作为后端神经网络的输入。

模拟特征提取器由两个组成模块构成：模拟滤波器组和模拟包络检波器组，每个模块都有各自的架构参数。包络检波器的架构参数，即窗口长度和窗口重叠，已在最先进的模拟特征提取器芯片中形成共识，其典型值分别为 20–32ms 和 10–16ms [7, 11, 20]。我们选择不探究这些参数，不仅是因为已有这种共识，也因为这些参数对包络检波器功耗的影响较小。

2.2 模拟滤波器组

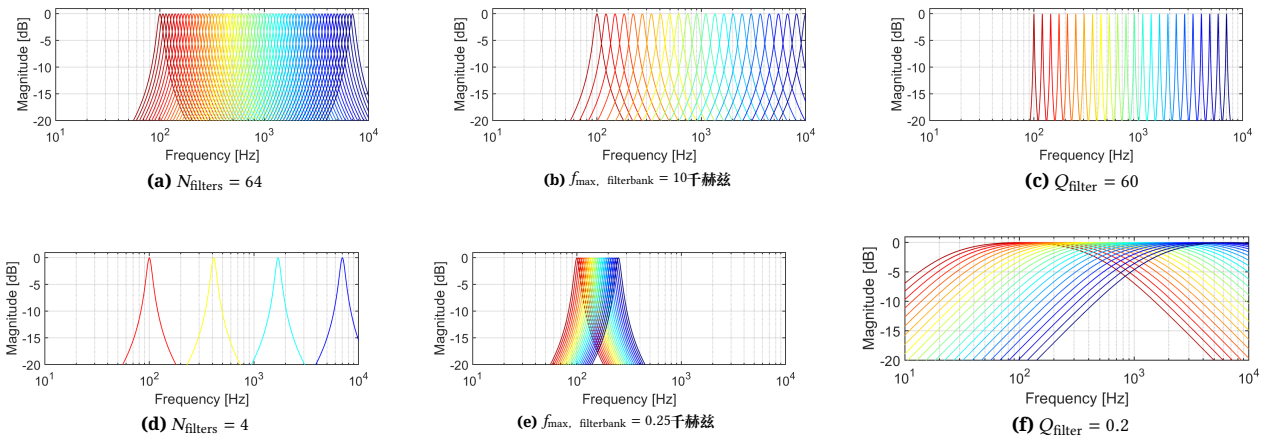
另一方面，如表1所示，滤波器组架构参数应设为何值尚未形成共识。对于 tinyML 而言，这种缺乏共识很重要，因为从一阶近似来看，滤波器组的功耗与我们关注的三个[2] 架构参数成正比 [18]： $N_{\text{滤波器}}$、$f_{\text{最大值，滤波器组}}$，$Q_{\text{滤波器}}$。我们要明确区分这些架构参数与用于描述模拟滤波器组非理想特性的参数，即噪声和失真；后两者也会显著影响功耗。由于在模拟设计过程中，通常会先确定架构参数，再规定噪声和失真的要求，因此本文选择先探究架构参数，将噪声和失真留待未来研究。

为更好地理解所关注的架构参数，图 2 展示了滤波器组在极端参数值下的频率响应。请注意，在每个子图的说明中，如果未注明某个参数值，则该参数取表1中规定的典型值。参数 $N_{\text{滤波器}}$ 是构成滤波器组的带通滤波器数量。参数 $f_{\text{max，滤波器组}}$ 是滤波器组的最高频率，更具体地说，是滤波器组中频率最高的带通滤波器的中心频率。参数 $Q_{\text{滤波器}}$ 是滤波器组中所有带通滤波器的品质因数。品质因数用于衡量带通滤波器的频率选择性，定义为其中心频率与其 3 dB 带宽之比；也就是说，品质因数是分数带宽的倒数。

除了所关注的三个架构参数外，还有一些次要参数；不过，对于这些参数的取值已有共识，或者它们对滤波器组功耗的影响较小。次要架构参数及其典型取值范围如下：

- 最低频带通滤波器的中心频率。在最先进的模拟特征提取器芯片中，通常设为 30–100 Hz [7, 11, 20, 22]。它对滤波器组功耗的影响较小，因为最低频带通滤波器的功耗只占滤波器组中所有带通滤波器总功耗的一小部分。
- 中心频率的分布。在最先进的模拟特征提取器芯片中，通常设为指数分布[3]，其依据是生物耳蜗的启发 [11, 22]。不过，也有一些采用 Mel [8]或 ERB [7] 分布。
- 带通滤波器的阶数。在最先进的模拟特征提取器芯片中，通常设为二阶[4] [7, 11, 22]。这是因为此类芯片已能达到足够的关键词识别准确率，而更高阶数会按比例消耗更多功耗。
- 品质因数方差。在最先进的模拟特征提取器芯片中，通常会将滤波器组中所有带通滤波器的品质因数设为相同的值[4]，

[1]假设每个滤波器的信噪比保持不变。[2]假设另外两个参数保持不变。[3]这就是为什么图 2 中的中心频率在对数频率刻度上呈线性分布。[4]这就是为什么在对数频率轴上，滤波器组中所有带通滤波器的 3 dB 带宽看起来都相同。



(a) $N_{\text{filters}} = 64$ **(b)** $f_{\text{max, filterbank}} = 10$千赫兹 **(c)** $Q_{\text{filter}} = 60$

(d) $N_{\text{filters}} = 4$ **(e)** $f_{\text{max, filterbank}} = 0.25$千赫兹 **(f)** $Q_{\text{filter}} = 0.2$

图2：模拟滤波器组的频率响应（理想情况），展示其架构参数取极端值时的情况，以说明各参数的含义。滤波器组功耗在一阶近似下与这些参数分别成正比。

这种设计受到生物耳蜗的启发 [7, 11, 22]，[8]。

讨论了感兴趣的三个模拟滤波器组特征参数——$N_{\text{滤波器}}$、$f_{\text{最大值，滤波器组}}$、$Q_{\text{滤波器}}$——并认识到滤波器组功耗在一阶近似下与各参数成正比后，接下来的问题是：在不显著降低后端神经网络的关键词识别准确率的前提下，各参数值可以降低到什么程度？为回答这个问题，我们逐一调整这三个参数，同时将其他两个参数保持在表1所列的典型值，并对每个调整点的关键词识别准确率进行仿真，从而确定准确率开始显著下降时对应的参数值。

3 仿真方法

3.1 特征提取前数据集

音频信号的特征提取前数据集由 Google语音命令数据集（GSCD）v0.02 构建，该数据集由时长 1 秒、16 位、采样率为 16 kS/s 的音频信号组成 [21]。在可用的 35 个词中，有 10 个词——yes、no、up、down、left、right、on、off、stop、go——被指定为关键词类别，其余 25 个词则被指定为第 11 个未知类别。每个关键词类别使用了 200 个训练样本，而未知类别使用了更多训练样本，共 800 个；否则，未知类别的分类准确率会低于关键词类别，这是因为未知类别包含许多（25 个）不同的词。总的来说，特征提取前训练集包含 2,800 个样本。

尽管 GSCD 为每个单词提供了至少 3,000 个示例，但只使用 200 个示例的原因在于，受可用计算资源的限制，这样可以缩短特征提取和训练时间。由于还剩下大量示例，我们决定在特征提取前的测试集（和验证集）中，额外为每个关键词使用 200 个示例，而不是使用例如 20 个示例。这样得到的训练集、验证集和测试集比例为 1:1:1；作出这一选择是为了利用剩余的示例来改进图 3、4、5 中所示的测试集准确率估计。

最后，虽然为了加快所需的 47 次训练会话，图 3、4、5 所用的数据集规模保持相对较小，仅包含 2,800 个训练样本，但图 6 和 7 中对“典型”模拟滤波器组与所提出的“微型”滤波器组进行最终比较时，使用了一个大 10 倍的数据集，其中包含 28,000 个训练样本。

3.2 前端模拟特征提取器

将特征提取前数据集（训练集/验证集/测试集）输入我们自定义的模拟特征提取器 MATLAB 模型，生成由频谱图组成的特征提取后数据集，并用该数据集训练和测试后端神经网络。如第 2 节所述，模拟特征提取器模型由可编程滤波器组和其后的固定包络检波器组组成。此外，如第 2.2 节所述，滤波器组的噪声和失真未纳入建模，留待未来研究；这样可以单独考察滤波器组架构参数对关键词识别准确率的影响。

包络检波器的固定参数为：窗口长度 20 ms，窗口重叠 10 ms。滤波器组的固定参数为：最低频带通滤波器的中心频率为 100 Hz；滤波器组中所有带通滤波器的中心频率呈指数分布；带通滤波器为二阶滤波器，且滤波器组中的所有带通滤波器具有相同的品质因数。

滤波器组的可编程参数为 $N_{\text{滤波器}}$、$f_{\text{最大，滤波器组}}$，以及 $Q_{\text{滤波器}}$。这三个参数各自在一定范围内进行扫描，其余两个参数则设为表1中指定的典型值。取值范围如下：

$N_{\text{滤波器}}$ ={1、2、4、6、8、10、12、14、16、20、24、28、32、48、64}，$f_{\text{最大，滤波器组}}$ ={0。25、0。5、1、2、3、4、5、6、7、8、9、10、12、16、20}千赫兹，以及 $Q_{\text{滤波器组}}$ ={0。2、0。4、0。6、0。8、1、2、3、4、6、8、10、15、20、25、30、40、60}这些范围经过选择，使其远低于和远高于表1所示的最先进的模拟特征芯片所采用的数值。

47个滤波器组参数扫描点中的每一个都对应于模拟特征提取器模型的一个不同实例，因此也对应于一个不同的特征提取后数据集；该数据集是通过使用特征提取器实例处理特征提取前数据集而生成的。

3.3 后端神经网络

针对 47 个特征提取后的训练集，分别对后端神经网络进行重新训练，随后在 47 个特征提取后的测试集上进行测试。此外，在每个扫描点上，神经网络都会经过三次试验的重新训练和重新测试，以构建该点关键词识别准确率的置信区间。我们这样做是因为，迄今发表的模拟特征提取器芯片论文尚未报告带有置信区间的关键词识别准确率 [7, 11, 20, 22]。

为了将滤波器组的影响与后端神经网络的影响区分开来，所有滤波器组参数扫描点均采用相同的神经网络架构。所用架构是 [16], 中的小型残差卷积神经网络 res16，具体而言，是我们使用 MATLAB 深度学习工具箱实现的版本 [17]。选择这一架构，是因为在关键词识别演示中，已发表的模拟音频特征提取器芯片里，它的功耗和面积效率最高。

为区分滤波器组本身与训练算法超参数的影响，所有滤波器组参数扫描点均使用相同的超参数值。所用超参数值是在特征提取后的验证集上通过手动调优得到的；该验证集由采用典型滤波器组参数值（$N_{滤波器}=24$, $f_{最大，滤波器组}=7$千赫兹, $Q_{滤波器}=8$）的特征提取器模型生成。调优得到的超参数值为：动量为 0.9，学习率初始值为 1，且每经过一个训练轮次便乘以 0.9，训练 25 个轮次，小批量大小为 64，L_2 正则化系数为 0.001；所用训练算法为带动量的随机梯度下降。

4 仿真结果

图 3、4、5 展示了中间仿真结果：10 词关键词识别准确率随三个关注的模拟滤波器组架构参数变化的曲线。这些曲线旨在帮助确定模拟滤波器组可以缩减到多小，同时又不会显著降低分类准确率。图 6 展示了本文的最终仿真结果，即从相对功耗和经后端神经网络处理后的下游准确率两方面，对比典型的最先进滤波器组与所提出的微型滤波器组

如第 3.1 节所述，这里再重申一次：图 6 的最终仿真结果采用了较大的 28,000 个样本训练集，而图 3、4、5 的中间仿真结果采用了较小的 2800 个样本训练集，这是为了加快生成这些图表数据所需的 141 次训练会话。由于训练集规模不同，而且神经网络更适合使用更多训练数据，图 6 中典型滤波器组的准确率为 94.2%，比图 3、4、5 红线处的 90.9% 高出 3.3%；尽管这两个准确率对应的是同一个滤波器组，即典型滤波器组。

4.1 滤波器数量

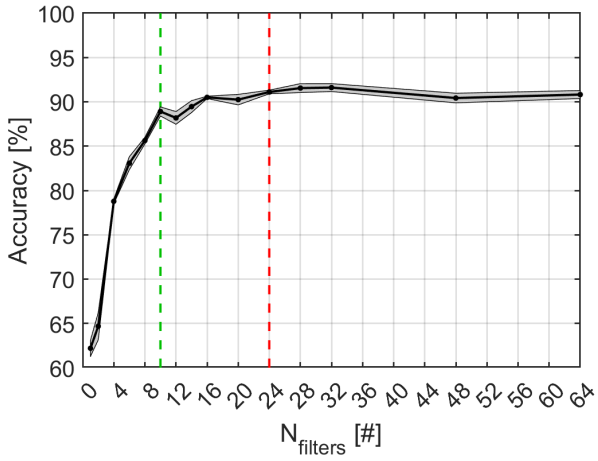


图3：10词关键词检测准确率与 $N_{滤波器}$数量的关系。结果基于2800个样本的训练集。图中显示了基于3次试验的95%置信区间上下界。红色表示最先进的模拟特征提取器芯片通常采用的选择，绿色表示本文提出的微型化选择。

图 3 展示了 10词关键词检测的分类准确率与滤波器组中的滤波器数量 $N_{滤波器}$之间的关系。正如预期，滤波器数量越少，准确率越低；滤波器数量越多，准确率越高，因为处理的频率子带分别更少和更多，这一点可从图2d和图2a中看出。根据启发式推理，特征数量增加时，准确率的边际收益会递减，这本在意料之中；但令人惊讶的是，准确率会出现平台期：多达 64 个滤波器时的准确率，与少至 16 个滤波器时的准确率大致相同。这表明，10词关键词检测使用滤波器数量超过 16 的滤波器组 $N_{滤波器}$ 会浪费功耗。最后值得注意的是，即使只有单个特征，准确率也达到 62.2%；作为对照，随机关键词检测的准确率为 9.1%（共 11 个类别，包括 10 个关键词和未知类别）。

4.2 滤波器组的最高频率

Fig. 图 4 展示了 10词关键词检测的分类准确率与滤波器组最高频率 $f_{max,filterbank}$ 的关系图。如预期所示，$f_{max,filterbank}$ 较低时准确率较低，而$f_{max,filterbank}$较高时准确率较高，因为分别处理的是语音带宽中较窄和较宽的部分，如图2e和图2b所示。与 $N_{filters}$ 关系图（图 3）中平台的出现令人惊讶不同，这种$f_{max,filterbank}$ 情况下的平台并不令人意外，因为语音带宽限制为4千赫； [8], 这意味着，如果$f_{max,filterbank}$ 设置得过高，最高频率滤波器输出就会是零值，传递零信息，因此不会提高准确率。然而，考虑到语音带宽为4千赫，令人惊讶的是，准确率直到低于仅2千赫时才会明显下降。这表明，对于10词关键词检测，若滤波器组的$f_{max,filterbank}$ 高于2千赫，就会浪费功耗。

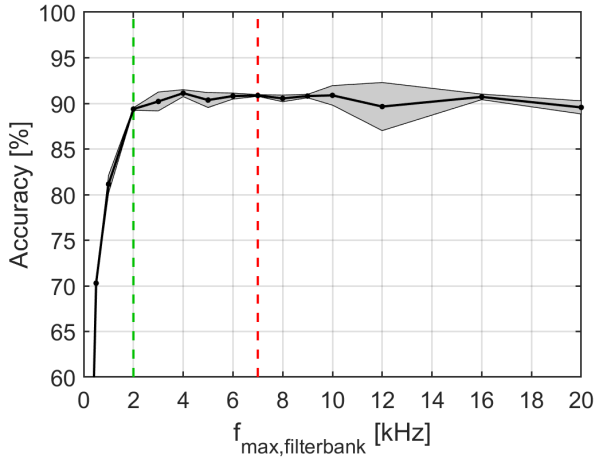


图4：10词关键词检测准确率与 $f_{\text{最大值、滤波器组}}$的关系。结果基于2800个样本的训练集。图中显示了基于3次试验的95%置信区间界限。最先进的模拟特征提取器芯片所采用的典型选择以红色表示，所提出的微型选择以绿色表示。

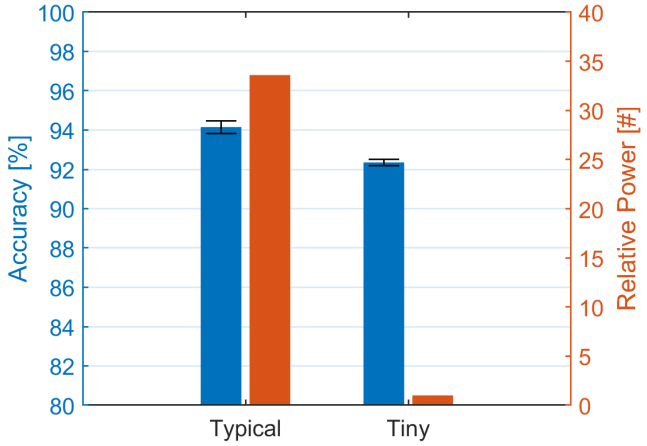


图6：典型的最先进模拟滤波器组与所提出的微型模拟滤波器组在相对功耗和通过后端神经网络实现的下游10词关键词检测准确率方面的比较。结果基于规模更大的28000个样本的数据集。图中标注了基于5次试验的95%置信区间。

4.3 滤波器的品质因数

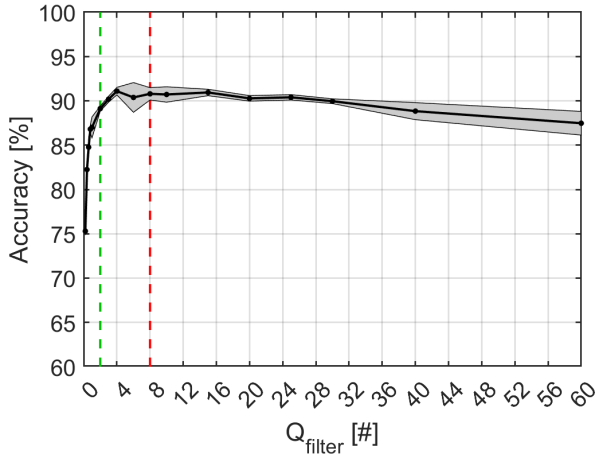


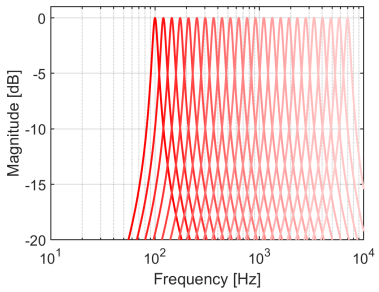
图5：10词关键词检测准确率与 $Q_{\text{滤波器}}$的关系。结果基于2800个样本的训练集。图中显示了基于3次试验的95%置信区间界限。最先进的模拟特征提取器芯片所采用的典型选择以红色表示，所提出的微型选择以绿色表示。

图 5 展示了10词关键词检测的分类准确率与滤波器组中滤波器的品质因数Q_{filter}之间的关系图。如预期所示，Q_{filter} 较低时准确率较低，因为滤波器的选择性较差，彼此之间会变得冗余，如图 2f 所示。现在请注意，Q_{filter} 过高时，准确率会下降。这同样在意料之中，因为当其过高时 $Q_{\text{滤波器}}$，滤波器组的通带极窄，因此通带间频率间隔极宽，如图 2c 所示，而信号能量会在这些间隔中损失。因此，应该存在一个最优的$Q_{\text{滤波器}}$，图 5 显示该值为 4；低于此值时，准确率会迅速下降（高于此值时，准确率则缓慢下降）。这表明，对于 10词关键词检测，采用$Q_{\text{滤波器}}$ 数量超过 4 的滤波器组会浪费功耗。

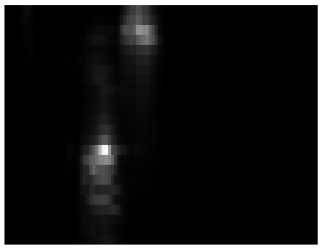
4.4 典型滤波器组与微型滤波器组

总而言之，图表明，模拟滤波器组架构参数的典型取值（以红线标示）可以大幅降低至绿线所示的取值，从而构成最终的“微型”滤波器组，其 $N_{\text{滤波器}} = 10$，$f_{\text{最大值,滤波器组}} = 2$ 千赫兹，以及 $Q_{\text{滤波器}} = 2$。由于这些微型取值相对于典型取值分别降低了 2.4×、3.5× 和 4×，因此微型滤波器组的功耗应低于典型滤波器组。回顾第 2.2 节，滤波器组功耗在一阶近似下与这三个参数中的每一个成正比。

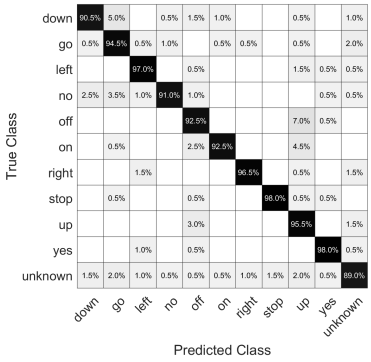
图 6 显示，与典型滤波器组相比，功耗低 33.6× 的微型滤波器组经由后端神经网络使下游 10词关键词检测准确率仅下降 1.8%，相对于如此显著的功耗节省而言，这是可以接受的代价。图 7 进一步从频率响应 (a)(d)、声谱图 (b)(e) 和混淆矩阵 (c)(f) 三个方面比较了典型滤波器组和微型滤波器组。频率响应图直观地显示，微型滤波器组 (d) 的滤波器数量更少、最高频率更低，品质因数也低于典型滤波器组 (a)。因此，微型滤波器组的声谱图 (e) 相较于典型声谱图 (b)“分辨率更低”。即便如此，两个



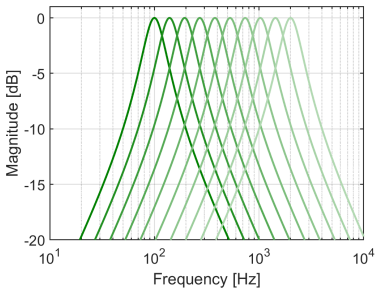
（a）典型滤波器组的频率响应



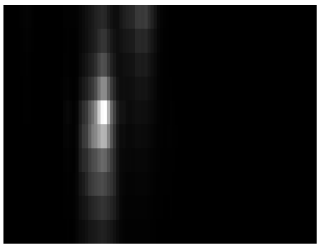
（b）使用典型滤波器组时“yes”的声谱图



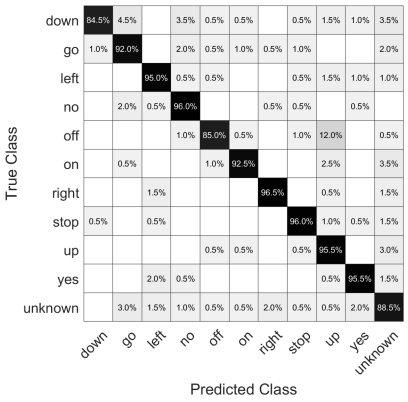
（c）使用典型滤波器组时的混淆矩阵



（d）微型滤波器组的频率响应



(e) 使用微型滤波器组得到的“yes”声谱图



(f) 使用微型滤波器组得到的混淆矩阵

图 7：典型的最先进的模拟滤波器组与所提出的微型滤波器组在频率响应方面的比较（a）（d）；一个示例声谱图，横轴为时间、纵轴为频率（b）（e）；以及使用以后端神经网络为输入、输入为频谱图的混淆矩阵（c）（f）。

区域仍然可以在微型声谱图 (e)[5] 中辨认出来：它们分别对应于“yes”中较低频的“ye..”部分和较高频的“...s”部分，而这两个区域在典型声谱图 (b) 中肉眼清晰可见。这一示例有助于理解微型滤波器组为何不会显著降低准确率。最后，为求完整，(c) 和 (f) 分别展示了典型滤波器组和微型滤波器组对应的混淆矩阵。

5 讨论

33.6× 滤波器组功耗降低幅度很大，而 1.8% 的准确率下降幅度很小，这表明这种权衡是值得的。不过，滤波器组功耗降低是否值得以 1.8% 的准确率下降为代价，最终取决于应用需求。此外，需要说明的是，本文结果针对的是所选用的后端神经网络，即 [16] 于 4 年前发表的 res15 小型残差卷积神经网络。虽然换用其他神经网络（例如更现代的网络）时结果可能会有所不同，但核心结论很可能仍然不变：以小幅准确率下降为代价，可以实现大幅功耗降低。

[5]不过，不得不承认，较高频率的“...s”区域很难辨认。

最后，我们提及一些有趣的相关研究：[3] 从信息论角度分析滤波器组特征，而 [9] 比较手工设计的滤波器组特征与学习得到的滤波器组特征。

6 结论

在采用前端模拟特征提取器和后端数字神经网络的关键词检测系统中，我们通过仿真发现，只需仔细选择其架构参数的取值，就能将最先进的模拟特征提取器芯片中典型模拟滤波器组的功耗降低 33.6×，而通过神经网络实现的下游10词关键词检测准确率仅下降1.8%。

在探究滤波器组架构参数对关键词识别准确率的影响之后，本研究为进一步探究由信噪比和线性度衡量的滤波器组非理想特性对关键词识别准确率的影响奠定了基础。此外，本研究还提出了许多新的

问题；例如，当关键词数量增加到 30 个时，结果会如何变化；或者，对于语音活动检测、唤醒词检测或说话人识别等其他音频识别任务，结果又会如何变化。最后，尽管本研究聚焦于前端特征提取器并对其进行了优化，但也为与后端神经网络进行协同优化铺平了道路。

致谢

感谢 Rui Xu 提供仿真支持，感谢 Rebecca Zhang 提供软件神经网络实现方面的支持。本研究部分由美国国家科学基金会 1704899 和亚德诺半导体资助。

参考文献

[1] Komail Badami、Kushal Dakshina Murthy、Pieter Harpe 和 Marian Verhelst。2018。用于 65NM 互补金属氧化物半导体低功耗传感应用的 0.6 V、54DB 信噪比模拟前端，THD 为 0.18%。载于*2018 IEEE* 超大规模集成电路研讨会。电气电子工程师学会，241–242。[2] Komail MH Badami、Steven Lauwereins、Wannes Meert 和 Marian Verhelst。2015。用于语音活动检测的 90 nm 互补金属氧化物半导体、6uW 功率比例型声学传感前端。*IEEE* 固态电路期刊 51，1（2015），291–302。[3] Boris Bergsma、Minhao Yang 和 Milos Cernak。2021。PEAF：用于音频识别的可学习高能效模拟声学特征。*arXiv* 预印本 *arXiv:2110.03715*（2021）。[4] Guoguo Chen、Carolina Parada 和 Georg Heigold。2014。使用深度神经网络进行小型关键词检出。载于*2014 IEEE*国际声学、语音与信号处理会议（*ICASSP*）。电气电子工程师学会，4087–4091。[5] Hiroshi Fuketa。2022。采用开关电容带通滤波器、最大值运算符和神经网络处理器的超低功耗关键词检出特征提取器。*IEEE*固态电路快报 5（2022），82–85。[6] Juan Sebastian P Giraldo、Steven Lauwereins、Komail Badami 和 Marian Ver-helst。2020。Vocell：一款用于 10-μW 关键词检出和说话人验证的 65-nm 语音触发唤醒片上系统。*IEEE* 固态电路期刊 55，4（2020），868–878。[7] Kwantae Kim、Chang Gao、Rui Graca、Ilya Kiselev、Hoi-Jun Yoo、Tobi Delbruck 和 Shih-Chii Liu。2022。一款采用环形振荡器时域特征提取的 23uW 关键词检出集成电路。*IEEE* 固态电路期刊 （2022）。[8] Philipos C Loizou。1998。模仿人耳。*IEEE*信号处理杂志 15，5（1998），101–130。[9] Iván López-Espejo、Zheng-Hua Tan 和 Jesper Jensen。2021。探索用于关键词检出的滤波器组学习。载于*2020*年第*28*届欧洲信号处理会议（*EUSIPCO*）。电气电子工程师学会，331–335。[10] Qianyun Lu 和 Boris Murmann。2022。使用对数梯度输入图像提高 tinyML 计算机视觉的能效和鲁棒性。*tinyML*研究研讨会论文集（2022）。[11] Subhajit Ray 和 Peter R. Kinget。2022。基于时域模拟滤波器组插值和时域模拟整流的超低功耗、小面积模拟音频特征提取。*IEEE* 固态电路期刊（ 2022 1 12 10 1109 2022 3227246），–。[12] 2015 https://doi.org/ 。/JSSC. 。Tara Sainath 和 Carolina Parada。。用于小型关键词检出的卷积神经网络。（2015）。[13] Rahul Sarpeshkar、Christopher Salthouse、Ji-Jon Sit、Michael W Baker、Serhii M Zhak、TK-T Lu、Lorenzo Turicchia 和 Stephanie Balster。2005。一款超低功耗可编程模拟仿生耳处理器。*IEEE* 生物医学工程汇刊 52，4（2005），711–727。[14] Weiwei Shan、Junyi Qian、Lixuan Zhu、Jun Yang、Cheng Huang 和 Hao Cai。2022。AAD-KWS：一款采用嵌入梅尔频率倒谱系数中的声学活动检测器和可调检测窗口的 28-nm 互补金属氧化物半导体亚-μW 关键词检出芯片。*IEEE* 固态电路期刊（2022），1–10。https://doi.org/10.1109/JSSC. 2022.3197838[15] Erjia Shi、Xian Tang 和 Kong Pang Pun。2020。一款用于常开语音活动检测的 270 nw 开关电容声学特征提取器。*IEEE* 电路与系统汇刊 *I*：常规论文 68，3（2020），1045–1054。[16] Raphael Tang 和 Jimmy Lin。2018。用于小型关键词检出的深度残差学习。载于*2018 IEEE*国际声学、语音与信号处理会议（*ICASSP*）。电气电子工程师学会，5484–5488。[17] The MathWorks Inc. 版本 14.2。深度学习工具箱。纳蒂克，马萨诸塞州，美国。https://www.mathworks.com/products/deep-learning.html [18] Yannis Tsividis。2018。信噪比、动态范围和功耗：关注它们之间的相互关系可极大助益模拟电路设计。*IEEE*固态电路杂志 10，4（2018），60–69。[19] Marian Verhelst 和 Ahmad Bahai。2015。模拟与数字的交汇：模拟到信息转换及其后续发展。*IEEE*固态电路杂志 7，3（2015），67–80。[20] Daniel Augusto Villamizar、Dante Gabriel Muratore、James B Wieser 和 Boris Murmann。2021。一款用于声音分类的 800 nW 开关电容特征提取滤波器组。*IEEE* 电路与系统汇刊 *I*：常规论文 68，4（2021），1578–1588。[21] Pete Warden。2018。语音命令：一个用于有限词汇语音识别的数据集。*arXiv* 预印本 *arXiv:1804.03209*（2018）。[22] Minhao Yang、Hongjie Liu、Weiwei Shan、Jun Zhang、Ilya Kiselev、Sang Joon Kim、Christian Enz 和 Mingoo Seok。2021。利用非线性模拟特征提取实现纳瓦级声学推理传感。*IEEE* 固态电路期刊 56，10（2021），3123–3133。[23] Minhao Yang、Chung-Heng Yeh、Yiyin Zhou、Joao P Cerqueira、Aurel A Lazar 和 Mingoo Seok。2019。借助定制软件模型辅助模拟特征提取的常开深度神经网络式 1uW 语音活动检测器设计。*IEEE* 固态电路期刊 54，6（2019），1764–1777。