

Dr. Paul Beckmann,
Aaron Grassian,
Matt Crowley



白皮书

超低功耗解决方案助力始终在线的语音命令系统应用



便携式和电池供电产品中应用语音命令系统的综合指引

内容

摘要	3
简介	7
什么是始终在线的语音系统?	7
对于便携式产品来说, 什么是最低功耗的解决方案?	7
始终在线语音命令的主要应用	9
遥控器	9
汽车	9
耳戴式设备	9
智能家居设备	9
可穿戴设备	10
便携式产品中始终在线语音命令的挑战	11
高功耗	11
电池寿命预期	11
互联网连接不可靠	12
外形尺寸妥协	12
环境因素影响	12
便携式产品中语音命令的硬件注意事项	14
麦克风阵列设计	14
音频处理器考虑	17
附加组件	18
工业设计考虑	20
便携式产品中用于语音命令的软件算法	21
基本算法结构	21
算法调整	24
参考设计	27
结论	28

摘要

现在，消费者比以往任何时候都更需要始终在线的语音命令产品，以实现他们与数字世界之间安全、自然的互动。COVID-19 在全球肆虐过后，人们正在转向无需手动操作的设备，以避免与公共屏幕、按钮和控制器发生物理接触。

由于人们对公共表面可能传播疾病的意识增强，始终在线的语音控制设备正在迅速应用于智慧城市、智能家居和工业应用中。

然而，直到高能效的硬件和软件开发方面取得进展，便携式和电池供电的产品才真正使用始终处于监听状态的语音命令系统。本文论述了在教育、技术和工艺方面的一些突破，使得超低功耗语音命令产品得以实现。

Ambiq® 的 SPOT®

在电池供电的产品中，外形尺寸和电池寿命是主要关注点，语音命令系统需要一个能够处理大量音频任务的超低功耗处理器。

在**电池供电的产品**中，语音命令系统需要一个能够处理大量音频任务的处理器。

Ambiq 的 Apollo SoC 系列是始终在线感知语音命令系统的超低功耗解决方案的优秀案例。这些 SoC 采用 Ambiq 的专有 SPOT® 平台（Subthreshold Power Optimized Technology，亚阈值功耗优化技术）设计，它们的运行电流不到传统音频处理器的 1/10。

具体来说，Apollo 产品（如图 1 所示）专注于超低功耗和始终在线感知的语音命令处理，这使得它们非常适合于超低功耗的耳戴式设备、可穿戴设备或其他移动应用。

图 1: Apollo3 Blue 产品



汽车、遥控器、智能家居设备和可穿戴设备是始终在线语音命令的一些热门应用。

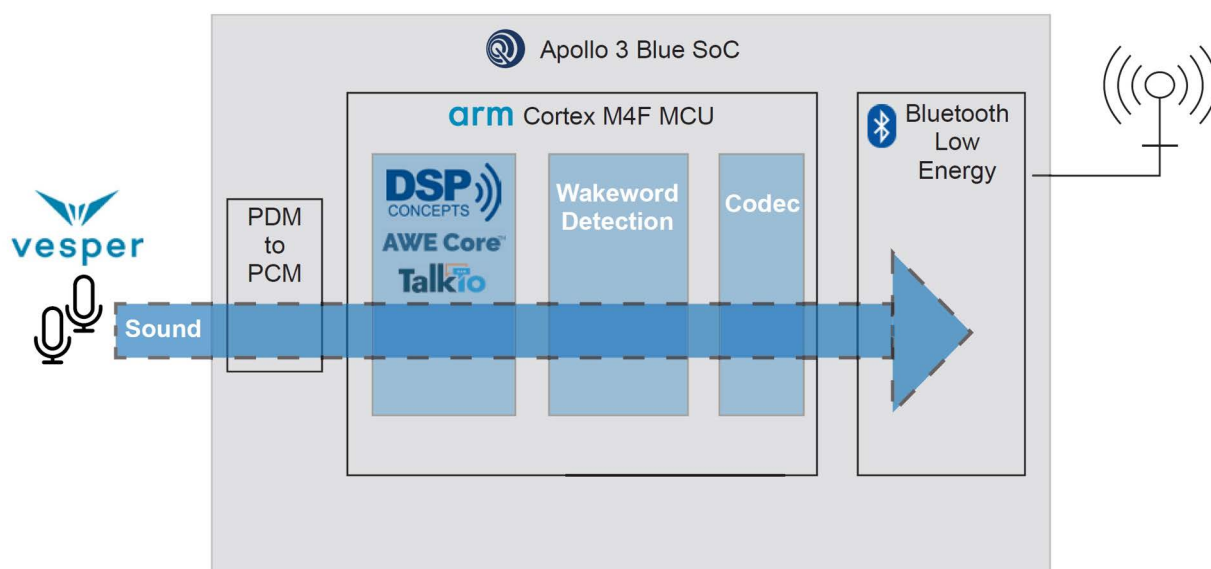
VESPER 的自适应 ZPL™ 技术

Vesper 的麦克风提供了前所未有的自适应 ZeroPower Listening™ 引擎，它可以主动监测声音级别，并根据特定的音频事件激活休眠中的音频处理器。这一强大功能使整个系统的功耗减少了 10 倍。当 SPOT® 和 ZPL™ 技术结合时，它们为由小电池供电的设备提供了极致超低功耗始终在线的解决方案。

DSP Concepts 的 Audio Weaver

DSP Concepts 的 Audio Weaver 平台使用简单的拖放式用户界面，可以在最先进的嵌入式处理器上快速构建支持语音命令的设计原型。他们的 TalkTo™ 音频前端算法即使在日常生活最嘈杂的环境中也能提供可靠的语音控制，支持消费电子、汽车以及医疗设备等应用。

图 2: 通过 Apollo 3 Blue SoC 的语音数据流



应用

尽管使用始终在线语音命令的概念并不新鲜，但直到最近，技术进步才使得便携式和电池供电的产品能够具有此功能。遥控器、智能手表、便携式智能音箱、联网家居设备和可穿戴技术是始终在线设备的一些热门应用。

汽车、遥控器、智能家居设备和可穿戴设备是始终在线语音命令的一些**热门应用**。

挑战

但是，对于提供始终在线语音命令的电池供电产品，制造商必须克服严峻的技术挑战。因为这些产品始终在监听，所以必须大幅降低功耗。

普通的科技消费者期望一次充电至少可以使用一整天，制造商也在不断受到提高电池续航性能的压力。像智能手表这样的一些产品预计可以持续数周，而像语音遥控器这样的其他产品至少必须持续一整年。

制造商可能需要在外形尺寸上做出妥协，以容纳超大电池，同时还要考虑到那些网络连接不稳定的用户。

硬件考虑

语音命令系统的核心硬件包括一个或多个并行工作的麦克风（麦克风阵列）以及从麦克风接收音频信号的处理器。选择麦克风时，尺寸、成本、可靠性和功耗是主要因素。对于音频处理器，它必须具有足够的计算能力来处理来自麦克风阵列的信号，并运行语音识别所需的所有算法。

选择麦克风时，**尺寸、成本、可靠性和功耗**是主要因素，而**计算能力**对音频处理器至关重要。

除了麦克风阵列和音频处理器，始终在新的语音命令产品还需要额外的组件，这些组件可能会根据产品和预期应用而有所不同。

然而，几乎所有的语音命令产品都需要一个无线接口，可以从云服务器发送和接收数据（通过访问互联网），以提供更多的功能。这些产品通常还具有音频反馈，可通过提示音或语音合成来确认用户的命令。产品的物理设计也会对其语音识别系统的性能产生巨大的影响。

软件算法

始终在线的语音命令产品中有许多不同的算法在工作。算法必须时刻监听唤醒词，当检测到声音时，从周围的噪音中隔离出用户的声音，然后为唤醒词检测引擎产生干净的信号，以可靠地识别唤醒词。

语音识别算法需时刻监听唤醒词，并能够**可靠地识别**唤醒词。

大多数语音命令产品都包含某种形式的用户反馈，以确认设备是否处于活动状态、是否正确理解了用户的命令以及是否执行了所需的操作。为了更好地减小声学差异，麦克风应尽可能以相同的配置进行安装。

在语音命令产品中，算法包需要包含必要的组件，如声音检测、降噪和滤波、语音方向到达检测、波束成形、回声消除 (AEC)、自适应干扰消除 (AIC)、唤醒词检测和本地命令集识别。大多数语音命令产品可能还会整合播放处理，以适应用户反馈，确认设备处于活动状态并且正确执行了用户意图所需的操作。由于这些算法具有复杂的功能，其中一些需要调整以获得最佳的语音识别精度。

简介

在过去的十年里，始终在线的语音命令系统已经在数千万台智能音箱中证明了它们的吸引力和可靠性。然而，直到最近，始终在线的语音命令系统才能有效地应用于便携式和电池供电的产品中。得益于超低功耗解决方案的进步，如今的消费产品制造商可以通过始终在线的移动语音命令系统提供令人满意的体验。

本文将探讨始终在线的语音命令及其在便携式和电池供电产品中的应用。它还将讨论尝试在便携式产品中实施始终在线的语音命令系统时可能出现的潜在挑战。最后，本文将描述始终在线语音命令系统在硬件和软件算法方面的考虑因素，使其应用在便携式和移动产品中的效果，与在专为家用设计的交流插座供电的设备一样。

什么是始终在线的语音系统？

在典型的语音命令系统中，用户必须按下按钮（所谓的即按即说）来“唤醒”系统，而不是使用唤醒词。但是，在一个始终在线的语音命令系统中，系统始终处于开启状态，不断地监听特定的唤醒词或触发词，如“Alexa”或“Ok Google”。目前的“智能音箱”，如 Amazon• Echo、Apple• HomePod 或 Google• Home，都是这样工作的。

在始终在线的语音命令系统中，系统始终监听**特定的唤醒词或触发词**。

如今，始终在线的系统主要用于家用设备。由于功耗要求较高，这些系统由交流壁式插座供电。然而，超低功耗解决方案的发展使得便携式产品可以使用始终在线的语音命令，而不会迅速耗尽其电池电量。

对于便携式产品来说，什么是最低功耗的解决方案？

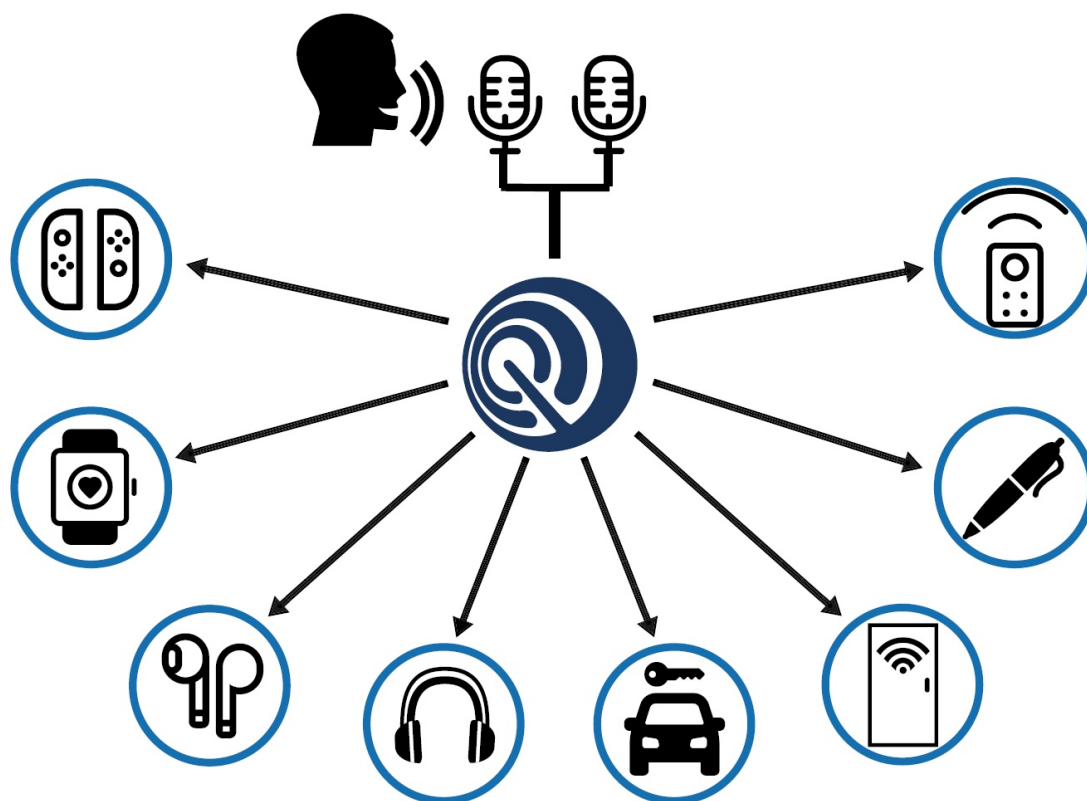
在 SPOT（亚阈值功耗优化技术）平台的支持下，Ambiq 能够帮助电子制造商延长电池寿命，并为小型化、电池供电的产品（例如智能手表、智能卡和物联网传感器）增加了新功能

Ambiq 的 Voice-on-SPOT[®]（简称 VoS），是耳戴式、可穿戴和其他便携式电池供电产品中始终在线的语音助手集成和命令识别的最低功耗解决方案。VoS 支持极低功耗的音频信号处理，为语音助手和设备上的语音用户界面（VUI）提供高质量的音频，同时对

电池寿命的影响最小。然后，用户可以在日常生活中随处部署高质量的唤醒词和命令短语识别。

Voice-on-SPOT[®] (VoS) 使用户能够随处部署高质量的唤醒词和命令短语识别。

图 3: Voice-on-SPOT[®] 使用案例



始终在线语音命令的主要应用

由于在便携式和电池供电产品中使用始终在线的语音命令的概念是新的，这项技术的应用才刚刚开始出现。其中一些可能性包括：

遥控器

目前大多数语音命令遥控器要求用户在说出命令之前按下按钮以唤醒系统。许多遥控器甚至要求用户将其靠近嘴才能捕获声音。而始终在线的系统允许用户在遥控器不在身边（甚至放在其他地方）时访问其功能。这种新一代的遥控器可以为 Netflix™ 带来全新的舒适体验。



汽车

在配备了始终在线的语音命令的车辆中，驾驶员无需将手从方向盘上移开，也无需摸索大多数语音命令系统中使用的唤醒按钮。这使得车辆操作更安全，并且更容易控制车辆的功能，例如 GPS、环境控制、娱乐和打开后舱门等遥控操作。此外，其他乘客也同样可以访问语音命令系统。



耳戴式设备

始终在线的语音命令允许用户开始 / 停止播放音频节目、选择、跳过 / 重复音乐曲目、接电话或访问个人助理功能。通过将耳机或耳麦（通过蓝牙）与联网的智能手机配对，所有这些任务都变得简单。

始终在线的语音命令允许用户开始 / 停止 / 跳过音乐曲目、接电话或访问个人助理功能。

由于无需按下按钮，用户可以腾出双手来执行其他任务，从而使语音命令耳机成为运动 / 健身、办公室工作或作为入耳式个人助理的理想选择。



智能家居设备

通过配备始终在线语音命令的壁挂式控制面板，用户可以在语音范围内从任何地方控制家庭系统，如 HVAC、照明和安全。用户不需要物理访问面板或使用智能手机调用必要的控制应用，所以不需要昂贵的触摸屏。此外，始终在线的

语音命令还可用于其他智能家居产品，包括自动开闭的垃圾桶或语音控制的窗帘和遮阳帘。



可穿戴设备

始终在线的语音命令系统可以使多种类型的可穿戴设备受益。在健身追踪器或项圈式设备中，始终在线的语音命令允许用户在跑步、步行或锻炼时控制设备，而无需伸手去触摸控制器。语音命令在较小的产品中尤其实用，这些产品可能没有足够的空间放置大型显示屏和控制按钮。

语音命令在体积较小的产品中尤其实用，这些产品可能没有足够的空间放置**大型显示屏和控制按钮**。

小型的可穿戴设备可以作为腕上私人助理，或作为与智能音箱或其他超出语音范围的设备的接口。数据可以通过蓝牙传输到配对的手机，或通过 Wi-Fi 传输到本地网络。

新兴的 AR 应用可以使用语音作为无缝的用户界面模式。锻炼时，用户可以不带智能手机，但不会失去智能手机的体验。用户可以将音乐传输到耳机，用声音开始和停止锻炼计时器，或从手表直接将音乐传输到耳机。

便携式产品中始终在线语音命令的挑战

目前，在许多便携式和电池供电的产品中实现语音命令面临着一些相当大的挑战。这些挑战包括：



高功耗

始终在线的语音命令系统必须一直处于活动状态，以随时捕捉用户的命令。对于插入交流电源的智能音箱，这不是问题。但在电池供电的产品中，这是一个更大的挑战，特别是当电池续航时间是科技消费者的主要关注点时。

对于工程师来说，减少功耗也是一个挑战，因为他们通常需要最小化电池大小以保持紧凑的外观。特别是，负责识别唤醒词的处理器必须快速响应，并且至少有一个麦克风始终处于活动状态。

在始终在线的系统中，负责识别唤醒词的**处理器**以及至少一个**麦克风**必须始终处于活跃状态。

在大型系统中，专用组件可以隔离其中的一些功能，这允许设备的大部分其他组件在设备空闲时关闭电源。而较小的便携式产品往往依赖于系统级芯片（SoC），其中单个组件执行设备的几乎所有功能。在这些产品中，可能很少或根本没有可以关闭的非活动组件。



电池寿命预期

大多数科技消费者期望产品至少可以使用一整天（8 小时）而不需要充电或更换电池。如今，大多数头戴式耳机和入耳式耳机可以运行 18 到 20 小时，而廉价型号通常可以运行 10 小时。一些最新的可穿戴设备，如 TWS 耳机和夹式无线扬声器，电池续航时间在 5 小时左右。无论如何，制造商都面临着来自消费者和评论员的巨大压力，要求提高电池续航时间，特别是在作为控制接口的语音命令产品中。

目前，消费者期望遥控器的电池至少能使用 10 到 12 个月。添加“始终在线语音”等新功能后，许多遥控器从 AA 电池迁移到电量减半的 AAA 电池，同样也需要维持这一使用寿命。智能家居系统的壁挂式控制面板通常使用一组 AA 或 AAA 电池运行大约一年（甚至两年）。在壁挂式控制面板中使用可充电电池是不切实际的，并且期望消费者经常更换这些电池也是不现实的。



互联网连接不可靠

家用产品可以依赖于几乎始终存在的互联网连接，将大部分语音识别处理任务上传到外部服务器。然而，便携式产品通常需要通过低功耗蓝牙与智能手机配对才能访问互联网，因为蜂窝数据连接在许多地方并不可靠甚至不可能。而像暖通空调和安防系统这样的关键应用在偶尔的互联网中断期间也需要正常工作。

由于在便携式应用中的互联网连接不可靠，使用语音命令的便携式产品必须自行识别和处理少量的语音命令，而无需外部服务器的帮助。这不仅需要更强大的处理能力，而且也限制了语音命令的可控功能。

使用语音命令的便携式产品必须自行识别并处理有限的语音命令词汇。



外形尺寸妥协

大多数便携式产品的紧凑尺寸可能需要在阵列中使用较少数量的麦克风。这也可能迫使工程师以牺牲其性能的方式放置麦克风，从而损害它们的性能，并使多个麦克风的响应和灵敏度的精确匹配变得困难或不可行。

穿戴设备的外形尺寸限制可能迫使产品设计人员使用更少的麦克风并选择更小的电池。

穿戴设备和其他紧凑型便携产品的外形尺寸限制也迫使产品设计人员选择更小的电池。例如，典型的 AA 碱性电池可以提供 3000 mAH 的电量。相比之下，许多微型科技产品中使用的 CR2032 纽扣电池仅能提供 220 mAH 的电量。因此，由 CR2032 供电的产品在 10 mA 的平均电流消耗下只能运行 22 小时。



环境因素影响

家用产品比便携产品更少地暴露在恶劣的环境中。穿戴设备至少必须是防汗的，这需要 IPx5 的防护等级，而用于崎岖户外的便携产品应该是完全可浸泡

的 (IPx7 等级)。消费者希望便携产品能够经受住偶尔的暴风雨或洗衣机漂洗的考验。

然而，为达到这些防护等级所需的密封性能可能会损害麦克风的功能并限制麦克风阵列的设置。此外，便携产品中的语音识别麦克风需要能够承受从至少腰部高度跌落到硬表面（如混凝土）上的潜在冲击。

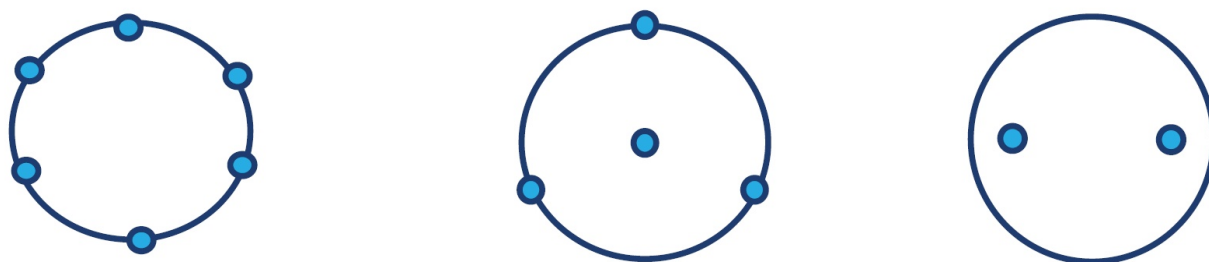
便携式产品中语音命令的硬件注意事项

声控接口的核心硬件由一个麦克风阵列和一个能够接收并处理麦克风音频信号的处理器组成。根据设备类型的不同，可能还需要各种其他组件，例如用于 BLE 或 Wi-Fi 的无线接口，加上扬声器、音频放大器、LED 和显示器以提供用户反馈。

麦克风阵列设计

尽管可以在语音命令产品中使用单个麦克风，但大多数这类产品使用两到七个麦克风的波束成形阵列，以更好地将说话者与环境噪音隔离开。该阵列允许音频处理器将麦克风的拾音模式集中在用户的声音上，从而提高用户的声​​音相对于周围环境噪音的信噪比。然而，便携式和电池供电产品的外形尺寸提出了许多在设计为插入式产品中未发现的挑战。

图 4: 典型的麦克风阵列模式



智能音箱设计中使用的典型麦克风模式包括：6 麦克风（高端）、3+1 麦克风（标准）和 2 麦克风（低成本）设备。麦克风阵列的直径可以在 40-75mm 之间。便携式智能音箱设计使用 4 麦克风或 2 麦克风配置，麦克风阵列位于产品的顶部。

图 5: 在电视等家用电器中常用的典型麦克风阵列模式



用于电视等家用电器的 2 或 4 麦克风线性阵列设计，2 麦克风的间距可以是 25-60mm，而 4 麦克风中每对麦克风之间的间距是 25mm，麦克风设计为面朝上或面朝前。

DSP Concepts 的白皮书“设计和优化麦克风波束成形”中提到，实现最佳的信噪比对于语音命令产品的准确性和可靠性至关重要。

该文献还发现，使用灵敏度公差更严格的麦克风可以通过使用 ± 1 dB 公差的麦克风（而不是更典型的 ± 3 dB）来帮助提高性能。由于阵列中的每个麦克风可能处于不同的声学

环境中（依赖于产品的物理设计），因此最好是匹配处理器的麦克风灵敏度，而不是匹配麦克风本身。同时，麦克风的灵敏度在产品的整个使用寿命中保持稳定也至关重要。

麦克风数量

DSP Concepts 的研究表明，增加麦克风的数量可以提高语音用户界面（UI）的可靠性。麦克风的灵敏度匹配得越一致，波束成形的性能就越好。

麦克风的灵敏度匹配得越一致，波束成形的性能就越好。

为了实现这一目标，最实用的方法是在安装麦克风后在硬件中平衡麦克风的灵敏度。这样，灵敏度调整不仅可以补偿麦克风的增益（通常精确到 ± 3 dB），还可以补偿外壳对麦克风产生的声学效应。

然而，很少有便携式产品和几乎没有穿戴设备有足够的空间容纳这样的阵列。例如，TWS 耳机通常每个耳机只有两个麦克风的可用空间，麦克风之间的可用间距只有 10 到 20 毫米。而且，这样的阵列所需的处理能力可能超出了大多数便携设备中使用的相对较小的处理器的能力。因此，执行波束成形和其他语音 UI 优化功能的软件算法必须具有针对两个或最多三个麦克风进行优化的能力。

麦克风选择

由于语音命令产品使用多个麦克风，因此这些产品中麦克风选择的主要因素通常是大小、成本和质量。然而，在便携式和电池供电的产品中，降低系统功耗变得至关重要。

选择语音命令产品的麦克风的主要因素通常是大小、成本、质量和功耗。

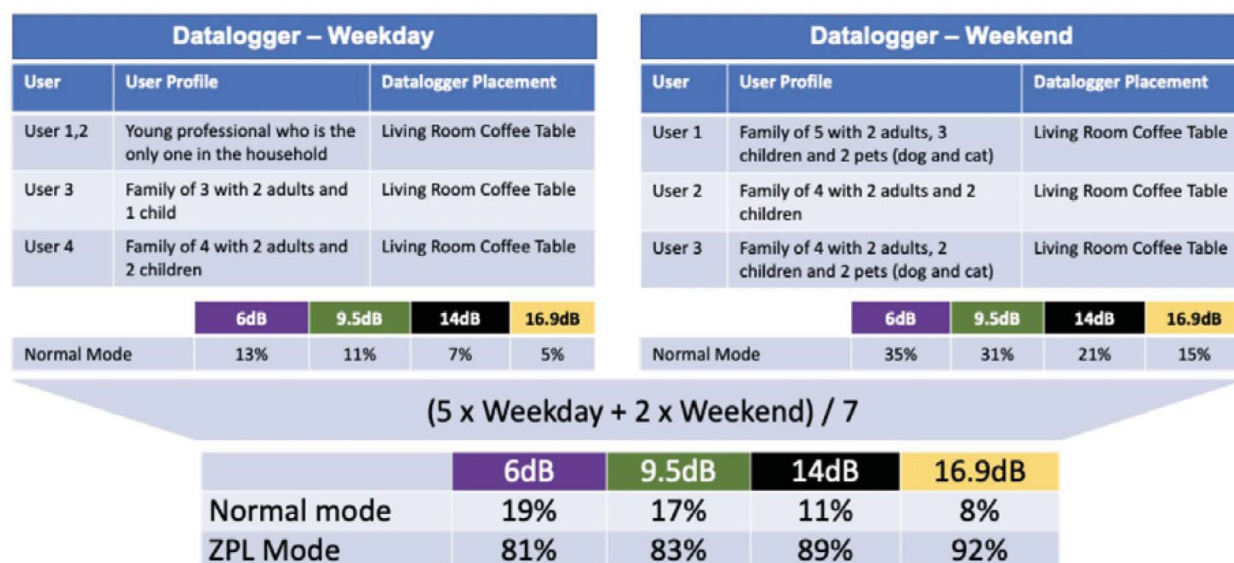
Vesper MEMS 压电麦克风已经在市场上以小尺寸、自适应 ZPL 和极具竞争力的成本推出，可满足便携式和电池供电产品的需求。Vesper 压电 MEMS 麦克风的另一个优点是它们非常稳定，因为灵敏度在回流焊、湿度或温度变化期间不会发生变化。所有 Vesper 麦克风均为 IP57 防护等级，启动时间为 200 微秒。

实际产品：Vesper VM3011 是首款采用第二代 Vesper ZPL（自适应型）技术的 MEMS 数字麦克风，处于“声音唤醒”模式时仅消耗 10 μ A 的电力。考虑到便携产品的电池即使在完全关机状态下通常也要消耗约 50 μ A 的电力，因此 VM3011 对便携产品的电池

寿命几乎没有任何影响。最重要的是，这些 ZPL 麦克风允许系统的其余部分在极低功耗模式下休眠，同时 ZPL 麦克风监测环境并在出现某些特定声音时唤醒处理器。

以下是 Vesper 进行的一项研究，通过 Vesper 数据记录器在工作日和周末从不同家庭收集数据数据（24 小时），每个数据记录器都装有 4 个具有不同的 WoS 阈值设置的 VM3011。如图 6 所示的结果表明，VM3011 允许始终在线，系统在 81% 到 92% 的时间内处于休眠状态。

图 6: Vesper VM3011 实现系统休眠时在线监听



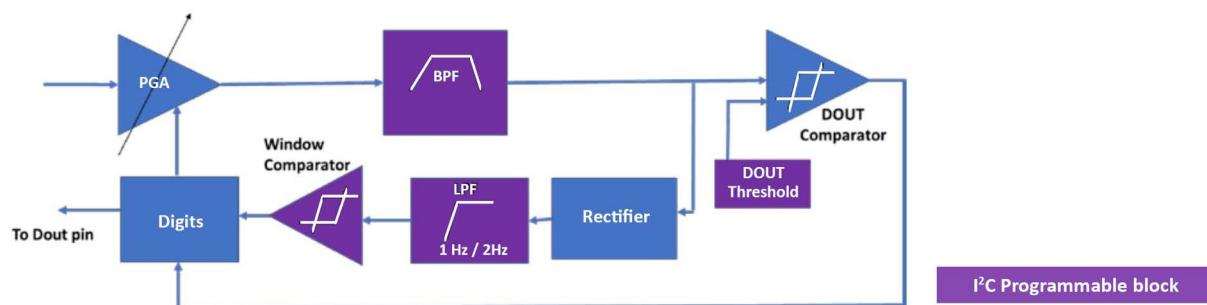
VM3011 中的压电麦克风元件由一个非常低功耗的模拟电路进行监控，该电路监测并跟踪环境噪声水平，仅当检测到背景噪声之上的人声时才激活系统。WoS（Wake On Sound）阈值允许为各种噪声环境优化 VM3011 以获得最佳性能。一个通带滤波器有助于缩小对 200 Hz 到 8 kHz 之间感兴趣的特定声音的选择，以更好地捕获人声并抑制机械隆隆声和风噪等环境噪声。该电路被称为“自适应 ZPL”。

自适应 ZPL 通过 I²C 提供多个可编程模块：

- 一. 带通滤波器，用于缩小特定声音的选择范围
- 二. WoS 阈值设定为比 RMS 水平高 2-8 倍
- 三. 低通滤波器，用于对 RMS 背景噪声的更新率进行改写
- 四. 窗口比较器，用于快速与慢速移动的背景噪声适应

图 7 以下是 ZPL 的简化框图：

图 7: 通过 I²C 接口实现多个 ZPL 可编程模块



单个压电麦克风可以触发语音命令产品的麦克风阵列、音频处理电路和网络连接（如有）。

音频处理器考虑

在任何语音命令产品中，音频处理器（无论是专用的 DSP 还是 SoC 内的一个处理核心）都必须具有必要的计算能力，以处理来自阵列中所有麦克风的信号，并运行语音识别所需的所有算法。

音频处理器必须能够处理所有麦克风的信号并运行语音识别算法。

算法越先进，芯片可容纳的麦克风越多，信噪比就越好，语音识别也就越准确。然而，在便携式和电池供电的产品中，处理器还必须消耗尽可能少的电力，以维持产品足够的电池寿命。

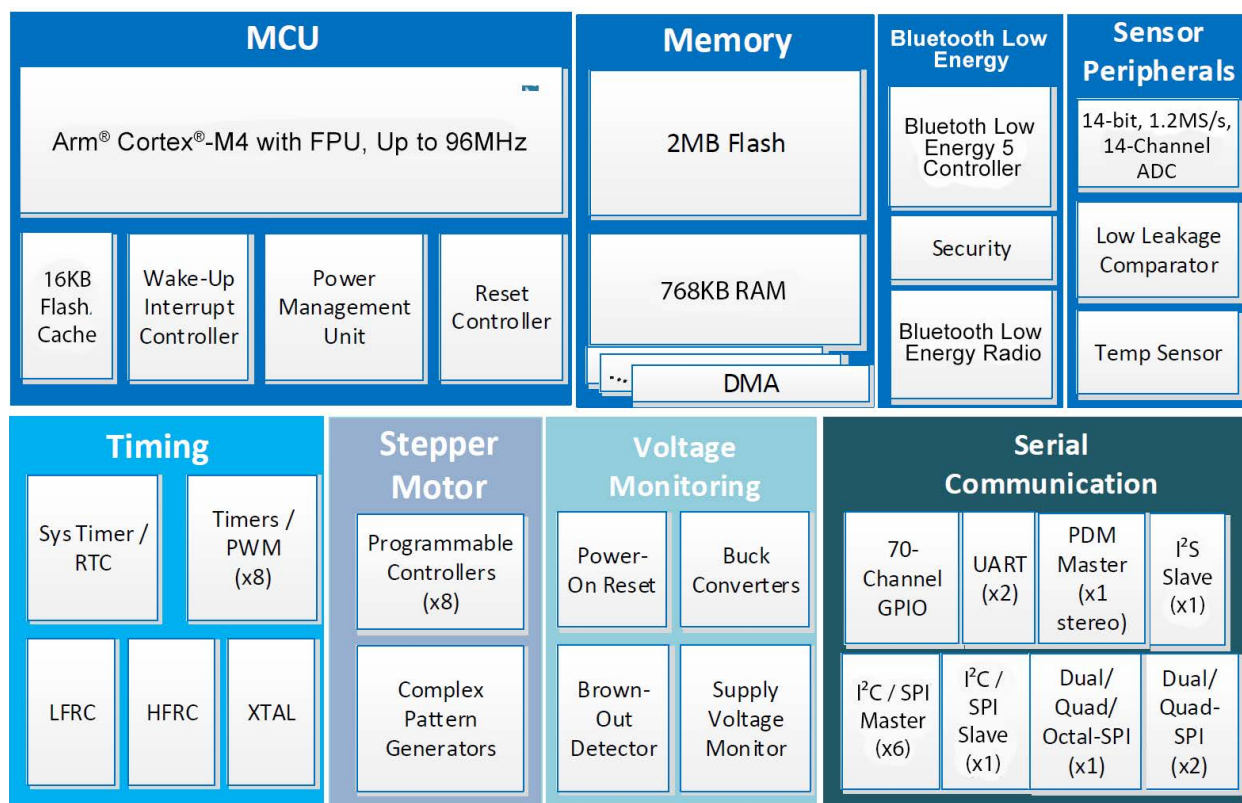
实际产品：为处理小型电池供电产品中的大量音频处理任务，Ambiq 推出了 Apollo 系列。这些微控制器单元 (MCUs) 和系统级芯片 (SoCs) 采用 Ambiq Micro 的 SPOT[®] (Subthreshold Power Optimized Technology, 亚阈值功耗优化技术) 平台设计，运行电流仅为传统音频处理器电流的 1/10。

基于 SPOT[®] 的 Ambiq Apollo2 系列是一款专注于传感器和语音处理的 48 MHz Arm[®] Cortex[®]-M4F 内核 MCU，动态运行功耗仅为 10 μ A/MHz。其中的 Apollo2 Blue 配有用于语音助手交互的低功耗蓝牙通道。

Apollo3 Blue 进一步将功耗降低至 6 μ A/MHz，并将主频提高到 96MHz。其低功耗蓝牙无线电部分符合蓝牙 5.0 标准。

Apollo2、Apollo2 Blue、Apollo3 Blue 和 Apollo3 Blue Plus 处理器能够使用 DSP Concept 的语音 UI 算法处理来自多麦克风阵列的信号，使它们完美适用于超低功耗耳戴产品、腕戴产品、遥控器以及其他移动终端应用。

图 8: Apollo3 Blue Plus 框图



Ambiq 将通过即将推出的处理器来扩大其在超低功耗领域的行业领先地位，这些处理器将大幅降低 $\mu\text{A}/\text{MHz}$ 并提高 F_{max} （主频）。新的处理器将实现更紧凑的波束成形模式、更好的信噪比和更高的语音识别准确度。

所有这些处理器都具有适合诸如手环、智能手表和耳塞式耳机等产品所需的紧凑尺寸，它们的尺寸视封装形态而有所不同，大小从 2.5 毫米到 5.3 毫米不等。

附加组件

除了麦克风阵列和音频处理器外，语音命令产品还需要其他组件。具体的组件要求取决于应用和外形尺寸，但几乎每个语音命令产品都使用了一些相似的组件。与麦克风和处理器一样，选择这些组件不仅要考虑其功能和性能，还要考虑其小尺寸和低功耗。

无线接口

语音命令产品需要从外部服务器发送和接收数据（通过访问互联网）以提供附加功能。为家庭使用而设计的智能音箱通过 Wi-Fi 连接到局域网，对于便携式语音命令产品，通过蓝牙或 BLE（低功耗蓝牙）与智能手机或平板电脑连接，然后通过蜂窝数据网络或 Wi-Fi 连接到互联网。

语音命令产品需要从外部服务器发送和接收数据，以提供比基本功能更多的功能。

用户反馈组件

大多数语音命令产品都融入了某种形式的用户反馈，以确认设备是否处于活跃状态、正确听到并理解用户的命令以及执行所需的操作。这些设备可以是 LED 灯，例如 Amazon Echo 和 Google Home 智能音箱顶部的闪烁灯。它们也可以是字母数字或图形显示，这些可以在许多遥控器和家居自动化面板上找到。

这些设备通常还具有音频反馈，可以通过警报音或语音合成来确认用户的命令，但这会增加处理器的负担。该设备必须使用一个放大器和某种类型的扬声器来再现语音或者警报音。有些产品甚至可能使用多个驱动器，结合波束成形算法，来将响应指向听众。

工业设计考虑

产品的物理设计对其语音识别系统的性能可能会产生巨大影响。根据 DSP Concepts 的研究，精确匹配麦克风的灵敏度对于可靠的波束成形器性能和准确的语音识别至关重要。

制造一致性

例如，如果阵列中的麦克风与产品边缘的距离不同，它们将具有不同的频率响应，因此在不同的频率上具有不同的灵敏度。为了更好地减少声学差异，应尽可能以相同的方式安装麦克风。围绕麦克风的任何密封件也必须在设计、材料和安装上保持一致。

制造商应仔细设计语音命令产品，以确保阵列中的每个麦克风都处于相似的声学环境中。制造商还可以通过在工厂单独对每个单元的麦克风进行电平匹配来确保语音命令产品的最佳性能。这一额外的质量控制步骤确保了因轻微的制造不一致性导致的麦克风性能差异不会影响到语音识别的准确性。

制造商应仔细设计语音命令产品，以确保阵列中的每一个麦克风都处于**相似的声学环境中**。

语音 UI 解决方案的成功设计依赖于稳定的麦克风阵列和传播信号链，以及优化的 MIPS 和内存使用。DSP Concepts Audio Weaver 软件配备了 RTASC（实时音频系统检查）模块，用于在设计周期中反复验证硬件和软件功能。借助易于使用的调试脚本，RTASC 实现了无风险且敏捷的产品开发流程，从而使支持语音命令的产品能够更快地上市。

便携式产品中用于语音命令的软件算法

始终在线的语音命令产品中有很多不同的算法在工作，所有这些算法都必须根据产品的设计和应用进行调整。这些算法必须全天候监听唤醒词，在检测到声音时将用户的声音从周围噪声中隔离出来，然后为唤醒词检测引擎产生干净的信号，以可靠地识别唤醒词。

基本算法结构

以下是语音命令算法包的基本组成部分，按照从麦克风端到最终信号输出的顺序呈现：

声音检测

通常，Vesper Adaptive ZPL 会监控来自单个麦克风的信号。当信号电平超过某个阈值（例如当用户说出唤醒词）时，麦克风会发送一个信号以启动系统的其余部分，这对于便携式产品非常重要，因为它允许关闭其他组件以节省电量。

当麦克风的信号电平**超过某个阈值时**，比较器会发送一个命令以启动系统的其余部分。

该功能还必须快速执行，以便系统可以接收唤醒词。例如，使用 Vesper 麦克风，麦克风在 200 微秒内唤醒，这远远小于说出任何关键词中第一个字母所需的时间。因此，不需要音频缓冲区。

降噪和滤波

为了在嘈杂的环境（例如典型的家庭场景）中实现语音检测 Vesper 设计了一个名为“Adaptive ZPL”的创新 IP。用户可以通过 I²C 接口来编程这个超低功耗模拟电路。Adaptive ZPL 的配置可以随时更改，且易于使用，因此可以与任何具有可用的 PDM 和 I²C 接口的应用处理器集成。

Adaptive ZPL 电路通过用户可编程的音频带滤波器过滤掉任何不需要的声音 / 噪声。同时，仅当检测到超过用户编程的 WoS 阈值的聲音时，它才会锁存 DOUT 引脚。在 Adaptive ZPL 中，WoS 阈值以 0.5 秒 / 1 秒的刷新率（也可由用户编程）持续跟踪并跟随 RMS 背景噪音，从而大大减少了误触发可能。

到达方向检测

由于语音命令产品使用多个麦克风，因此为这些产品选择麦克风的主要因素通常是尺寸、成本和质量。然而，在便携式和电池供电的产品中，降低系统功耗变得至关重要。

为了让麦克风阵列聚焦用户的语音，它必须首先确定用户相对于产品的位置。

麦克风阵列要聚焦用户的语音，必须首先确定用户相对于产品的位置。处理器通过比较来自麦克风的信号的相位信息来确定声音到达的方向。

波束成形

麦克风阵列可以处理来自多个麦克风的信号，使得该阵列具有方向性。它接受来自确定的到达方向的声音，同时拒绝来自其他方向的声音。

对于某些产品，如耳机和汽车音响系统，用户的语音相对于麦克风阵列的方向是已知的，因此波束成形器的方向可能是永久固定的。而在智能音箱、遥控器和家庭自动化墙面板等设备中，必须确定波束成形器所需的聚焦方向，并调整阵列的响应以便聚焦于用户的方向。。

回声消除 (AEC)

回声消除会拒绝来自设备本身的声音（如音乐或公告），以便麦克风阵列能更清晰地拾取用户的语音。由于原始信号和设备内部扬声器的响应是已知的，因此设备知道要拒绝通过麦克风返回的信号。

AEC 拒绝来自设备本身的声音，以便麦克风阵列能更清晰地拾取用户的语音。

选择具有高过载点的麦克风并最小化播放路径中扬声器的失真对于获得出色的 AEC 性能至关重要。同时，这会带来更好的音乐插播性能，特别是在大声播放低频内容时。DSP Concepts 的立体声 AEC 算法可消除音乐插播期间 35 dB 的回声，从而提高了唤醒词检测精度并改善了用户体验。

对于头戴式耳机和入耳式耳机等产品来说，AEC 不是必需的，因为来自产品扬声器的声音是封闭的，通常泄露的声音不足以影响产品麦克风的性能。

自适应干扰消除器 (AIC)

自适应干扰消除器可抑制诸如像客厅中播放的电视噪音或厨房里的微波炉噪音此类的干扰声音，这些干扰音很难用上面描述的传统波束成形器来消除。与其他自适应消除技术不同，DSP Concepts 的 AIC 算法不需要参考信号来消除干扰噪声。相反，它结合使用波束成形、自适应信号处理和机器学习来消除 30 dB 的干扰噪声，同时保留所需的语音信号。

AIC 对于遥控器和智能音箱等产品来说是必要的，它们通常在客厅环境中操作，那里存在干扰噪声和中度到高度的混响条件。例如，当用户操作电视遥控器时，AIC 能够从音频流中消除电视声音，以将语音信号呈现给唤醒词检测引擎，就好像不存在干扰噪声一样。

唤醒词检测

一旦系统检测到声音并开机，它必须记录传入的音频并将其与存储的唤醒词数字文件（例如 Amazon Echo 的 “Alexa”）进行比较。如果传入音频的波形与存储的文件足够接近，设备就会开始接受语音命令。

如果传入音频的波形与存储的文件足够接近，则设备可以**接受语音命令**。

与便携式产品相比，智能音箱只需要识别其唤醒词，因为它们将其他语音识别任务卸载给外部连接到互联网的服务器。唤醒词检测模型通常在设备本地运行，而亚马逊等一些服务提供商可以在云端启用额外的唤醒词检查。

本地命令集识别

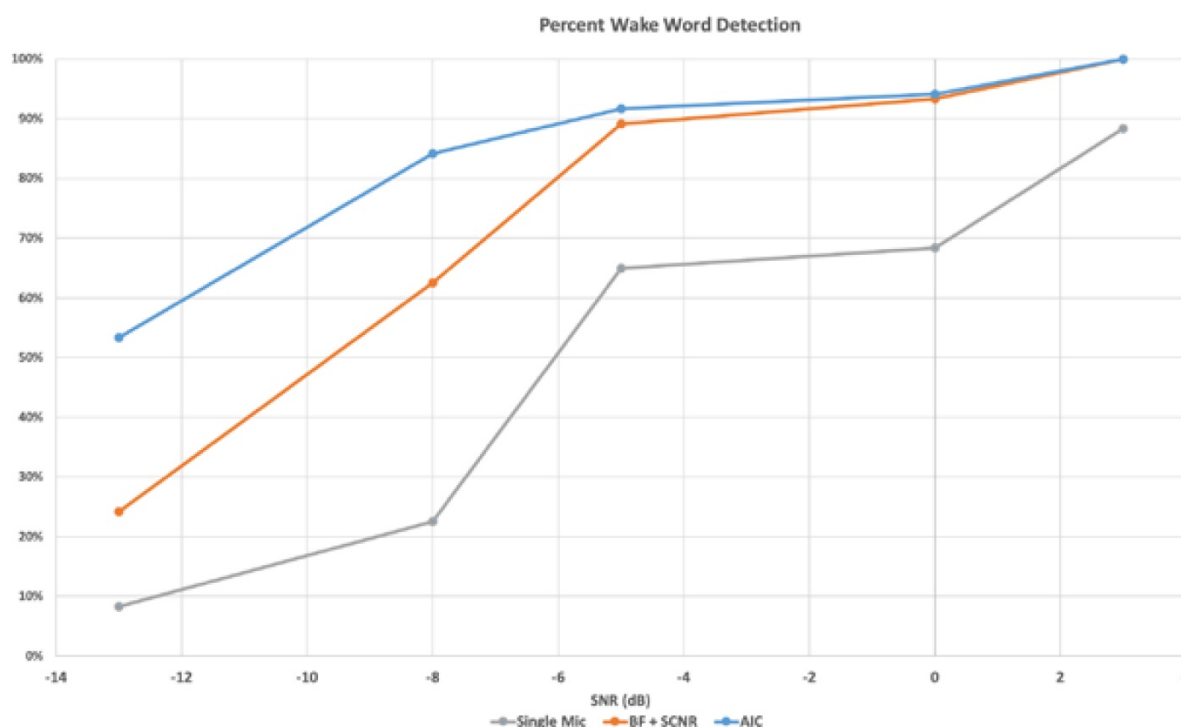
由于便携式产品无法像当今的智能音箱那样依赖互联网连接，所以它们需要在没有外部服务器帮助的情况下自行识别一定数量的基本功能命令。

便携式产品需要在**没有外部服务器的帮助下**识别一定数量的基本功能命令。

这些命令通常仅限于基本功能，例如播放、暂停、跳过曲目、重复和接听电话。这些命令的识别方式与唤醒词检测相同。然而，尽管命令集是有限的，但与智能音箱相比，本地命令集的识别增加了处理器的负担。例如可穿戴耳机这样的便携式设备可以通过蓝牙或 WiFi 与手机通信，然后由手机执行命令处理。

实际产品：图 9 显示了使用 DSP Concepts 的 TalkTo™算法在 Ambiq Apollo 3 SoC 上运行的遥控器参考设计的唤醒词检测率百分比。唤醒词检测百分比是衡量设计能准确识别多少唤醒词的度量。在测试期间，用户位于遥控器一米外，而音乐在离遥控器两米处的电视上播放。电视音量在 62-78 dB 声级之间变化，而麦克风阵列以 65 dB 声压播放语音，以达到图 9 所示的 SNR 值。

图 9: DSP Concepts TalkTo™软件配置遥控器用例的性能比较



结果表明，AIC 在信噪比极低的最具挑战性的噪声条件下表现最佳。单麦克风算法需要至少 3 dB 的 SNR 才能达到超过 80% 的唤醒词检测率。双麦克风波束成形器与单通道降噪 (SCNR) 算法在 0 dB SNR 时与 AIC 的表现相同。

然而，随着 SNR 的恶化，AIC 提供了明显更好的性能，在 -6 dB 时提高了 10%，在 -8 dB 时提高了 20%。在遥控器这样的应用中，由于无法使用参考信号来消除电视噪声，AIC 算法对非稳态噪声和混响的稳健性对于实现可靠的始终监听操作至关重要。

算法调整

上述每种算法的功能都很复杂。必须调整它们以适应应用场景，尤其是在便携式产品中，其环境和使用模式可能与家用产品有所不同。以下是需要调整以获得最佳语音识别准确性的算法功能：

检测 / 唤醒阈值

声音检测和唤醒词检测的阈值必须设置得足够高，以最小化设备的误触发，但又必须足够低，使用户能够以正常的说话音量对设备进行操作。这些唤醒阈值还取决于使用情况。

例如，与用户相距 2-3 英尺的遥控器的阈值应该设定得较低，而可穿戴设备的阈值则应该设定得较高以减少误报。特别是在便携式产品中，可能需要动态调整这些阈值，以便可以根据环境声音的不同水平进行性能调整。而动态补偿的功能本身也必须进行调整。

唤醒阈值必须**尽量减少设备误触发**，但仍能允许用户以平均说话水平来操作设备。

降噪 / 消噪

设备可以根据其应用进行调整以抑制不同类型的噪声。例如，制造商知道任何给定汽车在不同速度下的道路和发动机噪音的频谱，因此可以调整语音识别系统以拒绝这些声音。降噪 / 消噪算法还可以通过适应不断变化的环境来动态运行。然而，这种动态功能也需要调整。

始终处于语音命令状态的设备可以根据其应用来调整，**以抑制不同类型的噪声**。

波束成形器的波束宽度

波束成形器的波束宽度越窄，它就越能抑制环境声音和来自其他物体对用户声音的反射。但是，如果用户稍微移动，波束宽度设置得太窄会导致设备拒绝用户的声音。

波束成形器的**波束宽度越窄**，它就越能抑制环境声音和用户声音的反射。

在像耳塞和头戴式耳机这样的产品中，用户声音的到达方向不会变，因此可以设置一个较窄的波束宽度。但是，在像遥控器和家庭自动化面板这样的产品中，波束宽度必须设置得更宽，以适应用户说话时的移动。

唤醒 / 睡眠策略

最小化功耗的一个关键的目标是尽可能频繁地使设备进入休眠状态，并尽可能长时间地保持休眠状态。然而，这个目标需要权衡。如果设备在使用后过快进入休眠状态，它可能会错过跟随唤醒词的命令，并要求用户再次说出唤醒词，这通常会导致用户感到沮丧。但是，如果设备保持唤醒状态的时间超过了必要的时间，它则会消耗更多的电量。

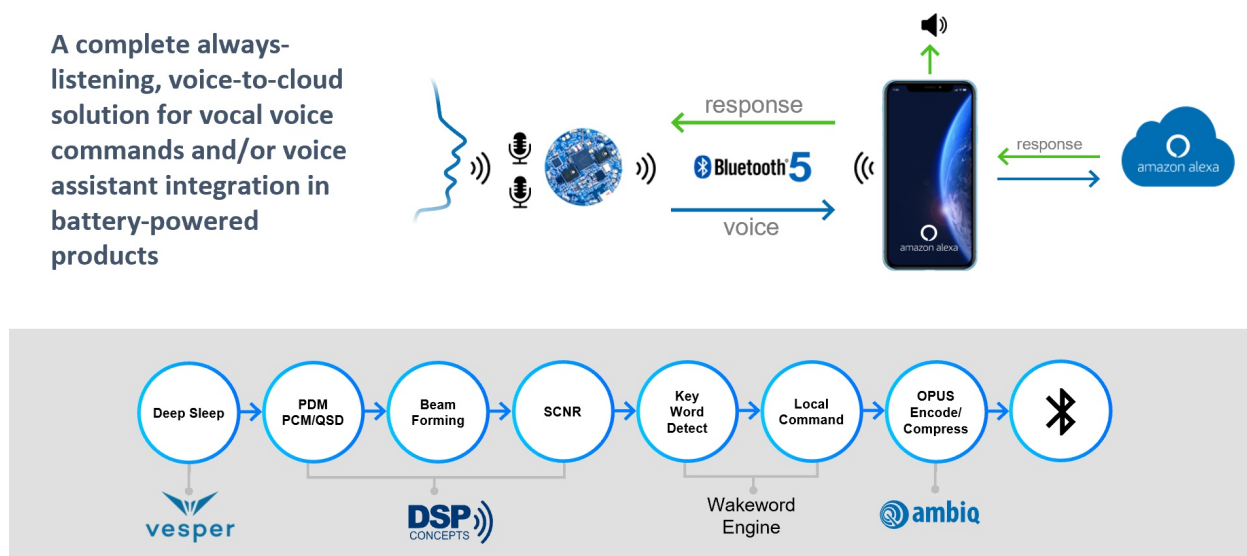
为了**最小化功耗**，语音命令设备应该尽可能经常休眠，并保持尽可能长时间的休眠。

参考设计

借助前面各节中描述的技术，我们可以为电池供电的产品提供一个完整的始终在线、语音到云的解决方案，以用于本地语音命令和 / 或语音助手集成。

具体来说，Voice-on-SPOT[®] 参考设计（如图 10 所示）将 Vesper 麦克风的 Wake-on-Sound 功能与运行 DSP Concepts 高效且灵活的 TalkTo™ 算法的 Ambiq Apollo SoC 结合在一起，为业界提供了领先的超低功耗语音体验。

图 10: Voice-on-SPOT[®] 参考示例



结论

由于功耗要求和外形尺寸限制，大多数便携式产品中使用的音频处理器的功能通常少于家用产品中使用的处理器的功能。

然而，像 Ambiq Micro 的 Apollo 微控制器单元 (MCUs) 和 Vesper 的 ZeroPower Listening™ 麦克风这样的超低功耗解决方案，其功耗可以低于传统音频处理器的 1/10。通过专注于超低功耗和始终在线的语音命令处理，Apollo 产品可以为车辆、遥控器、智能家居设备以及可听戴和可穿戴技术提供服务。

创建可以依靠电池供电运行数小时到数月的始终在线语音命令产品，同时达到与当今流行的智能音箱类似的功能，面临的挑战是巨大的。但得益于本文中描述的产品，这些挑战现在可以被克服。正确选择组件结合仔细调整以适应应用，可以使得便携式语音命令产品为消费者提供满意和可靠的体验。

使用 Ambiq SPOT® 平台的优势

借助 Ambiq 的 SPOT® (亚阈值功耗优化技术) 平台，您将拥有创建未来语音命令系统所需的超低功耗解决方案。更具体地说，您将拥有创建始终在线的、适用于便携式和电池供电产品的本地语音命令和语音助手集成的语音到云的解决方案。

如欲了解更多关于 Ambiq 的 SPOT® 平台如何为数千万商业产品赋能，请查看我们的专家资源库以获得更多内容。



Paul Beckmann, PhD — DSP Concepts 创始人 / 首席技术官

Beckmann 博士在开发音频产品和实施数值密集型算法方面拥有丰富的经验。他在 Bose 公司工作了 9 年，在那里他因对 Videostage 解码算法的贡献而获得《大众科学》颁发的“最佳最新”奖。此外，Paul 受 Bose 博士委托为 Bose 学院开设数字信号处理行业课程，并在信号处理技术方面拥有多项专利。他于 1989 年获得麻省理工学院电气工程学士和硕士学位，并于 1992 年获得麻省理工学院电气工程博士学位。要了解有关 DSP Concepts 强大的软件算法的更多信息，请访问 www.dspconcepts.com。



Aaron Grassian — Ambiq Micro 销售副总裁

Aaron 是一名国际业务发展主管，在为上市和私营科技公司建立、组织和管理全球团队和渠道合作伙伴方面拥有丰富的经验。在他 15 年的职业生涯中，他在通过多种渠道推动超低功耗解决方案、瞄准并赢得初始设计以及通过客户参与和合作伙伴关系快速增长收入方面取得了多次成功。他还在为初创公司建立和管理亚洲销售方面拥有全球客户经验。Aaron 拥有佛罗里达大学电子工程学士学位。要了解有关 Ambiq 业界领先的 SoC 技术的更多信息，请访问 ambiq.com。



Matt Crowley — Vesper 首席执行官

Matt 热衷于建立优秀的团队，将颠覆性技术推向市场。在他的领导下，Vesper 推出其第一款产品的速度几乎是行业平均水平的五倍。在加入 Vesper 之前，他在无晶圆厂微机电系统公司 Sand 9 开创了压电 MEMS 器件的大规模商业化。Matt 还曾为财富 500 强公司提供运营和战略方面的建议。他拥有普林斯顿大学物理学和科学哲学的跨科学学位。马特的日语也很流利。要了解有关 Vesper 节能麦克风的更多信息，请访问 vespermems.com。



Ambiq 文字商标和徽标是 Ambiq Micro, Inc. 的注册商标。蓝牙文字标志和徽标是蓝牙 SIG, Inc. 拥有的注册商标，对此类标志的任何使用均受许可。其他商标和商品名称是其各自所有者的商标和商品名称。

任何文件翻译成英语以外的语言仅为方便非英语阅读的公众，并不具有法律约束力。我们已尽量提供对英文原文的准确翻译，但也可能会存在细微的差异。在大多数翻译成非英文的文档中，均可提供英文原文的参考信息。

© 2023 Ambiq Micro, Inc. All rights reserved.
6500 River Place Boulevard, Building 7, Suite 200, Austin, TX 78730
www.ambiq.com
sales_china@ambiq.com
+1 (512) 879-2850

A-SOCAP0-WPGA01CN v1.0
2023年11月