[Cortana 设备建议](http://www.audio6.com/cortana1/)

[音频杂谈](http://www.audio6.com/audio/basicaudio/)  [小众声学](http://www.audio6.com/author/audiodb/)  3年前 (2016-08-11)  1142浏览  [0评论](http://www.audio6.com/cortana1/#respond)

Cortana 设备建议 – Windows 10 hardware dev

第 0 部分：文档

**Disclaimer:**本文档按“原样”提供。本文档中表示的信息和视图（包括 URL 和其他 Internet 网站引用）如有更改，恕不另行通知。在商业发行之前会发生实质性修改的、与预发布产品相关的一些信息。Microsoft 不对此处提供的信息作任何明示或暗示的担保。你需自行承担使用本文档所带来的风险。

这里提供的一些示例仅供演示，而且是虚构的。与真实世界无实际关联，也不应以此作为推断。

本文档中的所有建议和标准仅作为近似准则，并不保证所有设备在实际条件下的语音识别性能。语音识别是复杂的计算机学习过程，本质上并不完美。即使满足所有高级建议，也不保证 Cortana 和语音识别在设备上的工作完美无缺。

定义

有关相关术语、参考 [(ITU-p.10)](https://www.itu.int/rec/T-REC-P.10-200607-I/en) 的完整列表。

|  |  |
| --- | --- |
| **标题** | **定义** |
| ADC | 模拟数字转换器（简称 A/D） |
| AEC | 回声消除器 |
| AGC | 自动增益控制 |
| ASR | 自动语音识别 |
| BPF | 带通滤波 |
| 捕获（或发送） | 输入（近端）语音信号 |
| dB SPL | 分贝声压级 = 语音平台等式，其中 pref=20 微帕斯卡。0 dBPa 对应 94 dBSPL |
| dBFS（或 dBov） | 满刻度分贝值 = 语音平台等式，其中 Sref= 满刻度数字正弦波的 RMS  相对于其过载电平或最大电平的数字信号的信号电平由 dBov 指定。这通常也称为 dBFS（满刻度）。  例如，仅具有最大正数或最大负数的矩形函数电平为 0 dBov。对于最大刻度数字正弦信号，峰值电平是 0 dBov 并且 RMS 是 -3.01 dBov。(ITU-T G.100.1)。 |
| DI | 指向性指数（有关详细信息，请参阅本文档中的附录 A。） |
| DRC | 动态范围压缩 |
| DSP | 数字信号处理 |
| DUT（或 DRP） | 测试中的设备（或设备参考点） |
| FIR | 有限脉冲响应 |
| HATS | 人头和躯干模拟器 – 在 ITU-T 建议 P.58 [15] 中介绍 |
| IHV | 独立硬件供应商 |
| IIR | 无限脉冲响应 |
| MRP | 嘴参考点是仿真嘴的嘴唇平面前 25 毫米的点。这是在测量正常讲话、大声讲话和安静讲话之前校准语音级别的点 |
| NG | 环境噪声增益（有关详细信息，请参阅本文档中的附录 A。） |
| NGA | A 加权环境噪声增益（有关详细信息，请参阅本文档中的附录 A。） |
| 正规带 | 标称通频带为 300-3400 Hz 的语音信号 (ITU-p.10 N-3) |
| 呈现（或接收） | 输出（远端）语音和/或其他呈现的音频信号 |
| TCLw | 终端耦合损耗（加权） |
| THD+N | 总谐波失真（包括噪声）= ((signal-f₀))⁄f₀¹ |
| 宽频带 | 标称通频带为 100-7000 Hz 的语音信号 (ITU-p.10 W-3) |
| ƒ₀ | 基本频率 |

¹SpeechPlatformEquation1其中“—”通过陷波滤波完成

第 1 部分：摘要和大纲

语音平台用于增强 Windows 10 中的所有语音体验，例如 Cortana 和听写。 本文档提供旨在用于 Microsoft 语音平台的音频输入设备设计和开发的指南及建议。

本文档中提供的指南仅限于直接影响以下项的设计参数：1) 语音识别准确性 2) 基础语音处理算法的行为。本文档不是 Windows 认证计划的一部分，也不是设备认证所需的建议。建议仅用作设备设计的帮助指南和最佳做法。

Windows 10 中的语音识别

鼓励设备制造商将语音增强处理集成到其设备中并进行调整，以便相对于[语音识别测试条件](https://msdn.microsoft.com/zh-cn/library/windows/hardware/dn957009(v=vs.85).aspx" \o ")来优化性能。

对于未集成语音增强处理的设备，Microsoft 在 Windows 10 中提供了默认处理。Microsoft 中的语音增强处理不需要 IHV 进行特定于设备的调整。

Microsoft 语音增强管道包括以下功能 – 针对高质量语音识别进行调整：

* 宽带处理 (16 kHz)
* 回声抵消
* 麦克风阵列支持和波束形成
* 平稳噪声抑制

如果音频驱动程序未公开麦克风几何结构和语音的音频信号处理，将使用 Microsoft 语音增强管道。为了利用第三方增强功能，必须提供麦克风几何结构、支持音频输入上的“语音”信号处理模式，并确保音频驱动程序提供的效果或其 APO 至少包含噪声抑制和回声抵消。

设备建议

本文档中的建议分为三个部分进行讨论：

1. Device.SpeechRecognition（有关详细信息，请参阅本文档中的第 2 部分。）
2. Device.Audio（有关详细信息，请参阅本文档中的第 3 部分。）
3. Device.Audio.Acoustics（有关详细信息，请参阅本文档中的第 4 部分。）

Device.SpeechRecognition

本部分定义语音识别性能要求，以确保高质量语音体验。应根据这些性能要求测试所有设备。

Device.Audio

本部分提供指南，以便根据软件接口、通信协议和数据格式在主机操作系统上获得最佳的运行效果。应根据这些指南测试所有设备。

Device.Audio.Acoustics

本部分提供有关设备设计的声学及相关属性的建议和最佳实践。本部分最常用于使用 Microsoft 语音增强处理的设备。

合适的组件选择、声学设计和机械集成对于捕获尽可能最好的音频从而提供良好体验非常重要，而不是依赖于语音增强管道。建议未在设备上执行语音增强处理的设备（即，使用操作系统中提供的默认语音增强管道）遵循 Device.Audio.Acoustics 下的高级建议，以便最大程度地满足 Device.SpeechRecognition 体验要求，并与操作系统语音增强管道一起正常工作。

性能级别

有时，本文档中的功能指南将拆分为两个级别 – 标准和高级。进行这样的区分是为了明确存在一系列面向语音功能、跨目标价格和操作区域的设备。Microsoft 建议所有设备都以高级指南（如果指定）为目标。

设备必须满足所有类别中的高级建议才能视为高级设备。高级和标准类别不属于 Windows 认证计划，而属于已了解的最佳做法和指南。

高级

设备满足为出色使用语音识别建议的功能指南。设备在更具挑战性的条件下（如增加的噪声级别或更远的距离）也将正常工作。由于用户可以从大于 1 米的距离与设备交互，这将支持使用语音进行自然交互。用户将能够在具有挑战性的噪声环境中（环境噪声或由于设备播放）使用设备，并且这将支持自然交互，例如“你好小娜，下一首歌”。

标准

设备满足使用语音识别的基本功能指南。设备将在环境条件下正常工作。



语音识别和电话服务之间的区别

本文档仅包含使用语音平台的建议，但是，许多面向语音识别功能的设备也同时面向电话服务使用。相似点非常明显 – 两种情况都使用带有麦克风的设备接收人工语音、使用音频处理管道从环境中去除噪声并增强人工语音，并使用依赖于清晰语音信号的应用程序来理解语音。

区别在于由哪一方使用语音信号。电话服务的使用者为人类，对他们来说，可感知的语音质量和语言清晰度极其重要。语音识别的使用者为算法，其中由对语音信号的特定特征进行训练的机器学习来确定识别的内容，并且这些特征并不一定以感知规范为标准。

可感知的语音质量通常与语音识别的准确性关联起来，但并非总是如此。本文档侧重于评估和最大程度提高语音识别准确度的方法。建议支持“语音”信号处理模式并专门为语音识别调整该模式。

通过 [Skype/Lync 音频认证](https://technet.microsoft.com/office/dn788953)是良好的设备音频性能的强有力证明。

第 2 部分：Device.SpeechRecognition 建议

本部分提供有关设备在各种环境中的目标语音识别准确性的建议。所有建议都表示每种方案中可接受的用户体验的最低要求。测试条件和步骤在[测试设置](https://msdn.microsoft.com/library/windows/hardware/dn957009(v=vs.85).aspx" \o ")中指定。

设备应务必满足本部分中的建议，以便：

* 确保设备可以在各种环境中正常使用语音识别
* 确保设备将在 Microsoft 语音增强管道的可操作边界内正常工作（针对未使用其自己的语音增强管道的设备）

目标是语音的准确性；语音级别是 89 dB SPL @ MRP。

三个测试必须全部达到 **Standard** 才能归类为 **Standard**。

Device.SpeechRecognition.Quiet

安静测试表示环境噪声最小（本底噪声 < 35 dBA SPL）的理想环境。

|  |  |  |
| --- | --- | --- |
| **安静** | **高级** | **标准** |
| <= 35 dBA SPL | 95% | 90% |

Device.SpeechRecognition.AmbientNoise

环境噪声测试表示各种级别和类型的噪声环境，例如[咖啡馆和酒吧](http://go.microsoft.com/fwlink/p/?linkid=528451" \o ")。

|  |  |  |
| --- | --- | --- |
| **环境噪声 @ DUT** | **高级** | **标准** |
| >= 57 dBA SPL | 90% | 85% |

Device.SpeechRecognition.EchoNoise

回声噪声测试表示各种级别和类型的呈现播放场景（例如媒体播放）。

级别校准在侦听器位置 (LRP)。

|  |  |  |
| --- | --- | --- |
| **回声噪声 @ LRP** | **高级** | **标准** |
| >= 70 dBA SPL | 90% | 85% |

Device.SpeechRecognition.DefaultMicGain

为了确保最佳的用户体验并最大程度地保证 Windows 语音的准确性，麦克风增益设置应设为用于通过上述测试的最佳级别。然后，生成的增益设置可以在注册表项中进行如下所示：

|  |  |
| --- | --- |
| 名称 | **HKLM\SOFTWARE\Microsoft\Speech\_OneCore\AudioInput\MicWiz\DefaultDefaultMicGain** |
| 类型 | REG\_DWORD |
| 数据 | 0-10000 十进制，0-0x2710 十六进制 |
| 说明 | 最大增益百分比乘以 100 |

对于符合或超过所有三个测试标准指标且具有集成麦克风阵列的设备，由 IHV 设置此项。

**示例**

此示例将 MicGain 设置为 42.00%，即 0x1068：

[HKEY\_LOCAL\_MACHINE\SOFTWARE\Microsoft\Speech\_OneCore\AudioInput\MicWiz]

“DefaultDefaultMicGain”=dword:00001068

此示例将 MicGain 设置为 64.50% 即 0x1932：

[HKEY\_LOCAL\_MACHINE\SOFTWARE\Microsoft\Speech\_OneCore\AudioInput\MicWiz]

“DefaultDefaultMicGain”=dword:00001932

**条件**

* 该注册表项仅适用于集成的麦克风阵列，不应在使用外部麦克风时实现。
* 不应在不符合或未超过所有三个测试标准指标的设备中实现该注册表项。

第 3 部分：Device.Audio HLK 建议

本部分提供有关设备的软件和硬件接口、通信协议以及数据格式的建议。

设备应务必满足本部分中的建议，以便：

* 确保设备将在 Windows 音频管道框架中正常工作
* 确保设备将在 Microsoft 语音增强管道的可操作边界内正常工作（针对未使用其自己的语音增强管道的设备）

要使用语音识别功能的设备必须满足 [Windows 硬件认证要求](https://msdn.microsoft.com/library/windows/hardware/jj134354.aspx)下提供的所有 Device.Audio 要求。

语音识别功能最相关的 Device.Audio 要求将在以下部分中进行说明。

Device.Audio.Base.AudioProcessing

驱动程序必须通过 FXStreamCLSID、FXModeCLSID 和 FXEndpointCLSID APO（或代理 APO）公开所有音频效果。 APO 必须在系统查询时向其发送已启用效果的准确列表。 驱动程序必须支持 APO 更改通知，并且仅在 APO 更改发生时才通知系统。

不应存在不可发现或不可控制的硬件、固件或基于第三方软件的 BPF、AGC、DRC、AEC、波束形成、噪声抑制或任何其他从设备/向设备改变音频样本（例如非线性处理）的技术。

驱动程序配置验证工具可以用于在设备上进行验证。

Device.Audio.Base.StreamingFormats

语音识别在 StreamingFormats HLK 中定义的所有音频捕获和呈现流格式下工作，最佳状态为 16 kHz 并使用 24 位捕获和单声道呈现。 请注意，当单声道呈现对于回声抵消性能而言为最佳选择时，很明显它并不适用于媒体应用程序，因此此建议仅供语音输入是主要或独有功能的设备参考。 回声消除器将使用所有呈现格式，并且设备通常将具有立体声或更出色的呈现功能。

Device.Audio.Base.SamplePositionAccuracy

呈现和捕获音频信号应务必满足以下条件：1) 采样准确 2) 时间戳准确。

ADC 和 DAC 上的准确采样确保高保真的音频信号，这将支持良好的语音质量（样本误差将噪声添加到信号）和准确的回声抵消（样本误差将噪声添加到回声估计并创建信号误配准）。

准确的时间戳确保捕获和呈现信号可以针对回声抵消进行同步。

所有捕获通道都必须同步（对齐 t0 样本）。这包括所有在完全相同的时间点初始化的 ADC，以及任何要对齐且无数据的缓冲区。这对波束形成和声源定位器性能非常重要，后者要依赖于子样本对齐才能估计声源的位置。

漂移在[附录 A：计算](https://msdn.microsoft.com/library/windows/hardware/dn957009(v=vs.85).aspx" \l "appendix_a__calculations" \o ")中定义为与地面实况时钟有关的绝对时钟漂移。但是，回声消除器所看到的漂移实际是捕获和呈现路径之间的相对漂移。

等式

使用内置捕获和呈现路径的设备可以通过从同一个主时钟源派生捕获和呈现时钟来否定此相对漂移。这样，即使主时钟存在绝对漂移，但捕获和呈现时钟之间的相对漂移是 0%。

Device.Audio.USB.USB

所有 USB 音频输入设备都必须根据 [USB 设备类规范](http://www.usb.org/developers/docs/devclass_docs/)正确设置描述符。

|  |  |  |  |
| --- | --- | --- | --- |
| **终端类型** | **代码** | **I/O** | **说明** |
| 未定义的输入 | 0x0200 | I | 输入终端，未定义的类型。 |
| 麦克风 | 0x0201 | I | 不适合任何其他分类的通用麦克风。 |
| 桌面麦克风 | 0x0202 | I | 通常放置在桌面或集成到显示器的麦克风。 |
| 个人麦克风 | 0x0203 | I | 头戴式或领夹式麦克风。 |
| 全向麦克风 | 0x0204 | I | 专用于在相对较长范围内接收多个演讲者语音的麦克风。 |
| 麦克风阵列 | 0x0205 | I | 专用于使用基于主机的信号处理算法进行定向处理的一组麦克风。 |
| 处理麦克风阵列 | 0x0206 | I | 具有嵌入式信号处理器的一组麦克风。 |

驱动程序指南

WDM 音频驱动程序开发路线图：[http://www.microsoft.com/whdc/device/audio/wavertport.mspx](http://go.microsoft.com/fwlink/p/?linkid=528454" \o ")。

第 4 部分：Device.AudioAcoustics 建议

本部分提供有关设备的声学和相关属性（例如麦克风和扬声器放置、麦克风响应、从设备接收的噪声等等）的建议。麦克风选择、放置、集成和阵列设计是一些非常重要的因素，以支持高品质语音识别性能。

建议和测试相对于在语音增强处理之前但在麦克风均衡及麦克风增益固定之后的信号。

设备应务必满足本部分中的建议，以便：

* 确保设备将在 Microsoft 语音增强管道的可操作边界内正常工作（针对未使用其自己的语音增强管道的设备）。
* 确保设备将在 Microsoft 语音识别应用程序（例如 Cortana）的可操作边界内正常工作。
* 确保设备具有与其他用于使用语音识别应用程序的语音输入设备类似的特征。

Device.Audio.Acoustics.MicArray

麦克风阵列在支持语音增强管道去除以下形式的本地化噪声方面起着重要作用：

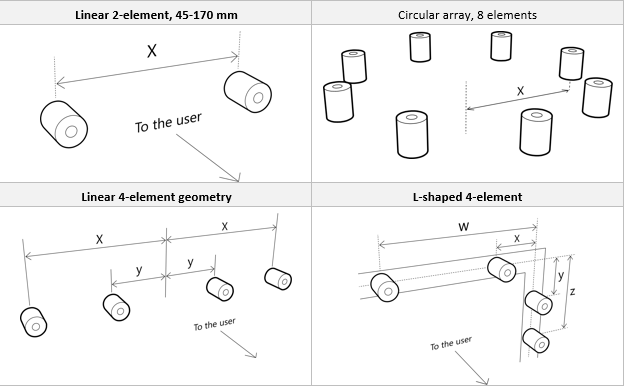
* 环境噪声
* 接收（回声）噪音
* 室内混响
* 设备生成的自噪声（如风扇）

请注意，良好的麦克风阵列设计将涉及许多参数而不仅是麦克风的数量，并且高度依赖于设备集成和使用情况。有关设计注意事项和实现指南（以及其他许多内容非常详尽的最佳做法），请参阅 [Windows 中的麦克风阵列支持](https://msdn.microsoft.com/library/windows/hardware/dn613960.aspx)。

音频驱动程序必须实现 [KSPROPERTY\_AUDIO\_MIC\_ARRAY\_GEOMETRY](https://go.microsoft.com/fwlink/p/?linkid=528456) 属性。然后可以通过 Windows.Devices.Enumeration API 访问 System.Devices.MicrophoneArray.Geometry 属性。USB 音频驱动程序将为在 USB 描述符中设置相应字段的 USB 麦克风阵列支持此属性。

驱动程序配置验证工具可以用于在设备上进行验证。

对于将使用 Microsoft 语音增强管道的设备，建议使用已知可正常工作的以下阵列几何结构。但是，Microsoft 语音增强功能可以使用麦克风阵列描述符中指定的任何阵列类型。

具有以下理论性能属性：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **麦克风阵列** | **麦克风** | **类型** | **NG (dB)** | **NGA (dB)** | **DI (dB)** |
| 线性（小型） | 2 | 单向 | -12.7 | -6.0 | 7.4 |
| 线性（大型） | 2 | 单向 | -12.9 | -6.7 | 7.1 |
| 线性 (4 el A) | 4 | 单向 | -13.1 | -7.6 | 10.1 |
| 4 L 形（可旋转） | 4 | 单向 | -12.9 | -7.0 | 10.2 |
| 5 (4 el B) | 4 | 全向 | -12.9 | -7.3 | 9.9 |

麦克风阵列描述符

设备必须使用麦克风阵列描述符描述其麦克风类型和几何结构。麦克风阵列描述符用于参数化 Microsoft 和第三方的语音增强管道中的波束形成器和声源定位器。

可以通过麦克风阵列描述符定义并支持与上述子部分中默认描述符不同的自定义描述符。

有关麦克风阵列描述符的详细信息如下：

* [麦克风阵列几何结构属性](http://go.microsoft.com/fwlink/p/?linkid=528457)
* [麦克风阵列几何结构描述符格式](https://go.microsoft.com/fwlink/p/?linkid=528458)
* [KSPROPERTY\_AUDIO\_MIC\_ARRAY\_GEOMETRY](https://go.microsoft.com/fwlink/p/?linkid=528459)
* [KSAUDIO\_MIC\_ARRAY\_GEOMETRY 结构](https://go.microsoft.com/fwlink/p/?linkid=528460)
* [KSAUDIO\_MICROPHONE\_COORDINATES 结构](https://go.microsoft.com/fwlink/p/?linkid=528461)

Device.Audio.Acoustics.MicSensitivity

麦克风敏感度建议确保可以支持足够电平的语音输入，而无需使数字音频输入路径饱和（即，削波），并确保输入语音信号位于足够高的电平以进行识别。

|  |  |  |  |
| --- | --- | --- | --- |
| **麦克风 + 数字接口** | **建议** | **测试信号** | **DUT 位置** |
| 最高电平 | ≤ -20 dBFS RMS  100-8000Hz | 连续对数扫频 3 秒  94 dB SPL @ MRP（“响亮”） | 最接近 MRP 到 DUT 支持的方案（由 IHV 定义） |
| 最低电平 | ≥ -55 dBFS RMS  100-8000Hz | 连续对数扫频 3 秒  69 dB SPL @ MRP（“安静”） | 最远的 MRP 到 DUT 支持的方案 |

“最高”建议设置为能够支持视为“响亮”的语音输入电平（70 dB SPL RMS @ 1m，波峰因数为 12），其中已针对在 94 dB SPL @ MRP 回放的对数扫频测试信号测量 -20 dBFS 的阈值以便与目标电平的真实语音相关联。请注意：选择此阈值和输入信号以简化典型音频生成器测试/分析测试设置。

“最低”建议设置为能够支持视为“安静”的语音输入电平（55 dB SPL RMS @ 1m，波峰因数为 12），其中已针对在 69 dB SPL @ MRP 回放的对数扫频测试信号测量 -55 dBFS 的阈值以便与目标电平的真实语音相关联。

如果上述方案中接收的电平通过简单的线性增益调整后未适应于最大和最小阈值之间，应使用设备特定的 AGC 调整语音电平以适应阈值。

阵列中所有麦克风上的麦克风敏感度匹配非常重要，对于 **Standard** 应在 +/- 3 dB 之内，对于 **Premium** 应在 +/- 1 dB 之内。系统级别性能在 Device.Audio.Acoustics.MicMagnitudeResponse 中指定并测试。MEMS 麦克风具有较低的制造公差，建议用于最佳的麦克风到麦克风匹配特性。

请注意，如果不使用 MEM 麦克风，则敏感度将在装配期间发生变化，也将由于环境条件而在使用期间发生变化。

Device.Audio.Acoustics.MicIntegration

谨慎进行麦克风的机械集成非常重要，这是一个常见的失败根源。

必须集成麦克风，以便确保麦克风和设备机壳之间以及沿麦克风端口声管（如果适用）的良好声密封。最大程度地减少系统和麦克风之间的噪声和振动也非常重要。两种典型的解决方案是使用橡胶套或衬垫。

无论选择哪种方法，都要检查声密封是否足以适应所有生产公差以及环境和生命周期的改变。

Device.Audio.Acoustics.MicPlacement

一般情况下，遵循以下准则来放置麦克风阵列：

* 尽可能远离噪声源，例如扬声器、风扇、键盘、硬盘驱动器和用户的手。
* 尽可能接近演讲者的嘴部。

Device.Audio.Acoustics.MicSelfNoise

麦克风组件也是噪声源，称为自噪声或等效噪声。它可能由各种源产生，但它在麦克风内部创建，并且实际独立于 SPL。等效噪声表示麦克风动态范围的最低点，并且对于获得良好的波束形成性能而言尤为重要。使用高品质麦克风将最大程度地减小麦克风内部噪音。

对于 **Standard**，建议使用标称 SNR 至少为 61 dB 的麦克风，对于 **Premium** 则为 63 dB。

Device.Audio.Acoustics.MicReceivedNoise

接收噪声的两个主要来源是声噪声和电气噪声。声噪声可能由外部设备产生，或在设备内部（由于风扇，硬盘等）产生。声噪声还可以通过设备机械装置传递。

可以使用数字麦克风而不是模拟麦克风来最大程度地减小电气噪声。

Device.Audio.Acoustics.MicMagnitudeResponse

**Premium** 和 **Standard** 掩码适用于 Device.Audio.Acoustics.Bandwidth 下的所有设备层。例如，设备可以具有 **Standard** 带宽（窄频带）和该频带内的 **Premium** 幅值响应。

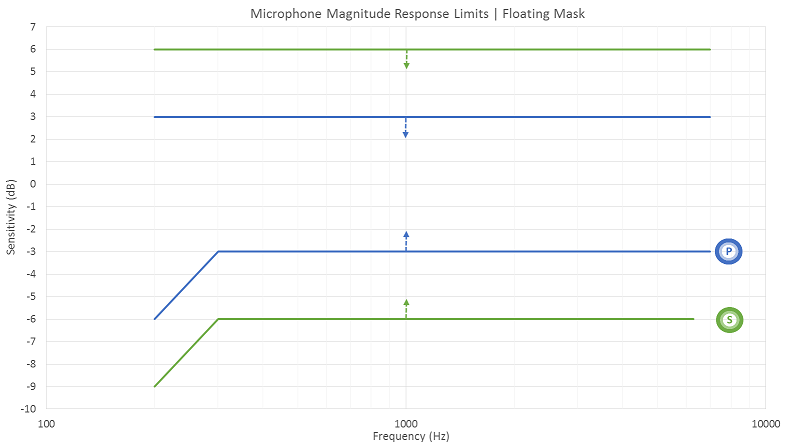
出于两个原因，麦克风幅值响应容差掩码非常重要：

* 减少由于通道响应而在阵列中的麦克风之间产生的差异
* 减少由于通道响应而在设备之间产生的差异

这会带来以下结果：

* 用于训练语音声学模型的数据变化更少
* 语音增强管道中的波束形成性能更一致
* 低频率的能量更少，SNR 区域更低 -> 由于低频噪声，信号中的噪声更少并且饱和度的可能性更小
* 可感知应用程序的语音听起来更自然

出于对这些因素的考虑，Microsoft 鼓励满足下方响应掩码（“平坦”表示理想目标），从而针对随着时间发展的声学模型，使响应在设备之间融合。



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **高级** | **高级** | **标准** | **标准** |
| **频率 (Hz)** | **上限 (dB)** | **下限 (dB)** | **上限 (dB)** | **下限 (dB)** |
| 200 | 3 | -6 | 6 | -9 |
| 300 | 3 | -3 | 6 | -6 |
| 5000 | 3 | -3 | 6 | -6 |
| 6300 | 3 | -3 | 6 | -6 |
| 7000 | 3 | -3 | 6 |  |

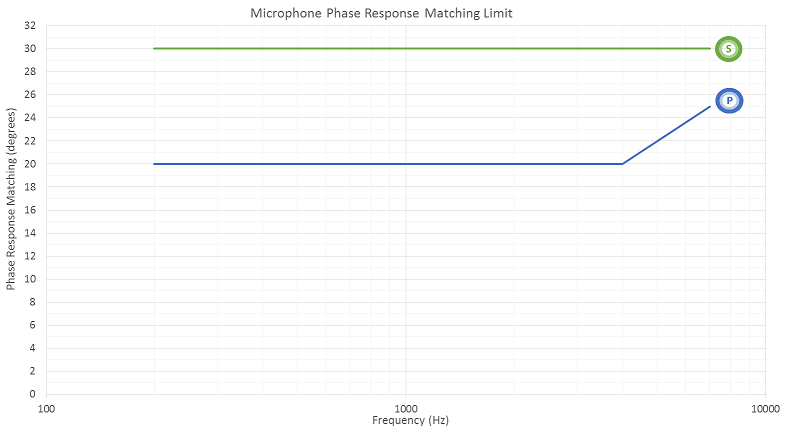
所有麦克风都必须同时满足掩码。掩码是浮动的，即绝对增益并不是其中一个因素。

Device.Audio.Acoustics.MicPhaseResponseMatching

麦克风相位响应匹配限制非常重要，以确保通过阵列中麦克风元素所接收信号之间的临时关系与与阵列中麦克风元素的物理几何结构保持一致。

这对于 Microsoft 语音增强管道中的波束形成器和声源定位器处理非常重要，这将使用每个麦克风通道的语音信号之间的相对延迟完成以下操作：

* 估计定位声音的方向
* 抑制来自非侦听方向的噪声。



|  |  |  |
| --- | --- | --- |
|  | **高级** | **标准** |
| **频率 (Hz)** | **上限 (deg)** | **上限 (deg)** |
| 200 | 20 | 30 |
| 1000 | 20 | 30 |
| 4000 | 20 | 30 |
| 7000 | 25 | 30 |

Device.Audio.Acoustics.MicDistortion

要求扩展到有效带宽的 ½，这时第一个谐波将超过奈奎斯特速率。

出于两个原因，满足麦克风失真和噪声限值非常重要：

* 确保语音在进入语音识别器之前相对未失真
* 保持回音路径上的非线性关系最小以实现良好的回声抵消性能

建议使用 SDNR（脉冲噪声信号对失真和噪声比）测量失真，尽管也已提供 THD 目标。有关 SDNR 测试方法的详细信息，请参阅 IEEE 269-2010 Annex L。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **高级** | **高级** | **标准** | **标准** |
| **频率** | **THD** | **SDNR** | **THD** | **SDNR** |
| 250 | 2.50% | >= 32 | 3.20% | >= 30 |
| 1000 | 2.50% | >= 32 | 3.20% | >= 30 |
| 4000 | 2.50% | >= 32 | 3.20% | >= 30 |
| 5000 | 4.00% | >= 28 | 4.00% | >= 28 |
| 6000 | 6.30% | >= 24 | 6.30% | >= 24 |

Device.Audio.Acoustics.MicBandwidth

捕获信号的采样率是决定语音信号有效带宽的主要因素。由于语音平台在语音识别器中使用 16 kHz 声学模型，建议使用 16 kHz 最小采样率。300 Hz 是语音识别器的有效下端，但针对同时也面向语音通信的设备，建议使用 200 Hz 声学限制。

滤波还可以改变设备的有效带宽，例如 ADC 中的模拟 FIR 低通滤波器、管道较后阶段中的数字带通滤波器，或者甚至是由于麦克风元素或电子系统的响应产生的衰减。在设计过程中应考虑这些因素。

语音平台使用 8 kHz 声学模型，仅为提供对传统蓝牙音频设备的支持。

Device.Audio.Acoustics.RenderDistortion

请注意：Device.Audio.Acoustics.RenderDistortion 仅适用于具有内置扬声器的设备。

出于以下原因，满足扬声器失真限制非常重要：

* 保持回音路径上的非线性关系最小以实现良好的回声抵消性能

建议使用 SDNR（脉冲噪声信号对失真和噪声比）测量失真，尽管也已提供 THD 目标。有关 SDNR 测试方法的详细信息，请参阅 IEEE 269-2010 Annex L。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **高级** | **高级** | **高级** | **高级** | **标准** | **标准** | **标准** | **标准** |
|  | **电平：-22dBFS** | **电平：-22dBFS** | **电平：-16dBFS** | **电平：-16dBFS** | **电平：-22dBFS** | **电平：-22dBFS** | **电平：-16dBFS** | **电平：-16dBFS** |
| **频率** | **THD** | **SDNR** | **THD** | **SDNR** | **THD** | **SDNR** | **THD** | **SDNR** |
| 300 | 6.3% | >= 24 | 6.3% | >= 24 | NA | NA | NA | NA |
| 500 | 6.3% | >= 24 | 6.3% | >= 24 | NA | NA | NA | NA |
| 600 | 5% | >= 26 | 5% | >= 26 | 10% | >= 20 | 10% | >= 20 |
| 800 | 5% | >= 26 | 5% | >= 26 | 8% | >= 22 | 8% | >= 22 |
| 1000 | 4% | >= 26 | 5% | >= 26 | 6.3% | >= 24 | 6.3% | >= 24 |
| 1500 | 4% | >= 26 | 5% | >= 26 | 5% | >= 26 | 6.3% | >= 24 |
| 3000 | 4% | >= 26 | 5% | >= 26 | 5% | >= 26 | 6.3% | >= 24 |
| 4000 | 5% | >= 26 | 5% | >= 26 | 5% | >= 26 | 6.3% | >= 24 |
| 5000 | 5% | >= 26 | 5% | >= 26 | 6.3% | >= 24 | 6.3% | >= 24 |
| 6000 | 5% | >= 26 | 5% | >= 26 | 6.3% | >= 24 | 6.3% | >= 24 |

Device.Audio.Acoustics.RenderPlacement

若要使声学回声消除器正常工作，设备扬声器应放置在距离麦克风最远的位置，或直接放置朝空的扬声器。

参考

原文：https://msdn.microsoft.com/zh-cn/library/windows/hardware/dn957008(v=vs.85).aspx

|  |  |
| --- | --- |
| **标题** | **链接** |
| 有用的定义和指标 | [http://www.rane.com/note145.html](http://go.microsoft.com/fwlink/p/?linkid=528462) |
| Windows 中的麦克风阵列支持 | [http://msdn.microsoft.com/library/windows/hardware/dn613960.aspx](http://go.microsoft.com/fwlink/p/?linkid=528463) |
| ITU-p.10 参考术语 | [https://www.itu.int/rec/T-REC-P.10-200607-I/en](https://go.microsoft.com/fwlink/p/?linkid=528464) |
| 3GPP 终端声学特征 | [http://www.3gpp.org/DynaReport/26131.htm](http://go.microsoft.com/fwlink/p/?linkid=528465) |
| ETSI UMTS 语音电话服务终端声学测试规范  （3GPP TS 26.132 版本 11.4.0，第 11 版） | [http://www.etsi.org/deliver/etsi\_ts/126100\_126199/126132/11.04.00\_60/ts\_126132v110400p.pdf](http://go.microsoft.com/fwlink/p/?linkid=528466) |
| ETSI EG 202 396-1 | [http://www.etsi.org/deliver/etsi\_eg/202300\_202399/20239601/01.02.02\_60/eg\_20239601v010202p.pdf](http://go.microsoft.com/fwlink/p/?linkid=528467) |

转载请注明：[小众声学](http://www.audio6.com/) » [Cortana 设备建议](http://www.audio6.com/cortana1/)