Steve F. Temme[1]

(1. Listen, Inc., Boston, MA 02118 USA)

*

摘要：语音控制的智能设备（例如智能扬声器、听觉设备和车辆信息娱乐系统）非常难以测试。它们具有从有线到无线的多样连接方式，并且在接收端和重放端使用了诸多信号处理技术。这意味着它们的特性会随着使用环境的“现实世界”条件（例如背景噪声、播放级别和室内声学条件）的不同而变化。 此外，它们的多功能特性意味着可能需要测试该设备的许多方面，包括语音识别、音乐播放、作为免提电话或听觉设备或助听器使用时的性能。由于其复杂的非线性使用情况，这些设备通常需要在不同级别和不同环境条件下进行测试。本文重点介绍在各种实际条件下准确测量此类设备的音频性能的工具和技术。

关键词：听觉设备，汽车信息娱乐，智能扬声器，测试

The challenges of testing voice-controlled audio systems

Steve F. Temme[1]

(1. Listen, Inc., Boston, MA 02118 USA)

Abstract: Smart devices that are voice-controlled such as smart speakers, hearables, and vehicle infotainment systems are notoriously complex to test. They have numerous connections from wired to wireless and contain much signal processing, both on the record and the playback side. This means that their characteristics change according to ‘real world’ conditions of the environment that they are used in, such as background noise, playback levels, and room acoustics. Furthermore, their multifunctional nature means that there are many aspects of the device that may need to be tested, ranging from voice recognition to music playback, operation as a hands-free telephone, and in the case of hearables, hearing assistance. Due to their complex non-linear use cases, these devices often need to be tested at different levels and different environmental conditions. This paper focuses on tools and techniques to accurately measure the audio performance of such devices under the many various real-world conditions in which they are used.

Keywords: hearables, automotive infotainment, smart speakers, smartphones, test

---

* 译者：陈立，南京大学声学研究所。

0 □ □

智能扬声器、听觉设备和汽车信息娱乐系统等智能设备的测试正面临着日益严峻的挑战。它们具备诸多可能的交互方式，包括硬连线、无线连接（蓝牙、云端连接）、智能手机、语音控制（“Hey Siri”、“OK Google”、“Alexa”）等，对于汽车，还包括 USB 储存卡与 CarPlay/Android Auto 车载交互系统的连接方式。这些设备常常应用了诸多信号处理技术，例如接收端的波束形成、背景噪声过滤、语音活动度检测与重放端的响度控制、压缩、均衡和主动降噪等。这意味着它们的特性会随着“现实世界”条件（例如物理环境和背景噪声）的不同而变化。此外，它们的多功能特性意味着可能需要测试该设备的许多方面，从语音识别到音乐播放，甚至需要考虑其用作免提电话、通话耳机或助听器时的工作情况。这些设备通常需要在不同级别和环境条件（例如不同的物理设置以及有/无背景噪声、不同信号等）下进行测试。

尽管目前尚无用于测试大多数智能设备的标准，但我们可以借鉴许多其他音频设备的测试情况，参考例如用于扬声器和听觉设备测试的 IEC 和 BS EN 标准[1,2]、用于耳机测试的 IEEE 标准、用于电话测试的 IEEE/TIA/ITU 标准[3,4,5,6]、用于助听器的 ANSI 和 IEC 标准[7,8]和用于背景噪声的 ETSI 标准[9]。因此，为了全面表征设备的各项特性，测试系统的灵活性和测试各种声学设备的经验至关重要。

本文介绍了如何实施基本声学测试和更复杂的实际测试，以及可能使用到的技术和标准。本文讨论的大多数测试都与所有智能设备（包括智能扬声器、听觉设备和汽车信息娱乐系统）相关，还详细介绍了一些针对听觉的附加测试。最后，我们列出了测试系统功能的检查清单，当你需要全面表征设备的特性时，您可以查阅此清单来选择系统进行测试。

1 □ □ □ □ □ □

1.1 □ □ □ □ □ □ □ □

基本的声学测试会在静态条件下进行以获取一个或多个传声器以及扬声器或听觉设备的特性，例如频率和脉冲响应、失真、以及立体声听觉设备的左右跟踪性能。衡量其有线和无线性能以进行比较是一种有效方式，这需要用于测量有线响应的经过校准的输入设备（例如听觉设备放大器或辅助输入设备）和用于无线测量的经过校准的蓝牙接口（图 1）。

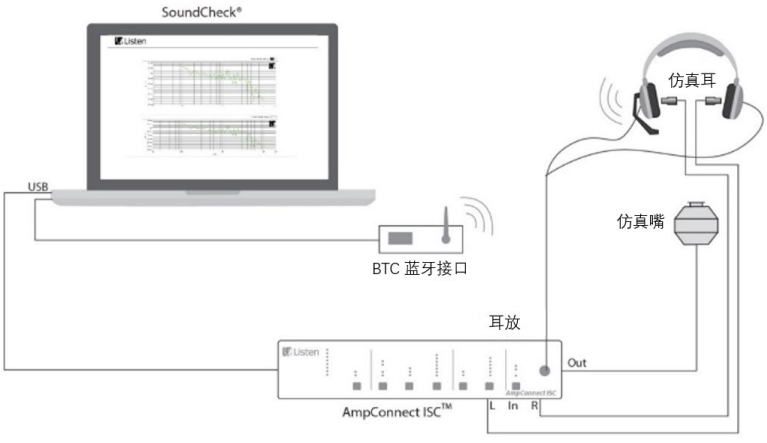

图 1. 基本有线和无线（蓝牙）耳机测量的测试配置。对于汽车信息娱乐系统，请替换为 6 传声器阵

列并通过主机使用汽车的扬声器。

Fig 1. Test configuration for basic wired and wireless （BT） headset measurement. For automotive infotainment systems, substitute a 6-mic array and use the car's speakers via the head unit.

此外，输出端和输入端的声学测量都需要使用经过校准的仿真口或声源扬声器以及经过校准的测量传声器或仿真耳，前者相当于一个频响恒定的声源，用以获得恒定的声压输出电平频响。尽管通常只需一个测量传声器就可以完成智能扬声器测试，但是头和躯干模拟器才是可听和信息娱乐系统测量的理想选择，因为它具有经过校准的仿真口和仿真耳，可以真正代表典型的听众。但是，对于听觉设备，耳耦合器通常用作低成本的选择；而对于汽车测试，设备通常会替换为 6 个测量传声器组成的阵列。

1.2 □□□□

尽管在所有情况下，测量本身都类似于标准的扬声器和传声器测试，但输入激励信号、记录响应信号仍然是一种挑战。通常，进行此操作的唯一方法是像实际用户一样通过语音经由云端进行控制。

头和躯干模拟器中通常包含一个仿真口以播放语音命令激活待测设备的传声器，该传声器通常经由蓝牙连接到智能手机和/或主机。在测试智能扬声器的情况时，可以使用仿真口或小型扬声器。并且由于设备本身可以联网，因此不需要通过智能手机进行网络连接。智能手机与云端专有语音服务（例如 Alexa 语音服务，AVS）进行通信，上传测试信号，通过蓝牙经由智能手机传输或者播放至待测设备。测试系统使用触发记录来拾取扬声器的重放信号并对其进行分析。

尽管各个设备的实际物理测试设置非常相似，但是对于每种设备，都必须了解如何无线地传送信号。测试信号必须传至云端以确保能够重放。每个厂商从云端重放信号的方式都有所不同，有些要求上传本地录音，有些则要求在 Spotify 这种媒体流平台上进行操作。对于传声器测试，某些系统（如 Alexa）允许访问自己制作的录音，其他系统则出于安全和隐私原因禁止这类操作，这使得传声器测试具有挑战性。每个设备都需要使用不同的唤醒词进行激活，它们具有不同的延迟补偿，并且在听到唤醒字之后记录的时间也不同，因此我们需要理清每个待测扬声器的具体情况（主要是通过反复试验）。我们可以参考针对于该设备的测试标准来进行物理设置。例如，对于耳机，我们可以参考 IEEE 269[4]推荐的物理设置，将其置于头和躯干模拟器上进行测试。对于智能扬声器，由于其实际使用情况类似于扬声器，因此 IEEE 1329-2010 [3]推荐的物理几何结构是一个不错的起点，它涉及在消声室内的桌上进行测试的情况。在所有 3 种情况下，仿真口（可能包含在头和躯干模拟器中）都会播放语音命令以激活待测设备传声器，该传声器通常通过蓝牙连接到智能手机（如果使用智能扬声器，则直接连接到云端）。智能手机与云端专有语音服务（例如 Alexa 语音服务，AVS）进行通信，上传测试信号，通过蓝牙经由智能手机传输或者播放至待测设备。测试系统使用触发记录来拾取扬声器的重放信号并对其进行分析。示例测试配置如下图 2 所示。

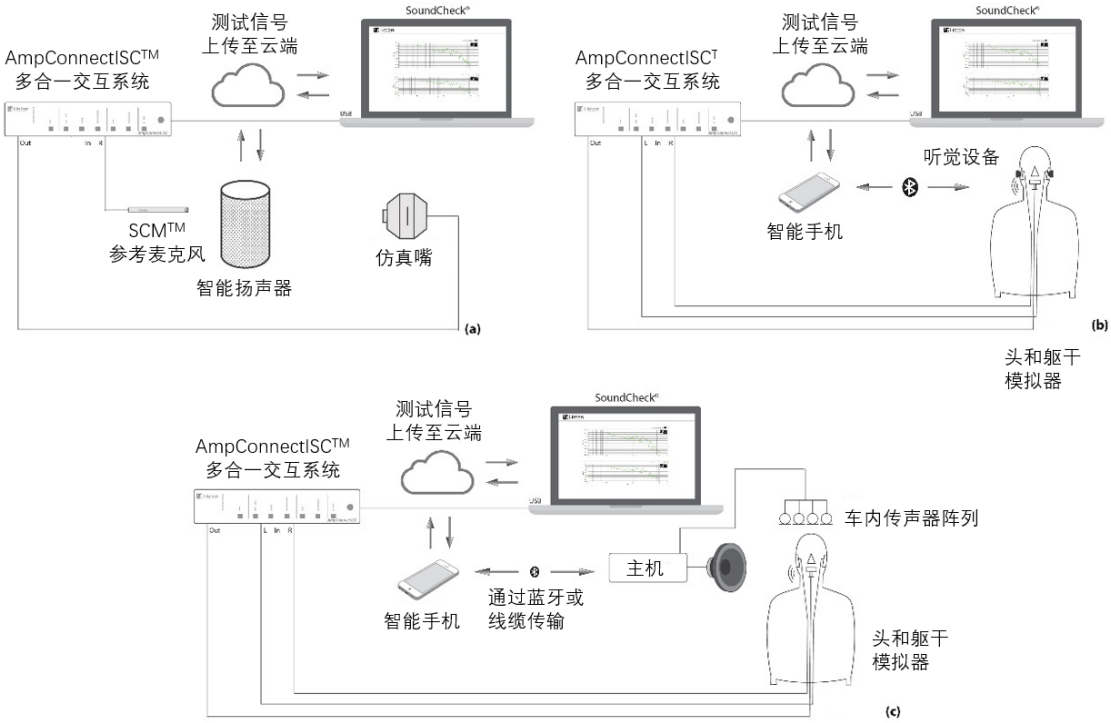

图 2 基于云端进行的基本扬声器测试，a）智能扬声器；b）听觉设备；c）汽车信息娱乐系统

Fig. 2. Test configuration for basic speaker measurements via the cloud for 2） smart speaker; b） hearable; and c） automotive infotainment

1.3 □ □

下面讨论使用语音激活功能创建重放测试所需的步骤。理想情况是使测试尽可能自动化，因此应预先录制唤醒词和测试激励信号，并将其上传到云端的音乐服务中，以便由测试系统触发。首先记录唤醒词，例如，“Alexa，播放测试信号”。理所当然地，我们应该使用适合系统的唤醒词。如果测试来自不同厂商的各种语音识别系统，则有必要将不同的语音唤醒单词与命令分开记录，以便可以选择任何组合。

另外，还需要创建测试激励信号并将其上传到云端。需要谨记的是，信号在上传时可能会被压缩而导致些许失真。通过对测试信号进行编解码，并在原始信号和编码版本之间进行传递函数分析，可以证明 MP3 编码本身不会降低测试信号的质量。测试信号将在下文进行更详细的讨论，但是对于基本的重放性能测试，首选标准步进正弦扫描信号。

一旦创建并上传了激励信号和语音命令，测试序列只需要播放唤醒词，然后通过仿真口发出激活命令，并使用参考传声器或仿真耳捕获响应即可。然后像传统测量中那样比较激励信号，进行分析，以客观地得到频率响应、脉冲响应和失真等特性。

通过测试系统处理所有这些交互信息，可以以图像显示听觉设备、汽车信息娱乐系统或智能扬声器的频响和失真，显示方式与基础的有线或蓝牙扬声器的相同。

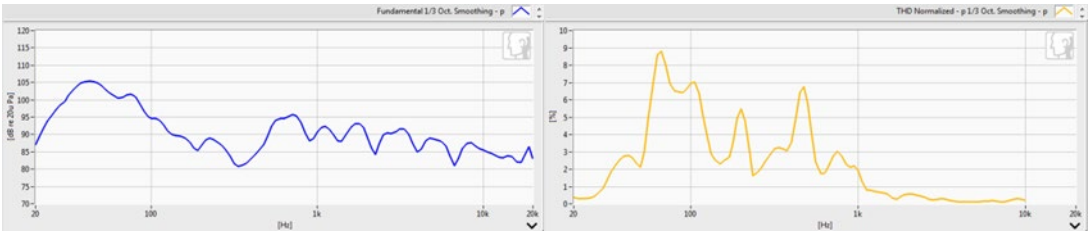

图 3 信息娱乐多通道播放系统的频率响应和失真响应

Fig. 3. Frequency and distortion responses of an infotainment multichannel playback system

1.4 □□□□

测试系统厂商在开发精确的智能系统测试中面临的主要挑战之一是克服采样率误差。传统的测量系统依赖于被测设备具有同步输入和输出，然而听觉设备、信息娱乐系统和智能扬声器本质上是开环的，这意味着尽管设备可以记录或播放信号，但没有同步信号路径。这可能会引入采样率误差，换句话说，当设备以 44.1 kHz 的采样率将信号记录到文件中时，由于用于驱动采样器的时钟晶体的实际速率发生偏斜，实际的采样率可能为 44.09 kHz（图 4）。在重放测试信号时也可能存在类似的问题，即可以以 44.1 kHz 的采样率对测试激励进行采样，但是由于重放采样率中的错误，文件实际上是以稍快或稍慢的速率进行重放的。这种采样率误差会导致测试激励信号的频率成分移至较高或较低的频点上。由于激励信号和响应信号之间的一致性损失，这种偏移会导致测量误差，也会改变相位响应。

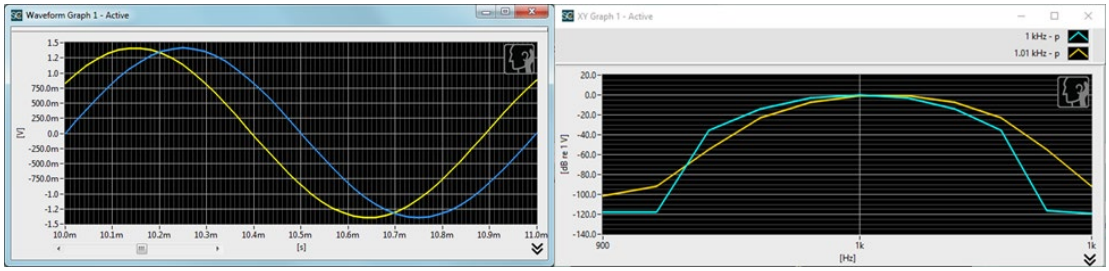

图 4.具有不同采样率的开环测量，以及重新采样以匹配测试信号采样率的记录波形

Fig. 4. Open loop measurements with differing sampling rates, and the Resampled Recorded Waveform to Match Sample Rate of Test Signal

我们运用了一种算法来克服这种采样率误差，该算法在响应波形的开头搜索“触发器”以提供参考点，从而对齐和移动对激励信号的响应。该触发器可以是预设电压和频率下的稳态正弦曲线（图 5），为了得到更稳定的性能，避免受到错误触发器的影响，也可以设置为连续对数扫频信号。然后使用外差滤波器将信号移至直流分量，滤除所有其他频率。外差滤波器的输出包含相位信息，可用于最终估计实际重放信号或记录响应信号的采样率。利用此信息，可以在分析之前将整个响应波形重新采样到正确的激励采样率。上述移频步骤校正了设备中的采样率误差，降低了测试的难度。

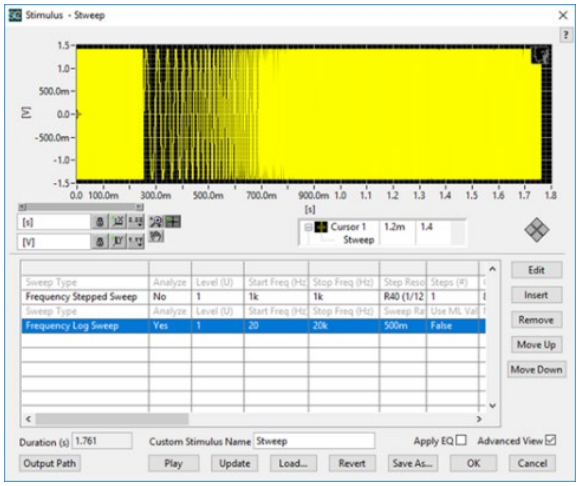

图 5：在步进正弦激励之前加上 1kHz 频率，以计算采样率并简化记录触发

Fig. 5: 1kHz frequency tone prepended to stepped sine stimulus to calculate sampling rate and simplify record triggering

1.5 □ □

当然，还需要测量智能设备中的传声器或传声器阵列。其设置非常相似，不同的是在这种情况下，唤醒词和测试激励是通过仿真口或者声源扬声器播放、由设备的传声器捕获的，激励信号通过调用记录的波形文件或使用虚拟声卡从云端下载得到，并传回测试系统（图 6）。以这种方式将信号输入到测试系统后，应像重放测试一样，通过适当的计算来克服频移和误差，并按常规手段进行分析。

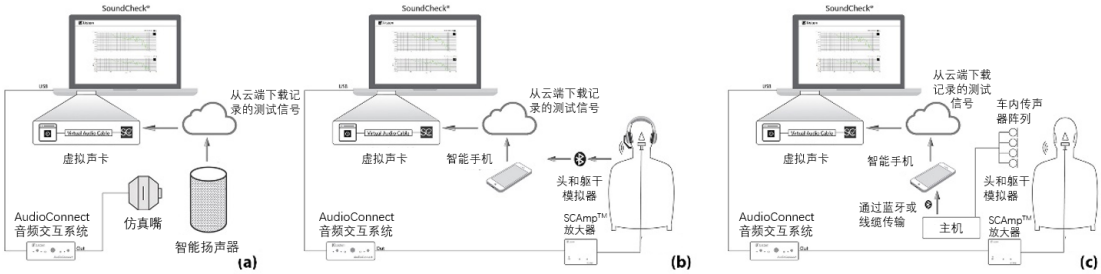

图 6 传声器的测试配置，a）智能扬声器；b）听觉设备 ；c）汽车信息娱乐系统

Fig. 6: Test configurations for measurement of the microphone within a） a smart speaker; b） a hearable; and c） an automotive infotainment system

2 □ □ □ □ □ □ □ □ “ □ □ □ □ ” □ □

2.1 □ □ □ □ □

尽管上述基本测试非常适合测量频率响应和失真，但它们难以达到人们想在研发实验中对智能设备（如智能扬声器，听觉或娱乐信息设备）进行测试的广度和深度。考虑到智能设备和系统通常是非线性的，它们的响应会根据许多因素而变化，例如背景噪声的类型、大小和设备是用于播放音乐还是传达语音等。仅考虑这些因素就需要大量测试，这些测试可能会带来有关系统性能的有价值的信息。尽管执行的特定测试会根据厂商的设计目标而有所不

同，但我们应讨论一些可能使用的测试方法和能提供有用起点的行业标准。

与基本测试的显著区别是，高级测试通常需要使用语音或音乐作为测试信号，这意味着能够支持此类测试信号的测试系统必不可少，而且该系统测量基于语音的信号的语音活跃度的能力也很重要，这可以在例如电话测量的工作中确保校准和标准化工作的进行。

2.2 语音识别

语音识别测试对于评估系统理解语音命令的能力非常重要，它的物理设置与上述基本测试中的设置相同。一种测试方法是创建一系列带有特殊标题的音乐曲目并上传到例如 iTunes 的云端服务中，曲目的标题为待测单词或短语，其实际内容为单音、双音或多音以便通过音频内容来识别轨道。通常来说，可以将“Harvard Sentences”（音节平衡的语句的集合）用作曲目标题。它们使用特定的、使用频率与英语中出现频率相同的音素，来源于“IEEE Recommended Practices for Speech Quality Measurements for VoIP, Cellular and other telephone systems”[10]的推荐，并已广泛用于电信语音识别之中。我们可以在线下载其预录制版本，也可以录制自己的版本，比如可以使用不同的口音测试设备。这些曲目标题用于命令播放信号，我们在实验中可以通过添加限制来检测是否播放了正确的信号或者实际上是否播放了信号（图 7）。

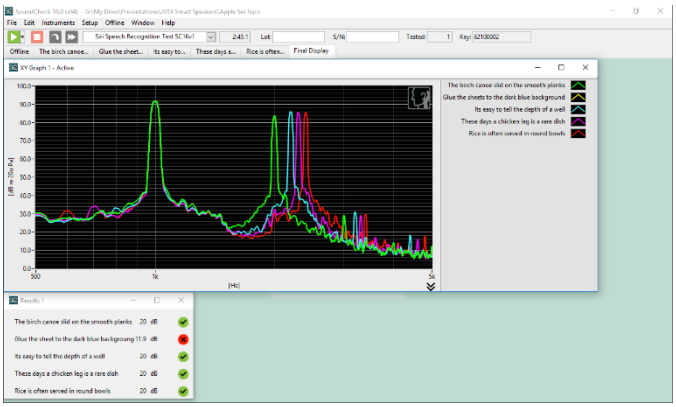

Fig. 7. Graphs showing results of end-to-end speech recognition

图 7 端到端语音识别结果示意图

此测试有很多变体，例如测试不同的声音和口音、测试活跃的讲话者与干扰的讲话者或背景噪声水平、使用不同类型的背景噪声进行测试等。

2.3 背景噪声

背景噪声对于各种智能系统和设备测试都是必须的，这些测试包括语音识别、传声器阵列的信噪比优化、波束形成方向性研究等等。当需要背景噪声信号时，我们可以参考 ETSI 标准 ES 202 396-1[9]，它包含不同噪声环境下进行的双耳录音，例如咖啡厅噪音、地铁噪音、道路噪音、交通噪音、警笛声等。我们可以借助昂贵的专业仪器或者足够的专业知识，通过自己编程来形成播放方法。但是，某些系统提供了集成的 ETSI 背景噪声模块，它们将 ETSI 标准库完全集成至系统中，以经过校准的声压级播放背景噪声，我们可以将其作为预编程测试序列的一部分进行控制和调整。该测试序列可能包括以逐渐增加的音量或改变的噪声循环重复测试的功能，这降低了我们创建测试的难度，便于进行提高背景噪声的固定声压级直到无法准确识别语音为止这样的实验。除了非常经济高效以外，它还可以显著简化物理测试的设置并减少测试开发时间。

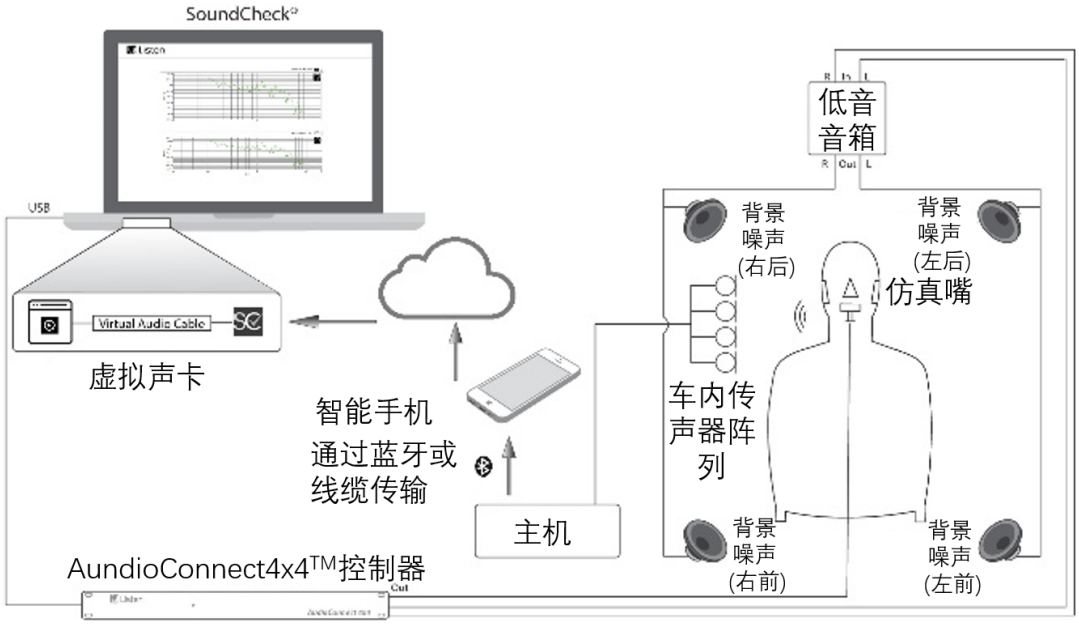

图 8.使用 ETSI 标准背景噪声进行免提测量的测试配置。 在听觉设备测试设置中，设备的传声器将取代车载阵列，并且信号将直接从该阵列传至智能手机。

Fig. 8. Test configuration for hands-free measurements with ETSI standard background noises. In a hearable set-up the device's microphone would replace the in-car array, and the signal would go directly from that to the smartphone.

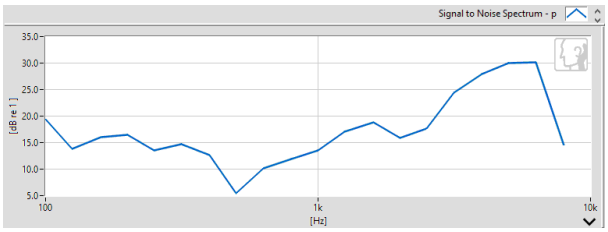

图 9.用 ETSI 标准背景噪声测量的传声器阵列的信噪比

Fig. 9. SNR for a microphone array measured with ETSI standard background noises

2.4 语音质量

智能扬声器通常被用作免提通信设备。在这种情况下，我们可以大量依靠电话测试序列，例如，我们可以参考涵盖免提设备（如在汽车应用和智能扬声器中的设备）和耳机设备（如听觉设备）的 TIA 标准 TIA920-B[5]。实际上，我们可以完全按照 TIA 标准中的定义，将智能扬声器和信息娱乐系统作为免提电话进行测试，将听觉设备作为耳机进行测试。

尽管每种情况下的测试设置都与上面概述的基本声学测试设置相同，但是按照标准进行测量的测试序列非常复杂，或者需要该领域的大量专业知识来创建测试，也可能需要购买作为整体的现成软件包。

2.5 噪声消除

尽管降噪测试最常用于听觉设备，但它们在汽车信息娱乐系统中也可用于减少机舱背景

噪声。测量降噪耳机的测试配置如图 10 所示。测试过程可以分为 3 个测量步骤和 3 个计算步骤。首先，将耳机从头部和躯干模拟器上移开，播放噪声信号，并通过开启状态下的仿真耳测量频谱，未被遮挡的仿真耳记录的频谱将被用作噪声的基准。接下来，将耳机放在头和躯干模拟器上，再次播放噪声信号并通过仿真耳测量噪声来测量被动隔噪情况。最后，在打开有源噪声消除（Active Noise Cancellation，ANC）电路的情况下重复进行测量。

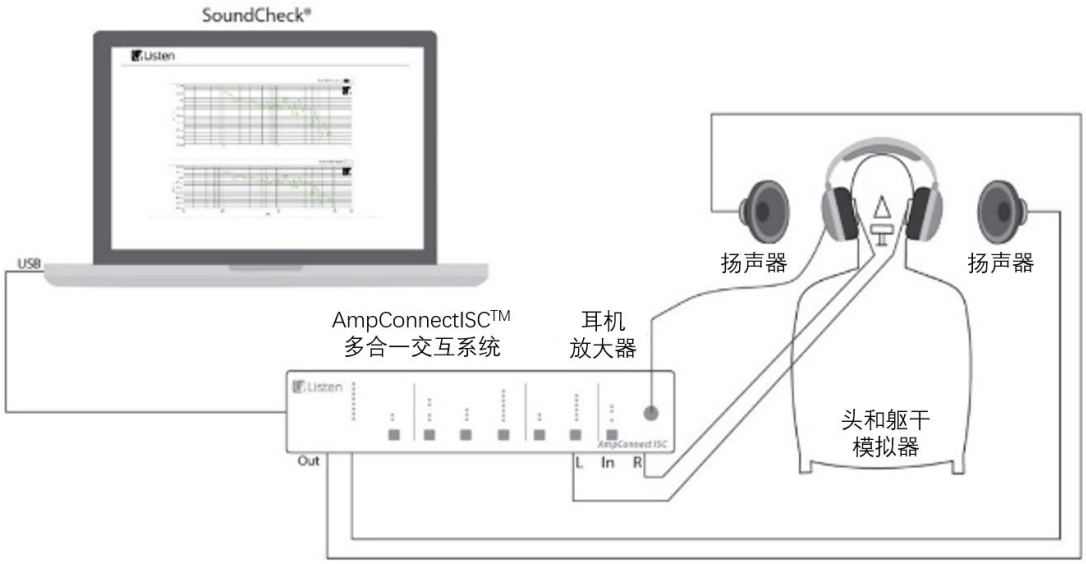

图 10.测量主动降噪耳机的测试配置

Fig. 10. Test configuration for measuring ANC headphones

这些测量将产生 3 个频谱，用于计算 3 个衰减量。被动隔离量可通过从安装耳机的第二次测量结果中减去未遮挡的基准曲线计算得到，它表征在未开启降噪功能时简单地戴着耳机所衰减的噪声量。被动隔离量在整个频带上可能都很显著，但在较高频率下最突出。耳罩式耳机和入耳式耳机可提供较合理的被动隔离量，而压耳式耳机耳机和耳塞式耳机通常提供的隔离度要低得多。

主动衰减量可以通过第三次测量值（降噪功能开）减去第二次测量值（降噪功能关）获得，它表征在开启降噪功能时减少（有时是增加）的噪声量，通常在低频较为显著。

最后，总衰减量可以通过从不带耳机的基准测量结果中减去打开降噪功能的曲线计算得到。它将无源和有源组件结合在一起以衰减背景噪声（图 11），代表了最终用户对设备的体验。

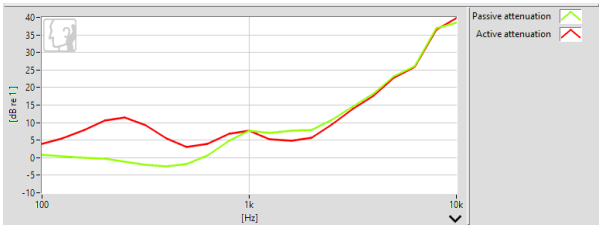

图 11 典型的噪声消除测量结果

Fig. 11. Typical Noise Cancelling Measurement Results

通常可以通过多次执行各测量步骤并平均所得频谱来优化此测试，这说明了耳机贴合度的变化，对于某些设备这可能会影响降噪性能的被动和主动部分。某些降噪耳机通常允许通过智能手机应用程序控制降噪等级。这应该是考虑到不同的收听环境，例如飞机上可以完全消除噪声，而在繁忙街道旁的人行道上行走时，应该更加了解周围环境，以免被汽车、自行车甚至另一个人撞到。在这种情况下，显示不同环境下降噪等级的频响的测量是一个有效方法。在这种情况下，同样应使用校准的模拟背景噪声系统（如 ETSI 指定）以易于重复的方式重建不同测试环境以进行实验。

同样，当启用降噪功能时，某些降噪耳机实际上会改变正在播放的音乐的频率响应。为了测试这种效果，可以使用上面概述的方法（无论是否启用降噪）测量频率响应，并比较最终的响应曲线。可能存在的差异通常出现在 1-3 kHz 范围内。

2.6 助听器

听觉设备相当特殊的地方在于其整合了听觉增强功能，因此它们还可以充当“个人声音放大产品（Personal Sound Amplification Products，PSAP）”或助听器，并且需要测量其放大或压缩算法以及相应的起音和释放时间。这些测量通常是根据 ANSI / IEC 标准[7,8]在消声测试箱（图 12）中进行的，在该消声箱中，扬声器被均衡以具有平坦的频率响应，而听觉设备则被插入 2 毫升的耳耦合器或者仿真耳中进行测试。

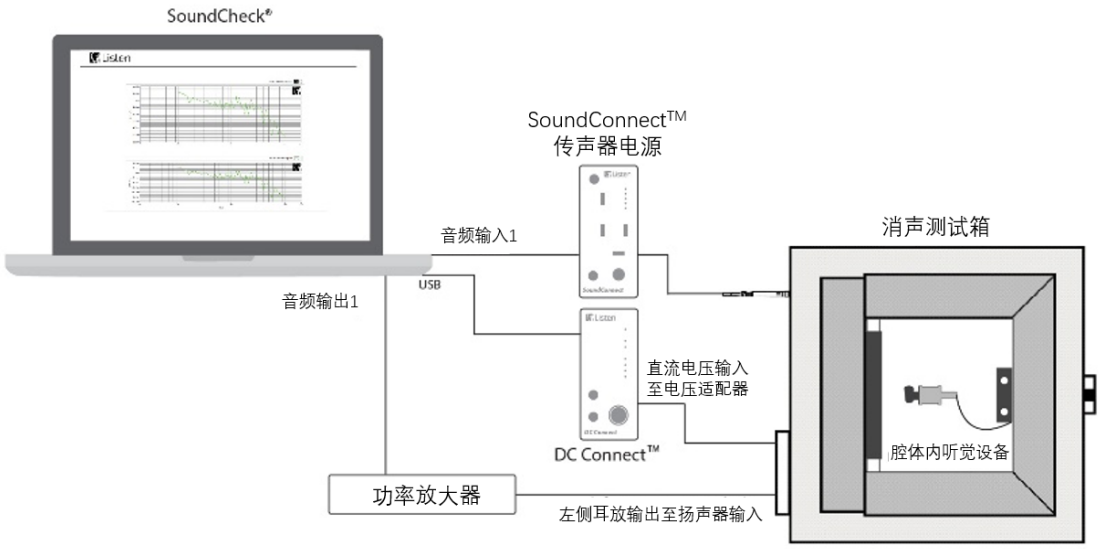

图 12 测量听觉设备是否符合助听器标准的实验设置

Fig. 12. Test configuration for measuring hearable to hearing aid standards.

ANSI S3.22 标准[7]描述了测试具有自动增益控制（Automatic Gain Control，AGC）功能的助听器的方法，它可以应用到提供此功能的听觉设备上。待测设备置于消声箱中，经由测试箱扬声器播放 50-90dB 声压级的幅度扫频信号，在 500Hz 到 4000Hz 之间的、相隔一个倍频程的四个频率点重复扫频。绘制助听器的输出与图表上的输入的关系曲线，以量化对低电压信号的增益和为听力损失补偿而在高电平处产生的压缩（图 13）。

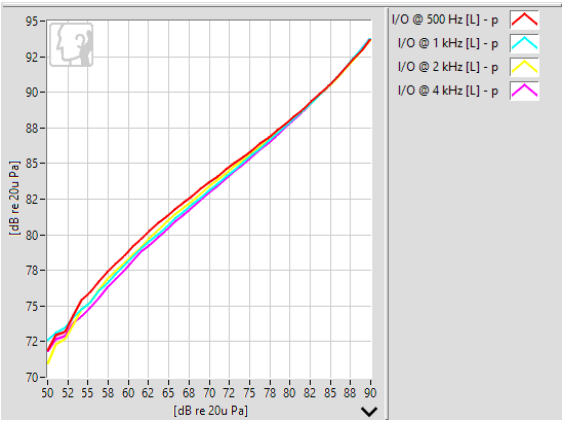

图 13 听觉设备输入输出测量结果

Fig. 13. Hearable measurement of Input vs. Output

3 □□□□□□□□□□

本文介绍了多种测量现代智能设备音频性能的方法，这些设备包括智能扬声器、耳机/听觉设备和汽车信息娱乐系统。由于其中大多数系统都具有语音控制和内置算法的智能手机集成模块，因此测量有必要包含这些功能。这些复杂的功能需要包括整个信号路径的开环测量，并在现实条件下使用语音和音乐作为测试信号。这些测量代表了扬声器、耳机、传声器、电话、助听器等的测试的集合，包含了大量的测试激励信号、算法、分析选项和信号路径选项，更不用说也包含了许多“有用”功能，例如集成的背景噪声模块、易于连续多次迭代序列的测试配置器等。以下是在选择测试系统之前应评估的功能清单：

- 可以测试设备的输入和输出组件（即扬声器或耳机和传声器）
- 可以创建具有一定频率和/或对数线性扫频触发器的复合激励激励信号，以实现准确的开环测试
- 可以设置为仅分析部分响应信号以便从分析中消除触发音
- 能够将语音和音乐用作测试信号、校准音量并均衡仿真口（要求包括主动语音算法）
- 背景噪声生成——尽管可以从外部生成，但是如果可以由测试系统创建并与测试序列集成，则测试将更加简单快捷
- 可以通过虚拟声卡接收信号，以便将云端信号传输至本地 Windows 多媒体接口
- 如果需要按照电话标准测试设备，则可以使用预先编写的 TIA 测试序列（可以基于足够的专业知识手动编程）
- 测试编程的灵活性 - 允许简单重复子序列和循环，这在重复进行不同级别的测试时非常有帮助
- 提供已校准的蓝牙接口
- 如果要测试汽车智能设备（例如车内语音控制），可以兼容 Mentor A2B 接口和控制测试系统内部

□□□□

[1] IEC 60268-7: Sound system equipment – Part 7: Headphones and earphones

[2] BS EN 50332-1 Sound system equipment – Headphones and earphones associated with portable audio equipment – Maximum SPL measurement methodology and limit considerations

[3] IEEE 1329-2010, “Standard Method for Measuring Transmission Performance of Speakerphones”, October 2010
[4] IEEE 269: Standard Methods for Measuring Transmission Performance of Analog and Digital Telephone Sets, Handsets, and Headsets
[5] TIA 920.130: Telecommunications Communications Products Transmission Requirements for Digital Interface Communications Devices with Headsets
[6] ITU-T Recommendation P.51, “Artificial Mouth”, August 1996
[7] ANSI S3.22 Specification of hearing aid characteristics
[8] IEC 60118-7: Electroacoustics - Hearing aids - Part 7: Measurement of the performance characteristics of hearing aids for production, supply and delivery quality assurance purposes
[9] ETSI ES 202 396-1: Speech and multimedia Transmission Quality （STQ）; Speech quality performance in presence of background noise; Part 1: Background noise simulation technique and background noise database
[10] IEEE Recommended Practice for Speech Quality Measurements
[11] Karlheinz Brandenburg, “MP3 and AAC Explained”, presented at AES17th International Conference on High-Quality Audio Coding, August 1999
[12] Steve Temme et al., “The Challenges of MP3 Player Testing”, presented at the AES 122nd Convention, May 2007
[13] Glenn Hess et al., “Challenges of IoT Smart Speaker Testing”, presented at the AES 143rd Convention, October 2017
[14] Steve Temme, “Testing Audio Performance of Hearables”, presented at the International Conference on Headphone Technology, August 2019
[15] IEC 61094-4, “Measurement Microphones Part 4: Specifications for Working Standard Microphones”, November 1995

□ □ □ □

Steve Temme 是 SoundCheck 音频测试系统的制造商——ListenCheck 公司的创始人兼总裁。Steve 于 1995 年成立公司，在过去的 24 年中，公司始终处于音频测量研究的最前沿，定期引入新的测量技术、算法和硬件，以测试从基本换能器到复杂音频系统的各种设备。近年来，Listen 的产品开发重点一直放在创建算法、接口和测试方法上，致力于测量当今多种多样的无线和云通信设备。在创立 Listen 之前，Steve 在 Brüel&Kjær 担任声学测试和测量应用工程师，并在 Apogee Acoustics 担任扬声器设计工程师多年。他拥有 Tufts 大学的 BSME 学位，撰写了许多有关声学测试的论文，并且在世界各地进行了广泛的演讲。