

沉浸式音频的未来： 计算机视觉辅助下的 声音重现

Santosh Singh, 消费电子系统应用高级工程师
Aravind Navada, 消费电子系统应用亚太区总监

摘要

消费类娱乐对沉浸式体验的要求越来越高，用户希望获得身临其境的体验并消费内容，声音则是临场感的重要组成部分。在未来，我们将更了解人类大脑处理和定位声音的方式，由此开发出创新的声音重现技术，助力基于视觉智能的音频系统。ADI公司将依托前沿的飞行时间(ToF)成像器和先进的DSP技术组合，为新一代沉浸式音频系统构建理想的平台。

新时代消费电子娱乐设备频频提到“沉浸”一词，但其真正含义是什么呢？在1999年爆火的电影《黑客帝国》中，Morpheus询问Neo他能闻到、尝到或触摸到的东西是否真实，并展示他所知的真实世界不过是用计算机愚弄人的感官而已。这就是真正的沉浸式体验，亦是人工沉浸式体验要达成的目标。

要让自己确信已置身于某个场景中，声音和感触方式是影响整个体验的关键。声音会激活大脑，决定着我们对情境的第一反应。大脑利用声音来更清晰地构建所处的环境或情境。声音说服大脑相信人工构建的沉浸式体验，在提供预期的沉浸式体验方面发挥着至关重要的作用。

多年来，声音重现技术取得了巨大飞跃，从基础的单声道音频系统到如今的环绕音频系统，从适合家庭影院的小型5.1 (6声

道) 或7.1 (8声道) 配置到适合影院屏幕的大型64声道和更高级别的配置。但在这些系统中，声音的空间感和精度受到扬声器数量和所处位置的限制。

新型声音重现技术基于对大脑声音处理和定位方式的深入了解，助力构建新一代沉浸式音频系统，无需在听众周围部署大量扬声器。即可为家庭影院带来360度沉浸式声音体验，但此类系统由于缺乏对听众及听音环境的感知，这便也成为了沉浸式音频需求的主要障碍。视觉智能与声音重现技术的组合可应对这一挑战，真正打造出下一代沉浸式音频系统。

在真实场景中自然接收声音时，我们的大脑会基于传到左右耳的音频信号来提取有关声源的空间线索。这与我们的双眼视觉系统的工作原理非常相似，大脑也是通过结合左右眼所看到的图像来感知深度。大脑处理到达左右耳的声音，通过比较振幅和时间延迟来推算声源位置。这是人类在进化过程中形成的能力，也是原始社会的关键生存技能。

双耳声音重现技术旨在通过新型信号处理，在左右耳生成与真实场景相同的左右音频信号，再现声音的自然体验(图1)。但在实践中实现这一目标并非易事，会面临重重问题。

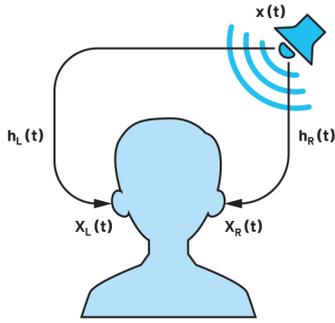


图1. 来自声源 $x(t)$ 的自然接收场景, $x_L(t)$ 表示到达左耳的音频信号, $x_R(t)$ 表示到达右耳的音频信号。

记录双耳音频的一种简单方法是在真实环境中人的左右耳各部署一个麦克风, 然后记录到达每只耳朵的声音信号, 这种方法称为双耳记录。然后通过耳机重现声音, 传至听众的耳朵。那这种方式效果如何呢? 在针对同一位听众进行捕捉和回放时确实有效, 但由于每个人大脑定位声音的方式不同, 这一技术并非百试百灵。我们的头/耳廓/身体对声音的影响会在频域中留下特定的特征, 帮助我们的大脑定位声音。这种特征因人而异, 被称为头部相关传递函数(HRTF)。如想让双耳技术真正有效, 须在声音重现过程中听众的耳朵上准确再现HRTF对声音的影响。

因此, 我们需要针对每位听众测量并定制HRTF, 不能采用通用的解决方案。研究表明, 当人们体验用其他人的HRTF制作的音频时, 其在体验期间的声音定位能力会显著降低。^{12,3}

在扬声器上实现双耳音频还会面临更大的挑战。首先, 来自多个扬声器的声音信号会相互干扰, 即所谓的串扰效应(图2)。其次是听音环境, 在声音到达听众耳朵之前, 它不可避免地会对声音产生一些影响。

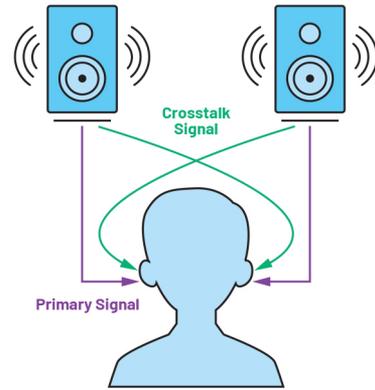


图2. 立体声扬声器中的串扰效应。

在实现真实模拟自然声音接收体验时, 扬声器串扰、HRTF个性化需求以及房间/听音环境的影响是主要的阻碍因素。而视觉系统能够捕捉到有关听众和听音环境的所有细节, 有助于解决双耳声音重现所面临的挑战。

例如, 可以构建为计算机视觉算法提供数据的摄像头来捕捉声音接收环境的三维架构信息(即房间形状、不同表面的几何测量细节以及存在的物体), 用于计算听音环境对声音的影响。然后, 可以在声音重现系统中适当增加滤波器和滤波器系数以消除不良影响。虽然家庭影院音频已采用此类系统, 但它通常依赖于在校准期间使用麦克风捕捉房间对声音的影响, 如果接收位置或房间结构发生变化, 则需要重新进行校准。

视觉系统可以进一步捕捉人体测量数据, 比如身体位置和结构细节⁴, 通过必要的计算将HRTF个性化, 以呈现准确的空间线索(图3)。使用听众头部位置相对于扬声器的信息和头部尺寸, 部署串扰消除算法, 在扬声器中呈现实时双耳音频, 听众能够随意移动, 同时保持理想的的声音体验(图4)。

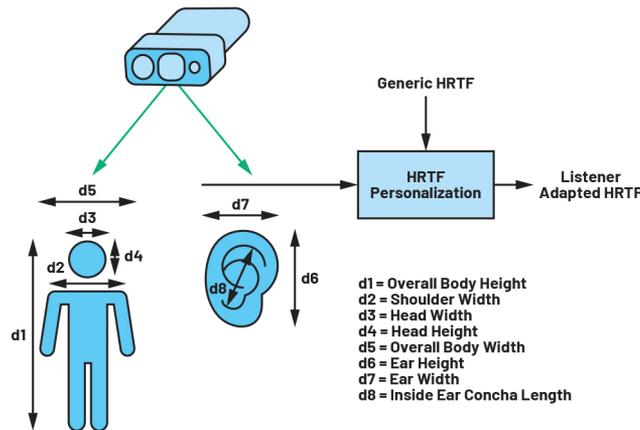


图3. 通过人体测量实现HRTF个性化。

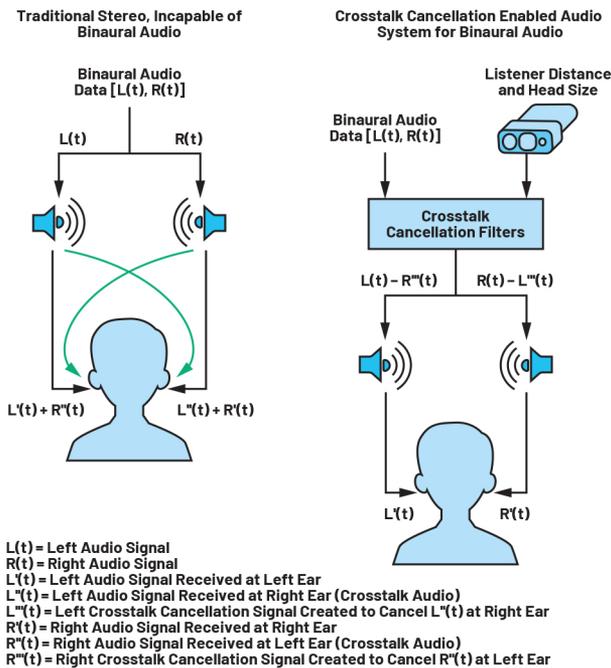


图4. 采用串扰消除技术，通过自由场扬声器系统实现双耳声音再现。

使用音频系统存在隐私问题，但从视觉系统获取的摄像头数据会经过实时处理，无需存储或传输到另一台远程机器，因此，使用专用的计算处理器在边缘处理视觉数据分析可以保护用户隐私。

ADI最新的多核SHARC® DSP和先进的ToF成像器提供了硬件平台实现音视频融合所需的关键部件，以创建下一代沉浸式音频系统（图5）。

我们的ADSP-SC598 SOC搭载SHARC双核和一个A55 Arm®内核，由大型片内存储器器和外部存储器DDR接口提供支持，可满足低延迟和内存密集型计算要求，是实现真正沉浸式音频的理想平台（图6）。

SHARC DSP上的计算资源，例如ADSP-SC598，可以将与音频解码相关的工作负载划分至DSP内核上，在第二个SHARC内核上实现音频回放的后处理和个性化。Arm A55可用于进行多种控制处理。⁶图5所示的视觉系统可以组合使用RGB和深度摄像头或单独使用深度摄像头。我们的高分辨率100万像素ToF深度成像器ADSD3100可以捕捉毫米分辨率级别的深度图，且能在不同的照明条件下工作，为之前所述的个性化算法（串扰消除、房间均衡、HRTF人性化等）提供了所需的高精度几何测量数据。

ADTF3175是基于ADSD3100 ToF深度成像器的100万像素、75 × 75度视场(FOV) ToF模块，它还集成了用于成像器的透镜和光学带通滤波器、红外光源（包含光学元件、激光二极管、激光二极管驱动器和光电探测器）、闪存和功率调节器以生成本地电源电压。该模块在多个范围和分辨率模式下进行完全校准。如需完善深度测量系统，可以将来自ADTF3175的原始图像数据通过主机系统处理器或深度ISP进行外部处理。ADTF3175图像数据输出接口通过4通道移动行业处理器接口(MIPI)、摄像头串行接口2 (CSI-2)变送器接口与主机系统进行电气接口。该模块编程和操作通过4线式SPI和I2C串行接口进行控制。

我们当前提供的EVAL-MELODY-8/9开发平台板、EV-2159X/SC59x-EZKIT板板和CrossCore®Embedded Studio（一款基于eclipse的编辑工具）可以帮助您部署和运行我们的ADSP SOC，以实时部署和调试应用。⁷

Melody平台是ADI为AVR和条形音箱应用提供的完整信号链解决方案。它将视频、DSP、音频、电源和软件方面的一流ADI组件结合到组合系统解决方案中，使客户能够利用最新技术快速进入市场，以达到他们的年度升级窗口。⁸

ToF模块ADTF3175可以连接至视觉计算平台，并连接至Melody板，为下一代沉浸式音频系统构建硬件平台（图7）。RGB摄像头可耦合至ADTF3175 ToF模块，构建RGBD摄像头来进行强化视觉分析。

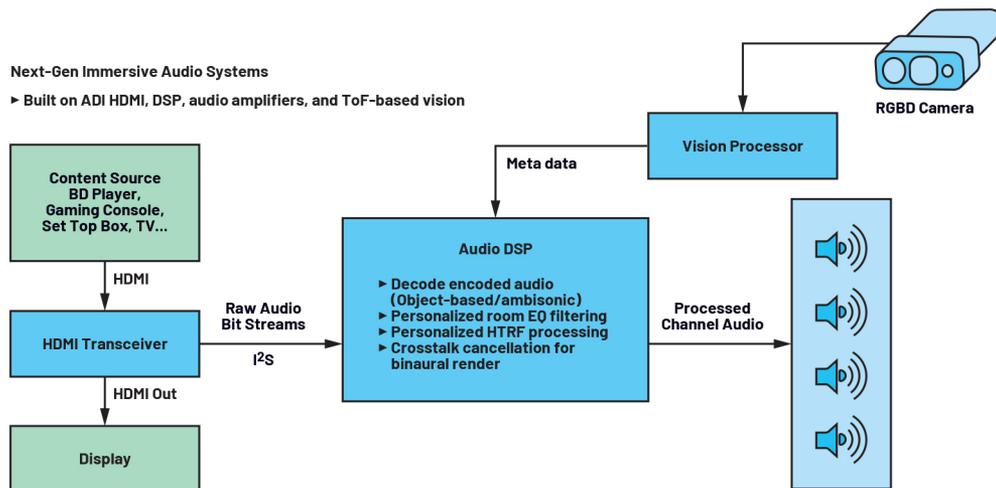


图5. 下一代沉浸式音频系统。

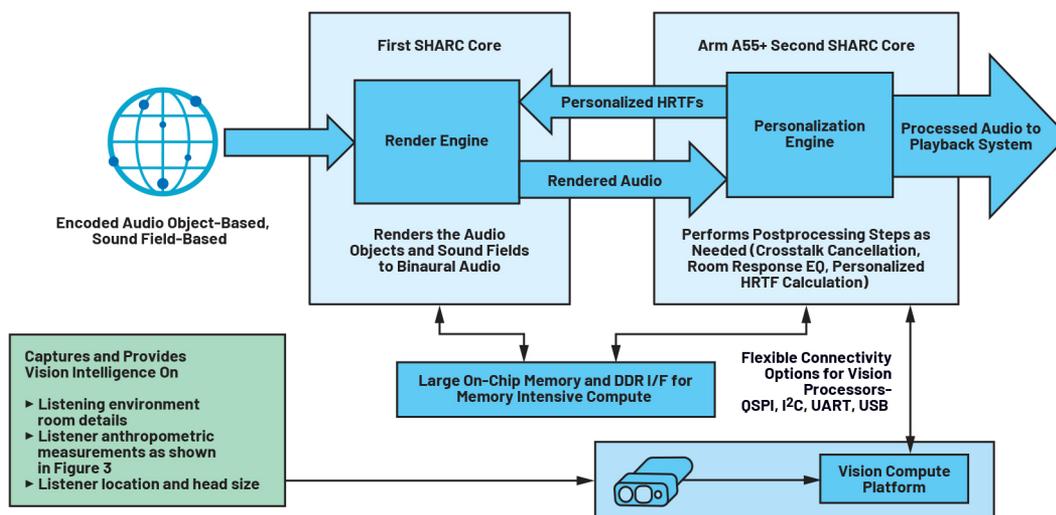


图6. 下一代沉浸式音频系统的系统分区。

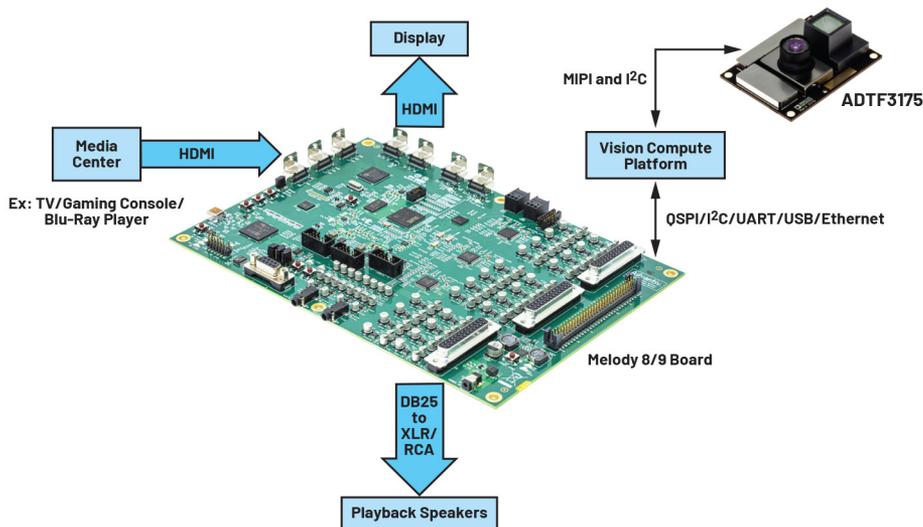


图7. 使用ADI平台实现沉浸式音频系统。

结论

ADI借助包含DSP、HDMI收发器、D类放大器和ToF成像器的解决方案系列，持续不懈地追求实现真正沉浸式的音频系统，力求提供与真实世界一般无二的声音。

参考资料

- ¹ Philipp Paukner、Martin Rothbucher、Klaus Diepold, [不同HRTF个性化方法的声音定位性能比较](#), 慕尼黑工业大学, 2014年4月。
- ² Parham Mokhtari、Ryouichi Nishimura、Hironori Takemoto, [实现HRTF个性化: 对模拟和实测HRTF的听觉感知评估](#), 国际听觉显示会议 (ICAD), 2008年7月。

- ³ Jenny Claudia、Christoph Reuter, [个性化头部相关的转换功能在虚拟现实中的可用性: 矢状面声音定位中知觉属性的实证研究](#), JMIR Serious Games, 2020年9月。
- ⁴ Geon Woo Lee、Hong Kook Kim, [基于深度神经网络使用人体测量数据和耳朵图像构建的个性化HRTF模型](#), 应用科学, 2018年11月。
- ⁵ Sanket Nayak、Mitesh Moonat, [EE-436: 使用ADSP-SC59x/2159x高性能FIR/IIR加速器](#), ADI公司, 2022年6月。
- ⁶ [ADSP-SC59x/2159x SHARC系列双倍提升产品系列的性能, 为复杂音频应用提供可用平台](#), ADI公司。
- ⁷ [CrossCore® Embedded Studio](#), ADI公司。
- ⁸ [家庭影院和游戏](#), ADI公司。

作者简介

Santosh Singh于2016年毕业于比尔拉科学技术学院，获得电子和通信学士学位。在过去的6年里，他一直在ADI公司工作，从大学实习生转为全职员工。他的职业生涯始于数字设计师，致力于为消费市场设计各种尖端技术产品。Santosh目前担任消费电子事业部的高级系统应用工程师。他主要负责解决关键系统挑战，帮助弥合应用和技术之间的差距。

Aravind K. Navada管理着消费电子系统应用团队，特别关注亚洲客户。他拥有20多年的IC设计经验，曾将各种混合信号物联网和消费电子SOC推向市场。他目前专注于为消费电子市场开发平台和解决方案，从而最大限度地提高ADI产品渗透率。Aravind拥有班加罗尔瓦拉亚大学工程学院(UVCE)电子和通信学士学位，以及德克萨斯大学达拉斯分校微电子硕士学位。

在线支持社区



访问ADI在线支持社区，中文技术论坛
与ADI技术专家互动。提出您的棘手设计问题、浏览常见问题解答，或参与讨论。

请访问ez.analog.com/cn



超越一切可能™

如需了解区域总部、销售和分销商，或联系客户服务和技术支持，请访问analog.com/cn/contact。

向我们的ADI技术专家提出棘手问题、浏览常见问题解答，或参与EngineerZone在线支持社区讨论。请访问ez.analog.com/cn。

©2023 Analog Devices, Inc. 保留所有权利。商标和注册商标属各自所有人所有。

“超越一切可能”是ADI公司的商标。

T24192sc-4/23



请访问analog.com/cn