

沉浸式音訊的未來： 電腦視覺輔助下的原音重現

消費性娛樂對沉浸式體驗的要求越來越高，使用者也希望能獲得更身歷其境的視聽體驗。聲音是臨場感的重要組成部分，未來，我們將更瞭解人類大腦處理和定位聲音的方式，由此開發出創新的聲音重現技術，助力基於視覺智慧的音訊系統。ADI 將根據尖端的飛時測距 (ToF) 成像器和先進的 DSP 技術組合，為新一代沉浸式音訊系統構建理想的平台。

■文：Santosh Singh / ADI 消費性系統應用資深工程師
Aravind Navada / ADI 亞洲消費性系統應用總監

新時代消費性電子娛樂設備頻頻提到“沉浸”一詞，但其真正含義是什麼呢？在 1999 年爆火的電影《駭客任務》(The Matrix) 中，Morpheus 詢問 Neo 他能聞到、嘗到或觸摸到的東西是否真實，並展示他所知的真實世界不過是用電腦愚弄人的感官而已。這就是真正的沉浸式體驗，亦是人工沉浸式體驗要達成的目標。

要讓自己確信已置身於某個場景中，聲音和感觸方式是影響整個體驗的關鍵。聲音會啟動大腦，決定我們因應情境的第一反應。大腦利用聲音來更清晰地構建所處的環境或情境。聲音說服大腦相信人工構建的沉浸式體驗，在提供預期的沉浸式體驗方面發揮著非常重要的作用。

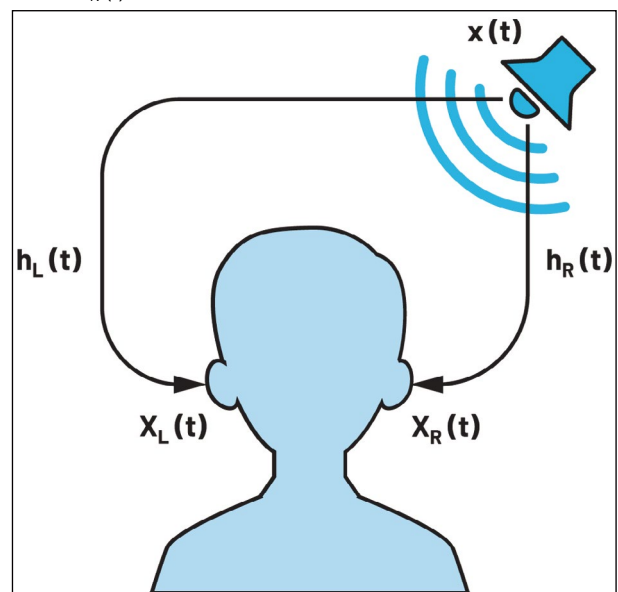
多年來，聲音重現技術取得了巨大躍進，包括從基礎的單聲道音訊系統到如今的環繞音訊系統，以及從適合家庭影院的小型 5.1 (6 聲道) 或 7.1 (8 聲道) 配置到適合影院螢幕的大型 64 聲道和更高級別的配置。但在這些系統中，聲音的空間感和精度受到揚聲器數量和所處位置的限制。

新型聲音重現技術基於對大腦聲音處理和定位方式的深入瞭解，助力建構新一代沉浸式

音訊系統，無需在聽眾周圍部署大量揚聲器。即可為家庭影院帶來 360 度沉浸式聲音體驗，但此類系統由於缺乏對聽眾及聽音環境的感知，這便也成為了沉浸式音訊需求的主要障礙。視覺智慧與聲音重現技術的組合可應對這一項挑戰，以打造出真正的下一代沉浸式音訊系統。

在真實場景中自然接收聲音時，我們的大腦會基於傳到左右耳的音訊訊號來提取有關聲源的空間線索。這與我們的雙眼視覺系統的工

圖 1：來自聲源 $x(t)$ 的自然接收場景， $X_L(t)$ 表示到達左耳的音訊訊號， $X_R(t)$ 表示到達右耳的音訊訊號。

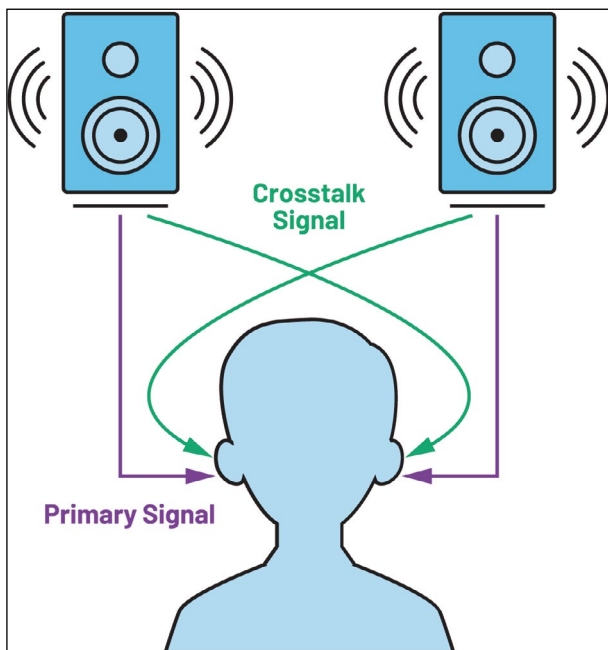


作原理非常相似，大腦也是透過結合左右眼所看到的圖像來感知深度。大腦處理到達左右耳的聲音，透過比較振幅和時間延遲來推算聲源位置。這是人類在進化過程中形成的能力，也是原始社會的關鍵生存技能。

雙耳聲音重現技術旨在透過新型訊號處理，在左右耳生成與真實場景相同的左右音訊訊號，再現聲音的自然體驗 (圖 1)。但在實踐中實現這一目標並非易事，會面臨重重問題。

記錄雙耳音訊的一種簡單方法是在真實環境中人的左右耳各部署一個麥克風，然後記錄到達每只耳朵的聲音訊號，這種方法稱為雙耳記錄。然後透過耳機重現聲音，傳至聽眾的耳朵。那這種方式效果如何呢？在針對同一位聽眾進行捕捉和重播時確實有效，但由於每個人大腦定位聲音的方式不同，這一項技術並非完全靈驗。我們的頭 / 耳廓 / 身體對聲音的影響會在頻域中留下特定的特徵，協助我們的大腦定位聲音。這種特徵因人而異，被稱為頭部相關傳遞函數 (HRTF)。如想讓雙耳技術真正有效，須在聲音重現過程中聽眾的耳朵上準確再

圖 2: 身歷聲揚聲器中的串擾效應。



現 HRTF 對聲音的影響。

因此，我們需要針對每位聽眾測量並客製化 HRTF，通用的解決方案並不可行。研究顯示，當人們體驗用其他人的 HRTF 製作的音訊時，其在體驗期間的聲音定位能力會顯著降低。^{1,2,3}

在揚聲器上實現雙耳音訊還會面臨更大的挑戰。首先，來自多個揚聲器的聲音訊號會相互干擾，即所謂的串擾效應 (圖 2)。其次是聽音環境，在聲音到達聽眾耳朵之前，它不可避免地會對聲音產生一些影響。

在實現真實模擬自然聲音接收體驗時，揚聲器串擾、HRTF 個性化需求以及房間 / 聽音環境的影響是主要的阻礙因素。而視覺系統能夠捕捉到有關聽眾和聽音環境的所有細節，有助於解決雙耳聲音重現所面臨的挑戰。

例如，可以建構為電腦視覺演算法提供資料的攝影機來捕捉聲音接收環境的三維架構資訊 (即房間形狀、不同表面的幾何測量細節以及存在的物體)，用於計算聽音環境對聲音的影響。然後，可以在聲音重現系統中適當增加濾波器和濾波器係數以消除不良影響。雖然家庭影院音訊已採用此類系統，但它通常依賴於在校準期間使用麥克風捕捉房間對聲音的影響，如果接收位置或房間結構發生變化，則需要重新進行校準。

視覺系統可以進一步捕捉人體測量資料，例如身體位置和結構細節⁴，透過必要的運算將 HRTF 個性化，以呈現準確的空間線索 (圖 3)。使用聽眾頭部位置相對於揚聲器的資訊和頭部尺寸，部署串擾消除演算法，在揚聲器中呈現即時雙耳音訊，聽眾能夠隨意移動，同時保持理想的聲音體驗 (圖 4)。

使用音訊系統存在隱私問題，但從視覺系統獲取的攝影機資料會經過即時處理，無需儲存或傳輸到另一台遠端機器，因此，使用專用

圖 3: 透過人體測量實現 HRTF 個性化。

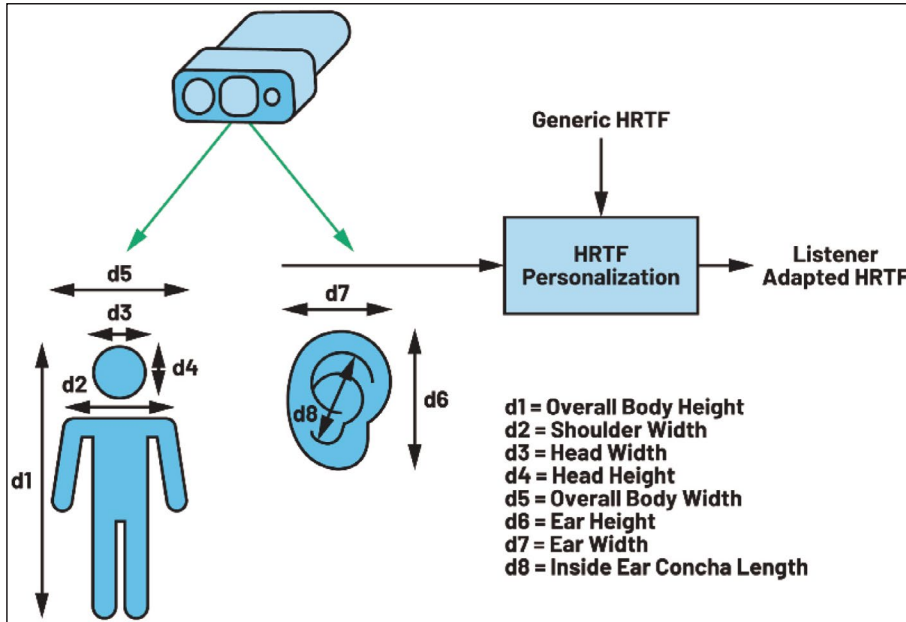
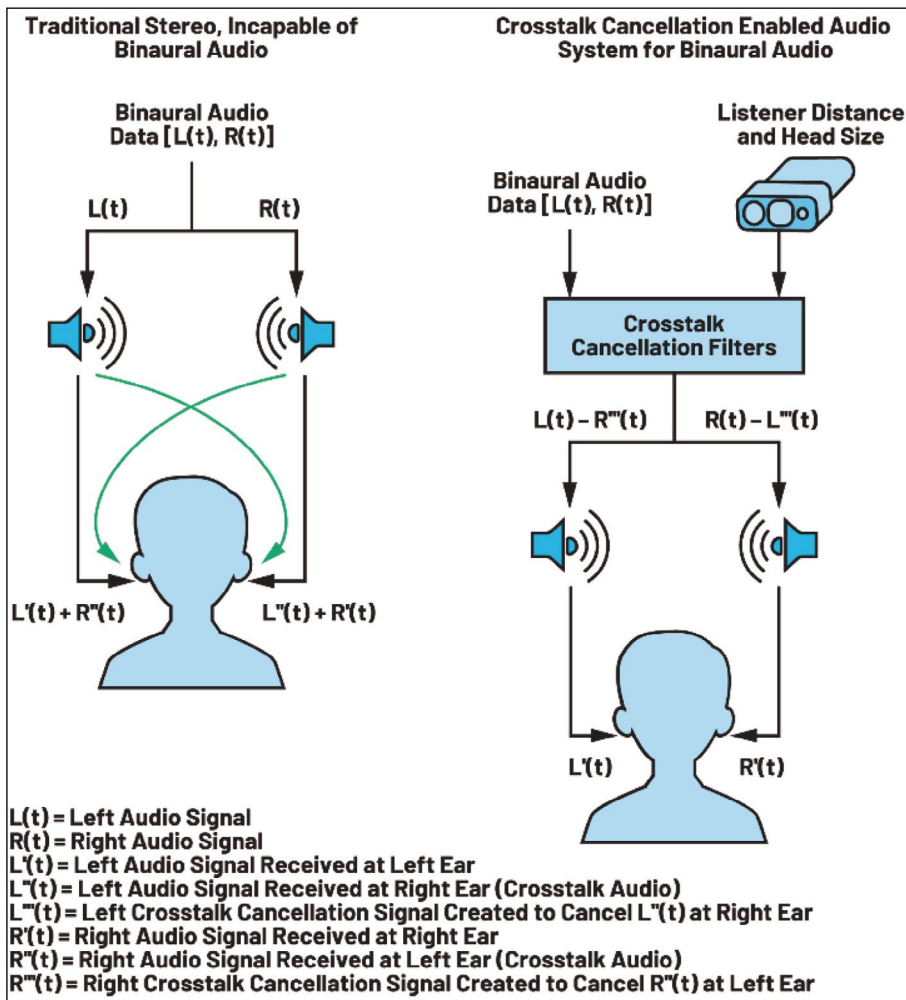


圖 4: 採用串擾消除技術，透過自由場揚聲器系統實現雙耳聲音再現。



的計算處理器在邊緣處理視覺資料分析可以保護使用者隱私。

ADI 最新的多核心 SHARC DSP 和先進的 ToF 成像器提供了硬體平台實現音視頻融合所需的關鍵部件，以創建下一代沉浸式音訊系統 (圖 5)。

ADI 的 ADSP-SC598 SOC 搭載 SHARC 雙核心和一個 A55 Arm 內核心，由大型晶片內記憶體和外部記憶體 DDR 介面提供支援，可滿足低延遲和記憶體密集型運算要求，是實現真正沉浸式音訊的理想平台 (圖 6)。SHARC DSP 上的運算資源，例如 ADSP-SC598，可以將與音訊解碼相關的工作負載劃分至 DSP 內核心上，在第二個 SHARC 核心上實現音訊重播的後處理和個性化。Arm A55 可用於進行多種控制處理。⁶ 圖 5 所示的視覺系統可以組合使用 RGB 和深度攝影機或單獨使用深度攝影機。ADI 的高解析度 100 萬像素 ToF 深度成像器 ADSD3100 可以捕捉毫米解析度級別的深度圖，且能在不同的照明條件下工作，為之前所述的個性化

圖 5: 下一代沉浸式音訊系統。

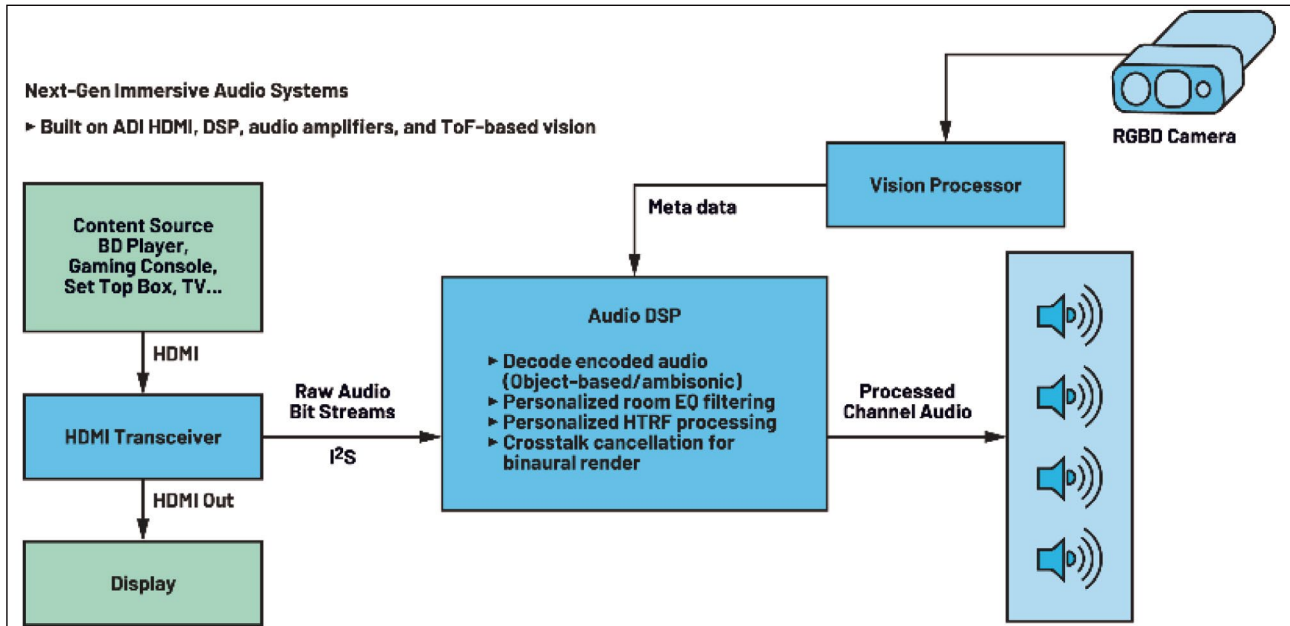
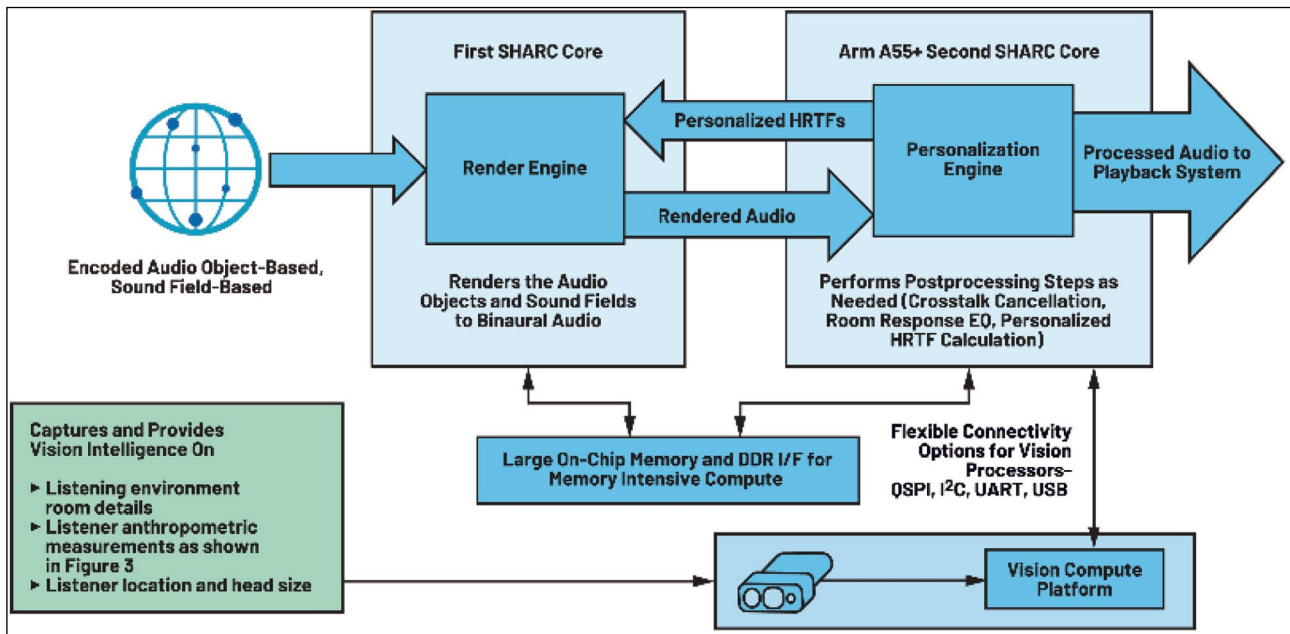


圖 6: 下一代沉浸式音訊系統的系统磁碟分割。



演算法 (串擾消除、房室均衡、HRTF 人性化等) 提供了所需的高精度幾何測量資料。

ADTF3175 是基於 ADSD3100 ToF 深度成像器的 100 萬像素、75 × 75 度視場 (FOV) ToF 模組，它還整合了用於成像器的透鏡和光學帶通濾波器、紅外光源 (包含光學元件、鐳射二極體、鐳射二極體驅動器和光電探測器)、

快閃記憶體和功率調節器以產生本地電源電壓。該模組在多個範圍和解析度模式下進行完全校準。如需完善深度測量系統，可以將來自 ADTF3175 的原始圖像資料透過主機系統處理器或深度 ISP 進行外部處理。ADTF3175 圖像資料輸出介面透過 4 通道行動產業處理器介面 (MIPI)、攝影機序列介面 2 (CSI-2) 變送器介面

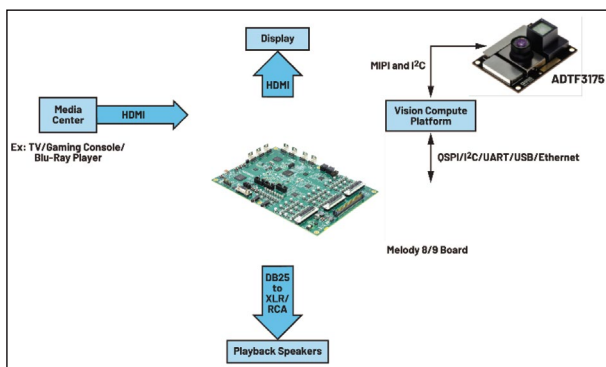
與主機系統進行電氣介面。該模組編程和操作是透過 4 線式 SPI 和 I²C 序列介面進行控制。

ADI 目前提供的 EVAL-MELODY-8/9 開發平台板、EV-2159X/SC59x-EZKIT 板和 CrossCore Embedded Studio(一款基於 eclipse 的編輯工具)可以協助使用者部署和運行 ADI 的 ADSP SOC，以即時部署和調試應用。⁷

Melody 平台是 ADI 為 AVR 和條形音箱應用提供的完整訊號鏈解決方案。其將視訊、DSP、音訊、電源和軟體方面的一流 ADI 元件結合到組合系統解決方案中，使客戶能夠利用最新技術快速進入市場，以達到其年度升級視窗。⁸

ToF 模組 ADTF3175 可以連接至視覺運算平台，並連接至 Melody 板，為下一代沉浸式音訊系統構建硬體平台(圖 7)。RGB 攝影機並可耦合至 ADTF3175 ToF 模組，透過建構 RGBD 攝影機來進行強化視覺分析。

圖 7: 使用 ADI 平台實現沉浸式音訊系統。



結論

借助包含 DSP、HDMI 收發器、D 類放大器和 ToF 成像器的解決方案系列，ADI 持續不懈地追求實現真正沉浸式的音訊系統，致力為使用者提供與真實世界一般無二的聲音。

參考資訊：

- ¹ Philipp Paukner, Martin Rothbucher, and Klaus Diepold. "Sound Localization Performance Comparison of Different HRTF-Individualization Methods." Technische Universität München, April 2014.
- ² Parham Mokhtari, Ryouichi Nishimura, and Hironori Takemoto. "Toward HRTF Personalization: An Auditory-Perceptual Evaluation of Simulated and Measured HRTFs." International Conference on Auditory Display (ICAD), July 2008.
- ³ Jenny Claudia and Christoph Reuter. "Usability of Individualized Head-Related Transfer Functions in Virtual Reality: Empirical Study With Perceptual Attributes in Sagittal Plane Sound Localization." JMIR Serious Games, September 2020.
- ⁴ Geon Woo Lee and Hong Kook Kim. "Personalized HRTF Modeling Based on Deep Neural Network Using Anthropometric Measurements and Images of the Ear." Applied Sciences, November 2018.
- ⁵ Sanket Nayak and Mitesh Moonat. "EE-436: Using ADSP-SC59x/2159x High Performance FIR/IIR Accelerators." Analog Devices Inc., June 2022.
- ⁶ "ADSP-SC59x/2159x SHARC Series Doubles Portfolio Performance, Enabling Platforms for Complex Audio Applications." Analog Devices, Inc.
- ⁷ "CrossCore Embedded Studio." Analog Devices, Inc.
- ⁸ "Home Theater and Gaming." Analog Devices, Inc.