音訊技術創新八方 “傾聽”

■文：編輯部

作為人體五感之一的聽覺系統，是我們用來接受外界聲音信號的特定裝備。所謂耳聽八方，就是人們接收一切可以聽到的聲音，從而瞭解周圍的一切，瞭解整個世界，並使得進一步與之交互成為可能。

物理學的深入研究揭示了聲音一種機械波的諸多特性，大家利用這些特性，使得音訊相關技術在多個領域取得了顯著的創新與突破。特別是 AI 技術的高速發展，具備智慧化的音訊技術不僅改變了我們通過聲音交互的方式，還在醫療、娛樂、安防等領域創造了許多新價值。

在 AI 這一具有里程碑意義的技術幫助下，聲音技術也迎來了新一波的創新。我們在用耳朵感知世界的同時，智慧的音訊技術也在默默“傾聽”著我們……

語音辨識與合成

更自然的語音助手：基於深度學習的語音合成技術，如 WaveNet、Tacotron、Transformer TTS 等，能夠生成接近人類的聲音，甚至可以模仿特定人的音色。比如微軟的 VALL-E 技術能夠通過少量語音樣本合成特定人物的聲音，廣泛應用于智慧客服和虛擬助手等領域。這些技術的應用使得語音助手更加自然、更加智慧，能夠更好地滿足用戶的需求。

多語言與即時翻譯：即時語音翻譯工具，這類工具在跨國旅遊的遊客中深受歡迎。比如 Google Translate 的對話模式，支援多種語言的即時交流。Meta 的 “Universal Speech Translator” 項目更是直接翻譯未訓練過的語言，極大地促進了跨語言溝通。這些技術的應用使得語言不再是交流的障礙，為更加廣泛地全球化交流提供了支援。

情感語音合成：AI 語音技術的發展，使得 AI 能夠生成帶有情感(如喜悅、悲傷)的語音，用於客服、有聲書或虛擬角色(如遊戲 NPC)。例如，一些虛擬客服系統通過情感語音合成技術，能夠根據使用者的情緒回饋調整語音語調，提升使用者體驗。這些技術的應用使得語音交互更加人性化、更加自

然。

音樂生成與創作

AI 已經開始參與作曲與編曲工作：Google 的 MusicLM、Stable Audio、OpenAI 的 Jukedeck(已關閉) 等工具能夠根據文本描述生成音樂片段，甚至模仿特定風格 (古典、電子、流行)。比如，MusicLM 可以根據使用者輸入的文本描述生成符合特定風格和情感的音樂。這些技術的應用使得音樂創作更加便捷、更加多樣化。

優化音訊品質：通過深度學習修復老舊錄音 (如披頭士歌曲的 AI 修復)，或從低品質音訊中還原清晰語音。例如，一些歷史錄音通過 AI 修復技術恢復了原始的音質。這些技術的應用使得音訊修復更加精准、更加高效。如 LANDR 公司利用 AI 自動優化音訊品質，iZotope 的外掛程式可智慧修復錄音缺陷。而且，這些技術能夠自動調整音訊的動態範圍、頻率響應等參數，提升音樂製作的效率。這類技術的應用使得音樂製作更加高效、更加專業。

互動式音樂體驗：在遊戲和 VR 中，AI 能夠根據使用者行為即時生成動態背景音樂。很多遊戲的環境音效能夠根據玩家的行為和環境變化即時調整，增強遊戲的沉浸感。這些技術的應用使得玩家更加身臨其境。

音訊處理與增強

降噪與語音增強：諸如 Krisp、nVIDIA RTX Voice 通過 AI 即時消除背景雜音 (鍵盤聲、風聲)，提升通話品質。這些技術的應用使得音訊通信更加清晰、更加可靠。在遠端辦公和線上教育中降噪和語音爭搶技術得到了廣泛應用。

聽力輔助技術：對於需要助聽器的使用者來說，並不是簡單的加大音量就能夠讓他們聽到聲音，很多病人只是在聽覺範圍的某些頻段無法感知到音訊信號，因此助聽器需要根據每個病人的情況單獨調整音訊信號，傳統方法製造這樣的助聽器不僅昂貴，而且無法隨著聽覺系統的病變進行調整。AI 演算法正在說明人們提升人工耳蝸的聲音解析能力，說明使用者更好地理解語言。這些技術能夠根據使用者的聽力特徵動態調整聲音處理參數，區分語音和噪音，並動態優化聲音場景，說明聽障人士更好地理解語言。這類技術的應用使得聽力輔助更加智慧、更加有效，且價格更加便宜，能夠惠及更多聽障人士。

聽覺健康與醫療

耳鳴治療：在醫療應用中，AI 分析患者腦電波，生成個性化聲波對抗耳鳴。例如，Neuromod Devices 的 Lenire

系統通過 AI 技術生成特定的聲波，幫助患者緩解耳鳴症狀。這些技術的應用使得耳鳴治療更加個性化、更加有效。

早期疾病診斷：通過分析咳嗽聲、呼吸聲檢測疾病，如前兩年 COVID-19 大流行期間，醫院透過簡單的聲音分析技術進行早期病例篩查，這項技術同樣對哮喘病例識別也有作用，現在很多醫院已經開始通過語音變化識別帕金森症。例如，一些研究通過分析患者的語音變化來早期識別帕金森症。這些技術的應用使得疾病診斷更加早期、更加準確。

聲音事件檢測與環境感知

智能家居與安防：識別玻璃破碎、嬰兒哭聲等異常聲音並觸發警報 (如 Amazon Alexa Guard)。這些技術能夠即時監測家庭環境中的聲音變化，保障家庭安全。這些技術的應用使得智慧家居更加智慧、更加安全。

生態監測：通過野外錄音識別動物叫聲 (如 Rainforest Connection 用 AI 保護雨林)。例如，Rainforest Connection 通過 AI 技術分析野外錄音，識別動物叫聲，用於生態監測和保護。這些技術的應用使得生態監測更加精准、更加高效。

工業檢測：分析機器運行聲音預測故障 (如軸承異響檢測)。這種預測性維護的技術已將廣泛使用在製造業環境中，借助對機器聲音的監聽，人們能夠即時監測機器的運行狀態，提前預測故障，減少停機時間。這些技術的應用使得工業檢測更加智慧、更加高效。

例如，西門子公司的研究人員借助 NVH 模擬器創建關鍵零件的頻率響應函數，故障監測提供依據。

虛擬實境 (VR) 與空間音訊

空間音訊通過多聲道技術模擬三維聲場，使聲音具有方向感和空間感。AI 類比頭部相關傳輸函數 (HRTF)，為 VR/AR 提供沉浸式空間音訊 (如 Meta 的 Project Cambria)。這一技術能夠根據使用者的頭部位置和方向動態調整音效，增強虛擬環境的真實感。這些技術的應用使得虛擬實境體驗更加沉浸、更加真實。

圖說：西門子 Simcenter Testlab 軟體幫助設計人員建立頻率響應函數

圖片來源：siemens.com

圖說：Ceva-RealSpace Elevate 多聲道空間音訊解決方案

圖片來源：ceva.com

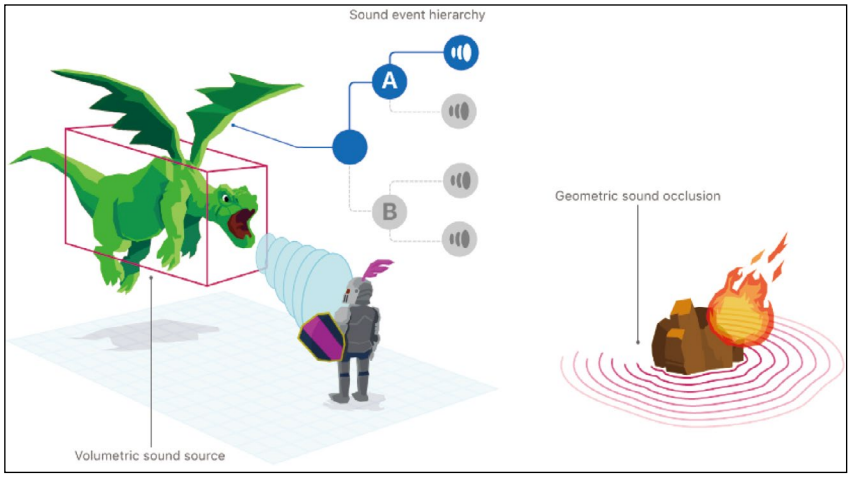
在虛擬環境中動態類比聲音傳播 (如反射、混響)，增強真實感。例如，一些 VR 平臺通過即時聲場建模技術，能夠根據虛擬環境的幾何結構動態調整聲音傳播。這些技術的應用使得虛擬環境更加真實、更加沉浸。

隨著消費者對高品質音訊體驗需求的增加，空間音訊技術在遊戲、電影、音樂等娛樂領域應用廣泛。例如，Ceva-RealSpace Elevate 多聲道空間音訊解決方案，能夠為安卓和視窗設備提供沉浸式體驗，支援從單聲道到多聲道和基於物件的音訊，還具備精確的頭部追蹤功能，當聽者移動頭部時，聲場保持固定，仿佛聲音來自周圍空間。這些技術的應用使得音訊體驗更加沉浸、更加真實。

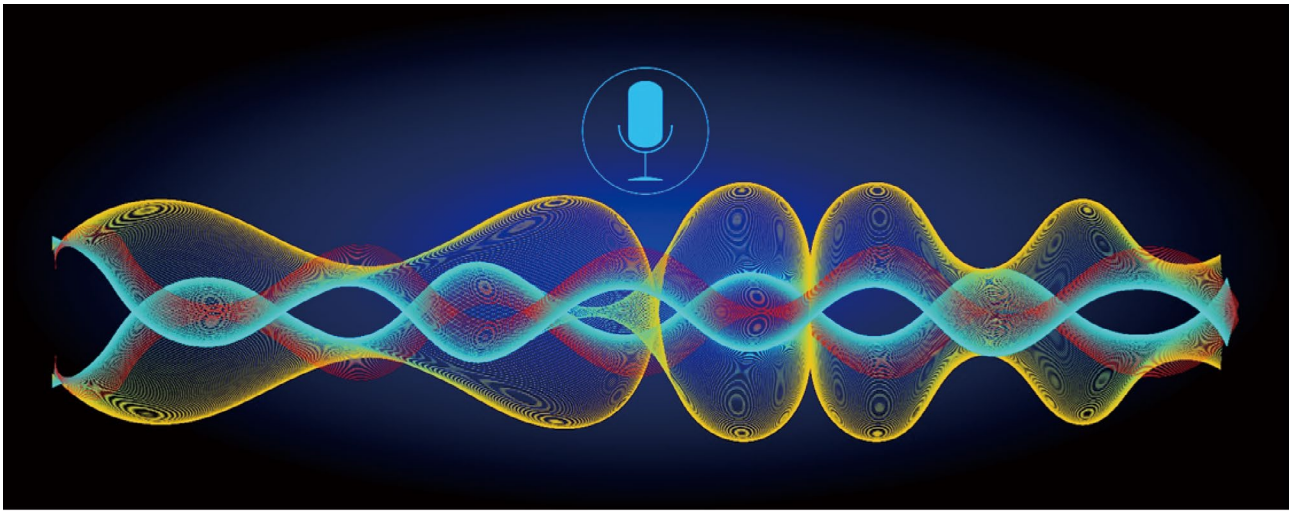
該項解解決方案是 Ceva 與聯發科技 (MediaTek) 兩家公司合作，將 Ceva-RealSpace Elevate 多聲道空間音訊與頭部追蹤解決方案，結合採用藍牙 LE 音訊的聯發科技天璣 (Dimensity) 9400 旗艦 5G 智慧手機晶片，為真無線立體聲 (TWS) 和藍牙 LE 音訊耳機提供身臨其境的音效體驗。用戶可以體驗到與 4K 串流顯示和遊戲的視覺清晰度相匹配的高品質音訊，同時受益於更好的效能、增強的人工智慧和更長的電池續航時間。

ARM、Google 和三星聯手打造的 Eclipsa Audio 是一種多頻道音訊環繞格式，利用 IAMF 技術提供沉浸式聆聽體驗，支援多達 28 個輸入通道，可渲染到一組輸出揚聲器或耳機，具備雙耳渲染功能，適用於多種消費電子產品。這些產品使得空間音訊體驗更加便捷、更加優質。

圖說：蘋果公司的空間音訊技術示意



圖片來源：developer.apple.com



音訊內容分析與理解

自動摘要與標籤化：AI 分析播客或會議錄音，生成文字

摘要和關鍵字標籤 (如 Otter.ai、Descript)。這些技術能夠快速提取音訊內容的核心資訊，方便使用者快速流覽。這些技術的應用使得音訊內容處理更加高效、更加便捷。

情感與意圖識別：通過語音語調分析使用者情緒，應用於客服質檢或心理健康評估。例如，Cogito 公司通過分析客服通話中的語音語調，識別客戶的情緒狀態，提升服務品質。

版權檢測與內容審核：識別音訊中的版權音樂 (如 Shazam)、敏感內容或虛假資訊 (如 Deepfake 音訊檢測)。這些技術能夠有效保護版權和防止虛假資訊傳播。這類技術使得音訊內容管理更加規範、更加安全。

腦機介面與神經音訊

通過解碼大腦信號直接合成語音，說明失語者溝通 (如 Neuralink 的研究)。這些技術為失語者提供了一種新的溝通方式。這些技術的應用使得溝通更加便捷、更加自然。

此外，借助腦機介面，研究人員正在利用特定聲波刺激大腦，改善睡眠 (如 Philips 的 SmartSleep) 或注意力。例如，Philips 的 SmartSleep 通過特定的聲波刺激，嘗試幫助用戶改善睡眠品質。

自動駕駛中的聽覺感知

為了使自動駕駛更加安全、可靠。車輛通過 AI 識別警笛聲、行人腳步聲或輪胎異常噪音，提升安全性 (如特斯拉的 FSD 系統)。這些技術能夠即時監測車輛周圍的環境聲音，配合其他的車輛感知系統，加強行車安全。

與人工智慧結合的多模態音訊技術

多模態音訊技術結合了音訊、視覺、文本等多種資訊，通過深度學習和 AI 演算法，實現更智慧的音訊處理和交互。例如，Arm 在推動多模態大模型的發展，使機器人等設備能夠通過多種感官模式感知環境，進行分析、推理和決策。這些技術的應用使得音訊交互更加智慧、更加自然。

隨著 AI 技術的快速發展，多模態音訊技術在智慧型機器人、自動駕駛、智慧家居等領域具有廣闊的應用前景。例如，Boston Dynamics 的機器狗 Spot 可以在博物館裡當導遊，與參觀者互動，展示多模態音訊技術在實際應用中的潛力。這些技術的應用使得智慧設備更加智慧、更加便捷。

目前市場上尚未有成熟的多模態音訊產品，但相關技術

正在不斷研發和探索中。例如，谷歌的 Gemini 模型和微軟的 LLaVA 等，都在嘗試將多模態技術應用於音訊和語音處理。這些技術的研發將為多模態音訊技術的發展提供有力支援。

小結

未來音訊技術將更多地結合視覺、語音和文本的跨模態理解。例如，視頻自動生成字幕、通過圖像識別輔助音訊降噪等。這種多模態融合將為用戶提供更加豐富和自然的交互體驗。這些趨勢的發展將使得音訊技術更加智慧、更加便捷。

CTA