

雙向開講， 「語音互聯網」成形

■文：任苙萍

互聯網正從指尖走向舌尖 (語音)。Juniper Research 最新報告指出，2024 年語音助理設備的數量將超過地球人口總數，達 84 億個，行動設備使用量增加是主要推力。語音控制亦是重要的人機介面 (HMI)，包括：手機、智能穿戴裝置、虛擬個人助理 (VPA)、智能揚聲器、智慧家電，乃至汽車、飛機等交通載具都是應用範疇。日前 2021 CES (國際消費電子展) 亦不乏相關產品亮相，例如：樂金 (LG) InstaView 聲控冰箱，以及可與 LG ThinQ、亞馬遜 (Amazon) Alexa 或谷歌 (Google) Assistant 等智能產品結合使用的進化版 Magic Remote 智慧遙控器。

比觸控直觀&安全！聲控躋身 HMI 主流

值得注意的是，基於音訊的健康解決方案與生物辨識驗證正在崛起，CES 可見多款智能耳機、助聽器大舉壓境，而音訊內容的有效性、高傳真度及個性化配置是關鍵指標。當聲控躋身 HMI 主流，「語音互聯網」隱然成形，亞馬遜和蘋果 (Apple) 更搶先投入分析語氣、

圖 1：德國 Continental 公司 (大陸集團) 的智能和自適應駕駛員輔助系統可實現自然對話，在一句話中理解多個問題並檢測邏輯連接



資料來源：<https://www.continental.com/en/press/press-releases/2019-07-18-smart-voice-assistant-178956>

判讀情感的研發；臉書 (Facebook) 亦不斷深耕語音助理，冀與亞馬遜和蘋果等先進者互別苗頭。與此同時，有人主張語音控制是相對安全的車用 HMI，避免因觸控選項而分心。研調公司 Research Dive 表示，組合式車載語音有助於維護駕駛安全。

文本語音轉換技術和語音辨識 (Speech Recognition) 讓駕駛員在操控資訊娛樂 (infotainment) / 導航系統或使用手機時依然可平視顯

示器，以確保駕駛員將注意力放在道路上。另有研究指出，未來將有 80% 汽車的 HMI 將集成語音辨識系統 (不包括 Apple Siri 等手機應用程式)，主要用於控制媒體播放器、設置導航目的地和連接智慧手機撥打電話。Research Dive 預估到 2026 年，汽車語音命令系統市值將達 49.9 億美元，谷歌、蘋果和紐安斯通訊 (Nuance Communications) 正在改變車載語音。

Apple CarPlay 附加在車輛觸

控螢幕的安全 iOS 版本，按下方向盤語音按鈕，Siri 可在導航、播放列表、發送簡訊和電子郵件之間切換；未來還可 Skype 打電話、WhatsApp 發訊息或用 Apple Maps 預訂晚餐。Google Android 用戶可透過 Android Auto 連接 USB 電纜，在車輛儀表板使用手機螢幕、駕駛中靠藍牙通話；手機插入汽車後，Android Auto 將暫時讓手機失能而強迫音控。Nuance 結合賽倫斯 (Cerence) 公司的 Dragon Drive 汽車助理和雲端平台，擁有自然語言模式傳播與語音生物辨識能力(註：Cerence 為 Nuance 分拆公司，專責汽車業務)。

機器學習引領，語音辨識大躍進

Nuance + Cerence Dragon Drive 可辨識當前說話者，甚至在用戶提出要求前就了解其道路行為並根據需求進行更新，還可與 Google Nest 匹配以學習自動調節房屋的最佳溫度設置。受惠於機器學習 (ML) 演算法的日新月異，語音辨識還能控制先進駕駛輔助系統 (ADAS) 和車內環境，雜訊 (噪聲) 消除和個性化音訊亦越發受到關注。蔚來汽車 (NIO)、梅賽德斯-賓士 (Mercedes-Benz) 和飛亞特汽車 (FIAT) 是產業先行者。2017 年，蔚來發佈首款旗艦 SUV ES8 所搭載的「NOMI」(取自英文 Know Me 諧音)，被視為全球首個「量產」的車載人工智慧 (AI) 系統。

第一代 NOMI Mate 初試啼聲

圖 2：蔚來 NOMI 集成語音交互系統和智能情感引擎，訴求「有禮貌、會傾聽、會思考、知冷暖」，創造一種全新的人車交互方式



資料來源：<https://www.nio.cn/nomi>

後，2018 年第二代 NOMI Halo 採用光量 + 聲音的簡潔形式實現語音互動，包括：播放歌曲、設置空調、語音搜索、模糊語義和亮度控制等。雖然一度瀕臨破產危機，蔚來日前所推出的首款自駕車 ET7，仍擄獲不少關注目光——憑藉四個麥克風陣列和專用的網路處理器

(NPU) 內核，NOMI 具有語音識別和準確的聲音本地化功能，以實現交互而不中斷。梅賽德斯第二代用戶體驗虛擬助理 MBUX2，沿用 Cerence 對話式 AI 平台控制汽車環境、娛樂和其他系統，支援多國自然語言理解和間接命令。

例如，當用戶說「很冷」，

圖 3：只要一聲「Hey Mercedes」，從導航、多媒體選項到座椅位置和理想溫度皆可語音辨識，做個性化設置



資料來源：<http://cerence.com/news-releases/news-release-details/cerence-powers-voice-and-ai-driven-features-next-generation>

會自動提升駕駛艙內的溫度。驅動程式可使用其他命令繼續完成對話框，一次給出多個指引並依序執行，包括觸控手寫；MBUX2 還借助 Cerence 語音生物辨識注入個性化元素以迎合不同駕駛員喜好。汽車製造商可在 Cerence Studio 平台自定義語音域，為駕駛員創造獨特品牌體驗，甚至擴展到汽車之外——車主可在開車時使用 MBUX 連接、控制智能家居設備。FIAT 亦選擇 Cerence 在第首款純電動車 FIAT 500 提供對話式 AI，只要聽到關鍵詞，就能語音連接與 TomTom 合作開發的功能，不必學習特定語言。

包括：導航、媒體娛樂、電話、簡訊、溫度控制並訪問 Amazon Alexa。先進的文本語音轉換功能使 New 500 可自然與駕駛員和乘客交流，以回答問題和完成請求。索尼 (Sony) 今年也在 CES 展示了為其 Vision S 概念車開發的語音識

別系統，可配合儀表板上的攝影機讀取唇形、臉部表情和手勢動作，以增強清晰度和駕駛員語音識別能力，幾乎可在任何環境或車款使用（包括敞篷車）。整合麥克風、放大器、揚聲器和先進數位訊號處理等技術，有助於降低背景噪音、使乘客之間的語音溝通更清晰，或在緊急狀況時能免持語音撥號。

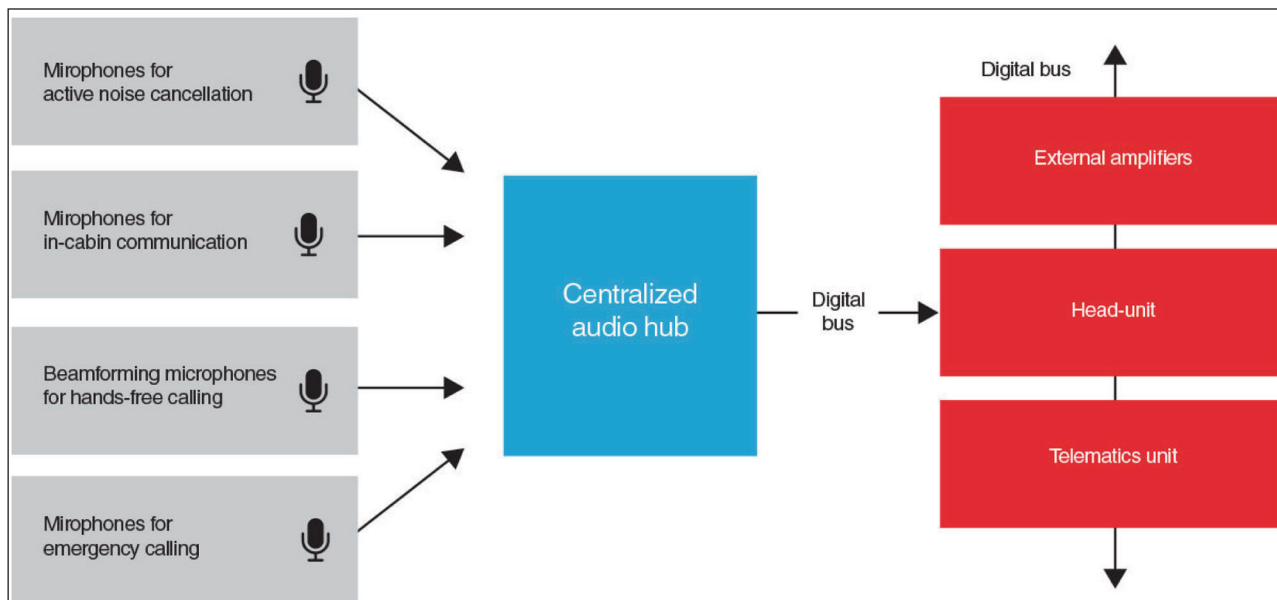
車載語音四大趨勢現蹤，維護飛安也用得著！

德州儀器 (TI) 談到車載語音有四大趨勢。一是主動降噪 (ANC) 系統。相較傳統降噪技術會讓車輛笨重且降低燃料效率，ANC 可顯著減輕車體重量，且由於跟音效播放的揚聲器是同一個，加裝 ANC 系統的新增成本相對較低——在整個車廂內部裝設 2～6 個麥克風以測量內部噪音、將音效資料傳輸至音效子系統，再發出反音效訊號到

內建揚聲器；入門車款約 2～4 個麥克風，高階車種最多有 8 個。二是車內通訊系統。策略性裝設 2～8 個麥克風接收每位乘客的聲音，主動調整並強化每個人發言、降低不必要噪音，再透過車內音效揚聲器系統播送。

三是緊急呼叫 (eCall) 與免持語音系統，通常有 1、2 個麥克風，發生緊急狀況時汽車和在地緊急救援服務單位可直接語音溝通；內含一個連網模組，能將麥克風訊號數位化並傳給報案受理者，對方回覆內容則會透過車內專用揚聲器播放。免持語音則有 1～8 個麥克風或一整個陣列的波束成形麥克風，以提供清晰的語音通話和指令功能。最後一個趨勢是：整合前述所有功能的「集中式音效中心」，將語音訊號數位化並傳送到各個音效子系統進一步處理。為此滿足不同層次需求，TI 已開發出 PCM6260-Q1 多頻道語音類比數

圖 4：車載集中式音效中心模組



資料來源：TI 提供

位轉換器 (ADC) 系列產品。

另一方面，由歐盟資助的「VOICI」專案意在借助語音辨識和 AI 維護飛航安全——偵聽駕駛艙內機組員之間、以及與空中交通管制的所有通訊，並加以辨識、解釋語音內容，包括了解航空術語的嵌入式座艙語音處理系統及一系列經優化陣列處理的低雜訊光學麥克風；它還具有語音合成功能，適用於航空術語和雜訊水平、應對嘈雜座艙的語音擷取和辨識技術以及智能對話系統。經由飛行員的頭戴式耳機和周圍的麥克風陣列擷取語音，所使用的深度神經網路 (DNN) 和對話系統之語音辨識是專為座艙環境開發的獨立系統，所有演算法已經實現並通過測試。

AI 語音互動走進醫療照護領域

Nuance 的 Dragon 系列另有

圖 5：Oticon More 是首個板載 DNN 的助聽器，可更自然地呈現所有聲音，提供完整且精確平衡的聲音場輕鬆地景，幫助用戶理解語音



資料來源：<https://www.ces.tech/Innovation-Awards/Honorees/2021/Honorees/Oticon-More%E2%84%A2.aspx>

遠程醫療方案——Medical One，全球已有逾 550,000 位醫師使用。去年第四季，Medical One 再獲美國退伍軍人事務部 (VA) 採用。帶有 PowerMic 行動麥克風應用程式與雲端語音辨識平台，讓 VA 醫師可利用電話或 VA Video Connect 平台虛擬問診。早在 2014 年，基於 Nuance 雲端的 Dragon Medical 系統就已打入 VA 標準化系統，可與 VA CPRS 和 Cerner Millennium 應用程式（聯邦政府電子病歷解決方案要員）相容。如今，醫師更將 Dragon Medical One 附加功能和行動靈活性用於遠程醫療服務。

不只醫療診斷，語音辨識也開始走進健康輔具領域。「全球首個具備 AI DeNoise 技術」的助聽器——Orka One，在微型晶片上運行功能強大的 AI 神經網路，在減少背景噪音的同時增強人聲。它還具有藍牙 5.0，可讓使用者微調、自定義聽力配置文件，並與電話和智慧助理集成，輕鬆接聽來電、請求、命令或從 Apple 設備播放音頻。今年在 CES 獲得創新獎殊榮的 Oticon More 助聽器，則利用板載 DNN 進行 1,200 萬種現實生活中的聲音訓練，並身懷應對空間和功率的能力，可提供完整、精確的聲音平衡以實現最佳感知。

另兩個奪下創新獎的助聽器產品，一是 WIDEX MOMENT 天然聲音助聽器，一是 Starkey Livio Edge AI 助聽器。WIDEX MOMENT 結合雙 AI 引擎——SoundSense Learn 以兩種方式

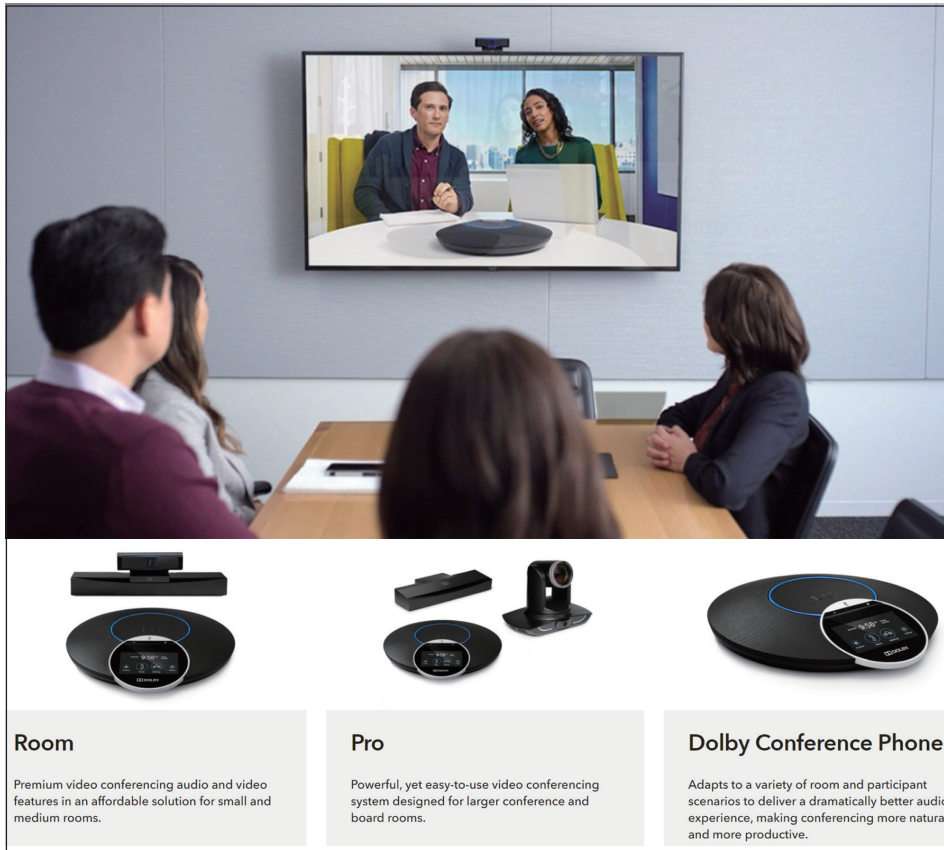
調整助聽器，Widex PureSound ZeroDelay 並行處理路徑將聲音延遲從 7 ~ 10 毫秒降至 0.5 毫秒，改善即時聆聽效果；透過分析設置及一系列 A、B 比較指導用戶了解周圍環境偏好，再利用儲存在雲端的數百萬個用戶設置協助個性化收聽體驗。Starkey Livio Edge AI 則是借助動態應用增益、輸出、噪聲管理、定向麥克風和其他功能，透過 AI 模型最佳化語音的可聽性和音質。

邊緣模式可確保為每種環境中的每個人優化聲音，對配戴口罩的聽障者助益尤大。此外，榮獲「無障礙類別之最佳創新獎」的以色列公司 Voiceitt，開發出一款可讓語言和運動障礙患者用自己聲音交流的應用程式，使用專有的機器學習和自動語音辨識 (ASR) 技術來識別、適應個人獨特的受損語音模式，例如：呼吸暫停和非語音，且可與 iPhone、iPad 和 Amazon Alexa 集成以通訊、控制智能設備。另一方面，遠距工作與線上教學風潮，亦讓智能語音互動的關注度跳升；為改善 PC-based 設備的音訊／視訊通話效果，杜比 (Dolby) 特為此發佈「語音」工具。

防疫帶旺線上會議&遠距教學，各路語音技術爭鋒

杜比語音 (Dolby Voice) 可消除背景噪聲和迴聲，並自動調整安靜或遠離麥克風的聲音電平，以優化麥克風和揚聲器性能；若

圖 6：杜比語音 (Dolby Voice) 可提供沉浸式、自然的會議體驗



資料來源：<https://professional.dolby.com/communications/>

線上會議的應用程式可提供立體聲，就能區分多人聲音、使聲音更清晰自然，聯想 (Lenovo) 最新的 ThinkPad X1 Carbon 和 X1 Yoga 筆記型電腦是最早使用杜比語音技術的系統之一，可利用空間音訊技術提供出色的會議體驗、分離語音並改善語音助理的語音辨識能力。其中，YOGA 筆記型電腦旗艦機還藉 Elevoc Vocplus PC 強化語音和通訊體驗，已通過 Microsoft Teams Certification、Microsoft Cortana 和 Amazon Alexa 測試認證。

Elevoc Vocplus PC 是運行於英特爾 (Intel) GNA (高斯神經加速器) 的第三方 AI 語音增強方案，

利用監督和數據驅動模仿人類聽覺處理機制，並結合 DNN 和訊號處理技術從背景雜訊智能擷取目標音訊，可消除 96% 動態噪音；借助深度學習和麥克風陣列，即使在嘈雜環境亦可大幅提升 PC-based 語音助理的喚醒和辨識率。Sony 則結合新的 MAS-A100 波束成形麥克風與 Edge Analytic 設備，開發 AI 遠程學習方案，支援 Dante 和乙太網供電 (PoE)，安裝容易、接觸點少，且可根據任何要求量身打造教學解決方案。

新型 MAS-A100 吊頂麥克風專為演講和演講環境而設計，透過將波束賦形技術與「智能反饋減少器」功能完美結合，可為語音增強

和錄音提供高級清晰的音頻質量功能。在擷取語音的同時，利用 Sony 的高性能數位訊號處理和獨特演算法來抑制不需要的反饋，無需手持或隨身佩戴的麥克風、也無需管理電池或設備。麥克風具有雙通道輸出，可同時錄製大範圍區域，以捕獲說話者和學生的聲音；捕獲語音後，麥克風的自動增益控制功能會自動調整輸出音量，使其與演示者所在的位置無關，並使演講和演示文稿更易於收聽。

同樣著眼於防疫商機，TDK 的 Chirp CH201 超音波飛時測距 (ToF) 距離感測器結合微機電 (MEMS) 壓電微機械超聲換能器

(PMUT) 和低功耗混合訊號專用晶片上的數位訊號處理器 (DSP)，可在任何光照條件下 (包括全日照以完全黑暗) 提供最遠 5 公尺的「毫米級」精確距離測量，與目標的顏色和光學透明度無關。與 CH101 匹配，CH201 具有可配置視場 (FoV) 和靈活的 DSP，能處理多種超音波訊號處理演算法。Chirp 提供完整感測器應用程式介面 (API) 和參考設計，可與支援藍牙低功耗 (BLE) 的微控制器 (MCU) 集成追蹤方案或保持社交距離。CTA