

語音互動靈敏、通話音質清晰

MEMS 麥克風技壓全場

■文：任苙萍

降低道路和車內噪音(包括人際對話、資訊娛樂應用程式音源和空調通風口等背景雜訊),接收語音命令的麥克風位置、數位訊號處理技術和濾波器至關重要。就系統層級來看,從接收端到音訊編解碼器、軟體演算法、數位訊號處理器(DSP)、音訊放大器和揚聲器(喇叭)的每個環節,都將影響語音互動的結果。受惠於超靈敏的聲感測器和可解釋語音的複雜機器學習演算法,語音識別技術的準確性越發精進。由上述可知,負責接收音訊的麥克風裝置是第一道關卡,它是將聲波轉為電訊號的「換能器」(Transducer)。

對音質、音場的要求越來越高,音訊訊號鏈中所參與的電子元件也變多;為實現輕薄短小的緊湊設計,微機電(MEMS)技術應用漸廣。Market Reports World 平台刊載一份研究預估,2024 年全球 MEMS 麥克風市值將達 6.683 億美元,2020 ~ 2024 年期間年複合成長率(CAGR)為 8%;單是 2021 年就較去年同期增長 7.01%,主要供應商有:瑞聲科技(AAC)、思睿邏輯(Cirrus Logic)、歌爾(Goertek)、英飛凌(Infineon)、樓氏電子(Knowles)、歐姆龍

(OMRON)、羅伯特博世(Robert Bosch)、意法半導體(ST)和東電化(TDK)一應美盛(InvenSense)。

類比 vs. 數位 MEMS 麥克風,齊攻「主動降噪」

坊間 MEMS 麥克風元件可分為類比、數位兩類,最大區別是:類比麥克風結構相對單純、尺寸較小,適用於空間有限的小型手持裝置;數位麥克風則是將類比數位轉換功能從編解碼器移至麥克風,打造從麥克風至處理器的「全數位音訊通道」以降低雜訊,有利於應對訊號干擾較大的電路設計,但整個器件尺寸自然也較大。若麥克風收音僅用於靜默監聽、無需播放,音訊鏈後端不必有編解碼器或音訊轉換器,那麼,數位 MEMS 麥克風顯然是首選;反之,牽涉到音訊播放或語音交互,類比 MEMS 則當仁不讓。

就整個 MEMS 麥克風出貨市況來看,市調顧問公司 Omdia 在去年底發佈的報告指出,英飛凌已躍居 MEMS 麥克風晶片市場龍頭,囊括 43.5% 市佔率,且領先第二名將近 4%、更遠超第三名 37% 以上。日前,英飛凌再推出新一代類

圖 1: IM73A135 麥克風具備 73 dB 的 SNR 和較高的 AOP (135 dB SPL), 功耗僅有 170µA



資料來源: 英飛凌提供

比式 MEMS 麥克風——XENSIV IM73A135,讓設計人員再也不必在高訊號雜訊比(SNR,信噪比)、精巧封裝、高聲學過載點(AOP)、低功耗,或 MEMS 與駐極體電容式麥克風(ECM)麥克風之間做取捨。IM73A135 可達到 ECM 獨有高水準的音頻效能,兼具 MEMS 技術固有優勢,可增強耳機主動降噪(ANC)功能。

IM73A135 麥克風具備 73 dB 的 SNR 和較高的 AOP (135 dB SPL),擁有高動態範圍和緊密的頻率曲線匹配,可實現最有效的音頻訊號處理並達到業界最低的 170µA 功耗,且體積小巧。此外,其低自有雜訊,特別適用於會議系統、攝影機或錄音機所需的高品質音頻擷取。XENSIV MEMS IM73A135 將於 2021 年 3 月進入經銷市場,適用穿戴裝置的全新 MEMS 麥克風技術將由「Infineon-inside」合作

夥伴隨後推出。事實上，類比式 MEMS 麥克風輸出阻抗相對較低，搭配數位式 MEMS 麥克風輸出，更適合電氣雜訊充斥的環境。

在類比 ANC 揚聲器驅動晶片穩定出貨逾十年的艾邁斯半導體 (ams)，雖已將 MEMS 麥克風專用晶片 (ASIC) 業務售予樓氏，仍致力耕耘數位 ANC 技術——其 AS3460「數位音訊放大器」協助全球首款雙晶片降噪爆款耳機「Padmate PaMu Quiet」協調管理 ANC 和聲音通透性，可放大語音、自動無縫執行預設交叉淡入、淡出，實現 40dB ANC 噪音消除和音訊播放 (未來上看 50dB+)，與藍牙 (Bluetooth) 搭配使用無須犧牲音訊或降噪效能。ams 強化聽覺引擎平台可為入耳式/半入耳式耳機導入自適應洩漏補償 (ALC)、自動預設選擇 (APS) 等功能。

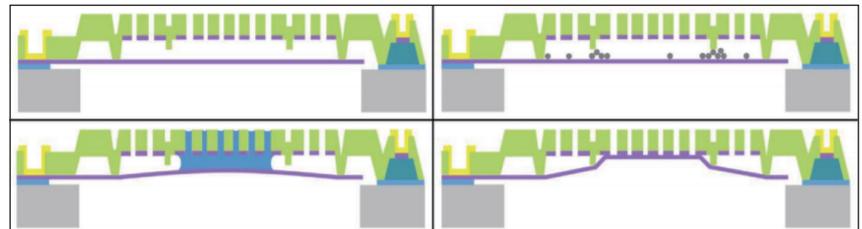
TDK InvenSense 在今年 CES (國際消費電子展) 推出一款 PDM (脈衝密度調變) 數位 MEMS 麥克風 T5838，是其 SmartSound 產品系列最新技術突破，具有 68dB SNR、133dB AOP，號稱業界功率最低——130 μ A 低功耗模式，330 μ A 高質量模式，新功能「聲學活動檢測」(AAD) 支援 20 μ A 的可配置的聲音喚醒功能，以因應常時啟動 (Always-On) 所需。TDK 表示，聲學活動檢測是一項新的超低功耗邊緣處理功能，以監視聲學環境並在檢測到活動時喚醒系統單晶片 (SoC) 或應用處理器 (AP)；T5838 還提供用戶可編程性，為不同應用優化性能。

後起之秀：「壓電式」MEMS，耐用、啟動快且可自供電

整體而言，MEMS 麥克風由於較傳統 ECM 擁有更好的音質、靈敏度和定向性，便於從數據獲取上、下文進行有效處理和渲染，且功耗、成本更低，體積也小巧得多。ECM 預先存在表面電荷，電容器可保持永久電壓以減少所需功率；然而當收音設備越來越精巧，SNR 和穩定性卻隨之下降，溫差大時尤其嚴重。於是，可在麥克風內部將類比轉換為數位訊號、較不受雜訊影響、設計自由度高且易於製造的數位 MEMS 麥克風日受歡迎。不過，為進一步增強抵抗粉塵/微粒、雨水/汗水、震動及高分貝噪音能力，「壓電式」(Piezoelectric) 麥克風成後起之秀。

「壓電式」MEMS 麥克風採用單層薄膜設計的壓電器件不會在背板間隙夾雜灰塵、空氣或雨水等污染物，影響靈敏度和頻率響應，亦避免膜片黏附於背板上。簡言之，它具備防水、防塵、超高聲學過載點、消除迴聲特性，耐用、啟動快且可自供電，對於細微聲響的收音效果更好，更適用於背景雜

圖 2：電容式 MEMS 麥克風易受粉塵/微粒 (右上)、水份 (左下) 污染，或使膜片黏附於背板 (右下)



資料來源：<https://vespermems.com/wp-content/uploads/2017/10/Vesper-Microphone-Arrays-whitepaper-1.pdf>

音大的空間、長距離、戶外或低頻對話，應用面廣泛，但薄膜器件製造不易是其缺點。可喜的是，美國 Vesper Technologies 借助石墨烯製程已有所突破，且已獲亞馬遜 (Amazon) Alexa 認可。

2018 年，Vesper 宣佈與 Unisem 合作開發壓電 MEMS 麥克風，次年即推出專為物聯網 (IoT) 設備設計的 VM3000 數位麥克風。Vesper 一語道破壓電和電容 MEMS 麥克風結構的主要區別在於：壓電 MEMS 不需背板、不存在背板和氣隙，將為 MEMS 壓電式麥克風提供可靠性和強固性。再者，壓電式的抑制超聲波頻率能力亦較優：MEMS 電容式麥克風的訊號是與移動的振動膜和固定背板之間間隙變化成比例，但壓電式訊號是在移動的振動膜中生成，利用壓電效應將聲波能量直接轉為電訊號，SNR 出色且可穩定匹配，非常適合大型陣列應用。

Vesper 先馳得點，VM3011 將「音場背景」納入考量

反觀在電容式 MEMS 麥克風，振動膜片的剛度變化會顯著

牽制麥克風的 SNR，無法實現更高的 AOP；但壓電式感測器在高聲壓級之前是高度線性（極限聲壓約 170 dB SPL），故這些麥克風上的 AOP 僅受 ASIC 電壓軌的限制。這種線性為壓電 MEMS 麥克風提供較高的抗靜電能力。當聲壓進入壓電麥克風的聲音端口，會使懸臂壓電振動板偏轉及彎曲，振動膜會產生應力並根據電路的電邊界條件產生電場或電位移（電場大小取決於聲壓和耦合係數所產生的應力），將直接經由壓電效應（Piezoelectricity）在 MEMS 換能器的輸出端產生電壓。

然後，由封裝內的 ASIC 電路放大。壓電材料乃響應機械應力而產生電荷，不需電荷泵即可產生偏置電壓，所以可以超低電流啟動並快速喚醒（啟動速度僅約 200 微秒），不會遺漏關鍵字；但電容式則需要時間才能將 MEMS 充電至高偏置電壓，啟動時間 > 250 倍！

Vesper 自有 ZeroPower Listening

(ZPL) 收聽技術便使用壓電效應讓換能器充當開關。ZPL 的工作原理是：當聲波撞擊壓電懸臂時將移動懸臂、產生電壓，由非常低功耗的比較器電路感測到後，該電路將喚醒訊號發送到系統的其餘部分。

Vesper 強調，雖然任何聲音都會在壓電 MEMS 懸臂上產生電壓，但並非所有聲音都會啟動伴隨電路，該電路設計僅會響應 250Hz ~ 6kHz（包括人聲）聲音，風聲或空調聲等超出範圍者並不會讓麥克風誤動作；要啟動麥克風，聲音還須超過 65 ~ 89 dB 的指定聲壓級 (SPL)，可經由外部電阻值設置。Vesper 於去年中推出世界首個具有自適應 ZPL 技術 + IP57 防護等級的智能數位 MEMS 麥克風——VM3011，會根據環境的背景水平自動調整聲學閾值，在同一封裝集成 ASIC 和壓電 MEMS 換能器，數位 PDM 接口可用於傳輸音訊資料。

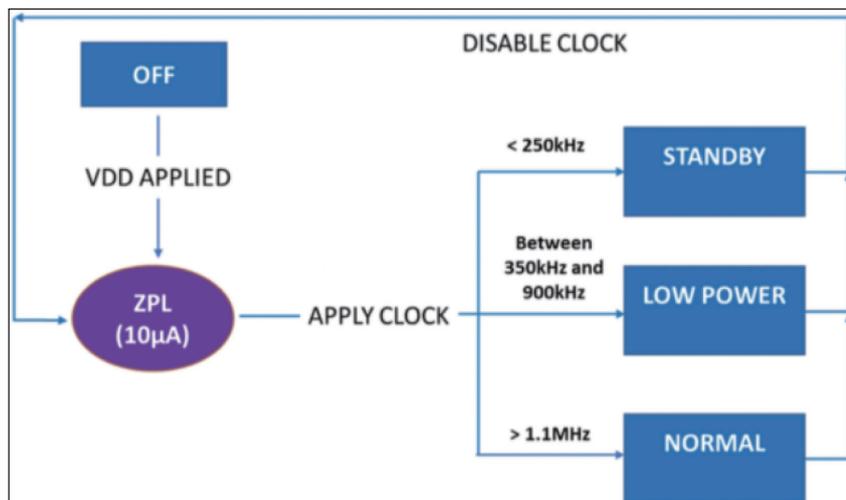
壓電式微機械超聲換能器，潛力大

藉由 PDM 時脈速度更改設備模式，I²C 介面則允許「即時」配置自適應 ZPL。當與壓電麥克風的超快速啟動結合使用時，系統可忽略背景雜音，僅在響應關鍵詞或需要聲學觸發才喚醒，90% 時間皆處於休眠狀態，可延長十倍電池壽命；搭配 DSP 還可運行 Sensory Truly Handfree 演算法做關鍵字檢查。ST 亦加快壓電 MEMS 研發腳步，於去年底宣佈與新加坡研究機構 A* STAR 的 IME 和日本製造工具供應商 ULVAC 合作，共同在 ST 的現有製造工廠內共同建立一條專注於 Piezo MEMS 技術的 8 吋 (200mm) 研發線——「Lab-in-Fab」。

此舉意在促進創新並加速新材料、製程技術及最終產品開發，加速壓電 MEMS 致動器新應用，包括：智能眼鏡、擴增實境 (AR) 擬真音場耳機、光達 (LiDAR) 系統的 MEMS 反射鏡、新興醫療應用的壓電式微機械超聲換能器 (PMUT)，以及商用／醫用壓電頭和工業 3D 列印。新廠擬於今年第二季投產、於 2022 年底量產。終端應用亦有新創意，德國設計工作室 WertelOberfell 去年底推出可由太陽能自行採集能量的自行車頭盔 ESub Tracks，集成印製壓電骨傳導揚聲器，熱壓到帽帶壓紋內層的皮帶上、再貼近騎士耳朵，以產生接近的聲音。

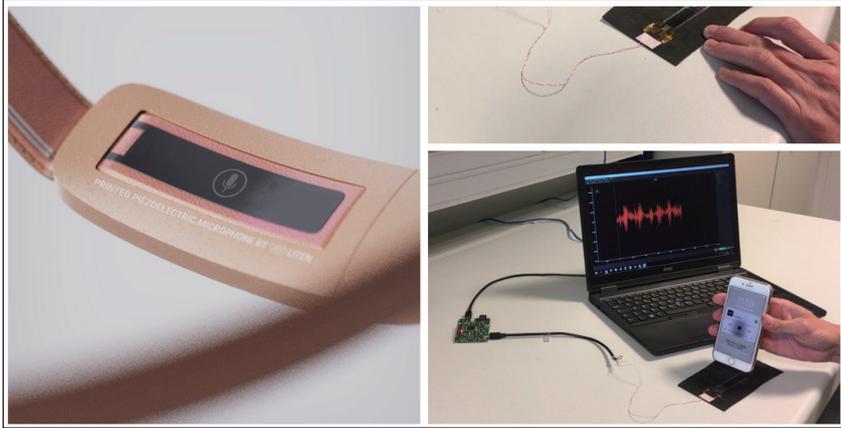
揚聲器與智慧手機的藍牙連接，以提供清晰音訊。在用戶的下

圖 3：VM3011 支援不同 PDM 時脈速度的多種操作模式，若用於聲音活動檢測的 ZPL 模式需關閉時脈，可使用 I²C 數據交換動態更改 ZPL 觸發器的閾值



資料來源：<https://vespermems.com/products/vm3011/>

圖 4：印刷的壓電麥克風能理解基本的語音命令並控制轉向訊號燈



資料來源：<http://www.werteloberfell.com/project/esub-tracks-helmet/>

巴下方，則有印刷壓電麥克風可理解語音命令；若騎士必須轉彎，麥克風可記錄轉彎訊號等語音命令並在帽沿以閃燈表示，無需手部動作示意。要將即時語音生成文本有兩大步驟：一是感測器必須將傳入的聲波轉換為電訊號，二是使用軟體確定單詞——將電訊號從類比轉換為數位後，由快速傅立葉轉換 (Fourier transform) 分析，以找出隨時間變化的頻率幅度；一旦聲音變化很大，必須透過機器學習提高時間準確性。有鑑於此，韓國學術單位仿效人類耳膜構造、用壓電式聲波感測器收音。

喚醒詞太枯燥？訂做一個喜歡的吧！

沿著位於振膜上的各定點分析多個訊號，較僅就一個訊號做時間、頻譜轉換分析更準確（實驗準確率達 97.5%），且號稱可收到的資訊量是傳統電容式感測器的兩倍。解決了收音問題，要開啓語音互動前，如果厭倦了單調的喚醒

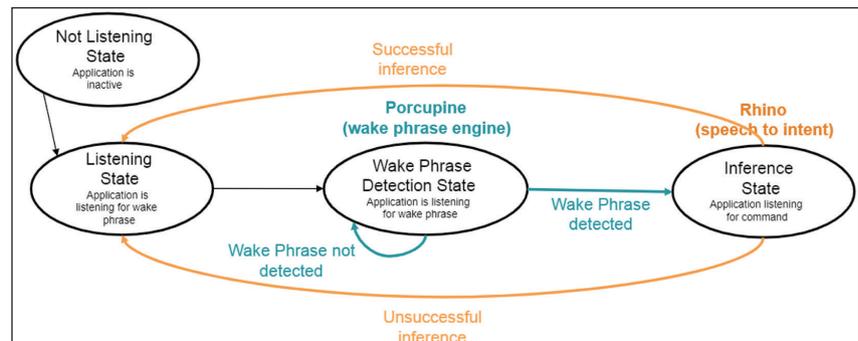
詞，現在也可「訂做一個它」。交互式語音技術新創公司 Picovoice 將公開共享他們的平台，並免費為公司許可 Alexa、Siri 和 Google Assistant 的喚醒詞模型。設備製造商既可提供自己的基於 Picovoice 平台的語音助理、亦可與最受歡迎的語音助理連接，不必付費使他們的設備了解喚醒詞。

Picovoice 提供一個建構自定義喚醒詞的平台，樂金 (LG) 和惠而浦 (Whirlpool) 等公司已將其整合到其產品中，以替代 Amazon 或 Google 的服務。語音引擎可針對非常特殊的用例生成定制的助理，

並帶有諸如洗衣服或打開微波爐等活動的命令。基於開放和可互操作願景，Picovoice 可確保語音助理都在設備上運行，而無需將任何數據發送到雲端；此種基於邊緣的方式意味著製造商不必依靠遠程伺服器添加語音交互，也不必擔心發送到雲端的數據受到破壞，所有數據都保留在本地和私有位置。

Picovoice 軟體幾乎可在微控制器 (MCU) 到 Web 瀏覽器的任何平台運行，透過設備上的語音辨識和自然語言理解 (NLU) 為邊緣語音 AI 提供支持，包括行動平台、桌面操作系統 (Linux/Mac/Windows)，以及 Raspberry Pi、BeagleBone 等嵌入式平台。他們斷言，「針對目的和品牌的語音助理，是語音作為介面發展的下一個邏輯步驟」，但這並不妨礙公司在設備上提供 Alexa 或 Google 助手，它只是不必自動執行。免費獲得喚醒詞模型許可讓設備製造商有更多選擇，例如，可為電梯和對講機等操作設備提供零接觸操作。CTA

圖 5：Picovoice 平台兩大功能：Porcupine 和 Rhino。前者負責傾聽並用於檢測熱門單詞；後者是語音到意圖引擎，可對語音數據執行自然語言理解 (NLU) 並推斷意圖，無需中間語音到文本



資料來源：<https://picovoice.ai/docs/tutorials/using-picovoice-engines-with-react/>