

效應 3：音訊處理器為聲控做第一線把關

Microsemi：提高辨識率與有效距離，音訊觸發用途廣

■文：任笠萍

亞馬遜 (Amazon) Alexa 的風行，顯示「語音」已成物聯網 (IoT) 重要人機介面；然而用語音發號施令的先決條件是：語音裝置須有良好的類比接收前端以及處理能力，才能提高辨識率，而「雞尾酒會效應」(Cocktail Party Effect) 一直是聲控專業埋首研究的重點，希望仿效人類聽力具有選擇性的原理，突破語音辨識盲點——儘管背景一片喧囂，還是能聽到想聽的內容。擅長類比／數位混合 IC 及客製化系統單晶片的美高森美 (Microsemi)，利用 FPGA 可現場編程、升級特性所開發的「Timberwolf」音訊處理器，就是一個關鍵元件。

Timberwolf 音訊處理器，智慧辨識主聲源方位，收音更清晰

Timberwolf 具備可運行三組「同步聲學回聲消除器」(AEC) 的特殊硬體加速器，能「摒除雜鳴」、採樣乾淨聲源——即使相距 4、5 公尺之遙，也能去除週遭電視、冰箱及冷氣機等家電噪音、或閒雜人畜的嘈雜聲，將主聲源確實



照片人物：Microsemi 語音事業部行銷暨應用總監 Shahin Sadeghi

收錄進去。Microsemi 語音事業部行銷暨應用總監 Shahin Sadeghi 介紹，基於「Timberwolf」音訊處理器所開發的模組，可應用在各式 IoT 裝置，例如：結合視訊建置門禁系統。此外，使用者可透過內建 Timberwolf 的語音裝置聲控家電、照明，或與手機結合進行遠距對話，為一心多用、分身乏術的使用者「解放雙手」。

Sadeghi 認為，在本地場域運用聲控的機會遠比對外連接到雲端多；「我們的使命是專注於聲音

擷取，確保語音從類比前端到節點裝置的通訊品質，順利將指令後送到區域內網或雲端，但不涉及聯外指令編碼的工作」，他詳述 Microsemi 在 IoT 語音應用的居間作用。特別一提的是，為辨識聲源方向，通常須採用多個麥克風陣列實現；例如，Amazon Echo 內嵌 6+1 麥克風陣列，可提示不同聲源的喚醒方向，但第一代 Google Home 只有兩個麥克風，僅能顯示是否正在偵聽、卻無法獲悉方向。然而，Timberwolf 只需單一麥克

圖 1 : Timberwolf ZL38063 最多可支援六個麥克風陣列



資料來源：Microsemi 官網

風就能搞定「指向性」問題。

Timberwolf 配備壓縮器／限幅器／擴展器 (CLE) 之麥克風拾音演算法，以及波束成形、全雙工雙向音訊、定態／非定態雜訊抑制技術，可提升語音控制／辨識的有效距離，並為 IoT 產品營造更多附加價值。智能家電只需建置一組麥克風，輔以演算法便可偵測、辨識玻璃碎裂或火警等特殊音頻，變身居家保全；一旦有異狀，可傳送訊號到終端使用者的手機。為避免誤動作，系統會在異常音頻出現三次後才警報。「現階段，煙霧與門窗

侵入警報是標配選項，亦可對門鈴聲、嬰兒啼哭、動物吠叫等在演算法做客製化回饋」，Sadeghi 補充。

難能可貴！與 Alexa 對答如流，吸引通訊服務營運商目光

為降低元件的相容性風險、讓用戶在設計及採購上有更多元的選擇，Microsemi 已與多家類比／數位麥克風的製造商完成產品匹配驗證；Timberwolf 音訊處理器有多種封裝版本，亦能應用到汽車、工業等自動化場域，但聲音資料庫會相對複雜。經過親身體驗，Sadeghi 在日前 COMPUTEX 2017 與 Alexa 互動，的確要比同場其他參展裝置要敏捷許多；即使正在播放音樂或有電視等其他聲源干擾，依然可對答如流、或遵從指示聲控音訊設備的音量，完全顛覆有些實機演示給人沈默半晌、呆若木雞，甚至不搭理人的負面觀感！

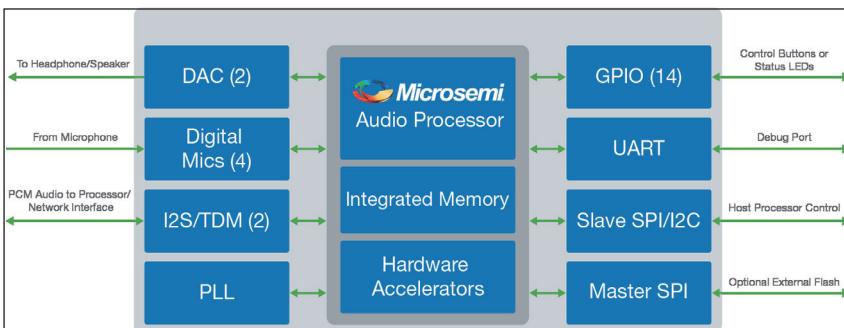
Microsemi 另針對 AVS 推出名為「AcuEdge」的開發套件，零售價格為 299 美元，已通過 Amazon Alexa 語音服務 (Alexa Voice Service, AVS) 相容性認證，

且與樹莓派 (Raspberry Pi, RPi) 相容。它可擴展支援兩組麥克風的雙向語音通訊，實現 180° 和 360° 拾音，優化免持通話品質、增強在惡劣音訊環境識別 Alexa 喚醒字的能力；另一方面，「打斷」功能允許用戶在播放音訊時中斷 Alexa 設備。AcuEdge ZLK38AVS 開發套件搭載 Microsemi ZL38063 Timberwolf 音訊處理器，讓開發者能便捷創建高品質語音辨識介面的電路原型，並進行相關演示。

Timberwolf 提供 SPI、I²S 和分時多工 (TDM) 等多種連接介面；在樹莓派運作 AVS 用戶端和觸發 Alexa 喚醒詞監測的同時，Timberwolf 也在執行各種音訊處理功能。ZL38063 專為電視機、機上盒和智慧音箱等懸掛式／桌上型衍生產品而設計，預載雙向語音通訊和「自動語音辨識」(Automatic Speech Recognition, ASR) 輔助演算法兩項韌體，可針對特定工作模式運作不同韌體，在正常運行期間動態切換工作模式。Sadeghi 透露，COMPUTEX 展出期間已有多家通訊服務營運商因有外包硬體設備需求，主動參訪 Timberwolf 語音處理方案並表達高度興趣。

借助 MiTuner ZLS38508 或 ZLS38508LITE GUI 套裝軟體，可在視覺化畫面上設定路由配置、直觀使用下拉式選單調整參數，並在發送 (Tx) 和接收 (Rx) 音訊通道做功能編程及設置類比／數位增益。另額外選配 ZLE38470BADA 自動調試套件，提供自動調試和簡易手動微調功能。**CTA**

圖 2 : ZL38063 Timberwolf 音訊處理器區塊圖



資料來源：Microsemi 官網