AI聽覺進化----智慧音箱

人類與人工智慧進行語音交互的夢想已經很久了，但直到21世紀之初，這個夢想仍然只是在影視作品和遊戲中不斷完善。技術進步的道路並非一帆風順，幾代科學家在艱難中不斷探索。科技巨頭們早已在智慧語音交互應用中佈局，大家都感覺到了智慧語音交互時代即將來臨，只是不知以何種方式呈現，直到智慧音箱的出現。

從時間上來看，智慧音箱的出現與近些年快速發展的AI技術是同步的，可說是AI市場最為成功的一個落地應用。從2015年不到一百多萬台，到2018年的一億台出貨量，智慧音箱市場正在大規模爆發……

艱難中前行的語音交互技術

■文：編輯部整理

一直以來，通過語言與機器直接溝通，是很多技術人員追求的目標，可看似簡單的語音交互，卻經歷了長達半個多世紀的技術探索。這期間不管文學還是影視作品，都一次次描繪了人機語音交互的美好場景。但直到 21 世紀初，人與機器的語音交互仍然是令人抓狂的一種操作，遠遠沒有鍵盤和滑鼠來得方便。

原始發展階段

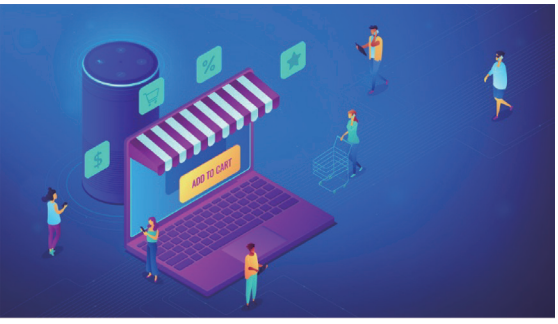
在 1952 年，貝爾實驗室研發出了 10 個孤立數位語音的識別系統，為人類的語音辨識開啓了篇章；20 世紀 60 年代開始，卡耐基

梅隆大學 Reddy 等人開展了連續語音辨識的研究，但是相關研究進展緩慢；1969 年，經歷了十幾年語音辨識研究的貝爾實驗室，也不得不承認在當時的技術條件下，語音辨識難度超乎想像，Pierce J 在公開信中將語音辨識列為短期內難以突破的技術難題。但是科學界仍然在尋找語音辨識的突破方法。

20 世紀 80 年代開始，以隱藏式馬可夫模型 (hidden Markov model，HMM) 方法為代表的基於統計模型方法逐漸在語音辨識研究中佔據了主導地位。HMM 模型能夠很好地描述語音信號短時平穩特性，將聲學、語言學、句法等知識集成到統一框架中。此後，HMM 的研究和應用逐漸成為了主流。

快速發展階段

當時在美國卡耐基梅隆大學讀書的臺灣人李開復在 HMM 模型的此基礎上研發出了 SPHINX 系統，這是技術人員首次嘗試“非特定人連續語音辨識系統”，其核心框架就是 GMM-HMM 框架，其中 GMM 是指 (Gaussian mixture model，高斯混合模型) 用來對語音的觀察概率進行建模，HMM 則對語音的時序進行建模。



同時期發展出的技術，還有 20 世紀 80 年代後期人工神經網路 (artificial neural network， ANN)，採用 ANN 技術進行語音辨識研究也成為了語音辨識的一個方向【而當 ANN 後來進化為深度神經網路 (deep neural network，DNN)，語音辨識技術才有了本質的突破】。

到了 20 世紀 90 年代，隨著電腦技術的快速發展，包括個人電腦在內的一大批設備開始嘗試使用語音辨識技術。這一時期劍橋發佈的 HTK 開源工具包大幅度降低了語音辨識研究的門檻。然而在接下來的一段時間，GMM-HMM 框架的技術局限性使得其應用效果差強人意。筆者清晰得記得，當時 IBM 推出的一款語音辨識軟體，安裝包就有幾張光碟，在硬碟容量寸土寸金的個人電腦中，語音辨識軟體的體積比很多當時的大型軟體還要大，除去存儲成本，更加麻煩的問題是安裝之後的訓練工作，僅僅識別一個人的語音就需要花上幾個小時來訓練，而且識別錯誤率還很高，最後不得不束之高閣。這可能是接下來在 21 世紀初的幾年中，語音辨識很少被人提及的原因。

語音交互技術實現突破

2006 年 Hinton 提出深度置信網路 (deep belief network，DBN)，解決了深度神經網路訓練過程中容易陷入局部最優的問題，為深度學習技術開啓新方向。2009 年，Hinton 和他的學生 Mohamed D 將 DBN 應用在語音辨識聲學建模中，並且在 TIMIT 這樣的小詞彙量連續語音辨識資料庫上獲得成功。

2011 年 DNN 在大詞彙量連續語音辨識上獲得成功語音辨識效果取得了近 10 年來最大的突破，並從此成為主流的語音辨識建模方式。

3 年以後的 2014 年 11 月，以 DNN 技術為基礎的亞馬遜憑藉 Echo 一舉開創出了智慧音箱這個全新的市場。

在語音辨識技術方向中，具有更強的長時建模能力的迴圈神經網路 (recurrent neural network，RNN)，卷積神經網路 (convolutional neural network，CNN)，以及在語音辨識領域獨樹一幟的科大訊飛公司提出的 DFCNN 技術相繼出現，從而使得人機語音交互的體驗越來越好，以智慧音箱為代表的語音交互設備開始受到越來越多消費者的歡迎。

值得一提的是科大訊飛在語音辨識技術方面的研究，其 DFCNN 框架的識別率相較以往的技術再次提升了 15% 以上，比傳統的 GMM-HMM 框架性能提升 30% ～ 60%，並與 Google 在語音辨識方面展開深度合作。在中文語音辨識方面，更是佔據超過成市場，是中國相當多智慧音箱中的首選語音辨識技術。

參考資料：

《語音辨識技術的研究進展與展望》科大訊飛股份有限公司人工智慧研究院

作者：王海坤， 潘嘉， 劉聰 CTA

智慧音箱軟硬生態圈

■文：徐俊毅

亞馬遜 (Amazon) Echo 的巨大成功，讓同樣在語音交互技術上探索出路的科技界忽然明白了市場的方向。蘋果、Goolge、Microsoft 等科技巨頭紛紛投身到智慧音箱市場，因為智慧音箱不再是一個孤立的消費產品，而是與千千萬萬消費者連接的語音交互入口，透過這個入口，大家能做的事情非常非常多。在大洋彼岸的中國大陸，科技精英們同樣意識到了智慧音箱潛在的價值，也迅速跟進，與既有的互聯網服務相結合，推動技術和市場的發展。

在語音交互產品方面，亞馬遜實際晚於蘋果，Google 和 Microsoft。

早在 2011 年，蘋果的 Jobs 堅持完成了對 Siri 的收購，使得 Phone4S 手機的語音交互應用成為產品一大亮點，但或許是手機市場實在太好的原因，Apple 並未在 Siri 上下太多功夫，讓 Siri 一直沒有更進一步的表現；Microsoft 則是專注於他們作業系統的生態，在 Windows 10 上面首次搭載了自家的語音助手 –Cortana，與使用者進行語音互動，Cortana 透過記錄使用者行為，學習使用者習慣，配合搜尋引擎回答問題，多媒體點播、讀取郵件等等工作，其實 Cortana 源自 Microsoft 遊戲 Halo 系列中的一個人工智慧角色，透過語音與人類進行資訊交互，雖然她在遊戲中的生日是 2549 年 11 月 7 日，但是實際上卻是在 2001 年就出現在遊戲中了。和其他人一樣，身負 Android 重任的 Google 同樣將功夫下在移動設備市場，Google Vocie Search 仍然是其為其搜索業務拓展市場服務，後整合到 Google assistant。

強強聯合打造各自生態系統

當業界將語音交互入口統一到智慧音箱的時候，這些已經存在的技術，配合各自的產品形成了智慧音箱不同的技術陣營，打造各自的生態系統。

Amazon Alexa 是與其 Echo 音箱同期發佈的開放平臺，於 2015 年 6 月 Echo 音箱正式發佈後不久，開放給協力廠商開發者，目前各類項目已經接近 4 萬個，遠超其他對手。Amazon Alexa 已在全球 38 個國家開通 (中國大陸暫未開通)，涵蓋英語、德語、法語、義大利語、西班牙語和日語 6 種語言。除了搭載 Echo 音箱產品，Alexa 也能與聯想、哈曼卡頓等音箱產品，amazon fire TV 智慧電視，amazon fire、HTC 等平板電腦、智慧手機，筆記型電腦、PC，智慧冰箱、智慧燈、智慧開關等智慧家居產品，智慧耳機、智慧手錶等可穿戴設備，以及具備 AI 功能的寶馬、雷克薩斯、豐田等品牌汽車連接，為使用者提供語音交互。

圖說：Amazon 的 Echo Show 和 Echo Spot 均以配備螢幕

Google Assistant 是谷歌推出的支援語音交互的虛擬助力服務，2017 年發佈軟體開發套件，不僅支援語音交互還支援視覺相應，比如透過手機的相機翻譯文字或說明。目前已經支援幾十種語言，除了可以配合自家的 Google Home 音箱、Pixel 智慧手機，還有 Sony、Panasonic、LG 等品牌的家電音箱產品連接，並在部分富豪汽車上嘗試提供語音交互服務。

Apple Siri 是蘋果自家智慧音

圖說：Google Home 智慧音箱

圖片來源：Google

箱 HomePod 的基礎語音平臺，早在 2015 年，Apple 就曾經嘗試過與自家的 Home Kit 智慧家居控制系統整合，結果是差強人意，如果不是智慧音箱市場的崛起，Siri 可能還會停留在僅僅是一個 iOS 的 App 這樣的位置。2016 年蘋果開放了 Siri 介面，允許開發者在新增的 Sirikit 協議下調用 Siri 實現音訊交互，目前僅能在蘋果旗下各產品中使用。

Microsoft 的 Cortana 在 2014 年微軟開發者大會首次發佈，後端使用 bing 搜尋引擎，以 Windows 10 作業系統為基礎，為使用者提供語音交互服務，如前面提到的搜索問題，控制智慧家居產品，連接自家智慧音箱。2017 年開放給協力廠商開發者，目前支援包括中文在內的 10 來種語言。

高通與 Line 合作的 Clova 智慧雲則是另外一種方式的合作，一個是晶片供應商，另一個則是擁有龐大的線上互動使用者的即時通信平臺。

在大陸市場，由於科大訊飛在中文語音辨識技術上的主導作用，一大批智慧音箱、智慧家居、智慧汽車產品與之合作。比如位居中國智慧音箱銷售前列的－叮咚音箱 (京東的智慧音箱) 就是採用了科大訊飛的語音交互技術。

其他幾家中國的頂尖科技公司在智慧音箱市場也無一缺席，比如：百度 DuerOS– 小度，2017 年 7 月面世，依託百度搜尋引擎提供多種服務；阿里巴巴的 AliGenie 語音開發平臺配合自家的天貓精靈智慧音箱，語音交互技術來自另一家技術供應商思必馳，為用戶提供購物、出行、智慧家居等各種場景的服務；小米的語音交互平臺被命名為小愛同學，自己的智慧音箱就叫小愛音箱，語音交互技術同樣來自思必馳，最大優勢是小米在家電，IoT 設備龐大的硬體產品群以及大量的使用者，“小愛同學”可以直接與這些設備進行語音交互操作，而且小米家硬體產品種類還在不斷擴充。

圖說：百度智能音箱 2018 下半年快速成長小度在家智慧視頻音箱

圖片來源：baidu.com

TI、MTK 佔據優勢供應商位置

從結構上看，智慧音箱與傳統音箱的差別除了增加麥克風或者麥克風陣列，最大差異就是後端需要一個小型的計算系統，與雲端進行資料交互，如果去掉音訊交互硬體部分，這個小型的計算系統與其他的移動設備沒有任何區別，需要有 CPU、存儲、無線通訊 (藍牙、WiFi)、電源等等各種模組，而且新一代智慧音箱已經配備了螢幕和觸控操作功能。

在以智慧音箱為主的語音助理設備晶片領域，聯發科目前佔據絕對優勢地位，市場佔有率達到 70% 以上。作為排名第一的晶片供應商，MTK 為亞馬遜 Echo 系列 (Echo、Dot、Plus、Spot)，Google Home，百度小度在家，阿里天貓精靈，叮咚等主流產品提供解決方案。此外包括 TI、NXP、Realtek、Marvell、Amlogic、全志、瑞芯微、國芯在內一些移動領域的供應商也在為智慧音箱提供產品。

而在音訊交互的硬體設備上，無論是亞馬遜 Echo、Google Home 或者叮咚、天貓精靈等產品，ADC，數位音訊功率放大器、LED 驅動中，都不乏德州儀器各類產品的身影，TI 可說是在音訊信

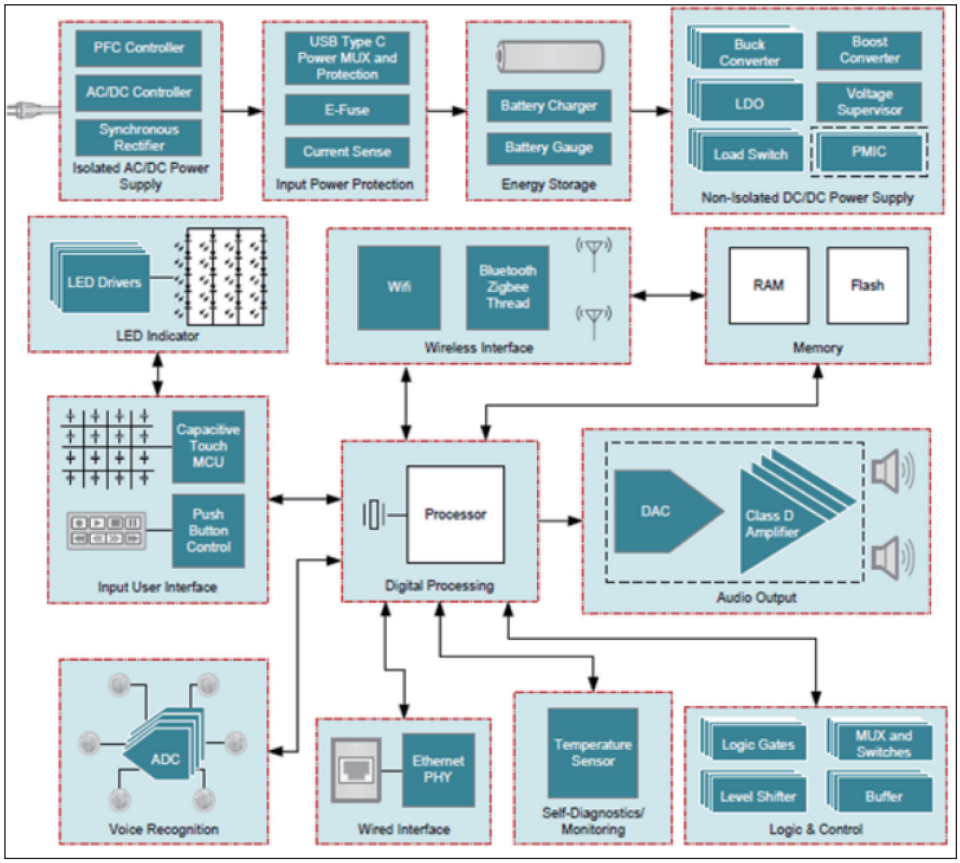
號鏈佔據了相當有利位置。

新一代智慧音箱更具交互功能

德州儀器認為：在智慧語音助力設備設計製造領域，總體成本、電池續航力、熱效應、回升消除和音質方面是未來競爭的重點。為此，TI在10 W以下的音訊系統，用鐵氧體磁珠取代電感器，一個典型系統中的4個電感器的平均成本為0.16美元，而4個鐵氧體磁珠的平均成本僅為0.04美元。僅這一項，設計人員就可在一個雙聲道系統中平均節約0.12美元；同時TI從音訊放大器著手控制系統功耗，運用混合調製模組音訊放大器可調整輸出脈衝寬度調製的工作週期，控制從電感器到音箱的電流，從而減少50%的閒置電流；在音質方面，透過簡單的D類音訊放大器系濾除雜訊和背景雜訊，改進演算法和模型提升音訊放大器的輸出品質。

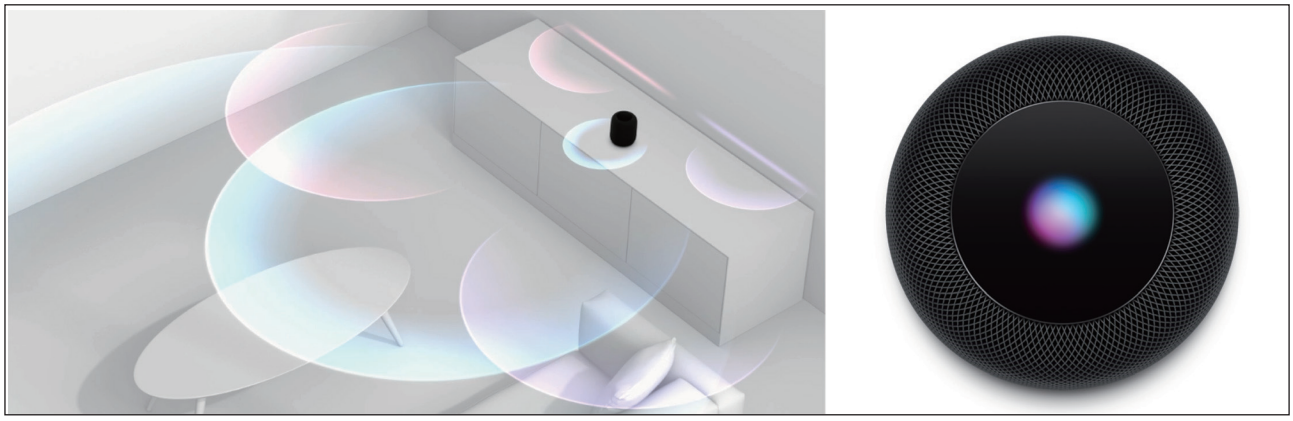
除了高保真的音質，配備電池和彩色螢幕的新一代智慧音箱對硬體系統提出了更高的要求，觸控功能將會出現在越來越多的智慧音箱上，藍牙語音交互也即將成為市場標配，這些新增加的功能將增加系統複雜度，並對功耗的控制更為嚴苛，這要求晶片及方案業者也要適時跟進市場的需求。CTA

圖說：新一代智慧音箱系統框圖



圖片來源：德州儀器

圖說：2019年初即將發佈的HomePod主打音質



圖片來源：apple.com

2018 智慧音箱市場爆漲 3 倍 2019 動力十足

■文：徐俊毅

從 AI 經濟的角度來看，目前在市場上最為成功的 AI 專案應當是智慧音箱。

來自不同機構的資料顯示，截至到 2018 年 7 月，有 4300～5000 萬美國人已經擁有了一台智能音箱。Strategy 報告顯示，2018 第三季全球智慧音箱同比增長 197%，達到 2270 萬台。2018 年 12 月，來自 RBC 的資料指出，到 2018 年末美國的智慧音箱普及率已經達到 40%，相比 2017 年已經翻倍。預估 2018 年全年的全球智慧音箱設備數量出貨量超過一億台，而 2017 年全球的智慧音箱出貨量為 3200 萬部 (其中 1800 萬部來自 2017 年第四季)。

鑒於龐大的需求和巨大的市場潛力，投資人估計，科技巨頭們未來將會花費年度研發預算的 10%用於語音辨識，總計超過可能 50 億美元。Insight 資料顯示，預計 2024 年全球智慧音箱的市場規模將達到 300 億美元。

亞馬遜和 Google 是 2018 年最大贏家營業成長曲線

截止到 2018 年 11 月的資料，亞馬遜總共賣出約 4700 萬套 Echo 設備，占美國市場 66%，據全球市場 42% 份額，Echo 音箱同時服務著 80 個國家的消費者，平均每天處理 1.3 億個問題。就在剛剛過去的耶誕節期間，亞馬遜為 Echo 智慧音箱提供的雲服務，一度因為不堪重負而罷工。值得注意的是，Amazon Alexa 生態系統的成長迅速，僅僅 2018 年下半年，Echo 就新增 2 萬多種技能，這些包括大量協力廠商開發者參與的專案，為 AmazonEcho 提供了大量方便快捷的應用，提升消費體驗，使其穩坐領頭羊位置。

排名第二的 Google 在 2018 年也有十分搶眼的表現，統計顯示 Google 的智慧音箱，在美滲透率從 2017 年的 8% 躍升至 23%，占美國市場 29%。單從增長速度來看已經超越亞馬遜，Google 背後龐大的生態系統和內容服務是其最大的潛在實力。

MIC 產業分析師曾巧靈表示，

2019 年智慧音箱市場可望持續成長，以美中兩大市場為核心，4 大廠包括亞馬遜 (Amazon)、穀歌 (Google)、阿里巴巴、百度四強鼎立。其中 Amazon 與 Google 幾乎囊括 8 成以上美國市場。Amazon 與 Google 未來將瞄準美國以外市場，例如：Google 加入更多國家的語言支援，包含全球第二多人口使用的西班牙語，並提供雙語支援以增加產品全球普及率；Amazon 則宣佈全球第二大音樂串流服務 Apple Music 加入 Echo 服務，增加對數千萬 Apple music 用戶的吸引力；除此還有產品線擴充，維持機海戰術策略等。

大陸市場血拼價格

從出貨情況來看，2018 年的中國大陸智慧音箱市場是相當喜人的，相較 2017 年 150 萬台智慧音箱的總量，2018 年猛增至將近 1000 萬台。但是價格戰讓業者們的生存情況變得非常嚴峻，從 2017 年底“雙 11“，阿里巴巴直接將天貓精靈 X1 售價從 499 降到 99 元開始，京東緊隨其後將 399 的智慧音箱價格拉低至 49 元，一場空前慘烈的價格在 2018 年的智慧音箱市場上演，百度 89 元的小度音箱，小米小愛智能音箱 mini 限時價 99，家底雄厚的品牌紛紛以遠低於成本的價格搶佔市場，2017 年“百箱”大戰的盛況不復存在。在強勢資金補貼的情況下，阿里巴巴、百度、小米、叮咚四家拿下大陸約 90% 的市場佔有率，剩餘 10% 的市場讓其他中小業者瑟瑟發抖。有業內人士表示：大陸智慧音箱市場業者在 2018 年的生存狀況，用十不存一來形容，可能並不為過。

MIC 產業分析師曾巧靈認為，中國智慧音箱業者佈局，阿里巴巴、百度等大廠皆以低於百元人民幣的價格取得市占，其中，阿里巴巴厘用其電商優勢，整合零售、支付等服務，主打語音購物；百度則持續強化 DuerOS 語音助理功能，並結合智慧型手機品牌廠商，奠定語音助理使用者基礎。此外，阿里巴巴和百度雖有不錯市占，但仍存在其他如叮咚、小米等本土品牌競逐者，市場普及程度也未及 Amazon、Google 於美國市場的表現。雖然中國智慧音箱市場在 2018 年大幅躍升，但以 2018 年銷售近千萬台的成績而言，2019 年仍有巨大成長空間。

智慧語音市場爆發持續力待觀察

在語音辨識技術方面，MIC 資深產業分析師楊政霖表示，明年情感辨識發展將進入情緒辨識 2.0，呈現應用多元與情緒優化情境。明年後應用領域可延伸到影視、零售、醫療、教育、電話客服等領域。2020 年後可期待例如家用型機器人、智慧音箱等、甚至聊天機器人具備情感辨識功能。

有調查顯示，目前美國的消費者使用智慧音箱除了提問題 (查詢時間，天氣)，更多的時候是用來開關燈，可見人們是多麼不想去按那個開關，這是一個良好的開始，智慧音箱與智慧家居和汽車等設備的結合是各大供應商正在搶佔的熱點領域；中國消費者則抱怨一些廉價智慧音箱遠場識別差，誤動作多，語義理解差，音質不佳等等，新鮮過後就失去了興趣，這可能是中國市場總體普及率不高的一個原因，不過隨著 2018 年智慧家居設備的快速發展，智慧音箱對智慧家居設備的控制使其開始顯現實用價值，消費者對智慧音箱的興趣正在快速上升。

綜合來看，智慧音箱市場在 2018 年第三季迎來爆發 (這點與 2017 年類似)，將整個市場的成長速度提升了一個檔位元，未來市場預期看漲。但值得注意的是，相比起智慧手機，消費者對智慧音箱的換機動力並不強勁，現有產品已經滿足大部分使用需求，因此智慧音箱市場的持續爆發力可能不如智慧手機，除非不斷創造出更多需求。從整個 AI 語音交互的角度來看，智慧語音交互功能也在迅速向智慧音箱以外的設備市場擴展，交通、工業、家居、醫療等等市場都有語音交互需求和明顯的市場成長，市場整體的成長潛力依然可期。 CTA