



作者：徐業良(2009-10-12)；推薦：徐業良(2009-10-12)。

附註：本文發表於汽車購買指南雜誌，2009年十一月號，史丹福專欄。

## 淺談語音控制技術—建立人車伙伴關係的關鍵元素

您最近一定注意到電視上一支很有科技感的廣告，裕隆自創品牌的新車納智捷 Luxgen，化身為一個類似電影「機械公敵(I, Robot)」中的機器人，配上旁白：「這是人，這是她，這個人走向她，一百年走過，這個人終於走向她」，「她思考，她說話，她對這個人說話，這個人對她說話」…

廣告中 Luxgen 是「她」喔！廣告最後「人」和「她」還有一個頗為前衛的接吻鏡頭…

噢，扯太遠了。這裡沒有要討論 Luxgen 的科技（或性別），只是想指出 Luxgen 廣告中想要凸顯人和汽車之間的「伙伴關係」。這個「伙伴關係」中很重要的一個元素，就是「她對這個人說話，這個人對她說話」，人和汽車之間的對話。

像我這個年紀的 LKK，馬上就聯想起二十多年前的電視影集「霹靂遊俠李麥克」中，那部會和李麥克說話聊天的 Trans Am 跑車，名喚「伙計」的霹靂車。

「人機介面」指的是人藉以與機械互動的所有方式，而在許多應用上「語音控制系統(voice-activated system)」的確是最合理、最方便的人機介面。語音控制早已不是什麼創新概念了，不只在汽車的應用，在許多其他領域產品的人機介面，如手機、家電、機器人、乃至智慧住宅，語音控制的應用都備受期待。

但是我個人覺得，語音控制技術應用在汽車的人機介面上，角色尤其重要。前面提到人和汽車之間的伙伴關係要靠人車之間的對話來維繫（Luxgen 倒是沒有特別強調語音控制技術），這已經是比較抽象、概念層次的需求。在操作方便性的實務層次上，越來越多的科技，無線通訊、網際網路、影音娛樂等等系統通通被搬上汽車，固

然帶給駕駛人許多方便和樂趣，但無形中也帶給駕駛人許多新的潛在危險，駕駛人必須一面開車，一面操作撥號、開關、鍵盤等等。語音控制技術使得駕駛人操控使用這些配備時，手不必移開方向盤、眼睛不用離開路面，的確有助於提升行車安全性。

除了安全性之外，語音控制技術還有其他重要價值，例如駕駛人動動口就可以控制各種配備，取代按一大堆按鍵的動作，操作簡單快速許多，車主也不必花很多時間了解汽車上這一堆配備到底要怎麼操作。語音控制技術甚至可能為汽車內裝和儀表板設計帶來新的革命，汽車內裝設計的大難題，就是如何在小小的汽車面版和駕駛人觸手可及的車室空間內，擠進一大堆儀表和控制按鍵，再有天分的設計師可能也很難做出完美的設計。使用語音控制技術的話，很多功能的控制按鍵可以被「藏起來」，給了設計師更大的自由度，汽車的內裝、儀表板也可以清爽得多。

語音控制技術的可能性眾家車廠都看到了，量產汽車應用語音控制技術也已經有相當長的歷史。汽車上語音控制應用一開始主要還是在車用電話，豐田 Lexus 旗艦車 LS400 早在 1993 年便已經配置語音控制的車用電話；Mercedes Benz 的 S-Class 在 1996 年德國市場銷售的車款也裝上了 LINGUATRONIC 第一代語音控制系統，主要也是車用電話的語音控制，這個系統聽得懂大約 30 個語音指令；2000 年 LINGUATRONIC 第二代語音控制系統聽得懂的字彙已經增加到 300 個，特別的是 LINGUATRONIC 還可以聽得懂德文、英文、義大利文、西班牙文、法文等多國語言，語音指令範圍也從車用電話控制擴大到音響、空調等操作。積架 2000 年推出的 S-Type 轎車車款，一共聽得懂 44 個語音指令，包括撥打和接聽電話、選擇收音機電台、播放 CD、調整空調，都可以用語音指令完成。近年來許多汽車導航系統也可以接受語音指令，駕駛人唸出一個地名（或者一個字母一個字母拼出一個地名），導航系統便可在地圖上找到這個目的地。不是只有駕駛人單向說話而已，汽車也會對駕駛人說話，像是您所熟悉的導航系統會隨時以語音方式提醒駕駛人路況，有的車用電話系統接到簡訊還會以語音方式讀出簡訊內容，不過目前為止還做不到人車真正的「對話」。

這麼好用的語音控制技術，出現十幾年之後卻還是沒有看到在汽車上普及應用，電腦的語音辨識能力是最大問題。汽車語音控制系統期望讓駕駛人以輕鬆自然的說話方式，來控制汽車各項系統，不過受限於電腦的語音辨識能力，現行語音控制系統（您也許有使用手機語音控制功能的經驗）使用者往往還是需要一個字一個字降低速度、咬字清楚地讀出特定指令，即使如此，系統卻還是常常會聽不懂或辨識錯誤。高檔一點的手機都會有語音控制功能，但似乎沒有人真正在意，很少看到有人用語音控制功能打電話，實用性似乎還有段距離。在汽車的應用上，語音控制也還不能取代手控，更談不上人車之間可以流暢對話的「伙伴關係」建立了。

這陣子在學校做一些機械人方面的研究計畫，也想試試看應用語音控制建立人機介面，深究之下才發現語音控制技術確實不簡單。

「語音辨識(voice recognition)」是語音控制系統技術的核心。要辨識您的語音指令，系統電腦首先必須將聲音數位化，轉變到頻率範圍來顯示，您一定在電視影集看過，說話過程中電腦螢幕上每個頻段都有數值上上下下跳動著。電腦不斷分析每一個頻段，尋找一些聲音的特徵，叫做「音素(phoneme)」，聲音的元素，也就是最小的聲音單元。每一種語言都有自己的音素集合，舉例來說，英文和德文都有大約 40 個音素，而每一個字彙則都是由這些音素所組合而成。找到不同頻率範圍的這些音素之後，再把音素組合起來，成為一個數位模式(digital pattern)，和儲存在系統記憶體裡的字彙作比對，看看是否儲存有類似的數位模式，嘗試辨識出這個語音指令。

這整個過程在現代電腦中處理只需要幾個微秒的時間，辨識出語音指令後，系統再將語音指令轉變成為控制訊號，操控汽車上各項系統，或者找出預先設定好對應的答案回答駕駛人。

說起來容易，在這個過程中有許多技術的細節，使得提高語音辨識正確率相當困難。首先是麥克風的品質要好不能失真，才能夠完整蒐集到語音指令在各個頻率範圍的數值；其次是背景噪音的處理，人腦是非常聰明的，您在嘈雜的環境和別人對話，接收一大堆聲音之後，您的大腦會自動辨識、濾除掉環境中的噪音，只留下您想聽的部分，電腦系統要模擬這樣的功能，可就困難了。語音辨識系統往往在實驗室安靜環境中辨識正確率表現優異，一到真實嘈雜環境就差異很大，然而汽車使用的環境先天上又特別嘈雜，周邊交通噪音、路面噪音、風切聲，使得噪音濾除能力在車用語音控制系統中顯得更為重要。

前面提到電腦的語音辨識主要在找出每一個頻率範圍中的音素，再組合起來與記憶體中的儲存的字彙作數位模式的比對。問題是即使是同樣的字彙，每個人說話的方式語調、口音都不同，其實不容易有標準模式，甚至每個人說話的速度和斷句的方式，例如「臺北市基隆路」，和「臺北市」、「基隆路」，都會造成語音系統辨識的困難。

您實際與人對話時要「聽懂」別人的談話，主要靠的其實不只是您所聽到的「音素」，而是您對談話內容上下文語意的瞭解，搭配對話者表情、手勢的輔助，因此不同人說話的語調、口音、速度，對您來說可能不是特別的困擾，甚至講話含糊一些，一句話沒有說完整，您也多半可以猜得出意思，絕對不會把「伙計」聽成「火雞」。電腦系統完全沒有這些額外的幫助，要電腦系統能夠瞭解談話的語意，這牽涉到更高層次的人工智慧、自然語言的研究，就更複雜了。

一般語音控制系統多半會提供一些訓練功能來克服這個問題。大體來說，就是讓使用者對著系統反覆讀出特定語音指令，或者一篇預先設計好的文章，讓系統學習這個使用者專屬的音素數位模式。

語音辨識和語音控制技術不斷在進步，但似乎進步有些緩慢。二十多年前的科幻影集「霹靂遊俠李麥克」許多人車對話的場景，到現在看起來還是，很科幻。