



作者：徐業良(2007-06-13)；推薦：徐業良(2007-06-13)。

附註：本文發表於汽車購買指南雜誌，2007年七月號，史丹福專欄。

淺談汽車語音辨識系統

大約十年前開始吧，汽車就不再是一個機械的產品了，微處理器、行車電腦逐漸掌握了汽車所有的系統，電子元件勢力範圍擴展到汽車引擎、傳動、煞車等等原本很「機械」的系統，導航系統、影音娛樂、以及各種各樣的通訊裝置，更是一波一波地入侵汽車車室這個小小空間，改變了汽車機械的風味。

我這個機械系的教授，聽起來很酸對不對？

汽車這種改變，對駕駛人也造成了兩個後遺症，一是各種儀表版、按鍵、控制鈕越來越多，駕駛人面對像是飛機駕駛座艙般複雜的儀表版，越來越手忙腳亂。

第二個後遺症是，這些電腦、微處理器掌握大權後不久，紛紛開始「說話」。語音也好、警示鈴聲也好，您開車時要是有什麼做得不對的小地方，像是沒有繫上安全帶、車門沒關好、下車時忘記關大燈之類的，一定會聽到汽車上這些電子系統很雞婆地發出聲音提醒您。

會不會覺得怪怪的？到底誰是老大？！我們要的應該不是一部會說話的汽車，而是一部會「聽話」的汽車。

所以您知道，汽車語音辨識系統的發展，是汽車科技發展上很自然的下一步。經過十幾年的發展，建立在語音辨識軟體上的聲控系統，已經成為豪華車的必要配備，是取代車內各種通訊裝置、娛樂系統、衛星導航系統上，眾多按鍵、搖桿、螢幕觸控等控制方式一個絕佳的替代選擇，也是駕駛環境「眼睛不離開路面、手不離開方向盤」的嚴苛要求下，人機介面設計最好的解決方式。

語音辨識軟體的發展讓人們可以和電腦對話。well,「對話」的層次可能還太高了一些,但至少已經有好幾家電腦公司成功開發出電腦軟體,可以讓個人電腦接受口述的指令,汽車上也開始有能接受有限基本指令的語音辨識軟體。

好吧,我的車子陽春得很,我從來沒有使用汽車上語音辨識聲控系統的經驗,我最接近的經驗是使用手機的語音辨識。您知道,金城武在電視廣告中對著手機說「拉麵」,手機就自動撥號給拉麵店,不過在真實生活裡可不是如此,我對著手機說「媽媽」,說了十幾次,嘗試了各種腔調,手機還是聽不懂,打給媽媽的電話也撥不出去。

當然手機上的語音辨識系統微處理機的處理速度和記憶容量都很受限制,早期汽車上的語音辨識系統也受到硬體限制,準確率不高,但是微處理機的成本越來越低,車上電腦的複雜度相對提高,這些計算和記憶上的限制大部分都消失了。許多公司開始針對汽車開發語音辨識軟體,例如 Acura 的導航系統的就有很先進的聲控裝置,是由 IBM 個人電腦版的 ViaVoice 軟體演進而來的,其能力和在個人電腦上運作一樣好,甚至更強,可以玩類似「查號台」的 trick,駕駛人口述一個電話號碼,系統便能從儲存的通訊錄中找到對應的地址,在導航系統上 show 出來。

在汽車上使用語音辨識軟體面臨特別的挑戰,是汽車車室內常有各種各樣的噪音。在汽車車室內經常性出現的路面噪音和氣流噪音已經不是大問題了,語音辨識軟體都有能力辨別這些固定噪音加以濾除。但是車室內突然發生的尖銳噪音仍然會讓語音辨識系統失誤,像是小孩在後座哭鬧的聲音、大聲關門的聲音、雨刷刮玻璃的尖叫聲等等,所以語音辨識系統還是要針對每一部不同的車子做校正。舉例來說,麥克風必須被放在車內適當的位置,才能正確地抓到駕駛人的聲音,但是又必須要離車內噪音來源越遠越好,像是風扇或轉向訊號聲。

IBM 的研究員又向前多走一步,增加了車室內攝影機來改進語音辨識系統的準確性。藏在後照鏡後方的攝影機拍攝的影像經過軟體處理,可以找到駕駛人的臉和嘴唇來分辨駕駛人是否有在說話,駕駛人有說話時才啟動語音辨識。加了這台攝影機甚至還可以讀駕駛人的唇語來幫助語音辨識系統,人說話的時候有些音聽起來好像差不多,但是嘴唇的形狀會不一樣。

此外攝影機對駕駛人影像的監測,也可以和汽車安全系統結合,例如汽車偏離車道時,從駕駛人的頭部位置可以判斷駕駛人是想要變換車道,還是在打瞌睡。除此之外,在車室內的攝影機甚至可以加入影像辨識的功能,家中不同駕駛人坐在駕駛座上時,駕駛人一個字也不用說,車子便會自動調整駕駛座、後照鏡位置等等。軟體甚至可以設定成,如果有一個不熟悉的臉孔坐在駕駛座上,沒有經過許可,汽車將拒絕發

動。事實上這樣的系統在技術上都是可行的，只是其成本和複雜度都還太高，看起來短期內還沒有辦法被用在量產汽車上。

OK，再回到語音辨識的討論。提高語音辨識的準確度，是汽車聲控系統的第一個嚴苛的挑戰，語音辨識有 90% 的準確率都不夠高，畢竟聲控是要取代按鍵手控，而按鍵手控的準確率可是 100% 的。

討論到這裡，不曉得您是否還是覺得怪怪的。沒錯，汽車可以聽得懂您的指令了，不過您似乎也要開始學一套汽車的火星文，才能正確地下指令給汽車。

要求駕駛人記得幾十個甚至上百個特定指令，可能很令人分心、甚至是有些危險的。想像您正以一百公里的時速行駛，突然下起暴雨，您還要一面想著，打開雨刷的指令到底是雨刷、開雨刷、啟動雨刷、芝麻開門、還是什麼碗公？

更重要的問題是，我為什麼要學汽車的火星文？到底誰是老大啊？！汽車應該能夠聽懂我的話才對！

所以汽車語音辨識系統下一步發展，是要讓汽車能夠「聽懂」駕駛人自然的談話，能夠從上下文中闡釋駕駛人的意願。舉例來說，駕駛人說出「交通」二字，語音辨識系統必須能根據上下文判定駕駛人要是將收音機轉到交通電台，還是要知道附近交通的狀況？

這件事情技術上的困難度又高了一個層次。前面談到語音辨識的準確度，主要是要將使用者口述的指令和先前儲存的指令做比對，而要能夠真正聽懂駕駛人的語意，語音辨識軟體需要儲存大量的字彙，處理器需要處理結合了語言學和統計學、非常複雜的語言模型，才能闡釋口述語言的意思。另一項困難的地方，是聽懂駕駛人的指令之後，聲控系統還要能天衣無縫地在幾個系統之間切換控制，舉例來說，將收音機轉到交通電台是對車內娛樂系統的控制，提供附近交通的狀況，又是屬於導航系統的工作。

技術上難歸難，一步一步還是有進展，現今豪華車的導航系統像是 Acura 的 MDX，已經可以對不甚明確的指令，像是「找一家最近的加油站」，做出正確的回應。看到一篇報導對於 IBM 發展 VoiceBox 最新版軟體的測試，語音辨識軟體已經可以順暢地闡釋駕駛人的語意，在不同的系統之間切換控制。估計這種可以聽懂駕駛人談話內容先進的語音辨識系統，在一兩年之內可能就會出現。

也許五年後您買一部新車時，不必再拿著一本教科書一樣大小的車主手冊苦讀，嘗試了解新車各部的功能，而只是瀟灑地坐在駕駛座上，問這部汽車，“What can you do for me?”

會不會覺得，這樣您比較像是老大？！