【專欄】基於人工智慧之語音溝通輔具

2019-06-20 | 數理科學, 漫步科研, 科普專欄

第1693期

作者 / 曹昱（本院資訊科技創新研究中心副研究員）

語音溝通（包括聽話跟說話）是人與人最便利的訊息交流方式，人類語音的複雜度遠遠超過其他動物的溝通形式，也是人類能夠成就高度文明的重要因素。聖經上巴別塔的故事中說到：「看哪，他們都是一樣的人，說著同一種語言，如今他們既然能做起這事，以後他們想要做的事就沒有不成功的了」，更看出語音溝通對人類文明及科技發展的重要。因此當一個人失去了與他人的語音溝通能力，即使在科技發達的現代，生活還是會受到嚴重地影響。就聽覺而言，長年的聽損會造成年長者與他人產生隔閡，造成生活上的不便，失智風險亦隨之上升。學齡兒童的聽損，導致學業成績落後及與其同儕互動不良，對於兒童學習及社交能力發展具負面影響。就口語及發音而言，發聲構造異常或受損是言語清晰度降低（構音異常）最常見的成因。構音異常影響語者與他人的溝通效能及其自身生活品質。筆者近年的研究是以人工智慧演算法改進現有的語音溝通輔具，期能夠幫助語音溝通障礙者提升語音溝通效能、進一步改善其生活品質。接下來我們將分別介紹近年聽覺及發聲輔助科技的進展，以及本實驗室近年的工作。

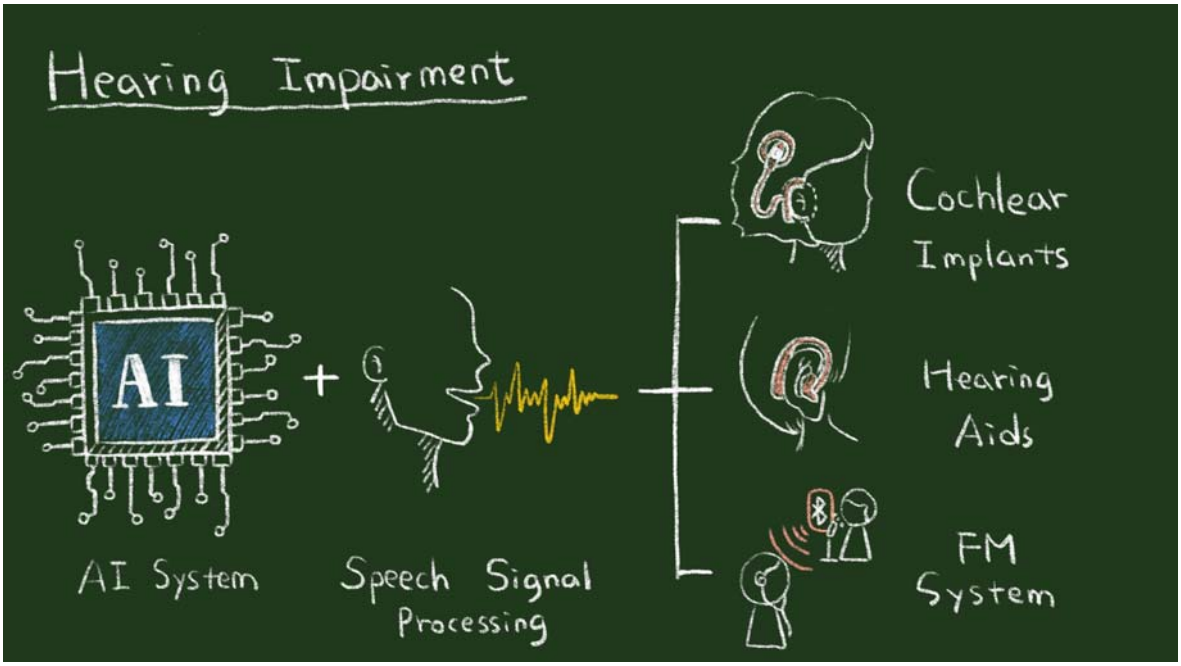
由於人類壽命增長、環境噪音頻繁出現、過度使用耳機，近年來聽損人數逐年升高，而聽損族群的年齡卻逐年下降。根據最新的國際報導指出，超過70歲以上的長者，2/3的人有不同程度的聽損，由於不易發覺，聽損初期通常會被忽略。就算是確定有聽損，由於聽覺輔具所費不貲，年長者常常會因為不願意花晚輩的錢而打消購買念頭。目前國際上有許多學者正在提倡降低聽覺輔具的價格，以提升聽障者購買及配戴聽覺輔具的意願。除了價格之外，聽覺輔具還有另一個重要的問題：噪音情境使用下效果不佳的問題。根據國際研究指出，有95%使用者指出當在有噪音情境下使用聽覺輔具時通常無法得到理想的聲音品質以及語音理解度。由於環境噪音（特別是與人類語音特性相近的噪音）通常難以準確估測；因此，要能夠有效消除雜訊是語音訊號處理相當棘手也是多年懸而未決的問題。好消息是，近年來人工智慧（特別是深度學習理論）的進步對解決這個問題帶來了曙光。

基於深度學習理論，學者們提出了多項新穎的語音訊號處理演算法應用於消除加乘式噪音、摺機式噪音（空間混響）、以及收音設及備通道不匹配問題，進而還原出高品質的語音訊號，讓聽者聽得更懂、聽得更舒服。相較於傳統機器學習模型，深度學習模型擁有更強的特徵參數抽取及非線性轉換能力，因此可以達成比傳統方法更佳的效能。近年來，筆者實驗室致力於開發基於深度學習模型的語音訊號處理演算法為基礎的聽覺輔具科技，包括FM無線調頻系統、助聽器、及人工電子耳。實驗結果證實，在各類聽覺輔具上面，基於深度學習模型的系統確實可以提供比傳統方法更優異的效能。目前這個研究方向除了在學研界逐漸受到重視以外，產業界也開始投入，然而在系統實踐上仍有一些困難需要克服。目前筆者實驗室努力的方向為：

1. 開發以任務導向的語音訊號處理技術：在機器學習中，訓練模型時所使用的目標函數會影響最終的分類或是迴歸能力。在基於深度學習的語音訊號處理架構下，我們也需要定義一個目標函數來訓練模型參數。傳統上，我們使用估測訊號以及乾淨訊號的均方差（Mean Squared Error，MSE）當作目標函數來訓練模型，然而MSE並不是基於人類的聽覺理解模型設計，因此以往的語音訊號處理系統未必能有效地提升語音理解度。在近期的研究中，我們嘗試基於聽覺理解模型來設計目標函數。實驗結果證實，設計出的語音處理系統可以同時提升正常聽力者以及聽損者的語音理解度。
2. 語音訊號處理模型壓縮技術：為了達到良好的效能，我們常使用複雜的深度學習模型來處理語音訊號，然而聽覺輔具運算資源通常有限，因此無法使用極複雜的深度學習模型。為克服此一限制，我們需要模型壓縮技術來簡化模型複雜度。近年已有許多研究團隊提出深度學習模型壓縮技術，而筆者團隊嘗試於

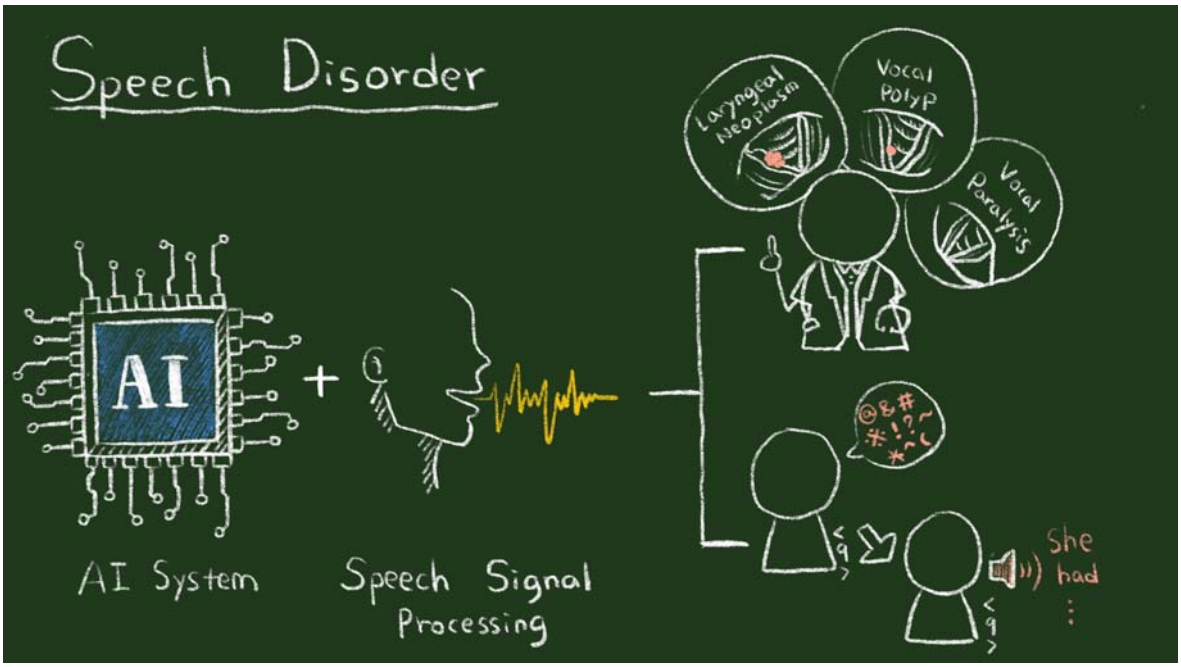
語音訊號處理的是 Computation-Performance Optimization （CPO） 壓縮技術，主要的設計概念是基於最後的效能動態消除深度學習模型的參數。另一種技術為Parameter Quantization （PQ），此技術是基於Quantization演算法減少參數的精度，藉以壓縮深度學習模型，同時加速線上運算效能。

3. 結合多模態之語音訊號處理技術：人與人的溝通包含口語與非口語的部分，發話端傳遞口語訊息時，收話端的聽者除了專注於聲音本身外，也接收相關的視覺訊息來協助了解語音的內容。一般而言，視覺資訊構築非口語的部分，包含語者說話時的發音動作、臉部表情以及肢體語言，在某些語音技術中，圖像及聲音訊號的結合能有效地幫助訊息傳遞以及人機介面的高效設置。由此發想，我們研究結合視覺與聲音訊號的方法，提出了新穎的多模態語音增強演算法，實驗結果證實相對基於單一模態（語音訊號）的系統，多模態語音增強系統能夠提供更佳的語音理解效能。



基於人工智慧之聽覺輔具

人類語音訊號是肺部送出的氣流，流經氣管、聲門，在口腔或鼻腔的中形成共振。發音過程中，需要唇、舌、齒等器官適時地阻斷氣流，產生不同語音。當這條路徑的任何一部份出現異常，皆會造成語音品質以及理解度下降。近年來，有越來越多針對發生異常的診斷以及改善語音的理解度的研究。針對構音異常診斷，通常是請說話者發出一個長母音或是產生一個句子，經由分析聲音的特性判斷語音正常與否。若為構音異常，則再進一步分析異常的類別。數項研究指出基於深度學習的判斷器能夠較傳統機器學習判斷器提供更佳的診斷效能。此外，由於職業、性別、年齡、生活習慣（是否有菸、酒習慣）對於發聲也會有影響，因此最近筆者團隊開發出一套新穎的多模態聲音診斷系統。此診斷系統是基於深度學習模型優異的融合功能，結合聲音以及病史兩種截然不同的資料型態。實驗結果證實此系統可以提供比單獨聲音為主的判斷系統提供更準確的診斷。除了發聲診斷系統，筆者團隊亦投入開發新穎的構音輔具技術，目標是提升構音異常人士的語音理解度，增進其與其他人的溝通效率。我們嘗試使用新穎的深度學習模型（包括非負矩陣分解以及對抗式生成模型）對構音異常語音增強，實驗證明能夠有效地提升口腔癌術後的語音理解度。未來我們將會繼續嘗試結合自動語音辨識、更新穎的語音訊號處理技術在構音異常輔助優化的任務上。



基於人工智慧之構音異常偵測及增強輔具

近年人工智慧技術大幅進步，各種新穎的技術大量應用在提高機器影像辨識器、語音辨識器、下棋及電玩、對話及問答系統上，確實讓很多任務的效能可以超越人類的能力。然而筆者認為，相對於追求開發超越人類的機器，我們或許可以運用人工智慧來發展輔具，提供給需要幫助的障礙者，這樣的研究或許能讓人工智慧的進步對人類社會更有實質上的助益，也讓科學研究更有溫度。

相關論文請參考：https://www.citi.sinica.edu.tw/pages/yu.tsao/publications_en.html