



台灣聽力語言學會電子學報

The Speech-Language-Hearing Association, Taiwan

- 主題文章：AI 於人工電子耳應用之經驗分享－『噪音消除』
- 撰 稿 者：賴穎暉



主題文章

AI 於人工電子耳應用之經驗分享－『噪音消除』

賴穎暉

國立陽明大學生物醫學工程學系

● 人工智慧(artificial intelligence, AI)發展

人工智慧(artificial intelligence, AI)於近年已不斷的被證明在各種應用領域(例如:醫學[1, 2]、影像辨識[3]、語音辨識[4]...等)有優異的信號處理能力。它的概念最早於1956年達特茅斯會議被提出，其精神是讓電腦能具備人類的知識與行為，並有:學習、推理、判斷、知識儲存記憶、瞭解人類語音與行為之能力。在眾多的人工智慧技術中，目前又以深度學習(deep learning)理論最為熱門。

深度學習主要模仿生物大腦運作方式所設計之演算法(algorithm)，它讓資料透過多層(layer)神經網路來自動抽取具代表性之特徵(feature)，進而透過所抽取到之特徵來提升機器識別能力。圖1為一個典型之深層類神經網路(deep neural network, DNN)架構。當輸入資料(例如:語音)經過處理後，這些資料(如圖1中的 x_1, x_2, \dots, x_N)將送入DNN中的隱藏層(hidden layer)進行運算(i.e., 權重之乘法運算及偏差值調整)。而上述之神經網路中的權重 w (w_1, w_2, \dots, w_N)及偏差值參數 b (b_1, b_2, \dots, b_N)主要是透過給定任務之訓練資料予以訓練得來(詳細之訓練技術可參考[5, 6])。計算完成之結果(z)將送入激發函數 $\sigma(-)$ 再進行計算，進而獲得該神經元於隱藏層計算後的結果" a "。隱藏層1中所有計算出之結果(a_1, a_2, \dots, a_N)將

會作為隱藏層2之輸入值並重復上述之步驟。最後，隱藏層 L 值將送到輸出層而完成所有計算。當完成運算後，此神經網路將會依照使用者於訓練階段所給定之任務(例如:語音辨識、聲音情境辨識及語音回歸...等)給予最終之輸出結果來幫助人們進行辨識及信號處理任務。除了DNN架構外，近年也有多種網路結構不斷的被提出來，例如:卷積神經網路(convolutional neural network, CNN)[7]、遞歸神經網路(recurrent neural networks, RNN)[8]及長短期記憶神經網路(long short-term memory, LSTM)[8, 9]...等。而上述這些網路結構均有其優點並在不同的應用情境下有所優勢。

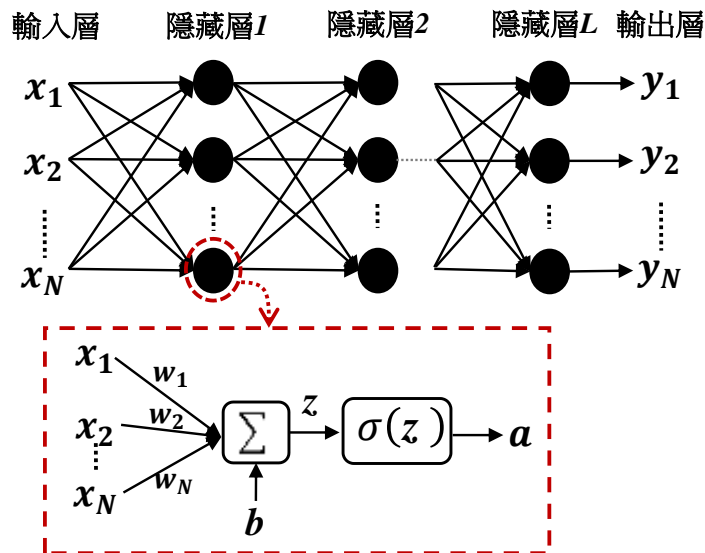


圖1、深層類神經網路架構

人工智慧的成功，除了演算法不斷進步外，巨量資料(*big-data*)也是關鍵因素之一。由於網際網路、社群軟體的發展，使得巨量資料(例如:語音、音樂、圖片、影像...等)不再難以取得，進而能提供更豐富之資料讓電腦進行學習。此外，電腦的運算能力(*computing-power*)也是近年人工智慧技術成功的重要一環。舉例來說，透過電腦運算能力的提升讓過去可能需要訓練超過一年以上之網路模型可縮短至幾天就完成，進而加快研究進展。總結來說，人工智慧技術的成功主要受到上述A(*algorithm*)、B(*big-data*)及C(*computing-power*)等三方面的研究基礎所支持。因此，對於期望採用人工智慧技術進行研究之聽語人員，建議可從此ABC三方面來著手思考及深入研究。

● 人工電子耳面臨的重要挑戰之一：噪音

人工電子耳是全聾患者目前重要的聽覺輔具之一，其原理是將所收到的語音信號透過語言處理機(*speech processor*)進行處理，處理後之信號透過傳送器(*transmitter*)送至植入耳蝸中的電極中以刺激聽覺神經，進而讓全聾患者重新聽見聲音。在過去已有許多信號處理方法被提出(例如:編碼策略、噪音消除演算法...等)以提升人工電子耳使用者的語音聽辨能力。而在近期的研究成果指出(如圖2)[10]，當全聾患者植入人工電子耳後並配合適當的復健，將能有效的提升使用者於安靜環境下之語音聽辨能力;但於噪音環境聆聽條件下，仍具有有很大的改善空間。例如圖2中的單字詞於噪音(*Word Noise*) 測試情況，10

位受測者的語音理解力均明顯下降(i.e.,相較於Word Quiet)。

為能有效的提升人工電子耳患者於噪音情況下之聆聽效益，可透過噪音消除 (noise reduction) 技術來改善。噪音消除的目的是消除語音訊號中的雜訊(noise)成份並提升語音的訊噪比(signal-to-noise ratio, SNR)，進而增進語音的理解度(speech intelligibility)、語音品質(speech quality)及降低聆聽負擔(listening effort) [11]。然而目前於人工電子耳裝置中常見之噪音消除演算法，經臨床測試結果發現仍有進步的空間[12]。

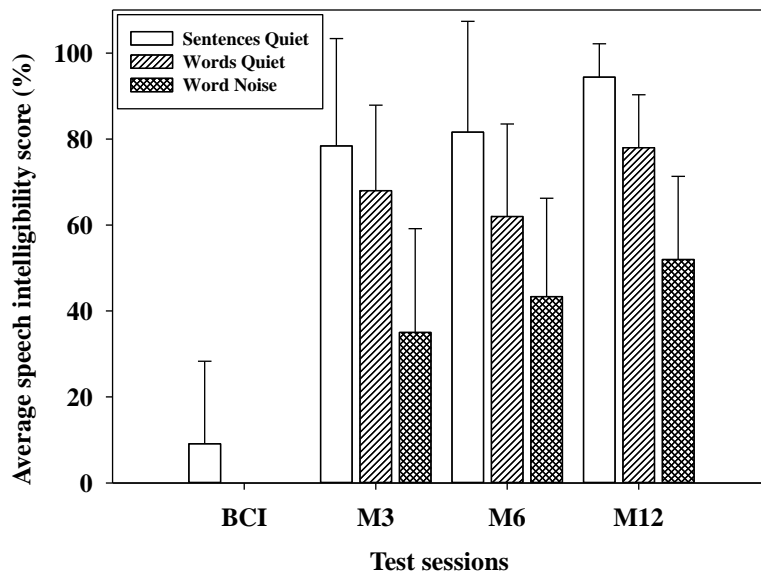


圖2、10位人工電子耳使用者於不同階段下之語音聽辨能力。X軸中的BCI表示在植入電子耳前之平均語音聽辨能力；M3、M6及M12分別表示植入3、6及12個月後之平均識能力、Y軸表示語音聽辨能力[10]。

● 以AI為核心之噪音消除技術來提升人工電子耳使用者語音理解力表現

近年，以深度學習架構為基礎之噪音消除方法(例如:[13-16])不斷的被證明它能比傳統方法有更佳之語音增強能力。有鑑於此，我們團隊基於DNN架構來設計噪音消除演算法(稱deep denoising autoencoder, DDAE[13])，首先透過Vocoder模擬[17]證明此架構對人工電子耳患者之效益[18]。隨後，我們將DDAE結合情境辨識單元再提出noise classifier (NC) + DDAE (NC+DDAE)[19]之噪音消除架構，來更進一步提升人工電子耳使用者於噪音下之聆聽效益。

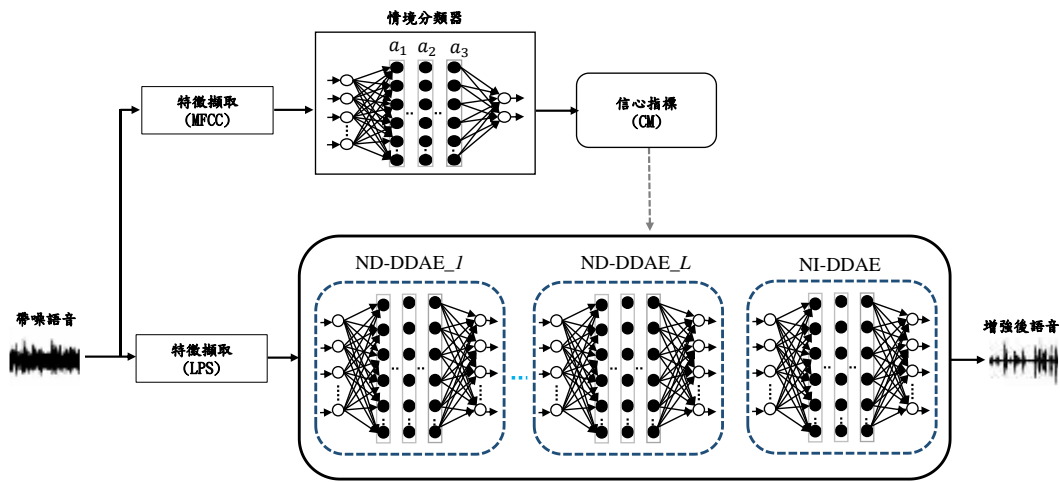


圖3、本團隊提出之NC+DDAE語音增強方法[19]

圖3為NC+DDAE噪音消除法信號處理流程圖。此架構運用神經網架構對當前環境噪音進行分類處理，進而再以對應之回歸分析(Regression)神經網路達成噪音消除的功能。更具體地來說，我們透過一個情境分類器進行背景噪音辨識，進而從分類結果在多個模型池中選擇最佳模型(如圖3中的ND+DDAE_L 或 NI+DDAE)進行語音增強以進一步提高效率。圖4呈現出本研究對九位CI使用者進行臨床聽辨力測試之實驗平均結果，X軸代表測試條件，Y軸代表詞語正確率。於圖4之實驗結果我們觀察到，提出系統(NC+DDAE)皆比未處理前之語音(noisy)及傳統噪音消除法(e.g., logMMSE [20], KLT[21-23])處理後有更佳的語音理解力表現。更具體地說，提出方法(NC+DDAE)最高能提升CI使用者之語音理解度表現相較於傳統方法（即logMMSE與KLT）在2T_0dB噪音條件下提升了41.7%，在2T_5dB噪音條件下提升了34.2%，在CJ_0dB噪音條件下提升了48.4%，在CJ_5dB噪音條件下提升了34.1%。由上述之實驗結果可以觀察到，以人工智慧技術為基礎之噪音消除技術能有效的提升人工電子耳於噪音情境下之語音聽辨能力表現。而這結果也直接證明AI於聽覺輔具上的應用潛力。註:詳細之NC+DDAE技術說明，可以參考[19]。

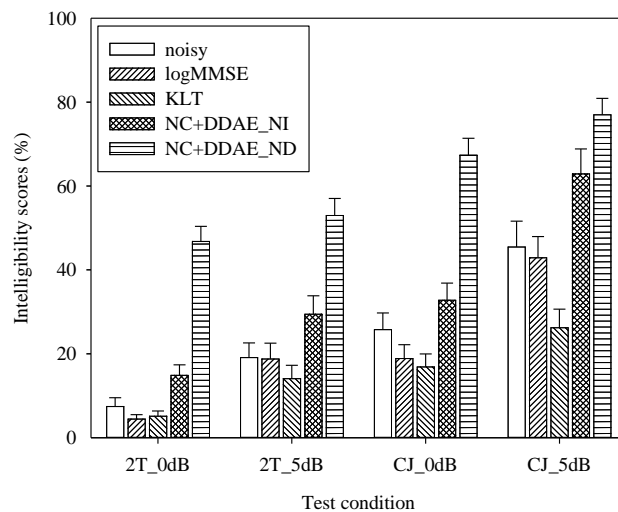


圖4、臨床實驗評分結果。註: 2T_0dB(雙人人聲噪音於訊噪比0dB)、2T_5dB(雙人人聲噪音於訊噪比5dB)、CJ_0dB(建築噪音於訊噪比0dB)及CJ_5dB(建築噪音於訊噪比5dB)等測試情況。

● 結論

本文章簡介我們團隊近期以AI技術為核心所開發之噪音消除法，並應用於人工電子耳裝置中來提升語音理解力之成果。由臨床研究成果證明，AI技術能顯著的改善人工電子耳使用者於噪音情境下之語音聽辨能力。而這也代表著AI技術未來在人工電子耳或其它聽覺輔具未來應用之可能性。最後也希望透過此例子之分享，讓聽語專業人員未來於AI技術於臨床應用上有所幫助。

● 感謝

在此感謝中研院資創中心曹昱博士、振興醫院力博宏主任及所有參與此計畫的研究人員們共同合作及努力，將三方專長（醫學工程、資訊科學及臨床醫學）予以整合，來完成此項研究成果。此外，也要感謝楊順聰教授過去在聽語科學等相關專業知識的啟發，才能讓本人能持續深入該領域之研究。最後，也感謝家人無怨無悔的包容及支持，讓我能全心全力進行研究工作。協助聽障人士改善聆聽效益是我持續研究的方向，本人也期許自己更加努力，並以臨床需求為核心，來幫助聽障人士重拾聽覺自信。

◇ 參考文獻

- [1] A. Esteva *et al.* (2017), "Dermatologist-level classification of skin cancer with deep neural networks," vol. 542, no. 7639, p. 115.
- [2] K. Wiggers. (2018). "Google AI claims 99% accuracy in metastatic breast cancer detection," Available: <https://venturebeat.com/2018/10/12/google-ai-claims-99-accuracy-in-metastatic-breast-cancer-detection/>
- [3] K. He, X. Zhang, S. Ren, and J. Sun (2016), "Deep residual learning for image recognition," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 770-778.
- [4] G. Hinton *et al.* (2012), "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97.
- [5] G. E. Hinton, S. Osindero, and Y.-W. The (2006), "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527-1554.
- [6] Y. LeCun, Y. Bengio, and G. J. n. Hinton (2015), "Deep learning," *Nature*, vol. 521, no. 7553, p. 436.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton (2012), "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097-1105.
- [8] A. Graves, A. R. Mohamed, and G. Hinton (2013), "Speech recognition with deep

- recurrent neural networks," *in Proc. ICASSP*, pp. 6645-6649.
- [9] S. Hochreiter and J. J. N. c. Schmidhuber (1997), "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780.
- [10] D. Borger *et al.* (2015), "One-year follow up of auditory performance in post-lingually deafened adults implanted with the Neurelec Digisonic® SP/Saphyr® Neo cochlear implant system," *Audiology research*, vol. 5, no. 2, pp. 139.
- [11] I. Brons, R. Houben, and W. A. Dreschler (2014), "Effects of Noise Reduction on Speech Intelligibility, Perceived Listening Effort, and Personal Preference in Hearing-Impaired Listeners," *Trends in hearing*, vol. 18, pp.10.
- [12] F. Chen, Y. Hu, and M. Yuan (2015), "Evaluation of Noise Reduction Methods for Sentence Recognition by Mandarin-Speaking Cochlear Implant Listeners," *Ear and hearing*, vol. 36, no. 1, pp. 61-71.
- [13] X. Lu, Y. Tsao, S. Matsuda, and C. Hori (2013), "Speech enhancement based on deep denoising autoencoder," *in Proc. Interspeech*, pp. 436-440.
- [14] A. Narayanan and D. Wang (2013), "Ideal ratio mask estimation using deep neural networks for robust speech recognition," *in Proc. ICASSP*, pp. 7092-7096.
- [15] Y. Xu, J. Du, L. R. Dai, and C. H. Lee (2015), "A Regression Approach to Speech Enhancement Based on Deep Neural Networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 7-19.
- [16] J. C. Hou, S. S. Wang, Y. H. Lai, Y. Tsao, H. W. Chang and H. M. Wang (2018), "Audio-Visual Speech Enhancement using Multimodal Deep Convolutional Neural Networks," *IEEE Transactions on emerging topics in computational intelligence*, vol. 2, pp. 117-128.
- [17] F. Chen and A. H. Lau (2014), "Effect of vocoder type to Mandarin speech recognition in cochlear implant simulation," *in Proc. ISCSLP*, pp. 551-554.
- [18] Y.-H. Lai, F. Chen, S.-S. Wang, X. Lu, Y. Tsao, and C.-H. Lee, "A Deep Denoising Autoencoder Approach to Improving the Intelligibility of Vcoded Speech in Cochlear Implant Simulation (2017)," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1568-1578.
- [19] Y.-H. Lai *et al.*, "Deep Learning based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients (2018)," *Ear and Hearing*, vol. 39, no. 4, pp. 795-809.
- [20] Y. Ephraim and D. Malah (1985), "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 443-445.
- [21] A. Rezayee and S. Gazor (2001), "An adaptive KLT approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 2, pp. 87-95.
- [22] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise (2003)," *IEEE Transactions on Speech and Audio Processing*,

vol. 11, no. 4, pp. 334-341.

- [23] P. C. Loizou, A. Lobo, and Y. Hu (2005), "Subspace algorithms for noise reduction in cochlear implants," *The Journal of the Acoustical Society of America*, vol. 118, no. 5, pp. 2791-2793.



編輯

發行單位：台灣聽力語言學會

發行人：葉文英

主編：曾尹霆

編輯顧問：曾進興

助理編輯：陳奕秀

網址：www.slh.org.tw

發行日期：2019.04.01

聽語學報：第八十五期

副主編：吳詠渝、陳孟好、席芸、
姚若綺、鄭秀蓮、王靖嵐、
李善祺、薛偉明

美術編輯：李善祺