The 2017 Conference on Computational Linguistics and Speech Processing
ROCLING 2017, pp. 287-294

以語音能量特性發展即時語速偵測裝置-前導型研究

Real-time monitoring device of phonation speed and volume based on speech energy: A pilot study

王榮德 Chi-Te Wang
亞東紀念醫院耳鼻喉科 主治醫師
Department of Otolaryngology
Far Eastern Hospital
drwangct@gmail.com

林峯全 Feng-Chuan Lin
亞東紀念醫院耳鼻喉科 語言治療師
Department of Otolaryngology
Far Eastern Hospital
autoioio@gmail.com

鄭惟中 Wei-Zhung, Zheng
元智大學電機工程學系
Department of Electrical Engineering
Yuan-Ze University
s1010654@ee.yzu.edu.tw

方士豪 Shih-Hau Fang
元智大學電機工程學系
Department of Electrical Engineering
Yuan-Ze University
shfang@saturn.yzu.edu.tw

曹昱 Yu Tsao
中央研究院 資訊科技創新研究中心
Research center for information technology innovation
Academia Sinica
yu.tsao@citi.sinica.edu.tw

賴穎暉 Ying-Hui Lai
國立陽明大學生物醫學工程學系
Department of Biomedical Engineering
National Yang-Ming University
yh.lai@ym.edu.tw

摘要

嗓音問題如聲帶結節、息肉等，是現代社會中十分常見的健康疾病。常見的危險因子包括性別(女性)、用聲習慣(過度或不當使用)、環境噪音(背景值 65 分貝以上)及個人特質(如 A 型人格)等。其中，又以錯誤的用聲習慣為疾病常見起因。過去的研究指出，適度的給予患者在錯誤語速(或音量應)產生時給予提醒，將能有效的提升臨床治療效益。有鑑於此，本計畫提出一套低運算需求之即時嗓音監測系統來幫助患者在不當用聲時，給予患者即時之提醒(例如振動、閃燈)。計畫提出之系統包括:(1)語音訊號預處理、(2)噪音消除、(3)語音能量胞絡線偵測、(4)動態發聲閾值調整及(5)即時回饋等五個部份。由實驗結果證明，本研究所發展之系統於噪音情境下之語速偵測準確率可達到 95.4%。此外，由於本研究所提出之系統運算需求量小，未來將會以微型化為目標將其實踐於嵌入式系統中以方便於臨床治療之應用。

關鍵詞：語音訊號預處理、噪音消除、語音能量胞絡線偵測、動態發聲閾值調整

一、緒論

嗓音異常是教師常見之職業疾病[1]。根據過去的研究顯示，教師出現嗓音異常的比率明顯高於非教師，且在症狀的程度上也較為嚴重[2,3]。近年，以問卷的方式調查美國Iowa州教師嗓音異常的盛行率，結果發現在554位中小學教師中，自覺有嗓音異常的比率為32 %，顯著高於非教師的1 %；其中有60 %的教師提到在過去的一年中，曾經因為工作出現嗓音異常的情形，其中又以嘶啞聲、嗓音疲憊等最常出現。此外，Roy等學者於2004[4]調查美國猶他州 (Utah) 和愛荷華州 (Iowa)的1243位教師和1288位非教師的嗓音狀況，結果發現教師嗓音異常之盛行率顯著高於非教師。此外，Śliwińska-Kowalska等學者於2006[5]研究425位教師嗓音異常的盛行率，結果也發現到教師發生嗓音異常的機率是非教師的二至三倍，症狀也較為嚴重。由上述幾項研究可發現，教師嗓音異常的盛行率顯著高於非教師，而嗓音異常症狀也較非教師多且嚴重。有鑑於此，教師的嗓音保健及治療將是一個重要之研究課題[2-4]。

教師嗓音異常的原因以嗓音誤用(或濫用)為主 [4,6]，係指長時間說話且語速過快、

於背景噪音下大聲說話等行為[6]。以語音信號的角度來看，便是指我們常聽見之語速頻率過高及語音能量振幅過大。更具體的來說，嗓音誤用（vocal misuse）係指不正確的發聲習慣，如提高說話音調、清喉嚨或不正確的呼吸方式等行為[6]。這些錯誤的發聲行為是造成嗓音異常最主要的原因 [1,6]。Preciado 等學者於 2005 [7]對 579 名嗓音異常教師和 326 名嗓音正常教師進行嗓音研究。結果發現，嗓音異常教師的濫用行為比無嗓音異常教師多(i.e.,74.8%比 67.1%)。換言之，嗓音濫用將是此疾病重要的發病原因之一[8,9]。

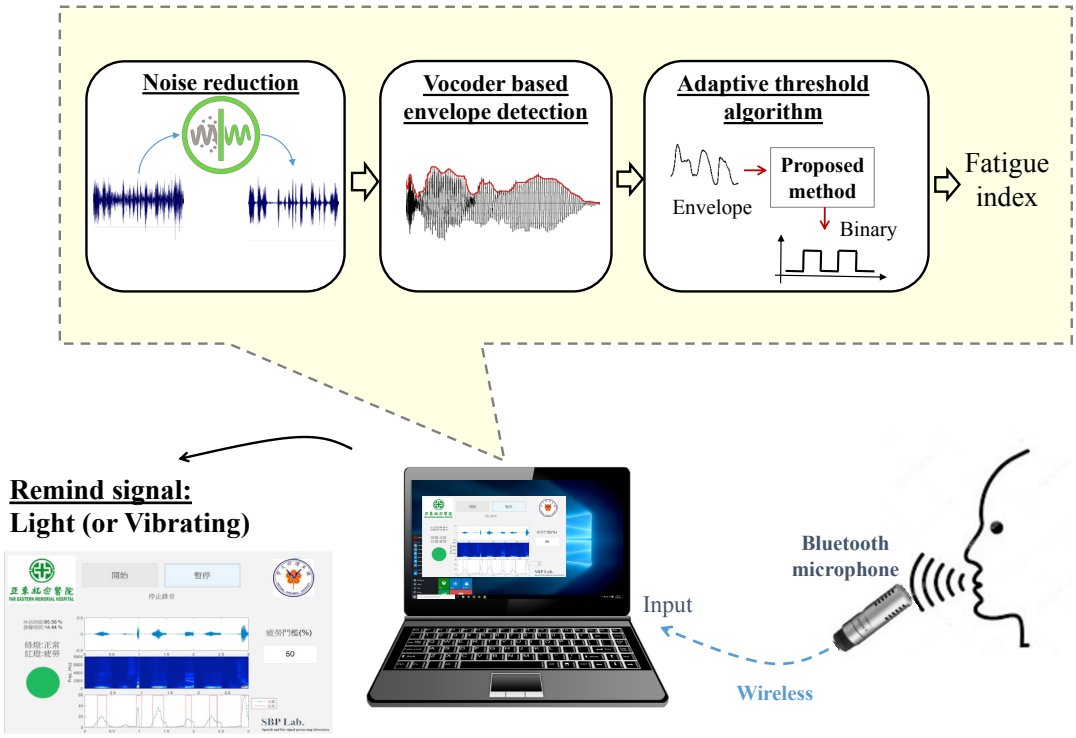
臨床上為有效的幫助嗓音異常患者能獲得有效之治療，最為常見的方法是進行嗓音治療訓練，以減少錯誤的發聲機會。根據臨床觀察，嗓音治療在類化到日常生活中容易發生困難。換言之，患者雖然能在治療期間正確的使用嗓音，但離開了治療的場域將會不自主的回復到錯誤的嗓音應用情況。因此，讓患者走出治療室後也能持續正確應用嗓音將是最為根本之治療方法。有鑑於此，近年已開始著手於音量監控之裝置開發研究。Van Stan 等將音量監控設備『Ambulatory voice biofeedback』應用在聲帶結節的患者，讓使用者在治療室以外的地方配合使用，以控制音量過大的情況[14]。結果顯示，透過平時不斷的協助患者控制平日說話音量及語速，將會顯著的提升嗓音治療成效。然而上述之方法應用於臨床治療仍有困難(例如成本較高且不易隨身攜帶)。此外，當患者使用環境處於較挑戰時(例如:環境噪音不斷變動)，其方法仍有很大的進步空間。有鑑於上述之問題，本研究提出一套以語音能量特性為基礎之即時語速及音量偵測暨回饋系統(詳細技術可參考下)，來幫助嗓音異常患者在日常生活或工作場合中之適當調整發聲習慣，以增進臨床治療成效。

二、方法

(一) 系統架構

本研究提出一套以語音能量特性為基礎之即時嗓音監測系統來幫助患者在不當用聲時(i.e., 語速過快及音量過高)，給予即時之提醒以提升臨床治療效果。而本研究所提出之訊號處理流程如下圖一所示。由於患者所處之環境往往都容易存在許多噪音(例如:冷氣、電冰箱、電視…等)，而這些噪音也會直接的影響語速偵測之準確性。有鑑於此，本研究提出之系統將採用非監督式噪音消除法(i.e., logMMSE[11])做為前端信號處理以消除噪音。接著，處理後之語音將透過語音能量特性進行語音能量胞絡線提取動作。此外，本系統提出一個適應性調動值演算法(adaptive threshold algorithm)來依據使用者所

處之環境適時的調動偵測閾值。進而透過此閾值與語音胞線信號間之關係轉換成二元編碼資訊(i.e., binary code)來預估其輸入信號是否為語音成份。接下來，我們更進一步的把這二元編碼資訊轉換成臨床所需之嗓音疲勞指標(i.e., fatigue index)。註:此嗓音疲勞指標將透過醫師依照患者之病情進行參數設定，進而讓患者能即時的進行個人化之語速偵測(i.e., 超速與否)。當患者的語話過快時，系統將會即時的提出警示信號來提醒患者減慢語說速度以提升臨床嗓音復健之治療效益。



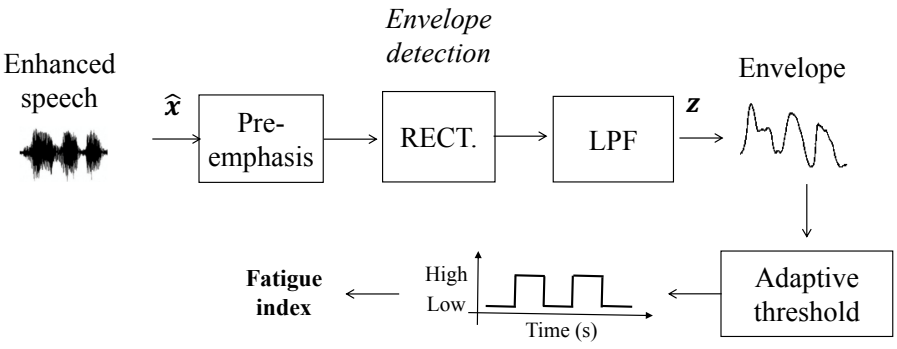
圖一、訊號處理流程

本研究將採用之演算法概念如圖二所示。當一個帶有噪聲之語音被上述的噪音消除法處理後(i.e., enhanced speech)，此信號$\hat{x}$將會道先透過 pre-emphasis 處理。接著我們採用整流器(i.e., rectifier)之概念對語音信號之上半部訊息保留後，接下來再採用一個低通濾波器(i.e., LPF)進行語音低頻信號之保量。註:語音胞絡線低頻部份為人類發聲時振動之基頻信號，而也是臨床上用來判別聲音動作與否之重要指標。而其各頻帶間(i.e., 低至高頻)之權重關係將會依據華語語言特性進行權重調整。接著，取出之胞絡線信號將基於一個動態調整閾值(i.e., adaptive threshold, AT)來將此胞線線轉換成方波信號進行單

位時間下之語音與否之預估。其動態調整閾值之設計方法如下(1)式:

$$\text{AT} = ax_1 + b^2x_2 + c^3x_3 \quad (1)$$

其中x_1、x_2及x_3分別是當下音框及前兩個音框資訊，而 a、b 及 c 分別是優化參數。註:此三個優化參數我們採用基因演算法[10]進行最佳化參數搜尋，以優化對每一個音框所用之 AT 來提升系統的準確性。接著，我們將透上述說明之二元編碼方法進行單位時間中之語音速度百分比，並再將此資訊轉換成臨床所需之"fatigue index"以做為患者語速是否符合醫師建議之評估依據。當患者語速過高時，系統將透過燈號(或振動)適時之提醒患者，以達到治療之目的。



圖二、語速偵測之信號處理流程圖

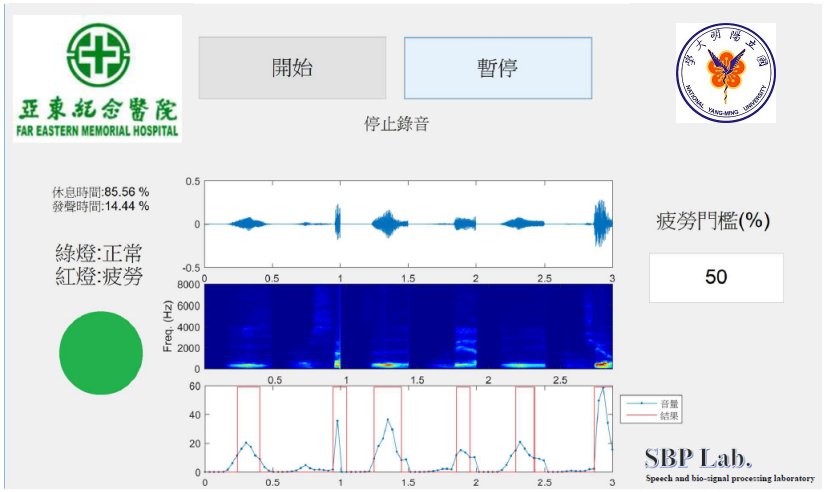
(二) 實驗設計與流程

本實驗之語句的語音與非語音標記採人工標記，每一句以人工將有語音及沒語音之百分比，共 20 句，每句 10 秒，所混之噪音類型為 SSN、訊噪比為 0dB，共分成 3 組做測試，(1)為音訊未套用 Noise Reduction(NR)便直接代入固定閾值公式進行語音判別之百分比，(2)為音訊經 NR 後代入固定閾值公式進行語音判別之百分比 (註: logMMSE 噪音消除法於此研究被採用 [11]);(3)為音訊經 NR 後代入由基因演算法(GA)最佳化後的動態閾值(AT)公式進行語音判別之百分比，之後將這三組所估出的閾值百分比與人工所標示的答案做比對，即可計算出本研究所提出之架構對於語音判別的準確率，詳細結果請參見圖四。

三、結果與討論

(一) 系統介面

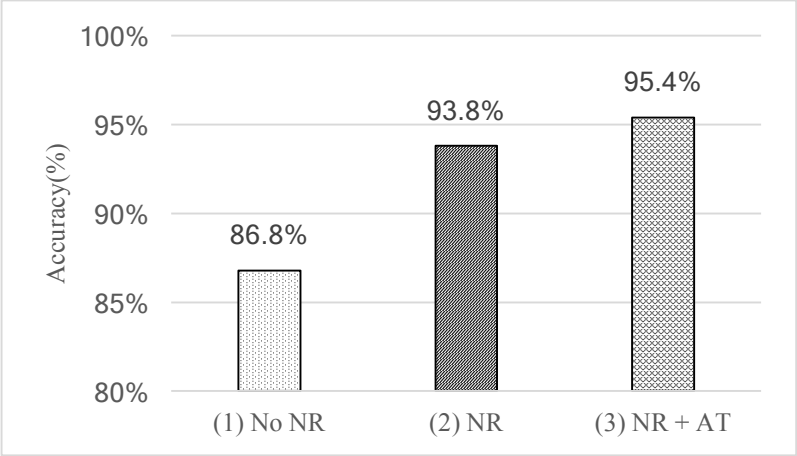
圖三為本研究所開發出系統介面，此系統已能讓臨床醫師進行疲勞門檻參數設定，並讓患者能即時的進行個人化之語速偵測(i.e., 超速與否)。更具體的來說，臨床醫師可以依照患者的嗓音傷害情況來個別化的調整疲勞門檻參數設定。當患者的語話過快時而超過此疲勞門檻時，系統將會即時的產生警示信號(綠燈：正常、紅燈：疲勞)來提醒患者減慢語說速度以提升臨床嗓音復健之治療效益。此外，本系統也提到視覺化之介面讓使用者(或家人)能即時自我觀察患者當前之嗓音使用情況(i.e., 時域及頻域語音信號的視覺化)，本系統亦有使用期間內整體嗓音之休息比例及使用比例的數值化呈現，可以讓使用者們可以更即時的掌握嗓音使用率的掌控。



圖三、本研究中所開發出即時語速偵測系統介面圖(Matlab 軟體實現)

(二) 語音識別率

圖四為本研究之語速偵測實驗結果，X 軸表示三種不同信號處理方法(i.e.,未使用 NR、採用 NR 及採用 NR+AT 之動態閾值調整法)。實驗結果發現，患者語音於噪音情境下(i.e., SSN 噪音類型、0dB SNR level)，本系統在未使用 NR 處理時的準確率為 86.8%; 當採用噪音消除法時，系統準確率為 93.8%; 最後本系統採用 NR+AT 方法時，系統準確率為 95.4%。由此結果我們觀察到以下幾項結論:(1)噪音消除法將能有效的提升本系統於噪音環境下之預估能力、(2)本研究提出之 AT 動態閾值調整法能更進一步的提升本系統之語速偵測準確性。



圖四、語音偵測準確率(%)

四、結論

於本研究之結果證明，噪音消除法能有效的提升以語音能量為基礎之語速偵測能量。換言之，一個良好的噪音消除演算法將能使本系統有更好的表現。近年 Lu 等學者[12, 13]提出一套監督式噪音消除法，稱為深層降噪自動編碼演算法（DDAE）。其主要運用深類類神經網路訓練架構進行噪音消除任務。於過去之研究也證明此方法能比傳統之噪音消除法有更佳之效益，因此未來也將嘗試採用此新式架構來提升本系統之語速偵測能力。

參考文獻

[1] Stemple, J. C., Glaze, L. E., & Klaben, B. (2010). *Clinical voice pathology: Theory and management.* San Diego, CA: Plural Publishing.

[2] Smith, E., Gray, S. D., Dove, H., Kirchner, L., & Heras, H. (1997). Frequency and effects of teachers' voice problems. *Journal of Voice, 11*, 81-87.

[3] Smith, E., Lemke, J., Taylor, M., Kirchner, H. L., & Hoffman, H. (1998). Frequency of Voice Problems Among Teachers and Other Occupations. *Journal of Voice, 12*(4), 480-488.

[4] Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D., & Smith, E. M. (2004).

Prevalence of voice disorders in teachers and the general population. *J Speech Lang Hear Res, 47*(2), 281-293. doi:10.1044/1092-4388(2004/023)

[5] Śliwińska-Kowalska, M., Niebudek-Bogusz, E., Fiszer, M., Łoś-Spychalska, T., Kotyło, P., & Sznurowska-Przygocka, B. (2006). The prevalence and risk factors for occupational voice disorders in teachers. *Folia Phoniatrica et Logopaedica, 58*(2), 85-102.

[6] Boone, D. R., McFarlane, S. C., Von Berg, S. L., & Zraick, R. L. (2013). *The voice and voice therapy (9th ed.)*. Boston, MA: Allyn & Bacon.

[7] Preciado, J., Pérez, C., Calzada, M., & Preciado, P. (2005). Function vocal examination and acoustic analysis of 905 teaching staff of La Rioja, Spain. *Acta otorrinolaringológica española, 56*(6), 261-272.

[8] Duffy, O. M., & Hazlett, D. E. (2004). The impact of preventive voice care programs for training teachers: a longitudinal study. *J Voice, 18*(1), 63-70. doi:10.1016/S0892-1997(03)00088-2.

[9] Södersten, M., Granqvist, S., Hammarberg, B., & Szabo, A. (2002). Vocal behaviour and vocal loading factors for preschool teachers at work studied with binaural DAT-recordings. *Journal of Voice, 16*, 356-371.

[10] Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*: MIT press.

[11] Ephraim, Y., & Malah, D. (1985). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 33*(2), 443-445.

[12] Lu, X., Tsao, Y., Matsuda, S., & Hori, C. (2013). *Speech enhancement based on deep denoising autoencoder.* Paper presented at the Interspeech.

[13] Lu, X., Tsao, Y., Matsuda, S., & Hori, C. (2014). *Ensemble modeling of denoising autoencoder for speech spectrum restoration.* Paper presented at the Interspeech.

[14] Van Stan, J. H., Mehta, D. D., Sternad, D., Petit, R., & Hillman, R. E. (2017). *Ambulatory Voice Biofeedback: Relative Frequency and Summary Feedback Effects on Performance and Retention of Reduced Vocal Intensity in the Daily Lives of Participants With Normal Voices*. J Speech Lang Hear Res, 60(4), 853-864.