

# 應用聲學韻律學及語意標籤之多分類器於情感語音之情緒辨識

吳宗憲\*、梁維彬

國立成功大學資訊工程系

[chunghsienwu@gmail.com](mailto:chunghsienwu@gmail.com)

IEEE Trans. Affective Computing, VOL. 2, NO. 1, JANUARY-MARCH 2011, pp. 10-21.

## 【105年科技部傑出研究獎】得獎人專刊

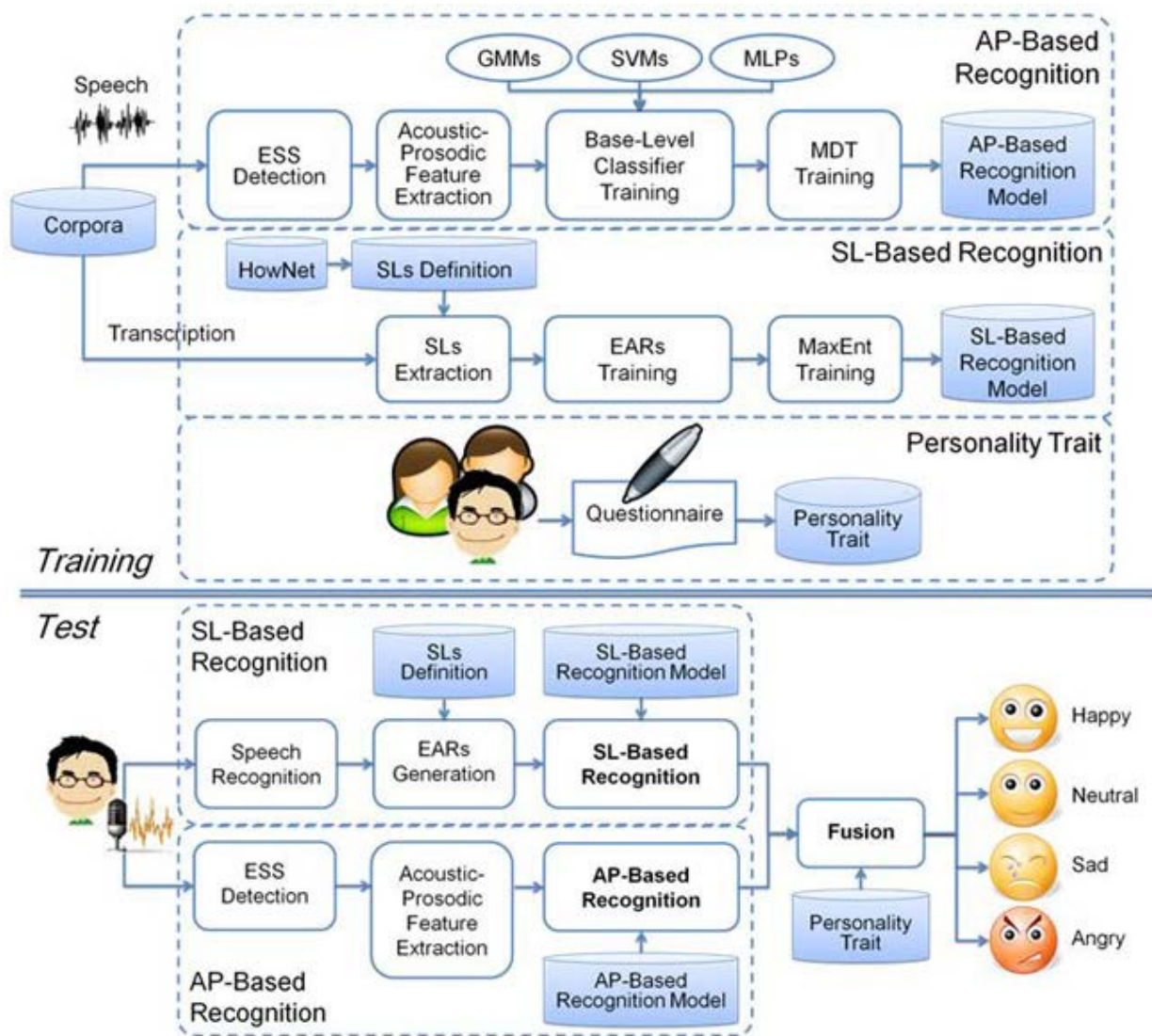
### 語

音為人類溝通最自然且最基本之方式。近年來，口述對話系統已成功應用在多個領域上。然而其應用仍局限在簡單資訊對話系統上，例如導航及旅遊資訊系統等<sup>[1]</sup><sup>[2]</sup>。為了能將其應用在較複雜之對話系統(例如家庭照護<sup>[3]</sup>教育/訓練，及聊天<sup>[4]</sup>)，情感互動將是一重要成分。然而要透過語音做情感互動，在技術上仍存在相當多之問題。本研究提出一應用聲學韻律學及語意標籤之多分類器於情感語音之情緒辨識。



圖一為應用聲學韻律學及語意標籤之多分類器於情感語音之情緒辨識之架構圖。在此研究中，對於採用聲學韻律學參數之部分，首先我們取出輸入語音中之情緒重要區段。而後利用此區段之參數建立高斯混合模型(GMM)，支持向量機模型(SVM)及多層感知模型(MLP)當成基礎辨識器。而後利用超決策樹來結合此三辨識模型。另一方面，應用語音辨識後之文字訊息，建立一語意標籤模型。此語意標籤乃透過中文知識庫知網(HowNet<sup>[5]</sup>)來建立。而後利用文字探勘技術擷取出情緒關聯規則(EAR)。最後，我們利用最大熵模型<sup>[6]</sup>來輸出情緒辨識結果。最後整合兩種辨識器，以得到最後情緒辨識結果。此外我們亦考慮人格特質之因素以提高情緒辨識之效能。

在效能評量中，我們蒐集了2,033 情緒語句包含中性，快樂，生氣及悲傷四種情緒。評量結果如表一所示。根據艾森克人格問卷，語者A(Speaker A)是外向型，辨識結果顯示其在高興及生氣情緒有較強之表現。語者B(Speaker B)既非外向亦非內向，評量結果顯示在四種情緒之表現差異不大。從評量結果可驗證本研究所提方法有其可靠性。總結而言，本系統在考量人格特質之條件下，其辨識準確率可達85.79%，再次肯定本方法之有效性。



圖一: 應用聲學韻律學及語意標籤之多分類器於情感語音之情緒辨識之架構圖

表一: 應用聲學韻律學及語意標籤之多分類器於情感語音之情緒辨識之評量結果

	MDT+MaxEnt ( $\lambda_{AP} = 0.4$ ) (Accuracy %)				
	Neutral	Happy	Sad	Angry	Average
Speaker A	75.80%	85.97%	83.35%	87.81%	83.23%
Speaker B	80.40%	81.61%	88.49%	84.93%	83.86%
Copus B	78.10%	83.79%	85.92%	86.37%	83.55%
	Proposed (MDT+MaxEnt+PT) (Accuracy %)				
	Neutral	Happy	Sad	Angry	Average
Speaker A	76.30%	88.79%	83.17%	89.91%	84.54%
Speaker B	86.48%	84.99%	91.49%	85.17%	87.03%
Copus B	81.39%	86.89%	87.33%	87.54%	85.79%

參考文獻:

1. J. Liu, Y. Xu, S. Seneff, and V. Zue, "CityBrowser II: A Multimodal Restaurant Guide in Mandarin," in Proc. *International Symposium Chinese Spoken Language Processing (ISCSLP)*, pp. 1-4, 2008.
2. C.-H. Wu and G.-L. Yan, "Speech Act Modeling and Verification of Spontaneous Speech with Disfluency in a Spoken Dialogue System," *IEEE Trans. on Speech and Audio Processing*, Vol.13, pp.330-344, May 2005.
3. N. Roy, J. Pineau, and S. Thrun, "Spoken Dialogue Management Using Probabilistic Reasoning," in Proc. *Annual Meeting on Association for Computational Linguistics (AM-ACL)*, pp. 93-100, 2000.

4. D. Jurafsky, R. Ranganath, D. McFarland, "Extracting Social Meaning: Identifying Interactional Style in Spoken Conversation," in Proc. *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT)* , pp. 638-646, 2009.
5. Z. Dong, and Q. Dong, HowNet [Online] Available: <http://www.keenage.com/>
6. A. Berger, S. Della Pietra, and V. Della Pietra, "A Maximum Entropy Approach to Natural Language Processing," *Computational Linguistics*, Vol.22, No. 1, pp. 39-71, 1996.

*Copyright 2017 National Cheng Kung University*