

利用潛在語言空間模型及增量貝氏資訊準則於語言轉換偵測

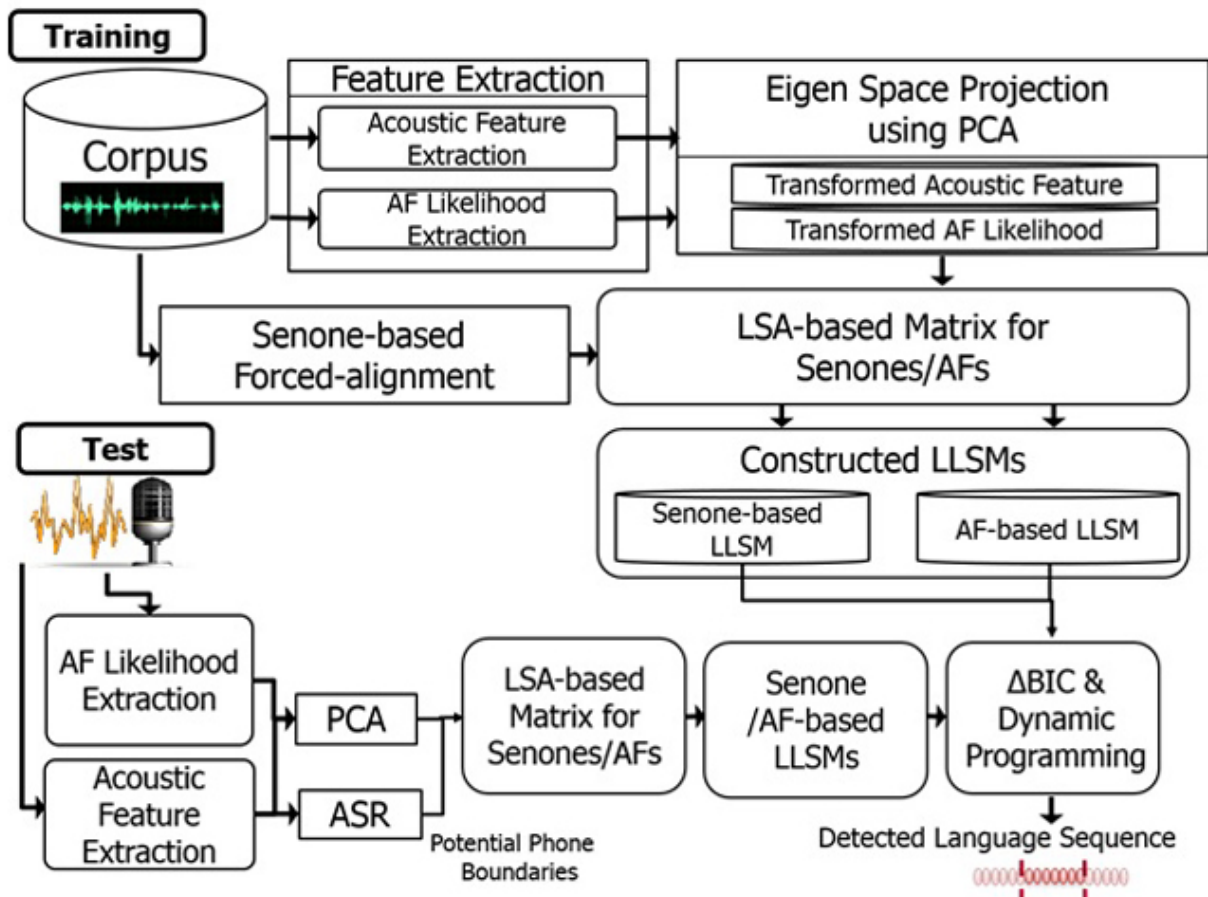
吳宗憲*、沈涵平、許純珊

國立成功大學資訊工程學系

chunghsienwu@gmail.com

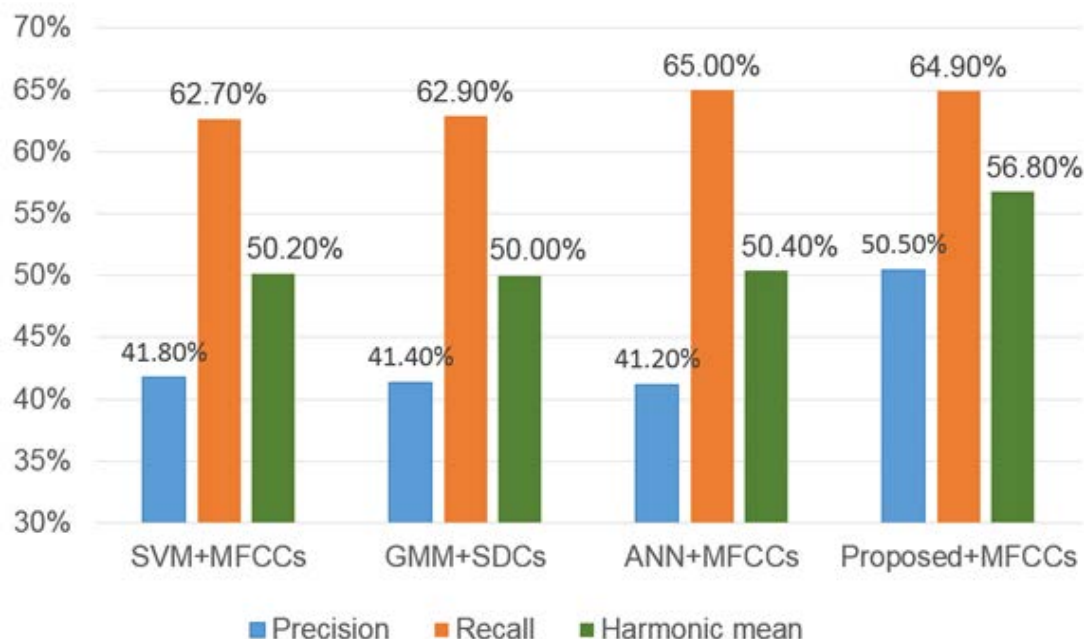
IEEE/ACM Trans. Audio, Speech, and Language Processing, Vol. 22, No. 10, October 2014, pp.1558-1570.

本研究提出了一基於潛在語言空間模型和增量貝氏資訊準則於語言轉換偵測。在所提出的方法中，首先計算出每個狀態音段中聲學特徵和發音參數之機率，並且通過使用主成分分析將其轉換為在特徵空間的主成分（特徵向量）。通過潛在語義分析建造一個矩陣來模擬在特徵空間中每個語言單元音的發音參數和各主成分的重要性。將主成分分析轉化的特徵值表示的單元音（或發音參數）之間的空間關係用來建構一個潛在語言空間模型來表示語言特性。基於這種概念，在檢測階段，一個輸入語音經識別後的單元音的聲學特徵（或發音參數）可表示為主成分分析轉化到特徵空間所形成的單元音（或發音參數）所表示的子潛語言空間模型。我們可將每個輸入語音的潛語言空間模型與目標語言的潛語言空間模型做比較。此外，對於語言轉換檢測，增量貝氏資訊準則用於估計兩個連續視窗之間的差異的分割能力表現出令人滿意之結果。因此增量貝氏資訊準則被採用來計算在輸入語音中語言的每個潛在的變化點。為了避免在所有的變化的點的窮盡搜索，從自動語音識別器識別出的單音邊界被視為潛在的語言變化點。最後，利用動態規劃算法確定基於從潛在語言空間模型和增量貝氏資訊準則估計的相似的最可能的語言序列。圖1示出所提出的語言轉換的系統框架。



圖一 系統架構圖

為了進行評估，以支持向量機、高斯混合模型及人工神經網絡為基礎的方法進行了比較。這些方法被用於將輸入語音標記化到語言序列。圖2表示所有利用這些方法所獲得的結果。評量結果顯示，所提出的方法在精確度，召回率及調和平均和語音長度方面均優於其他三種方法。



圖二 所提方法與支持向量機、高斯混合模型及人工神經網絡為基礎的方法之效能比較

參考文獻：

- [1] H. Y. Su, "Code-switching between Mandarin and Taiwanese in three telephone conversation: The negotiation of interpersonal relationships among bilingual speakers in Taiwan," in *Proc. The Symposium about Language and Society*, Apr 2001.
- [2] C. Chen, "Two types of code-switching in Taiwan," in *Proc. Sociolinguistics Symposium 15 (SS15)*, Newcastle upon Tyne, United Kingdom., Apr 2004.
- [3] H. Halmari, *Government and Code-Switching: Explaining American Finnish*. Amsterdam: John Benjamins, 1997.
- [4] J. Weiner, N. T. Vu, D. Telaar, F. Metze, T. Schultz, D.-C. Lyu, E.-S. Chng and H. Li, "Integration Of Language Identification Into A Recognition System For Spoken Conversations Containing Code-Switches," in *Proceedings of The third International Workshop on Spoken Languages Technologies for Under-resourced Languages (SLTU'12)*, 2012.
- [5] A. Hanani, "Human and Computer Recognition of Region Accents and Ethnic Groups from British Speech," PhD Thesis, The University of Birmingham, March 2012.
- [6] C.-H. Wu and C.-H. Hsieh, "Multiple change-point audio segmentation and classification using an MDL-based Gaussian model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 647–657, Mar 2006.
- [7] C.-H. Wu, H.-P. Shen and Y.-T. Yang, "Chinese-English Phone Set Construction for Code-Switching ASR Using Acoustic and DNN-Extracted Articulatory Features," *IEEE/ACM Trans. Audio, Speech, and Language Processing*, Vol. 22, No. 4, April 2014, pp. 858-862.