The 2015 Conference on Computational Linguistics and Speech Processing
ROCLING 2015, pp. 100-102

調變頻譜分解之改良於強健性語音辨識

Several Refinements of Modulation Spectrum Factorization for Robust Speech Recognition

張庭豪 Ting-Hao Chang, 洪孝宗 Hsiao-Tsung Hung, 陳柏琳 Berlin Chen
國立臺灣師範大學資訊工程學系
Department of Computer Science and Information Engineering
National Taiwan Normal University
{60247029S, 60047064S, berlin}@ntnu.edu.tw

陳冠宇 Kuan-Yu Chen, 王新民 Hsin-Min Wang
中央研究院資訊科學研究所
Institute of Information Science, Academia Sinica
{kychen, whm}@iis.sinica.edu.tw

摘要

絕大多數的自動語音辨識(Automatic Speech Recognition, ASR)系統常因為訓練與測試環境的不匹配而致使效能嚴重地下降。有鑒於此，語音強健性(Robustness)技術的發展長久以來一直是一個相當重要且熱門的研究領域。本論文之目的在於探索新穎的語音強健性技術，期望透過簡單且有效的語音特徵調變頻譜處理[1-3]來擷取較具強健性的語音特徵。為達此目的，本論文使用非負矩陣分解(Nonnegative Matrix Factorization, NMF)[4-6]以及一些改進方法來分解調變頻譜強度成分，以獲得較具強健性的語音特徵。本論文有下列幾項特色：(1)我們嘗試結合稀疏性的想法[7]，冀望能夠獲取到較具調變頻譜局部性的資訊以及重疊較少的 NMF 基底向量表示；(2)藉助於局部不變性的概念[8]，我們希望發音內容相似的語句之調變頻譜強度成分能在 NMF 空間有越相近的向量表示，以保留兩兩語句之間的關連程度；(3)在測試階段經由正規化 NMF 之編碼向量，更進一步提升語音特徵之強健性；(4)我們結合上述三種 NMF 的改進方法。本論文的所有實驗皆於國際通用的 Aurora-2 連續數字語音語料庫進行[9]；一系列的實驗結果顯示出，相較於僅使用梅爾倒頻譜特徵(Mel-frequency Cepstral Coefficients, MFCC)之基礎系統，我們所提出的新穎語音強健性技術能夠顯著地增進語音辨識效能，最終獲得 63.18%的相對詞錯誤率降低。另一方面，本論文也嘗試將我們所提出的改進方法與一些知名的特徵強健技術做比較和結合，以驗證我們所提

出語音強健性技術之實用性。例如，當其與統計圖等化法(Histogram Equalization, HEQ)[10]結合時，能較僅使用統計圖等化法的語音辨識系統有 19.90%的相對詞錯誤率降低；而當其與進階前端標準方法(Advanced Front-End Standard, AFE)[11]結合時，能較僅使用進階前端標準方法的語音辨識系統有 2.73%的相對詞錯誤率降低。

關鍵詞： 語音辨識、雜訊、強健性、調變頻譜、非負矩陣分解

致謝

本論文之研究承蒙教育部－國立臺灣師範大學邁向頂尖大學計畫(104-2911-I-003-301)與行政院科技部研究計畫(MOST 104-2221-E-003-018-MY3, MOST 103-2221-E-003-016-MY2, NSC 103-2911-I-003-301)之經費支持，謹此致謝。

參考文獻

[1] H. Hermansky, “Should recognizers have ears?” Invited Tutorial Paper, in *Proc. ESCA-NATO Tutorial and Research Workshop,* 1997.

[2] N. Kanedera, T. Arai, H. Hermansky and M. Pavel, “On the importance of various modulation frequencies for speech recognition,” in *Proc. European Conference on Speech Communication and Technology*, 1997.

[3] S. Greenberg, “On the origins of speech intelligibility in the real world,” in *Proc. ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels,* 1997.

[4] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature,* vol. 401, 788–791, 1999.

[5] W. Y., Chu, Y. C. Kao and B. Chen, “Probabilistic modulation spectrum factorization for robust speech recognition,” in *Proc ROCLING XXIII: Conference on Computational Linguistics and Speech Processing*, 2011.

[6] Y. C. Kao, Y. T. Wang and B. Chen. “Effective modulation spectrum Ffactorization for robust speech recognition.” in *Proc. the Annual Conference of the International Speech Communication Association*, 2014.

[7] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann and R. D. Pascual-Marqui, “Nonsmooth nonnegtive matrix facotorization (nsNMF),” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 403–415, 2006.

[8] D. Cai, X. He, J. Han, T. S. Huang, ”Graph regularized nonnegative matrix

factorization for data representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, 1548–1560, 2011.

[9] H. G. Hirsch and D. Pearce, “The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions,” in *Proc. ISCA ITRW ASR,* 2000.

[10] A. D. L. Torre, A. M. Peinado, J. C. Segura, J. L. Perez-Cordoba, M. C. Benitez, and A. J. Rubio, “Histogram equalization of speech representation for robust speech recognition,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 3, pp. 355–366, 2005.

[11] D. Macho, L. Mauuary, B. Noé, Y. M. Cheng, D. Ealey, D. Jouvet, H. Kelleher, D. Pearce and F. Saadoun, “Evaluation of a noise-robust DSR front-end on Aurora databases”, in *Proc. the Annual Conference of the International Speech Communication Association*, 2002.