雜訊環境下應用線性估測編碼於特徵時序列之強健性語音辨識

Employing linear prediction coding in feature time sequences for robust speech recognition in noisy environments

范顯騰 Hao-teng Fan, 曾文俞 Wen-yu Tseng, 洪志偉 Jeih-weih Hung
國立暨南國際大學電機工程學系
s99323904@mail1.ncnu.edu.tw, s100323553@mail1.ncnu.edu.tw, jwhung@ncnu.edu.tw

摘要

近幾十年來，無數的學者先進對於此雜訊干擾問題提出了豐富眾多的演算法，略分成兩大類別：強健性語音特徵參數表示法(robust speech feature representation)與語音模型調適法(speech model adaptation)，第一類別之方法主要目的在抽取不易受到外在環境干擾下而失真的語音特徵參數，或從原始語音特徵中儘量削減雜訊造成的效應，比較知名的方法有：倒頻譜平均值與變異數正規化法(cepstral mean and variance normalization, CMVN)[1]、倒頻譜統計圖正規化法(cepstral histogram normalization, CHN)[2]、倒頻譜平均值與變異數正規化結合自動回歸動態平均濾波器法(cepstral mean and variance normalization plus auto-regressive-moving average filtering, MVA)[3]等；第二類別之方法，則藉由少量的應用環境語料或雜訊，來對原始的語音模型中的統計參數作調整，降低模型之訓練環境與應用環境之不匹配的情況。較有名的語音模型調適技術包含了：最大後機率法則調適法(maximum a posteriori adaptation, MAP)[4]、平行模型合併法(parallel model combination, PMC)[5]、向量泰勒級數轉換(vector Taylor series transform, VTS)[6]等。本論文較集中討論與發展的是上述的第一類方法，我們提出一套作用於倒頻譜時間序列域的強健性技術，稱作線性估測編碼濾波法(linear prediction coding-based filtering, LPCF)，此方法主要是應用線性估測(linear prediction)[7]的原理，來擷取語音特徵隨著時間變化的特性、進而凸顯語音的成分、抑制雜訊的成分。在 LPCF 法中，將一段時域(time domain)上的訊號 $x[n]$ 用以下數學式表示：

$$x[n]=\sum_{k=1}^{P}a_k x[n-k]+e[n], \tag{1}$$

進而將上式的 $x[n]$ 經過 LPC 分析所得到的新特徵時間序列，表示爲 $\hat{x}[n]$，作法是先將原始語音特徵時間序列以 $x[n]$ 作 P 階之線性估測，求取式(1)之最佳之線性估測係數 $\left\{a_k, 1\le k\le P\right\}$。之後，經由下式求得新的特徵時間序列：

$$\hat{x}[n]=\sum_{k=1}^{P}a_k x[n-k], \tag{2}$$

上述的新方法雖然看似簡易，卻有許多合理的原因可顯示新序列相對於原始序列而言，包含了較少的失真、或對於雜訊更具強健性。語音分析中，原始訊號 $x[n]$ 與預估訊號 $\hat{x}[n]$ 之間的誤差訊號可能是週期性訊號或是白色雜訊，一般的線性迴歸模型(auto-regression model, AR model)也是建立在誤差訊號本身是白色雜訊的假設下。將其套用於我們這裡分析的語音特徵時間序列 $x[n]$ 中，可合理推測線性預估序列 $\hat{x}[n]$ 相當於扣除了 $x[n]$ 其中

部份無法線性估測的近似雜訊成份或週期性訊號成份，然而一般從語音特徵時間序列的軌跡，很少出現週期性的現象，因此我們可較確定的是，藉由 LPC 對於原始特徵序列 $x[n]$ 的分解，我們可將其分佈於全頻帶的白色雜訊成份加以消除或減低，而使新特徵序列 $\hat{x}[n]$ 包含較少的失真成份。另外，誤差序列 $e[n]$ 可能是週期性訊號(頻譜亦成週期性)或白色雜訊(頻譜呈現平坦之形狀)，但根據許多前人的研究，語音特徵時間序列其頻譜(即調變頻譜)的主要成份是集中於中低頻率上，因此誤差序列不太可能包含語音特徵序列的重要資訊，亦即將其扣除，至少無損於語音辨識的精確度。

本論文之實驗中所採用的語音資料庫爲歐洲電信標準協會(European Telecommunication Standard Institute, ETSI)所發行的 AURORA 2.0[8]語音資料庫，內容包含美國成年男女以人工方式錄製的一系列連續英文數字字串。辨識結果顯示，所提出的方法 LPCF 優於 MFCC，平均進步率達 4%；同時，我們也結合三種知名時間序列處理技術：CMVN、CHN 與 MVA，這裡我們將 LPCF 法作用於經 CMVN、CHN 或 MVA 法預處理後的 MFCC 特徵上，觀察 LPCF 法是否能夠使它們的辨識率進一步提升，LPCF 法能使 CMVN、CHN 與 MVA 預處理之特徵分別提升了 3.38%、2.2%與 0.87%，此代表了 LPCF 能與這些著名的時序域強健性技術有良好的加成性。

關鍵詞：線性估測編碼、特徵時間序列、雜訊強健性。
Keywords: noise robustness, speech recognition, linear predictive coding, temporal filtering.

參考文獻

[1] S. Tiberewala and H. Hermansky, “Multiband and adaptation approaches to robust speech recognition,” *in Proceedings of European Conference on Speech Communication and Technology,* 25(1-3), pp. 2619-2622, 1997.

[2] F. Hilger and H. Ney, “Quantile based histogram equalization for noise robust large vocabulary speech recognition,” *IEEE Transactions on Audio, Speech and Language Processing,* 14(3), pp. 845–854, 2006.

[3] C. P. Chen and J. Bilmes, “MVA processing of speech features,” *IEEE Transactions on Audio Speech and Language Processing,* 15(1), pp. 257-270, 2007.

[4] J. L. Gauiain and C. H. Lee, “Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains,” *IEEE Transactions on Speech and Audio Processing,* 2(2), pp. 291-298, 1994.

[5] J. W. Hung, J. L. Shen and L. S. Lee, “New approaches for domain transformation and parameter combination for improved accuracy in parallel model combination techniques,” *IEEE Transactions on Speech and Audio Processing,* 9(8), pp. 842-855, 2001

[6] P. J. Moreno, B. Raj, and R. M. Stern, “A vector Taylor series approach for environment-independent speech recognition,” *in Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing,* 2, pp. 733-736, 1996.

[7] 王小川, “語音訊號處理,” *全華科技圖書,* 2004.

[8] H. G. Hirsch and D. Pearce, “The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions,” *in Proceedings of the 2000 Automatic Speech Recognition Challenges for the new Millenium,* pp. 181-188, 2000.