

以高階隱藏式馬可夫模型作語音辨認之研究

李佳蓓、李立民

E-mail: 9510173@mail.dyu.edu.tw

摘要

為改善傳統一階隱藏式馬可夫模型的缺失，我們提出高階隱藏式馬可夫模型作語音辨識。在我們的模型中，狀態轉換及觀測輸出與當前及過去數個時間之狀態有關，因此，在捕捉語音單位之音長與頻譜動態軌跡上更加精確，配合高階隱藏式模型，我們發展擴充型Viterbi演算法來訓練模型與進行辨識。我們以一系列的實驗來驗證所提出方法之性能，實驗結果顯示，狀態轉換及觀測輸出均高階相依之系統不論在進行國語數字獨立音、國語數字混淆音、雜訊語音下之辨識均表現良好，錯誤率均隨著階數的增加而降低。在受雜訊干擾之混淆音的辨識表現上尤其顯著，平均可降低約達12%之錯誤率。高階隱藏式馬可夫模型確實能達到改善傳統一階隱藏式馬可夫模型的缺失。

關鍵詞：語音辨識；高階隱藏式馬可夫模型；動態軌跡；擴充型Viterbi演算法

目錄

封面內頁 簽名頁 授權書	iii	中文摘要	
.	iv	英文摘要	
.	v	誌謝	vi
.	vii	圖目錄	x
目錄	xii	第一章 緒論	
.	1	1.1 研究目的	1
.	1	1.2 研究背景	1
.	2	1.3 研究方法	2
.	3	1.4 章節概要	3
.	3	第二章 語音特徵參數擷取	4
辨識系統概要	4	2.1 語音	4
梅爾頻率倒頻譜參數擷取	5	2.2 語音特徵參數	5
.	12	2.3 梅爾頻率倒頻譜參數擷取	12
.	15	3 第三章 隱藏式馬可夫模型	15
.	16	3.1 隱藏式馬可夫模型	16
.	15	3.2 建立隱藏式馬可夫模型	15
.	16	3.2.1 初始狀態分佈	16
.	17	3.2.2 狀態轉移分佈	17
.	19	3.2.3 觀測符號機率分佈	19
.	21	3.4 Viterbi演算法	21
.	23	4 第四章 高階隱藏式馬可夫模型	23
.	23	4.1 高階隱藏式馬可夫模型	23
.	25	4.2 擴充型Viterbi演算法	25
.	25	4.3 模型訓練	25
.	30	5 第五章 實驗結果與討論	30
.	30	5.1 語音資料庫	30
.	31	5.2 模型建立	31
獨立音實驗結果	31	5.3 獨立音實驗	31
雜訊之實驗結果	33	5.3.1 乾淨語音之	33
.	35	5.3.2 語音摻雜20dB訊雜比白	35
.	35	5.3.2.1 語音摻雜20dB訊雜比白	35
.	35	5.3.2.2 語音摻雜10dB訊雜比白	35
.	37	5.4 混淆音實驗	37
實驗結果	39	5.4.1 未加雜訊之混淆音實	39
.	39	5.4.2 雜訊語音之混淆音實	39
.	40	5.4.2.1 混淆音摻雜20dB訊雜比白	40
.	41	5.4.2.2 混淆音摻雜10dB訊雜比白	41
.	42	5.5 實驗結果比較	42
.	47	6 第六章 結論與未來研究方向	47
.	47	6.1 結論	47
.	47	6.2 未來研究方向	47
.	48	參考文獻	48

參考文獻

- [1] S. Furui, "Speaker independent isolated word recognition using dynamic features of speech spectrum," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 52 – 59, 1986.
- [2] J.-F. Mari, J.-P. Haton, and A. Kriouile, "Automatic word recognition based on second-order hidden Markov models." IEEE Transactions on Speech and Audio Processing, vol. 5 no. 1, pp. 22 – 25, 1997.
- [3] Y. He, "Extended Viterbi algorithm for second-order hidden Markov process." In Proceedings of the IEEE 9th International Conference on

Pattern Recognition, pp.718 – 720, 1988.

[4] J.A. du Preez, “ Algorithms for high order hidden Markov modeling, ” Proceedings of the IEEE South African Symposium on Communications and Signal Processing, 9-10 Sept. pp. 101 -106, 1997.

[5] L. Deng, M. Aksmanovic, D. Sun, and C. F. J. Wu. “ Speech recognition using hidden Markov models with polynomial regression functions as nonstationary states, ” IEEE Transactions on Speech and Audio Processing, Vol. 2, No. 4, pp. 507-520, October, 1994.

[6] 王小川，語音訊號處理，全華科技，台北市，2005。

[7] 鄭智寬，“語音特徵抽取方法對連續音辨認影響之研究”，大葉大學碩士論文，彰化，民國93年6月。

[8] 廖子傑，“國語連續數字辨認之研究”，大葉大學碩士論文，彰化，民國93年6月。

[9] T. F. Quatieri , Discrete-Time speech signal processing principles and practice, Prentice Hall PTR, 2002.

[10] S. E. Levinson, L. R. Rabiner, M. M. Sondhi, “ An Introduction to the Application of the Theory of Probabilistic Function of a Markov Process to Automatic Speech Recognition, ” The Bell System Technical Journal, Vol.62, No.4, April 1983.

[11] X.D. Xuang, Y. Ariki, M.A. Jack, Hidden Markov Models for Speech Recognition. Edinburgh University Press, pp. 187-205, 1990.

[12] S. Young, The HTK Book, Version 3.2, Cambridge University Engineering Department, 2002.

[13] S. Davis, P. Mermelstein, “ Comparing of Parametric Representation for Monisyllable Word Recognition in Continuously Spoken Sentence, ” IEEE Trans. On Acoustic, Speech and Signal Processing, pp.357-366, 1980.

[14] Lee L. M. and Lee J. C. “ A Study on High-Order Hidden Markov Models and Applications to Speech Recognition, ” IEA/AIE 2006, Springer Lecture Notes in Artificial Intelligence, vol. 4031, pp. 682 – 690, Jun. 2006.