

# 語音特徵與模型架構之選取對語音辨認之影響

李坤旺、李立民

E-mail: 9607417@mail.dyu.edu.tw

## 摘要

本研究嘗試尋找最佳語音特徵參數抽取方法及選擇最佳隱藏式馬可夫模型之架構，以獲得最佳辨識率。在模型架構選取中，我們分別探討與比較全字模型、聲韻母模型、音素模型之性能。在第一部分的全字模型架構中，改變句子中的間隔音及靜音，並調整各音節狀態數以求性能提升。第二部份採聲韻模型架構，並以字內三音來模塑前後音的影響。第三部份為音素模型，以跨字關聯音的方式來精確表示語音，並且增加分群門檻值的語音特徵參數進行實驗，有效的降低錯誤型狀態的發生，第四部份語音特徵參數的調整，利用三角帶通濾波器之個數變化與語音音框長度的調整，進行實驗及比較，可得知語音特徵參數的改變對語音辨識效果的影響，在實驗結果顯示語音特徵與模型架構的選取對語音辨識有明顯成長的結果。

關鍵詞：語音辨識；分群門檻值；三角帶通濾波器之個數

## 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 目錄 . . . . .		vii 圖目錄 . . . . .	x
表目錄 . . . . .	xiv	第一章 前言 . . . . .	1
主題 . . . . .	1	1.1 研究背景 . . . . .	1
. . . . .	1	1.2 研究目的 . . . . .	2
. . . . .	3	1.3 研究目的 . . . . .	2
. . . . .	3	1.4 章節概要 . . . . .	3
. . . . .	4	第二章 語音特徵參數擷取 . . . . .	4
. . . . .	4	2.1 語音辨認流程簡介 . . . . .	4
. . . . .	5	2.2 抽取語音特徵參數 . . . . .	4
. . . . .	5	2.3 梅爾頻率倒頻譜參數 . . . . .	7
. . . . .	12	2.4 本章結論 . . . . .	7
. . . . .	12	第三章 語音參考模型的架構 . . . . .	13
. . . . .	13	3.1 國語數字語音之聲學模型架構 . . . . .	13
. . . . .	13	3.2 決策樹 . . . . .	17
. . . . .	15	3.3 隱藏式馬可夫模型的建立 . . . . .	17
. . . . .	22	3.3.1 國語數字語音之聲學模型架構 . . . . .	17
. . . . .	22	3.3.2 逆算程序 . . . . .	24
. . . . .	25	3.4 模型參數重估 . . . . .	24
. . . . .	25	3.5 模型比對 . . . . .	28
. . . . .	31	第四章 實驗結果與分析 . . . . .	31
. . . . .	31	4.1 語音資料庫簡介 . . . . .	31
. . . . .	31	4.2 聲學模型的建立與狀態數的分配 . . . . .	35
. . . . .	37	4.3 錯誤類型 . . . . .	35
. . . . .	37	4.4 各模型架構語音辨識實驗結果 . . . . .	37
. . . . .	37	4.4.1 全字模型架構 . . . . .	37
. . . . .	37	4.4.1.1 句首尾靜音實驗 . . . . .	37
. . . . .	37	4.4.1.2 字間間隔音實驗 . . . . .	41
. . . . .	41	4.4.1.3 字間間隔音加句首尾靜音實驗 . . . . .	46
. . . . .	53	4.4.1.4 每字音不同狀態數模型實驗 . . . . .	53
. . . . .	59	4.4.2 聲韻模型架構 . . . . .	59
. . . . .	59	4.4.2.1 聲韻模型實驗 . . . . .	59
. . . . .	59	4.4.3 跨字關聯音模型架構 . . . . .	65
. . . . .	65	4.4.3.1 跨字關聯音模型實驗 . . . . .	65
. . . . .	65	4.4.3.2 分群門檻值實驗 . . . . .	72
. . . . .	74	4.4.4 語音特徵架構 . . . . .	74
. . . . .	74	4.4.4.1 三角帶通濾波器個數實驗 . . . . .	74
. . . . .	74	4.4.4.2 音框長度實驗 . . . . .	74
. . . . .	75	第五章 結論及未來研究方向 . . . . .	78
. . . . .	78	5.1 結論 . . . . .	78
. . . . .	78	5.2 未來研究方向 . . . . .	78
. . . . .	80	參考文獻 . . . . .	80

## 參考文獻

- [1] Bogert, B. and Healy, M. and Tukey, J. " The Quefrency Analysis , " Proc. Symp. On Time series Anslsysis, New York, Wiley, J. pp.209-243, 1963.
- [2] Davis, S. and Mermelstein, P. " Comparing of Parametric Representations for Monosyllable Word Recognition in Continuously Spoken Sentence, " IEEE Trans. On Acoustic, Speech and Signal Processing, pp.357-366, 1980.
- [3] Beulen, K. and Ney, H. " Autumatic Question Generation for Decision Tree Based State Tying, " In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp805-808 Vol.2, 1998.
- [4] Baum, L. E. and Petrie, T. and Soules, G. R. and Weiss, N. " A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, " Ann. Math., Stat., vol. 41, no.1, pp 164-171, 1970.

- [5] Levinson, S. E. and Rabiner, L. R. and Sondhi, M. M. " 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.
- [6] Viterbi, A. J. " Error bounds for convolutional codes and an asymptotically optimal decoding algorithm, " IEEE Trans. Information Theory, IT-13:260-269, 1967.
- [7] Huang, Alex Acero, Hsiao-Wuen Hon: " Spoken Language Processing: A Guide to Theory, Algorithm and System Development Spoken language processing, " Prentice Hall, 2001.
- [8] Enrique Vidal and Andres Marzal, " A Review And New Approaches For Automatic Segmentation Of Speech Signals, " in "Signal Processing V: Theories and Applications". Eds. L.Torres, E.Masgrau, M.A.Lagunas. Elsevier, 1990.
- [9] Itakura, F. and Saito, S. " Analysis Synthesis Telephony Based on the Maximum Likelihood Method, " Proc. 6th Int. Congress on Acoustics ,Tokyo , Japan, 1968.
- [10] Svendsen, T. FK Soong, " On the Automatic Segmentation of Speech Signals, " ICASSP, pp. 77-80, 1987.
- [11] Thomas F. Quatieri, " Discrete-Time Speech signal processing principles and practice, " Prentice-Hall PTR, 2000 [12] Sagayama, S. and Itakura, F.: " On Individuality in a Dynamic Measure of Speech, " Proc. ASJ Spring Spring Conf. 1979, 3-2-7, pp. 589-590, June 1979 [13] Steve Young, " The HTK book, " Version 3.2, Cambridge University Engineering Department, 2002.
- [14] Rabiner L. R. and Schafer, R. W. " Digital Processing of Speech Signals, " Prentice- Hall Inc., 1978.
- [15] Boll, S. F. " Suppression of acousitic noise in speech using spectral subtraction, " IEEE Trans. ASSP, Vol.ASSP-27,NO.2,pp.113-120, Apr. 1979.
- [16] ATAL, B.S. (1985) " Linear Predictive Coding of Speech, " in F. FALLSIDE - W.A. WOODS (Eds) Computer Speech Processing. Englewood Cliffs, N.J. : Prentice Hall Intenational pp. 81-124.
- [17] Markel, J.D. and Gray Jr, A.H. " Linear Prediction of Speech, " Springer-Verlag, New York, 1976.
- [18] 鄭智寬, " 語音特徵抽取方法對連續音辨認影響之研究, " 碩士論文, 私立大葉大學電信工程研究所, 彰化, 2004,6月 [19] 李佳蓊, " 以高階隱藏式馬可夫模型作語音辨認之研究, " 碩士論文, 私立大葉大學電信工程研究所, 彰化, 2005,6月 [20] 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.