

A Study of Speaker Verification System

黃威勳、李立民

E-mail: 9707896@mail.dyu.edu.tw

ABSTRACT

The main purpose of speaker verification is to identify the speaker according to the related information of voice signals, and it requires a lot of steps to catch the differences between these signals by computers. In this thesis, Mel-Frequency Cepstrum Coefficients, MFCCs, are used as voice characteristic coefficients to match the characteristics of human pronunciation and hearing. Gaussian mixture model is widely used in the field of text independent speaker verification. However, differences of various speakers' voice are not only caused by different oral cavity shapes and vocal cords, but also the articulation speed. Because Gaussian mixture model does not consider the difference of articulation speed, high-order ergodic Gaussian model is adopted in this thesis to implement the text independent speaker verification system. These two models are tested under the same condition, and the results show that high-order ergodic Gaussian model can improve the performance. Equal Error Rate reduces 3.8 percentages.

Keywords : gaussian mixture model ; high-order ergodic gaussian model ; speaker verification

Table of Contents

目錄 封面內頁 簽名頁 授權書	iii	中文摘要	
.	iv	英文摘要	v
.	vi	誌謝	
.	vii	目錄	
.	viii	圖目錄	
.	ix	表目錄	
.	x	第一章 緒論	
.	1	1.1 研究動機	
.	1	1.1.1 研究動機	
.	1	1.1.2 語者驗證系統概論	
.	2	1.3 研究背景	
.	3	1.4 研究目的	
.	4	1.5 章節概要	
.	5	第二章 語音特徵參數擷取	
.	6	2.1 語者驗證系統概要	
.	6	2.2 量化誤差	
.	7	2.3 端點偵測	
.	7	2.4 梅爾倒頻譜參數	
.	10	第三章 向量量化	
.	18	3.1 簡介	
.	18	3.2 K-means分群法	
.	18	3.2 分裂分群法	
.	21	第四章 建立語者模型	
.	23	4.1 適用於語者驗證的模型	
.	23	4.2 高斯混合模型	
.	24	4.3 目標函式最佳化	
.	26	4.4 高階統計式高斯模型	
.	27	第五章 實驗結果	
.	33	5.1 語音資料庫簡介	
.	33	5.2 實驗流程設定	
.	33	5.3 系統評估方式	
.	34	5.2 實驗結果	
.	36	第五章 結論及未來研究方向	
.	49	6.1 結論	
.	49	6.2 未來研究方向	
.	50	參考文獻	
.	51		

REFERENCES

[1] D. A. Reynolds, " Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models ", IEEE Trans. Speech and Audio Processing, Vol. 3, No. 1, pp. 72-83, 1995.

[2] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, " Speaker Verification Using Adapted Gaussian Mixture Models ", Digital Signal Processing, Vol. 10, pp. 19-41, 2000.

[3] S. Davis, and P. Mermelstein, " Comparing of Parametric Representations for Monosyllable Word Recognition in Continuously Spoken Sentence ", IEEE Trans. On Acoustic, Speech and Signal Processing, pp. 357-366, 1980.

[4] L. R. Rabiner, and M. R. Sambur, " An algorithm for determining the endpoints of isolated utterances ", The Bell System Technique Journal, Vol. 54, pp. 297 - 315, 1975.

[5] T. K. Moon, " The Expectation-Maximization Algorithm ", IEEE, Signal Processing Magazine, Nov. 1996.

[6] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, " The DET curve in assessment of detection task performance ", In

Proceedings of the European Conference on Speech Communication and Technology, pp. 1895-1898, 1997.

[7] S. Young, The HTK book, Version 3.4, Cambridge University Engineering Department, 2002.

[8] 鍾偉仁, “語者辨認與驗證之初步研究”, 國立台灣大學碩士論文, 民國90年6月。

[9] 吳金池, “語者辨識系統之研究”, 國立中央大學碩士論文, 民國90年6月。

[10] 王小川, “語音訊號處理”, 全華科技圖書有限公司, 2005。

[11] 李佳蓓, “以高階隱藏式馬可夫模型作語音辨認之研究”, 大葉大學碩士論文, 民國95年6月。