中華民國計算語言學學會
通訊 ACLCLP
The Association for Computational Linguistics and Chinese Language Processing
第二十四卷第三期
發行人：王新民
主　編：王新民
執行編輯：黃　琪

E-Mail：aclclp@hp.iis.sinica.edu.tw
Website：http://www.aclclp.org.tw
地 址：台北市研究院路二段128號中研院資訊所
電 話：(02)2788-3799 ext:1502
傳 眞：(02)2788-1638
劃 撥：19166251

本 期 要 目

Information on upcoming conferences

ACL 2014

The 52nd Annual Meeting of the Association for Computational Linguistics

- Conference Date: June 22- 27, 2014
- Submission Deadline:
  Long paper: January 10, 2014
  Short paper: March 12, 2014
- Location: Baltimore, USA
- http://www.cs.jhu.edu/ACL2014/

ACM SIGIR 2014

The 37th Annual ACM SIGIR Conference

- Conference Date: July 6-11, 2014
- Submission Deadline: January 20, 2014
- Location: Gold Coast, Australia
- http://sigir.org/sigir2014/

COLING 2014

The 25th International Conference on Computational Linguistics

- Conference Date: August 23-29, 2014
- Submission Deadline: March 21, 2014
- Location: Dublin, Ireland
- http://www.coling-2014.org/

ICASSP 2014

2014 IEEE International Conference on Acoustics, Speech, and Signal Processing

- Conference Date: May 4-9, 2014
- Submission Deadline: Closed
- Location: Florence, Italy
- http://www.icassp2014.org/

INTERSPEECH 2014

15th Annual Conference of the International Speech Communication Association

- Conference Date: September 14-18, 2014
- Submission Deadline: March 24, 2014
- Location: Singapore
- http://www.interspeech2014.org/

IsCLL-14

The 14th International Symposium on Chinese Languages and Linguistics

- Conference Date: June 4-6, 2014
- Submission Deadline: February 10, 2014
- Location: Taipei, Taiwan
- http://iscll-14.ling.sinica.edu.tw/

ISCSLP 2014

The 9th International Symposium on Chinese Spoken Language Processing

- Conference Date: September 12-14, 2014
- Submission Deadline: TBD
- Location: Singapore
- http://www.iscslp2014.org/

KDD 2014

The 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining

- Conference Date: August 24-27, 2014
- Submission Deadline: February 13, 2014
- Location: New York, USA
- http://www.kdd.org/kdd2014/

Sep. 25-26, 2014
National Central University, Taiwan
http://sites.google.com/site/rocling2014

ROCLING 2014

The 26th International Conference on Computational Linguistics and Speech Processing

Conference Scope

The 2014 conference on computational linguistics and speech processing is the 26th annual conference sponsored by the Association for Computational Linguistics and Chinese Language Processing (ACLCLP). ROCLING 2014 will be held in National Central University, located in Jhongli, Taiwan. We invite paper submissions reporting original investigation results and system development experience on all areas of computational linguistics, natural language processing, and speech processing, including, but not limited to the following topic areas.

Topics

Computational Linguistics
Cognitive/Psychological linguistics
Computational semantics
Computational pragmatics
Computational morphology
Computational phonology
Discourse processing
Dialogue system
Syntax and parsing Word segmentation/POS tagging

Information Understanding
Information extraction
Information retrieval
Language learning
Machine translation/Multi-lingual systems
Named entity recognition
NLP tools/resources
Opinion mining and sentiment analysis
Question answering
Summarization

Speech Processing
Spoken language processing Speech analysis
Speech understanding
Speech synthesis/conversion Speech/speaker/language recognition
Speech based affective computing
Speech summarization Speech enhancement
Speech coding

Paper Submission

Each submission will be reviewed based on originality, significance, technical soundness, and relevance to the conference. Accepted papers will be presented orally or as poster presentations. Both oral and poster presentations will be published in the ROCLING 2013 conference proceedings and included in the ACL Anthology. A number of papers will be selected as Journal versions and invited for extension and publication in a special issue of the International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP).

Important Dates

Paper Submission Due: July 4 2014
Author Notification: Aug. 15, 2014
Camera Ready Due: Aug. 22, 2014

Hosts

National Central University, Taiwan
Association for Computational Linguistics and Chinese Processing

Honorary Chair
■ Jing-Yang Jou, National Central University

General Chairs
■ Chia-Hui Chang, National Central University
■ Hsin-Min Wang, Academia Sinica

Program Chairs
■ Jen-Tzung Chien, National Chiao-Tung University
■ Hung-Yu Kao, National Cheng-Kung University

Publicity Chair
■ Richard T.H. Tsai, National Central University

Industry Track Chair
■ Wen-Hsiang Lu, National Cheng-Kung University

Local Arrangement Chair
■ Jia-Ching Wang, National Central University

Publication Chair
■ Shih-Hung Wu, Chaoyang University

國立中央大學
National Central University

基於稀疏表示之語者辨識之研究

王光耀，林昶宏，王家慶
中央大學資訊工程研究所

1. 前言

以生物特徵作為辨識基礎的研究已經有長達數十年的歷史，包含人臉、語音和指紋。其中聲音因具有取得容易、非侵入性、運算量少、輸入具有便利性等優點，語者辨識一直是近年來熱門的主題。語者辨識利用語者語音的特性來辨識使用者身份，大致上的研究方向分成語音特徵擷取以及分類演算法兩部分，其中以分類演算法的研究最多，包含的層面也最廣，包括語者模型的建立、分類機制、因為周邊環境的影響所需的補償辦法等研究。目前來說高斯混和模型(Gaussian Mixture Model, GMM)[1][2]和支持向量機(Support Vector Machine, SVM)[3]是經典的分類方法。特徵參數方面，會將一段語音切割成數個音框，音框代表此語者在短時間內的發聲特徵，幾種常見的有梅爾倒頻譜係數(Mel-scale Frequency Cepstral Coefficients, MFCC)[5]、感知線性預測係數(Perceptual Linear Prediction Coefficients, PLPC)[6])等。

基於 GMM 超級向量(GMM-Supervector)[7-10]的辨識方法是一種綜合 GMM 模型表示以及 SVM 分類方法優點的語者辨識演算法。GMM 超級向量是一種語者模型的向量表示法，通用背景模型(Universal Background Model, UBM)經過輸入語音調適後成為 GMM 語者模型，其中各個高斯成份(Components) 的期望值被串在一起形成一個超級向量。此向量表示可以用來代表整個語音檔的特徵向量，最後，再以 SVM 作為分類方法來達到辨識語者的目的。然而語者辨識系統會因為錄製工具和背景不同，造成通道差異、語音內容差異(Session)和環境差異的干擾，導致辨識系統的辨識率低落。擾動屬性投影(Nuisance Attribute Projection, NAP)是一種有效補償通道干擾的方式[7][11]，藉由設計一個最佳化問題使干擾最小，配合著前面提到的方法，更有效地提高辨識系統的強健性。在 NAP 之後陸續有聯合因素分析(Joint Factor Analysis , JFA)[12][13]、i-vector[14-16]等技術被用來補償上述提到的差異，透過對語音參數的拆解，將一些語者不相關的資訊刪除，得到去除干擾後最能代表語者的特徵。

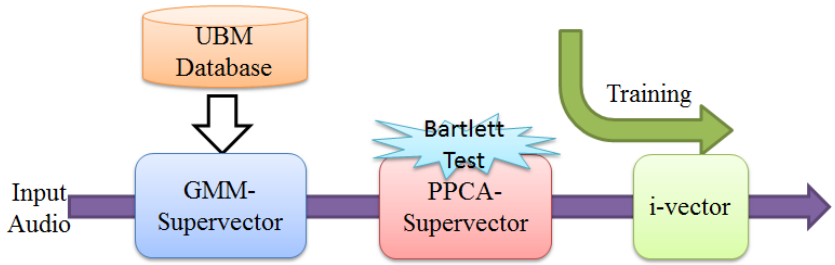
近年來對於複雜的資訊(例如訊號、圖像)，往往希望可以用較簡化的方式呈現，特別在訊號處理的部分，分析時經常先將資料轉換至不同的定義域，並且假設其在轉換後，會呈現稀疏分布[26]。在近期稀疏表示的發展中，稀疏表示分類器(Sparse Representation Classifier, SRC)在[27][28]中被提出，SRC 是一個 Nonparametric 學習方法，不需訓練過程

但是需要訓練資料，以及可以直接參考訓練資料對測試資料進行分類的動作。實驗結果顯示出在人臉辨識的應用上，SRC 有著比 KNN (K-Nearest Neighbors)[29]以及 Nearest Subspace(NS)[30][31]更好的辨識率。近幾年也有少數研究將 SRC 應用於語者驗證[16-19]的問題上，然而此方面的研究仍屬於剛起步的階段，仍有許多問題值得我們探討。

本專文提出一套基於稀疏表示的語者辨識系統，文章的編排共分成五個部分:第一部分為前言，第二部分為參數擷取，以機率型主成分分析(Probabilistic Principle Component Analysis, PPCA)建構超級向量(PPCA 超級向量)取代 GMM 超級向量[20][33]，並且以巴雷特檢定(Bartlett Test)的方式調整特徵值的選取，第三部分，描述如何利用第二部分的特徵參數建立稀疏表示分類器，以及依據變異的蒐集方式建構噪音字典的方法，第四部分，以貝氏機率概念的 Approximate Bayesian Compressed Sensing (ABCS)[24][25]求解稀疏係數。第五部分為實驗部分，展示提出之改良方法是否具有其必要性。

2. 參數擷取

在本專文中，我們利用機率型主成分分析建構超級向量[20][33]，並以提出巴雷特檢定(Bartlett Test)主成分的個數。傳統上，超級向量是由高斯混和模型建構而成，這裡加入主成分分析的概念，並希望能以機率分布模型的形式與高斯模型對應，使得資料點由原本高斯混和模型轉成 PPCA 混和模型，再透過 Latent Factor 的轉換，形成新的超級向量，稱作 PPCA 超級向量[20][21][33]，其中，主軸的挑選，我們引入巴雷特檢定(Bartlett Test)[22][23]的概念，建立假說，找到臨界的特徵值。而目前語者辨識問題中，i-vector 是表現最好的參數之一，藉由訓練出總體變異矩陣，將原本的超級向量轉到更低維的空間，使 i-vector 更加表現出語者及通道的資訊，因此，我們將 PPCA 超級向量轉換到總體變異空間上，希望得到更具鑑別力的 i-vector 使辨識效能提升。本文中所採用的特徵擷取流程如下圖所示。



《圖一》特徵擷取流程。

2.1. 基於機率型主成分分析之因素分析模型

在[21][33]中提到傳統上的 GMM 超級向量，並沒有考量到其聲學特徵參數有高度的冗餘性，因此應該採用更低的子空間來表示。在[21][33]中比較 Factor Analysis 與 PCA 的關聯性，觀察 Factor Analysis 的數學式：

$$x = Wz + \mu + \varepsilon \tag{1}$$

其中 x 代表高維的資料，z 代表低維變數，又稱 Latent Factor。回想主成分分析的數學式，找出主成分 V 使資料降維：

$$x = \mu + Vz \tag{2}$$

我們可將 Factor Analysis 視成對資料 x 做 PCA 加上一個噪音項，並且導入機率的概論解釋。

在式(1)中，x 是 $d \times 1$ 的原始資料，W 為 $d \times m$ 的轉換矩陣，其中 $m < d$，而 Latent Factor z 假設為一高斯分布 $N(0, \mathrm{I})$，噪音 $\varepsilon \sim N(0, \sigma^2 \mathrm{I})$，根據以上的假設可以知道原始資料 x 能建模(Model)成 $N(\mu, \sigma^2 \mathrm{I} + \mathrm{WW}^{\mathrm{T}})$，稱作 PPCA 模型。當遇到更複雜的資料時，希望藉由 Mix 多組 PPCA 模型，如下式：

$$p(x) = \sum_{c=1}^{M} w_c \, p(x \mid c) \tag{3}$$

$$p(x \mid c) = N(\mu_c, {\sigma_c}^2 \mathrm{I} + W_c {W_c}^{\mathrm{T}}) \tag{4}$$

其中 c 是高斯成份的個數，這可以跟高斯混合模型做對應，一般表示資料以多個高斯模型描述，這裡的基本單位以 PPCA 模型取代。

因此，我們希望找出 W_c 及 σ_c 來推算 Latent Factor，方法是利用最大概似估計(Maximum Likelihood Estimation, MLE)，因為已經知道 x 的分布，所以藉由 MLE 得到的 W_c 及 σ_c 可進而推出 Latent Factor z_c

$$z_c = ({W_c}^T W_c + {\sigma_c}^2 \mathrm{I})^{-1} {W_c}^T (x - \mu_c) \tag{5}$$

當原始資料 x 進入系統後，會先以 PPCA 混合模型，並且透過式(5)的轉換得到屬於每個高斯成份的 Latent Factor。當所有的參數都做了 PPCA 後，則原始以 MFCC 為基礎的 UBM 也必須調整，形成以 Latent Factor 為參數下構建的新 UBM：

$$p(z \mid c) = \sum_{c=1}^{M} w_c N(\mu_{z,c}, \Sigma_{z,c}) = \sum_{c=1}^{M} w_c N(0, I - \sigma_c{}^2 \Lambda_c{}^{-1}) \tag{6}$$

$$\text{其中} \quad \mu_{z,c} = E\{(W_c{}^T W_c + \sigma_c{}^2 \mathrm{I})^{-1} W_c{}^T (x - \mu_c)\} = 0 \tag{7}$$

$$\begin{aligned} \Sigma_{z,c} &= E\{z_c z_c{}^T\} - \mu_{z,c} \mu_{z,c}{}^T \\ &= (W_c{}^T W_c + \sigma_c{}^2 \mathrm{I})^{-1} W_c{}^T E\{(x - \mu_c)(x - \mu_c)^T\} W_c (W_c{}^T W_c + \sigma_c{}^2 \mathrm{I})^{-T} \\ &= \Lambda_c{}^{-1} (\Lambda_c - \sigma_c{}^2 \mathrm{I})^{T/2} \Lambda_c (\Lambda_c - \sigma_c{}^2 \mathrm{I})^{1/2} \Lambda_c{}^{-T} = I - \sigma_c{}^2 \Lambda_c{}^{-1} \end{aligned} \tag{8}$$

2.2. 巴雷特檢定

在上個小節中提到的 PPCA 中，有一個問題是需要討論的，那就是主軸個數的挑選。我們假設後面不要的特徵值，其數值小到足以視為相同，因此，定義假說 $H_0 : \lambda_{k+1} = \lambda_{K+2} = \ldots = \lambda_n$，希望迴圈由 n-1 往前找到門檻值 k。而如果我們將特徵值的大小視為高斯分布，則巴雷特檢定[22][23]可以整理成相似度的檢驗，如下式：

$$T = \frac{\prod_{q=k+1}^{n} \lambda_q}{\left(\frac{\sum_{q=k+1}^{n} \lambda_q}{n-k} \right)^{n-k}} \tag{9}$$

當觀測資料數 m 夠多的話，可將 T 視為一個 X^2 分布，且將上式近似於下式：

$$T \approx \left(m - (2n-11)/6\right) \left((n-k) \log \bar{\lambda} - \sum_{q=k+1}^{n} \log \lambda_q \right) \tag{10}$$

其中 $\bar{\lambda}$ 是後 n-k 個特徵值的平均，在檢定過程中，當 $T > X^2{}_{\alpha,n}$ 時，則否決假說 H_0，檢定終止，q 即為我們挑選的主軸個數，其中 α 為 X^2 的顯著水平。

2.3. i-vector

啟發於早期 JFA 在語者辨識上的應用，Dehak et al. 提出的一個新的分析方法 i-vector [14-16]，不像 JFA 將語者和通道分開，i-vector 僅用一個總體變異性空間，他發現 JFA 中的通道部分仍包含了能用來識別語者的資訊，所以將 JFA 中分開的變異部分，合併成一個單一的總體變異性空間超級向量，藉由合併錄音方式的變異性，提升其辨識性，表示如下：

$$\mu = m + Tw \tag{11}$$

m 是 UBM 超級向量，與 JFA 中使用的相同；T 代表的是所有變異性的矩陣；i-vector 代表的是總體變異性元素 w。

最後，總結參數擷取的整體架構，參數擷取包含三個部分，第一，由 Universal Background Data 首先訓練出 UBM，並透過 PPCA 將 UBM 轉換成以 Latent Factor 為參數的 UBM，第二，當輸入語音進入系統後，擷取其 Latent Factor，接著，對新的 UBM 調適產生 PPCA 超級向量。最後，在基於 PPCA 超級向量參數下訓練總變異矩陣，並將輸入語音轉換成 i-vector。

3. 稀疏表示分類器

我們利用稀疏表示具鑑別力的特性，應用於語者辨識問題上，利用第 2.3 節得到的 i-vector 作為特徵參數，以固定維度的向量表示語音訊號。其稀疏表示分類器[16-19]架構及演算法如《圖二》所示，假設有 k 個不同語者類別(Class)的訓練資料，每一筆訓練資料為一個 i-vector，我們需要建構一個跨類別的(Global)字典 D，作法是將所有語者類別的字典組在一起

$$D = [D_1\ D_2 ..... D_k] \in \Re^{m \times n} \tag{12}$$

其中，D_i 表示第 i 個類別的字典，由類別 i 的 i-vector 串接而成。此外，m 代表參數維度，n 是訓練資料個數。在測試資料 y 與字典 D 已知的情況下，我們希望求出稀疏係數 $x = [x_1\ x_2 \ldots x_k]$，使原訊號與重建訊號之間的誤差能越小越好，且 x 要符合稀疏特性，如下式：

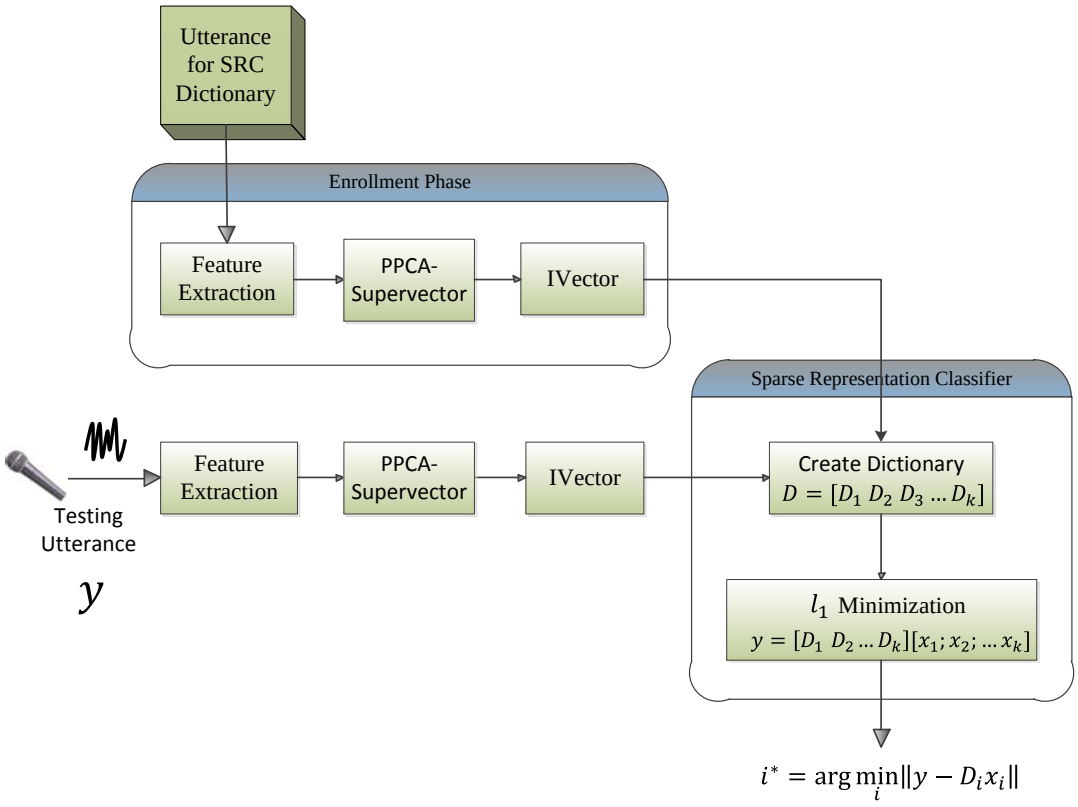
$$\min_x \|y - Dx\|_2^2 + \lambda \|x\|_1 \tag{13}$$

在解出稀疏係數 x 後，決策上，將 k 類別各自的字典 D_i 及係數 x_i 還原 y，並與 y 計算誤差，獲得最小誤差者即為所屬類別：

$$i^* = \arg\min_i \left\| y - D_i x_i \right\|_2^2 \tag{14}$$

3.1 噪音字典

噪音字典的建立方式是擴增一組雜訊基底至 SRC 的字典中，目的是吸收輸入語音未能被原先字典所表示的殘值。假設輸入語音不能被原本的基底所完整的線性組合表示，則剩下的殘值可由雜訊基底所吸收。如此一來，透過雜訊基底的吸收，可排除掉額外的變異。我們建構噪音字典的方法是蒐集變異量，找出每個 Atom 與自身語者類別平均的差值，即為雜訊基底。假設原始字典有 k 個語者類別，$D = [D_1\ D_2 ..... D_k]$，每個語者類別有 n 個 Atom，定義 N_1 等於 D_1 每個 Atom 扣除 D_1 的平均，因此加入雜訊基底的字典可表示成 $D_{new} = [D_1 ..... D_k\ N_1 ..... N_k]$。



《圖二》稀疏表示分類器架構。

4. Approximate Bayesian Compressed Sensing(ABCS)

在本篇專文中，對於求解 SRC 係數的問題，我們利用[24][25]中提到 ABCS 來求解。在[24][25]中以機率化的概念解釋，並且以 Semi Gaussian 作為係數的 Prior，限制其 Sparse 的特性。

$$y = H\beta \qquad s.t. \qquad \|\beta\|_1^2 < \varepsilon \tag{15}$$

其中 y 是一筆維度 m 的測試資料，H 是 n 個 Atom 的字典，且 m << n。假設 y 滿足以下由 H 做線性組合表示：$y = H\beta + \varsigma$，ς 假設為高斯分布 $N(0, R)$。則我們可以對 $p(y \mid \beta)$ 亦做高斯的假設，分布如下：

$$p(y \mid \beta) \propto \exp(-1/2(y - H\beta)^T R^{-1}(y - H\beta)) \tag{16}$$

此外，如果我們假設 β 的 Prior $p(\beta)$ 是已知，則我們可以由 Maximum a Posteriori (MAP) 估測出 β。在 ABCS 的推導中，$p(\beta)$ 是兩個 Prior Constraints 的乘積，如以下：

$$\begin{aligned} \beta^* &= \arg\max_{\beta} p(\beta \mid y) = \arg\max_{\beta} p(y \mid \beta) p(\beta) \\ &= \arg\max_{\beta} p(y \mid \beta) p_G(\beta) p_{SG}(\beta) \end{aligned} \tag{17}$$

其中 $p_G(\beta)$ 是 Gaussian Constraint，$p_{SG}(\beta)$ 是 Semi Gaussian constraint，限制係數滿足 Sparse 的特性。在[24][25]中對式(17)進行兩步驟的求解，最終推導出一疊代的解析解。

5. 實驗數據以及討論

我們以 NIST2005 [32] 做為語者資料庫，NIST 每年都會錄製語者資料庫，許多產業界、學術界都會以此作為評估效能的標準，而 2005 年發布的資料庫以電話對話語音數據為主，同時收集一些輔助麥克風接收的數據，這些數據主要來自英語演講，並包括四種額外語言。我們挑選其中 Male 8con-1con 的 Condition 做為測試資料庫，其中 8con 為訓練資料，而 1con 是測試資料。

為了驗證巴雷特檢定是否能進一步改善辨識系統，我們定義另一套主軸個數選取方式：各個高斯成份在統一主軸個數下與透過檢定方式挑選適合主軸個數下的差別，最後再經由轉換到 i-vector 上。藉由比較巴雷特檢定與我們定義的主軸選取方式來判斷巴雷特檢定的必要性。在我們的實驗中，固定每個高斯成份的主軸個數=27/Component 來到最高 76.91%，不過隨著主軸個數的減少辨識率也降低。而加入巴雷特檢定後，我們設定巴雷特檢定的 $\alpha = 0.05$。各個高斯成份降維的數目不固定，有效的提升辨識率到 77.01%。在接

下來的實驗中，我們會以這個辨識率為 77.01%的系統作為後續實驗的基準(Baseline)，測試提出採用的改進方法是否有效。

對於噪音字典擴增的實驗，根據實驗數據顯示，加入噪音字典在效果上降為 76.81%，分析原因，可能是因為係數必須滿足稀疏的性質，假設係數只有 50 個有值，則這 50 個理論上會落在原本的字典上，大多數不會落在噪音字典，因此，並未發揮吸收噪音的效果。

參考文獻

[1] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture models," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 1, pp. 72–83, Jan. 1995.

[2] B. L. Pellom and J. H. L. Hansen, "An efficient scoring algorithm for Gaussian mixture model based speaker identification," *IEEE Signal Process. Lett.*, vol. 5, no. 11, pp. 281–284, Nov. 1998.

[3] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer, and P. A. Torres-Carrasquillo, "Support vector machines for speaker and language recognition," *Comput. Speech Lang.*, vol. 20, pp. 210–229, 2006.

[4] J. L. Gauvain and C. H. Lee, "Maximum *a posteriori* estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 291–298, 1994.

[5] M. R. Hasan, M. Jamil, M. G. Rabbani, and M. S. Rahman, "Speaker identification using Mel frequency cepstral coefficients," *3rd international Conference on Electrical & Computer Engineering ICECE 2004*, 28-30 December 2004, Dhaka, Bangladesh.

[6] H. Hermansky. "Perceptual Linear Predictive (PLP) Analysis of Speech," *Journal of the Acoust. Society ofAmer.*, 87: 1738- 1752, April, 1990.

[7] W. Campbell, D. Sturim, D. Reynolds, and A. Solomonoff, "SVM based speaker verification using a GMM supervector kernel and nap variability compensation," in *Proc. ICASSP*, Toulouse, France, 2006, pp. 97–100.

[8] W. Campbell, D. Sturim, and D. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 308–311, May 2006.

[9] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Commun.*, vol. 52, no. 1, pp. 12–40, 2010.

[10] B. G. B. Fauve, D. Matrouf, N. Scheffer, J.-F. Bonastre, and J. S. D. Mason, "State-of-the-art performance in text-independent speaker verification through open-source software," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 7, pp. 1960–1968, 2007.

[11] A. Solomonoff, W. M. Campbell, and C. Quillen, "Channel compensation for SVM speaker recognition," *in Proc. Odyssey04*, 2004, pp. 57–62.

[12] O. Glembek, L. Burget, N. Brummer, and P. Kenny, "Comparison of scoring methods used in speaker recognition with joint factor analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 4057–4060.

[13] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A study of interspeaker variability in speaker verification," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 5, pp. 980–988, Jul. 2008.

[14] A. Kanagasundaram, R. Vogt, D. Dean, S. Sridharan, and M. Mason, "i-vector based Speaker Recognition on Short Utterances," in *Interspeech*, 2011.

[15] P. Matejka, O. Glembek, F. Castaldo, O. Plchot, P. Kenny, L. Burget, and J. Cernocky, “Full-covariance ubm and heavy-tailed plda in i-vector speaker verification,” *Proc. ICASSP ’11*, pp. 4828–4831, 2011.

[16] J.M.K. Kua, J. Epps, E. Ambikairajah, “i-vector with sparse representation classification for speaker verification, ” *Speech Commun*, 2013.

[17] K. Huang and S. Aviyente, “Sparse Representation for Signal Classification,” *Neural Information Processing Systems*, 2006.

[18] J. M. K. Kua, E. Ambikairajah, J. Epps, and R. Togneri, “Speaker verification using sparse representation classification,” in *Proc. ICASSP*, May 2011, pp. 4548–4551.

[19] R. Saeidi, A. Hurmalainen, T. Virtanen, and D. A. van Leeuwen, “Exemplar-based Sparse Representation and Sparse Discrimination for Noise Robust Speaker Identification,” in *Odyssey speaker and language recognition workshop*, Singapore, 2012.

[20] T. Hasan and J. H. L. Hansen, “Factor analysis of acoustic features using a mixture of probabilistic principal component analyzers for robust speaker verification,” *in Proc. Odyssey*, Singapore, Jun. 2012.

[21] M. Tipping and C. Bishop, “Mixtures of probabilistic principal component analyzers,” *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.

[22] 李孟穎，「感知因素分析法應用於語音強化」，成功大學資訊工程學系博士專文，2004 年。

[23] M.S. Bartlett, “Tests of significance in factor analysis,” *British Journal of Psychology*, Statistical Section 3, 77–85, 1950

[24] A. Carmi, P. Gurfil, D. Kanevsky, and B. Ramabhadran, “ABCS: Approximate Bayesian Compressed Sensing,” *Tech. Rep., Human Language Technologies*, IBM, 2009.

[25] T. N. Sainath, A. Carmi, D. Kanevsky, and B. Ramabhadran, “Bayesian compressive sensing for phonetic classification,” in *Proc. Int. Conf. Audio, Speech, Signal Process.*, 2010, pp. 4370–4373.

[26] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: an algorithm for designing overcomplete dictionaries for Sparse Representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[27] A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry, “Feature selection in face recognition: A sparse representation perspective,” EECS Dept., Univ. California, Berkeley, CA, Tech. Report UCB/EECS-2007-99, 2007.

[28] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Transaction Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–226, 2009.

[29] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd edition New York: Wiley, 2000.

[30] S. Z. Li, “Face recognition based on nearest linear combinations,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 839–844, 1998.

[31] K. C. Lee, J. Ho, and D. Kriegman, “Acquiring linear subspaces for face recognition under variable lighting,” *IEEE Transaction Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[32] NIST 2005 Speaker Recognition Evaluation plan, http://www.nist.gov/speech/tests/spk/2005/sre-05_evalplan-v6.pdf.

[33] T. Hasan and J. H. L. Hansen, “Acoustic factor analysis for robust speaker verification,” *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 21, no.4, pp.842-853, Apr. 2013.