

語者辨識

Speaker Recognition

專題組員:劉禮毅 蕭翊安

專題編號: PRJ-NTPUCSIE-101-004

執行期間:101 年 8 月 至 102 年 6 月

1. 摘要

本研究的目的是在於辨識出不同的語者，靠的是將各種不同韻母的聲音特性加以分析，來辨識出各個人之頻譜差異。

我們使用 Matlab 做為開發平臺，並用 Excel 做為資料庫之儲存及管理。

未來可應用在身分核對上，例如保險箱之聲紋密碼鎖、領取證件或重要物品文件等。以及個人物品使用之身分確認，例如手機、平板及個人電腦。

2. 簡介

人們一般說話會包含兩種因素，分為內容以及音色。內容就是說話的內文，音色指的則是不同語者的聲音特色。因此，只要固定說話語句的內容，將不同韻母的影響因素固定，即可依此區分出不同人音色之不同。

我們之所以能夠區分不同的人聲、母音，主要就是依靠判斷共振峰不同的分布位置。共振峰，顧名思義就是聲音頻譜中能量相對集中的一些區域。人聲受到每個人的生理構造不

同之影響：如鼻腔、咽喉、口腔，而有自身的共振峰區。透過改變這些共鳴空間形狀和大小不同的變化，聲音就會擁有不同的共振峰。

本專題主要目標就是辨識出語者是誰。

3. 專題進行方式

我們專題實驗收集的樣本共有 10 人，分為 5 位男生及 5 位女生；設計出來的語句有五種，每個人會分別對每一種語句錄音 7 次，其中 5 句當作資料庫之樣本，另外 2 句當作測試句。每一種語句都包含 Y、一、ㄨ、ㄟ、又，5 個音，並分別對這 5 個音做 LPC 頻譜分析，設計的語句如下：

(1) 骨頭比誰大

(2) 七組咖啡豆

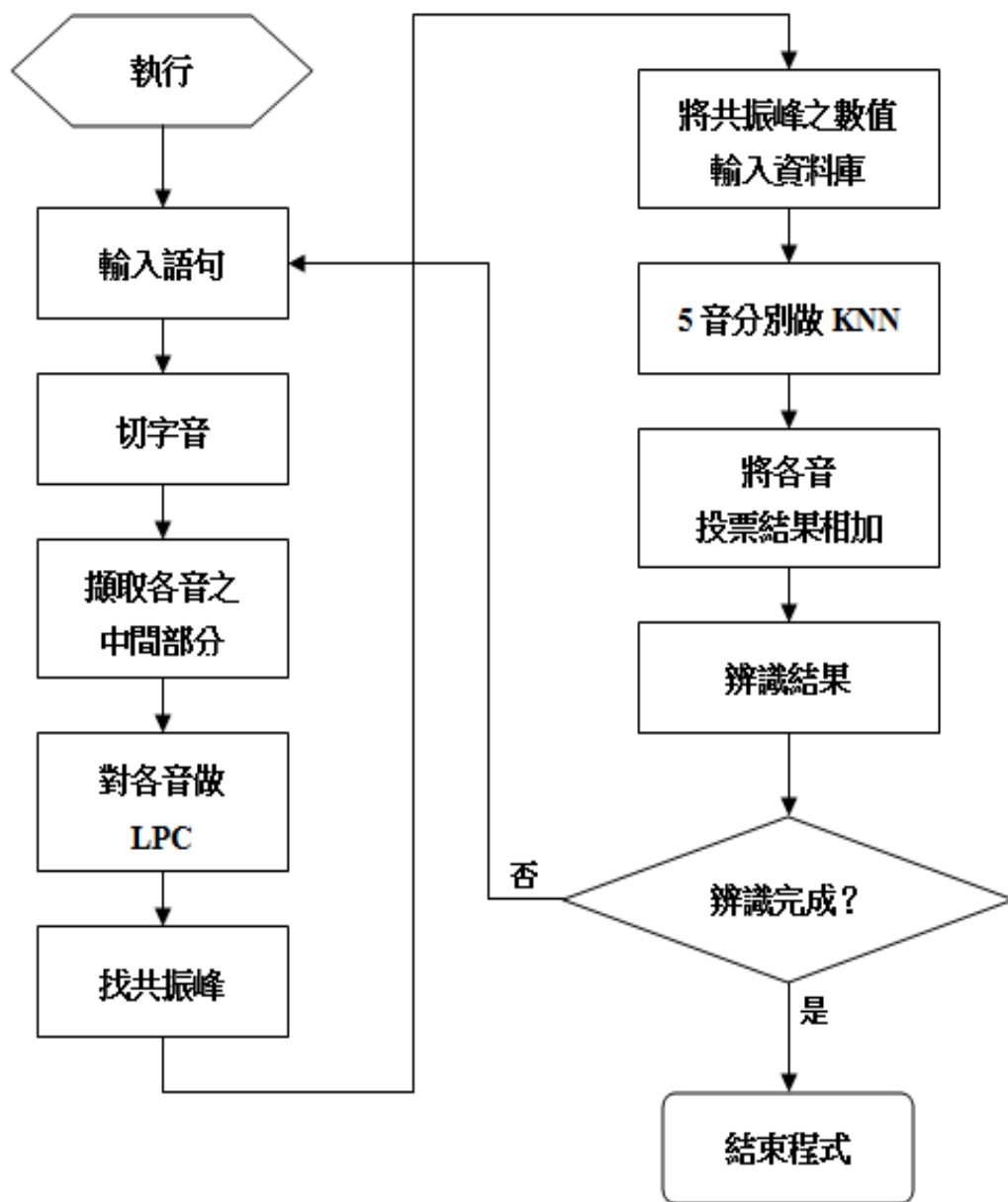
(3) 黑狗撲大地

(4) 土地大黑手

(5) 不給他豆皮

可以看出每一句話皆包含了上述 5 個音。

方法流程圖如下：



3.1 LPC 頻譜分析

(Linear Prediction Coefficient)

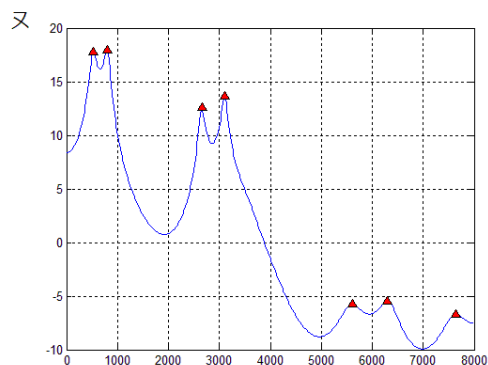
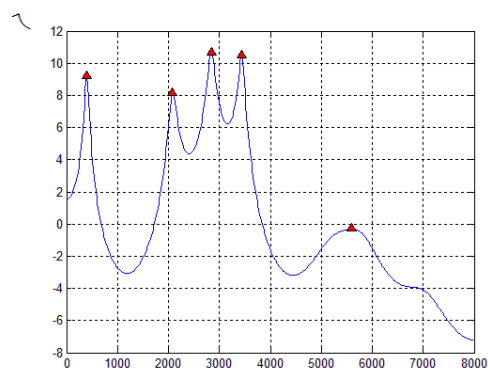
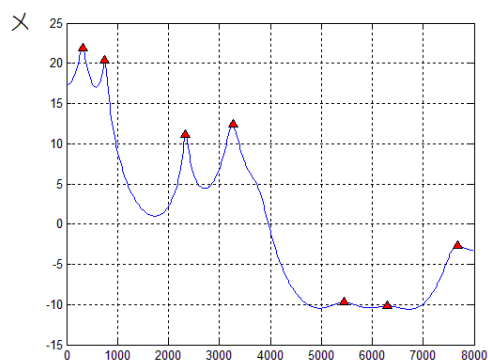
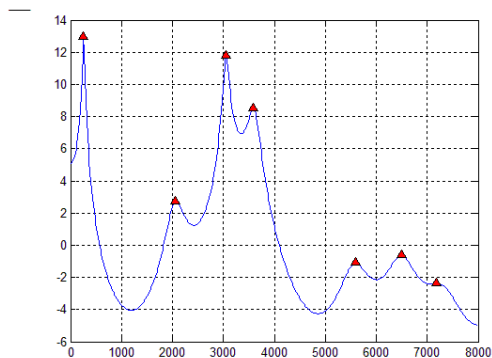
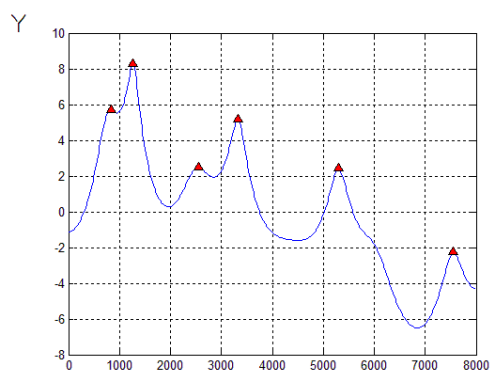
線性預測基本原理，就是因為大部分的語音訊號都具有週期性。所以假設目前聲音之取樣值，可由在前面的 P 個取樣值以線性組合來預測。公式如下：

$$\tilde{s}(n) = \sum_{k=1}^p \alpha_k s(n-k)$$

而我們依靠下列公式幫助我們描繪出頻譜圖，以便找出共振峰點：

$$\tilde{H}(e^{j\omega}) = \frac{G}{1 - \sum_{k=1}^p \alpha_k e^{-j\omega k}}$$

下圖為 Y、一、x、ㄟ、又之頻譜圖，可以很明顯的找出每個音的共振峰點分布位置，並以此區分出各音之不同：



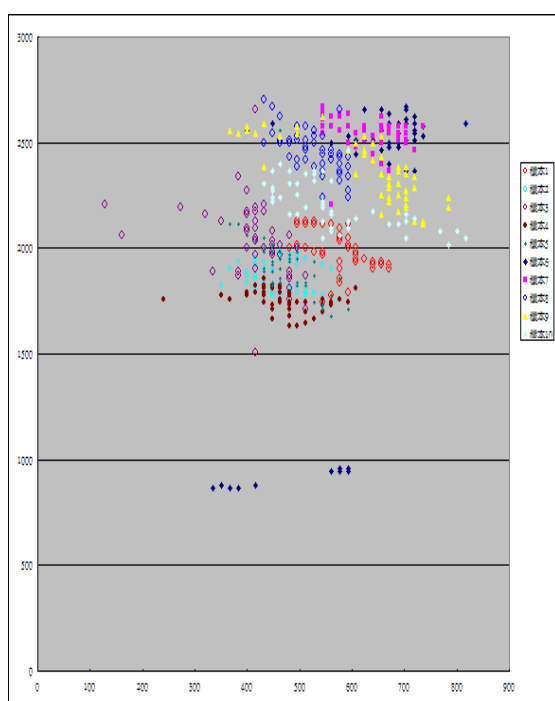
3.2 KNN 分群法

(K-Nearest Neighbors)

KNN 的基本原理，就是假設相同類別之物件在空間中的點，距離會比較接近。所以只要找出在資料庫中和未知類別資料，最接近的 K 個點最多者，即可判定是相同的類別。

所以我們抓取前三個共振峰：f1, f2, f3 做為 KNN 之比較標準。

下圖為本實驗資料庫內，ㄟ這個音共振峰點分布狀況，可以看出不同人各有各的分布區域，且有分群現象。橫軸為 f1，縱軸為 f2，單位皆為 Hz。

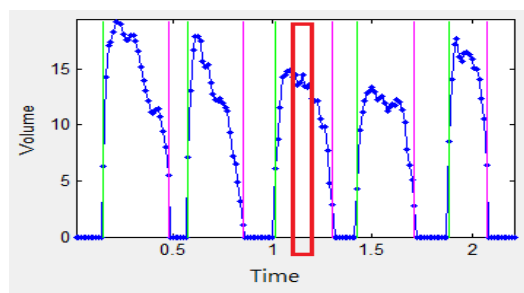


3.3 實驗流程說明

將一段語句輸入後，將輸入的語音訊號做切字音的動作，將低於門檻值的訊號濾除，並保留高於門檻值的部分，使程式能夠分別對每個音做判斷。

接著擷取各個音之中間部分，目的是在於去除前面子音和氣音的干擾

因素，並得到每個音中較為穩定的部分，如下圖：



然後對擷取出來的部分做 LPC 頻譜分析，會得到各音之共振峰數值，並將這些數值輸入到資料庫中做比對。

我們比對的方式分為 f1f2、f2f3、f1f3 以及立體空間的投入點比較，並將各個音選出最近的 K 個點之後，判斷出所屬的類別，這樣算為一票。接著對各音加入不同的權重，使分群效果明顯的音有較高權重，並讓不易辨識之音的影響力降低。各音之權重如下：

ㄚ：1.5

一：2.8

ㄨ：0.5

ㄟ：2.8

又：0.2

將各音之票數乘上權重後相加，並合併 4 種比對方式最後的票數，即可得到總票數，票數最多者即為判斷出來的結果。

4. 主要成果與評估

4.1 實驗結果

實驗結果如下表：

	辨識成功之投票率 (%)					成功率
	語句a	語句b	語句c	語句d	語句e	
樣本1	78.38	85.71	64.86	48.65	78.95	100%
	76.92	89.19	97.37	54.05	57.89	
樣本2	60.53	70.27	72.97	57.89	75	100%
	60.53	86.49	83.33	57.14	60	
樣本3	51.28	80	42.11	55.56	21.62	80%
	52.78	70.59	80	37.84	37.84	
樣本4	94.74	67.57	67.57	57.5	60.53	90%
	65.79	86.11	57.89	47.37	79.49	
樣本5	50	35.14	56.76	51.43	44.74	80%
	58.82	59.46	39.47	35.14	36.11	
樣本6	55.26	81.08	76.32	73.68	71.05	100%
	71.05	48.72	76.92	72.22	79.49	
樣本7	38.89	37.84	27.03	42.11	56.76	40%
	10.81	8.1	8.1	26.32	35.14	
樣本8	42.11	72.97	81.08	21.62	61.11	80%
	86.84	50	41.67	5.56	64.86	
樣本9	50	13.89	23.08	42.11	32.43	40%
	17.94	37.84	7.89	56.26	22.5	
樣本10	2.63	0	54.05	41.67	27.03	50%
	14.28	35.14	22.22	64.86	41.67	
Total	80%	85%	75%	75%	65%	76%

註：表中之數字為該樣本單一語句之正確投票率。藍底黑字的格子表示判斷正確，紅字則表示判斷錯誤。

若將各種語句做不同的組合，判斷正確率如右圖：

(A)

	a	b	c	d	e
a		70%	70%	100%	80%
b			70%	100%	80%
c				90%	90%
d					100%
e					
平均					85%

(B)

	成功率
abc	70%
abd	100%
abe	80%
aod	90%
ace	90%
ade	100%
bcd	90%
boe	90%
bde	100%
ode	90%
平均	90%

(C)

	成功率
abod	100%
abce	90%
abde	90%
aode	90%
bode	90%
平均	92%

(D)

	成功率
abcde	100%

註：

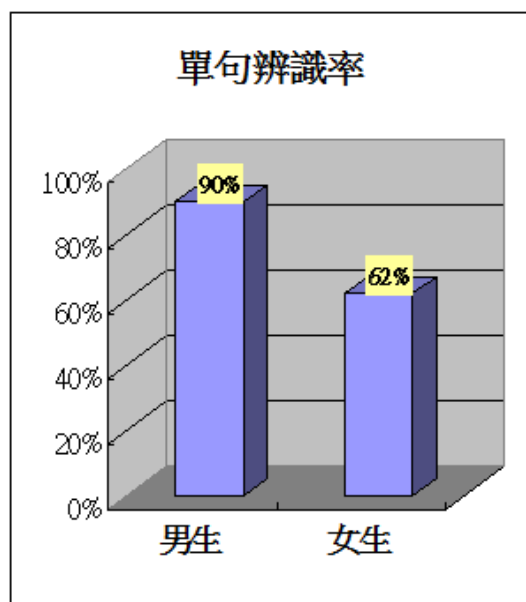
圖(A)為語句兩兩配對之成功率。

圖(B)為三句一組之成功率。

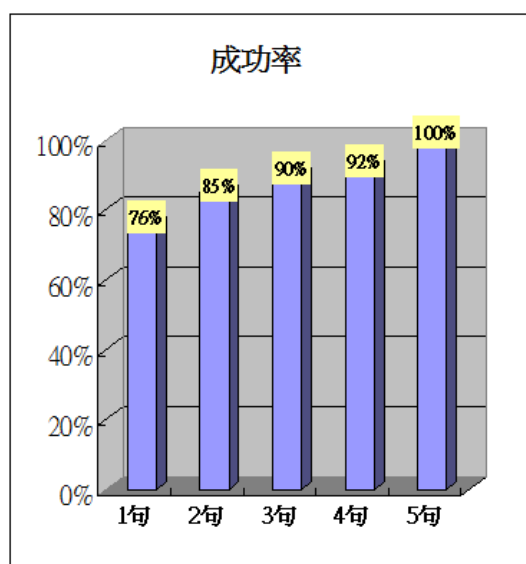
圖(C)為四句一組之成功率。

圖(D)為五句一組，成功率達到 100%

若以單句辨識率來看，男生之辨識率會較女生為高，可能原因為本實驗為抓取前三個共振峰做為比對之標準。而男生的低頻較穩定，故其 KNN 較為準確，結果如下：



綜合來看，不難看出增加語句可以提升正確率。結果如下：



4.2 討論

辨識率的高低，通常會受到幾個因素影響：

(1) 峰值是否有漏抓或抓錯。

(2) LPC 之 P 值，男女生及各音適用不同。

(3) 錄音時的因素。例：能量衰弱造成音框數目不足或共振峰不明顯。

(4) 某些語句的發音會使得分群的效果不明顯。例：三聲容易壓低聲音發聲，可能造成共振峰過少或能量不易集中，使 KNN 效果不佳。

可能之改善方法如下：

(1) 增加資料庫之樣本數，進而提高資料的群聚效果，使得 KNN 之準確率提升。

(2) 改良各音的權重。

5. 結語與展望

未來希望能夠達到不固定語句的辨識，使程式之可用性更高、應用層面更廣，並移植到不同的平臺上運作，例如：手機、平版的加解密，不再只侷限於電腦上。

6. 銘謝

我們在這為期一年的專題製作過程中，經歷了很多，也學到了很多，特別要感謝的是我們的指導教授，耐心的指導我們一步接著一步完成這個專題，有困難時適時地給予我們幫助，收穫良多。

7. 參考文獻

[1] MATLAB 函式庫說明

<http://www.mathworks.com/help/matlab/index.html>

[2] <Audio Signal Processing and Recognition (音訊處理與辨識)> 張

智星教授

[3]

<音量分析斷字音> 張智星教授

[http://neural.cs.nthu.edu.tw/jang/books/audioSignalProcessing/basicFeatureVolume.asp?title=5-2%20Volume%20\(%AD%B5%B6q\)](http://neural.cs.nthu.edu.tw/jang/books/audioSignalProcessing/basicFeatureVolume.asp?title=5-2%20Volume%20(%AD%B5%B6q))

[http://neural.cs.nthu.edu.tw/jang/books/audioSignalProcessing/basicFeatureZeroCrossingRate.asp?title=5-3%20Zero%20Crossing%20Rate%20\(%B9L%B9s%B2v\)](http://neural.cs.nthu.edu.tw/jang/books/audioSignalProcessing/basicFeatureZeroCrossingRate.asp?title=5-3%20Zero%20Crossing%20Rate%20(%B9L%B9s%B2v))

[4]

<KNN 演算法概念> WIKI

https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[5]

<KNN MATLAB 範例> 張智星教授

<http://neural.cs.nthu.edu.tw/jang/books/dcpr/prKnn.asp?title=5-2%20K-nearest-neighbor%20Classifiers>

[6]

<LPC 線性預測編碼技術>

<http://file.lw23.com/6/60/605/605855f5-cea3-44e9-b5d9-077ae26320b3.pdf>

