

國立臺灣大學工學院應用力學研究所

博士論文

Graduate Institute of Applied Mechanics

College of Engineering

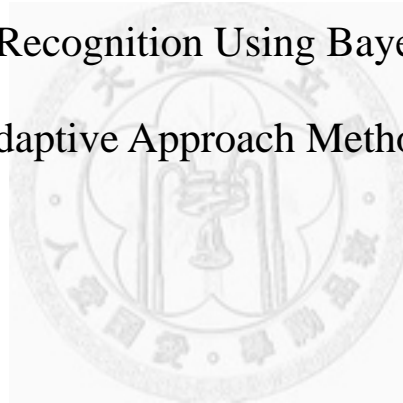
National Taiwan University

Ph.D. Thesis

應用貝氏網路及適應性調適方法於語音情緒辨識之研究

Speech Emotion Recognition Using Bayesian Network and

Adaptive Approach Methods



游志源

Chih-Yuan Yu

指導教授：劉佩玲 博士

Advisor: Pei-Ling Liu, Ph.D.

中華民國 100 年 1 月

January, 2011

誌 謝

本論文得以順利完成，首先要感謝恩師 劉佩玲教授多年來的悉心指導與關懷。從碩士班到博士班的求學過程中，老師本著傳道、授業、解惑的指導精神，除了在研究上訓練我獨立思考與解決問題的能力外，老師在研究上實事求是的治學態度及為人處事，亦使我記憶深刻並永存於心，在此僅向恩師獻上衷心的敬意與感謝。

論文定稿期間，承蒙成功大學 吳宗憲教授、中正大學 余松年教授、大同大學 包蒼龍教授、南台科大 陳有圳助理教授、崑山科大 黃志賢助理教授、工研院 石明于博士以及廖志彬博士對於本文的指正及提供寶貴的意見，使得本論文能夠更臻完備，也讓我在專業領域的知識上獲益良多，在此表達本人誠摯的謝意。

在實驗室的研究生活期間，特別感謝葉柏涼學長在課業以及生活上的相互提攜，每當研究遇到瓶頸，學長總是給於適時的鼓勵與協助，真心的感謝您。同時亦感謝實驗室人文、力立、志峰、威靈、智雄、士芳以及嘉文等學弟的熱心協助，在此亦一併致謝。

感謝工研院南分院在我工作服務的那段期間，提供語料作為測試與驗證使用，同時感謝俊昇以及珠娟在實驗上的協助，淵翔、信嘉、昇照、秦安、育賢、毅翔一起為院長盃四連霸奮戰的日子，濤哥、士傑、鳴立、蕭大哥、蔡大哥、佳婉、紹興每個禮拜運動練球的日子，

龍哥、俊斌、時瑜、洋銘、祥傑在執行工研院計畫及技術發展時，給予的協助與幫忙，在此也一併感謝。

最後感謝我的家人、女友欣君、紀爸爸、紀媽媽以及每一個關懷我的朋友們，在求學的這段期間對我的支持與鼓勵，讓我能夠安心並持之以恆的完成目標，僅將本文獻給所有關心我的人們。



中文摘要：

本研究之主要目的在發展一貝氏網路自動化語音情緒辨識方法，透過情緒語音之相關特徵參數計算，並與資料庫中各情緒之資料作比對，將語者的情緒狀態從語音訊號中辨識出來。

首先，將語者之情緒語音訊號以統計方式計算音高(Pitch)、音框能量(Frame energy)、共振峰(Formants)以及梅爾倒頻譜係數(Mel-scale Frequency Cepstral Coefficients, MFCC)等相關之語音情緒特徵，並以各特徵參數在中性情緒下之資料庫平均值為正規化特徵參數因子，將音高、音框能量以及共振峰利用正規化特徵參數因子進行正規化，得到正規化後之特徵參數，以消除不同語者之間差異。

各特徵參數對情緒之辨識能力不同，如音高可大約分辨悲傷及中性，快樂及生氣會被視為同一群。由於沒有任一參數可明顯分辨四種情緒，故本研究採用分層解析的方式，先將特徵參數分群，同一群特徵參數具有相似的情緒分類效果，並以分群結果建立多層貝氏網路(Multi-Layered Bayesian Network, MLBN)，第一層的輸入參數都只能辨識兩群情緒，第二層的輸入參數可辨識三群情緒，較無明顯分群效果的參數則置於第三層，以辨識四種情緒。

由於特徵參數之間具有相關性，因此，本論文將 MLBN 延伸，將特徵參數之相關性納入考量並提出多層共變異數貝氏網路

(Multi-Layered Bayesian Network with Covariance, MLBNC)。

當語者資料不在訓練資料中時，其辨識效果通常不佳。為改善此狀況，本研究提出適應性 MLBN 及適應性 MLBNC 語音情緒辨識調適方法，在調適過程中，當辨識結果與語者情緒不同時，則根據語者情緒語音所得之參數值來調整 MLBN 及 MLBNC 資料庫中各群之平均值與標準差或共變異數，以符合語者真實的情緒狀態。

為驗證本研究所提出之方法，我們使用德國情緒語料庫 EMO-DB 當作訓練與測試語料，並以 KNN、SVM、MLBN 以及 MLBNC 分別進行 Inside 與 Outside Test。同時，我們亦以 EMO-DB 為訓練語料，並以工業技術研究所自行錄製之情緒語料為測試語料，分別對 KNN、SVM、MLBN 以及 MLBNC 進行不同語系之測試。而在適應性 MLBN 及 MLBNC 的驗證上，我們以 EMO-DB 為訓練語料，並以工研院之情緒語料為調適與測試語料，分別對適應性 KNN、MLBN 及 MLBNC 進行調適前後的測試。

實驗結果顯示，本研究提出之 MLBN、MLBNC 與單純使用貝氏決策之 Inside Test 辨識率分別為 81.1%、88.8% 以及 70.8%，顯示透過各參數分層分群的辨識方式，可以有效提高語音情緒之辨識率，而考慮特徵參數間相關性之 MLBNC，其結果亦優於 MLBN。在 Outside Test 的部分，KNN、SVM 以及 MLBN 使用原始參數時，其辨識率分

別為 78.2%、89.1%以及 69.9%，而使用正規化參數時，辨識率分別為 82.6%、91.7%以及 77.6%，顯示正規化特徵參數可以有效縮小語者之間在特徵參數上的差異。而當訓練與測試語料為不同語系時，KNN、SVM、MLBN 以及 MLBNC 之辨識率分別為 34.21%、46.92%、39.33%以及 52.08%，此結果顯示，當發音方式或表達情緒方式與資料庫有所差異時，各分類器之辨識效果均不佳。

在調適實驗部分，由調適前後之辨識結果顯示，KNN 經調適過後，辨識率從 34.2%提升至 73.7%，而 MLBN 及 MLBNC 經調適後，其辨識率分別從 37.8%提升至 82.4%以及 51.6%提升至 81.2%，本研究提出之適應性 MLBN 與適應性 MLBNC 語音情緒辨識方法於資料庫修正後，其辨識效果明顯優於適應性 KNN。而當調適次數增加時，MLBN 及 MLBNC 經調適後，其辨識率則分別從 39.3%提升至 88.9%以及 52.1%提升至 90.0%，顯示經由本論文所提出之調適方法，經調適後，確實可以真正的反映語者的真實狀況，並得到良好的調適後辨識結果。

關鍵字：語音情緒辨識、特徵參數、正規化、多層貝氏網路、多層共變異數貝氏網路、適應性。

ABSTRACT :

The objective of this study is to develop an automatic speech emotion recognition method using Bayesian Network. By calculating the relevant features of emotion speech and comparing the features with emotion database, the speaker's emotion state can be identified.

Firstly, we calculate the statistical features of pitch, frame energy, formants, mel-scale frequency cepstral coefficients (MFCC). Then we use the mean value of neutral emotion in corpus as normalized factor for each feature, and calculate the normalized features of pitch, frame energy and formants. The normalized features can reduce the feature difference between speakers.

Each feature has different ability of emotion recognition. For example, the normalized pitch mean can recognize sad and neutral, and happy and angry can consider as the same cluster. No features can obviously recognize the four emotions, so we use different cluster to recognize the four emotions layer by layer. We cluster the features which have similar ability of emotion recognition and establish the Multi-Layered Bayesian Network (MLBN) method for speech emotion recognition. The features of layer 1 can recognize two clusters of emotion.

The features of layer 2 can recognize three clusters of emotion. The features which have no obvious clusters are put on layer 3 and recognize the four emotions.

There are some relations between each feature. Therefore, we extend the MLBN method and establish the Multi-Layered Bayesian Network with Covariance (MLBNC) method, which consider the relations between each feature, for speech emotion recognition.

The recognition rate will be poor if the training data of recognizer did not contain speaker's speech emotion data. Therefore, we propose adaptive MLBN and MLBNC method for speech emotion recognition. In the adaptive MLBN and MLBNC process, we adjust the mean and standard deviation or covariance of clusters in the MLBN or MLBNC database to fit speaker's real emotion status when the recognition result is wrong.

To verify the proposed method in this research, we use German emotional database (EMO-DB) as training and testing data for inside and outside test of KNN, SVM, MLBN and MLBNC recognizer. We also use EMO-DB as training data and ITRI emotional database as testing data for different corpus test. In the adaptive tests, we use EMO-DB as training

data and ITRI emotional database as adaptive and testing data for adaptive KNN, MLBN and MLBNC recognizer.

The inside test recognition rate of MLBN, MLBNC and Bayesian Decision (BD) are 81.1%, 88.8% and 70.8% respectively. It shows that cluster of features layer by layer can effectively increase the recognition rate and it will be better when regards of the relations between each feature. In outside test, the recognition rate of KNN, SVM and MLBN are 78.2%, 89.1% and 69.9% respectively using original features and 82.6%, 91.7% and 77.6% respectively using normalized features. It shows that normalized features can reduce the feature difference between speakers and increase the recognition rate. In testing corpus is different with training, the recognition rate of KNN, SVM, MLBN and MLBNC are 34.21%, 46.92%, 39.33% and 52.08% respectively. It shows if speaker's pronunciation or emotion presentation is different with training data, the recognition result is bad for each recognizer.

For adaptive emotion recognition test, adaptive KNN method can increase the recognition rate from 34.2% to 73.7%, adaptive MLBN method can increase from 37.8% to 82.4% and adaptive MLBNC method can increase from 51.6% to 81.2%. The proposed adaptive MLBN and

MLBNC method of this study is better than adaptive KNN method. When adjustment times increase, the recognition rate of MLBN can increase from 39.3% to 88.9% and MLBNC can increase from 52.1% to 90.0%. It shows that adaptive MLBN and MLBNC method can really reflect the real status of speaker's emotion state and get good recognition results after appropriate adjustment.

Keywords : speech emotion recognition 、 features 、 normalization 、

MLBN 、 MLBNC 、 adaptive.



目 錄：

誌謝-----	I
中文摘要-----	III
ABSTRACT-----	VI
目錄-----	X
圖目錄-----	XIII
表目錄-----	XVI
第一章 導論-----	1
1.1 前言-----	1
1.2 文獻回顧-----	2
1.3 本文簡介-----	7
第二章 語音情緒特徵參數計算-----	9
2.1 前處理-----	9
2.1.1 音框(Frame)與視窗(Windows)-----	9
2.1.2 預強調(Pre-emphasis)-----	10
2.1.3 快速傅立葉轉換(Fast Fourier Transform, FFT)-----	11
2.2 語音特徵(Speech Feature)計算-----	12
2.2.1 音高(Pitch)-----	12
2.2.2 共振峰(Formant)-----	14
2.2.3 音框能量(Frame Energy)-----	14
2.2.4 梅爾頻率倒頻譜係數(Mel-frequency Cepstral coefficient, MFCC)-----	15
2.3 小結-----	16
第三章 特徵參數之統計計算與正規化-----	23
3.1 情緒語音資料庫-----	23

3.2 特徵統計計算-----	24
3.2.1 語音特徵平均值(Mean)與標準差(Standard deviation)-	24
3.2.2 語音特徵正規化計算(Normalization)-----	28
3.3 小結-----	32
第四章 KNN 與 SVM 於語音情緒辨識之實驗與分析-----	50
4.1 KNN 語音情緒辨識-----	50
4.1.1 第 K 個最近鄰(K-Nearest Neighbor, KNN)-----	50
4.1.2 KNN 語音情緒辨識實驗結果與分析-----	51
4.2 SVM 語音情緒辨識-----	56
4.2.1 支持向量機(Support Vector Machine, SVM)-----	56
4.2.2 SVM 情緒辨識實驗結果與分析-----	58
4.3 小結-----	62
第五章 多層貝氏網路與多層共變異數貝氏網路語音情緒辨識-----	71
5.1 決策樹、貝氏決策與貝氏網路介紹-----	71
5.1.1 決策樹 (Decision Tree) -----	71
5.1.2 貝氏決策(Bayes Decision)與貝氏網路(Bayesian Network)	
-----	72
5.2 多層貝氏網路(Multi-Layer Bayesian Network, MLBN)語音情	
緒辨識-----	75
5.2.1 語音情緒特徵分群分析-----	75
5.2.2 多層貝氏網路(MLBN)-----	83
5.2.3 多層貝氏網路(MLBN)語音情緒辨識實驗結果與分析	
-----	94
5.3 多層共變異數貝氏網路(Multi-Layer Bayesian Network with	
Covariance, MLBNC)語音情緒辨識-----	98
5.3.1 多層共變異數貝氏網路(MLBNC)-----	98

5.3.2	MLBNC 實驗結果與分析	111
5.4	蒙地卡羅模擬(Monte Carlo Simulation)與分析	115
5.5	小結	120
第六章	調適性語音情緒辨識	149
6.1	適應性 KNN 語音情緒辨識實驗與分析	149
6.2	適應性 MLBN 語音情緒辨識實驗與分析	152
6.2.1	適應性 MLBN	152
6.2.2	適應性 MLBN 實驗結果與分析	166
6.3	適應性 MLBNC 語音情緒辨識實驗與分析	171
6.3.1	適應性 MLBNC	172
6.3.2	適應性 MLBNC 實驗結果與分析	183
6.4	小結	189
第七章	結論與未來展望	211
	參考文獻	218
	附錄 A	225
	附錄 B	232
	附錄 C	240

圖目錄：

圖 2-1 語音情緒辨識流程圖-----	18
圖 2-2 語音訊號音框化後之波形-----	18
圖 2-3 Hamming and Rectangular window -----	19
圖 2-4 人耳聽覺各頻率最小音壓曲線-----	19
圖 2-5 預強調前後之語音波形-----	20
圖 2-6 預強調前後之語音頻譜-----	20
圖 2-7 語音之音高變化-----	21
圖 2-8 單一音框之共振峰示意圖-----	21
圖 2-9 語音之音框能量變化-----	22
圖 3-1 工研院情緒語音錄製環境-----	34
圖 3-2 不同語者在不同情緒之音高平均值分佈-----	35
圖 3-3 不同語者在不同情緒之原始特徵參數分佈-----	39
圖 3-4 正規化後音高平均值在不同語者不同情緒下之分佈-----	40
圖 3-5 正規化後音高標準差在不同語者不同情緒下之分佈-----	40
圖 3-6 正規化後各特徵參數在不同語者不同情緒下之分佈-----	42
圖 3-7 使用正規化特徵參數之語音情緒辨識流程-----	43
圖 4-1 第 K 個最近鄰(KNN)-----	64
圖 4-2 KNN 語音情緒辨識流程-----	64

圖 4-3 KNN 不同 K 值之辨識率-----	65
圖 4-4 二維支持向量機-----	65
圖 4-5 SVM 語音情緒辨識流程-----	66
圖 5-1 簡易交通工具分類流程-----	123
圖 5-2 四類別之貝氏決策示意圖-----	123
圖 5-3 兩類別之貝氏決策示意圖-----	124
圖 5-4 四種情緒類別之正規化後音高分佈-----	124
圖 5-5 四種情緒類別之正規化特徵參數分佈-----	128
圖 5-6 悲傷與其他情緒之正規化後音高平均值分佈-----	129
圖 5-7 多層貝氏網路語音情緒辨識架構-----	130
圖 5-8 多層貝氏網路第一層-----	131
圖 5-9 MFCC C8 平均值原始分佈及分群後分佈-----	131
圖 5-10 多層貝氏網路第一層分群後分佈-----	132
圖 5-11 多層貝氏網路第二層-----	133
圖 5-12 多層貝氏網路第二層分群後分佈-----	134
圖 5-13 多層貝氏網路第三層-----	135
圖 5-14 多層貝氏網路語音情緒辨識流程-----	135
圖 5-15 多層共變異數貝氏網路第一層-----	136
圖 5-16 多層共變異數貝氏網路第二層-----	136

圖 5-17 多層共變異數貝氏網路第三層-----	137
圖 5-18 MFCC C1 與 C4 平均值四情緒之特徵參數落點分佈-----	137
圖 5-19 MFCC C1 與 C4 平均值四情緒之雙變數常態分佈圖-----	138
圖 5-20 MFCC C1 與 C4 平均值四情緒之雙變數常態分佈投影----	138
圖 5-21 多層共變異數貝氏網路語音情緒辨識流程-----	139
圖 6-1 適應性 KNN 調適流程-----	196
圖 6-2 適應性 MLBN 調適流程-----	196
圖 6-3 MLBN 第一層三種分群方式-----	197
圖 6-4 MLBN 第二層兩種種分群方式-----	198
圖 6-5 $\alpha = 0.5$ 時, 資料庫調適後變化圖-----	198
圖 6-6 $\alpha = 0.9$ 時, 資料庫調適後變化圖-----	199
圖 6-7 適應性 MLBNC 調適流程-----	199
圖 6-8 MLBNC 第一層三種分群方式-----	200
圖 6-9 MLBNC 第二層第一種分群方式-----	200

表目錄：

表 3-1 德國情緒語料庫語料數量-----	44
表 3-2 工業技術研究院語料庫語料數量-----	45
表 3-3 工業技術研究院語料庫語料詞句-----	46
表 3-4 使用之語音特徵參數-----	48
表 3-5 使用之正規化後語音特徵參數-----	49
表 4-1 使用正規化特徵參數 KNN Inside Test 鑑別矩陣-----	67
表 4-2 使用原始特徵參數 KNN Outside Test 鑑別矩陣-----	67
表 4-3 使用正規化特徵參數 KNN Outside Test 鑑別矩陣-----	68
表 4-4 KNN 不同語系測試之鑑別矩陣(正規化特徵參數)-----	68
表 4-5 使用正規化特徵參數 SVM Inside Test 鑑別矩陣-----	69
表 4-6 使用原始特徵參數 SVM Outside Test 鑑別矩陣-----	69
表 4-7 使用正規化特徵參數 SVM Outside Test 鑑別矩陣-----	70
表 4-8 SVM 不同語系測試之鑑別矩陣(正規化特徵參數)-----	70
表 5-1 正規化後音高平均值之正確辨識率矩陣-----	140
表 5-2 正規化後音高之標準差各情緒之正確辨識率矩陣-----	140
表 5-3 情緒類別分群與對應之特徵參數-----	141
表 5-4 使用正規化特徵參數 MLBN Inside Test 鑑別矩陣-----	142
表 5-5 使用正規化特徵參數 BD Inside Test 鑑別矩陣-----	142

表 5-6 使用正規化特徵參數 MLBN Outside Test 鑑別矩陣-----	143
表 5-7 使用原始特徵參數 MLBN Outside Test 鑑別矩陣-----	143
表 5-8 MLBN 不同語系測試之鑑別矩陣(正規化特徵參數)-----	144
表 5-9 使用正規化特徵參數 MLBNC Inside Test 鑑別矩陣-----	144
表 5-10 使用正規化特徵參數 MLBNC Outside Test 鑑別矩陣-----	145
表 5-11 MLBNC 不同語系測試之鑑別矩陣(正規化特徵參數)-----	145
表 5-12 KNN 蒙地卡羅模擬鑑別矩陣-----	146
表 5-13 SVM 蒙地卡羅模擬鑑別矩陣-----	146
表 5-14 MLBN 蒙地卡羅模擬鑑別矩陣-----	147
表 5-15 MLBNC 蒙地卡羅模擬鑑別矩陣-----	147
表 5-16 各分類器蒙地卡羅模擬結果-----	148
表 6-1 KNN 調適前後之鑑別矩陣(工研院 P1)-----	201
表 6-2 KNN 不同語系調適前後之鑑別矩陣(一半調適，一半測試)--	202
表 6-3 MLBN 德國語料庫同語系調適前後之鑑別矩陣-----	203
表 6-4 使用正規化特徵參數 MLBN Dependent Test 鑑別矩陣-----	204
表 6-5 MLBN 不同語系調適前後之鑑別矩陣(一半調適，一半測試) -----	205
表 6-6 調適次數增加時，MLBN 不同語系調適前後之鑑別矩陣-----	206
表 6-7 MLBNC 德國語料庫同語系調適前後之鑑別矩陣-----	207

表 6-8 MLBNC 不同語系調適前後之鑑別矩陣(一半調適，一半測試)	208
表 6-9 調適次數增加時,MLBN 不同語系調適前後之鑑別矩陣	209
表 6-10 各分類於不同測試條件下之整體辨識結果整理	210
表 A-1 KNN Outside Test 各情緒辨識正確率(原始特徵參數)	225
表 A-2 KNN Outside Test 各情緒辨識正確率(正規化特徵參數)	226
表 A-3 SVM Outside Test 各情緒辨識正確率(原始特徵參數)	227
表 A-4 SVM Outside Test 各情緒辨識正確率(正規化特徵參數)	228
表 A-5 MLBN Outside Test 各情緒辨識正確率(原始特徵參數)	229
表 A-6 MLBN Outside Test 各情緒辨識正確率(正規化特徵參數)	230
表 A-7 MLBNC Outside Test 各情緒辨識正確率(正規化特徵參數)	231
表 B-1 正規化後音高平均值各情緒之正確辨識率矩陣	232
表 B-2 正規化後音高之標準差各情緒之正確辨識率矩陣	232
表 B-3 正規化後能量平均值各情緒之正確辨識率矩陣	232
表 B-4 正規化後能量之標準差各情緒之正確辨識率矩陣	233
表 B-5 正規化後第一共振峰平均值各情緒之正確辨識率矩陣	233
表 B-6 正規化後第一共振峰之標準差各情緒之正確辨識率矩陣	233
表 B-7 正規化後第二共振峰平均值各情緒之正確辨識率矩陣	234
表 B-8 正規化後第二共振峰之標準差各情緒之正確辨識率矩陣	234
表 B-9 正規化後第三共振峰平均值各情緒之正確辨識率矩陣	234

表 B-10 正規化後第三共振峰之標準差各情緒之正確辨識率矩陣	235
表 B-11 MFCC C1 平均值各情緒之正確辨識率矩陣	235
表 B-12 MFCC C2 平均值各情緒之正確辨識率矩陣	235
表 B-13 MFCC C3 平均值各情緒之正確辨識率矩陣	236
表 B-14 MFCC C4 平均值各情緒之正確辨識率矩陣	236
表 B-15 MFCC C5 平均值各情緒之正確辨識率矩陣	236
表 B-16 MFCC C6 平均值各情緒之正確辨識率矩陣	237
表 B-17 MFCC C7 平均值各情緒之正確辨識率矩陣	237
表 B-18 MFCC C8 平均值各情緒之正確辨識率矩陣	237
表 B-19 MFCC C9 平均值各情緒之正確辨識率矩陣	238
表 B-20 MFCC C10 平均值各情緒之正確辨識率矩陣	238
表 B-21 MFCC C11 平均值各情緒之正確辨識率矩陣	238
表 B-22 MFCC C12 平均值各情緒之正確辨識率矩陣	239
表 B-23 MFCC C13 平均值各情緒之正確辨識率矩陣	239
表 C-1 KNN 不同語系 P1 ~ P20 調適前後之辨識結果(一半調適，一半 測試)	240
表 C-2 MLBN 德國語料庫同語系 P1 ~ P20 調適前後之辨識結果	242
表 C-3 MLBN 不同語系 P1 ~ P20 調適前後之辨識結果(一半調適，一 半測試)	244

表 C-4 調適次數增加時，MLBN 不同語系 P1 ~ P20 調適前後之辨識
結果-----246

表 C-5 MLBNC 德國語料庫同語系 P1 ~ P20 調適前後之辨識結果--248

表 C-6 MLBNC 不同語系 P1 ~ P20 調適前後之辨識結果(一半調適，
一半測試)-----250

表 C-7 調適次數增加時，MLBN 不同語系 P1 ~ P20 調適前後之辨識
結果-----252



第一章 導論

1.1 前言：

在人與人的溝通互動行為(Human communication)當中，除了文字的內容外，語調(Prosody)、手勢(Hand gestures)以及臉部表情(Facial expression)的呈現，也是溝通行為當中，重要的資訊傳遞方式，透過這樣的資訊傳遞，能夠讓對方明白的感受到彼此要給對方的資訊。在溝通過程傳達的資訊內容中，除了詞面的意義外，人類可以藉由與生俱來的感知能力，敏銳的洞悉許多隱藏在內部的事物，這些內部的相關資訊，深深的影響著詞面上的意義，不同的表達方式與語氣代表著不同的意思，例如一句”你這句話是什麼意思”，如果是以詢問的方式說”你這句話是什麼意思?”，則表示希望講話者重新解釋之前話中的意思，但如果是以”你這句話是什麼意思!”來表示，則表示是質疑對方講話的內容；從上述的例子中，相同的一段話，不同的說話語調會讓聽者有不同的感受，在人類的交談中，常常透過語調、語韻以及音量等不同的變化與修飾，將隱含的資訊附加在語句當中，希望對方能夠從中得到講話者欲表達的資訊。

近幾年來，隨著電腦科技的日新月異，許多研究紛紛投入人機介面(Human Machine Interface)、電腦代理人(Computer Agent)、自然語音辨識(Natural Speech Recognition)等相關領域，這些研究使得電腦與

人之間的互動更加真實，這種人機互動的運行方式，使得人們對於電腦的概念產生變化，電腦不再只是固定執行人給的指令，而是可以跟人產生互動的虛擬機器人。在目前的人類生活中，電腦資訊科技已經扮演了一個不可或缺的角色，舉凡手機、PC、NB 或車用電子，皆使我們的生活更加便利，相信在不久的將來，智慧生活科技將在人類生活中扮演著舉足輕重的角色，而在智慧生活科技之人機互動介面上，機器對人類情緒的感知與理解也是智慧生活科技重要課題之一，如何讓機器具有像人類一樣的情緒感知能力，使得機器能夠像人一樣，更加自然與人性化，將是未來智慧生活科技人機互動上發展的重要方向之一。

1.2 文獻回顧：

Picard[51]於 1998 年針對” Affective Computing” 提出了幾個重要的議題與方向，從電腦辨識之人類情感模型、電腦輔助學習、藝術、娛樂、健康以及人機互動等方向上提出建議與看法，Pantic and Rothkrantz [46]於 2003 年發表的文章中，說明了情緒辨識在人機介面上的重要性以及利用多個不同的感知方式來辨識情緒，Cowie et al.[14]描述了在表達情緒時，語音與臉部表情的情緒特徵與反應，以及在特定的情緒觸發條件下，語音特徵與特定情緒反應之間的對應關係，Scherer[55]針對 2003 年語音情緒辨識的研究現況做回顧，並對語音

辨識研究之未來方向所討論與建議。目前，自動化情緒辨識方法與系統開發已成為人機情感介面上重要的研究主題之一，對於情緒辨識的研究，目前仍處於發展的階段，相關的研究主題主要分為幾個部分：

1. 語音情緒辨識(Speech Emotion Recognition)
2. 臉部影像情緒辨識(Facial Image Emotion Recognition)
3. 語音臉部影像雙模情緒辨識(Audio-Visual Emotion Recognition)
4. 人機互動(Human Machine Interaction)

首先，在語音情緒辨識上，Banse and Scherer[3]針對口語上，情緒所表現之語音特徵做深入探討及研究，Cowie and Cornelius[15]則是描述語音特徵與情緒狀態之間的關連性。由於語音情緒在辨識上，需要以各情緒之語料建置辨識分類所需之資料庫，因此，在語料庫的建置上，Douglas-Cowie et al.[18]建立了 100 個人之情緒資料庫，內容同時包含聲音與影像兩個部分，Schiel et al.[56]利用自行開發的 SmartKom 設備，錄製具有聲音、影像以及觸覺輸入的情緒資料庫，Makarova and Petrushin[38]建置了包含 61 個人五種情緒之俄羅斯語情緒資料庫 RUSLANA，Burkhardt et al.[4] 建置了包含 10 個人 800 個語句之德語情緒資料庫 EMO-DB，Wu et al.[68]則是建置了 68 個人 10 個句子 5 種情緒之華語情緒資料庫 MASC (Mandarin Affective Speech Corpus)，Johnstone [30]利用互動遊戲來當作情緒觸發的機

制，並記錄當時之語音情緒反應，以此為語音情緒語料庫，Schuller et al.[59] 於 2009 年將目前研究語音情緒的九種語料(ABC, AVIC, DES, EMO-DB, eINTERFACE,SAL, SmartKom, SUSAS, and VAM)作連結，並轉換到 Valence-Arousal 空間中加以探討。

在語音情緒辨識之研究上，Scherer[54]從九種不同國家之語料中，探討不同文化及語系之間情緒的表達方式是否有所差異，Montero et al.[41]以西班牙語建立語音情緒之模型並進行分析，Petrushin[49, 50] 使用 23 個人 700 個五種不同情緒之語料，以音高、第一、第二共振峰、能量、以及講話速度為特徵參數，進行不同分類器之辨識實驗，並作為情緒辨識代理人系統以及 Call Centers 之開發應用，Ang et al.[2] 利用語音訊號中之語調、語音模型以及講話者的講話方式，來辨識憤怒與沮喪情緒，Chuang and Wu[11]利用語調及語意，並透過 PCA(Principle Component Analysis)與 SVM 來辨識說話者的情緒狀態，Wu et al.[66]使用語意標示以及 Separable Mixture Models 來辨識語音訊號中的情緒反應，Kwon、Schuller et al.、New et al.、Lee et al.、Jiang and Cai 以及 Lin and Wei[33, 58, 44, 35, 29, 36]使用隱藏式馬可夫模型(Hidden Markov Model, HMM)並以語調相關特徵來進行語音辨識情緒，Fragopanagos and Taylor、Busso et al.以及 Iliou et al.[21, 5, 28]使用類神經網路(Neural Network, NN)來進行語音辨識情緒，Casale

et al.、Chandrakala et al.及 Kandali et al.[6, 7, 32]則使用高斯混合模型(Gaussian Mixture Models) 及支持向量機(Support Vector Machine, SVM)來進行語音辨識情緒，Pao et al.[48]利用最短距離(Nearest class mean)分類法，並使用特徵篩選與特徵結合的方式來改善語者相關之中文語音情緒辨識，Zhang[73] 則是使用模糊最小平方支持向量機來進行語音情緒辨識，並使用不同之訊噪比進行實驗。

除了以分類器來做語音情緒辨識的研究外，有些研究方向則著重於語音情緒特徵參數的計算，Lee et al.以及 Wang et al.[34, 63]利用主成分分析(PCA)先將特徵參數作篩選後，再分別以不同之分類器做語音情緒辨識，Wu et al.及 Meshram et al.[67, 39]利用不同時間序列的特徵參數來進行語音情緒辨識，Espinosa et al.及 Dongrui et al.[19, 17]使用心理學上 VAD(Valence/Activation/Dominance)三維之情緒辨識方法，將特徵參數轉換成 VAD 值，再進行辨識，Xin et al.[69]則使用 HHT(Hilbert-Huang Transformation)方法計算新語音情緒特徵參數，稱為 ECC 參數，並用於語音情緒辨識中。

在臉部影像情緒辨識及語音臉部影像雙模情緒辨識之研究上，Lyons et al.[37]利用 Gabor Wavelet 來擷取臉部表情特徵，並利用 Gabor Wavelet coding 來建立臉部表情之分類器，Fasel and Luetttin[20]提出一自動臉部表情分析之方法，包含臉部圖像的正規化，臉部的動態反應

以及強度等特徵計算，Wilhelm et al.[65]使用統計與類神經網路方法，透過影像分析來辨識使用者表情與性別，Seyedarabi et al.[60]則利用臉部的特徵點當作追蹤點，並使用類神經網路及 FIS(Fuzzy Inference System)來進行臉部影像情緒辨識，Go et al.、Metallinou et al.、Wang et al.、Chen et al.、Das et al.、Zeng et al.以及 Mower et al.[24, 40, 64, 10, 16, 72, 42]使用臉部影像與語音雙模方式作情緒辨識。

在人機互動應用介面及其他應用系統上，Polzin and Waibel[52]建立能表達情緒之人機介面，Fujita et al.[22]介紹仿人機器人 SDR-4X，此機器人具有與人互動並學習的功能，Fujita[23]介紹能與人互動並表達情緒寵物狗 AIBO，Álvarez et al[1]則是開發具有情緒模型之導覽機器人。Huber et al.及 Yacoub et al.[26, 70]將語音情緒辨識方法應用在客服系統(Call Centers)上，Niimi et al.、Iida et al.、Schroder and Grice、Yanushevskaya et al.以及 Tao et al.[45, 27, 57, 71, 62]將情緒因子加入到語音中，使語音合成之訊號帶有情緒反應，Pao and Chen[47]則將中文之語音情緒辨識應用在中文聽障電腦輔助語言教學系統中。

以目前語音情緒辨識的研究發展而言，辨識的效果十分仰賴資料庫的建置，當測試的語者資料不在訓練資料庫中或測試語者所使用的語系與訓練資料有所差異時，通常無法得到良好的辨識效果，因此，

如何透過適當的資料庫調適過程，使資料庫更貼近使用者的真實狀態，將是本論文的重點研究之一。

1.3 本文簡介：

本論文主要在發展一自動化語音情緒辨識系統，透過情緒語音訊號之偵測，將使用者之情緒狀態給辨識出來，在情緒之特徵參數方面，透過音高、能量、共振峰以及梅爾倒頻譜係數之計算，並利用縮小個人差異之正規化過程，將帶有情緒反應之語音正規化特徵參數給計算出來，在情緒辨識方面，本論文使用 KNN、SVM 以及我們所發展之多層貝氏網路與多層共變異數貝氏網路語音情緒辨識方法，對計算之正規化語音情緒特徵進行辨識，並得到辨識後之情緒狀態。

此外，本論文亦提出一針對我們所發展之多層貝氏網路與多層共變異數貝氏網路語音情緒之調適方法，此調適方法能針對使用者的真實語音情緒狀態進行資料庫調適，並得到更貼近使用者情緒狀態的辨識效果。

本文之內容安排如下：

第一章 導論。

第二章 介紹語音訊號之前處理過程與特徵參數之計算方式。

第三章 介紹特徵參數之統計計算與正規化計算方法，並介紹本論文所使用之語料庫。

第四章 介紹 KNN 及 SVM 分類器，並利用 KNN 及 SVM 分類器進行
語音情緒辨識。

第五章 介紹本論文所發展之多層貝氏網路與多層共變異數貝氏網路
語音情緒辨識方法，並進行語音情緒辨識。

第六章 發展多層貝氏網路與多層共變異數貝氏網路語音情緒辨識之
調適方法，並進行適應性實驗。

第七章 總結全文並對未來研究方向提出建議。



第二章 語音情緒特徵參數計算

一般而言，語音情緒辨識的主要流程如圖 2-1 所示，首先，將欲辨識情緒之語音訊號經由訊號擷取裝置(麥克風)，得到語音之類比訊號，經由 A/D 轉換後，將類比訊號轉換成離散時間序列，接著，將所得之離散時間序列經過前處理(Pre-processing)，並將處理過後的離散數位序列經由相關的特徵計算，得到此語音訊號之對應情緒特徵，最後，將此語音訊號所對應之情緒特徵與建置之資料庫作比對，找出相對應之情緒類別。在本章之內容當中，將主要針對情緒辨識流程中的前置處理以及特徵擷取計算的部分，作更詳細的敘述。

2.1 前處理

2.1.1 音框(Frame)與視窗(Windows)

由於語音訊號為一時變性(Time varying)訊號，訊號隨時間的變化非常快速，因此，在處理上，通常會取一小段語音訊號來做處理分析，此種分析方式稱為短時距分析(Short-term analysis)，而此一小段語音訊號則稱為音框(Frame)，如圖 2-2 所示。一般而言，音框的長度通常為 10 ~ 30 毫秒(ms)，音框的移動距離會取音框長度的一半，約 5 ~ 15 毫秒，讓音框與音框之間有所重疊，這樣比較能夠看到語音特徵改變的延續性，取音框的數學表示式如 (2.1.1)式所示。

$$f_x(n; m) = x(n; m)w(n) \quad (2.1.1)$$

其中， m 為音框編號， $x(n; m)$ 為第 m 個音框之語音訊號序列， $w(n)$ 為視窗的加權，一般常用的加權視窗如矩形窗(Rectangular window)與漢明窗(Hamming window)[25]，其數學式與圖分別如(2.1.2)、(2.1.3)及圖 2-3 所示：

$$w(n) = \begin{cases} 1 & 0 \leq n \leq N-1 \\ 0 & otherwise \end{cases} \quad (2.1.2)$$

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2n\pi}{N-1}\right) & 0 \leq n \leq N-1 \\ 0 & otherwise \end{cases} \quad (2.1.3)$$

2.1.2 預強調(Pre-emphasis)

由於人耳聽覺在不同頻率所能接收的最小音強(Intensity)不同，如圖 2-4 所示，人耳對於 1kHz 到 5kHz 的聲音最為敏感，但對於 1kHz 以下與 5kHz 以上的聲音，就需要較大的音強才聽的到，這條臨界曲線稱為聽覺閾(Hearing threshold)。而預強調的最主要工作就是強化高頻訊號部分，使得預強調後的語音訊號能更貼近人類的聽覺系統，其作法為將訊號經過一高通濾波器(High pass filter)，調整高頻成分的權重(Weight)，其計算式如式(2.1.4)所示：

$$H(z) = 1 - a \times z^{-1} \quad (2.1.4)$$

$H(z)$ 為濾波器頻率響應之 z 轉換，經過預強調的音頻訊號，可表示成：

$$S_2[n] = S_1[n] - a \cdot S_1[n-1] \quad (2.1.5)$$

其中 n 為訊號之取樣數， $S_2[n]$ 為經過預強調後的輸出訊號。圖 2-5 為某一段音頻訊號及其經過預強調之結果，圖 2.1-5 為預強調前後語音訊號的頻譜結果，由結果可以看出，在預強調之後，原本所抑制的高頻能量值將被增強，以符合人類的聽覺系統。

2.1.3 快速傅立葉轉換(Fast Fourier Transform, FFT)

由於語音訊號在時域(Time domain)上的多變性與時變性，一般而言，在時域上不容易找出語音相關特徵參數，通常的作法是將訊號從時域轉換到頻域(Frequency domain)上，然後在頻域上找出語音的相關特徵參數。由於語音訊號為數位訊號，因此，在時域轉換到頻域的過程中，需使用到離散傅立葉轉換(Discrete Fourier Transform, DFT)，離散傅立葉轉換的定義如下：設一點數為 N 的有限離散訊號 $x(n)$ ， $0 \leq n \leq N-1$ ，其離散傅立葉轉換為 $X(k)$ ，轉換式形式為：

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-i \frac{2\pi}{N} nk} = \sum_{n=0}^{N-1} x(n) W^{nk} \quad k = 0, 1, \dots, N-1 \quad (2.1.6)$$

離散逆轉換為

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{i \frac{2\pi}{N} nk} = \frac{1}{N} \sum_{k=0}^{N-1} X(k) W^{-nk} \quad k = 0, 1, \dots, N-1 \quad (2.1.7)$$

其中 $W = e^{-i2\pi/N}$ 。將上列兩式改寫成矩陣形式，如(2.1.8)式及(2.1.9)式所示：

$$\mathbf{X} = \mathbf{W}^{nk} \mathbf{x} \quad (2.1.8)$$

$$\mathbf{x} = \frac{1}{N} \mathbf{W}^{-nk} \mathbf{X} \quad (2.1.9)$$

式中 \mathbf{X} 與 \mathbf{x} 都是 N 維向量， \mathbf{W}^{nk} 與 \mathbf{W}^{-nk} 為 $N \times N$ 維矩陣。

離散傅立葉轉換讓傅立葉轉換能實際應用於電腦上，但卻因為計算量太大，使得應用上受到限制。要計算 N 點的 $X(k)$ 需要 N^2 次的複數乘法以及 $N(N-1)$ 次的複數加法，且因為一次複數乘法等於四次實數相乘與兩次實數相加，一次複數相加等於兩次實數相加，當 N 很大時，其計算量相當可觀。舉例來說，若處理點數為 1024 點的訊號，需要做 1048576 次的複數相乘，也就是 4194304 次的實數相乘，計算量相當驚人。

由於計算量過大，離散傅立葉轉換並未被廣泛使用，直到 1965 年由 Cooley 和 Tukey 提出了快速傅立葉轉換(Fast Fourier Transform, FFT)的計算方法[12]，這才使得離散傅立葉轉換被大量應用於工程分析上，FFT 的計算方式是利用 \mathbf{W} 的週期性與對稱性來減少計算量，其詳細推導過程可以參考數位訊號處理的相關書籍。

2.2 語音特徵(Speech Feature)計算

2.2.1 音高(Pitch)

在一段語音信號的形成過程中，首先會先由聲音源(Sound source)發出一定頻率的弦波訊號後，接著聲音經由口腔、舌和嘴唇等所形成

的濾波器(Filters)，將聲源信號做調變，便可以發出不同聲調的聲音。而所謂的音高(或者稱為基頻, Pitch)，意指發音源最原始的振動頻率；通常要精確的去找出音高是不太容易的事，一般常用的方法，有自相關函數分析(Autocorrelation Function Analysis) [53]、倒頻譜演算法(Cepstrum Algorithm) [43]以及小波轉換(Wavelet Transform)[31]等。

音高為語音辨識中重要的特徵之一，只有在有聲語音(Voiced signal)中，才會有音高的存在，而人類的情緒則是隱藏於有聲的語音信號中，圖 2-7 為一段語音訊號經由自相關函數分析所計算出來的音高輪廓(Pitch contour)，本文之音高計算將採用自相關函數分析作為音高的計算方法，其主要步驟如下：

- (1)、取出一段經過音框化之語音訊號，將此語音訊號序列經過快速傅立葉轉換後，得到一頻譜參數序列。

$$FSpectrum(n;m) = FFT(x[n]) \quad (2.2.1)$$

其中 m 表示第 m 個音框， $x[n]$ 表示第 m 個音框內的語音訊號序列。

- (2)、設定一頻寬(50~1000Hz)，將此頻率範圍內之音框頻譜做自相關分析，如式(2.2.2)所示。

$$\begin{aligned} ACF_FSpectrum(i;m) \\ = \sum_{i=1}^{n-1-\tau} FSpectrum(i;m)FSpectrum(i+\tau;m) \end{aligned} \quad (2.2.2)$$

- (3)、找出自相關分析 $ACF_FSpectrum(i;m)$ 除了第一個值外，最大值的位置，並由式(2.2.3)計算出對應的頻率。

$$pitch = \frac{Fs}{N \cdot (k-1)} \quad (2.2.3)$$

其中， N 為音框長度， k 表示除了第一個值外，最大值的位置點。

2.2.2 共振峰(Formant)

人類的聲帶振動經過聲道以及口腔後，會產生共振發出聲音，取一段經音框化後之音頻訊號，將音頻訊號利用快速傅立葉轉換將訊號轉換到頻譜上，將傅立葉轉換之頻譜取其包絡線，可得到一條較為緩和的頻譜曲線，在曲線上可以看到若干個高點，這些高點代表能量集中的頻率位置，這些能量集中的位置就是共振峰(Formant)之所在位置。如圖 2-8 所示，第一個高點通常為音高位置，標示為 F_0 ，第二個高點稱為第一共振峰，標示為 F_1 ，第三個高點稱為第二共振峰，標示為 F_2 ，依次類推。如果發出的聲音是聲帶不動的輔音，則在頻譜上就看不到基頻與共振峰，而且能量會比較集中在高頻處。

2.2.3 音框能量(Frame Energy)

在語音特徵中，聲音強度的變化是相當重要的訊息，聲音強度與波形振幅有關，振幅越大音強越大，在固定音框長度的情況下，音框能量可以以式(2.2.4)表示

$$En(m) = \sum_{n=mL-N+1}^m |x(n)|^2 \quad (2.2.4)$$

其中， m 為音框編號，音框與音框相距為 L 。

由於人耳對音強的感知並非線性，而是接近於對數的曲線，所以

一般時候我們會將能量以對數方式表示，

$$En(m) = \log \left[\sum_{n=mL-N+1}^m |x(n)|^2 \right] \quad (2.2.5)$$

我們從訊號起始端開始，第一個音框編號為 1，以此類推，我們可以得到一音框能量變化的序列，圖 2-9 為一語音訊號的音框能量曲線，

2.2.4 梅爾頻率倒頻譜係數(Mel-frequency Cepstral coefficient, MFCC)

人耳在頻域上的感知並非全頻域有相同的敏感度，在正常的情況下，對於低頻有較高的解析度，也就是在低頻可以分辨較小的頻率差異，此外還有臨界頻帶的現象，在 1kHz 頻率以下的臨界頻帶寬度約為 100Hz，1kHz 頻率以上的臨界頻帶寬度成指數增加。因此，配合人耳聽覺特性，在頻域中以梅爾(Mel-frequency)劃分頻帶，將屬於一個頻帶中的頻率成分，合在一起當作一個能量強度，然後將這些頻帶強度，以離散餘弦轉換(DCT)，轉換成倒頻譜，其轉換方法如下，首先，設計一組梅爾頻率的帶通濾波器，來得到通過帶通的音強，用以計算倒頻譜，圖 2-10 為三角形濾波器所組成的梅爾濾波器組，梅爾頻率刻度是以 1kHz 以下為等間距，1kHz 以上為對數間隔，在 4kHz 範圍內設計成 20 個頻帶，其中心頻率設定成；100，200，300，400，500，600，700，800，900，1000，1148，1318，1514，1737，1995，2291，2630，3020，3467，4000Hz。以數學式表示，第 m 個濾波器

的函數式如下：

$$B_m(k) = \begin{cases} 0 & k < f_{m-1} \\ \frac{k - f_{m-1}}{f_m - f_{m-1}} & f_{m-1} \leq k \leq f_m \\ \frac{f_{m+1} - k}{f_{m+1} - f_m} & f_m \leq k \leq f_{m+1} \\ 0 & f_{m+1} < k \end{cases} \quad 1 \leq m \leq M \quad (2.2.6)$$

$B_m(k)$ 表示是第 m 個頻帶的三角形濾波器， f_m 為第 m 個頻帶的中心頻率， f_{m-1} 與 f_{m+1} 就是前後兩個頻帶的中心頻率， M 為全部的頻帶數目。

將各頻率的能量，乘上三角形濾波器，然後累加起來，就是通過這個濾波器的能量，取對數值，得到：

$$Y(m) = \log \left[\sum_k |X(k)|^2 B_m(k) \right] \quad (2.2.7)$$

對全部 M 個濾波器輸出的對數能量，做離散餘弦轉換 (Discrete Cosine Transform, DCT)，得到梅爾頻率倒頻譜。

$$c_x''(n) = \frac{1}{M} \sum_{m=1}^M Y(m) \cos \left(\frac{\pi n (m - 1/2)}{M} \right) \quad (2.2.8)$$

$c_x''(n)$ 就是訊號 $x(n)$ 的梅爾倒頻譜係數 (Mel-frequency Cepstral Coefficient, MFCC)，在本論文中只用前 13 個，即 $n = 1, 2, \dots, 13$ 來作為語音的頻譜特徵。

2.3 小結

在本章的內容中，主要針對特徵擷取前的分析以及之後的特徵擷

取計算，由於人類對於不同振幅以及頻率，也就是音量以及音調的感知是相當敏感的，且人耳對於聲音的接收也是透過頻率的轉換來接收，因此，頻譜上的特徵計算在情緒特徵擷取上是相當重要的；而由於語音訊號常常伴隨著雜訊，而雜訊在頻譜上往往與語音訊號是互相重疊的，因此，環境與麥克風的品質亦深深影響語音的特徵擷取。此外，人與人之間在語音特徵上的差異也是造成目前語音情緒辨識率不佳的問題所在，如何縮小人與人之間在語音特徵上的差異，並經由適當的調適方法來改善辨識上的問題，將是本研究發展的另外一個課題所在，以下章節將持續針對縮小特徵差異、辨識分類器以及調適方法作更深入的探討。



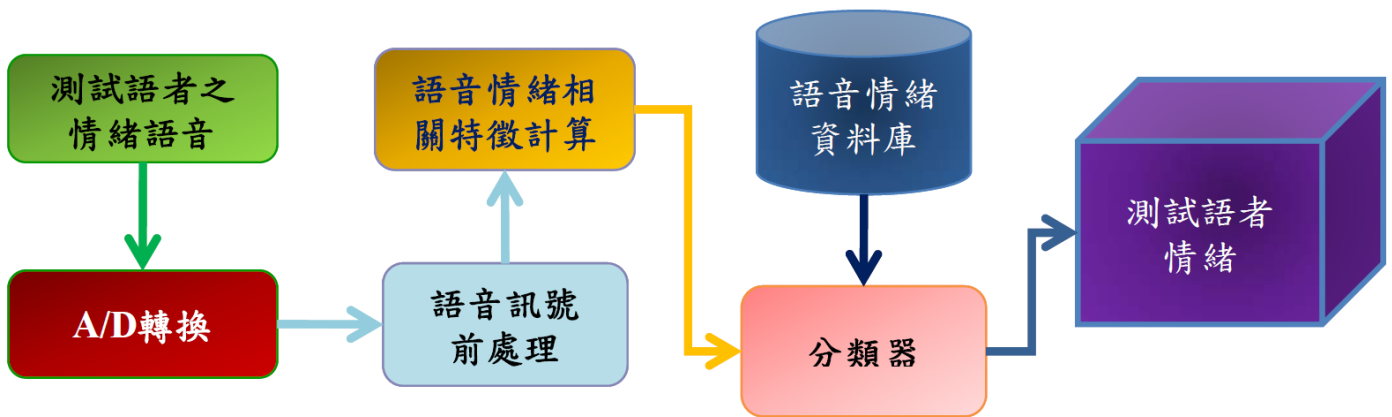


圖 2-1 語音情緒辨識流程圖

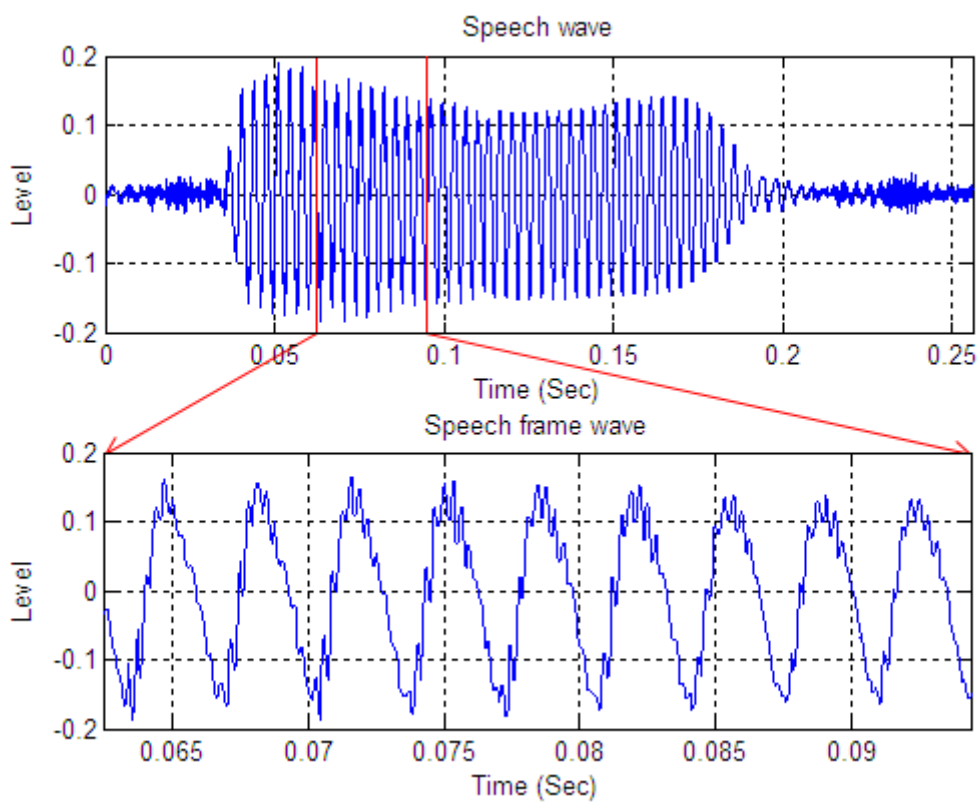


圖 2-2 語音訊號音框化後之波形

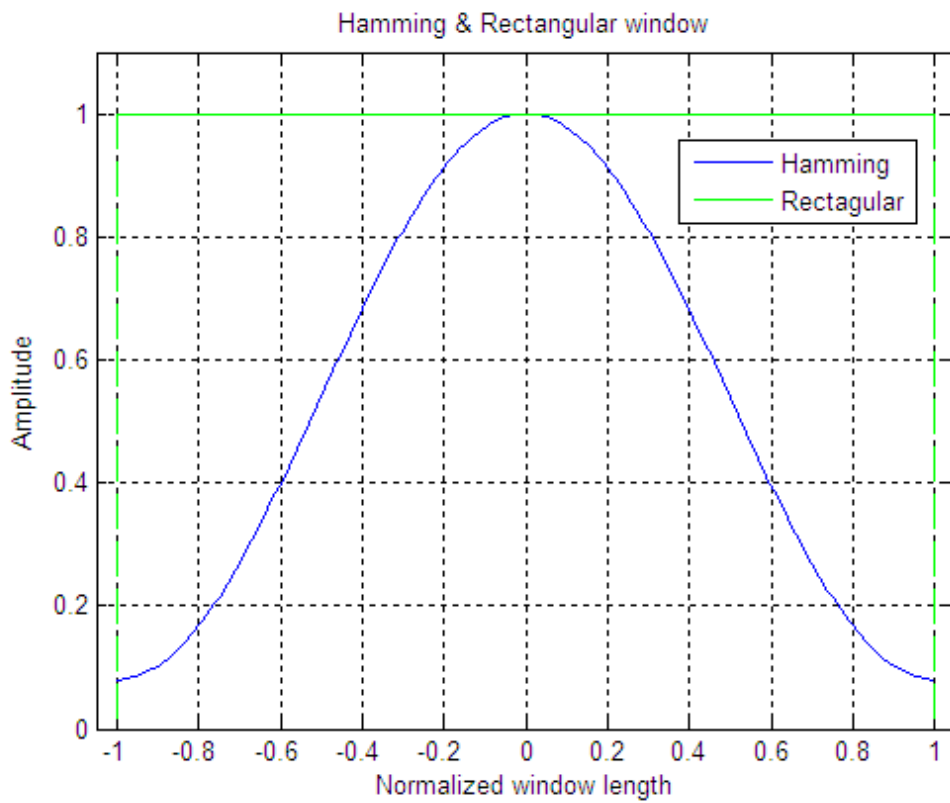


圖 2-3 Hamming and Rectangular window

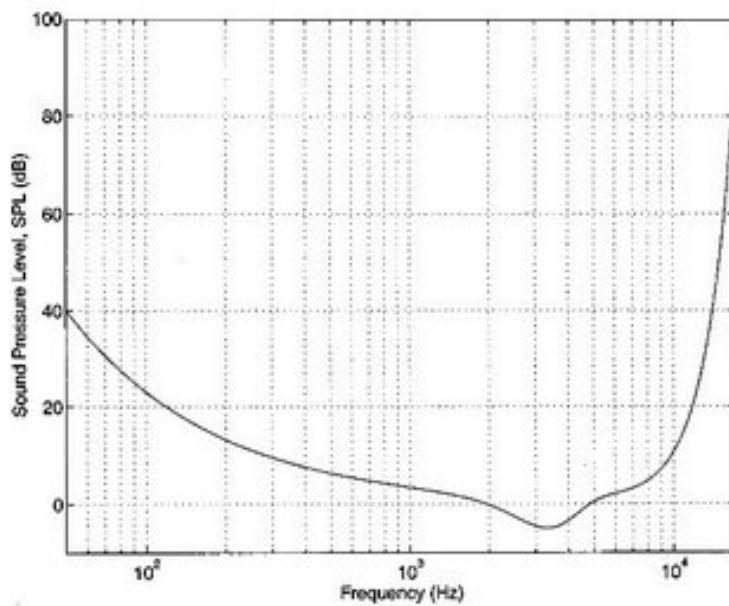


圖 2-4 人耳聽覺各頻率最小音壓曲線

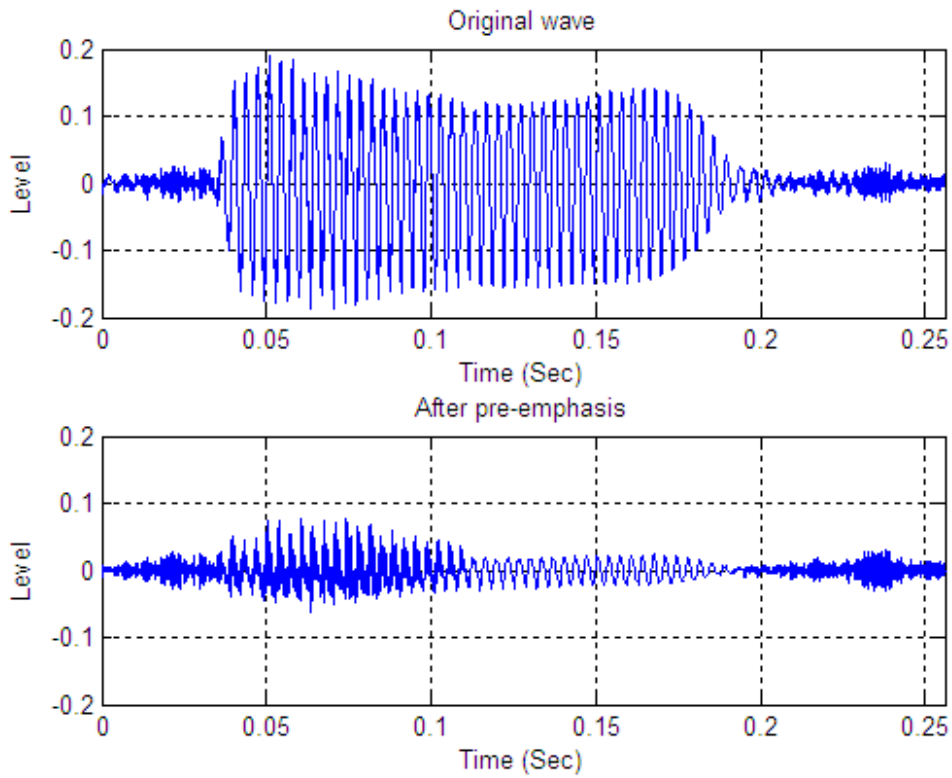


圖 2-5 預強調前後之語音波形

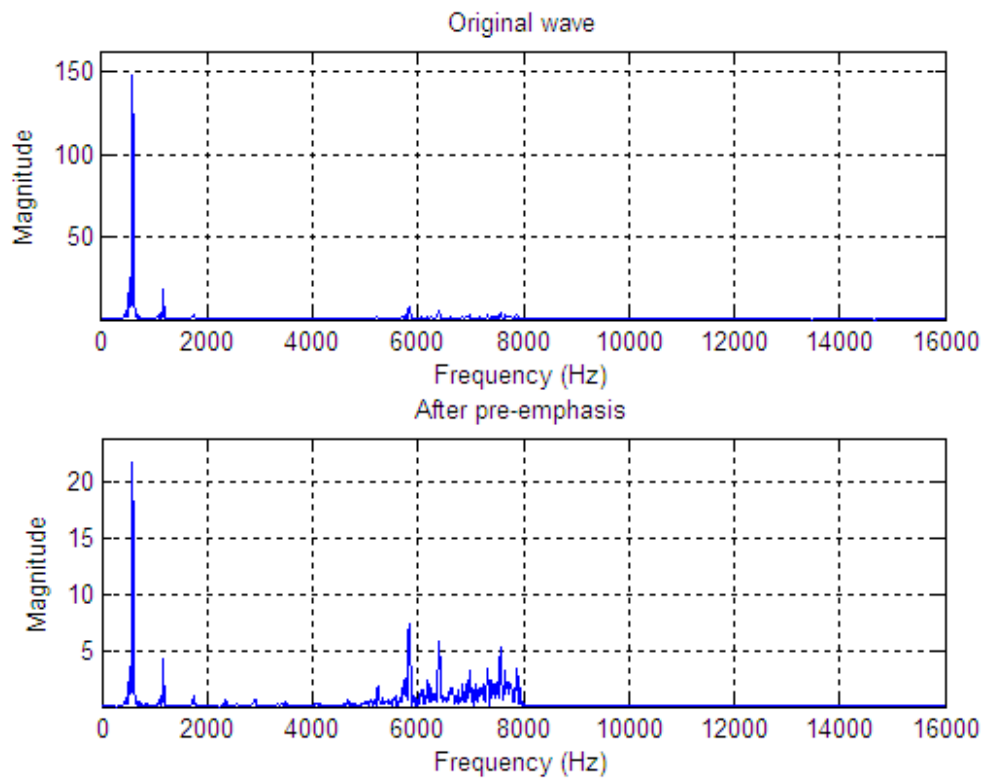


圖 2-6 預強調前後之語音頻譜

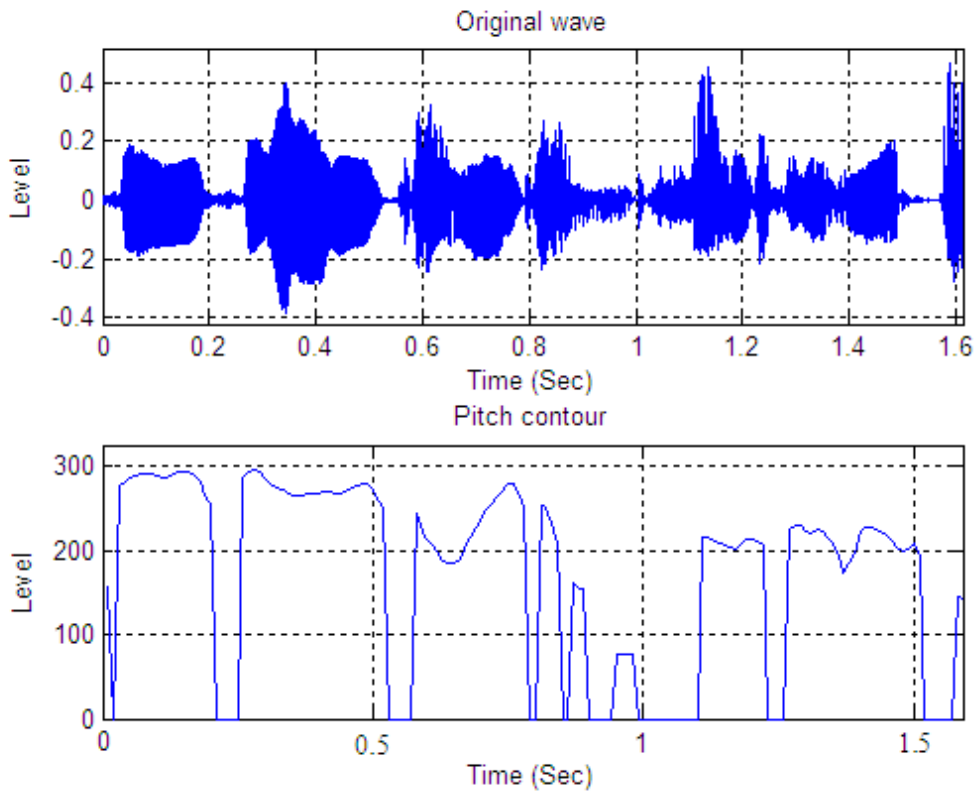


圖 2-7 語音之音高變化

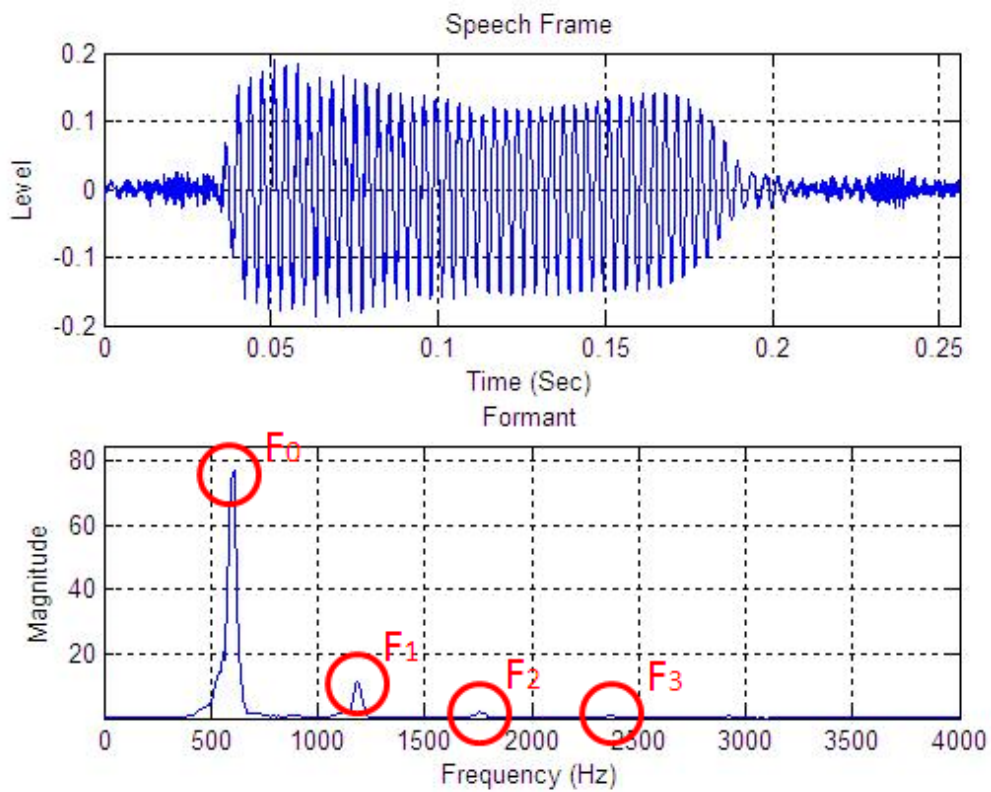


圖 2-8 單一音框之共振峰示意圖

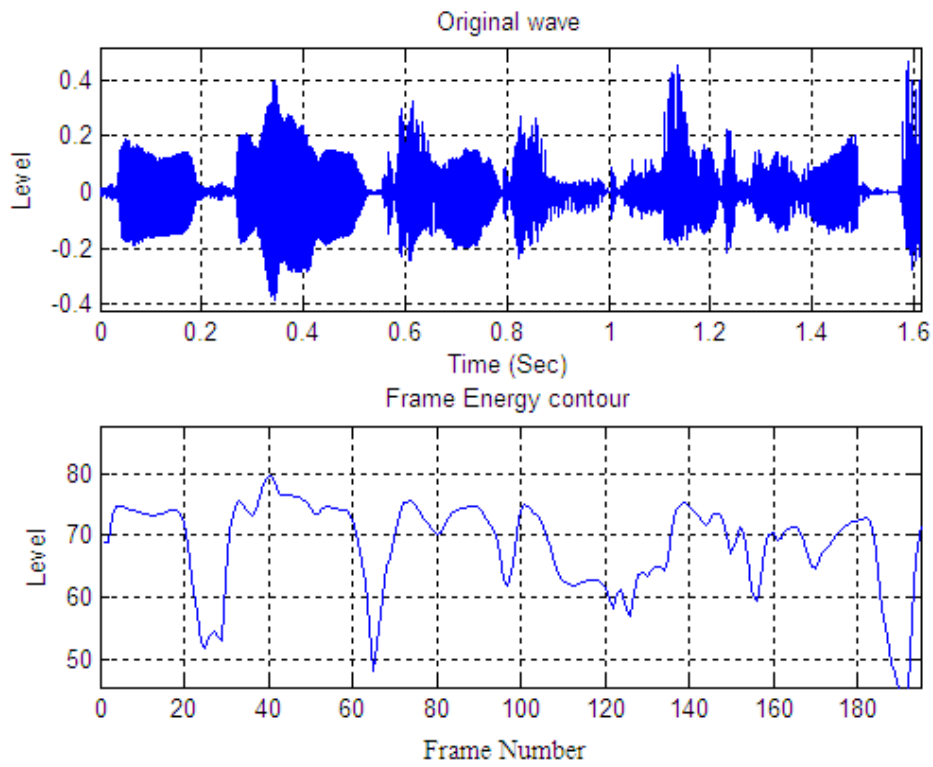


圖 2-9 語音之音框能量變化

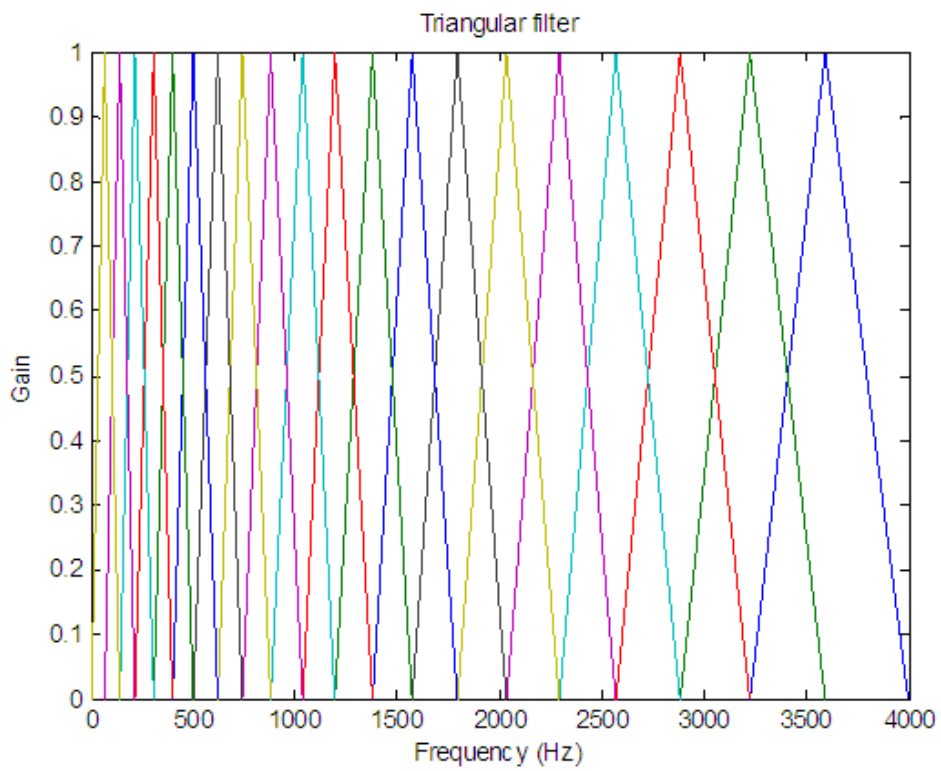


圖 2-10 三角形濾波器組成的梅爾頻率濾波器組

第三章 特徵參數之統計計算與正規化

在語音情緒辨識的流程中，其中一個重要的環節就是如何從現有的資訊中找出線索，並與資料庫作比對，前面一章所講到的特徵計算，對一段長度約 3 秒的語音訊號而言，計算出來的特徵參數是非常多的，因此，在特徵計算與資料庫特徵建置上，本文將採用統計的方式來得到相關的特徵。此外，由於不同語者在語音特徵上的差異，容易造成辨識的困難，因此，本章亦將提出一正規化的計算方法，縮小不同語者在語音特徵上的差異，期望對後續的辨識分類有所助益。

3.1 情緒語音資料庫

本論文所使用之情緒語音語料庫是 W. F. Sendlmeier 等人於 Technical University of Berlin 所蒐集的情緒語料庫，此情緒語料庫為開放資源，內容包含十名演員(五名男士及五名女士)，分別就十個單字與五個句子，以不同情緒發音，語料庫中包含悲傷 62 筆、中性 79 筆、快樂 71 筆及生氣 127 筆，共 339 筆語料(如表 3-1 所示)，主要當作訓練與辨識用。此外，本文亦使用工業技術研究院(Industrial Technology Research Institute, ITRI)所自行錄製之情緒語料庫，此情緒語料庫為工研院版權所有，內容包含女性十名，男性十名，每名語者

四種情緒各 30 筆語料(如表 3-2 所示)，各情緒之語料語句則如表 3-3 所列，錄製環境如圖 3-1 所示，錄音設備則包含 RODE NT-3 麥克風、E-MU 0404 USB 外接音效卡及一台桌上型 PC，錄音軟體則是利用自行撰寫之 MATLAB 程式，以單聲道、取樣頻率 16kHz 及取樣解析度 16bit 錄製語料，在錄製的過程中，會先讓錄音語者聆聽事先以語音情緒表達能力較高之語者所先行錄製之語音情緒範例，並解釋各情緒的可能表現方式，接著，會讓錄音語者對欲錄製的情緒類別試講數次，待情緒表達明確後，再進行錄音語者之情緒語音錄製，在錄製的過程中，若錄音語者情緒表達不明顯時，則會對該語句進行重錄，直至情緒表現明確為止。

3.2 特徵統計計算

3.2.1 語音特徵平均值(Mean)與標準差(Standard deviation)

針對一段長度 1.5 ~ 3 秒，取樣頻率 16kHz 的語音訊號而言，若音框長度為 256 點，音框與音框之間重疊的部分為 128 點，其一段語音訊號的音框數約為 188 ~ 376 個，因此，將計算所得之語音參數利用統計方式，計算出與統計分布情況有關的統計量值，如：最小值、最大值、中間值及標準差等統計參數，並利用設定閾值(Threshold)的方式，將雜訊或不必要的資料去除，以得到與情緒相關性較大的統計量值，是本文採用的特徵計算方法。以下將針對資料庫中不同語音特

徵的計算與處理方式作說明：

首先，計算資料庫中的音高特徵，將每一段資料庫中的語音訊號，以音框長度 256 點，音框與音框之間重疊部分 128 點，計算出每一個音框的音高值，接著，設定音高的上閾值(Upper bound)為 550Hz，下閾值(Lower bound)為 75Hz，如式(3.2.1)所示，這是由於人類發音的母音音高通常介於 75 ~ 550Hz 之間，利用閾值的限制，可以去除非語音以及環境雜訊的干擾。

$$P_{i,j}^{(E)}(m) = \begin{cases} P_{i,j}^{(E)}(m) & \text{if } 75 \leq P_{i,j}^{(E)}(m) \leq 550 \\ 0 & \text{otherwise} \end{cases} \quad (3.2.1)$$

其中， (E) 表示情緒類別， i 表示第 i 個人的語音， j 表示在 Person i 及情緒類別 (E) 下，第 j 筆語音情緒資料， m 則表示那一筆語音情緒資料中第 m 個音框的音高值，舉例來說， $P_{3,2}^{(H)}(100)$ 表示第 3 個人中，情緒類別為”快樂”資料下，第 2 筆資料第 100 個音框的音高值。

此外，我們建立一個以 $P_{i,j}^{(E)}(m)$ 為判斷基準的旗標(Flag)，並藉此旗標來決定此音框是否為說話音框的根據，其算式如(3.2.2)所示：

$$flag_{i,j}^{(E)}(m) = \begin{cases} 1 & \text{if } P_{i,j}^{(E)}(m) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.2.2)$$

當 $flag_{i,j}^{(E)}(m) = 1$ 表示此音框為說話音框， $flag_{i,j}^{(E)}(m) = 0$ 則否， $flag_{i,j}^{(E)}(m)$ 中各符號的意義與 $P_{i,j}^{(E)}(m)$ 相同。

接著，將有語音部分的音高值以統計方式，計算其平均值(Mean)

與標準差(Standard deviation)，其算式如(3.2.3)及(3.2.4)所示：

$$\overline{P}_{i,j}^{(E)} = \frac{1}{N} \sum_{m=1}^n P_{i,j}^{(E)}(m) \cdot flag_{i,j}^{(E)}(m) \quad (3.2.3)$$

$$\sigma_{P,i,j}^{(E)} = \sqrt{\frac{1}{N} \sum_{m=1}^n \left[\left(P_{i,j}^{(E)}(m) \right) - \left(\overline{P}_{i,j}^{(E)} \right) \right]^2} \quad (3.2.4)$$

其中， N 表示 $flag_{i,j}^{(E)}(m)=1$ 之音框總數。最後，我們得到音高的平均值與標準差，資料庫中的每一筆情緒語音皆會得到一個音高的平均值與標準差，舉例來說，在 Person i 中，我們有四種情緒(喜、怒、哀以及中性)各五筆資料，將音高的平均值設為第一個特徵參數，標準差設為第二參數，則我們的資料庫中第一個特徵參數將有四個類別，而每個類別則分別有五個音高平均值特徵值，第二個特徵也同樣具有四個類別，每個類別分別有五個音高標準差特徵值，以此類推，以下的各個特徵參數將以同樣的統計計算方式建立資料庫。

接著，計算音框能量及音框能量標準差兩個特徵參數，先計算每一個音框的能量序列 $En_{i,j}^{(E)}(m)$ ，當 $flag_{i,j}^{(E)}(m)=1$ 時，將所對應之音框能量以統計方式，計算其平均值，得到平均音框能量 $\overline{En}_{i,j}^{(E)}$ 及其標準差 $\sigma_{En,i,j}^{(E)}$ ，如式(3.2.5)及(3.2.6)所示。

$$\overline{En}_{i,j}^{(E)} = \frac{1}{N} \sum_{m=1}^n En_{i,j}^{(E)}(m) \cdot flag_{i,j}^{(E)}(m) = 1, \quad (3.2.5)$$

$$\sigma_{En,i,j}^{(E)} = \sqrt{\frac{1}{N} \sum_{m=1}^n \left(En_{i,j}^{(E)}(m) \right)^2 - \left(\overline{En}_{i,j}^{(E)} \right)^2} \quad (3.2.6)$$

接下來，與音高以及音框能量之計算方式相同，計算語音當中的

共振峰值，並取出與情緒關連性較大的第一至第三共振峰值($F1 \sim F3$)，將第一至第三共振峰值以統計方式，計算其平均值與標準差，其算式如(3.2.7) ~ (3.2.12)所示

$$\overline{F1}_{i,j}^{(E)} = \frac{1}{N} \sum_{m=1}^n F1_{i,j}^{(E)}(m) \cdot flag_{i,j}^{(E)}(m) = 1 \quad (3.2.7)$$

$$\sigma_{F1,i,j}^{(E)} = \sqrt{\frac{1}{N} \sum_{m=1}^n (F1_{i,j}^{(E)}(m))^2 - (\overline{F1}_{i,j}^{(E)})^2} \quad (3.2.8)$$

$$\overline{F2}_{i,j}^{(E)} = \frac{1}{N} \sum_{m=1}^n F2_{i,j}^{(E)}(m) \cdot flag_{i,j}^{(E)}(m) = 1 \quad (3.2.9)$$

$$\sigma_{F2,i,j}^{(E)} = \sqrt{\frac{1}{N} \sum_{m=1}^n (F2_{i,j}^{(E)}(m))^2 - (\overline{F2}_{i,j}^{(E)})^2} \quad (3.2.10)$$

$$\overline{F3}_{i,j}^{(E)} = \frac{1}{N} \sum_{m=1}^n F3_{i,j}^{(E)}(m) \cdot flag_{i,j}^{(E)}(m) = 1 \quad (3.2.11)$$

$$\sigma_{F3,i,j}^{(E)} = \sqrt{\frac{1}{N} \sum_{m=1}^n (F3_{i,j}^{(E)}(k))^2 - (\overline{F3}_{i,j}^{(E)})^2} \quad (3.2.12)$$

最後，計算梅爾頻率倒頻譜係數(MFCC)，並取前 13 個係數當作語音情緒辨識之特徵參數，同樣的，以音高所得到之旗標 $flag_{i,j}^{(E)}(m)$ 來當作 MFCC 取捨的依據，將 $flag_{i,j}^{(E)}(m)=1$ 時，所對應之音框的 MFCC 各係數以統計方式，計算其平均值，得到 $\overline{MFCC}_{C1,i,j}^{(E)}$ ， $\overline{MFCC}_{C2,i,j}^{(E)}$ ，.....， $\overline{MFCC}_{C13,i,j}^{(E)}$ ，如式(3.2.13)所示， $\overline{MFCC}_{Ct,i,j}^{(E)}$ 中各符號的意義與 $P_{i,j}^{(E)}(m)$ 相同， i 表示第 i 個人的語音， (E) 表示情緒類別， j 表示在 Person i 及情緒類別 (E) 下，第 j 個語音情緒資料， Ct 表示 MFCC 中第 t 個係數。

$$\overline{MFCC}_{Ct,i,j}^{(E)} = \frac{1}{N} \sum_{n=1}^n MFCC_{Ct,i,j}^{(E)}(m) \cdot flag_{i,j}^{(E)}(m) = 1, \quad (3.2.13)$$

表 3-4 為本文所採用之語音特徵參數列表，在本論文中所使用之語音特徵參數包含音高之平均值 $\overline{P}_{i,j}^{(E)}$ 與標準差 $\sigma_{P,i,j}^{(E)}$ 、音框能量平均值 $\overline{En}_{i,j}^{(E)}$ 及其標準差 $\sigma_{En,i,j}^{(E)}$ 、第一至第三共振峰之平均值 $\overline{F1}_{i,j}^{(E)} \sim \overline{F3}_{i,j}^{(E)}$ 與標準差 $\sigma_{F1,i,j}^{(E)} \sim \sigma_{F3,i,j}^{(E)}$ 、以及 MFCC 前十三個係數 $\overline{MFCC}_{C1,i,j}^{(E)}$ ， $\overline{MFCC}_{C2,i,j}^{(E)}$ ，……， $\overline{MFCC}_{C13,i,j}^{(E)}$ 平均值，共 23 個語音特徵參數。

3.2.2 語音特徵正規化計算(Normalization)

雖然人與人之間表達情緒的模式會因為成長環境的不同而有些許差異，但在語音上，其各情緒在語調上的差異卻是相近的，每個人在不同情緒的音調分佈趨勢，基本上是很相近的，唯一的差別是音調與發音頻率位置的不同，例如，男生普遍聲音較為低沈，而女生則較為高亢，圖 3-2 為德國語料庫中十個不同說話者在四種情緒上其音高之分佈狀況，圖中之橫軸為音高值，縱軸為語料庫中不同說話者之代號，由圖中可以發現，不同說話者之不同情緒在音高上的分佈是非常分散不集中的，例如 Person 1 的四種情緒其音高分佈範圍為悲傷 93.33 ~ 123.35Hz、中性 109.2 ~ 135.4Hz、快樂 168.2 ~ 255.02Hz、生氣 172.12 ~ 258.31Hz，而 Person2 的四種情緒其分佈範圍則為悲傷 143.72 ~ 172.78Hz、中性 180.99 ~ 222.05Hz、快樂 208.53 ~ 332.47Hz、生氣 277.36 ~ 391.64Hz，若單純利用音高來將四種情緒作分類，則 Person

1 與 Person 2 之間，相同情緒的音高值差異太大，容易造成分類的困難，進而影響辨識結果，圖 3-3(a) ~ (w) 為 23 個不同特徵參數在不同語者之間的分布狀況，由圖中可以看出，由於人與人之間因發音習慣不同進而造成特徵參數之間的差異，因此，如何利用計算方法縮小人與人之間在語音特徵上的差異，是辨識上很重要的課題之一。

在此，本論文提出一對於語音情緒特徵正規化的計算方法，藉此計算方法縮小人與人之間在特徵參數上的差異，透過這樣的計算，能讓不同說話者在各情緒之間的特徵參數差異性縮小，增加各情緒的辨識機率。本正規化的計算方式，是以不同語者之中性特徵參數為正規化的基準，首先，先計算資料庫中，中性情緒之所有資料庫特徵參數的平均值，以音高為例，其計算式如(3.2.14)所示：

$$\overline{P}_{i,mean}^{(N)} = \frac{1}{r} \sum_{j=1}^r \overline{P}_{i,j}^{(N)} \quad (3.2.14)$$

其中 i 表示第 i 個人的語音， (N) 表示情緒類別為中性， $mean$ 表示在 Person i 及中性情緒類別下，其所有中性資料的平均值， r 則表示中性語音情緒資料 (N) 中具有 r 筆資料，舉例來說， $\overline{P}_{3,mean}^{(N)}$ 且 $r=5$ ，表示 $\overline{P}_{3,mean}^{(N)}$ 此值為第 3 個人其資料庫中 5 筆中性語音資料音高的平均值。

接著，利用不同語者中性情緒之平均音高值的來對資料庫中不同語者各情緒的原始音高資料 $\overline{P}_{i,j}^{(E)}$ 作正規化，其算式如(3.2.15)：

$$\overline{P}_{i,normj}^{(E)} = \frac{\overline{P}_{i,j}^{(E)}}{\overline{P}_{i,mean}^{(N)}} \quad (3.2.15)$$

其中， $\overline{P}_{i,normj}^{(E)}$ 表示資料庫 $\overline{P}_{i,j}^{(E)}$ 經過正規化之音高值，從式(3.2.15)來看，其代表的意義是以每一個語者的中性音高為基準，將每個人的其他情緒音高特徵在同一基準下展開。原始特徵的資料庫中音高平均值經正規化後，其分佈如圖 3-4 所示，由圖中可以看出，悲傷(S)、中性(N)明顯形成兩個群聚(Cluster)，但快樂以及生氣在音高的分佈上還是較廣，沒有明顯的群聚效果。式(3.2.15)中是以 $\overline{P}_{i,mean}^{(N)}$ 為基準，這是由於中性的音高值在分佈上較為集中，而其他情緒在分佈上較廣，以 $\overline{P}_{i,mean,j}^{(N)}$ 作為正規化的基準，比其他情緒更具客觀性。

音高標準差 $\sigma_{P,i,j}^{(E)}$ 正規化的計算方式與音高計算方式相似，一樣採用中性情緒之平均值當作基準，如式(3.2.16)所示，而其正規化計算方法則如式(3.2.17)所示：

$$\overline{\sigma}_{P,i,mean}^{(E)} = \frac{1}{r} \sum_{j=1}^r \sigma_{P,i,j}^{(E)} \quad (3.2.16)$$

$$\overline{\sigma}_{P,i,normj}^{(E)} = \frac{\sigma_{P,i,j}^{(E)}}{\overline{\sigma}_{P,i,mean}^{(N)}} \quad (3.2.17)$$

其中， $\overline{\sigma}_{P,i,normj}^{(E)}$ 表示資料庫 $\sigma_{P,i,j}^{(E)}$ 經過正規化之音高標準差值，圖 3-5 為音高標準差正規化後的分佈圖，由結果可以看出，與音高平均值類似，悲傷(S)、中性(N)明顯形成兩個群聚，但快樂以及生氣則沒有明顯的群聚效果。接著，以同樣的計算方式對音框能量以及共振峰等參

數作正規化，其算式如(3.2.18) ~ (3.2.25)所示。

$$\overline{En}_{i,normj}^{(E)} = \frac{\overline{En}_{i,j}^{(E)}}{\overline{En}_{i,mean}^{(N)}} \quad (3.2.18)$$

$$\overline{\sigma}_{En,i,normj}^{(E)} = \frac{\sigma_{En,i,j}^{(E)}}{\overline{\sigma}_{En,i,mean}^{(N)}} \quad (3.2.19)$$

$$\overline{F1}_{i,normj}^{(E)} = \frac{\overline{F1}_{i,j}^{(E)}}{\overline{F1}_{i,mean}^{(N)}} \quad (3.2.20)$$

$$\overline{\sigma}_{F1,i,normj}^{(E)} = \frac{\sigma_{F1,i,j}^{(E)}}{\overline{\sigma}_{F1,i,mean}^{(N)}} \quad (3.2.21)$$

$$\overline{F2}_{i,normj}^{(E)} = \frac{\overline{F2}_{i,j}^{(E)}}{\overline{F2}_{i,mean}^{(N)}} \quad (3.2.22)$$

$$\overline{\sigma}_{F2,i,normj}^{(E)} = \frac{\sigma_{F2,i,j}^{(E)}}{\overline{\sigma}_{F2,i,mean}^{(N)}} \quad (3.2.23)$$

$$\overline{F3}_{i,normj}^{(E)} = \frac{\overline{F3}_{i,j}^{(E)}}{\overline{F3}_{i,mean}^{(N)}} \quad (3.2.24)$$

$$\overline{\sigma}_{F3,i,normj}^{(E)} = \frac{\sigma_{F3,i,j}^{(E)}}{\overline{\sigma}_{F3,i,mean}^{(N)}} \quad (3.2.25)$$



最後，由於梅爾頻率倒頻譜係數(MFCC)本身為梅爾頻率倒頻譜之係數值，各情緒之間並無明顯的趨勢，因此，梅爾頻率倒頻譜係數(MFCC)在此不做任何正規化的計算。圖 3-6(a) ~ (i)為音高平均值、音高標準差等特徵參數正規化後資料庫各特徵資料的分佈狀況，由圖中顯示，正規化的過程使得特徵在人與人之間的差異性縮小，將有助於後續的分類運算。

表 3-5 為本論文所使用之正規化後語音特徵參數列表，其中包含正規化後音高之平均值 $\overline{P}_{i,normj}^{(E)}$ 與標準差 $\sigma_{P,i,normj}^{(E)}$ 、正規化後音框能量平均值 $\overline{En}_{i,normj}^{(E)}$ 及其標準差 $\sigma_{En,i,normj}^{(E)}$ 、正規化後第一至第三共振峰之平均值 $\overline{F1}_{i,normj}^{(E)} \sim \overline{F3}_{i,normj}^{(E)}$ 與標準差 $\sigma_{F1,i,normj}^{(E)} \sim \sigma_{F3,i,normj}^{(E)}$ ，至於梅爾頻率倒頻譜係數(MFCC)在此採用原始 MFCC 前十三個係數 $\overline{MFCC}_{C1,i,j}^{(E)}$ ， $\overline{MFCC}_{C2,i,j}^{(E)}$ ，……， $\overline{MFCC}_{C13,i,j}^{(E)}$ 平均值，因此，與原始特徵參數個數相同，正規化後之特徵參數共有 23 個。

圖 3-7 為使用正規化特徵參數時，語音情緒辨識的主要流程圖，由於計算各個正規化特徵參數值均需要正規化特徵參數因子(Normalized factor)，因此，在語音情緒辨識的流程上，每一個語者均需要事先以中性的語調講一段話作為正規化特徵參數因子的計算依據(圖 3-7 下方虛線)，當正規化特徵參數因子計算完後，即可回到原始語音情緒辨識的主要流程(圖 3-7 上半部流程)，並進行使用者的語音情緒辨識。

3.3 小結

本章首先介紹論文中所使用的語料庫，內容包含兩個不同語系的語料，一個為 W. F. Sendlmeier 等人於 Technical University of Berlin 所蒐集的情緒語料庫，內容包含十名演員(五名男士及五名女士)，分別就十個單字與五個句子，以不同情緒發音，語料庫中包含悲傷 62

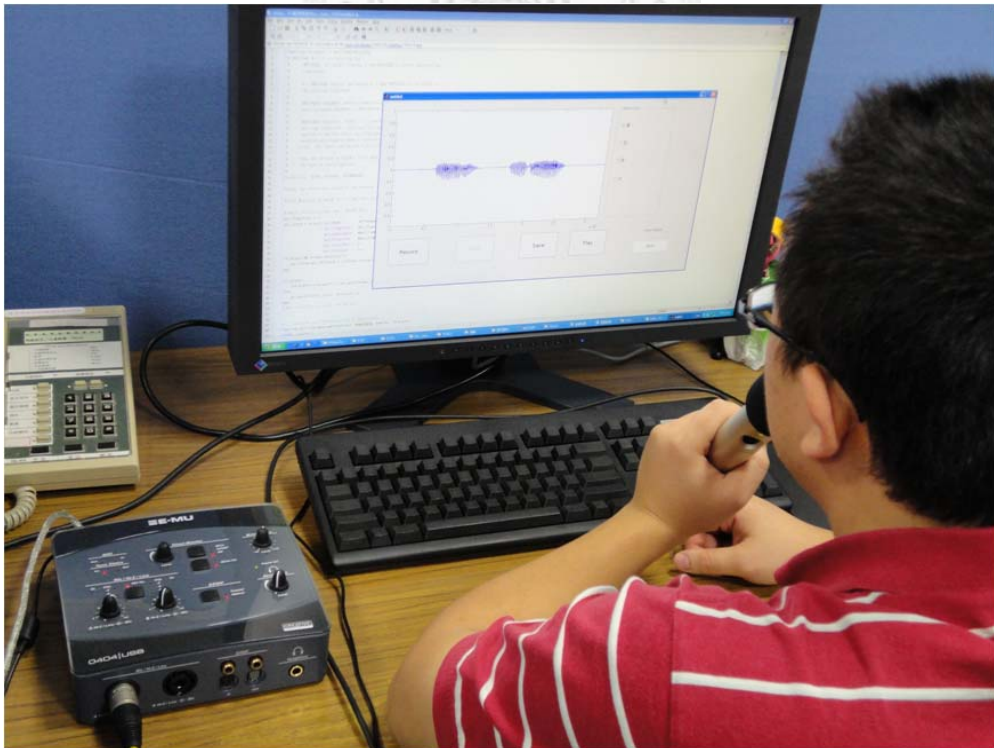
筆、中性 79 筆、快樂 71 筆及生氣 127 筆，共 339 筆語料，主要當作訓練與辨識用，另外，本文亦使用工業技術研究院(Industrial Technology Research Institute, ITRI)所自行錄製之情緒語料庫，內容包含女性十名，男性十名，每名語者四種情緒各 30 筆語料。

由於不同語者之間在語音特徵上的差異，造成資料庫建置的困難性，為了減少不同語者之間的差異，增加資料庫建置的可信度，使後續之分類器在處理上更加精準，本章介紹一特徵參數正規化之計算方法，透過此計算方法，可以縮小不同語者之間在情緒特徵上的差異，以利後續之分類辨識使用。





(a)



(b)

圖 3-1 工研院情緒語音錄製環境

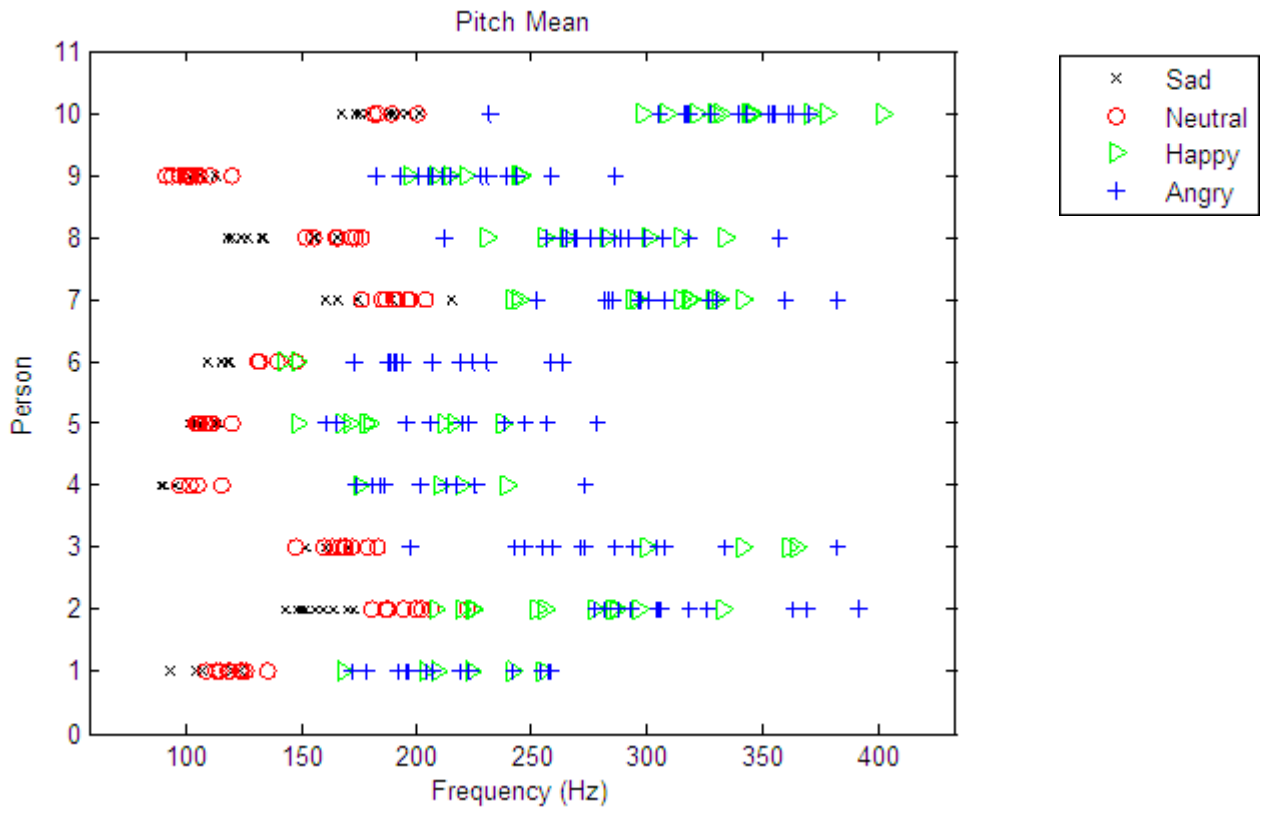
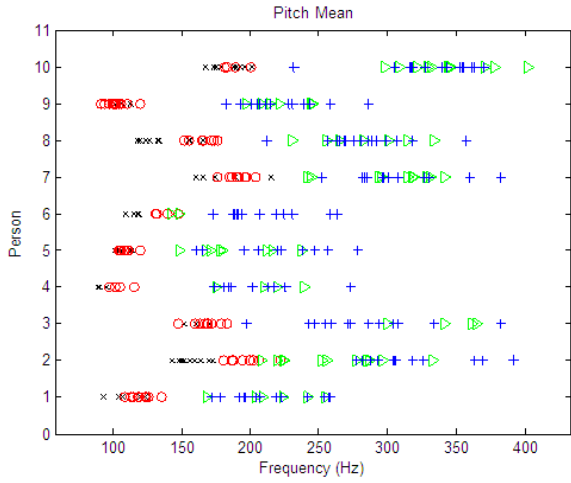
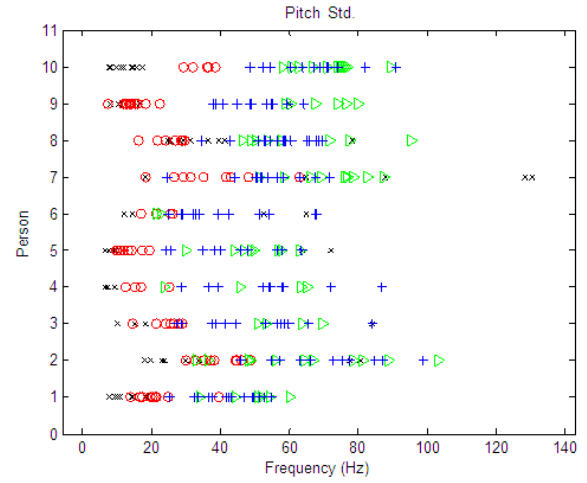


圖 3-2 不同語者在不同情緒之音高平均值分佈

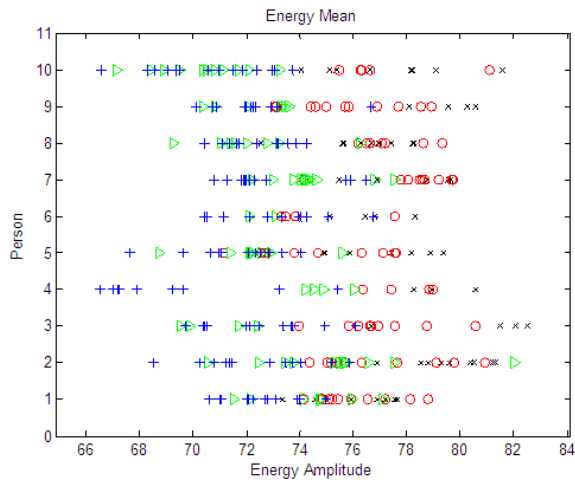




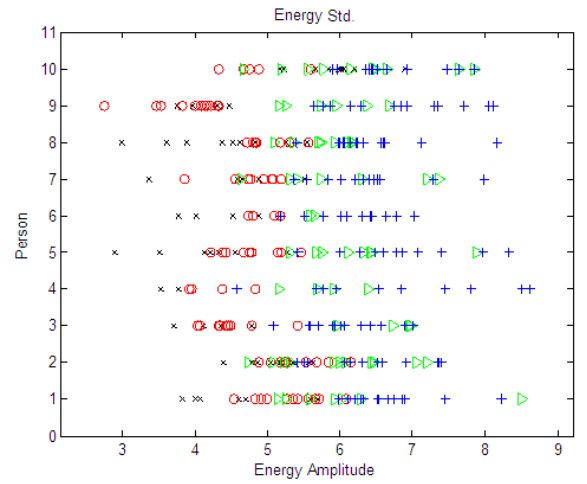
(a)



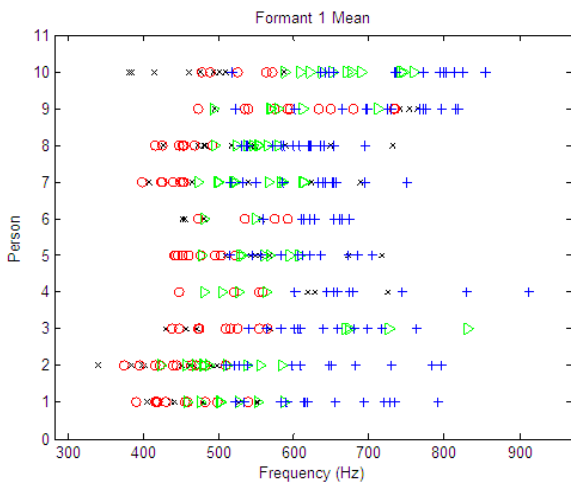
(b)



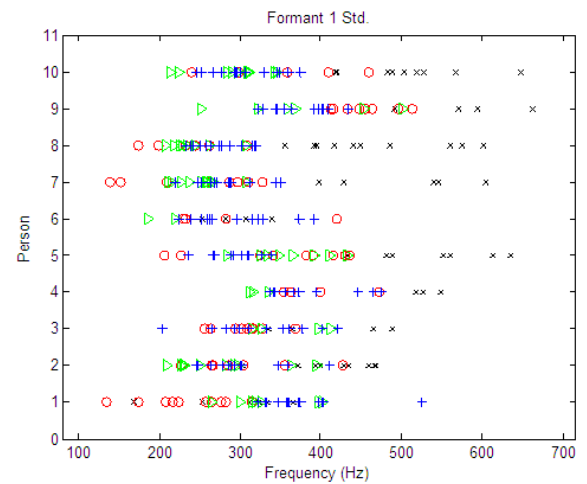
(c)



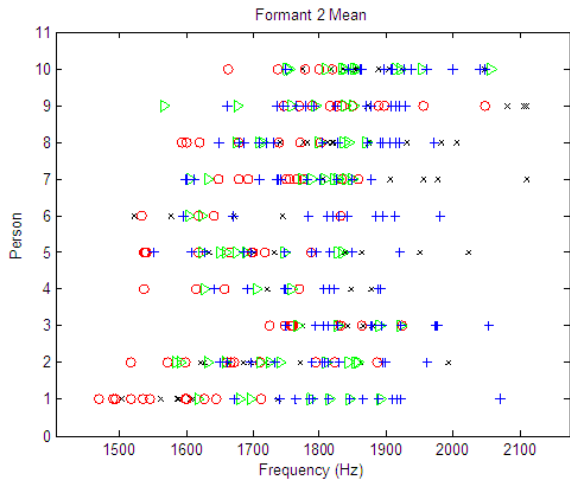
(d)



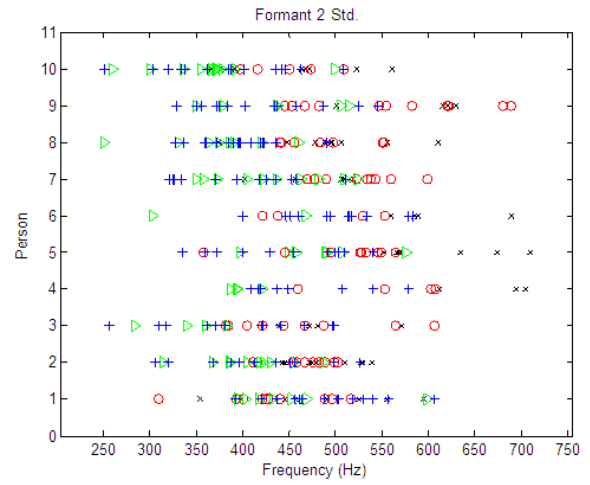
(e)



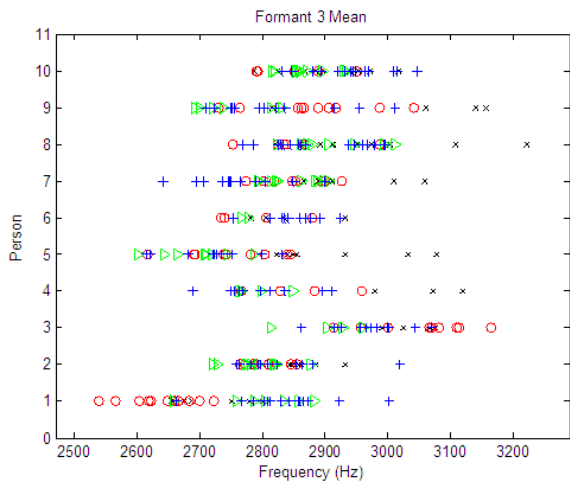
(f)



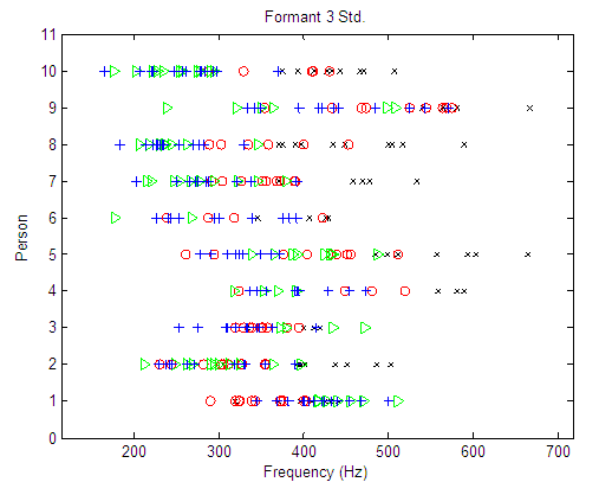
(g)



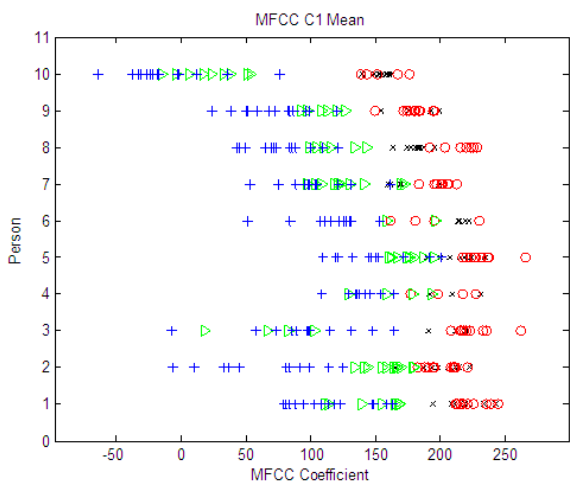
(h)



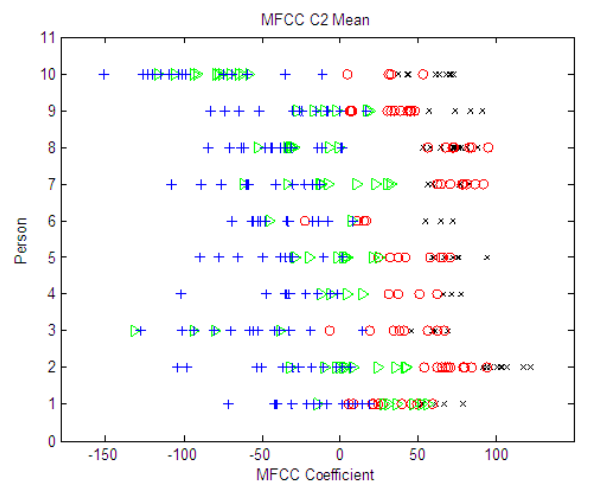
(i)



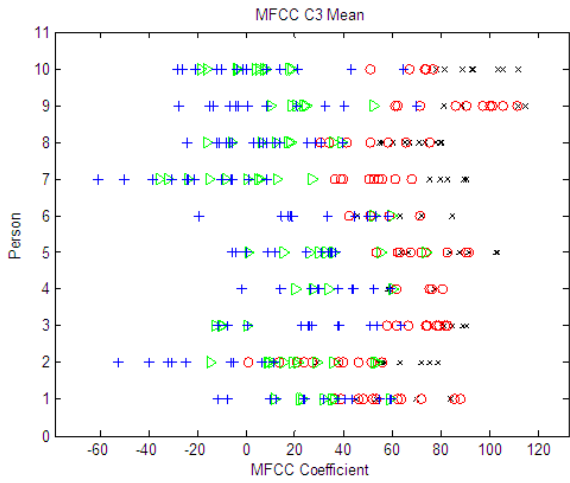
(j)



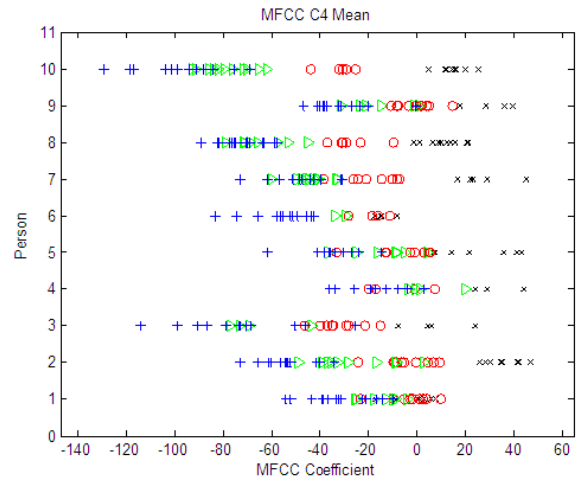
(k)



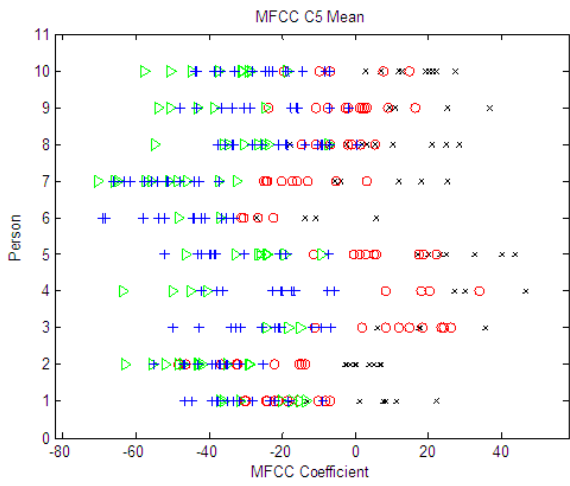
(l)



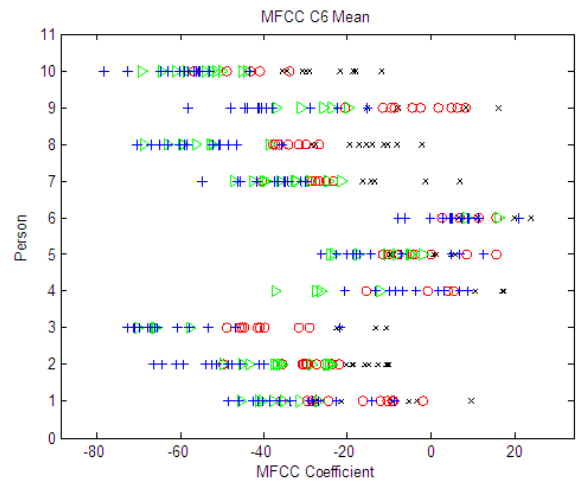
(m)



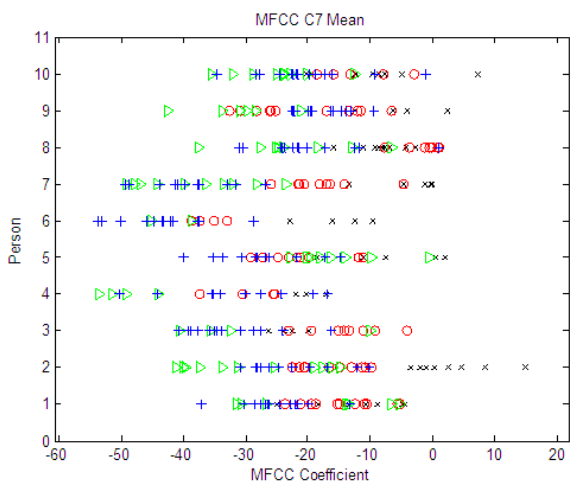
(n)



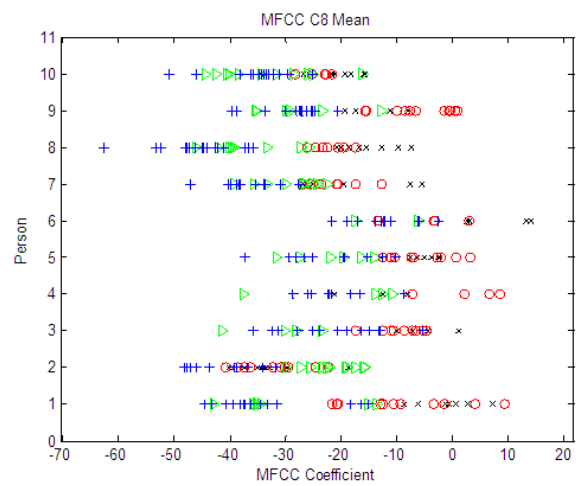
(o)



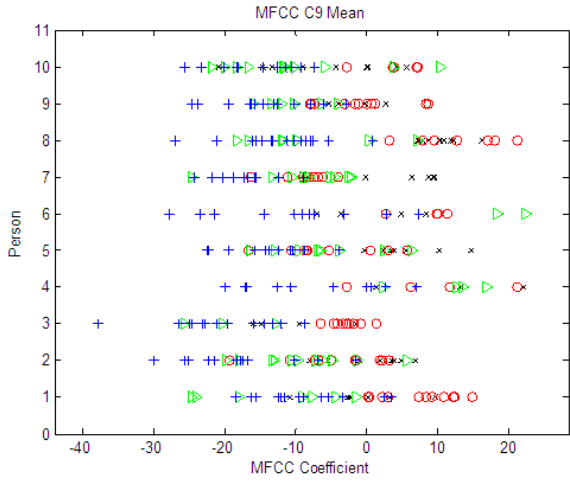
(p)



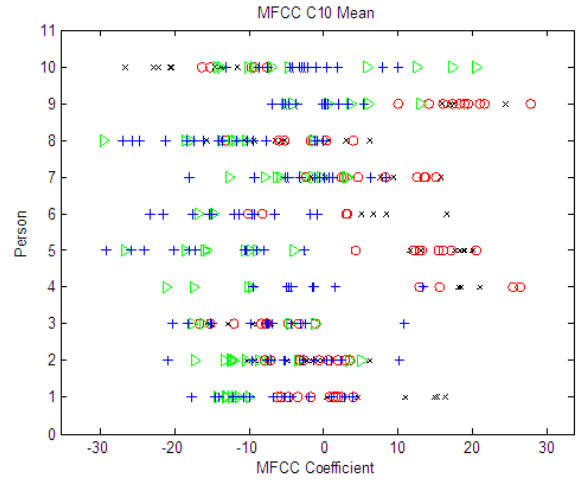
(q)



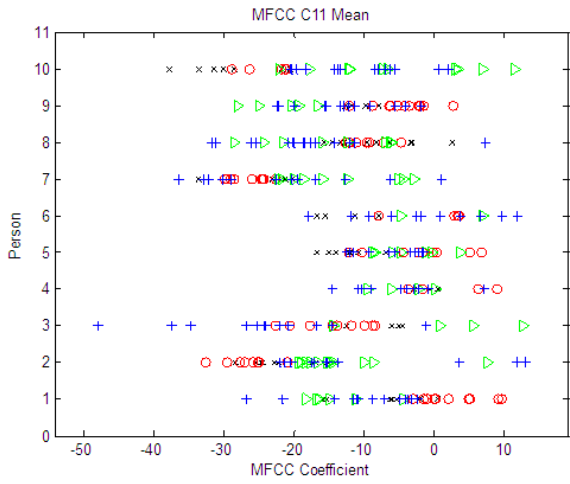
(r)



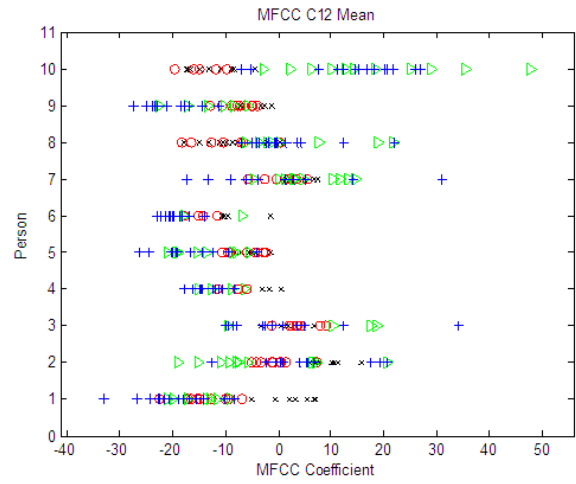
(s)



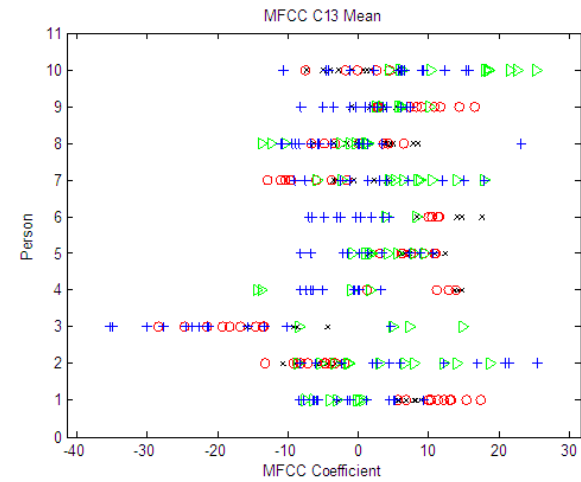
(t)



(u)



(v)



(w)

圖 3-3 不同語者在不同情緒之原始特徵參數分佈

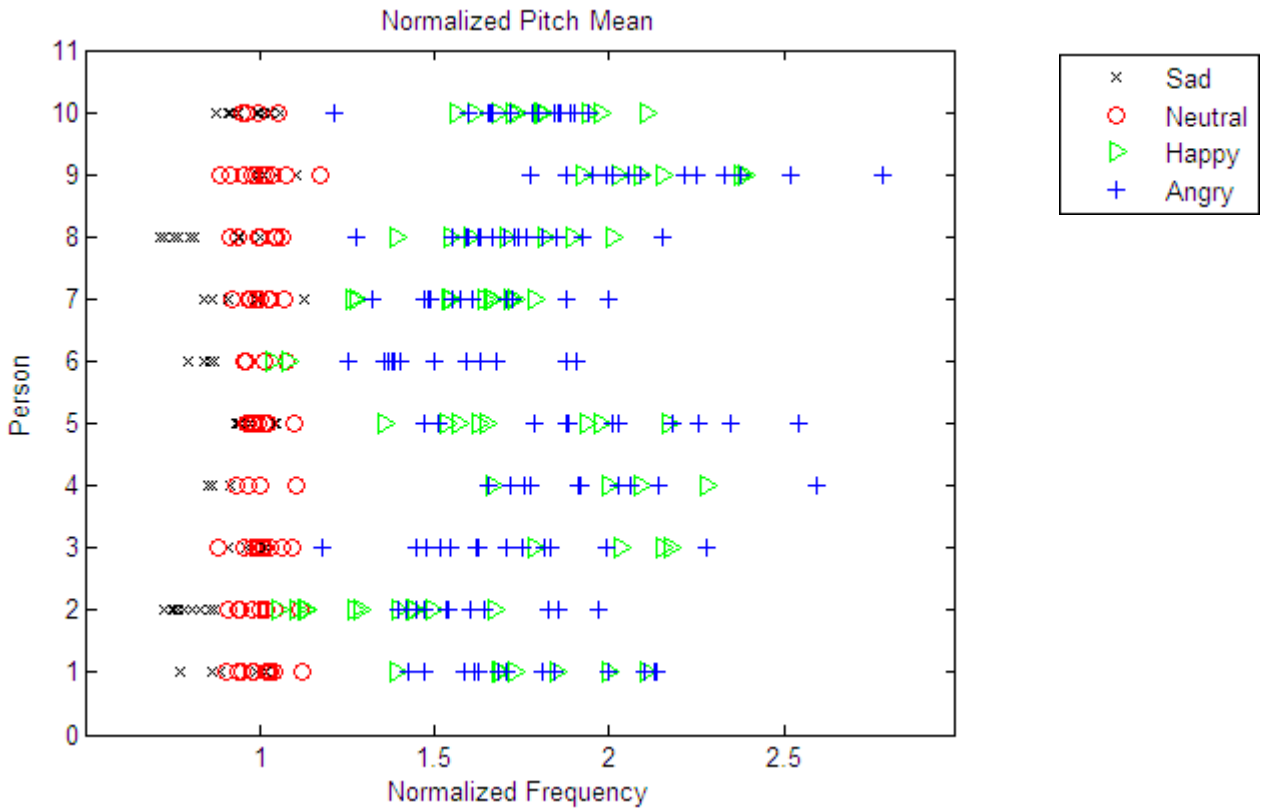


圖 3-4 正規化後音高平均值在不同語者不同情緒下之分佈

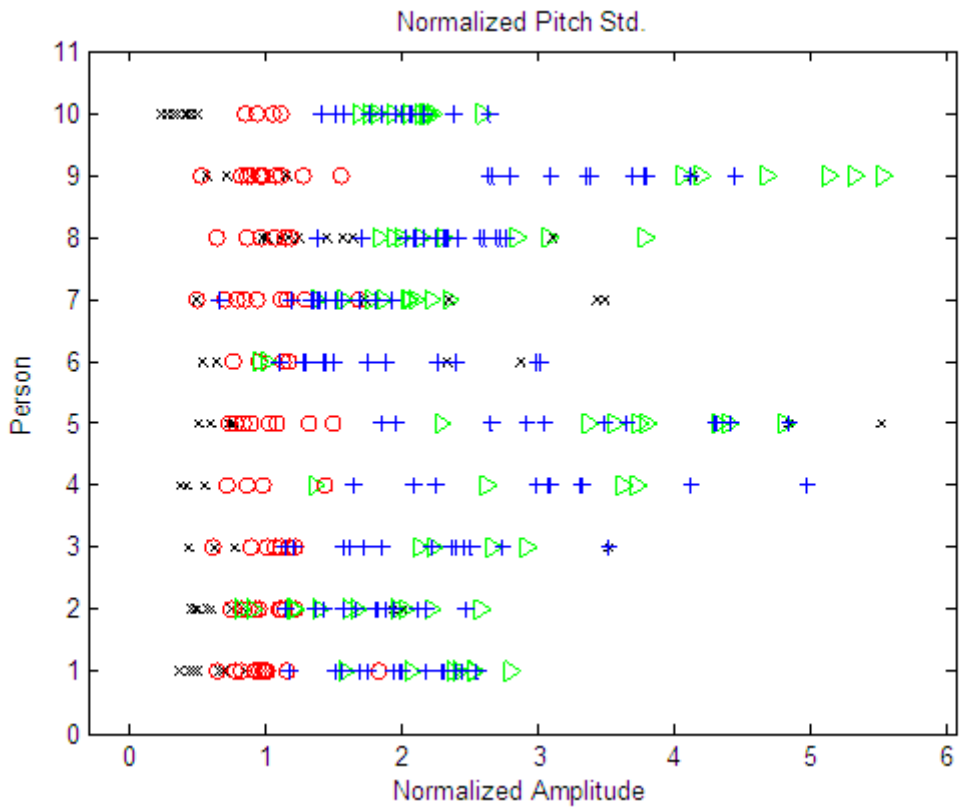
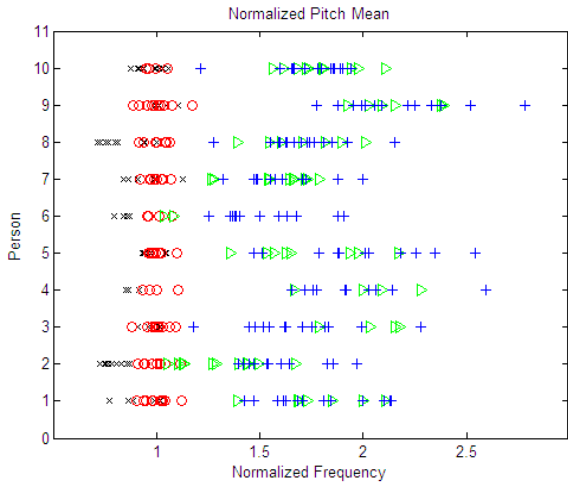
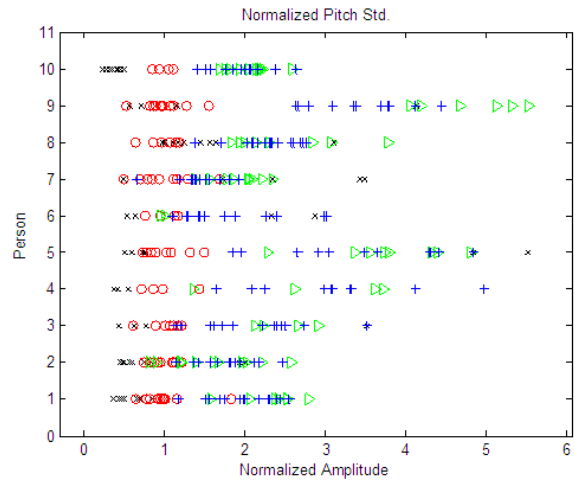


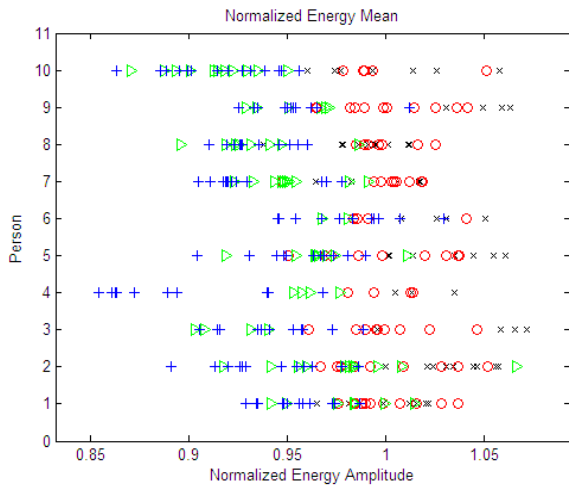
圖 3-5 正規化後音高標準差在不同語者不同情緒下之分佈



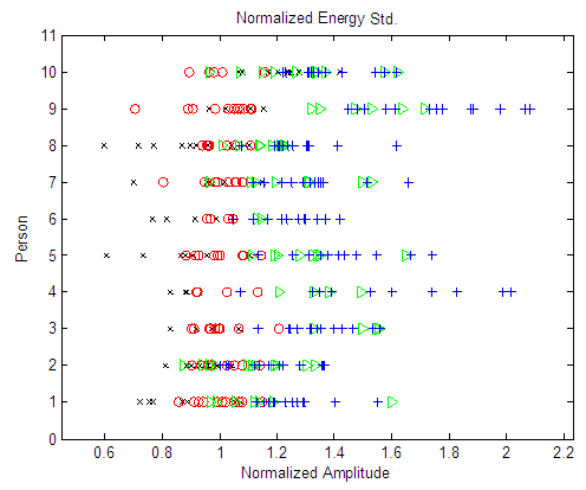
(a)



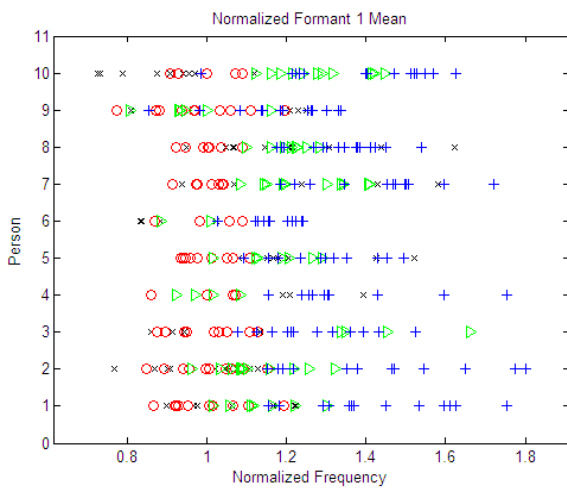
(b)



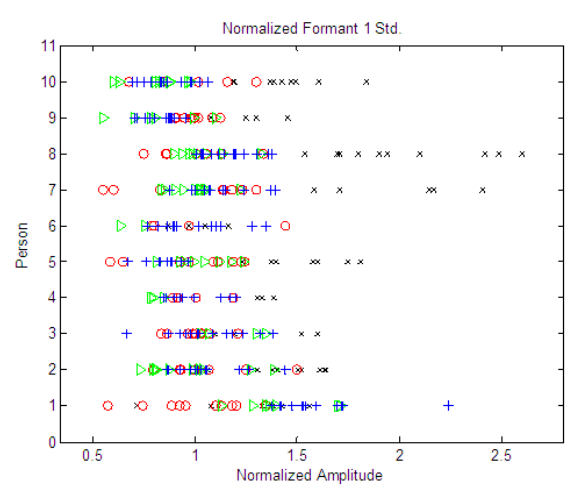
(c)



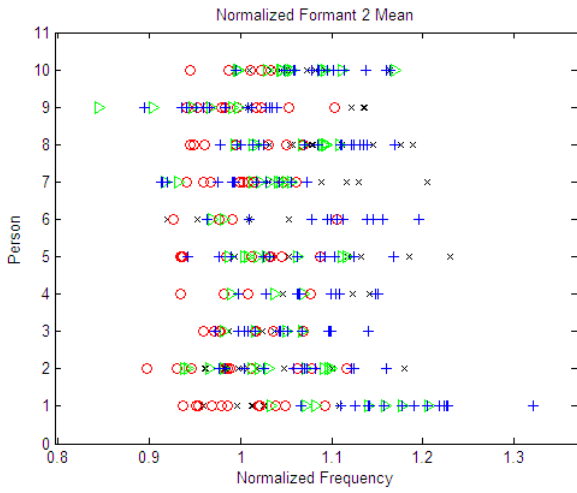
(d)



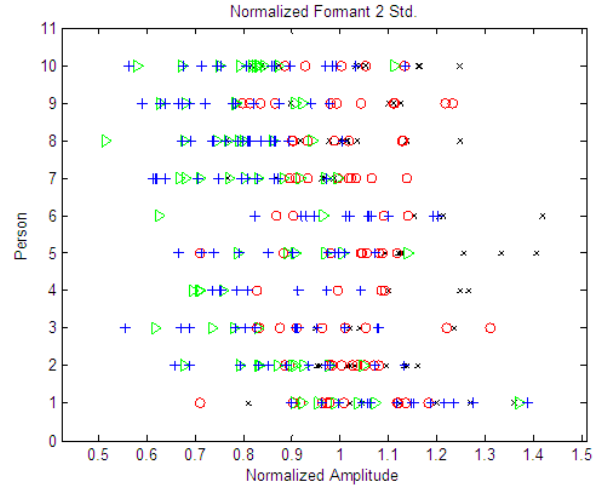
(e)



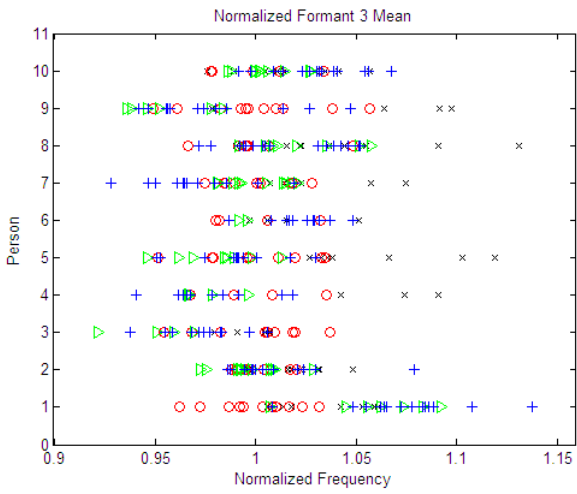
(f)



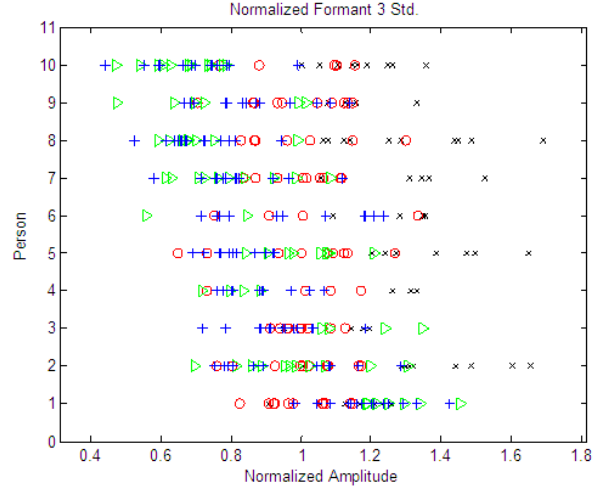
(g)



(h)



(i)



(j)

圖 3-6 正規化後各特徵參數在不同語者不同情緒下之分佈

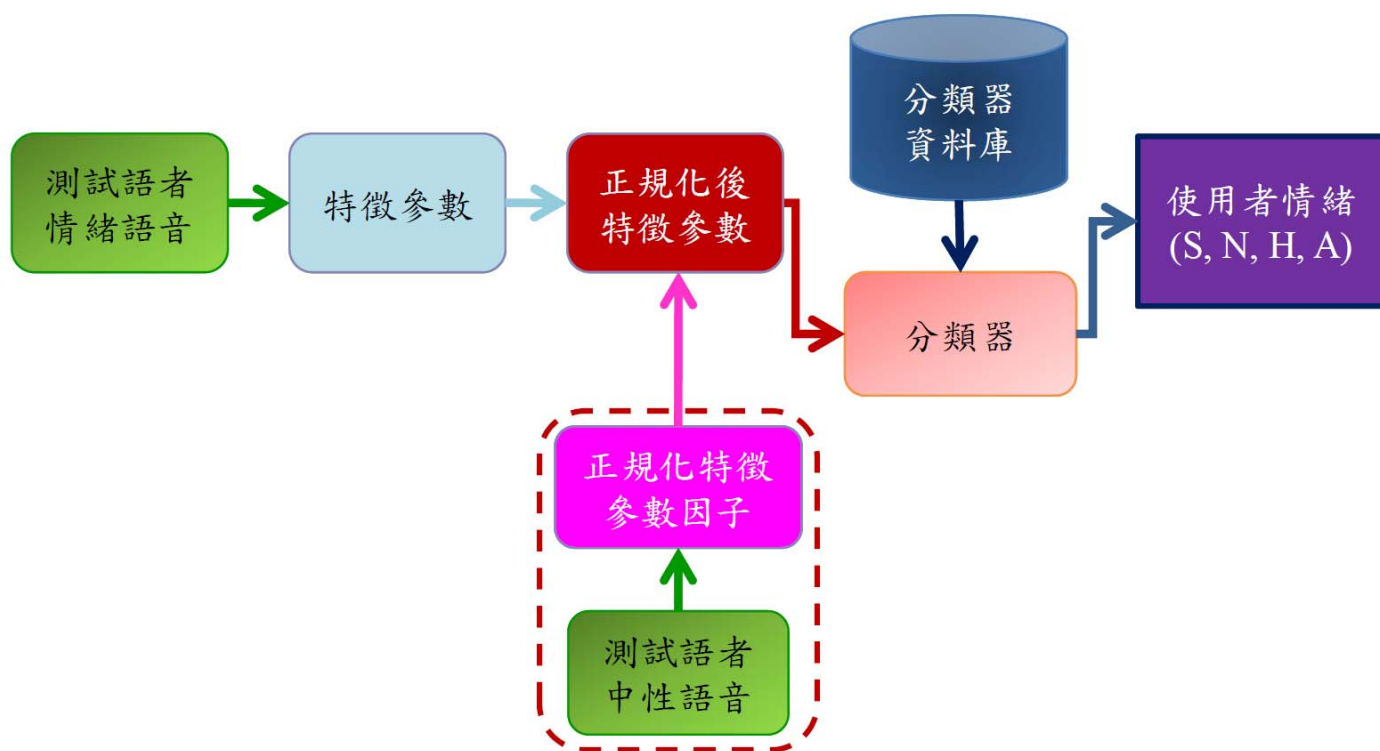


圖 3-7 使用正規化特徵參數之語音情緒辨識流程

表 3-1 德國情緒語料庫語料數量

錄音者	語料數量			
	悲傷	中性	快樂	生氣
P1(男性)	7	11	7	14
P2(女性)	9	10	11	12
P3(女性)	4	9	4	13
P4(男性)	3	4	4	10
P5(男性)	7	9	8	11
P6(男性)	4	4	2	12
P7(女性)	5	9	10	12
P8(女性)	10	7	8	16
P9(男性)	4	11	6	13
P10(女性)	9	5	11	14
總語料數量	62	79	71	127



表 3-2 工業技術研究院語料庫語料數量

錄音者	語料數量			
	悲傷	中性	快樂	生氣
P1(女性)	30	30	30	30
P2(女性)	30	30	30	30
P3(女性)	30	30	30	30
P4(女性)	30	30	30	30
P5(女性)	30	30	30	30
P6(女性)	30	30	30	30
P7(女性)	30	30	30	30
P8(女性)	30	30	30	30
P9(女性)	30	30	30	30
P10(女性)	30	30	30	30
P11(男性)	30	30	30	30
P12(男性)	30	30	30	30
P13(男性)	30	30	30	30
P14(男性)	30	30	30	30
P15(男性)	30	30	30	30
P16(男性)	30	30	30	30
P17(男性)	30	30	30	30
P18(男性)	30	30	30	30
P19(男性)	30	30	30	30
P20(男性)	30	30	30	30
總語料數量	600	600	600	600

表 3-3 工業技術研究院語料庫語料詞句

工業技術研究院語料庫語料詞句	
<p>悲傷：</p> <ol style="list-style-type: none"> 1. 我的狗不見了 2. 今天一直被罵，心情不是很好 3. 你這樣做讓我好難過 4. 什麼事情讓你看的比我還重要 5. 好衰喔，我的錢包被偷走了 6. 對你，我已不想再理會 7. 我已經不想再聽你解釋了 8. 你讓我對你感到好失望 9. 我不再感受到你對我的愛 10. 不用跟我多說什麼，你覺得好就好 11. 原來你的朋友比我還重要 12. 原來認不認真已不是那麼重要 13. 我想出事了 你才會想到我 14. 也許分手才是一種好的結果 15. 感冒了感覺好不舒服喔 16. 再多的努力，也沒人會在意 17. 我想這樣子就夠了 18. 我要怎麼做你才會滿意 19. 這生活方式讓我覺得好累 20. 你值得讓我付出嗎 21. 你要離開就離開吧 22. 也許放棄才是一件明智的選擇 23. 用這樣子的手段真令人心寒 24. 也許我們本來就不該這麼的堅持 25. 強求的愛情不會有好結果的 26. 難到我做的還不過多嗎 27. 你逼我只會讓我更痛苦 28. 我做這麼多，你還不滿意嗎 29. 跟你在一起我好痛苦 30. 這樣的結果真令人沮喪 	<p>中性：</p> <ol style="list-style-type: none"> 1. 沒有壓力的生活真好 2. 今天的天氣真好 3. 明天放假我要出去玩 4. 真是令人開心的一天 5. 好久沒吃到這麼特別的餅乾了 6. 今天的工作進行的真順利 7. 謝謝你送我禮物 8. 下了好多的雨終於停了 9. 我刮中一千元的彩券 10. 這次老師出的作業真簡單 11. 你這底有沒有認真在看 12. 你看你做了什麼好事 13. 東西到底是誰拿走的 14. 你為什麼把我的檔案刪掉了 15. 這花瓶到底是誰打破的 16. 你為什麼讓我等這麼久 17. 這單子是誰亂劃的 18. 你這樣做我會放心嗎 19. 不要老是造成別人的負擔 20. 今天的工作怎麼沒有做完 21. 你這麼做讓我很難過 22. 我已不想再聽你解釋 23. 我想這樣子就夠了 24. 你要離開就離開吧 25. 原來你的朋友比我還重要 26. 也許我們本來就不該這麼的堅持 27. 強求的愛情不會有好結果 28. 你讓我感到失望 29. 這樣的結果真讓人沮喪 30. 也許放棄才是一件明智抉擇

工業技術研究院語料庫語料詞句

快樂：

- 1.今天見到你好開心
- 2.新買的包包真漂靚
- 3.今天的工作真順利
- 4.沒有壓力的生活真好
- 5.你今天穿的好漂亮啊
- 6.我今天考試考一百分
- 7.這隻手機真好看
- 8.這糖果真好吃
- 9.我真是世界上最幸福的人
- 10.你長的好可愛喔。
- 11.今天的天氣真好
- 12.明天放假我要出去玩了
- 13.這碗麻醬麵好香喔
- 14.好久沒吃到這麼特別的餅乾
- 15.我終於畢業了
- 16.這衣服穿起來的質感很棒
- 17.看到成果，總算有代價了
- 18.很高興可以為你服務
- 19.今天的工作進行的真順利
- 20.這裡的風景好美喔
- 21.今天我中了一千元發票
- 22.這次老師的作業真簡單
- 23.我好喜歡你喔
- 24.真是令人開心的一天
- 25.這遊戲真刺激
- 26.這支股票讓我賺了不少呢
- 27.這隻狗好可愛阿
- 28.謝謝你送我的禮物
- 29.我刮中一千元的彩卷喔
- 30.下了好多天的雨，終於停了

生氣：

- 1.你到底有沒有認真看
- 2.今天的工作怎麼沒有做完
- 3.字怎麼寫的亂七八糟
- 4.整個行程怎麼亂成一團
- 5.看你做了什麼好事
- 6.我討厭你
- 7.東西到底是誰拿走的
- 8.為什麼你老是出狀況
- 9.這種結果叫做好嗎
- 10.你這樣做叫我怎麼能安心
- 11.老是依賴別人，這樣對嗎?
- 12.你以為我喜歡這樣對你嗎?
- 13.不要老是造成別人的負擔
- 14.你看起來真令人討厭
- 15.你是沒事做了嗎?
- 16.你真的好機車耶
- 17.你為什麼把我的檔案刪掉
- 18.你小心一點好不好，不要弄壞我的東西了
- 19.不要老是叫人家等你好嗎
- 20.你以為一句對不起就可以算了嗎
- 21.你這樣做真的是為我好嗎?
- 22.這花瓶到底是誰打破的
- 23.你這樣做我會放心嗎?
- 24.你不要太超過了喔
- 25.你好吵喔，可不可以小聲一點。
- 26.你為什麼讓我等這麼久
- 27.你以為等你是應該的嗎
- 28.這單子是誰亂畫的阿
- 29.你到底要我說幾次你才明白
- 30.不要每次都想靠別人好不好

表 3-4 使用之語音特徵參數

1	音高平均值(Pitch Mean)	13	梅爾頻率倒頻譜係數 C3 平均值(MFCC C3 Mean)
2	音高之標準差(Pitch Std.)	14	梅爾頻率倒頻譜係數 C4 平均值(MFCC C4 Mean)
3	能量平均值(Energy Mean)	15	梅爾頻率倒頻譜係數 C5 平均值(MFCC C5 Mean)
4	能量之標準差(Energy Std.)	16	梅爾頻率倒頻譜係數 C6 平均值(MFCC C6 Mean)
5	第一共振峰平均值(Formant 1 Mean)	17	梅爾頻率倒頻譜係數 C7 平均值(MFCC C7 Mean)
6	第一共振峰之標準差 (Formant 1 Std.)	18	梅爾頻率倒頻譜係數 C8 平均值(MFCC C8 Mean)
7	第二共振峰平均值(Formant 2 Mean)	19	梅爾頻率倒頻譜係數 C9 平均值(MFCC C9 Mean)
8	第二共振峰之標準差 (Formant 2 Std.)	20	梅爾頻率倒頻譜係數 C10 平均值(MFCC C10 Mean)
9	第三共振峰平均值(Formant 3 Mean)	21	梅爾頻率倒頻譜係數 C11 平均值(MFCC C11 Mean)
10	第三共振峰之標準差 (Formant 3 Std.)	22	梅爾頻率倒頻譜係數 C12 平均值(MFCC C12 Mean)
11	梅爾頻率倒頻譜係數 C1 平均值(MFCC C1 Mean)	23	梅爾頻率倒頻譜係數 C13 平均值(MFCC C13 Mean)
12	梅爾頻率倒頻譜係數 C2 平均值(MFCC C2 Mean)		

表 3-5 使用之正規化後語音特徵參數

1	正規化後音高平均值 (Normalized Pitch Mean)	13	梅爾頻率倒頻譜係數 C3 平均值(MFCC C3 Mean)
2	正規化後音高之標準差 (Normalized Pitch Std.)	14	梅爾頻率倒頻譜係數 C4 平均值(MFCC C4 Mean)
3	正規化後能量平均值 (Normalized Energy Mean)	15	梅爾頻率倒頻譜係數 C5 平均值(MFCC C5 Mean)
4	正規化後能量之標準差 (Normalized Energy Std.)	16	梅爾頻率倒頻譜係數 C6 平均值(MFCC C6 Mean)
5	正規化後第一共振峰平均值 (Normalized Formant 1 Mean)	17	梅爾頻率倒頻譜係數 C7 平均值(MFCC C7 Mean)
6	正規化後第一共振峰之標準差 (Normalized Formant 1 Std.)	18	梅爾頻率倒頻譜係數 C8 平均值(MFCC C8 Mean)
7	正規化後第二共振峰平均值 (Normalized Formant 2 Mean)	19	梅爾頻率倒頻譜係數 C9 平均值(MFCC C9 Mean)
8	正規化後第二共振峰之標準差 (Normalized Formant 2 Std.)	20	梅爾頻率倒頻譜係數 C10 平均值(MFCC C10 Mean)
9	正規化後第三共振峰平均值 (Normalized Formant 3 Mean)	21	梅爾頻率倒頻譜係數 C11 平均值(MFCC C11 Mean)
10	正規化後第三共振峰之標準差 (Normalized Formant 3 Std.)	22	梅爾頻率倒頻譜係數 C12 平均值(MFCC C12 Mean)
11	梅爾頻率倒頻譜係數 C1 平均值(MFCC C1 Mean)	23	梅爾頻率倒頻譜係數 C13 平均值(MFCC C13 Mean)
12	梅爾頻率倒頻譜係數 C2 平均值(MFCC C2 Mean)		

第四章 KNN 與 SVM 於語音情緒辨識之實驗與分析

4.1 KNN 語音情緒辨識

4.1.1 第 K 個最近鄰(K-Nearest Neighbor, KNN)

KNN 是一種監督式的(Supervised)學習法[61]。首先，替每一筆資料計算出其對應位置，以二維特徵參數為例(如圖 4-1 所示)，其目的是要讓機器學會分出四種顏色的點(圖 4-1 中之綠色、藍色、橘色、粉紅)。每個點在平面上的位置，分別用(x,y)來代表，機器從建立的資料中已事先知道哪些點是綠色、藍色、橘色、粉紅。當一筆新的資料輸入機器以後(圖 4-1 中間號沒有顏色的點)，KNN 演算法就先計算這個點和其他已經知道四種顏色點資料之間的距離，如式(4.1.1)所示。

$$dist(X,Y) = \sqrt{(X - x_1)^2 + (Y - y_1)^2} \quad (4.1.1)$$

若 KNN 的 K 如果設定成 1，也就是(1-NN)的話，代表機器會找距離最近的點，然後看這個點是四種顏色的哪一種，並將此未知的資料歸類為此種顏色，如果這個點是粉紅，那麼機器就預測剛才打問號的這個點(新的資料點)是粉紅點，之後這個問號點就被歸類為粉紅色，然後繼續反覆同樣的動作找下一個新輸入的資料點，預測新的問號點是綠色、藍色、橘色或是粉紅。

如果 $K=3$ ，也就是使用 3-NN 學習，機器同樣先計算圖中打問號的點和各個顏色的點的距離，接著選出前三名距離最近的點，然後看看裡面那種顏色的點比較多，如果粉紅色點比較多，就判斷這個新問號點，是粉紅點。同理，當 K 值不同，如 $K=5$ 、 $K=7$ 或 $K=11$ 時，則會有 5-NN、7-NN 或 11-NN 學習法，端看問題的需要而定。

一般而言，在使用 KNN 分類法時，會先將資料中每一個特徵參數先經過正規化處理，使得資料中之所有特徵參數值皆落在 $-1 \sim 1$ 之間，以便於後續之計算處理，以二維特徵參數為例，KNN 之參數正規化計算如式(4.1.2) 及(4.1.3)所示

$$x_{i,norm} = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (4.1.2)$$

$$y_{i,norm} = \frac{y_i - \min(y_i)}{\max(y_i) - \min(y_i)} \quad (4.1.3)$$

其中 $(x_{i,norm}, y_{i,norm})$ 為 KNN 之參數正規化值， $\min(x_i)$ 與 $\min(y_i)$ 為特徵參數 x 與 y 於資料庫中之最小值， $\max(x_i)$ 與 $\max(y_i)$ 為特徵參數 x 與 y 於資料庫中之最大值。

4.1.2 KNN 語音情緒辨識實驗結果與分析

本論文使用 KNN 作為語音情緒辨識的主要流程如圖 4-2 所示，首先，測試語者必須先講一段中性語調之語句，由此計算出正規化特徵參數因子(如流程中之虛線所示)，接著，進入語音情緒辨識流程，

在得到語者一段情緒語音訊號後，先計算出原始特徵參數，並經由正規化特徵參數因子對原始特徵參數做正規化計算，最後，將正規化特徵參數與 KNN 正規化特徵資料庫作比對與分類，辨識出使用者的情緒。

由於 KNN 分類器的分類效果取決於最近鄰是屬於哪一個分類的多寡來決定，因此，K 的選定會對分類結果有直接的影響，在此，我們利用德國語料庫所有語料隨機取各情緒一半當作訓練資料，另一半則當作測試資料，圖 4-3 為不同 K 值所得之辨識結果，由辨識結果顯示，當 K 大於 20 時，辨識率開始降低，而 K 值若取太小時，由於參考樣本不足，也容易造成誤判的情形發生，因此，本文中所使用之 K 值在參考其他文獻與測試結果後，將固定 K 值為 11，以避免過大或過小產生辨識效果不佳或誤判的情況。以下將分別針對 KNN 分類器在語者相關(User dependent)以及語者獨立(User independent)的情況進行實驗測試與分析，同時，也將針對 KNN 分類器在訓練語料與測試語料不同的情況進行實驗測試與分析。

首先，先針對訓練資料庫與測試資料庫為語者相關做實驗測試，以德國語料所有資料當作訓練資料，也就是悲傷有 62 筆、中性有 79 筆、快樂有 71 筆、生氣有 127 筆訓練資料，並以訓練資料當作測試資料，對德國語料作 Inside Test。表 4-1 為使用正規化特徵參數 KNN

Inside Test 鑑別矩陣(Discriminant matrix)，由辨識結果顯示，使用正規化特徵參數做 KNN Inside Test 時，悲傷 62 筆測試資料有 52 筆資料被辨識為悲傷，10 筆資料被辨識為中性，悲傷之辨識正確率為 84%，中性 79 筆測試資料有 79 筆資料被辨識為中性，中性之辨識正確率 100%，快樂 71 筆測試資料有 2 筆資料被辨識為中性，45 筆資料被辨識為快樂，24 筆資料被辨識為生氣，快樂之辨識正確率 63%，生氣 127 筆測試資料有 1 筆資料被辨識為中性，1 筆資料被辨識為快樂，125 筆資料被辨識為生氣，生氣之辨識正確率 98%，整體之辨識正確率為 88.8%，從鑑別矩陣顯示，悲傷與中性在識別上，比較容易產生混淆，而快樂則與生氣易產生混淆，造成悲傷與快樂之辨識度較低。

接著，考慮語者獨立的情況，也就是使用者資料不在訓練資料內的情況，同樣的，以德國語料庫當作實驗語料，首先，將 Person 1(P1) 的語料當作測試語料，並以 Person 2 ~ Person 10 語料當作訓練語料，因此悲傷有 7 筆、中性有 11 筆、快樂有 7 筆、生氣有 14 筆測試語料，而在訓練語料中，悲傷有 55 筆、中性有 68 筆、快樂有 64 筆、生氣有 113 筆，接著，以 Person 2 當測試資料，另外 9 人(P1 以及 P3 ~ P10) 當作訓練資料，進行 P2 KNN Outside Test，以此類推，對 P3 至 P10 進行 KNN Outside Test，表 4-2 為 P1 ~ P10 使用原始特徵參數 KNN

Outside Test 鑑別矩陣(P1 ~ P10 各情緒之辨識正確率詳見附錄 A)，由鑑別矩陣結果顯示，悲傷 62 筆測試資料中有 35 筆資料被辨識為悲傷，27 筆資料被辨識為中性，悲傷之辨識正確率為 56%，中性 79 筆測試資料有 2 筆資料被辨識為悲傷，77 筆資料被辨識為中性，中性之辨識正確率 97%，快樂 71 筆測試資料中有 3 筆資料被辨識為中性，32 筆被辨識為快樂，36 筆資料被辨識為生氣，快樂之辨識正確率 45%，生氣 127 筆測試資料中有 2 筆資料被辨識為中性，4 筆被辨識為快樂，121 筆資料被辨識為生氣，生氣之辨識正確率 95.3%，整體之辨識正確率為 78.2%。當測試資料不在訓練資料中時，明顯的，悲傷與快樂的辨識率大幅度下降，悲傷有一大部分資料容易被辨識成中性，而快樂則大部分容易被辨識成生氣，KNN 分類方法對於悲傷與快樂並沒有良好的分類效果。

同樣的，使用正規化特徵參數進行 KNN Outside Test，表 4-3 為 P1 ~ P10 使用正規化特徵參數 KNN Outside Test 鑑別矩陣(P1 ~ P10 各情緒之辨識正確率詳見附錄 A)，由鑑別矩陣結果顯示，悲傷 62 筆測試資料中有 41 筆資料被辨識為悲傷，21 筆資料被辨識為中性，悲傷之辨識正確率為 66%，中性 79 筆測試資料中有 79 筆資料被辨識為中性，中性之辨識正確率 100%，快樂 71 筆測試資料中有 5 筆資料被辨識為中性，39 筆被辨識為快樂，27 筆資料被辨識為生氣，快樂之

辨識正確率 55%，生氣 127 筆測試資料中有 1 筆資料被辨識為悲傷，1 筆資料被辨識為中性，4 筆被辨識為快樂，121 筆資料被辨識為生氣，生氣之辨識正確率 95.3%，整體之辨識正確率為 82.6%，整體之辨識正確率比使用原始特徵參數做 KNN Outside Test 時提升約 4.4%，由原始與正規化特徵參數之 KNN Outside Test 辨識結果可以看出，使用正規化特徵參數可以有效縮小語者之間在特徵參數上的差異，增加各情緒之間的分類效果。若從正規化特徵參數之 KNN Outside Test 鑑別矩陣來看，其結果與原始參數結果相近，悲傷容易被辨識成中性，而快樂則容易被辨識成生氣，此結果顯示 KNN 分類方法不論使用原始或正規化特徵參數皆對悲傷及快樂沒有良好的分類效果。

最後，由於不同地區、不同種族對於語音的發音以及表達情緒的方式皆不同，當以特定語系訓練好之分類器應用在不同語系或不同國家時，由於各情緒在特徵參數之間的差異，辨識效果將大為不同，因此，我們同樣針對 KNN 分類器在訓練語料與測試語料為不同語系情況下進行辨識，在此，我們以德國語料庫所有資料當作訓練資料，並以工業技術研究院所錄製之語料(表 3-2)當作測試語料，表 4-4 為 KNN 使用正規化特徵參數在訓練與測試為不同語系下之鑑別矩陣，由辨識結果顯示，KNN 對於中性有極佳之辨識效果，辨識率為 95.8%，而悲傷情緒只有 16.2% 被辨識為悲傷，有 83.3% 的悲傷情緒會被辨識為

中性，此結果表示中文語系悲傷之特徵參數與德國語料庫之中性特徵參數相近，因此，若以德國語料庫當作訓練資料時，中文語系之悲傷情緒容易被辨識為中性，而快樂與生氣情緒辨識結果與悲傷相當類似，快樂與生氣情緒各有 65.8% 以及 90.2% 會被辨識為中性，快樂被正確辨識為快樂的機率有 22.5%，而生氣僅有 2.3% 的機率被正確辨識為生氣，整體之正確辨識率為 34.2%，由以上各情緒之辨識結果可以發現，工研院四種情緒皆容易被辨識為中性，此結果與台灣人個性較為內斂有關，在表達情緒時，各情緒之特徵不若德國演員各情緒之呈現，因此，在辨識上，容易被辨識為中性。

4.2 SVM 語音情緒辨識

4.2.1 支持向量機(Support Vector Machine, SVM)

支持向量機 (Support Vector Machine) 是一種分類(Classification) 演算法，由 Cortes and Vapnik 等根據統計學習理論提出的一種新的機器學習方法[13]。

簡單來說，SVM 想要解決以下的問題：找出一個超平面(Hyperplane)，使之將兩個不同的集合分開，而超平面意指在高維度中的平面。以二維的例子來說，如圖 4-4 示，我們希望能找出一條線能夠將紅點和綠點分開，而且我們還希望這條線距離這兩個集合的邊界(Margin)越大越好，這樣我們才能夠很明確的分辨這個點是屬於那

個集合，否則在計算上容易因精度的問題而產生誤差。

假設有一堆點集合 $\{x_i, y_i\}$ ， $i = 1, 2, \dots, n$ ， $x_i \in R^d$ ， $y_i \in \{+1, -1\}$ ，我們希望能找到一條直線 $f(x) = w^T x - b$ ，使所有 $y_i = -1$ 的點落在 $f(x) < 0$ 這一邊，且使 $y_i = +1$ 的點落在 $f(x) > 0$ 這一邊，這樣我們就可以根據 $f(x)$ 的正負號來區分這個點是屬於這兩個集合之中的哪個集合，此超平面我們稱為 Separating hyperplane，而距離兩邊邊界最大的就稱為 Optimal Separating Hyperplane (OSH)，詳細之理論推導可以參考文獻上所列之 SVM 介紹[74]。

經整理後，SVM 相當於求解一最佳化問題

$$\partial_w L_p = 0 \quad \Rightarrow \quad w - \sum_{i=1}^N \alpha_i y_i x_i = 0 \quad (4.2.1)$$

$$\begin{aligned} \partial_b L_p = 0 \quad \Rightarrow \quad \sum_{i=1}^N \alpha_i y_i &= 0 \quad (4.2.2) \\ y_i (w^T x_i - b) - 1 &\geq 0 \end{aligned}$$

$$\text{Lagrange multiplier condition } \alpha_i > 0 \quad (4.2.3)$$

$$\text{Complementary slackness } \alpha_i [y_i (w^T x_i - b) - 1] = 0 \quad (4.2.4)$$

(4.2.1) 式 ~ (4.2.4) 式 這些條件又被稱為 KKT 條件 (Karush-Kuhn-Tucker conditions)，當我們把訓練資料(Training data)丟進來時，有些點會滿足上面的 KKT 條件，而這些點就是剛好落在 Support hyperplane 上面的點，這些點又被稱為 Support vector，並且它們的 α_i 會大於 0，在我們求得 Support vector 之後，我們就可以用

它們來判斷新加入的點是屬於那一邊的集合。

$$f(x) = w^{*T}x - b^* = \sum_i \alpha_i y_i x_i^T x - b^* \quad (4.2.5)$$

$$b^* = \left(\sum_j \alpha_j y_j x_j^T x_i - y_i \right) \quad i \text{ is a support vector} \quad (4.2.6)$$

基本上，SVM 是一個二元的分類器(Binary classifier)，只能處理兩種類別的分類，而在現實應用上，常常是需要多元的分類，因此，需要使用一些策略來使得，SVM 在多元的分類上也可以使用，SVM 在多元分類上的策略主要有兩種方式：

(1)、一對多(One-against-rest)：當有 k 個不同類別時，使用 k 個 SVM 分類器來將 k 個類別分開，第 m 個 SVM 分類器可以將第 m 個類別與其他類別分開，也就是使用分辨屬於特定類別的 SVM 分類器。

(2)、一對一(One-against-one)：針對任意兩個不同類別皆使用一個 SVM 分類器，因此，使用一對一策略時，共需要 $k(k-1)/2$ 個 SVM 分類器，最後，再藉由此 $k(k-1)/2$ 個 SVM 分類器將最可能發生的類別判斷出來。

本文所使用之 SVM 分類器採用臺灣大學資訊系林智仁老師所開發之 LIBSVM[8]為主要 SVM 分類辨識工具，論文所使用之 SVM Kernel function 為 Radial Basis Function (RBF)。

4.2.2 SVM 情緒辨識實驗結果與分析

本論文使用 SVM 作為語音情緒辨識的主要流程如圖 4-5 所示，與 KNN 辨識流程的不同之處，在於利用 SVM 做分類時，必須先由特徵參數資料庫訓練出一組 SVM Model，再由此 Model 與測試資料特徵作計算與分類，而 SVM 語音情緒辨識的其餘流程則與 KNN 流程相同，當使用原始特徵參數進行辨識時，直接計算出原始特徵參數，並與 SVM 原始特徵資料庫所訓練出來的 SVM Model 作比對與分類，辨識出使用者的情緒，而若使用正規化特徵參數做語音情緒辨識時，則流程與 KNN 相同，必須加入正規化特徵參數運算(如圖中虛線所示)，先由使用者的一段中性語調語句中計算出正規化特徵參數因子，接著，進入語音情緒辨識主要流程，在得到使用者一段情緒語音訊號後，先計算出原始特徵參數，並經由正規化特徵參數因子對原始特徵參數做正規化計算，最後，將正規化特徵參數與 SVM 正規化特徵資料庫所訓練出來的正規化 SVM Model 作比對與分類，辨識出使用者的情緒。

SVM 分類器是依據所給定的特徵參數資料庫，去找出一個最佳的多維度分類函數，進而將資料有效分類，在事前訓練上，必須先將資料庫透過訓練，產生此資料的最佳分類 Model，因此，在分類器使用上，KNN 分類器只需將特徵參數資料庫建置好即可，而 SVM 分類器，則需在特徵參數資料庫建置完後，再經由多維度分類函數的最佳

化，找出特徵參數資料庫的最佳分類模型，不同的特徵參數資料庫都需要重新再訓練出一組新的分類函數；以下將針對 SVM 分類器分別在語者相關以及語者獨立的情況下進行實驗測試、分析與討論。

首先，與 KNN 特徵參數資料庫相同，先針對 SVM 分類器在語者相關的情況下進行實驗測試，以德國語料所有資料當作訓練資料，悲傷有 62 筆、中性有 79 筆、快樂有 71 筆、生氣有 127 筆訓練資料，並以訓練資料當作測試資料，對資料作 Inside Test。表 4-5 為使用正規化特徵參數 SVM Inside Test 鑑別矩陣，由辨識結果顯示，所有測試資料皆完全辨識正確，整體之辨識正確率為 100%，顯示當測試資料在訓練資料中時，SVM 提供了極佳的分類效果。

接著，考慮語者獨立的情況，首先，將 Person 1(P1)的語料當作測試語料，並以 Person 2 ~ Person 10 語料當作訓練語料，產生一組訓練後 SVM Model，並對 P1 語料進行 Outside Test，接著，以 Person 2 當測試資料，另外 9 人(P1 以及 P3 ~ P10)當作訓練資料，再產生另一組訓練後 SVM Model，並對 P2 進行 Outside Test，以此類推，對 P3 至 P10 做 Outside Test，表 4-6 為 P1 ~ P10 使用原始特徵參數 SVM Outside Test 鑑別矩陣(P1 ~ P10 各情緒之辨識正確率詳見附錄 A)，由整體辨識結果顯示，悲傷 62 筆測試資料中有 44 筆資料辨識為悲傷，辨識正確率為 71%，中性 79 筆測試資料有 79 筆資料辨識為中性，辨

識正確率 100%，快樂 71 筆測試資料有 60 筆辨識為快樂，辨識正確率 85%，生氣 127 筆測試資料有 119 筆辨識為生氣，辨識正確率 93.7%，整體之辨識正確率為 89.1%，除了一部分悲傷與快樂測試資料容易被辨識為中性及生氣外，中性與生氣皆維持良好的分類效果。

同樣的，使用正規化特徵參數進行 SVM Outside Test，表 4-7 為 P1 ~ P10 使用正規化特徵參數 SVM Outside Test 鑑別矩陣(P1 ~ P10 各情緒之辨識正確率詳見附錄 A)，由鑑別矩陣結果顯示，悲傷 62 筆測試資料中有 51 筆資料被辨識為悲傷，悲傷之辨識正確率為 82%，中性 79 筆測試資料中有 79 筆資料被辨識為中性，中性之辨識正確率 100%，快樂 71 筆測試資料中有 57 筆被辨識為快樂，快樂之辨識正確率 80%，生氣 127 筆測試資料中有 124 筆資料被辨識為生氣，生氣之辨識正確率 97.6%，整體之辨識正確率為 91.7%，比使用原始特徵參數做 SVM Outside Test 時，整體之辨識正確率約提升 2.6%，由原始與正規化特徵參數之 SVM Outside Test 辨識結果可以看出，使用正規化特徵參數可以有效增加各情緒之間的分類效果。

最後，與 KNN 相同，考慮 SVM 分類器在訓練語料與測試語料為不同語系情況下進行辨識，同樣的，我們以德國語料庫所有資料當作訓練資料，並以工業技術研究院所錄製之語料(表 3-2)當作測試語料，表 4-8 為 SVM 使用正規化特徵參數在訓練與測試為不同語系下之鑑

別矩陣，由辨識結果顯示，SVM 對於悲傷與中性之辨識效果類似，悲傷與中性皆相對容易被辨識為中性，辨識率分別為 43.7% 與 67.3%，而快樂則與生氣之辨識效果類似，快樂與生氣皆相對容易被辨識為生氣，辨識率分別為 71.0% 與 70.3%，此結果表示中文語系悲傷與中性之特徵參數與德國語料庫之中性特徵參數相近，而快樂與生氣之特徵參數則與德國語料庫之生氣特徵參數相近，因此，若以德國語料庫當作訓練資料時，中文語系之悲傷情緒容易被辨識為中性，而快樂情緒則容易被辨識為生氣，整體之正確辨識率為 46.9%，此結果顯示，當 SVM 分類器在訓練語料與測試語料為不同語系時，SVM 分類器同樣無法得到良好的分類效果。

4.3 小結

在本章節中，首先，先針對所使用之 KNN 及 SVM 分類器做簡單的原理介紹，接著，根據上一章所提出之正規化特徵參數，分別以 KNN 與 SVM 分類器，對德國語料庫進行正規化特徵參數之 Inside Test，由辨識結果顯示，KNN 在使用正規化特徵參數作 Inside Test 情況下，整體之辨識正確率為 88.8%，而 SVM 則為 100%，此結果顯示，SVM 在測試資料在訓練資料中時，SVM 提供了極佳的分類效果。

而在 Outside Test 情況下時，KNN 使用原始與正規化特徵參數之整體辨識正確率分別為 78.2% 與 82.6%，SVM 則分別為 89.1% 與

91.7%，由辨識結果也可以發現，使用正規化特徵參數可以有效縮小語者之間在特徵參數上的差異，增加各情緒之間的分類效果。與 Inside Test 做比較，當測試資料不在訓練資料中時，SVM 與 KNN 之整體辨識正確率皆約下降 10% 左右，而 SVM 分類器之整體辨識效果仍優於 KNN 分類器。

最後，針對 KNN 及 SVM 分類器在訓練語料與測試語料為不同語系之情況下進行辨識，以德國語料庫所有資料當作訓練資料，並以工業技術研究院所錄製之語料當作測試語料，由辨識結果顯示，KNN 與 SVM 分類器在訓練語料與測試語料為不同語系時，兩者皆無法得到良好的分類效果，而由各情緒之辨識結果可以發現，悲傷與中性皆相對容易被辨識為中性，快樂與生氣皆相對容易被辨識為生氣與中性，此辨識結果與台灣人個性較為內斂有關，在表達情緒時，各情緒之特徵不若德國演員各情緒之呈現。

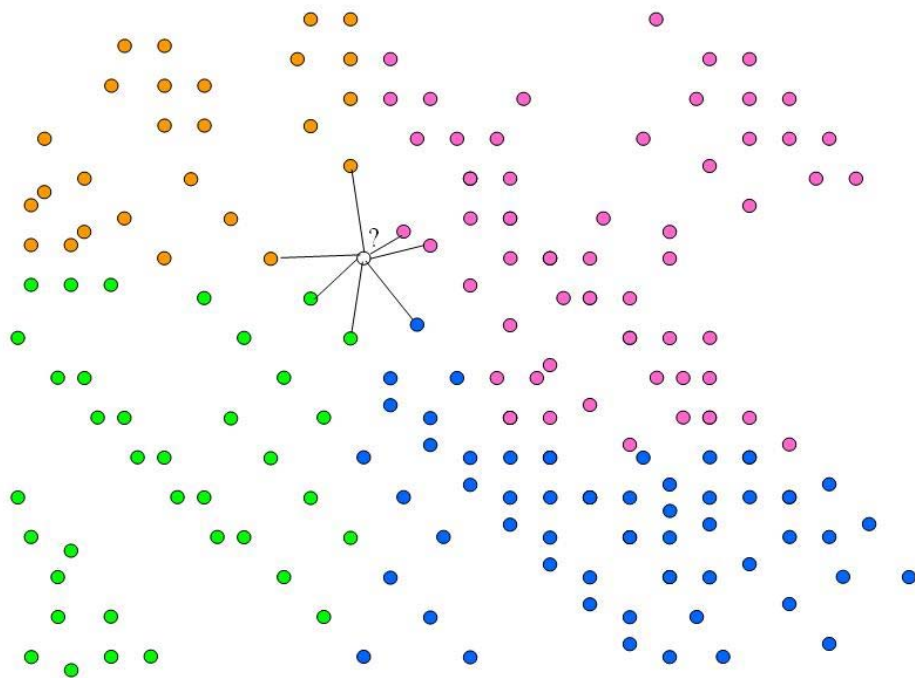


圖 4-1 第 K 個最近鄰(KNN)

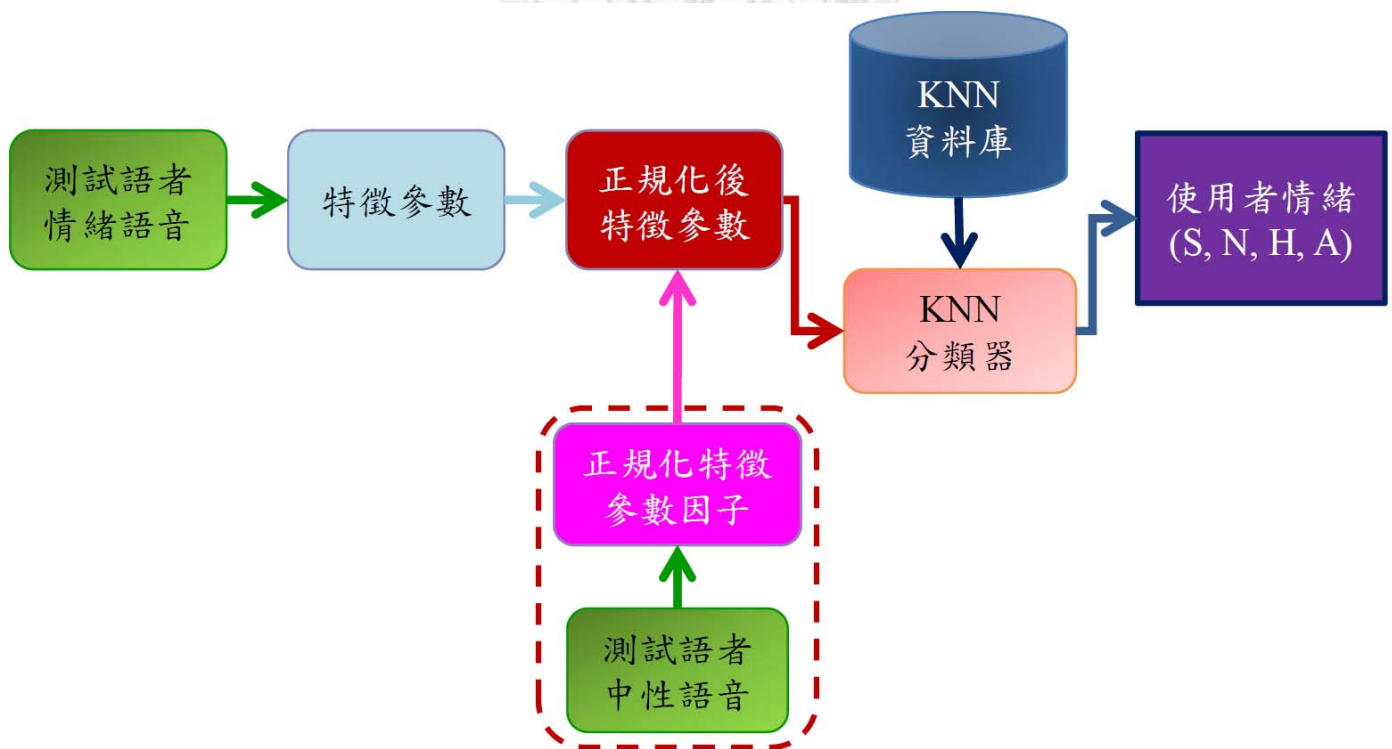


圖 4-2 KNN 語音情緒辨識流程

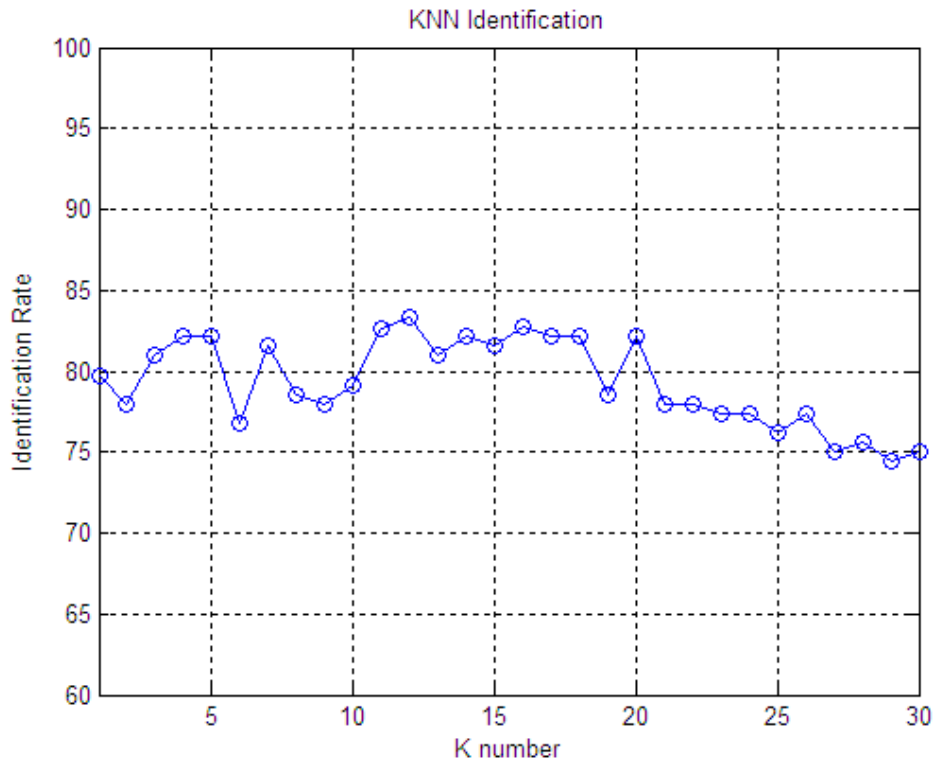


圖 4-3 KNN 不同 K 值之辨識率

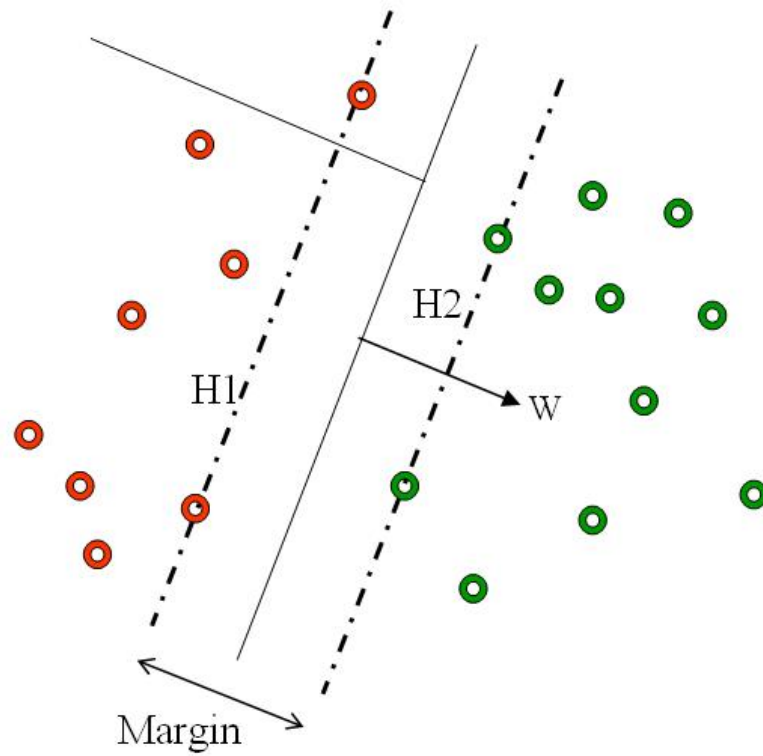


圖 4-4 二維支持向量機

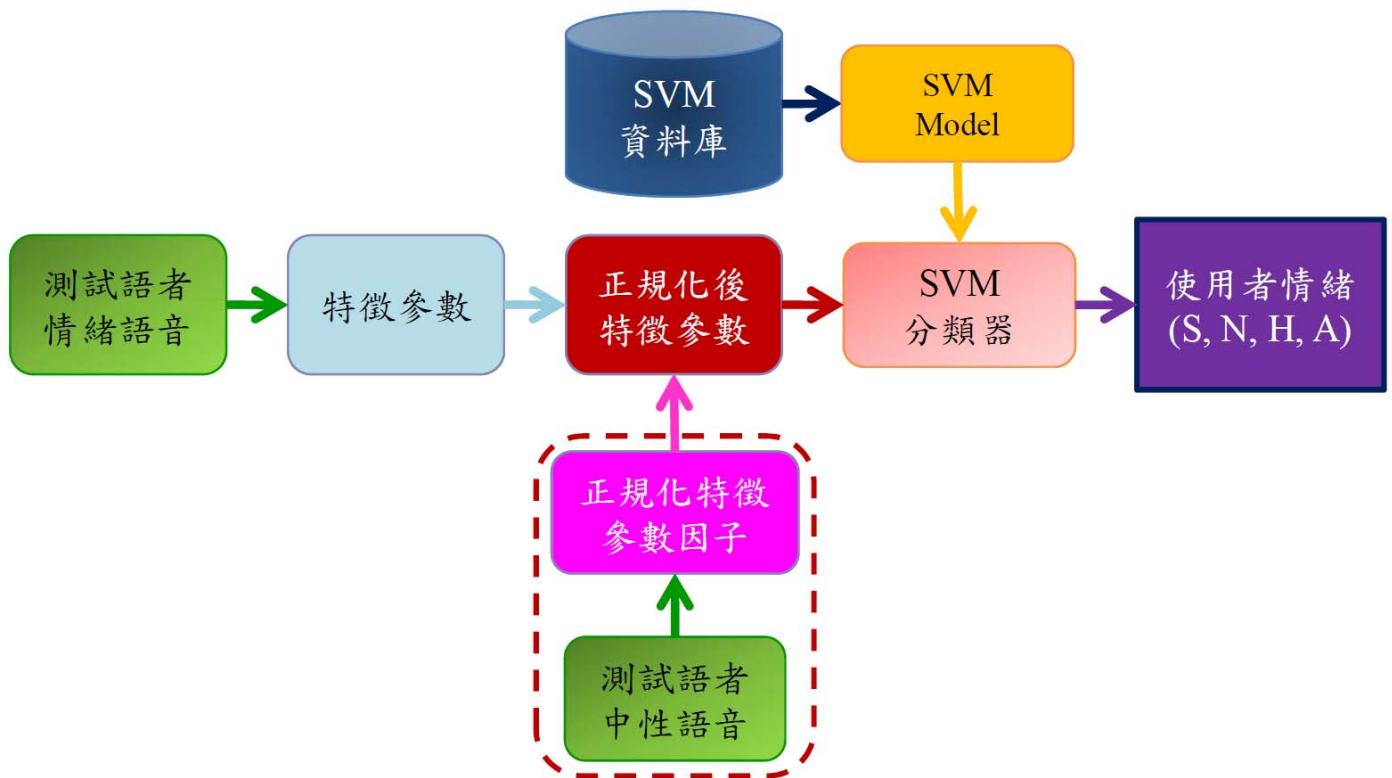


圖 4-5 SVM 語音情緒辨識流程

表 4-1 使用正規化特徵參數 KNN Inside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	84% (52/62)	16% (10/62)	0	0	84 (52/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	3% (2/71)	63% (45/71)	34% (24/71)	63% (45/71)
生氣(127)	0	0.8% (1/127)	0.8% (1/127)	98.4 (125/127)	98.4% (125/127)
平均辨識率					88.8% (301/339)

表 4-2 使用原始特徵參數 KNN Outside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	56% (35/62)	44% (27/62)	0	0	56% (35/62)
中性(79)	3% (2/79)	97% (77/79)	0	0	97% (77/79)
快樂(71)	0	4% (3/71)	45% (32/71)	51% (36/71)	45% (32/71)
生氣(127)	0	1.5% (2/127)	3.2% (4/127)	95.3% (121/127)	95.3% (121/127)
平均辨識率					78.2% (265/339)

表 4-3 使用正規化特徵參數 KNN Outside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	66% (41/62)	34% (21/62)	0	0	66% (41/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	7% (5/71)	55% (39/71)	38% (27/71)	55% (39/71)
生氣(127)	0.8% (1/127)	0.8% (1/127)	3.2% (4/127)	95.3% (121/127)	95.3% (121/127)
平均辨識率					82.6% (280/339)

表 4-4 KNN 不同語系測試之鑑別矩陣(正規化特徵參數)

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(600)	16.2% (97/600)	83.3% (500/600)	0.5% (3/600)	0	16.2% (97/600)
中性(600)	3.3% (20/600)	95.8% (575/600)	0.7% (4/600)	0.2% (1/600)	95.8% (575/600)
快樂(600)	3.7% (22/600)	65.8% (395/600)	22.5% (135/600)	8.0% (48/600)	22.5% (135/600)
生氣(600)	4.7% (28/600)	90.2% (541/600)	2.8% (17/600)	2.3% (14/600)	2.3% (14/600)
平均辨識率					34.21% (821/2400)

表 4-5 使用正規化特徵參數 SVM Inside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	100% (62/62)	0	0	0	100% (62/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	0	100% (71/71)	0	100% (71/71)
生氣(127)	0	0	0	100% (127/127)	100% (127/127)
平均辨識率					100% (339/339)

表 4-6 使用原始特徵參數 SVM Outside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	71% (44/62)	26% (16/62)	3% (2/62)	0	71% (44/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	1% (1/71)	85% (60/71)	14% (10/71)	85% (60/71)
生氣(127)	0	0	6.3% (8/127)	93.7% (119/127)	93.7% (119/127)
平均辨識率					89.1% (302/339)

表 4-7 使用正規化特徵參數 SVM Outside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	82% (51/62)	18% (11/62)	0	0	82% (51/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	1% (1/71)	80% (57/71)	18% (13/71)	80% (57/71)
生氣(127)	0	0	2.4% (3/127)	97.6% (124/127)	97.6% (124/127)
平均辨識率					91.7% (311/339)

表 4-8 SVM 不同語系測試之鑑別矩陣(正規化特徵參數)

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(600)	32.8% (197/600)	43.7% (262/600)	3.7% (22/600)	19.8% (119/600)	32.8% (197/600)
中性(600)	16.7% (100/600)	67.3% (404/600)	5.3% (32/600)	10.7% (64/600)	67.3% (404/600)
快樂(600)	6.0% (36/600)	5.8% (35/600)	17.2% (103/600)	71.0% (426/600)	17.2% (103/600)
生氣(600)	8.3% (50/600)	12.8% (77/600)	8.5% (51/600)	70.3% (422/600)	70.3% (422/600)
平均辨識率					46.92% (1126/2400)

第五章 多層貝氏網路與多層共變異 數貝氏網路語音情緒辨識

5.1 決策樹、貝氏決策與貝氏網路介紹

5.1.1 決策樹 (Decision Tree)

決策樹(Decision Tree)，一般也稱為判定樹，在決策樹的方法中，首先，先從實際範例集合中，針對資料進行處理，利用歸納學習法建立規則與決策樹，然後使用決策對新資料進行分析與分類，因此，在本質上，決策樹是透過一系列的規則來對資料進行分類的過程。

歸納學習法是決策樹中找出資料模式與規則的核心，其知識來源是一組描述某事物中各情況的範例，此範例稱為『訓練資料庫』，訓練資料庫應該盡量涵蓋事物的各種情況，並且在統計上具有良好的取樣特徵，能夠真實的代表母體的特質，才能得到有效的歸納結果。而歸納的推理過程則是從若干個事實的特徵、特性或屬性中，透過比較概括而得的一個規律性結論，歸納推理會從事物的一部分或整體的特定觀察中得到一個完整的描述，也就是從觀察的事件中找出普遍規律性的結論，另外，在推理的過程中，必須找到有效的歸納結果，一般而言，有效的歸納結果指的就是使用歸納推理出的分類規則與實際測試的預測結果具有很高的相同機率，而人類知識的增長過程中，主要

的部分也是來自於歸納學習，。

歸納學習法由於依賴經驗資料，因此，又稱做經驗學習，歸納學習一般可分為由下往上、由上往下以及雙向搜尋三種方式，而決策樹的基本組成部分有決策節點、分支以及葉節點，決策樹中最上面的節點稱為根節點(Root)，是整個決策樹的開始，每個分支為一個新的決策節點，或者是樹的葉節點，每一個決策節點可能代表一個問題或決策，每一個葉節點則代表一個可能的分類結果，在沿著決策樹從上到下的過程中，再每個節點都會遇到一個測試，在每個節點上，不同的測試結果將會導致不同的分支，最後到達一個葉節點。決策樹分類的過程，主要是決策樹在進行分類的過程中，利用若干個變數來判斷屬性，最後得到所屬的類別。圖 5-1 為簡單的交通工具分類範例，在決策判斷分類的過程中，車輪數量、車身高度以及加油與否為測試的節點，當車輪數量大於或等於四且車身高度大於或等於 2m 時，車子被判斷歸類為大型車輛，當車輪數量大於或等於四但車身高度小於 2m 時，車子被判斷歸類為家庭房車，而當車輪數量小於四且車子需要加油時，則車子被判斷歸類為摩托車，當車輪數量小於四且車子不需要加油時，則車子被判斷歸類為腳踏車，大型車輛、家庭房車、摩托車以及腳踏車分別為交通工具分類樹中的葉節點。

5.1.2 貝氏決策(Bayes Decision)與貝氏網路(Bayesian Network)

貝氏決策為統計樣式識別(Statistical Pattern Recognition)的基礎，此方法建立在以機率方式來表示各個類別可能發生的機率，並以此作為以機率判斷樣式類別的基礎。

當樣式空間中具有 $C_1, C_2, C_3, \dots, C_n$ 之 n 個類別，在觀察事件後，給定一含有 m 個特徵之觀察參數 $X_1, X_2, X_3, \dots, X_m$ ，根據貝氏定理，可以得到下式(5.1.1)

$$P(C_i|\mathbf{X}) = \frac{P(\mathbf{X}|C_i)P(C_i)}{P(\mathbf{X})} \quad (5.1.1)$$

其中， $P(C_i)$ 為事前機率(Prior Probability)，代表 C_i 這個類別發生的機率， $P(C_i|\mathbf{X})$ 為事後機率(Posterior Probability)，表示在觀察到特徵參數向量 \mathbf{X} 的前提下，類別為 C_i 的機率， $P(\mathbf{X}|C_i)$ 為條件機率，表示類別 C_i 出現特徵參數 \mathbf{X} 的機率， $P(\mathbf{X})$ 則為特徵參數向量 \mathbf{X} 出現的機率，其算式如(5.1.2)所示。

$$P(\mathbf{X}) = \sum_{i=1}^n P(\mathbf{X}|C_i)P(C_i) \quad (5.1.2)$$

圖 5-2 為四種類別 C_1, C_2, C_3, C_4 時，在二維空間情況下之示意圖，所謂的貝氏決策則是依據貝氏定理，找出在觀察到特徵參數 \mathbf{X} 的前提下，找出屬於某種類別的最大機率，也就是

$$k = \arg \max_i P(C_i|\mathbf{X}) \quad (5.1.3)$$

由於 $P(\mathbf{X}) = \sum_{i=1}^n P(\mathbf{X}|C_i)P(C_i)$ 在每個類別中均相同，因此式(5.1.3)可以改寫成

$$k = \arg \max_i P(C_i | \mathbf{X}) = \arg \max_i P(\mathbf{X} | C_i) P(C_i) \quad (5.1.4)$$

式(5.1.4) 稱之為貝氏決策規則(Bayes's decision rule)。

以一維之常態分布為例，若樣式空間中有兩個不同的類別，如圖 5-3 所示，分別為 C_1 、 C_2 ，假設類別 C_1 的事前機率 $P(C_1)$ 為 0.5，而類別 C_1 之常態分佈平均值與標準差分別為 $\mu_1 = 142.95$ ， $\sigma_1 = 34.61$ ，類別 C_2 的事前機率 $P(C_2)$ 為 0.5，而類別 C_2 之常態分佈平均值與標準差分別為 $\mu_2 = 261.77$ ， $\sigma_2 = 57.27$ ，根據貝氏決策規則(5.1.4)，我們可以找到一決策邊界(Decision boundary)，由於類別 C_1 的事前機率 $P(C_1)$ 與類別 C_2 的事前機率 $P(C_2)$ 皆為 0.5，因此，決策邊界恰好為兩常態分佈的交點位置，根據貝氏決策規則，當觀測值大於此決策邊界值時，則 $P(\mathbf{X} | C_2) P(C_2) > P(\mathbf{X} | C_1) P(C_1)$ ，因此觀測值將被歸類為類別 C_2 ，反之，若當觀測值小於此決策邊界值時，也就是 $P(\mathbf{X} | C_2) P(C_2) < P(\mathbf{X} | C_1) P(C_1)$ ，則觀測值將被歸類為類別 C_1 。

而貝氏網路則是一種以貝氏機率為基礎，並運用圖形化來描述變數之間的條件機率關係的方法，又稱為有向非循環圖形模型(Directed Acyclic Graphical model)，藉由有向非循環圖形(Directed Acyclic Graphs, DAGs)得知隨機變數 $\{x_1, x_2, \dots, x_n\}$ 及其條件機率分配(Conditional Probability Distributions, CPDs)，因此，貝氏網路可說是貝氏決策的延伸，將各變數之間的關係以圖形化的方式來描述其彼此

之間的關係。

5.2 多層貝氏網路(Multi-Layer Bayesian Network, MLBN)語音情緒辨識

在分類過程中，由於特徵參數與特徵參數之間對於不同的情緒類別具有不同的分類效果，有些特徵參數對於悲傷與中性的分類效果不佳，有些則是對於快樂與生氣的分類效果差，但有些特徵參數對於某種情緒類別則有特別強的分類效果，例如正規化後之平均音高值對於中性則有比較強的分類效果(如圖 3-4)，因此，本章透過貝氏決策與決策樹結合的方式，利用決策樹分層解析的優點結合貝氏決策，把不同分類效果的情緒特徵以分層的方式，發展一貝氏網路情緒辨識方法，將不同的情緒類別辨識出來，以下將就多層貝式網路的分群與辨識過程加以說明。

5.2.1 語音情緒特徵分群分析

由於分類樹是透過歸納學習，將不同類別透過層層解析的方式給分離辨識出來，而貝氏決策則是利用貝氏機率運算來描述各類別發生的機率，因此，本章期望透過分類樹與貝氏決策的優點，將不同特徵參數以分層解析方式，並結合貝氏決策之機率運算，以貝氏網路的方式將不同之語音情緒類別給分辨出來，在結合分類樹與貝氏決策之前，由於每個特徵參數對於不同情緒類別的分類效果不同，因此，必

須先將各特徵參數進行分層與歸類，首先，先針對所使用之特徵參數進行統計與歸納分析，以表 3-5 之第一個特徵正規化後音高平均值為例，將德國語料庫十位演員語料所計算得到之正規化後音高平均值以統計方式計算每一種情緒類別之平均值與標準差，也就是將十位演員之正規化後音高平均值(圖 3-4)中各情緒的點資料做整體統計，計算其平均值與標準差，圖 5-4 為四種不同情緒正規化後音高平均值分佈，由圖中可以發現，在正規化後音高平均值的分佈上，悲傷與中性可以很明顯的被分辨出來，形成獨立的一群，而快樂與生氣則無法明顯分辨，形成另外一群，圖 5-5(a) ~ (w) 為正規化後特徵參數經統計分析後之分佈狀態，由圖中可以發現，有些特徵參數悲傷與中性無法明顯被分辨出來，有些則是快樂與生氣無法明顯被分辨出來，而有些則是對於四種情緒類別均無法明顯分辨，因此，本節首先針對每一個特徵參數之各情緒類別先進行分群，並以前一節所介紹之貝式決策(BD)作為類別分群的依據。

由式(5.1.4)來作為特徵參數之類別分群時，首先，先假設四種情緒類別之事前機率 $P(C_i)$ 皆相同，因此，式(5.1.4)可改寫成

$$k = \arg \max_i P(X|C_i) \quad (5.2.1)$$

根據式(5.2.1)，在特徵參數之類別分群上，我們僅需考量各特徵參數在值域範圍內各情緒類別之機率密度(Probability density)。由於

分類樹是將分類結果透過層層解析的方式給分辨出來，因此，在分類樹之使用上必須先經由歸納處理，將不同分類效果之特徵參數作歸類。依據各特徵參數各情緒類別之分佈狀態，本論文提出一分群歸類之方法，將各特徵參數做分層歸類，以下將針對個別特徵參數分層分群的方式加以說明。

首先，我們分別計算在固定特徵參數且考慮只有兩個情緒類別的情況下(如圖 5-6 為例，只單純考慮悲傷與其他情緒類別之間)，各情緒類別之間彼此的正確辨識機率，以正規化後音高平均值為例，我們先計算只有悲傷與中性兩個情緒類別(如圖 5-6(b))時，在貝式決策分類下悲傷之正確辨識機率；根據常態分佈之機率密度函數(5.2.2)，在固定參數下，兩個常態分佈之交點位置為符合式(5.2.3)之解，經整理後，相當於求解一元二次方程式之解，如式(5.2.4)所示。

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (5.2.2)$$

$$\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} - \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} = 0 \quad (5.2.3)$$

$$(\sigma_1^2 - \sigma_2^2)x^2 + 2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2)x + (\mu_2^2\sigma_1^2 - \mu_1^2\sigma_2^2 + 2\sigma_1^2\sigma_2^2 \ln \frac{\sigma_2}{\sigma_1}) = 0 \quad (5.2.4)$$

假設欲辨識之主要情緒類別分佈為 $N_1(\mu_1, \sigma_1^2)$ ，另一情緒類別分佈為 $N_2(\mu_2, \sigma_2^2)$ ，由式(5.2.4)，當 $\sigma_1 = \sigma_2 = \sigma$ 時，則式(5.2.4)可以簡化為

$x = \frac{\mu_1 + \mu_2}{2}$ ，兩常態分佈交於一點，因此，若 $\mu_1 > \mu_2$ ，根據貝氏決策，當辨識特徵值 $x > \frac{\mu_1 + \mu_2}{2}$ 時，則 x 將被辨識為主要情緒類別 $N_1(\mu_1, \sigma_1^2)$ ，根據累積分佈函數(Cumulative Distribution Function CDF)，主要情緒類別之辨識正確率如式(5.2.5)所示。

$$\text{主情緒類別正確辨識率} = \int_{\frac{\mu_1 + \mu_2}{2}}^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \cdot \left(\frac{x - \mu_1}{\sigma_1}\right)^2\right] dx \quad (5.2.5)$$

反之，若 $\mu_1 < \mu_2$ ，根據貝氏決策，當辨識特徵值 $x < \frac{\mu_1 + \mu_2}{2}$ 時，則 x 將被辨識為主要情緒類別 $N_1(\mu_1, \sigma_1^2)$ ，主要情緒類別之辨識正確率如式(5.2.6)所示。

$$\text{主情緒類別正確辨識率} = \int_{-\infty}^{\frac{\mu_1 + \mu_2}{2}} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \cdot \left(\frac{x - \mu_1}{\sigma_1}\right)^2\right] dx \quad (5.2.6)$$

當 $\sigma_1 \neq \sigma_2$ 時，根據一元二次方程式之通解，則兩常態分佈之交點位置為式(5.2.7)所示

$$x_{1,2} = \frac{-2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) \pm \sqrt{4(\mu_1\sigma_2^2 - \mu_2\sigma_1^2)^2 - 4(\sigma_1^2 - \sigma_2^2)(\mu_2^2\sigma_1^2 - \mu_1^2\sigma_2^2 + 2\sigma_1^2\sigma_2^2 \ln \frac{\sigma_2}{\sigma_1})}}{2(\sigma_1^2 - \sigma_2^2)} \quad (5.2.7)$$

經整理後

$$x_{1,2} = \frac{-2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) \pm \sqrt{4\sigma_1^2\sigma_2^2 \left[(\mu_1 - \mu_2)^2 + 2(\sigma_2^2 - \sigma_1^2) \ln \frac{\sigma_2}{\sigma_1} \right]}}{2(\sigma_1^2 - \sigma_2^2)}$$

(5.2.8)

由於 $4\sigma_1^2\sigma_2^2\left[(\mu_1 - \mu_2)^2 + 2(\sigma_2^2 - \sigma_1^2)\ln\frac{\sigma_2}{\sigma_1}\right] \geq 0$ ，因此，當 $\sigma_1 \neq \sigma_2$ 或

$\mu_1 \neq \mu_2$ 時，則兩常態分佈必有兩相異交點。

同樣的，根據貝氏決策判斷式(5.2.1)式，當 $x_2 \leq x \leq x_1$ 且

$\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} > \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}$ 時，則 $x_2 \leq x \leq x_1$ 之值域將被辨識為

主要情緒類別 $N_1(\mu_1, \sigma_1^2)$ ，而 $x \geq x_1$ 及 $x \leq x_2$ 之值域則被辨識為次要情

緒類別 $N_2(\mu_2, \sigma_2^2)$ ，反之，當 $x_2 \leq x \leq x_1$ 且

$\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} < \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}$ 時，則 $x_2 \leq x \leq x_1$ 之值域將被辨識為

次要情緒類別 $N_2(\mu_2, \sigma_2^2)$ ，而 $x \geq x_1$ 及 $x \leq x_2$ 之值域將被辨識為主要情

緒類別 $N_1(\mu_1, \sigma_1^2)$ 。

為降低計算複雜度，並增進自動化辨識流程，在此，我們以兩交點之中間值作為 $x_2 \leq x \leq x_1$ 之分類依據，令 $\bar{x} = \frac{x_1 + x_2}{2}$ ，當

$\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu_1}{\sigma_1}\right)^2} > \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu_2}{\sigma_2}\right)^2}$ ，則 $x_2 \leq x \leq x_1$ 之值域將被辨識為主

要情緒類別 $N_1(\mu_1, \sigma_1^2)$ ，而 $x \geq x_1$ 及 $x \leq x_2$ 之值域將被辨識為次要情緒

類別 $N_2(\mu_2, \sigma_2^2)$ ，因此，在此情況下，主要情緒類別之辨識正確率如

式(5.2.9)所示。

$$\text{主情緒類別正確辨識率} = \int_{x_2}^{x_1} \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2} \cdot \left(\frac{x-\mu_1}{\sigma_1}\right)^2\right] dx$$

(5.2.9)

而當 $\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu_1}{\sigma_1}\right)^2} < \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu_2}{\sigma_2}\right)^2}$ ，則 $x_2 \leq x \leq x_1$ 之值域將被

辨識為次要情緒類別 $N_2(\mu_2, \sigma_2^2)$ ，而 $x \geq x_1$ 及 $x \leq x_2$ 之值域將被辨識為主要情緒類別 $N_1(\mu_1, \sigma_1^2)$ ，在此情況下，主要情緒類別之辨識正確率如式(5.2.10)所示。

主情緒類別正確辨識率 =

$$\int_{-\infty}^{x_2} \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2} \cdot \left(\frac{x-\mu_1}{\sigma_1}\right)^2\right] dx + \int_{x_1}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2} \cdot \left(\frac{x-\mu_1}{\sigma_1}\right)^2\right] dx \quad (5.2.10)$$

由前面所推導之兩常態分佈貝氏決策之辨識正確率計算，以悲傷與中性兩個情緒類別之正規化後音高平均值(圖 5-6(b))為例，並以悲傷為主要辨識之情緒類別，中性為次要情緒類別，經由計算， $N_1(\mu_1, \sigma_1^2) = N(0.8811, 0.0098)$ ， $N_2(\mu_2, \sigma_2^2) = N(1, 0.0033)$ ，兩常態分佈之交點位置分別為 $x_1 = 1.1882$ 及 $x_2 = 0.9329$ ，中點位置為 $\bar{x} = 1.0606$ ， $p_1(\bar{x}) = 0.7788$ ， $p_2(\bar{x}) = 3.9828$ ，因此，在只有悲傷與中性兩情緒類別下，悲傷之正確辨識率以式(5.2.10)計算而得，悲傷之正確辨識率為 70.07%，同理，考慮悲傷與快樂兩情緒類別時(圖 5-6(c))，悲傷之正確辨識率為 99.13%，表 5-1 為正規化後平均音高值在單純考慮兩情緒類別時，各情緒類別之辨識正確率矩陣，由結果顯示，當考慮悲傷與中性兩情緒時，悲傷之辨識正確機率為 70.07%，而中性的正確辨識機率為 87.77%，顯示正規化後平均音高特徵對於

悲傷與中性有不錯之分類效果，而在悲傷與快樂兩情緒下，悲傷正確辨識機率 99.13%，快樂正確辨識機率 96.57%，此結果顯示，此特徵對於悲傷與快樂有極佳的分類效果，同樣的，悲傷與生氣、中性與快樂、中性與生氣亦具有極佳的分類效果，而當兩情緒為快樂與生氣時，快樂的正確辨識率為 41.30%，生氣的正確辨識率為 67.89%，此結果顯示，此特徵對於快樂與生氣之分類效果有限，在此，我們訂定一特徵參數分群的原則與步驟，以下為特徵參數分群的原則與步驟。

(1)、假設運算所得之個別特徵各情緒類別辨識正確率矩陣為 T ， T 為

一 4×4 之矩陣，其中 T_{12} 為悲傷與中性兩情緒類別時，悲傷之正確辨識率， T_{13} 為悲傷與快樂兩情緒類別時，悲傷之正確辨識率， T_{14} 為悲傷與生氣兩情緒類別時，悲傷之正確辨識率， T_{21} 為悲傷與中性兩情緒類別時，中性之正確辨識率，以此類推。

(2)、當 $T_{ij} + T_{ji} < 120\%$ ($i \neq j$) 時，也就是兩類別之平均辨識率結果小於

60% 時，且此兩情緒與其他情緒之正確辨識率和皆大於 120%，

則兩情緒視為同一群，如正規化後音高平均值之快樂與生氣，其辨識率和為 109.19%，因此，在此特徵下，可視為同一群。

(3)、當兩情緒類別 A 與 B 之正確辨識率和小於 120%，兩情緒類別 B

與 C 或 A 與 C 之正確辨識率和亦小於 120%，且 A、B 以及 C

與另一情緒類別 D 之正確辨識率和皆大於 120% 時，則 A、B 以

及 C 在此特徵下，視為同一群。

根據上面分群的原則與步驟，正規化後平均音高值 (表 5-1)，可以將快樂與生氣當成同一群，而悲傷與中性則分別為獨立的一群，因此，正規化後音高平均值為分三群(悲傷、中性、快樂與生氣三群)的特徵之一。同樣的，考慮正規化後音高之標準差，其各情緒類別之辨識正確率矩陣如表 5-2 所示，由結果顯示，與正規化後平均音高特徵相同、此特徵對於快樂與生氣分類效果不佳，根據分群的原則，快樂與生氣可視為同一群，正規化後平均音高之標準差特徵亦可視為分三群的特徵之一。

附錄 B 為各正規化後特徵參數之辨識正確率矩陣，根據分群的原則及矩陣結果，有六種不同分群的組合，如表 5-3 所示，首先，分兩群的主要有三種，第一種為悲傷與中性為同一群，快樂與生氣則為另一群，相關的特徵有 MFCC C8 平均值以及 MFCC C10 平均值；第二種為悲傷為獨立的一群，中性、快樂以及生氣為另一群，相關的特徵有正規化後能量之標準差以及正規化後第二共振峰平均值；第三種則為中性為獨立的一群，悲傷、快樂以及生氣為另一群，相關的特徵有正規化後第一共振峰平均值。

其次，特徵參數分三群的則有兩種，第一種為悲傷以及中性各為獨立的一群，快樂與生氣為第三群，相關的特徵有正規化後音高平均

值、正規化後音高之標準差、正規化後第一共振峰之標準差、正規化後第二共振峰之標準差、正規化後第三共振峰平均值、MFCC C3 平均值、MFCC C5 平均值、MFCC C6 平均值以及 MFCC C7 平均值；第二種則為悲傷與中性為一群，快樂以及生氣則各為獨立的一群，相關的特徵有 MFCC C9 平均值。

最後，則是悲傷、中性、快樂與生氣皆為獨立的一群，主要的特徵有正規化後能量平均值、正規化後第三共振峰之標準差、MFCC C1 平均值、MFCC C2 平均值、MFCC C4 平均值、MFCC C11 平均值、MFCC C12 平均值以及 MFCC C13 平均值。

根據以上之分群方式，本論文提出一多層貝氏網路分類方法，下一章節將就多層貝氏網路語音情緒分類方法加以說明。

5.2.2 多層貝氏網路(MLBN)

根據上一章節之分群結果，本論文所提出之結合決策樹與貝氏網路之多層貝氏網路語音情緒辨識架構如圖 5-7 所示，主要分成三層，在分層分群的過程中，先將資料庫中的語音情緒特徵資料以表 5-3 之分群方式重新歸類，並重新計算各特徵參數分群後之平均值與標準差，以第一層(圖 5-8)之特徵參數 MFCC C8 平均值為例，主要分為兩群，悲傷與中性為一群，快樂與生氣則為另一群，圖 5-9(a)為原始分佈狀態，圖 5-9(b)則為分群後之分佈狀態，從圖 5-9(b)中可以發現，

分群後各群之分佈明顯分開，將有助於後續分類辨識之進行，圖 5-10(a) ~ (e) 為多層貝氏網路第一層之各特徵參數分群後分佈狀態，與原先各情緒之分佈狀態(圖 5-5)比較後可以發現，特徵參數在分群後，各群之分佈可以明顯被分辨出來。多層貝氏網路第二層(圖 5-11)為分三群，主要有兩種，第一種為悲傷以及中性各為獨立的一群，快樂與生氣為第三群，另一種則為悲傷與中性為一群，快樂以及生氣則各為獨立的一群，圖 5-12(a) ~ (j) 為多層貝氏網路第二層各特徵參數分三群後之分佈狀態。第三層則為分四群(圖 5-13)，也就是最原始之情緒類別，其分佈狀態可參考圖 5-5 之正規化後各特徵參數分佈圖。

與 KNN 及 SVM 語音情緒辨識流程類似，多層貝氏網路語音情緒辨識之流程如圖 5-14 所示，首先，測試語者先講一段正規化用中性情緒語音，並利用此正規化用之中性情緒語音，計算出測試語者之正規化特徵參數因子，接著，進入語音情緒辨識流程，在得到使用者一段情緒語音訊號後，先計算出原始特徵參數，並經由正規化特徵參數因子對原始特徵參數做正規化，最後，將正規化後特徵參數與多層貝氏網路資料庫做分層分群運算，計算出每一情緒類別之發生機率。以下將就此多層貝氏網路語音情緒分類器，透過分層的方式分別加以說明。

假設測試語者帶有情緒之語音訊號經正規化計算後所得之特徵

參數(如表 3-5 所列)為一特徵參數向量 \mathbf{X} ， \mathbf{X} 為一 23×1 之向量， X_1 為正規化後音高平均值， X_2 為正規化後音高之標準差， X_{11} 為 MFCC C1 平均值，以此類推。首先，考慮多層貝氏網路語音情緒辨識第一層(圖 5-8)之特徵參數分群運算，主要有三種分群方式，第一種為悲傷與中性為同一群，快樂與生氣為另一群，為簡化運算式之表示式，我們將悲傷以英文 S 表示，中性以 N 表示，快樂與生氣則分別以 H 以及 A 表示，因此，多層貝氏網路語音情緒辨識第一種分群之機率運算式分別為

$$\begin{aligned} \text{Prob}_{1,1}^1 &= \sum_i w_i \cdot P(S, N | X_i) \\ &= \sum_i w_i \frac{P(X_i | S, N)P(S, N)}{P(X_i | S, N)P(S, N) + P(X_i | H, A)P(H, A)} \\ i &= 18, 20 \end{aligned} \quad (5.2.11)$$

$$\begin{aligned} \text{Prob}_{1,2}^1 &= \sum_i w_i \cdot P(H, A | X_i) \\ &= \sum_i w_i \frac{P(X_i | H, A)P(H, A)}{P(X_i | S, N)P(S, N) + P(X_i | H, A)P(H, A)} \\ i &= 18, 20 \end{aligned} \quad (5.2.12)$$

其中 $w_{18} + w_{20} = 1$ ，在本論文中，使用 $w_{18} = \frac{1}{2}$ ， $w_{20} = \frac{1}{2}$ ， $\text{Prob}_{1,1}^1$ 中之上標 1 表示此機率為多層貝氏網路中之第一層，也就是分兩群的分類方式，而下標 1,1 中第一個 1 表示該層中之第一種分群方式，第二個 1 則表示第一種分群方式中之第一群 S,H，以 $\text{Prob}_{1,1}^1$ 為例，此機率表示第一層中第一種分群方式之第一分群 S,H 發生之機率，也就是悲

傷與中性為同一群，快樂與生氣為另一群時，悲傷與中性發生之機率，而 $\text{Prob}_{1,2}^1$ 則表示第一層第一種分群方式中，快樂與生氣發生之機率。 $P(X_i | S, N)$ 則表示在悲傷與中性為同一群分佈下， X_i 發生之機率密度， $P(X_i | H, A)$ 表示在快樂與生氣為同一群分佈下， X_i 發生之機率密度， $P(S, N)$ 以及 $P(H, A)$ 則分別代表悲傷與中性以及快樂與生氣之事前機率，在此，我們假設各情緒發生之事前機率均相同，因此， $P(S, N) = \frac{1}{2}$ ， $P(H, A) = \frac{1}{2}$ 。

其次，考慮多層貝氏網路語音情緒辨識第一層第二種分群方式，悲傷為獨立的一群，中性、快樂以及生氣則為另一群，其分群之機率運算分別為

$$\begin{aligned} \text{Prob}_{2,1}^1 &= \sum_i w_i \cdot P(S | X_i) \\ &= \sum_i w_i \frac{P(X_i | S)P(S)}{P(X_i | S)P(S) + P(X_i | N, H, A)P(N, H, A)} \\ i &= 4, 7 \end{aligned} \quad (5.2.13)$$

$$\begin{aligned} \text{Prob}_{2,2}^1 &= \sum_i w_i \cdot P(N, H, A | X_i) \\ &= \sum_i w_i \frac{P(X_i | N, H, A)P(N, H, A)}{P(X_i | S)P(S) + P(X_i | N, H, A)P(N, H, A)} \\ i &= 4, 7 \end{aligned} \quad (5.2.14)$$

其中 $w_4 + w_7 = 1$ ，在本論文中，使用 $w_4 = \frac{1}{2}$ ， $w_7 = \frac{1}{2}$ ，而 $\text{Prob}_{2,1}^1$ 表示第一層第二種分群方式中悲傷發生之機率， $\text{Prob}_{2,2}^1$ 則表示第一層第二種分群方式，中性、快樂以及生氣發生之機率。 $P(X_i | S)$ 表示在悲

傷為獨立分群分佈下， X_i 發生之機率密度， $P(X_i | N, H, A)$ 表示在中性、快樂與生氣為同一群分佈下， X_i 發生之機率密度， $P(S)$ 代表悲傷之事前機率， $P(N, H, A)$ 則為中性、快樂與生氣為同一群之事前機率，因此， $P(S) = \frac{1}{4}$ ， $P(N, H, A) = \frac{3}{4}$ 。

最後，考慮多層貝氏網路語音情緒辨識第一層第三種分群方式，中性為獨立的一群，悲傷、快樂以及生氣為另一群，由於只有使用單一特徵，因此，其分群之機率運算為

$$\text{Prob}_{3,1}^1 = P(N | X_5) = \frac{P(X_5 | N)P(N)}{P(X_5 | N)P(N) + P(X_5 | S, H, A)P(S, H, A)} \quad (5.2.15)$$

$$\begin{aligned} \text{Prob}_{3,2}^1 &= P(S, H, A | X_5) \\ &= \frac{P(X_5 | S, H, A)P(S, H, A)}{P(X_5 | N)P(N) + P(X_5 | S, H, A)P(S, H, A)} \end{aligned} \quad (5.2.16)$$

$\text{Prob}_{3,1}^1$ 為多層貝氏網路語音情緒辨識第一層第三種分群方式中性發生之機率， $\text{Prob}_{3,2}^1$ 則為第一層第三種分群方式中，悲傷、快樂以及生氣發生之機率。而 $P(X_5 | N)$ 表示在中性為獨立分群分佈下， X_5 發生之機率密度， $P(X_5 | S, H, A)$ 表示在悲傷、快樂與生氣為同一群分佈下， X_5 發生之機率密度， $P(N)$ 代表中性本身之事前機率， $P(S, H, A)$ 則為中性、快樂與生氣為同一群之事前機率，因此， $P(N) = \frac{1}{4}$ ， $P(S, H, A) = \frac{3}{4}$ 。

接著，考慮多層貝氏網路語音情緒辨識第二層(圖 5-11)之特徵參

數分群運算，主要有兩種分群方式，第一種為悲傷以及中性各為獨立的一群，快樂與生氣為第三群，其分群之機率運算分別為

$$\begin{aligned}
 \text{Prob}_{1,1}^2 &= \sum_i w_i \cdot P(S | X_i) \\
 &= \sum_i w_i \left(\frac{1}{3} \frac{P(X_i | S) \frac{\text{Prob}_{1,1}^1}{2}}{P(X_i | S) \frac{\text{Prob}_{1,1}^1}{2} + P(X_i | N) \frac{\text{Prob}_{1,1}^1}{2} + P(X_i | H, A) \text{Prob}_{1,2}^1} \right. \\
 &\quad + \frac{1}{3} \frac{P(X_i | S) \text{Prob}_{2,1}^1}{P(X_i | S) \text{Prob}_{2,1}^1 + P(X_i | N) \frac{\text{Prob}_{2,2}^1}{3} + P(X_i | H, A) \frac{2 \cdot \text{Prob}_{2,2}^1}{3}} \\
 &\quad \left. + \frac{1}{3} \frac{P(X_i | S) \frac{\text{Prob}_{3,2}^1}{3}}{P(X_i | S) \frac{\text{Prob}_{3,2}^1}{3} + P(X_i | N) \text{Prob}_{3,1}^1 + P(X_i | H, A) \frac{2 \cdot \text{Prob}_{3,2}^1}{3}} \right) \\
 i &= 1, 2, 6, 8, 9, 13, 15, 16, 17 \tag{5.2.17}
 \end{aligned}$$

$$\begin{aligned}
 \text{Prob}_{1,2}^2 &= \sum_i w_i \cdot P(N | X_i) \\
 &= \sum_i w_i \left(\frac{1}{3} \frac{P(X_i | N) \frac{\text{Prob}_{1,1}^1}{2}}{P(X_i | S) \frac{\text{Prob}_{1,1}^1}{2} + P(X_i | N) \frac{\text{Prob}_{1,1}^1}{2} + P(X_i | H, A) \text{Prob}_{1,2}^1} \right. \\
 &\quad + \frac{1}{3} \frac{P(X_i | N) \frac{\text{Prob}_{2,2}^1}{3}}{P(X_i | S) \text{Prob}_{2,1}^1 + P(X_i | N) \frac{\text{Prob}_{2,2}^1}{3} + P(X_i | H, A) \frac{2 \cdot \text{Prob}_{2,2}^1}{3}} \\
 &\quad \left. + \frac{1}{3} \frac{P(X_i | N) \text{Prob}_{3,1}^1}{P(X_i | S) \frac{\text{Prob}_{3,2}^1}{3} + P(X_i | N) \text{Prob}_{3,1}^1 + P(X_i | H, A) \frac{2 \cdot \text{Prob}_{3,2}^1}{3}} \right) \\
 i &= 1, 2, 6, 8, 9, 13, 15, 16, 17 \tag{5.2.18}
 \end{aligned}$$

$$\begin{aligned}
\text{Prob}_{1,3}^2 &= \sum_i w_i \cdot P(H, A | X_i) \\
&= \sum_i w_i \left(\frac{1}{3} \frac{P(X_i | H, A) \text{Prob}_{1,2}^1}{P(X_i | S) \frac{\text{Prob}_{1,1}^1}{2} + P(X_i | N) \frac{\text{Prob}_{1,1}^1}{2} + P(X_i | H, A) \text{Prob}_{1,2}^1} \right. \\
&\quad + \frac{1}{3} \frac{P(X_i | H, A) \frac{2 \cdot \text{Prob}_{2,2}^1}{3}}{P(X_i | S) \text{Prob}_{2,1}^1 + P(X_i | N) \frac{\text{Prob}_{2,2}^1}{3} + P(X_i | H, A) \frac{2 \cdot \text{Prob}_{2,2}^1}{3}} \\
&\quad \left. + \frac{1}{3} \frac{P(X_i | H, A) \frac{2 \cdot \text{Prob}_{3,2}^1}{3}}{P(X_i | S) \frac{\text{Prob}_{3,2}^1}{3} + P(X_i | N) \text{Prob}_{3,1}^1 + P(X_i | H, A) \frac{2 \cdot \text{Prob}_{3,2}^1}{3}} \right) \\
i &= 1, 2, 6, 8, 9, 13, 15, 16, 17 \tag{5.2.19}
\end{aligned}$$

多層貝氏網路語音情緒辨識第二層之分群方式為第一層之延伸，將第一層中分類效果不佳之情緒類別透過第二層所使用之特徵參數再加以分類出來，因此，第二層中各特徵參數之分群機率為第一層中三種不同分群方式對此特徵參數之後分佈機率(Post-probability)加總而得，而多層貝氏網路語音情緒辨識第二層整體分群之機率則為各特徵參數在不同權重下之機率總和。

在式(5.2.17)、式(5.2.18)以及式(5.2.19)，第二層第一種分群方式各群之計算式中， $w_1 + w_2 + w_6 + w_8 + w_9 + w_{13} + w_{15} + w_{16} + w_{17} = 1$ ，由於各特徵參數分類之效果不同，在此，我們針對不同特徵參數使用不同權重， $w_1 = \frac{1}{6}$ ， $w_2 = w_6 = w_8 = w_9 = \frac{1}{8}$ ， $w_{13} = w_{15} = w_{16} = w_{17} = \frac{1}{12}$ ，計算

式中， $\text{Prob}_{1,1}^2$ 表示第二層第一種分群方式中悲傷發生之機率， $\text{Prob}_{1,2}^2$ 表示第二層第一種分群方式，中性發生之機率， $\text{Prob}_{1,3}^2$ 則表示第二層第一種分群方式中，快樂與生氣發生之機率。式(5.2.17)、式(5.2.18)以及式(5.2.19)之前分佈機率值為第一層貝氏網路之分群機率，根據各分群內之情緒類別數目計算而得，當分群內只有一情緒類別時，此類別之前分佈機率為原分群機率，當分群內有兩種情緒類別時，則各情緒類別發生之前機率為分群機率之 1/2，而當分群內有三種情緒類別時，則各情緒類別發生之前機率為分群機率之 1/3。

在式(5.2.17)、式(5.2.18)以及式(5.2.19)中， $P(X_i | S)$ 表示在悲傷為獨立分群分佈下， X_i 發生之機率密度， $P(X_i | N)$ 表示在中性為獨立分群分佈下， X_i 發生之機率密度， $P(X_i | H, A)$ 表示在快樂與生氣為同一分群分佈下， X_i 發生之機率密度。

接著，考慮多層貝氏網路語音情緒辨識第二層第二種分群方式，悲傷與中性為同一群，快樂以及生氣則各為獨立的一群，由於只有一個特徵參數 X_{19} ，因此，第二層第二種分群方式之機率運算分別為

$$\begin{aligned} \text{Prob}_{2,1}^2 &= P(S, N | X_{19}) \\ &= \frac{P(X_{19} | S, N)\text{Prob}_{1,1}^1}{P(X_{19} | S, N)\text{Prob}_{1,1}^1 + P(X_{19} | H)\frac{\text{Prob}_{1,2}^1}{2} + P(X_{19} | A)\frac{\text{Prob}_{1,2}^1}{2}} \end{aligned} \quad (5.2.20)$$

$$\begin{aligned}
\text{Prob}_{2,2}^2 &= P(H | X_{19}) \\
&= \frac{P(X_{19} | H) \frac{\text{Prob}_{1,2}^1}{2}}{P(X_{19} | S, N) \text{Prob}_{1,1}^1 + P(X_{19} | H) \frac{\text{Prob}_{1,2}^1}{2} + P(X_{19} | A) \frac{\text{Prob}_{1,2}^1}{2}}
\end{aligned} \tag{5.2.21}$$

$$\begin{aligned}
\text{Prob}_{2,3}^2 &= P(A | X_{19}) \\
&= \frac{P(X_{19} | A) \frac{\text{Prob}_{1,2}^1}{2}}{P(X_{19} | S, N) \text{Prob}_{1,1}^1 + P(X_{19} | H) \frac{\text{Prob}_{1,2}^1}{2} + P(X_{19} | A) \frac{\text{Prob}_{1,2}^1}{2}}
\end{aligned} \tag{5.2.22}$$

由於第一層第二種分群及第三種分群，已經將悲傷及中性分別獨立出來，因此，式(5.2.20)、(5.2.21)及(5.2.22)之機率運算只單獨運用第一層第一種分群來當作前分佈機率。而 $\text{Prob}_{2,1}^2$ 表示第二層中第二種分群方式中，悲傷與中性發生之機率， $\text{Prob}_{2,2}^2$ 表示第二層中第二種分群方式中，快樂發生之機率， $\text{Prob}_{2,3}^2$ 則表示第二層中第二種分群方式中，生氣發生之機率。 $P(X_i | S, N)$ 表示在悲傷與中性為同一分群分佈下， X_i 發生之機率密度， $P(X_i | H)$ 表示在快樂為獨立分群分佈下， X_i 發生之機率密度， $P(X_i | A)$ 表示在生氣為獨立分群分佈下， X_i 發生之機率密度。

最後，考慮多層貝氏網路語音情緒辨識第三層之特徵分群運算(如圖 5-13 所示)，也就是將所有的情緒類別之個別機率計算出來，在此，悲傷、中性、快樂與生氣皆為獨立的群，其個別情緒類別之機率運算

分別為

$$\begin{aligned}
\text{Prob}_1^3 &= \sum_i w_i \cdot P(S | X_i) \\
&= \sum_i w_i \left(\frac{3}{4} \frac{P(X_i | S) \text{Prob}_{1,1}^2}{P(X_i | S) \text{Prob}_{1,1}^2 + P(X_i | N) \text{Prob}_{1,2}^2 + P(X_i | H) \frac{\text{Prob}_{1,3}^2}{2} + P(X_i | A) \frac{\text{Prob}_{1,3}^2}{2}} \right. \\
&\quad \left. + \frac{1}{4} \frac{P(X_i | S) \frac{\text{Prob}_{2,1}^2}{2}}{P(X_i | S) \frac{\text{Prob}_{2,1}^2}{2} + P(X_i | N) \frac{\text{Prob}_{2,1}^2}{2} + P(X_i | H) \text{Prob}_{2,2}^2 + P(X_i | A) \text{Prob}_{2,3}^2} \right) \\
i &= 3, 10, 11, 12, 14, 21, 22, 23 \tag{5.2.23}
\end{aligned}$$

$$\begin{aligned}
\text{Prob}_2^3 &= \sum_i w_i \cdot P(N | X_i) \\
&= \sum_i w_i \left(\frac{3}{4} \frac{P(X_i | N) \text{Prob}_{1,2}^2}{P(X_i | S) \text{Prob}_{1,1}^2 + P(X_i | N) \text{Prob}_{1,2}^2 + P(X_i | H) \frac{\text{Prob}_{1,3}^2}{2} + P(X_i | A) \frac{\text{Prob}_{1,3}^2}{2}} \right. \\
&\quad \left. + \frac{1}{4} \frac{P(X_i | N) \frac{\text{Prob}_{2,1}^2}{2}}{P(X_i | S) \frac{\text{Prob}_{2,1}^2}{2} + P(X_i | N) \frac{\text{Prob}_{2,1}^2}{2} + P(X_i | H) \text{Prob}_{2,2}^2 + P(X_i | A) \text{Prob}_{2,3}^2} \right) \\
i &= 3, 10, 11, 12, 14, 21, 22, 23 \tag{5.2.24}
\end{aligned}$$

$$\begin{aligned}
\text{Prob}_3^3 &= \sum_i w_i \cdot P(H | X_i) \\
&= \sum_i w_i \left(\frac{3}{4} \frac{P(X_i | H) \frac{\text{Prob}_{1,3}^2}{2}}{P(X_i | S) \text{Prob}_{1,1}^2 + P(X_i | N) \text{Prob}_{1,2}^2 + P(X_i | H) \frac{\text{Prob}_{1,3}^2}{2} + P(X_i | A) \frac{\text{Prob}_{1,3}^2}{2}} \right. \\
&\quad \left. + \frac{1}{4} \frac{P(X_i | H) \text{Prob}_{2,2}^2}{P(X_i | S) \frac{\text{Prob}_{2,1}^2}{2} + P(X_i | N) \frac{\text{Prob}_{2,1}^2}{2} + P(X_i | H) \text{Prob}_{2,2}^2 + P(X_i | A) \text{Prob}_{2,3}^2} \right) \\
i &= 3, 10, 11, 12, 14, 21, 22, 23 \tag{5.2.25}
\end{aligned}$$

$$\begin{aligned}
\text{Prob}_4^3 &= \sum_i w_i \cdot P(A | X_i) \\
&= \sum_i w_i \left(\frac{3}{4} \frac{P(X_i | A) \frac{\text{Prob}_{1,3}^2}{2}}{P(X_i | S) \text{Prob}_{1,1}^2 + P(X_i | N) \text{Prob}_{1,2}^2 + P(X_i | H) \frac{\text{Prob}_{1,3}^2}{2} + P(X_i | A) \frac{\text{Prob}_{1,3}^2}{2}} \right. \\
&\quad \left. + \frac{1}{4} \frac{P(X_i | A) \text{Prob}_{2,3}^2}{P(X_i | S) \frac{\text{Prob}_{2,1}^2}{2} + P(X_i | N) \frac{\text{Prob}_{2,1}^2}{2} + P(X_i | H) \text{Prob}_{2,2}^2 + P(X_i | A) \text{Prob}_{2,3}^2} \right) \\
i &= 3, 10, 11, 12, 14, 21, 22, 23 \tag{5.2.26}
\end{aligned}$$

多層貝氏網路語音情緒辨識第三層之個別情緒類別機率為第二層之延伸，由於第二層第一種分群方式所使用之特徵參數較多，因此，第二層第一種分群方式對於第三層各情緒類別機率應具有較大之權重，而第二層第二種分群方式則應佔較小之權重，在此，我們給定第二層兩種分群方式的權重分別為 3/4 以及 1/4，而最後所得之各情緒類別機率則為第三層各特徵參數在不同權重下之總和。

在式 (5.2.23)、式 (5.2.24)、式 (5.2.25) 以及式 (5.2.26) 中， $w_3 + w_{10} + w_{11} + w_{12} + w_{14} + w_{21} + w_{22} + w_{23} = 1$ ，在本論文中，使用 $w_3 = w_{10} = w_{11} = w_{12} = w_{14} = w_{21} = w_{22} = w_{23} = \frac{1}{8}$ ，而 Prob_1^3 表示第三層中，悲傷發生之機率， Prob_2^3 表示第三層中，中性發生之機率， Prob_3^3 表示第三層中，快樂發生之機率， Prob_4^3 表示第三層中，生氣發生之機率。

本論文所提出之多層貝氏網路語音情緒分類方法係透過決策樹之分層運算方式結合貝氏網路架構，分別計算出各情緒類別發生之機率，由於最後所得到的是各情緒類別可能的機率，因此，在實際運用

上，可以根據使用平台的實際狀況，將計算所得之各情緒發生的機率做最適當的判斷與運用。以下將就多層貝氏網路語音情緒辨識之實驗與測試結果分別加以說明。

5.2.3 多層貝氏網路(MLBN)語音情緒辨識實驗結果與分析

根據本論文於上一節所提出之多層貝氏網路語音情緒分類方法，此方法是先依據資料庫語料計算所得之相關特徵參數，利用統計方式計算出相關特徵參數於分層分群後之平均值與標準差，並以此平均值與標準差建立多層貝氏網路資料庫(MLBN database)，接著，在得到使用者之情緒語音後，將此情緒語音所得之特徵參數與資料庫之資料進行機率運算，分別計算出測試者情緒語音於各情緒類別可能發生之機率，舉例來說，當我們得到一段使用者之情緒語音後，經由多層貝氏網路語音情緒辨識後，假設計算出來之各情緒類別之機率分別為 $\text{Prob}_1^3 = 0.0338$ ， $\text{Prob}_2^3 = 0.0172$ ， $\text{Prob}_3^3 = 0.3648$ ， $\text{Prob}_4^3 = 0.5842$ ，也就是悲傷發生之機率為 3.38%，中性發生之機率為 1.72%，快樂發生之機率為 36.48%，生氣發生之機率為 58.42%，由於生氣發生之機率最高，因此，我們將此語音情緒辨識為生氣，而在實際運用上，多層貝氏網路語音情緒辨識方法可以根據使用平台的實際狀況，將計算所得之各情緒發生的機率做最適當的判斷與運用。本論文後續之多層貝氏網路情緒辨識皆以最高發生機率之情緒為實際情緒辨識結果，以下將

就多層貝氏網路語音情緒辨識於不同資料庫情況作分析與討論。

與第四章 KNN 及 SVM 所使用之語料庫相同，以德國語料所有資料當作訓練資料，也就是悲傷有 62 筆、中性有 79 筆、快樂有 71 筆、生氣有 127 筆訓練資料，並分別計算出各情緒語料之正規化特徵參數(表 3-5 所列)，接著，根據 5.2.1 節之特徵分群方式，分別計算各分層各群之特徵參數平均值與標準差，並以此為多層貝氏網路語音情緒辨識資料庫，最後，以訓練資料之正規化特徵參數當作測試資料，對資料作 Inside Test。表 5-4 為 MLBN 正規化特徵參數 Inside Test 鑑別矩陣，由辨識結果顯示，當測試資料在訓練資料中時，悲傷 62 筆資料有 54 筆資料被辨識為悲傷，8 筆資料被辨識為中性，辨識正確率為 87%，中性 79 筆資料有 77 筆資料被辨識為中性，2 筆資料被辨識為悲傷，辨識正確率 97%，快樂 71 筆資料有 6 筆被辨識為中性，44 筆被辨識為快樂，21 筆被辨識為生氣，辨識正確率 62%，生氣 127 筆資料有 1 筆被辨識為中性，26 筆被辨識為快樂，100 筆被辨識為生氣，辨識正確率 78.7%，整體正確辨識率為 81.1%，從鑑別矩陣來看，悲傷與中性在識別上，比較容易產生混淆，悲傷有一小部分容易被辨識為中性，而快樂則與生氣易產生混淆，尤其是快樂，其大部分之語音特徵與生氣之重疊性相當高，且分佈較廣，以多層貝氏網路語音情緒分類器做分類時，其快樂與生氣所計算出來之機率非常相近，因此，

容易被辨識歸類為生氣，造成辨識度較低，而生氣也因為特徵參數分佈與快樂重疊性高，造成辨識率不佳之情況。

表 5-5 為單純使用貝氏決策(BD)之正規化特徵參數 Inside Test 鑑別矩陣，由辨識結果來看，當測試資料在訓練資料中時，悲傷辨識正確率為 68%，中性辨識正確率 89%，快樂辨識正確率 37%，生氣辨識正確率 80.3%，整體辨識率為 70.8%。從 BD Inside Test 辨識結果與 MLBN Inside Test 辨識結果比較來看，多層貝氏網路語音情緒分類器的辨識效果明顯優於貝氏決策，因此，透過多層貝氏網路分層分群的辨識方式，可以有效提升語音情緒辨識率，增加辨識的準確性。

接著，考慮使用者資料不在訓練資料內的情況，與 KNN 及 SVM Outside Test 相同，首先，將 Person 1(P1)語料當作測試語料，並以 Person 2 ~ Person 10 語料當作訓練語料，分別計算各層各群之統計平均值與標準差，並以此為多層貝氏網路語音情緒資料庫，對 P1 語料進行 Outside Test，接著，以 Person 2 當測試資料，另外 9 人(P1 以及 P3 ~ P10)當作訓練資料，再重新計算各層各群之統計平均值與標準差，並以此新統計計算值為新的多層貝氏網路資料庫，對 P2 進行 Outside Test，以此類推，對 P3 至 P10 做 Outside Test，表 5-6 為 P1 ~ P10 正規化特徵參數 Outside Test 鑑別矩陣(P1 ~ P10 各情緒之辨識正確率詳見附錄 A)，由整體辨識結果顯示，除了 P7 有較佳的辨識效果

外，其餘之辨識率則約在 75% 上下，由 MLBN Outside Test 鑑別矩陣來看，悲傷 62 筆資料有 52 筆資料被辨識為悲傷，10 筆資料被辨識為中性，辨識正確率為 84%，中性 79 筆資料有 77 筆資料被辨識為中性，2 筆資料被辨識為悲傷，辨識正確率 97%，快樂 71 筆資料有 1 筆被辨識為悲傷，10 筆被辨識為中性，38 筆被辨識為快樂，22 筆被辨識為生氣，辨識正確率 54%，生氣 127 筆資料有 1 筆被辨識為中性，30 筆被辨識為快樂，96 筆被辨識為生氣，辨識正確率 75.6%，整體之正確辨識率為 77.6%。MLBN Outside Test 之鑑別矩陣與 MLBN Inside Test 類似，悲傷有一小部分容易被辨識為中性，而快樂則易與生氣產生混淆，兩者之辨識效果均不佳。表 5-7 則為 P1 ~ P10 使用原始特徵參數 Outside Test 鑑別矩陣(P1 ~ P10 各情緒之辨識正確率詳見附錄 A)，與表 5-6 比較可以發現，與 KNN 及 SVM 相同，使用正規化特徵參數可以有效縮小語者之間在特徵參數上的差異，並增加各情緒之間的分類效果。

最後，與 KNN 及 SVM 相同，考慮 MLBN 語音情緒分類器在訓練語料與測試語料為不同語系情況下進行辨識，同樣的，我們以德國語料庫所有資料當作訓練資料，並以工業技術研究院所錄製之語料當作測試語料，表 5-8 為 MLBN 使用正規化特徵參數在不同語系測試下之鑑別矩陣，由辨識結果顯示，MLBN 對於悲傷與中性之辨識效果

類似，悲傷被辨識為中性的機率有 75.8%，中性則有 96.0% 機率被辨識為中性，而快樂則與生氣之辨識效果類似，快樂與生氣皆相對容易被辨識為悲傷，辨識率分別為 42.8% 與 59.3%，此結果表示在以 MLBN 為辨識分類器的情況下，中文語系悲傷與中性之特徵參數與德國語料庫之中性特徵參數相近，而快樂與生氣之特徵參數則與德國語料庫之悲傷特徵參數相近，因此，若以德國語料庫當作訓練資料時，中文語系之悲傷情緒容易被辨識為中性，而快樂與生氣情緒則容易被辨識為悲傷，整體之正確辨識率為 39.3%，此結果與 KNN 及 SVM 分類器相同，MLBN 分類器在訓練語料與測試語料為不同語系時，同樣無法得到良好的分類效果。

5.3 多層共變異數貝氏網路(Multi-Layer Bayesian Network with Covariance, MLBNC) 語音情緒辨識

5.3.1 多層共變異數貝氏網路(MLBNC)

由於多層貝氏網路語音情緒分類方法在相關特徵參數與特徵參數之間，是假設彼此各為獨立之特徵參數，特徵參數之間沒有任何的關連性，因此，各節點只需計算其分群後之平均值與標準差，而本節所介紹之多層共變異數貝氏網路語音情緒辨識則是假設各特徵參數彼此之間是有相關性的，因此，在分群運算上必須考慮相關特徵參數間之共變異數(Covariance)。本論文所提出之語音情緒辨識多層共變

異數貝氏網路(MLBNC)語音情緒辨識，其架構與多層貝氏網路(MLBN)語音情緒辨識相同(圖 5-7)，而計算上與 MLBN 語音情緒辨識最大之不同在於考量特徵參數與特徵參數之間的相關性，因此，在計算各層各群發生之機率時，是以多變數常態分佈的方式去計算，而多層共變異數貝氏網路資料庫則是以訓練語料庫於各層各群的特徵參數向量與共變異數來建置，多層共變異數貝氏網路為多層貝氏網路之延伸，因此多層共變異數貝氏網路(MLBNC)與多層貝氏網路(MLBN)採用相同之分群方式，第一層共有三種不同分群方式，第一種為悲傷與中性為同一群，快樂與生氣則為另一群，第二種為悲傷為獨立的一群，中性、快樂以及生氣為另一群，第三種則為中性為獨立的一群，悲傷、快樂以及生氣為另一群。多層共變異數貝氏網路第二層則有兩種不同分群方式，第一種為悲傷以及中性各為獨立的一群，快樂與生氣為第三群，第二種則為悲傷與中性為一群，快樂以及生氣則各為獨立的一群，最後，多層共變異數貝氏網路第三層則是悲傷、中性、快樂與生氣皆為獨立的一群，各情緒類別之發生機率將會在此被計算出來，MLBNC 之分層分別如圖 5-15、圖 5-16 以及圖 5-17 所示，。

一般而言，多變數常態分佈之機率密度計算如式(5.3.1)所示

$$f_{\mathbf{z}}(Z_1, \dots, Z_N) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{Z} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu})\right] \quad (5.3.1)$$

其中， \mathbf{Z} 為參數特徵向量， $\boldsymbol{\mu}$ 為多變數常態分佈之平均值向量， Σ 則為多變數常態分佈之共變異數，以單純只有 MFCC C1 平均值及 MFCC C4 平均值兩變數為例，四種情緒類別之落點分佈如圖 5-18 所示，從圖中可以清楚看出，悲傷與中性在此兩特徵參數下較具鑑別性，而快樂與生氣則在辨識上容易產生混淆，圖 5-19 與圖 5-20 則為 MFCC C1 與 C4 平均值雙變數四種情緒之常態分佈圖，由其機率密度分佈情況可以發現，主要分為三個區塊，除了中性情緒外，其餘三個情緒類別之 MFCC C1 及 C4 平均值兩特徵參數均具有相關性。

多層共變異數貝氏網路語音情緒之辨識流程與多層貝氏網路語音情緒辨識流程相似，如圖 5-21 所示，首先，測試語者先講一段正規化用中性情緒語音，並利用此正規化用之中性情緒語音，計算出測試語者之正規化特徵參數因子，接著，進入語音情緒辨識流程，在得到使用者一段情緒語音訊號後，先計算出各層各群所需之原始特徵參數向量，並經由正規化特徵參數因子對原始特徵參數向量做正規化，最後，將正規化後特徵參數向量與多層共變異數貝氏網路資料庫做分層分群運算，計算出每一情緒類別之發生機率。多層共變異數貝氏網路語音情緒分類器與多層貝氏網路語音情緒分類器最大的不同在於資料庫的建置與機率運算方式，多層貝氏網路之資料庫建置僅需計算各特徵參數之平均值與標準差，並透過單一變數之常態分佈機率密度

計算各情緒發生之機率，而多層共變異數貝氏網路之資料庫則必須計算各群之特徵參數平均值向量與共變異數，最後，再透過多變數常態分佈機率密度運算將各情緒發生之機率計算出來。以下將就多層共變異數貝氏網路語音情緒辨識各分層分群之機率計算分別加以說明。

首先，考慮多層共變異數貝氏網路語音情緒辨識第一層(如圖 5-15 所示)，主要有三種分群方式，第一種為悲傷與中性為同一群，快樂與生氣則為另一群，假設第一層第一種分群方式使用之兩特徵參數向量為 \mathbf{R}_1 ，其中 $\mathbf{R}_1 = [X_{18} \quad X_{20}]^T$ ，兩分群之多變數機率運算分別為

$$\begin{aligned} \text{Probc}_{1,1}^1 &= P(S, N | \mathbf{R}_1) \\ &= \frac{P(\mathbf{R}_1 | S, N)P(S, N)}{P(\mathbf{R}_1 | S, N)P(S, N) + P(\mathbf{R}_1 | H, A)P(H, A)} \end{aligned} \quad (5.3.2)$$

$$\begin{aligned} \text{Probc}_{1,2}^1 &= P(H, A | \mathbf{R}_1) \\ &= \frac{P(\mathbf{R}_1 | H, A)P(H, A)}{P(\mathbf{R}_1 | S, N)P(S, N) + P(\mathbf{R}_1 | H, A)P(H, A)} \end{aligned} \quad (5.3.3)$$

其中， $\text{Probc}_{1,1}^1$ 表示多層共變異數貝氏網路語音情緒辨識第一層第一種分群方式中，悲傷與中性發生之機率， $\text{Probc}_{1,2}^1$ 則表示第一層第一種分群方式中，快樂與生氣發生之機率。式(5.3.2)及(5.3.3)中之 $P(S, N)$ 及 $P(H, A)$ 分別表示悲傷與中性以及快樂與生氣之前分佈機率，與多層貝氏網路語音情緒辨識之前分佈機率相同，在此，我們假設各情緒發生之前機率均相同，因此， $P(S, N) = \frac{1}{2}$ ， $P(H, A) = \frac{1}{2}$ ，而 $P(\mathbf{R}_1 | S, N)$ 及 $P(\mathbf{R}_1 | H, A)$ 則分別表示在多變數常態分布下，各群

在特徵參數向量為 \mathbf{R}_1 之機率密度，其算式分別如式(5.3.4)及(5.3.5)所示。

$$P(\mathbf{R}_1|S, N) = \frac{1}{(2\pi)^{|\Sigma_{1,1}^1|} \frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{R}_1 - \boldsymbol{\mu}_{1,1}^1)^T (\Sigma_{1,1}^1)^{-1} (\mathbf{R}_1 - \boldsymbol{\mu}_{1,1}^1)\right] \quad (5.3.4)$$

$$P(\mathbf{R}_1|H, A) = \frac{1}{(2\pi)^{|\Sigma_{1,2}^1|} \frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{R}_1 - \boldsymbol{\mu}_{1,2}^1)^T (\Sigma_{1,2}^1)^{-1} (\mathbf{R}_1 - \boldsymbol{\mu}_{1,2}^1)\right] \quad (5.3.5)$$

其中， $\Sigma_{1,1}^1$ 表示多層共變異數貝氏網路資料庫第一層第一種分群方式中，第一群悲傷與中性之共變異數， $\Sigma_{1,2}^1$ 則表示多層共變異數貝氏網路資料庫第一層第一種分群方式中，第二群快樂與生氣之共變異數，而 $\boldsymbol{\mu}_{1,1}^1$ 及 $\boldsymbol{\mu}_{1,2}^1$ 則分別表示多層共變異數貝氏網路資料庫中，第一層第一種分群方式中第一群及第二群之特徵參數平均值向量。

接著，考慮多層共變異數貝氏網路語音情緒辨識第一層第二種分群方式，悲傷為獨立的一群，中性、快樂以及生氣則為另一群，假設第一層第二種分群方式所使用之兩特徵參數向量為 \mathbf{R}_2 ，其中

$\mathbf{R}_2 = [X_4 \quad X_7]^T$ ，兩分群之多變數機率運算分別為

$$\begin{aligned} \text{Probc}_{2,1}^1 &= P(S | \mathbf{R}_2) \\ &= \frac{P(\mathbf{R}_2 | S)P(S)}{P(\mathbf{R}_2 | S)P(S) + P(\mathbf{R}_2 | N, H, A)P(N, H, A)} \end{aligned} \quad (5.3.6)$$

$$\begin{aligned} \text{Probc}_{2,2}^1 &= P(N, H, A | \mathbf{R}_2) \\ &= \frac{P(\mathbf{R}_2 | N, H, A)P(N, H, A)}{P(\mathbf{R}_2 | S)P(S) + P(\mathbf{R}_2 | N, H, A)P(N, H, A)} \end{aligned} \quad (5.3.7)$$

其中， $\text{Prob}c_{2,1}^1$ 表示多層共變異數貝氏網路語音情緒辨識第一層第二種分群方式中，悲傷發生之機率， $\text{Prob}c_{2,2}^1$ 則表示第一層第二種分群方式中，中性、快樂與生氣發生之機率。式(5.3.6)及(5.3.7)中之 $P(S)$ 及 $P(N, H, A)$ 分別表示悲傷以及中性、快樂與生氣之前分佈機率，因此， $P(S) = \frac{1}{4}$ ， $P(N, H, A) = \frac{3}{4}$ ，而 $P(\mathbf{R}_2 | S)$ 及 $P(\mathbf{R}_2 | N, H, A)$ 則分別表示在多變數常態分布下，各群在特徵參數向量為 \mathbf{R}_2 之機率密度，其算式分別如式(5.3.8)及(5.3.9)所示。

$$P(\mathbf{R}_2 | S) = \frac{1}{(2\pi)^{|\Sigma_{2,1}^1|} |\Sigma_{2,1}^1|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{R}_2 - \boldsymbol{\mu}_{2,1}^1)^T (\Sigma_{2,1}^1)^{-1} (\mathbf{R}_2 - \boldsymbol{\mu}_{2,1}^1)\right] \quad (5.3.8)$$

$$P(\mathbf{R}_2 | N, H, A) = \frac{1}{(2\pi)^{|\Sigma_{2,2}^1|} |\Sigma_{2,2}^1|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{R}_2 - \boldsymbol{\mu}_{2,2}^1)^T (\Sigma_{2,2}^1)^{-1} (\mathbf{R}_2 - \boldsymbol{\mu}_{2,2}^1)\right] \quad (5.3.9)$$

其中， $\Sigma_{2,1}^1$ 表示多層共變異數貝氏網路資料庫第一層第二種分群方式中，第一群悲傷之共變異數， $\Sigma_{2,2}^1$ 則表示多層共變異數貝氏網路資料庫第一層第二種分群方式中，第二群中性、快樂與生氣之共變異數， $\boldsymbol{\mu}_{2,1}^1$ 及 $\boldsymbol{\mu}_{2,2}^1$ 則分別表示多層共變異數貝氏網路資料庫中，第一層第二種分群方式中第一群及第二群之特徵參數平均值向量。

最後，考慮多層共變異數貝氏網路語音情緒辨識第一層第三種分群方式，由於只有單一特徵參數，因此，多層共變異數貝氏網路語音情緒辨識第一層第三種分群方式之各群機率運算與多層貝氏網路語

音情緒辨識第一層第三種分群方式相同，其分群之機率運算為

$$\begin{aligned} \text{Probc}_{3,1}^1 &= P(N | X_5) \\ &= \frac{P(X_5 | N)P(N)}{P(X_5 | N)P(N) + P(X_5 | S, H, A)P(S, H, A)} \end{aligned} \quad (5.3.10)$$

$$\begin{aligned} \text{Probc}_{3,2}^1 &= P(S, H, A | X_5) \\ &= \frac{P(X_5 | S, H, A)P(S, H, A)}{P(X_5 | N)P(N) + P(X_5 | S, H, A)P(S, H, A)} \end{aligned} \quad (5.3.11)$$

其中， $\text{Probc}_{3,1}^1$ 為多層貝氏網路語音情緒辨識第一層第三種分群方式中性發生之機率， $\text{Probc}_{3,2}^1$ 則為第一層第三種分群方式中，悲傷、快樂以及生氣發生之機率，而 $P(X_5 | N)$ 表示在中性為獨立分群分佈下， X_5 發生之機率密度， $P(X_5 | S, H, A)$ 表示在悲傷、快樂與生氣為同一群分佈下， X_5 發生之機率密度， $P(N)$ 代表中性本身之事前機率， $P(S, H, A)$ 則為悲傷、快樂與生氣為同一群之事前機率，因此， $P(N) = \frac{1}{4}$ ， $P(S, H, A) = \frac{3}{4}$ 。

其次，考慮多層共變異數貝氏網路語音情緒辨識第二層(如圖 5-16 所示)，主要有兩種分群方式，第一種為悲傷與中性各為獨立的群，快樂與生氣則為另一群，假設第一種分群所使用之特徵參數向量為 \mathbf{U} ，其中 $\mathbf{U} = [X_1 \ X_2 \ X_6 \ X_8 \ X_9 \ X_{13} \ X_{15} \ X_{16} \ X_{17}]^T$ ，因此，各群之多變數機率運算分別為

$$\begin{aligned}
\text{Probc}_{1,1}^2 &= P(S | U) \\
&= \left(\frac{1}{3} \frac{P(U|S) \frac{\text{Probc}_{1,1}^1}{2}}{P(U|S) \frac{\text{Probc}_{1,1}^1}{2} + P(U|N) \frac{\text{Probc}_{1,1}^1}{2} + P(U|H,A) \text{Probc}_{1,2}^1} + \right. \\
&\quad \frac{1}{3} \frac{P(U|S) \text{Probc}_{2,1}^1}{P(U|S) \text{Probc}_{2,1}^1 + P(U|N) \frac{\text{Probc}_{2,2}^1}{3} + P(U|H,A) \frac{2 \cdot \text{Probc}_{2,2}^1}{3}} \\
&\quad \left. + \frac{1}{3} \frac{P(U|S) \frac{\text{Probc}_{3,2}^1}{3}}{P(U|S) \frac{\text{Probc}_{3,2}^1}{3} + P(U|N) \text{Probc}_{3,1}^1 + P(U|H,A) \frac{2 \cdot \text{Probc}_{3,2}^1}{3}} \right) \\
i=1, 2, 6, 8, 9, 13, 15, 16, 17 & \tag{5.3.12}
\end{aligned}$$

$$\begin{aligned}
\text{Probc}_{1,2}^2 &= P(N | U) \\
&= \left(\frac{1}{3} \frac{P(U|N) \frac{\text{Probc}_{1,1}^1}{2}}{P(U|S) \frac{\text{Probc}_{1,1}^1}{2} + P(U|N) \frac{\text{Probc}_{1,1}^1}{2} + P(U|H,A) \text{Probc}_{1,2}^1} + \right. \\
&\quad \frac{1}{3} \frac{P(U|N) \frac{\text{Probc}_{2,2}^1}{3}}{P(U|S) \text{Probc}_{2,1}^1 + P(U|N) \frac{\text{Probc}_{2,2}^1}{3} + P(U|H,A) \frac{2 \cdot \text{Probc}_{2,2}^1}{3}} \\
&\quad \left. + \frac{1}{3} \frac{P(U|N) \text{Probc}_{3,1}^1}{P(U|S) \frac{\text{Probc}_{3,2}^1}{3} + P(U|N) \text{Probc}_{3,1}^1 + P(U|H,A) \frac{2 \cdot \text{Probc}_{3,2}^1}{3}} \right) \\
i=1, 2, 6, 8, 9, 13, 15, 16, 17 & \tag{5.3.13}
\end{aligned}$$

$$\begin{aligned}
\text{Probc}_{1,3}^2 &= P(H, A | U) \\
&= \left(\frac{1}{3} \frac{P(U | H, A) \text{Probc}_{1,2}^1}{P(U | S) \frac{\text{Probc}_{1,1}^1}{2} + P(U | N) \frac{\text{Probc}_{1,1}^1}{2} + P(U | H, A) \text{Probc}_{1,2}^1} + \right. \\
&\quad \frac{1}{3} \frac{P(U | H, A) \frac{2 \cdot \text{Probc}_{2,2}^1}{3}}{P(U | S) \text{Probc}_{2,1}^1 + P(U | N) \frac{\text{Probc}_{2,2}^1}{3} + P(U | H, A) \frac{2 \cdot \text{Probc}_{2,2}^1}{3}} \\
&\quad \left. + \frac{1}{3} \frac{P(U | H, A) \frac{2 \cdot \text{Probc}_{3,2}^1}{3}}{P(U | S) \frac{\text{Probc}_{3,2}^1}{3} + P(U | N) \text{Probc}_{3,1}^1 + P(U | H, A) \frac{2 \cdot \text{Probc}_{3,2}^1}{3}} \right) \\
i &= 1, 2, 6, 8, 9, 13, 15, 16, 17 \tag{5.3.14}
\end{aligned}$$

與多層貝氏網路語音情緒辨識相同，多層共變異數貝氏網路語音情緒辨識第二層之分群方式為第一層之延伸，因此，第二層第一種分群方式之各群機率為第一層中三種不同分群方式對此特徵參數向量之後分佈機率(Post-probability)加總而得。式(5.3.12)、(5.3.13)及(5.3.14)中， $\text{Probc}_{1,1}^2$ 表示多層共變異數貝氏網路第二層第一種分群方式中，悲傷發生之機率， $\text{Probc}_{1,2}^2$ 表示第二層第一種分群方式，中性發生之機率， $\text{Probc}_{1,3}^2$ 則表示第二層第一種分群方式，快樂與生氣發生之機率， $\text{Probc}_{1,1}^1$ 、 $\text{Probc}_{1,2}^1$ 、 $\text{Probc}_{2,1}^1$ 、 $\text{Probc}_{2,2}^1$ 、 $\text{Probc}_{3,1}^1$ 及 $\text{Probc}_{3,2}^1$ 則分別表示前分佈機率，由第一層之各分群方式中計算求得，而 $P(U | S)$ 、 $P(U | N)$ 及 $P(U | H, A)$ 則分別表示在多變數常態分布下，特徵參數向量 U 在各分群之機率密度，其算式分別如式(5.3.15)、(5.3.16)及(5.3.17)

所示。

$$P(\mathbf{U}|S) = \frac{1}{(2\pi)^{9/2} |\Sigma_{1,1}^2|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{U} - \boldsymbol{\mu}_{1,1}^2)^T (\Sigma_{1,1}^2)^{-1} (\mathbf{U} - \boldsymbol{\mu}_{1,1}^2)\right] \quad (5.3.15)$$

$$P(\mathbf{U}|N) = \frac{1}{(2\pi)^{9/2} |\Sigma_{1,2}^2|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{U} - \boldsymbol{\mu}_{1,2}^2)^T (\Sigma_{1,2}^2)^{-1} (\mathbf{U} - \boldsymbol{\mu}_{1,2}^2)\right] \quad (5.3.16)$$

$$P(\mathbf{U}|H, A) = \frac{1}{(2\pi)^{9/2} |\Sigma_{1,3}^2|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{U} - \boldsymbol{\mu}_{1,3}^2)^T (\Sigma_{1,3}^2)^{-1} (\mathbf{U} - \boldsymbol{\mu}_{1,3}^2)\right] \quad (5.3.17)$$

其中， $\Sigma_{1,1}^2$ 表示多層共變異數貝氏網路資料庫第二層第一種分群方式中，第一群悲傷之共變異數， $\Sigma_{1,2}^2$ 表示多層共變異數貝氏網路資料庫第二層第一種分群方式中，第二群中性之共變異數， $\Sigma_{1,3}^2$ 則表示多層共變異數貝氏網路資料庫第二層第一種分群方式中，第三群快樂與生氣之共變異數，而 $\boldsymbol{\mu}_{1,1}^2$ 、 $\boldsymbol{\mu}_{1,2}^2$ 及 $\boldsymbol{\mu}_{1,3}^2$ 則分別表示多層共變異數貝氏網路資料庫中，第二層第一種分群方式中，第一群、第二群及第三群之特徵參數平均值向量。

接著，考慮多層共變異數貝氏網路語音情緒辨識第二層第二種分群方式，悲傷與中性為同一群，快樂以及生氣則各為獨立的一群，由於只有一個特徵參數 X_{19} ，因此多層共變異數貝氏網路語音情緒辨識第二層第二種分群方式之機率運算分別為

$$\begin{aligned}
\text{Probc}_{2,1}^2 &= P(S, N | X_{19}) \\
&= \frac{P(X_{19} | S, N)\text{Probc}_{1,1}^1}{P(X_{19} | S, N)\text{Probc}_{1,1}^1 + P(X_{19} | H)\frac{\text{Probc}_{1,2}^1}{2} + P(X_{19} | A)\frac{\text{Probc}_{1,2}^1}{2}}
\end{aligned}
\tag{5.3.18}$$

$$\begin{aligned}
\text{Probc}_{2,2}^2 &= P(H | X_{19}) \\
&= \frac{P(X_{19} | H)\frac{\text{Probc}_{1,2}^1}{2}}{P(X_{19} | S, N)\text{Probc}_{1,1}^1 + P(X_{19} | H)\frac{\text{Probc}_{1,2}^1}{2} + P(X_{19} | A)\frac{\text{Probc}_{1,2}^1}{2}}
\end{aligned}
\tag{5.3.19}$$

$$\begin{aligned}
\text{Probc}_{2,3}^2 &= P(A | X_{19}) \\
&= \frac{P(X_{19} | A)\frac{\text{Probc}_{1,2}^1}{2}}{P(X_{19} | S, N)\text{Probc}_{1,1}^1 + P(X_{19} | H)\frac{\text{Probc}_{1,2}^1}{2} + P(X_{19} | A)\frac{\text{Probc}_{1,2}^1}{2}}
\end{aligned}
\tag{5.3.20}$$

其中， $\text{Probc}_{2,1}^2$ 表示多層共變異數貝氏網路第二層第二種分群方式中，悲傷與中性發生之機率， $\text{Probc}_{2,2}^2$ 表示第二層第二種分群方式中，快樂發生之機率， $\text{Probc}_{2,3}^2$ 則表示多層共變異數貝氏網路第二層中第二種分群方式中生氣發生之機率。

最後，考慮多層共變異數貝氏網路語音情緒辨識第三層(如圖 5-17 所示)，悲傷、中性、快樂與生氣皆為獨立的分群，因此，所有情緒類別之個別發生機率將被計算出來，在此，假設多層共變異數貝氏網路第三層所使用之特徵參數組成一新特徵參數向量 \mathbf{V} ，其中，

$$\mathbf{V} = [X_3 \quad X_{10} \quad X_{11} \quad X_{12} \quad X_{14} \quad X_{21} \quad X_{22} \quad X_{23}]^T, \text{ 而多層共變異數貝}$$

氏網路第三層各情緒類別之機率運算則分別為

$$\begin{aligned}
 \text{Probc}_1^3 &= P(S | \mathbf{V}) \\
 &= \frac{3}{4} \frac{P(\mathbf{V} | S)\text{Probc}_{1,1}^2}{P(\mathbf{V} | S)\text{Probc}_{1,1}^2 + P(\mathbf{V} | N)\text{Probc}_{1,2}^2 + P(\mathbf{V} | H)\frac{\text{Probc}_{1,3}^2}{2} + P(\mathbf{V} | A)\frac{\text{Probc}_{1,3}^2}{2}} \\
 &\quad + \frac{1}{4} \frac{P(\mathbf{V} | S)\frac{\text{Probc}_{2,1}^2}{2}}{P(\mathbf{V} | S)\frac{\text{Probc}_{2,1}^2}{2} + P(\mathbf{V} | N)\frac{\text{Probc}_{2,1}^2}{2} + P(\mathbf{V} | H)\text{Probc}_{2,2}^2 + P(\mathbf{V} | A)\text{Probc}_{2,3}^2} \\
 &\hspace{15em} (5.3.21)
 \end{aligned}$$

$$\begin{aligned}
 \text{Probc}_2^3 &= P(N | \mathbf{V}) \\
 &= \frac{3}{4} \frac{P(\mathbf{V} | N)\text{Probc}_{1,2}^2}{P(\mathbf{V} | S)\text{Probc}_{1,1}^2 + P(\mathbf{V} | N)\text{Probc}_{1,2}^2 + P(\mathbf{V} | H)\frac{\text{Probc}_{1,3}^2}{2} + P(\mathbf{V} | A)\frac{\text{Probc}_{1,3}^2}{2}} \\
 &\quad + \frac{1}{4} \frac{P(\mathbf{V} | N)\frac{\text{Probc}_{2,1}^2}{2}}{P(\mathbf{V} | S)\frac{\text{Probc}_{2,1}^2}{2} + P(\mathbf{V} | N)\frac{\text{Probc}_{2,1}^2}{2} + P(\mathbf{V} | H)\text{Probc}_{2,2}^2 + P(\mathbf{V} | A)\text{Probc}_{2,3}^2} \\
 &\hspace{15em} (5.3.22)
 \end{aligned}$$

$$\begin{aligned}
 \text{Probc}_3^3 &= P(H | \mathbf{V}) \\
 &= \frac{3}{4} \frac{P(\mathbf{V} | H)\frac{\text{Probc}_{1,3}^2}{2}}{P(\mathbf{V} | S)\text{Probc}_{1,1}^2 + P(\mathbf{V} | N)\text{Probc}_{1,2}^2 + P(\mathbf{V} | H)\frac{\text{Probc}_{1,3}^2}{2} + P(\mathbf{V} | A)\frac{\text{Probc}_{1,3}^2}{2}} \\
 &\quad + \frac{1}{4} \frac{P(\mathbf{V} | H)\text{Probc}_{2,2}^2}{P(\mathbf{V} | S)\frac{\text{Probc}_{2,1}^2}{2} + P(\mathbf{V} | N)\frac{\text{Probc}_{2,1}^2}{2} + P(\mathbf{V} | H)\text{Probc}_{2,2}^2 + P(\mathbf{V} | A)\text{Probc}_{2,3}^2} \\
 &\hspace{15em} (5.3.23)
 \end{aligned}$$

$$\begin{aligned}
\text{Probc}_4^3 &= P(A | \mathbf{V}) \\
&= \frac{3}{4} \frac{P(\mathbf{V} | A) \frac{\text{Probc}_{1,3}^2}{2}}{P(\mathbf{V} | S) \text{Probc}_{1,1}^2 + P(\mathbf{V} | N) \text{Probc}_{1,2}^2 + P(\mathbf{V} | H) \frac{\text{Probc}_{1,3}^2}{2} + P(\mathbf{V} | A) \frac{\text{Probc}_{1,3}^2}{2}} \\
&+ \frac{1}{4} \frac{P(\mathbf{V} | A) \text{Probc}_{2,3}^2}{P(\mathbf{V} | S) \frac{\text{Probc}_{2,1}^2}{2} + P(\mathbf{V} | N) \frac{\text{Probc}_{2,1}^2}{2} + P(\mathbf{V} | H) \text{Probc}_{2,2}^2 + P(\mathbf{V} | A) \text{Probc}_{2,3}^2}
\end{aligned} \tag{5.3.24}$$

多層共變異數貝氏網路語音情緒辨識第三層所使用之權重與多層貝氏網路語音情緒辨識相同，分別為 3/4 以及 1/4，而 Probc_1^3 表示多層共變異數貝氏網路第三層悲傷發生之機率， Probc_2^3 表示中性發生之機率， Probc_3^3 表示快樂發生之機率， Probc_4^3 則表示生氣發生之機率。式(5.3.21)、(5.3.22)、(5.3.23)以及(5.3.24)中之 $P(\mathbf{V} | S)$ 、 $P(\mathbf{V} | N)$ 、 $P(\mathbf{V} | H)$ 及 $P(\mathbf{V} | A)$ 則分別表示在多變數常態分布下，特徵參數向量 \mathbf{V} 在各情緒類別之機率密度，其算式分別如式(5.3.25)、(5.3.26)、(5.3.27)以及(5.3.28)所示。

$$P(\mathbf{V} | S) = \frac{1}{(2\pi)^4 |\Sigma_1^3|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{V} - \boldsymbol{\mu}_1^3)^T (\Sigma_1^3)^{-1} (\mathbf{V} - \boldsymbol{\mu}_1^3) \right] \tag{5.3.25}$$

$$P(\mathbf{V} | N) = \frac{1}{(2\pi)^4 |\Sigma_2^3|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{V} - \boldsymbol{\mu}_2^3)^T (\Sigma_2^3)^{-1} (\mathbf{V} - \boldsymbol{\mu}_2^3) \right] \tag{5.3.26}$$

$$P(\mathbf{V} | H) = \frac{1}{(2\pi)^4 |\Sigma_3^3|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{V} - \boldsymbol{\mu}_3^3)^T (\Sigma_3^3)^{-1} (\mathbf{V} - \boldsymbol{\mu}_3^3) \right] \tag{5.3.27}$$

$$P(\mathbf{V} | A) = \frac{1}{(2\pi)^4 |\Sigma_4^3|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{V} - \boldsymbol{\mu}_4^3)^T (\Sigma_4^3)^{-1} (\mathbf{V} - \boldsymbol{\mu}_4^3) \right] \tag{5.3.28}$$

其中， Σ_1^3 、 Σ_2^3 、 Σ_3^3 及 Σ_4^3 分別表示多層共變異數貝氏網路資料庫

中第三層悲傷、中性、快樂以及生氣之共變異數，而 μ_1^3 、 μ_2^3 、 μ_3^3 以及 μ_4^3 則分別表示多層共變異數貝氏網路資料庫中第三層悲傷、中性、快樂以及生氣之特徵參數平均值向量。

本節所提出之多層共變異數貝氏網路語音情緒辨識與多層貝氏網路語音情緒辨識相同，係利用決策樹之分層運算方式，並考慮特徵參數間相關性之貝氏網路架構，分別計算出各情緒類別發生之機率，最後所得到的是各情緒類別可能的機率。

5.3.2 MLBNC 實驗結果與分析

根據上一節所提出之多層共變異數貝氏網路語音情緒分類方法，此方法是先依據資料庫語料計算所得之相關特徵參數，考慮相關特徵參數間之相關性，並利用統計方式計算出相關特徵參數於分層分群後之平均值向量與共變異數，並以各分群之平均值向量與共變異數建立多層共變異數貝氏網路資料庫(MLBNC database)，接著，在得到使用者之情緒語音後，將此情緒語音所得之各分層特徵參數向量與資料庫進行機率運算，分別計算出測試者情緒語音於各情緒類別可能發生之機率，在此，與 MLBN 情緒辨識相同，MLBNC 語音情緒辨識以最高發生機率之情緒做為實際情緒辨識結果，以下將就 MLBNC 語音情緒辨識於不同資料庫情況進行實驗分析與討論。

與 KNN、SVM 以及 MLBN 所使用之語料庫相同，以德國語料所

有資料當作訓練資料，悲傷有 62 筆、中性有 79 筆、快樂有 71 筆、生氣有 127 筆訓練資料，根據上一節所介紹之多層共變異數貝氏網路語音情緒辨識，將德國語料所有資料分別計算各分層各群之特徵參數平均值向量($\mu_{1,1}^1$ 、 $\mu_{1,2}^1$ 、 $\mu_{2,1}^1$ 、 $\mu_{2,2}^1$ 、 $\mu_{1,1}^2$ 、 $\mu_{1,2}^2$ 、 $\mu_{1,3}^2$ 、 μ_1^3 、 μ_2^3 、 μ_3^3 以及 μ_4^3)與共變異數($\Sigma_{1,1}^1$ 、 $\Sigma_{1,2}^1$ 、 $\Sigma_{2,1}^1$ 、 $\Sigma_{2,2}^1$ 、 $\Sigma_{1,1}^2$ 、 $\Sigma_{1,2}^2$ 、 $\Sigma_{1,3}^2$ 、 Σ_1^3 、 Σ_2^3 、 Σ_3^3 以及 Σ_4^3)，並以此建置多層共變異數貝氏網路資料庫，接著，以訓練資料之正規化特徵參數當作測試資料，對資料作 Inside Test。表 5-9 為使用正規化特徵參數 MLBNC Inside Test 鑑別矩陣，由辨識結果顯示，當測試資料在訓練資料中時，悲傷 62 筆資料有 60 筆資料被辨識為悲傷，2 筆資料被辨識為中性，辨識正確率為 97%，中性 79 筆資料有 79 筆資料被辨識為中性，辨識正確率 100%，快樂 71 筆資料有 2 筆被辨識為中性，51 筆被辨識為快樂，18 筆被辨識為生氣，辨識正確率 72%，生氣 127 筆資料有 1 筆被辨識為中性，15 筆被辨識為快樂，111 筆被辨識為生氣，辨識正確率 87.4%，整體正確辨識率為 88.8%，從辨識之鑑別矩陣來看，悲傷與中性皆有不錯的辨識效果，而快樂則與生氣易產生混淆，尤其是快樂，有 25% 被辨識為生氣，造成快樂之正確辨識率相對較低，而生氣也因為特徵參數分佈與快樂重疊性高，造成有一小部分被辨識為快樂。

接著，考慮使用者資料不在訓練資料內的情況，與 KNN、SVM

及 MLBN Outside Test 相同，首先，將 Person 1(P1)語料當作測試語料，並以 Person 2 ~ Person 10 語料當作訓練語料，分別計算各層各群之平均值向量與共變異數，並以此建置多層共變異數貝氏網路資料庫，並對 P1 語料進行 Outside Test，接著，以 Person 2 當測試資料，另外 9 人(P1 以及 P3 ~ P10)當作訓練資料，再重新計算各層各群之平均值向量與共變異數，並以此平均值向量與共變異數為新的多層貝氏網路資料庫，並對 P2 進行 Outside Test，以此類推，對 P3 至 P10 做 Outside Test，表 5-10 為 P1 ~ P10 正規化特徵參數 MLBNC Outside Test 辨識正確率與鑑別矩陣(P1 ~ P10 各情緒之辨識正確率詳見附錄 A)，由整體辨識結果顯示，除了 P2 與 P9 外，其餘之辨識率則約在 85% 上下，由 MLBNC Outside Test 鑑別矩陣來看，悲傷 62 筆資料有 51 筆資料被辨識為悲傷，10 筆資料被辨識為中性，1 筆資料被辨識為快樂，辨識正確率為 82%，中性 79 筆資料有 79 筆資料被辨識為中性，辨識正確率 100%，快樂 71 筆資料有 2 筆被辨識為中性，44 筆被辨識為快樂，25 筆被辨識為生氣，辨識正確率 62%，生氣 127 筆資料有 18 筆被辨識為快樂，109 筆被辨識為生氣，辨識正確率 85.8%，整體之正確辨識率為 83.5%。MLBNC Outside Test 之鑑別矩陣與 MLBN Outside Test 類似，悲傷與生氣分別有一小部分容易被辨識為中性及快樂，而快樂則有一部份容易被辨識為生氣，就整體辨識正確率來

看，MLBNC 之辨識效果優於 MLBN，顯示考慮特徵參數間之相關性對於辨識分類有相當程度的助益。

最後，考慮 MLBNC 語音情緒分類器在訓練語料與測試語料為不同語系情況下進行辨識，與 KNN、SVM 及 MLBN 作法相同，我們以德國語料庫所有資料當作訓練資料，並以工業技術研究院所錄製之語料當作測試語料，表 5-11 為 MLBNC 使用正規化特徵參數在不同語系測試下之鑑別矩陣，由辨識結果顯示，MLBNC 對於悲傷與中性之辨識效果類似，悲傷被辨識為中性的機率有 71%，只有 18% 之悲傷情緒被辨識正確為悲傷，中性則有 88% 機率被辨識正確為中性，而快樂則與生氣之辨識效果類似，快樂被辨識為生氣的機率有 53%，有 40% 之快樂情緒被辨識正確為快樂，生氣則有 62.5% 機率被辨識正確為生氣，15.5% 被辨識正確為中性，18.3% 被辨識正確為快樂，此結果與 SVM 測試結果相近，中文語系悲傷與中性之特徵參數與德國語料庫之中性特徵參數較為接近，而快樂與生氣之特徵參數則與德國語料庫之生氣情緒較為接近，因此，若以德國語料庫當作訓練資料時，中文語系之悲傷情緒容易被辨識為中性，而快樂與生氣情緒則容易被辨識為生氣，整體之正確辨識率為 52.1%，此結果顯示，MLBNC 分類器在訓練語料與測試語料為不同語系時，辨識效果較 KNN、SVM 以及 MLBN 分類器為佳，但整體辨識效果也僅在五成左右。

5.4 蒙地卡羅模擬(Monte Carlo Simulation)與分析

蒙地卡羅模擬法(Monte Carlo Simulation)為一種數值方法，藉由統計方式在設定條件下利用亂數取樣(Random sampling)，來模擬實際可能發生的量測值，因此，蒙地卡羅模擬法，是基於大量數值的驗證方法，當實驗次數越多，其平均值將會越趨近理論值，在此，我們將利用蒙地卡羅模擬法來驗證各分類器之辨識可靠度。

在蒙地卡羅隨機資料產生的過程中，使用德國語料所有語料來當作蒙地卡羅模擬的參考資料，也就是悲傷有 62 筆、中性有 79 筆、快樂有 71 筆、生氣有 127 筆資料，從每一筆情緒資料中計算表 3-5 所列之 23 個正規化特徵參數，因此，悲傷有 23×62 筆，中性有 23×79 筆，快樂有 23×71 筆，生氣則有 23×127 筆特徵參數資料，接著，計算每一種情緒之特徵參數平均值向量(為 23×1 之向量)與共變異數(為 23×23 之矩陣)，並以此特徵參數平均值向量與共變異數隨機產生一萬筆特徵參數資料，也就是 23×10000 筆隨機特徵參數為測試資料，因此，悲傷、中性、快樂以及生氣各有一萬筆測試資料，每一筆測試資料均有 23 個特徵參數。以下將利用蒙地卡羅模擬方法針對本論文所使用之各個分類器作蒙地卡羅模擬分析。

首先，使用德國語料所有資料當作訓練資料，也就是悲傷有 62 筆、中性有 79 筆、快樂有 71 筆、生氣有 127 筆訓練資料，接著，以

隨機產生之各情緒各一萬筆資料當作測試資料，並以 KNN 分類器對此隨機資料進行測試，表 5-12 為 KNN 分類器蒙地卡羅模擬之辨識結果，由 KNN 蒙地卡羅模擬鑑別矩陣顯示，KNN 分類器對於中性有極佳的辨識效果，辨識率為 99.64%，對於悲傷與生氣則維持中等的辨識率，辨識率分別為 82.04% 以及 88.39%，對於快樂則效果最差，辨識率只有 61.03%，整體之辨識率為 82.78%，從鑑別矩陣可以看出，KNN 分類器對於快樂與生氣比較容易產生混淆，尤其是快樂，容易被辨識歸類為生氣，而悲傷與生氣則分別有一部分易被辨識歸類為中性及快樂，整體來看，KNN 分類器之 Inside Test 辨識率為 88.79%，而蒙地卡羅模擬辨識率為 82.78%，這是由於使用 KNN Inside Test 時，測試資料中會有一筆資料與訓練資料完全相同，因此造成 KNN Inside Test 辨識率較蒙地卡羅模擬辨識率高，當 KNN 運用到實際狀況時，整體的辨識效果會變差，辨識率下降了約 6%。

接著，使用 SVM 分類器進行蒙地卡羅模擬，在使用 SVM 分類器作蒙地卡羅模擬前，首先，先使用德國語料庫所有資料當作訓練資料，並以此訓練資料產生一組 SVM Model，並找出最佳切割函數，接著，與 KNN 測試資料相同，以隨機產生之各情緒各一萬筆資料當作測試資料，並以 SVM 分類器對此隨機資料作測試，表 5-13 為 SVM 分類器蒙地卡羅模擬之辨識結果，由 SVM 蒙地卡羅模擬鑑別矩陣顯

示，SVM 分類器對於悲傷與中性有極佳的辨識效果，辨識率分別為 96.38% 以及 97.80%，對於生氣則維持不錯的辨識率，辨識率為 89.79%，對於快樂則效果相對較差，辨識率為 75.57%，整體之辨識率為 89.89%，從鑑別矩陣可以看出，SVM 分類器與 KNN 相同，對於快樂與生氣比較容易產生混淆，快樂有一部份容易被辨識歸類為生氣，整體來看，SVM 分類器之 Inside Test 辨識率為 100%，Outside Test 辨識率為 91.74%，而蒙地卡羅模擬辨識率為 89.89%，此結果顯示，SVM 對於特定資料的分類有相當不錯的分類效果，但當測試資料不在訓練資料中時，SVM 的辨識效果會有所下降，但仍維持不錯的辨識率，由 Outside Test 與蒙地卡羅模擬辨識結果可以發現，當訓練資料涵蓋較多的訓練語者時，SVM 之整體的辨識率將與 Outside Test 結果較為相近。

接著，使用 MLBN 分類方法進行蒙地卡羅模擬，在使用 MLBN 分類器作辨識前，與前面之 KNN 及 SVM 相同，使用德國語料庫所有語料當作訓練資料，計算訓練資料中各情緒類別正規化特徵參數之平均值與標準差，以此建置 MLBN 資料庫，接著，以隨機產生之各情緒各一萬筆資料當作測試資料，並以 MLBN 分類器對此隨機資料作可靠度測試，表 5-14 為 MLBN 可靠度測試之辨識結果，由 MLBN 蒙地卡羅模擬鑑別矩陣顯示，MLBN 分類器對於悲傷與中性有極佳的

辨識效果，辨識率分別為 94.53% 以及 97.83%，對於生氣則維持中等的辨識率，辨識率為 80.99%，對於快樂則效果相對較差，辨識率為 55.29%，整體之辨識率為 82.16%，從鑑別矩陣可以看出，與 SVM 分類器狀況較為相同，MLBN 對於快樂與生氣比較容易產生混淆，快樂有一部份容易被歸類為生氣，而生氣則有一小部分易被歸類為快樂，整體來看，MLBN 分類器之 Inside Test 辨識率為 81.12%，而整體之蒙地卡羅模擬辨識率為 82.16%，此結果顯示雖然 MLBN 對於特定資料的分類效果沒有 SVM 及 KNN 好，但在辨識上卻相對穩定，而 MLBN 最後所得到的是各情緒類別可能的發生機率，因此，在實際運用上，會比 SVM 及 KNN 有更大的應用空間。

最後，使用 MLBNC 分類方法來進行蒙地卡羅模擬，在使用 MLBNC 分類器作辨識前，與前面各分類器相同，使用德國語料庫所有資料當作訓練資料，計算各層各群之正規化特徵參數平均值向量與共變異數，以此建置 MLBNC 資料庫，接著，以隨機產生之各情緒各一萬筆資料當作測試資料，並以 MLBNC 分類器對此隨機資料作測試，表 5-15 為 MLBNC 分類器可靠度測試之辨識結果，由 MLBNC 蒙地卡羅模擬鑑別矩陣顯示，MLBNC 分類器與 SVM 及 MLBN 分類效果類似，對於悲傷與中性有極佳的辨識效果，辨識率分別為 97.28% 以及 99.10%，對於生氣則維持不錯的辨識率，辨識率為 87.33%，對

於快樂則相對辨識較差，辨識率為 65.79%，整體之辨識率為 87.33%，從鑑別矩陣可以看出，與前面幾個分類器狀況相同，MLBNC 對於快樂與生氣比較容易產生混淆，快樂有一部份容易被辨識為生氣，而生氣則有一小部分易被歸類為快樂，整體來看，MLBNC 分類器之 Inside Test 辨識率為 88.79%，而蒙地卡羅模擬辨識率為 87.33%，兩者差異不大，此結果顯示，雖然 MLBNC 對於特定資料的分類效果沒有 SVM 好，但整體辨識效果的穩定度卻與 SVM 相差不遠，且 MLBNC 最後所得到的是各情緒類別可能的發生機率，而不是單純的結果，因此，在實際應用面上，MLBNC 分類器可以提供語音情緒辨識上更大且可靠的應用空間。

表 5-16 為各分類器蒙地卡羅模擬測試之辨識正確率整理，從各分類器之辨識結果來看，四種分類器對於中性情緒均有極佳的辨識效果，而快樂則為最不容易辨識的情緒，各分類器對於快樂與生氣皆容易產生混淆的狀況，此結果可從圖 5-5 之正規化特徵參數分佈圖中看出，相關正規化特徵參數對於快樂與生氣的確缺乏有效的特徵來區隔此兩種情緒，因此，找出快樂與生氣在語音中能有效區隔此兩種情緒之特徵參數，亦為提升有效辨識的重點之一。從蒙地卡羅模擬測試辨識結果顯示，SVM 以及 MLBNC 均提供了良好的辨識率，由於 MLBNC 使用機率運算的方式來判別各情緒發生的機率，因此，在實

際應用面上，MLBNC 分類器比 SVM 分類器更可以提供較大且可靠的應用空間。

5.5 小結

本章首先針對分類樹、貝氏決策以及貝氏網路原理做簡單的介紹與說明，接著，根據分類樹原理，將本論文所使用之正規化特徵參數作分群分析，並依據 5.2.1 所提出之分群的原則與步驟，對各特徵參數做分群，並以此分群方式建立 MLBN 語音情緒分類器，MLBN 語音情緒分類器共有三層，第一層為分兩群，主要有三種分群方式，第一種分群方式中，悲傷與中性為同一群，快樂與生氣則為另一群，第二種分群方式，悲傷為獨立的一群，中性、快樂與生氣則為另一群，第三種分群方式，中性為獨立的一群，悲傷、快樂與生氣則為另一群；MLBN 第二層為分三群，主要有兩種分群方式，第一種分群方式中，悲傷以及中性各為獨立的一群，快樂與生氣為第三群，第二種分群方式則為悲傷與中性為同一群，快樂以及生氣則各為獨立的一群，最後，MLBN 第三層為分四群，也就是悲傷、中性、快樂與生氣皆為獨立的一群。

接著，利用 MLBN 語音情緒分類器，對德國語料庫進行 Inside Test 與 Outside Test，從 MLBN Inside Test 與單純使用貝氏決策之 Inside Test 結果比較，多層貝氏網路語音情緒分類器的辨識效果明顯優於貝

氏決策，因此，透過結合決策樹與貝氏網路分群的辨識方式，可以有效提升語音情緒辨識率，增加辨識的準確性，而從原始與正規化特徵參數之 MLBN Outside Test 結果比較可以發現，與 KNN 及 SVM 之結果相同，使用正規化特徵參數可以有效縮小語者之間在特徵參數上的差異，並增加各情緒之間的分類效果，此外，本章亦針對 MLBN 語音情緒分類器在訓練語料與測試語料為不同語系之情況下進行辨識，由結果顯示，與 KNN 及 SVM 分類器相同，MLBN 分類器在訓練語料與測試語料為不同語系時，同樣無法得到良好的分類效果。

本章亦提出一考慮特徵參數間相關性之多層共變異數貝氏網路分類器(MLBNC)，此 MLBNC 分類器與 MLBN 語音情緒分類器最大之不同在於考慮，不同分層分群之特徵參數間相關性，因此，在 MLBNC 各層各群之機率運算上採用多變數之機率密度計算，在本章中，同樣使用 MLBNC 語音情緒分類器，對德國語料庫進行 Inside Test 與 Outside Test，由辨識結果可以發現，MLBNC 之整體辨識正確率與 KNN 相近，優於 MLBN Inside Test，而從 MLBNC Outside Test 結果來看，其鑑別矩陣 MLBN Outside Test 類似，悲傷與生氣分別有一小部分容易被辨識為中性及快樂，而快樂則有一部份容易被辨識為生氣，而就整體辨識正確率來看，MLBNC 之辨識效果優於 MLBN，此結果顯示，考慮特徵參數間之相關性對於辨識分類有相當程度的助

益。同樣的，在訓練語料與測試語料為不同語系之情況下進行辨識，由結果顯示，當 MLBNC 分類器在訓練語料與測試語料為不同語系時，辨識效果較 KNN、SVM 以及 MLBN 分類器為佳，但整體辨識效果也僅在五成左右。

最後，本章利用蒙地卡羅模擬法對各分類器之辨識效果進行測試，由各分類器測試之辨識結果來看，四種分類器對於中性均有極佳的辨識效果，而快樂則為最不容易辨識的情緒，各分類器對於快樂與生氣皆容易產生混淆的狀況，因此，找出快樂與生氣在語音中能有效區隔此兩種情緒之特徵參數，亦為提升有效辨識的重點之一。而從測試辨識結果可以發現，SVM 以及 MLBNC 均提供了良好的辨識率，由於 MLBNC 使用機率運算的方式來判別各情緒發生的機率，而非只是一個判斷結果，因此，在實際應用面上，MLBNC 分類器將比 SVM 分類器更可以提供較大且可靠的應用空間。

從第四章與第五章之實驗結果來看，當測試語者之資料在訓練資料中時，各分類器之辨識效果均明顯優於訓練資料不含測試語者之資料時所得之結果，因此，如何透過適當的調適方式(Adaptive process)調整資料庫中的資料庫狀態，進而改善各情緒之分類效果將是語音情緒辨識的另一個發展重點，下一個章節將分別針對 MLBN 以及 MLBNC 分類器之適應性調適方法作說明與分析討論。

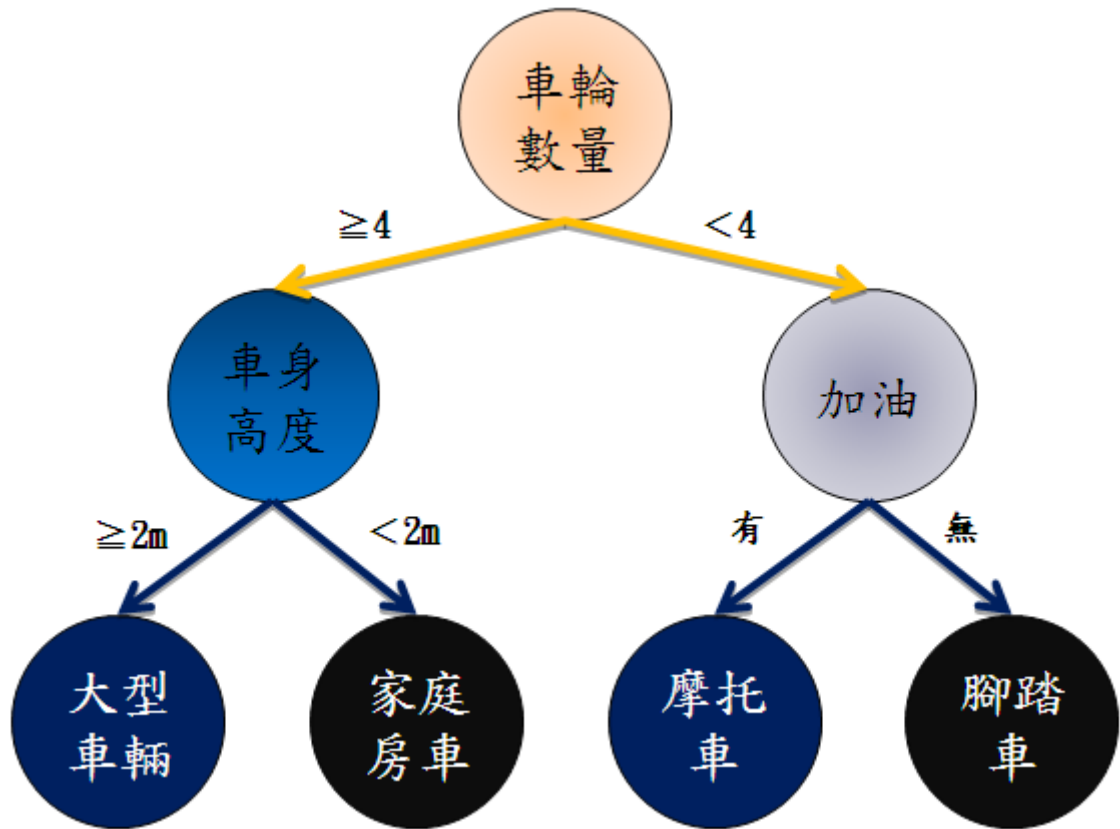


圖 5-1 簡易交通工具分類流程

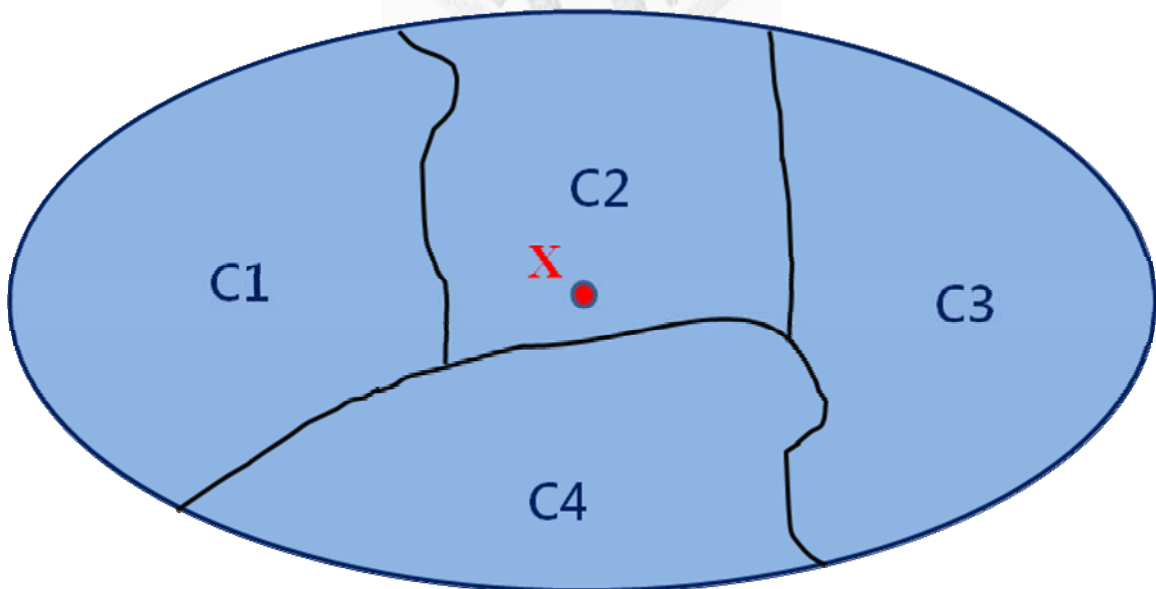


圖 5-2 四類別之貝氏決策示意圖

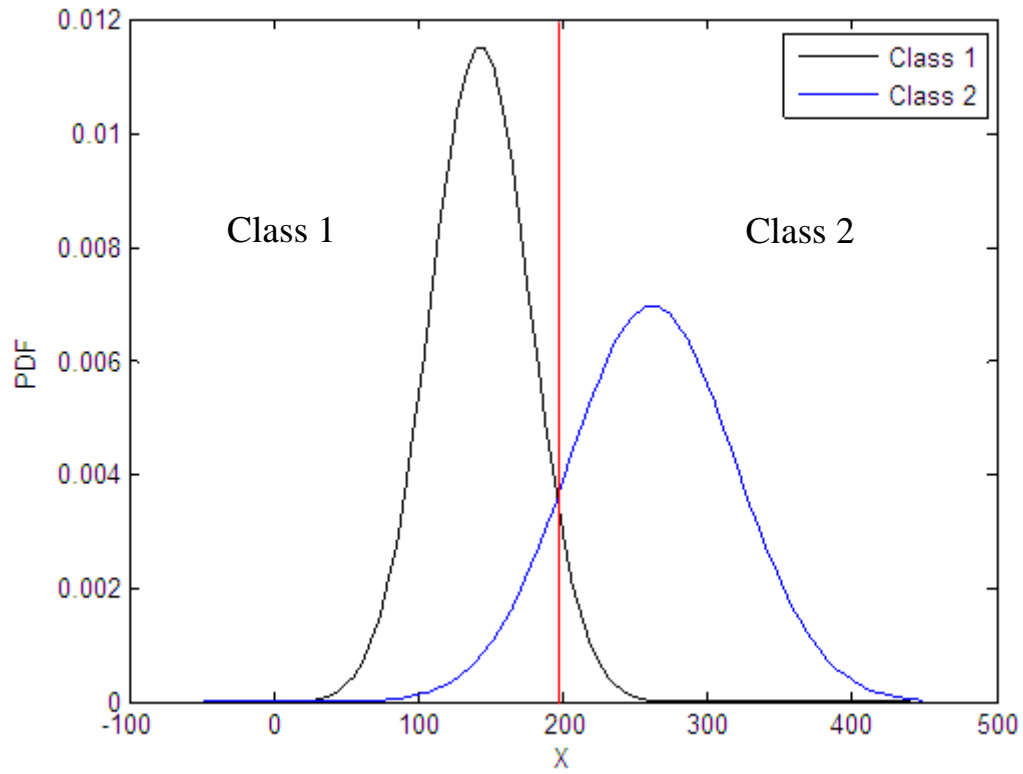


圖 5-3 兩類別之貝氏決策示意圖

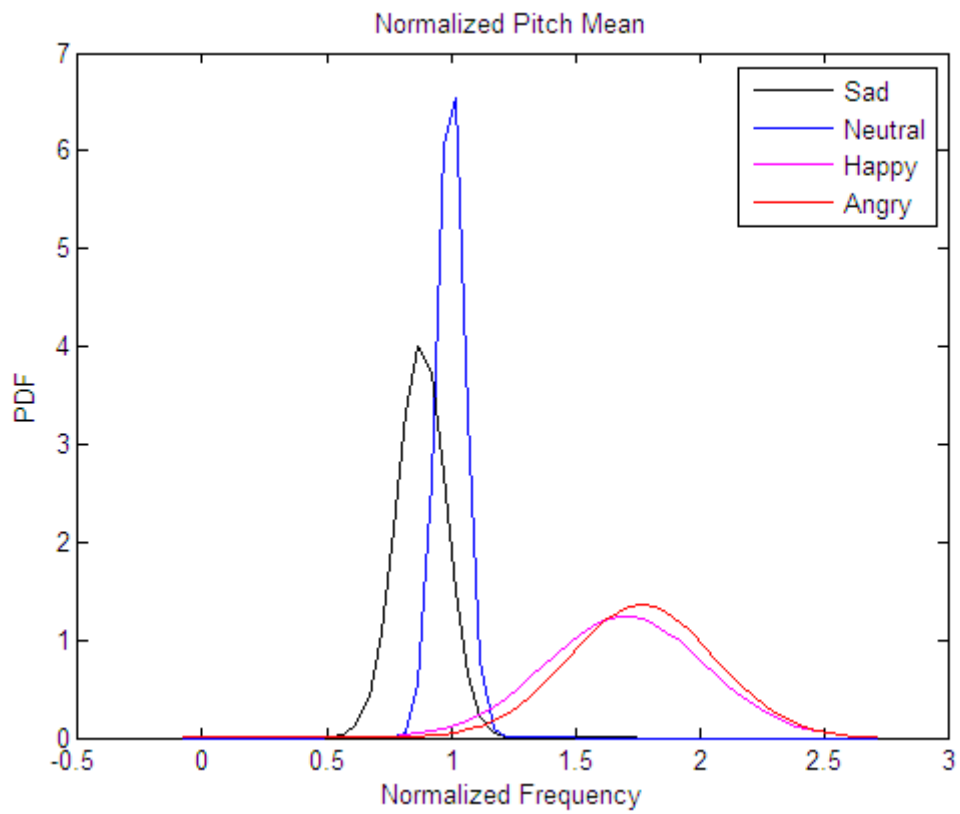
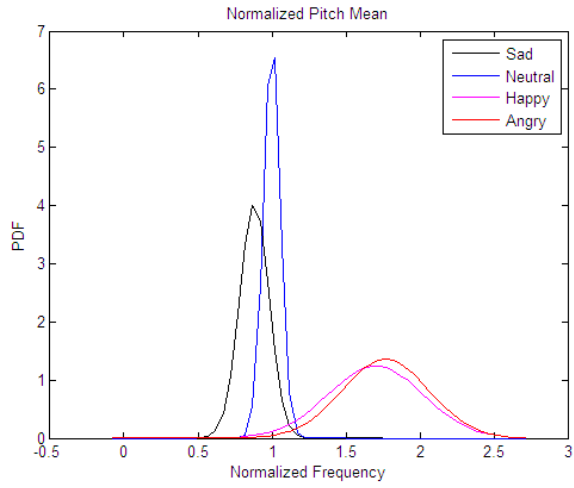
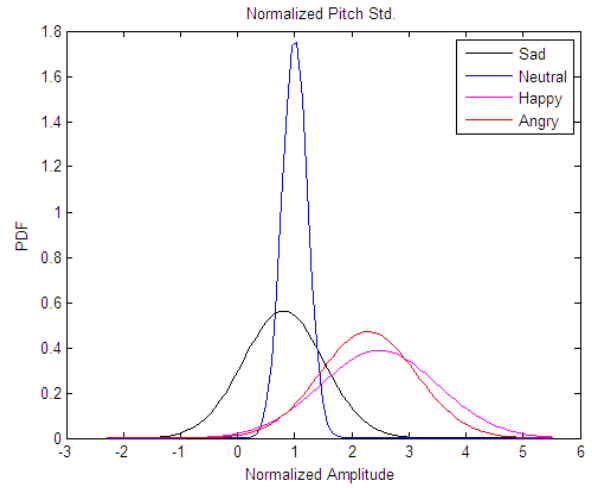


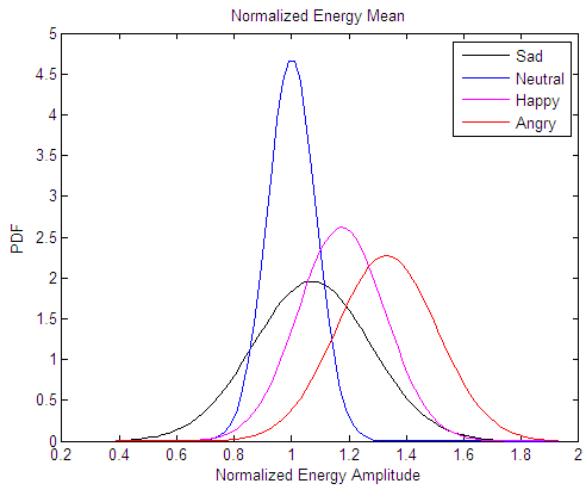
圖 5-4 四種情緒類別之正規化後音高分佈



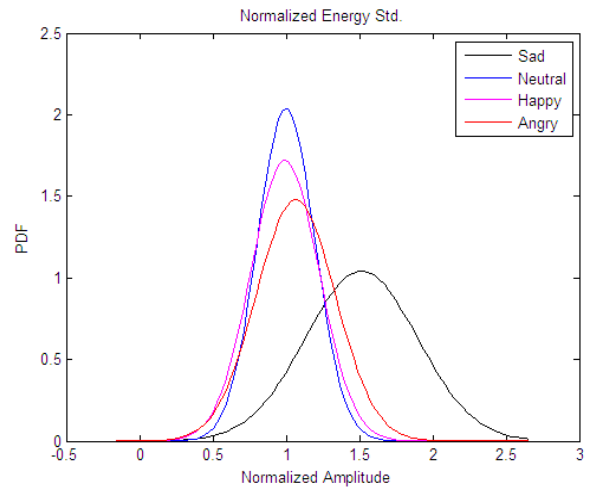
(a)



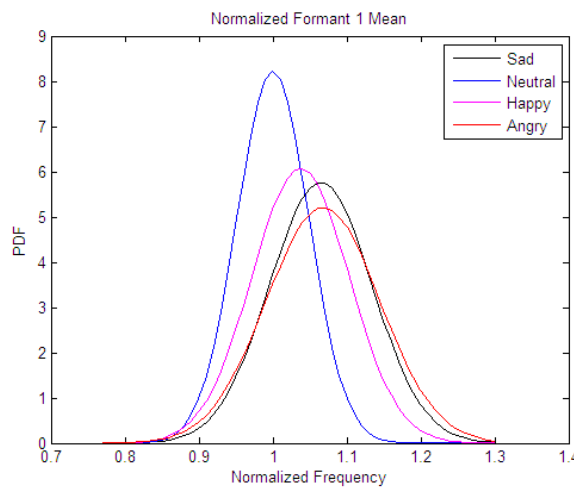
(b)



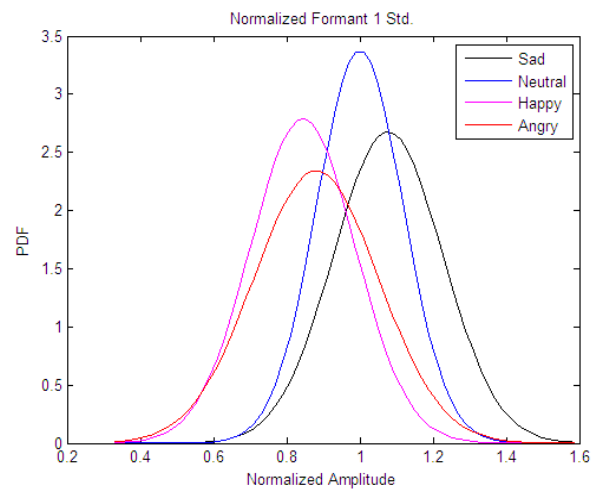
(c)



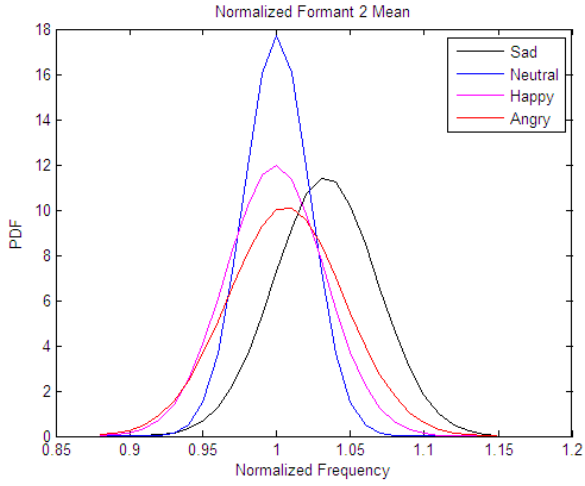
(d)



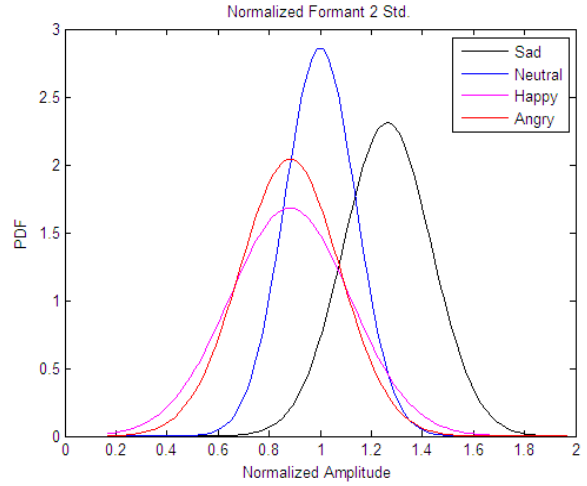
(e)



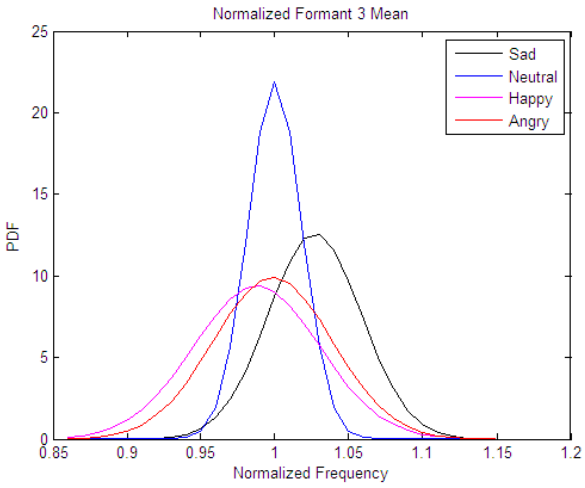
(f)



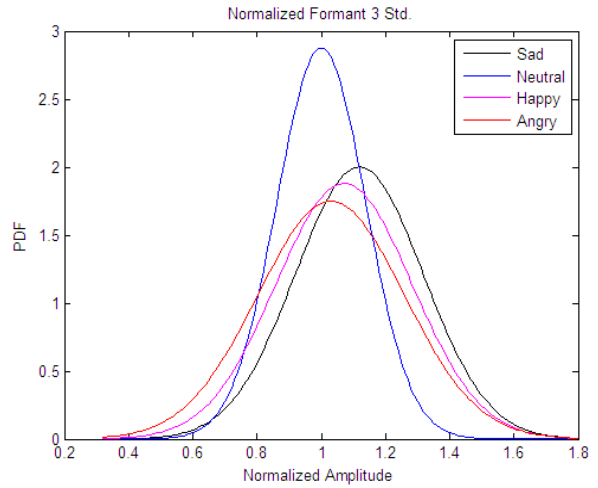
(g)



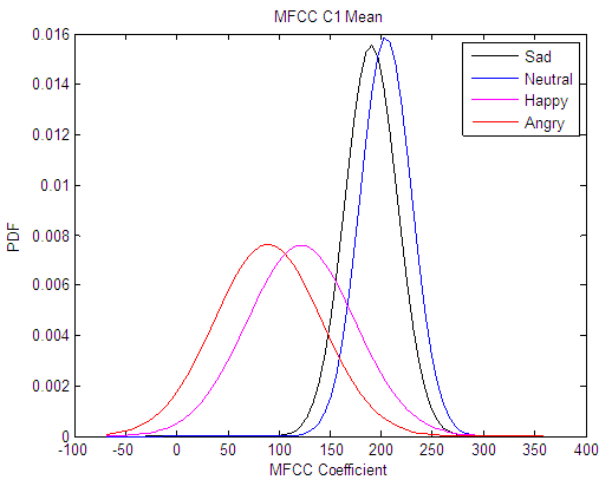
(h)



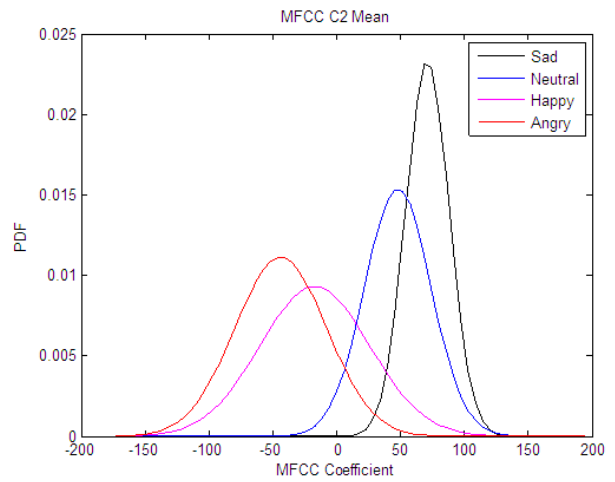
(i)



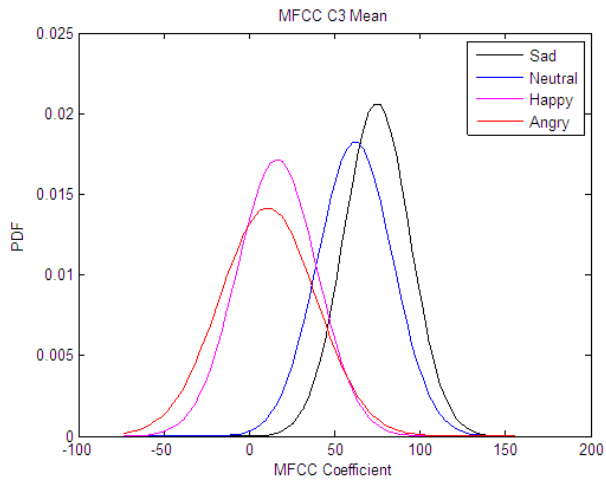
(j)



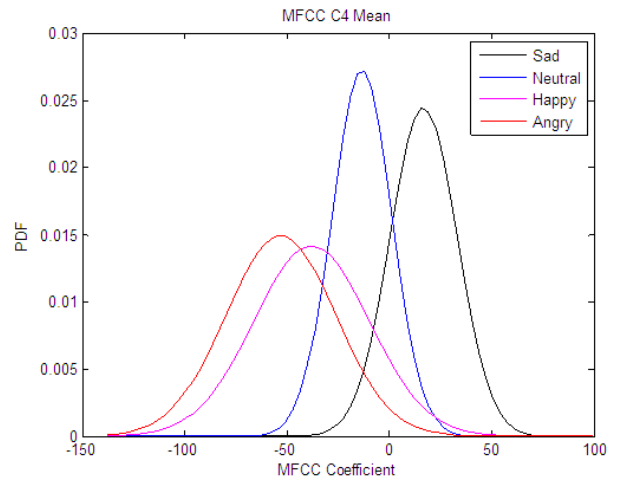
(k)



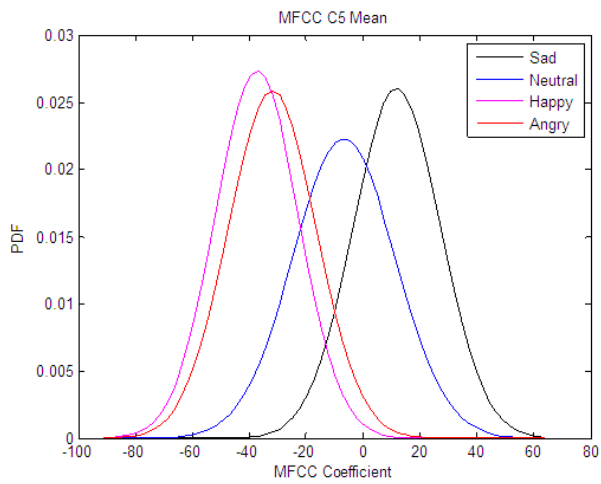
(l)



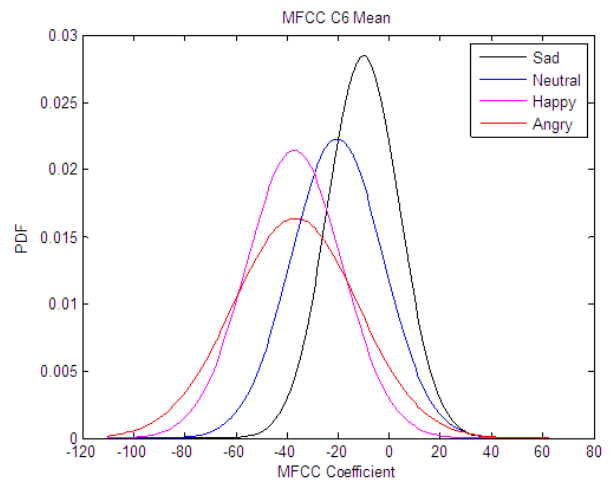
(m)



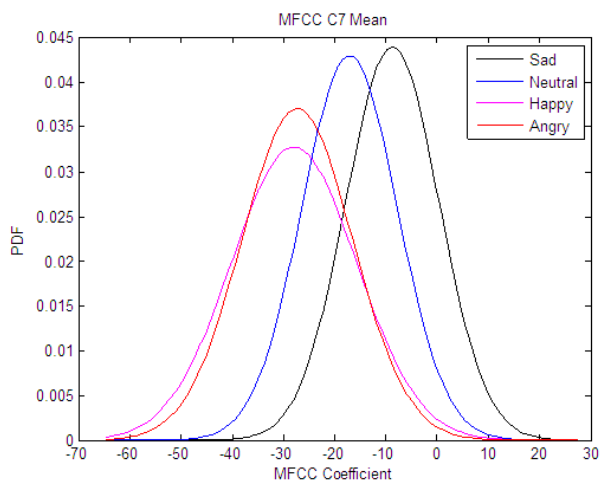
(n)



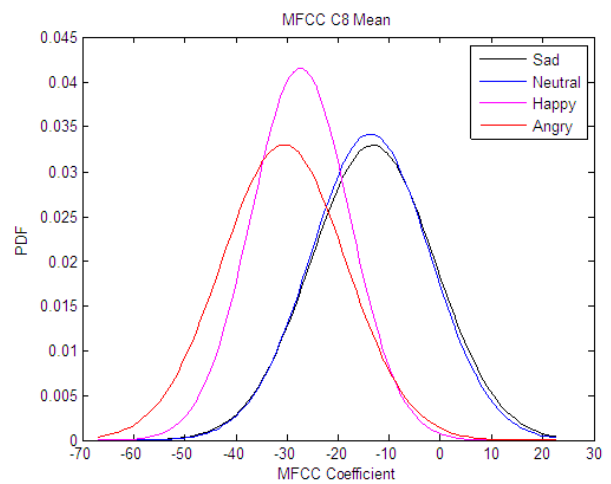
(o)



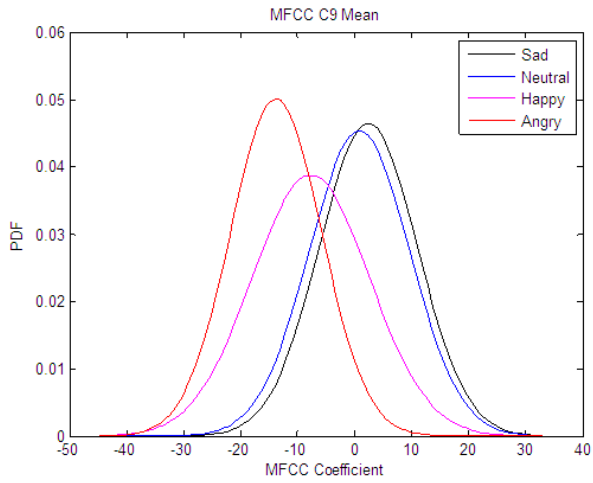
(p)



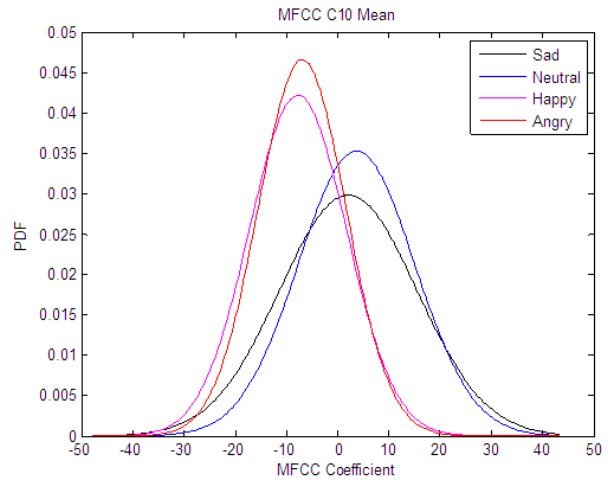
(q)



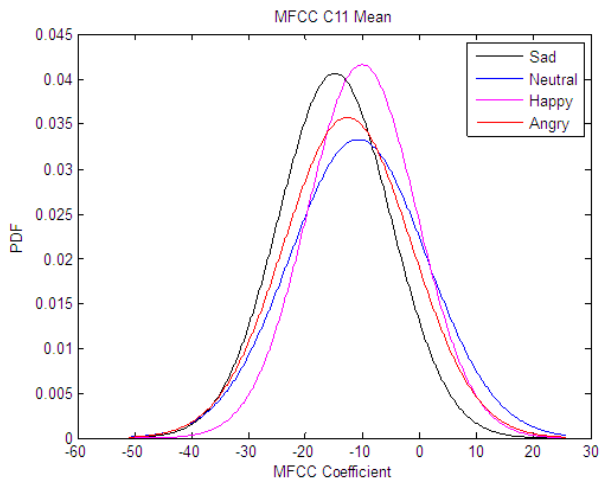
(r)



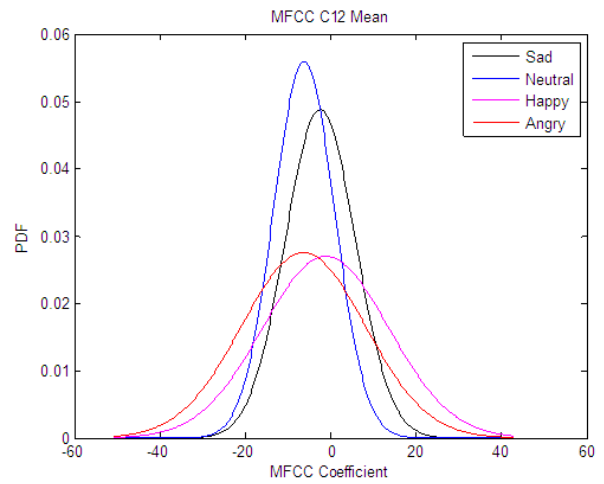
(s)



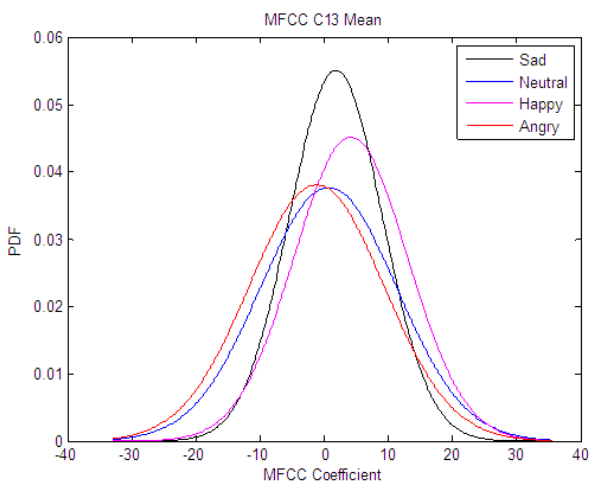
(t)



(u)

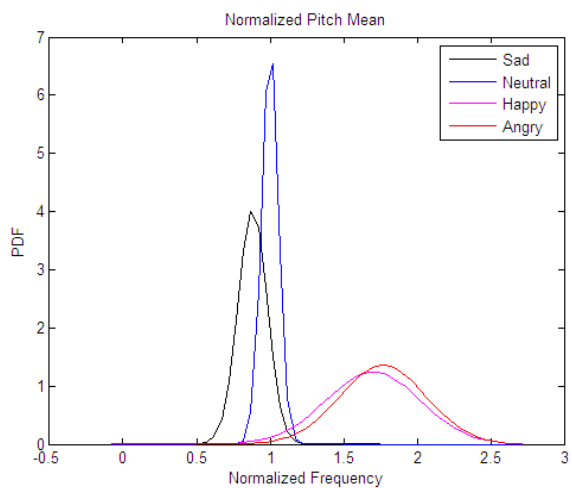


(v)

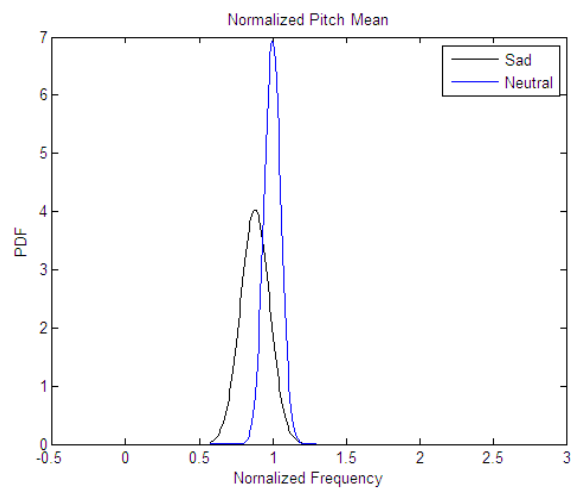


(w)

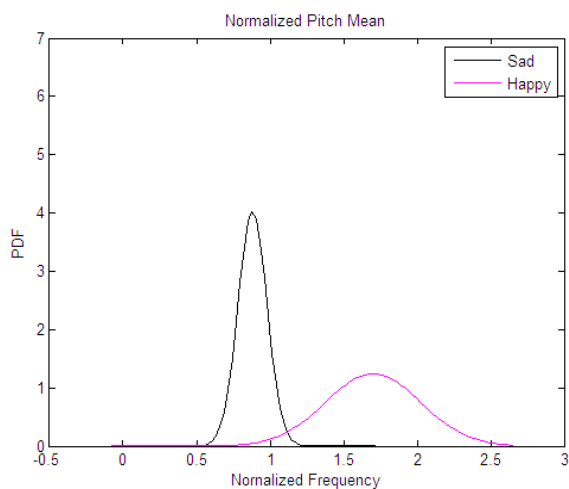
圖 5-5 四種情緒類別之正規化特徵參數分佈



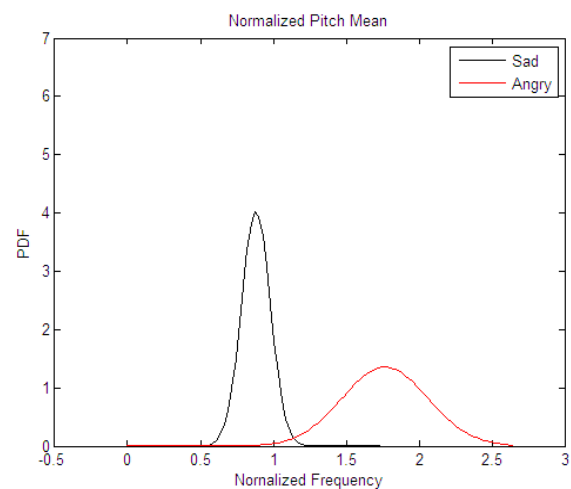
(a)



(b)



(c)



(d)

圖 5-6 悲傷與其他情緒之正規化後音高平均值分佈

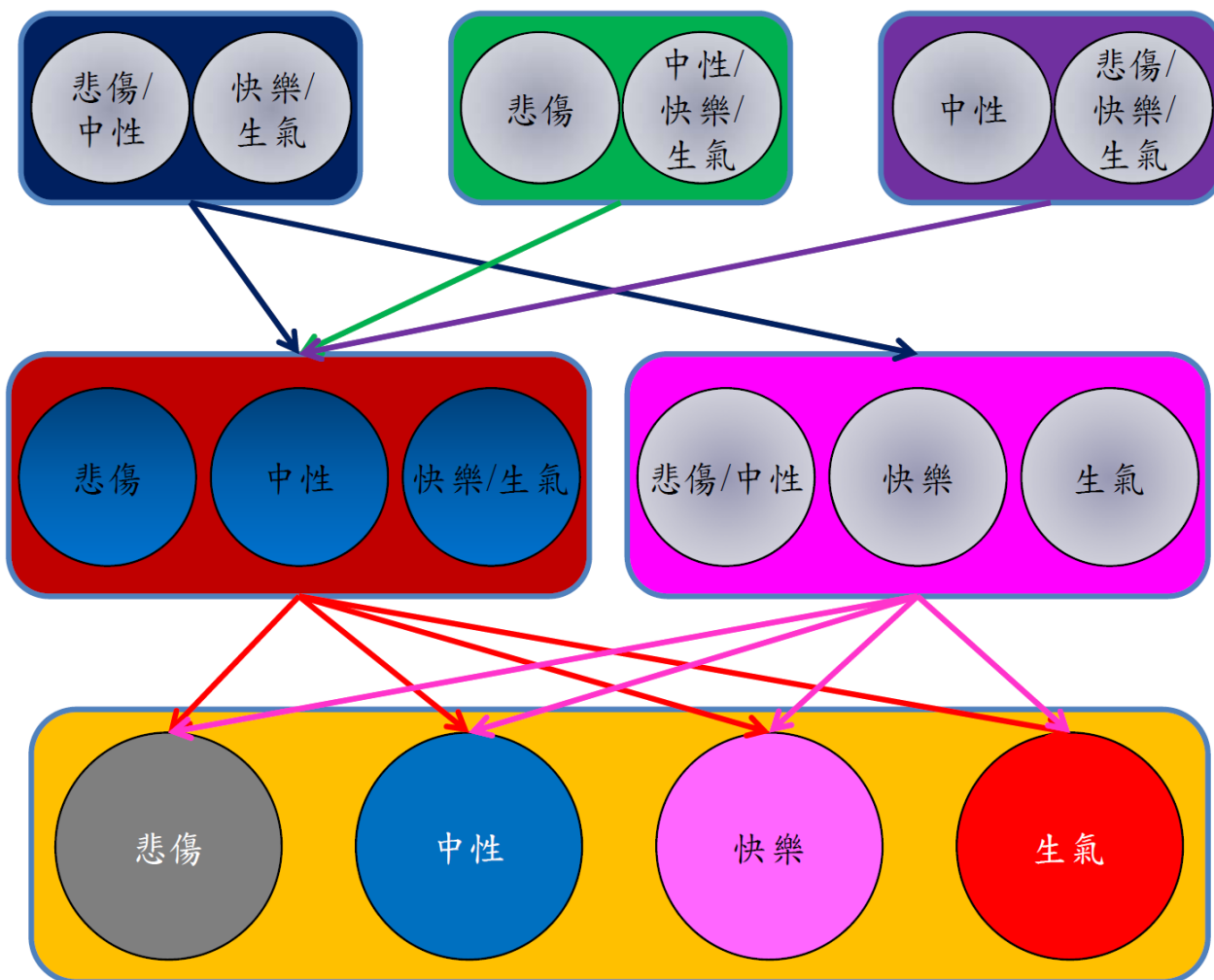


圖 5-7 多層貝氏網路語音情緒辨識架構

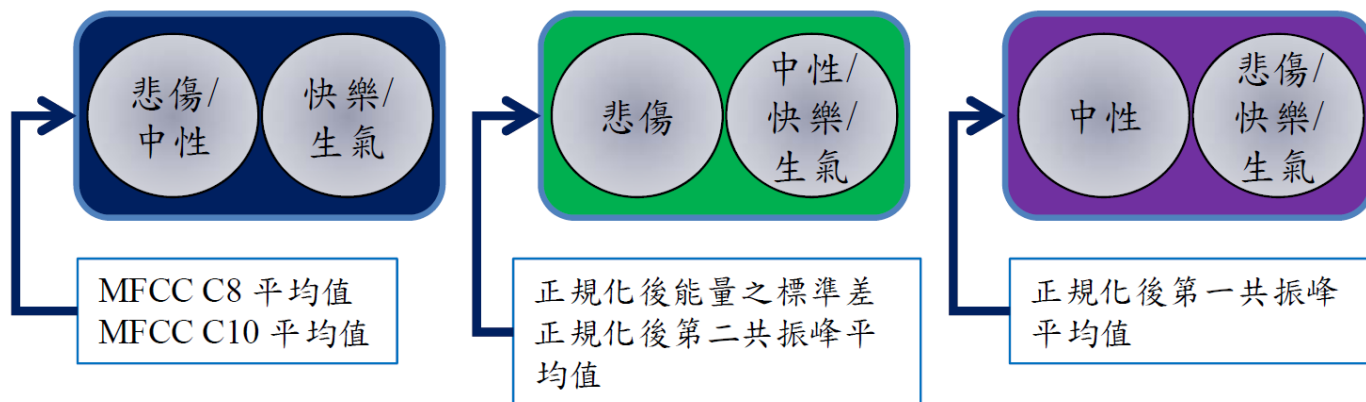
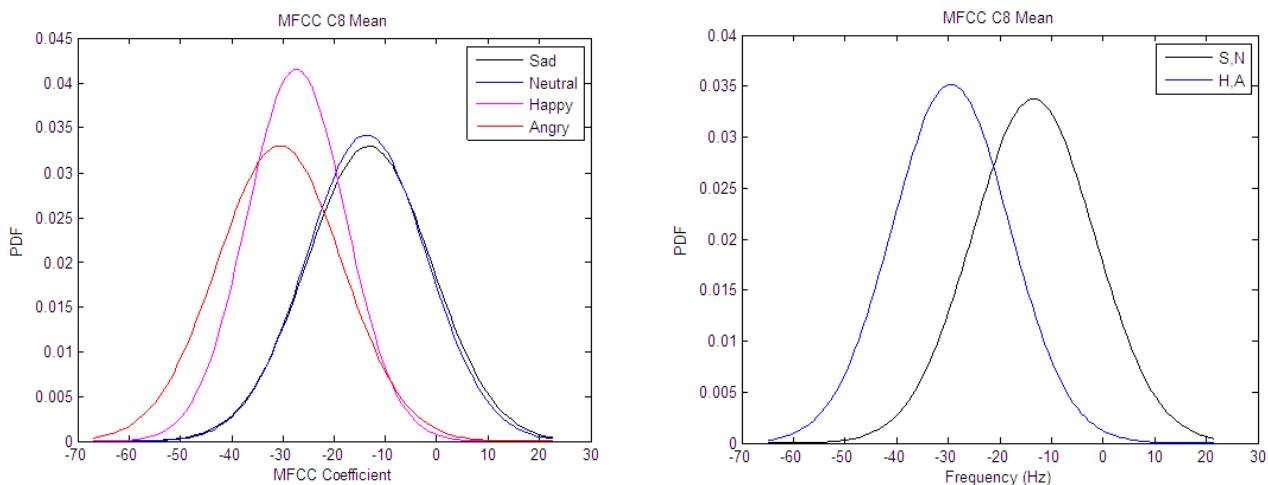


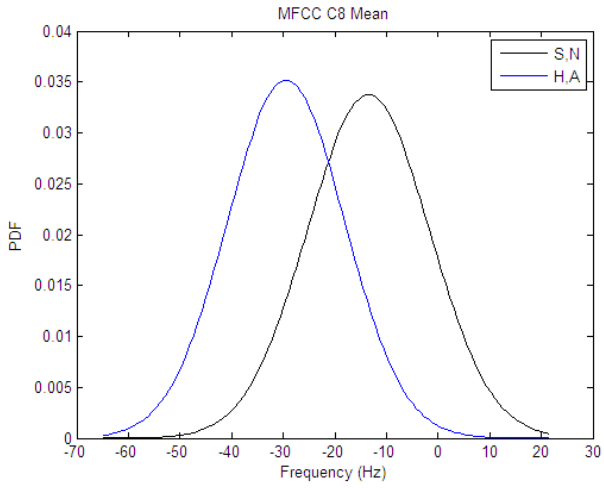
圖 5-8 多層貝氏網路第一層



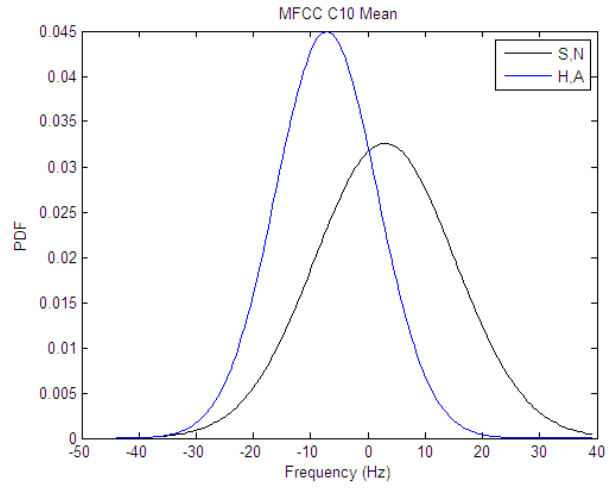
(a)原始分佈

(b)分群後分佈

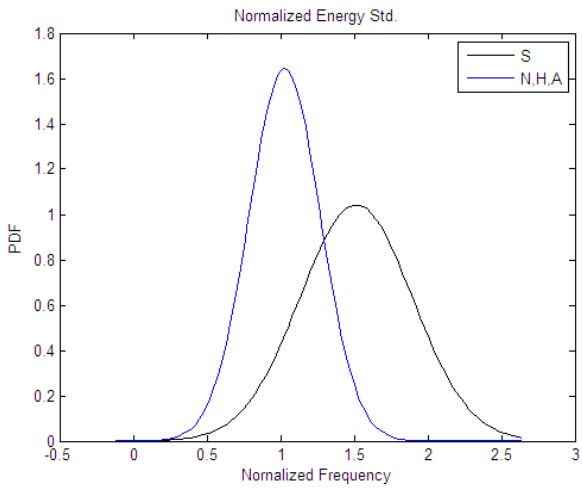
圖 5-9 MFCC C8 平均值原始分佈及分群後分佈



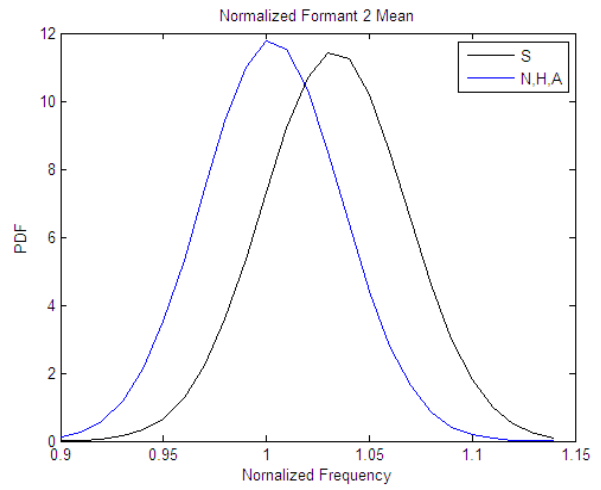
(a)



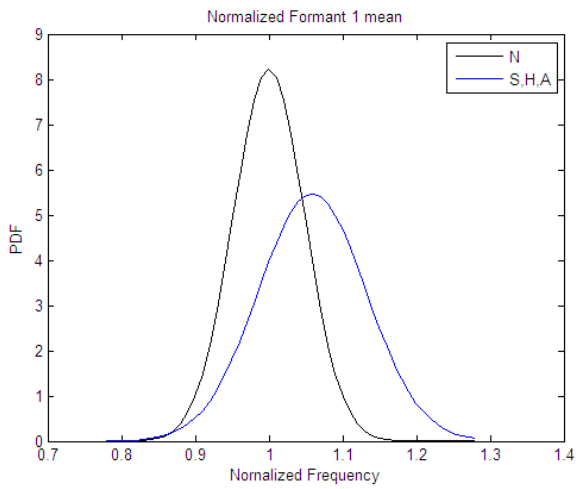
(b)



(c)



(d)



(e)

圖 5-10 多層貝氏網路第一層分群後分佈

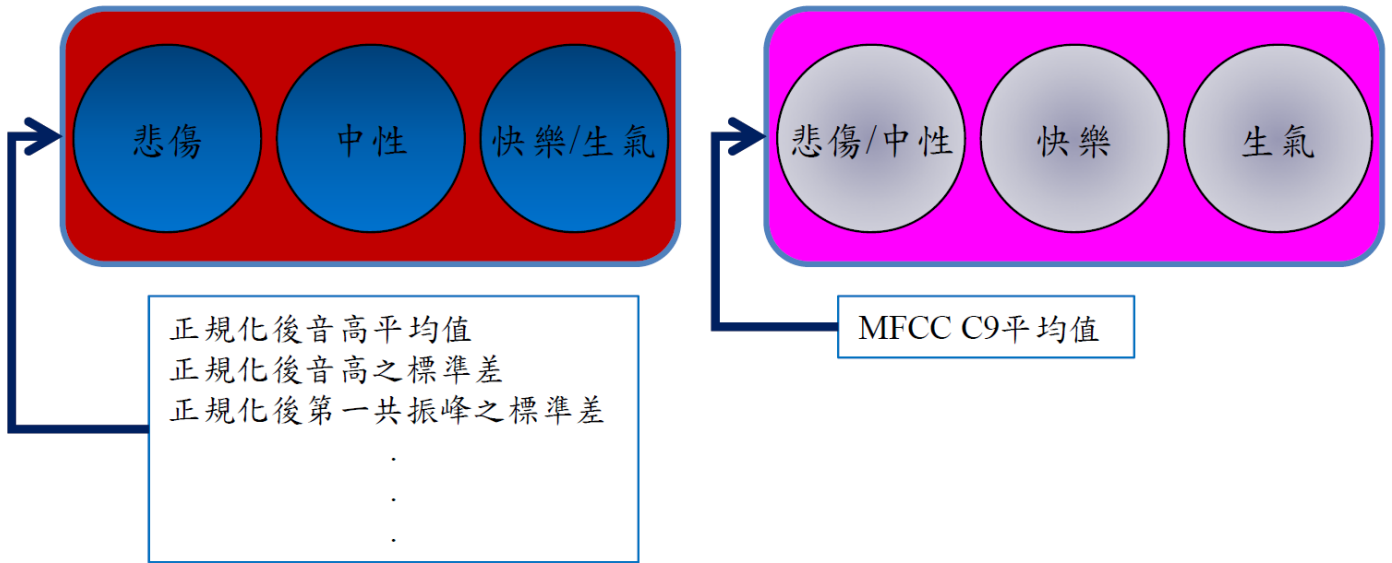
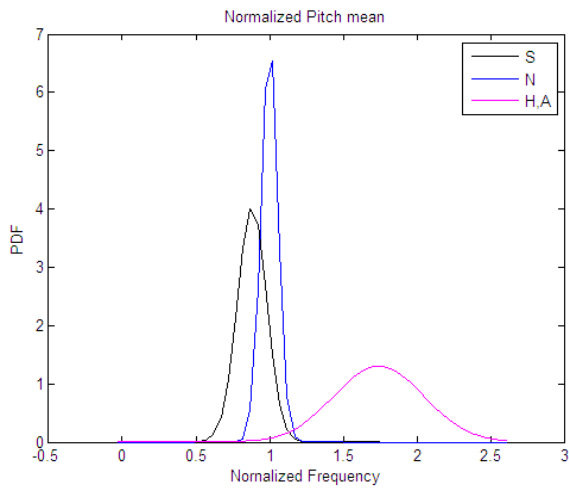
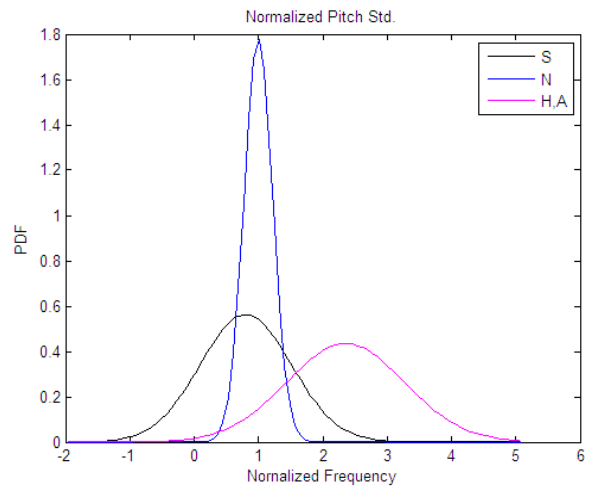


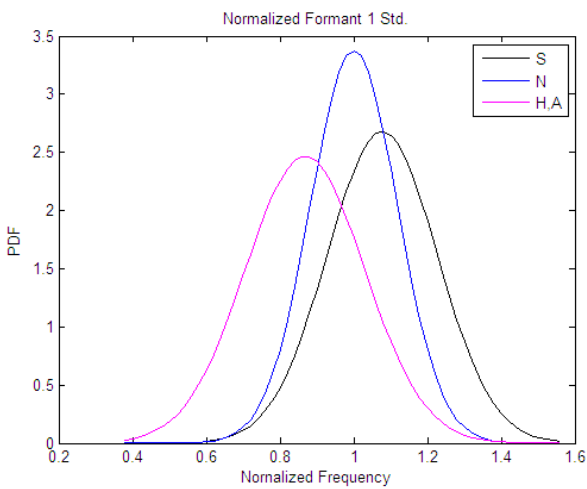
圖 5-11 多層貝氏網路第二層



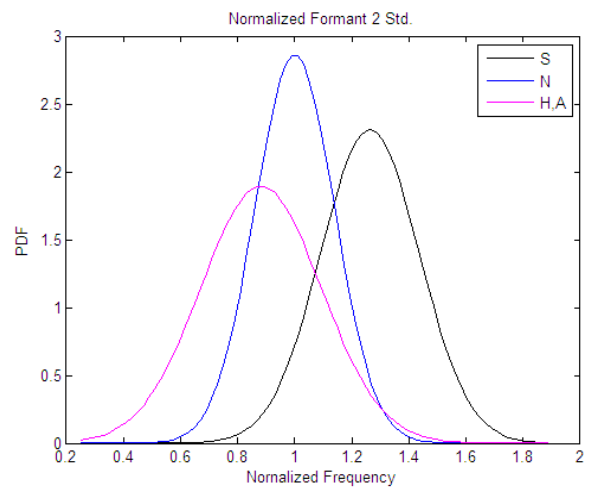
(a)



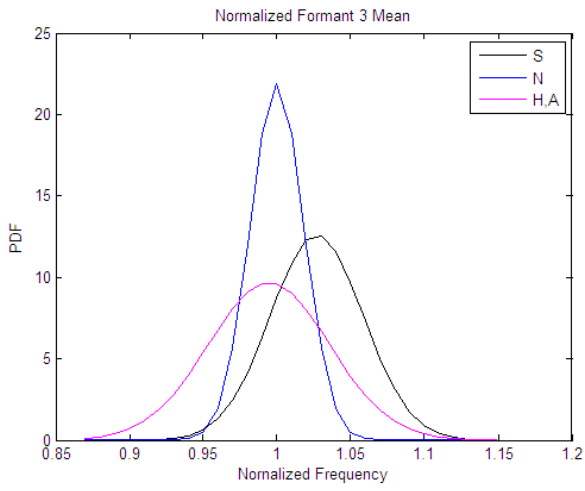
(b)



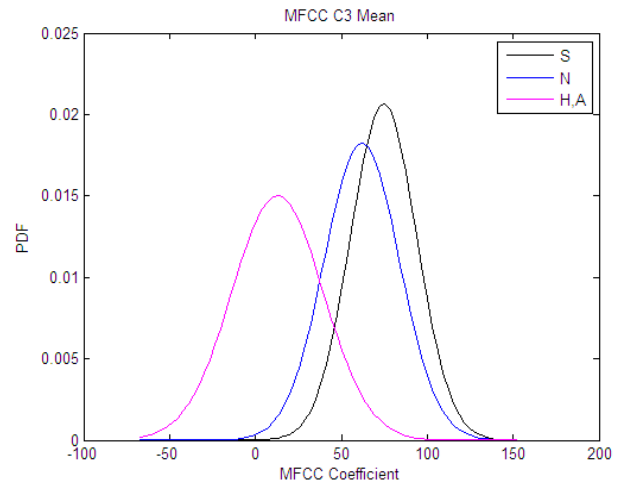
(c)



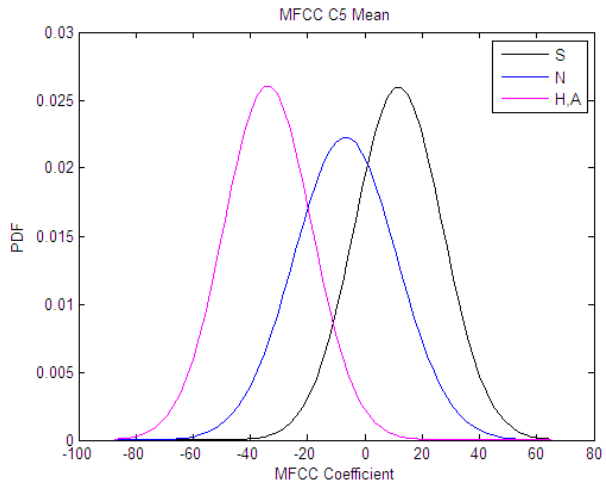
(d)



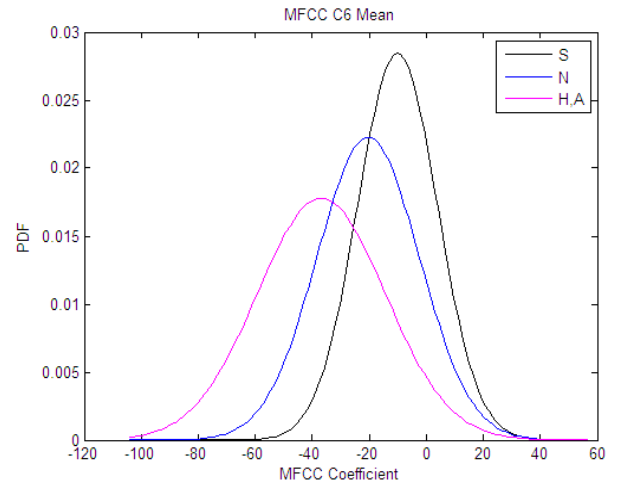
(e)



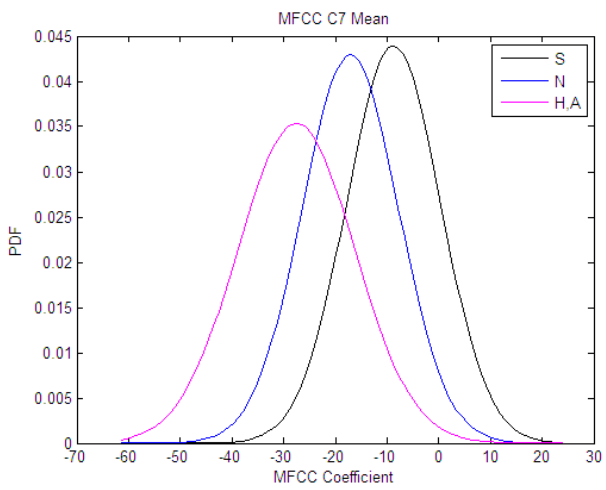
(f)



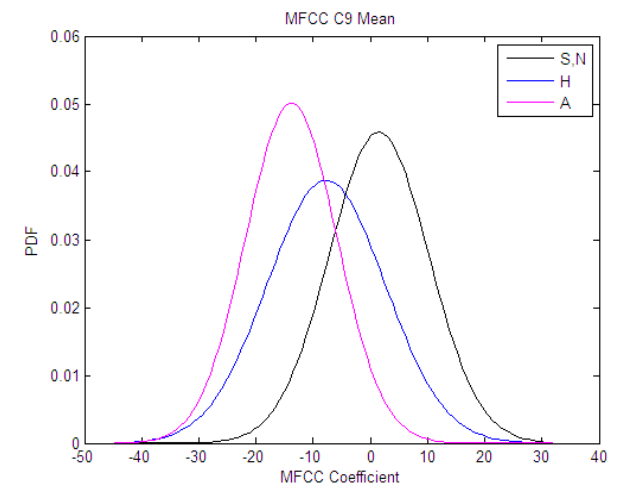
(g)



(h)



(i)



(j)

圖 5-12 多層貝氏網路第二層分群後分佈

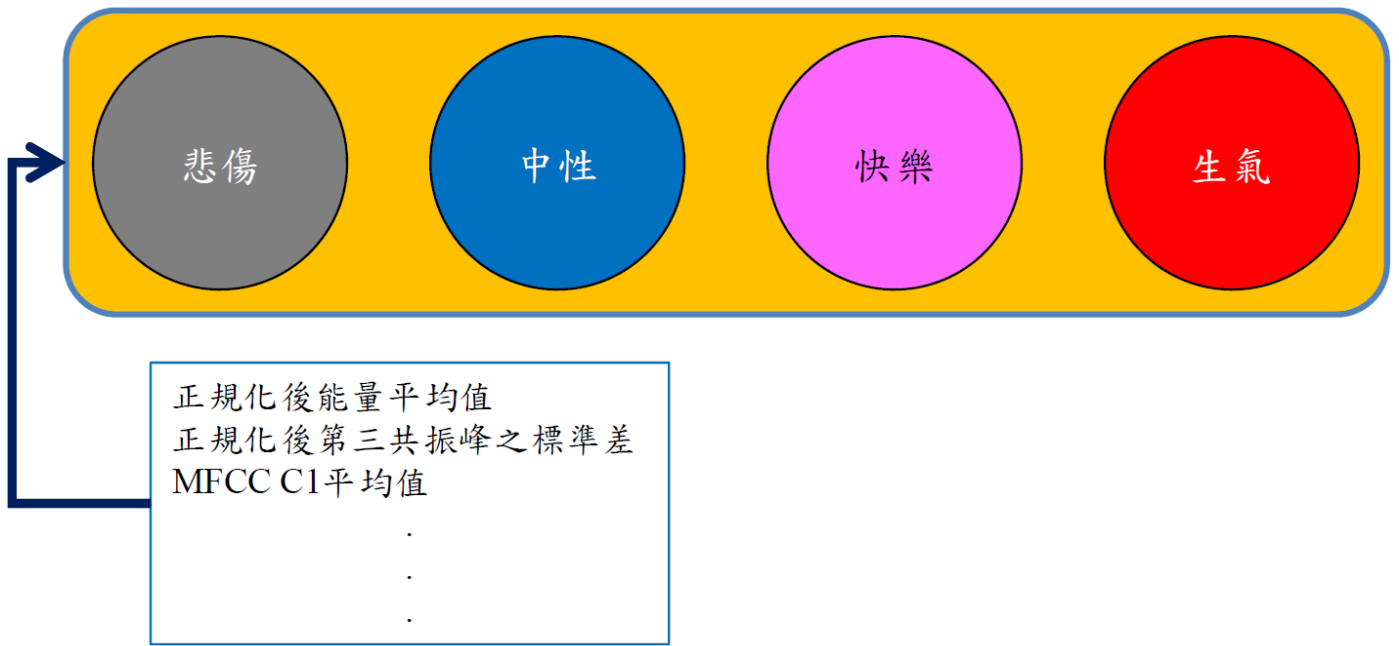


圖 5-13 多層貝氏網路第三層

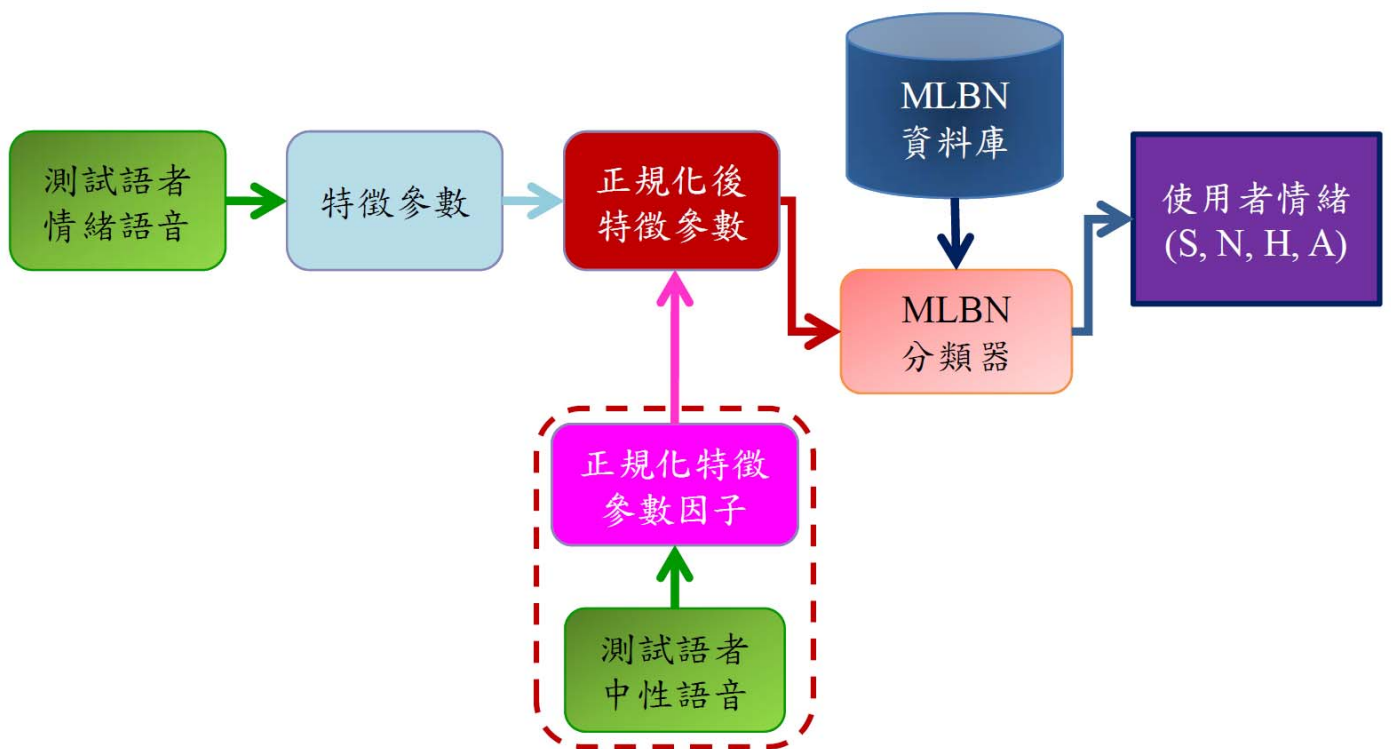


圖 5-14 多層貝氏網路語音情緒辨識流程

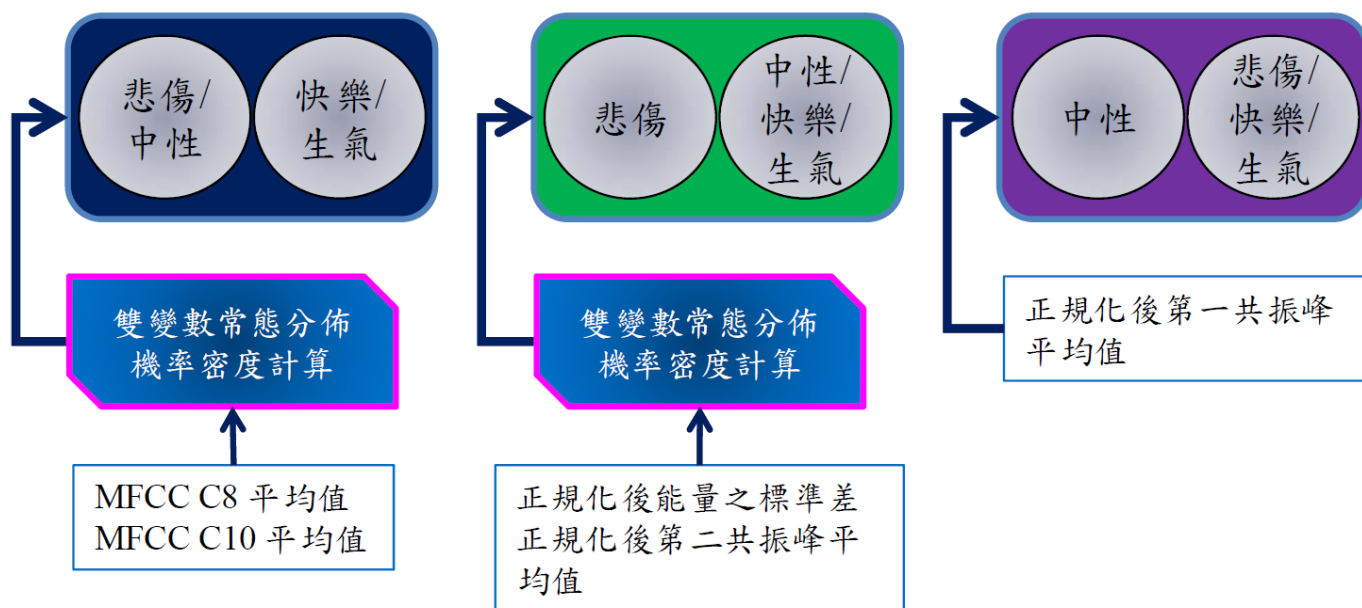


圖 5-15 多層共變異數貝氏網路第一層

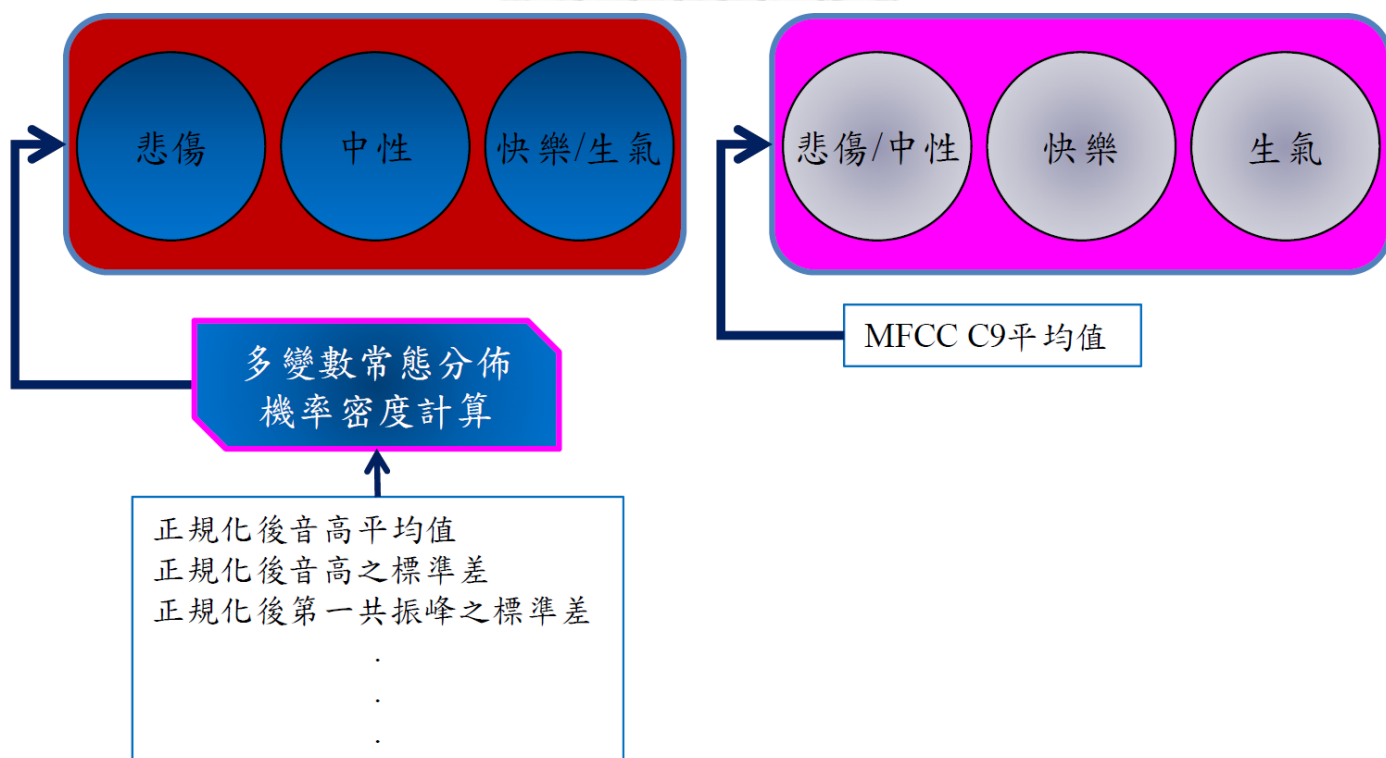


圖 5-16 多層共變異數貝氏網路第二層

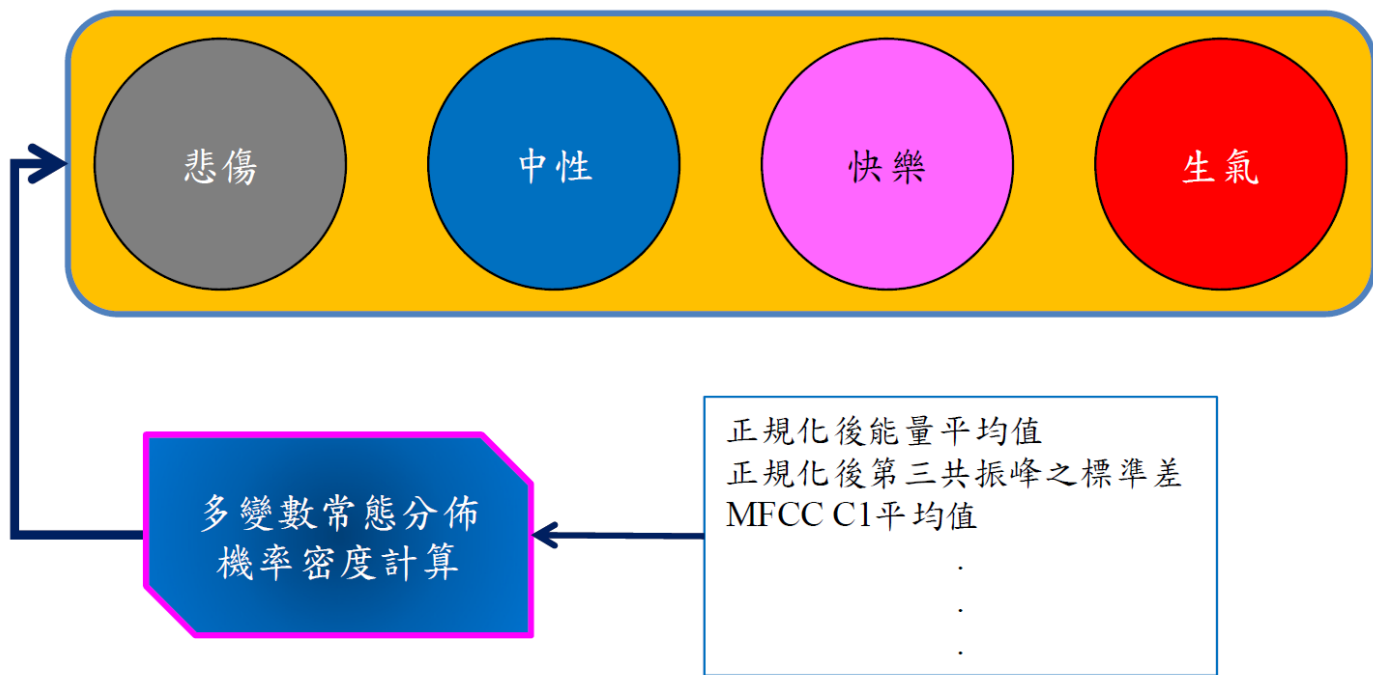


圖 5-17 多層共變異數貝氏網路第三層

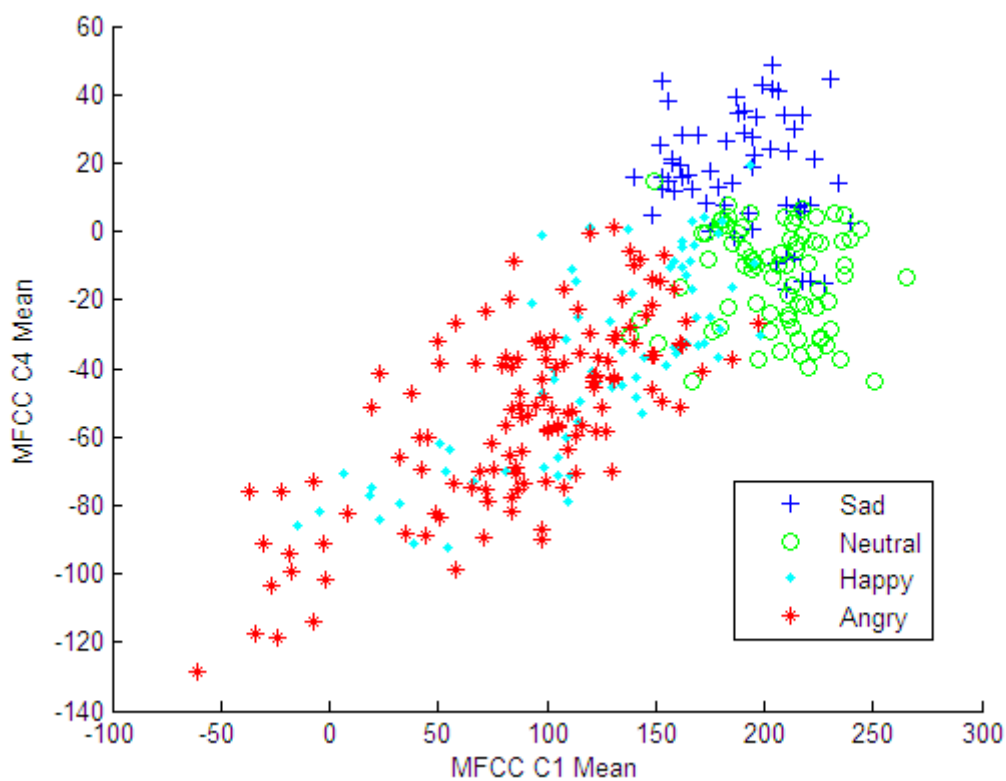


圖 5-18 MFCC C1 與 C4 平均值四情緒之特徵參數落點分佈

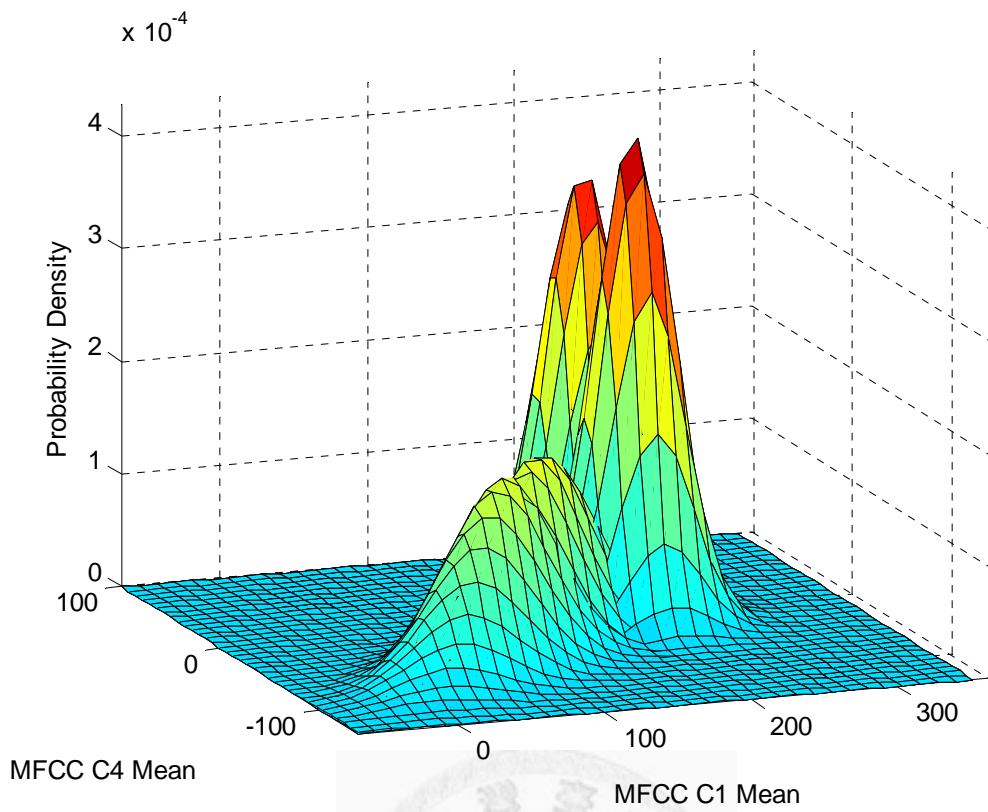


圖 5-19 MFCC C1 與 C4 平均值四情緒之雙變數常態分佈圖

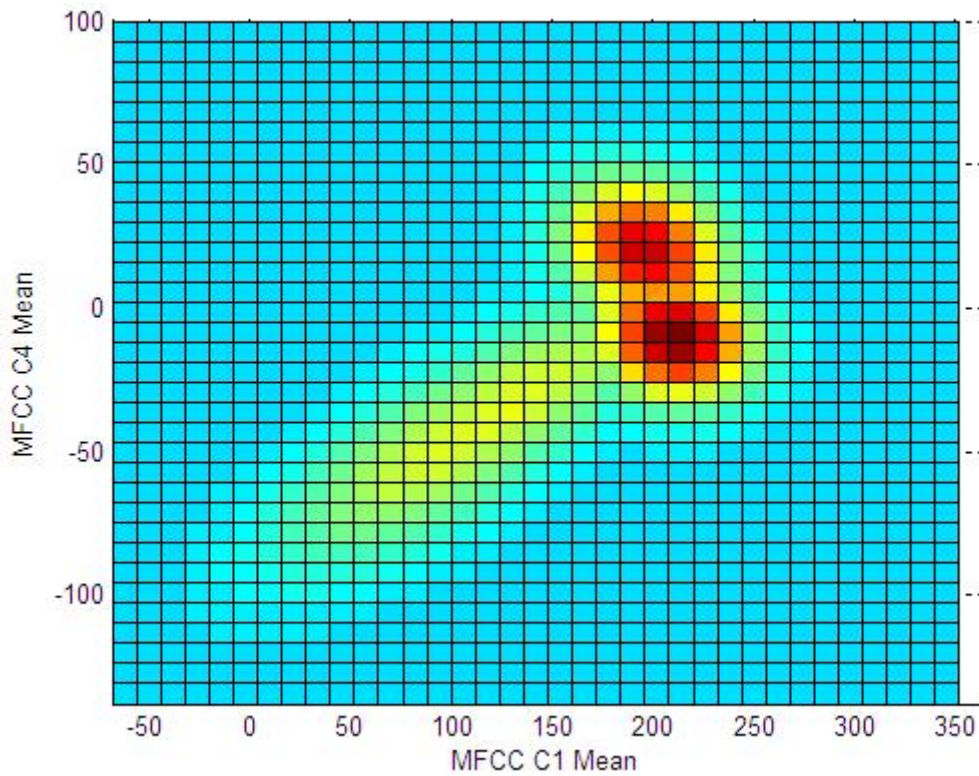


圖 5-20 MFCC C1 與 C4 平均值四情緒之雙變數常態分佈投影

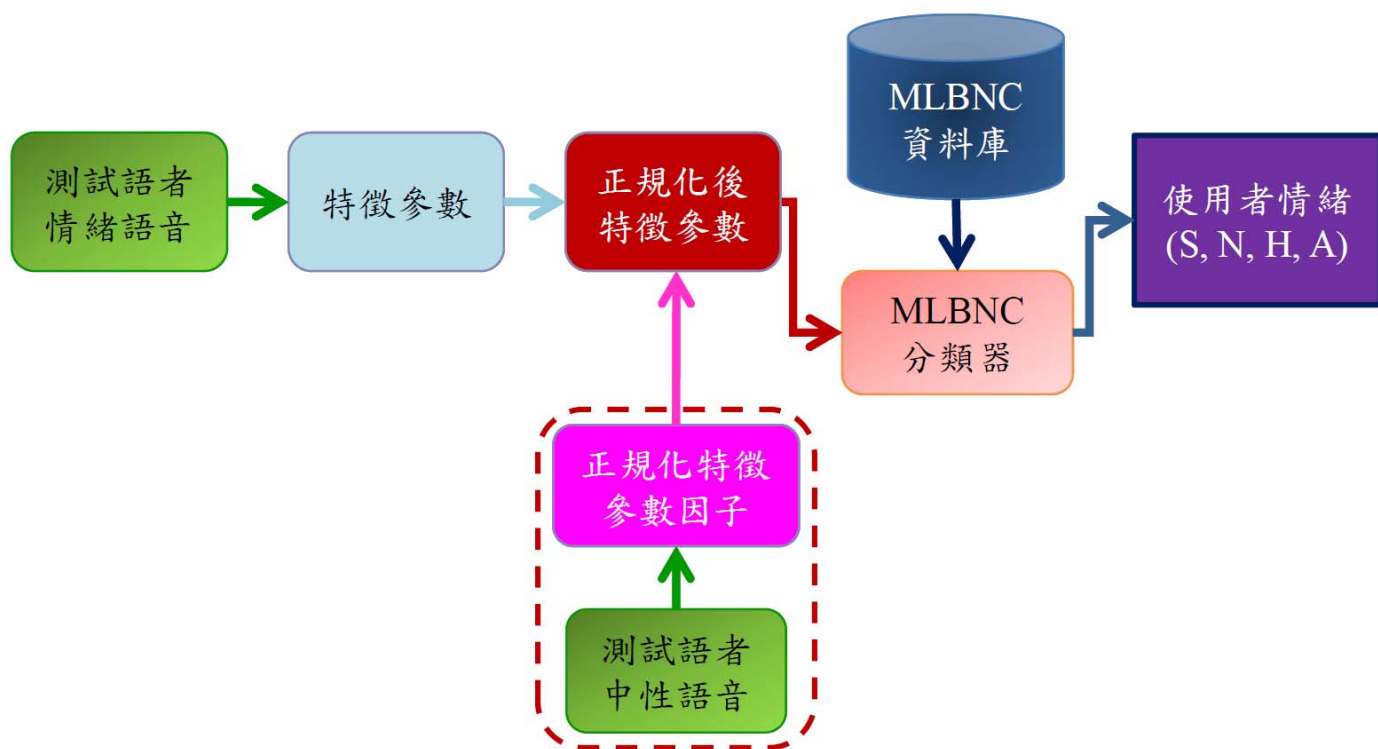


圖 5-21 多層共變異數貝氏網路語音情緒辨識流程



表 5-1 正規化後音高平均值之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	70.07%	99.13%	99.53%
中性	87.77%	1	99.40%	99.69%
快樂	96.57%	96.08%	1	41.30%
生氣	98.36%	98.10%	67.89%	1

表 5-2 正規化後音高之標準差各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	62.30%	88.25%	85.35%
中性	89.49%	1	97.41%	96.96%
快樂	79.22%	86.33%	1	38.55%
生氣	80.70%	85.52%	73.59%	1

表 5-3 情緒類別分群與對應之特徵參數

群數	分群	相關特徵參數
2 群	1.悲傷+快樂 2.快樂+生氣	MFCC C8 平均值 MFCC C10 平均值
	1.悲傷 2.中性+快樂+ 生氣	正規化後能量之標準差 正規化後第二共振峰平均值
	1.中性 2.悲傷+快樂+ 生氣	正規化後第一共振峰平均值
3 群	1.悲傷 2.中性 3.快樂+生氣	正規化後音高平均值 正規化後音高之標準差 正規化後第一共振峰之標準差 正規化後第二共振峰之標準差 正規化後第三共振峰平均值 MFCC C3 平均值 MFCC C5 平均值 MFCC C6 平均值 MFCC C7 平均值
	1.悲傷+中性 2.快樂 3.生氣	MFCC C9 平均值
4 群	1.悲傷 2.中性 3.快樂 4.生氣	正規化後能量平均值 正規化後第三共振峰之標準差 MFCC C1 平均值 MFCC C2 平均值 MFCC C4 平均值 MFCC C11 平均值 MFCC C12 平均值 MFCC C13 平均值

表 5-4 使用正規化特徵參數 MLBN Inside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	87% (54/62)	13% (8/62)	0	0	87% (54/62)
中性(79)	3% (2/79)	97% (77/79)	0	0	97% (77/79)
快樂(71)	0	8% (6/71)	62% (44/71)	30% (21/71)	62% (44/71)
生氣(127)	0	0.8% (1/127)	20.5% (26/127)	78.7% (100/127)	78.7% (100/127)
平均辨識率					81.1% (275/339)

表 5-5 使用正規化特徵參數 BD Inside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	68% (42/62)	24% (15/62)	2% (1/62)	6% (4/62)	68% (42/62)
中性(79)	9% (7/79)	89% (70/79)	3% (2/79)	0	89% (70/79)
快樂(71)	1.41% (1/71)	16.90% (12/71)	36.62% (26/71)	45.07% (32/71)	37% (26/71)
生氣(127)	0.8% (1/127)	3.9% (5/127)	15.0% (19/127)	80.3% (102/127)	80.3% (102/127)
平均辨識率					70.8% (240/339)

表 5-6 使用正規化特徵參數 MLBN Outside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	84% (52/62)	16% (10/62)	0	0	84% (52/62)
中性(79)	3% (2/79)	97% (77/79)	0	0	97% (77/79)
快樂(71)	1% (1/71)	14% (10/71)	54% (38/71)	31% (22/71)	54% (38/71)
生氣(127)	0	0.8% (1/127)	23.6% (30/127)	75.6% (96/127)	75.6% (96/127)
平均辨識率					77.6% (263/339)

表 5-7 使用原始特徵參數 MLBN Outside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	63% (39/62)	37% (23/62)	0	0	63% (39/62)
中性(79)	6% (5/79)	87% (69/79)	6% (5/79)	0	87% (69/79)
快樂(71)	0	24% (17/71)	42% (30/71)	34% (24/71)	42% (30/71)
生氣(127)	0	4.7% (6/127)	17.3% (22/127)	78.0% (99/127)	78.0% (99/127)
平均辨識率					69.9% (237/339)

表 5-8 MLBN 不同語系測試之鑑別矩陣(正規化特徵參數)

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(600)	20.7% (124/600)	75.8% (455/600)	1.3% (8/600)	2.2% (13/600)	20.7% (124/600)
中性(600)	3.5% (21/600)	96.0% (576/600)	0.2% (1/600)	0.3% (2/600)	96.0% (576/600)
快樂(600)	42.8% (257/600)	8.2% (49/600)	35.0% (210/600)	14.0% (84/600)	35.0% (210/600)
生氣(600)	59.3% (356/600)	15.8% (95/600)	19.2% (115/600)	5.7% (34/600)	5.7% (34/600)
平均辨識率					39.33% (944/2400)

表 5-9 使用正規化特徵參數 MLBNC Inside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	97% (60/62)	3% (2/62)	0	0	97% (60/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	3% (2/71)	72% (51/71)	25% (18/71)	72% (51/71)
生氣(127)	0	0.8% (1/127)	11.8% (15/127)	87.4% (111/127)	87.4% (111/127)
平均辨識率					88.8% (301/339)

表 5-10 使用正規化特徵參數 MLBNC Outside Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	82% (51/62)	16% (10/62)	2% (1/62)	0	82% (51/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	3% (2/71)	62% (44/71)	35% (25/71)	62% (44/71)
生氣(127)	0	0	14.2% (18/127)	85.8% (109/127)	85.8% (109/127)
平均辨識率					83.5% (283/339)

表 5-11 MLBNC 不同語系測試之鑑別矩陣(正規化特徵參數)

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(600)	18.0% (108/600)	71.5% (429/600)	6.5% (39/600)	4.0% (24/600)	18.0% (108/600)
中性(600)	7.0% (42/600)	88.2% (529/600)	4.0% (24/600)	0.8% (5/600)	88.2% (529/600)
快樂(600)	2.8% (17/600)	5.0% (30/600)	39.7% (238/600)	52.5% (315/600)	39.7% (238/600)
生氣(600)	3.7% (22/600)	15.5% (93/600)	18.3% (110/600)	62.5% (375/600)	62.5% (375/600)
平均辨識率					52.08% (1250/2400)

表 5-12 KNN 蒙地卡羅模擬鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷	中性	快樂	生氣	辨識正確率
悲傷(10k)	82.04%	17.90%	0.03%	0.03%	82.04%
中性(10k)	0.30%	99.64%	0.06%	0.00%	99.64%
快樂(10k)	0.03%	1.81%	61.03%	37.13%	61.03%
生氣(10k)	0.00%	0.24%	11.37%	88.39%	88.39%
平均辨識率					82.78% (33100/40k)

表 5-13 SVM 蒙地卡羅模擬鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷	中性	快樂	生氣	辨識正確率
悲傷(10k)	96.38%	3.29%	0.14%	0.19%	96.38%
中性(10k)	1.46%	97.80%	0.60%	0.14%	97.80%
快樂(10k)	0.02%	0.37%	75.57%	24.04%	75.57%
生氣(10k)	0.00%	0.06%	10.15%	89.79%	89.79%
平均辨識率					89.89% (35954/40k)

表 5-14 MLBN 蒙地卡羅模擬鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷	中性	快樂	生氣	辨識正確率
悲傷(10k)	94.53%	5.46%	0.01%	0.00%	94.53%
中性(10k)	2.16%	97.83%	0.01%	0.00%	97.83%
快樂(10k)	2.06%	3.80%	55.29%	38.85%	55.29%
生氣(10k)	0.15%	0.43%	18.43%	80.99%	80.99%
平均辨識率					82.16% (32864/40k)

表 5-15 MLBNC 蒙地卡羅模擬鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷	中性	快樂	生氣	辨識正確率
悲傷(10k)	97.28%	2.32%	0.40%	0.00%	97.28%
中性(10k)	0.84%	99.10%	0.06%	0.00%	99.10%
快樂(10k)	0.00%	0.05%	65.79%	34.16%	65.79%
生氣(10k)	0.00%	0.00%	12.67%	87.33%	87.33%
平均辨識率					87.38% (34950/40k)

表 5-16 各分類器蒙地卡羅模擬結果

分類器 \ 辨識結果	悲傷	中性	快樂	生氣	平均辨識率 (辨識個數/ 測試個數)
KNN	82.04%	99.64%	61.03%	88.39%	82.78% (33100/40k)
SVM	96.38%	97.80%	75.57%	89.79%	89.89% (35954/40k)
MLBN	94.53%	97.83%	55.29%	80.99%	82.16% (32864/40k)
MLBNC	97.28%	99.10%	65.79%	87.33%	87.38% (34950/40k)



第六章 適應性語音情緒辨識

由第四章與第五章，KNN、SVM、MLBN 以及 MLBNC 於不同語系之測試辨識結果(表 4-4、表 4-8、表 5-8 以及表 5-11)可以發現，當訓練語料庫與使用者之情緒表達方式有很大之差異時，四種分類器之分類效果均不佳，除了 MLBNC 在五成左右，其他分類器之辨識率皆約在四成上下，此外，從第四章與第五章各分類器之 Inside 與 Outside Test 辨識結果也可以發現，當使用者資料在訓練資料中時，其辨識效果優於使用者資料不在訓練資料中，因此，在語音情緒辨識實際使用時，如何修正訓練資料庫中的資料，藉以得到更貼近使用者各情緒特徵之新資料庫，並得到較佳的辨識效果將是語音情緒辨識另一個發展的重點所在，本章節首先將先針對 KNN 分類方法進行適應性實驗與分析，接著，針對本論文所提出之適應性 MLBN 語音情緒辨識以及適應性 MLBNC 語音情緒辨識方法與計算分別加以說明，最後，進行調適後 MLBN 與調適後 MLBNC 實驗與分析。

6.1 適應性 KNN 語音情緒辨識實驗與分析

由 4.1.1 第 k 個最近鄰(K-Nearest Neighbor, KNN)中所介紹之 KNN 分類方法，KNN 是根據資料庫中最接近測試資料之 K 個點資料所代表的類別來當成測試資料的辨識結果，因此，本章節所使用之適應性

KNN，其調適流程如圖 6-1 所示，當辨識結果正確時，KNN 資料庫不做任何更改變動，而當辨識結果錯誤時，則是將辨識錯誤之測試資料，於標示類別後加入訓練資料中，藉以擴充訓練資料以符合測試者實際的特徵狀態，以下將以圖 6-1 所示之適應性 KNN 分類方法針對不同語系之語音情緒辨識做調適，並驗證調適後之 KNN 辨識效果。

由於適應性方法是針對使用者個人做調適，因此，在適應性 KNN 語音情緒辨識實驗上，本論文將採用測試者個人邊調適邊辨識的方式進行，以下將就 KNN 調適與實驗過程分別加以說明，首先，在訓練資料中，我們以德國語料庫所有語料當作訓練資料，接著，將工研院語料(表 3-2)中不同測試者以個人方式，隨機取各情緒語料 15 筆，並以此當作調適語料，而剩餘之各情緒語料則當作測試語料，對調適後之 KNN 分類器做測試，以 P1 為例，其四種情緒語料各有 30 筆，從四種情緒語料中隨機各取 15 筆語料當作調適語料，也就是悲傷有 15 筆、中性 15 筆、快樂 15 筆、生氣 15 筆調適語料，另外，剩餘之四種情緒各 15 筆語料則為測試語料，對調適後之 KNN 分類器做調適後測試。

在 P1 KNN 分類器調適與調適後實驗的過程中，所使用的調適流程是將 P1 各情緒各 15 筆語料，共 60 筆語料，隨機取出一筆語料作測試，當辨識結果正確時，KNN 資料庫不做任何調適，並隨機取出

下一筆調適語料進行測試，而當調適語料辨識錯誤時，則將調適語料之特徵參數加入 KNN 分類器之資料庫中，也就是增加一筆資料庫資料，接著，再繼續從剩下的調適語料中隨機取下一筆調適測試語料，一直持續到 60 筆調適語料完全測試調適完為止，當所有語料測試調適完後，我們得到新的 KNN 資料庫，接著，以剩餘之四種情緒各 15 筆語料，共 60 筆測試語料，分別對未調適前及調適完後之 KNN 分類器做測試，表 6-1 為 P1 測試語料於調適前後之 KNN(K=11)鑑別矩陣，由辨識結果可以發現，未調適前，KNN 分類器對於悲傷、中性、快樂以及生氣四種情緒，絕大部分會被辨識為中性，辨識率僅有 25.00%，經調適後，快樂與生氣之辨識率上升到五成左右，而中性則下降到 66.67%，整體辨識正確率提升到 50.00%。

接著，與 P1 之調適測試實驗相同，同樣以德國語料庫所有資料當作訓練資料，分別對 P2 ~ P20 做 KNN 調適前後之辨識實驗，表 6-2 為 P1 ~ P20 之 KNN 分類器調適前後鑑別矩陣(P1 ~ P20 調適前後各情緒之辨識正確結果詳見附錄 C)，從鑑別矩陣結果可以發現，在調適前，除了悲傷有 16%被正確辨識為悲傷，快樂有 22.3%被正確辨識為快樂外，絕大部分情緒皆被辨識為中性，尤其是生氣情緒，有 92.7%被辨識為中性，調適前之整體辨識正確率為 34.17%，而調適過後，原先辨識結果集中在某一情緒的狀況有所改善，調適後主要分為兩個

族群，悲傷與中性為一族群，快樂與生氣則為另一個族群，悲傷、快樂及生氣之辨識正確率分別由 16.0%、22.3%、1.0% 提升至 73.7%、75.0% 以及 71.0%，悲傷有 26.3% 被辨識為中性，快樂有 15.7% 被辨識為生氣，生氣則有 14.3% 被辨識為快樂，而中性之辨識正確率則是下降，辨識正確率由 97.3% 下降至 75.0%，其中，有一部份原先辨識正確之中性情緒，在新 KNN 資料庫下，被辨識為悲傷，從鑑別矩陣來看，各情緒之辨識正確率在調適後趨於一致，顯示新資料庫對於各情緒之分類效果較趨穩定，調適後之整體辨識正確率為 73.67%，調適前後，整體辨識正確率從 34.17% 提升到 73.67%，因此，透過調適過程增加訓練資料數量以符合實際使用者的狀況，可以有效提升 KNN 分類器之整體辨識效果。

6.2 適應性 MLBN 語音情緒辨識實驗與分析

6.2.1 適應性 MLBN

適應性 MLBN 之調適流程如圖 6-2 所示，前半部之流程與 KNN 分類器調適流程一樣，將測試語音經正規化後之特徵參數，利用 MLBN 分類器計算各情緒發生之機率，當 MLBN 分類器辨識出來之情緒與調適測試語音之情緒相同時，MLBN 資料庫不做任何更改變動，而當辨識結果與調適測試語音之情緒不同時，前面所使用之 KNN 分類器調適過程是將測試資料加入訓練資料庫中，而本文提出之

MLBN 分類器調適過程，則是將 MLBN 分類器上各層各群之平均值與標準差做調適修正，產生新的 MLBN 資料庫，以下將就 MLBN 分類器資料庫各層各群之調適計算分別加以說明。

在適應性 MLBN 分類器調適的過程中，首先，將測試之情緒語音訊號，經計算所得之特徵參數 X 透過 MLBN 分類器分類後，當分類辨識之結果與測試語音之情緒不同時，則必須修正 MLBN 資料庫中各層各群之資料，在 MLBN 資料庫修正的過程中，首先，先考慮 MLBN 第一層第一種分群方式(如圖 6-3(a)所示)，假設未調適前，MLBN 第一層第一種分群方式資料庫中，各特徵參數各群之平均值與標準差分別為 $\bar{X}_{i,old}^{S,N}$ 、 $\bar{X}_{i,old}^{H,A}$ 、 $\sigma_{i,old}^{S,N}$ 以及 $\sigma_{i,old}^{H,A}$ ，其中， $i=18, 20$ ，由於辨識結果與測試語音之情緒不符，因此，必須對第一層第一種分群方式之特徵參數資料庫作修正，若測試語音之情緒為悲傷或中性，而 MLBN 辨識結果非悲傷或中性時，MLBN 第一層第一種分群方式中，第一群悲傷與中性之資料庫特徵參數修正式如式(6.2.1.1)與式(6.2.1.2)所示。

$$\bar{X}_{i,new}^{S,N} = \alpha_1 \bar{X}_{i,old}^{S,N} + (1 - \alpha_1) X_i$$

$$i = 18, 20 \quad (6.2.1)$$

$$\sigma_{i,new}^{S,N} = \sqrt{\alpha_1 \left[(\sigma_{i,old}^{S,N})^2 + (\bar{X}_{i,old}^{S,N} - \bar{X}_{i,new}^{S,N})^2 \right] + (1 - \alpha_1) (X_i - \bar{X}_{i,new}^{S,N})^2}$$

$$i = 18, 20 \quad (6.2.2)$$

其中， $\bar{X}_{i,old}^{S,N}$ 為未調適前，MLBN 資料庫第一層第一種分群方式

中，第一群悲傷與中性之特徵參數原始平均值， α_1 為適應性參數， $\bar{X}_{i,new}^{S,N}$ 為調適後，MLBN 資料庫第一層第一種分群方式中，第一群悲傷與中性之特徵參數新平均值，而 $\sigma_{i,old}^{S,N}$ 為未調適前，MLBN 資料庫第一層第一種分群方式中，第一群悲傷與中性之特徵參數原始標準差， $\sigma_{i,new}^{S,N}$ 則為調適後，MLBN 資料庫第一層第一種分群方式中，第一群悲傷與中性之特徵參數新標準差。

若測試語音之情緒為快樂或生氣，而 MLBN 辨識結果非快樂或生氣時，MLBN 第一層第一種分群方式中，第二群快樂與生氣之資料庫特徵參數修正式如式(6.2.3)與式(6.2.4)所示。

$$\bar{X}_{i,new}^{H,A} = \alpha_1 \bar{X}_{i,old}^{H,A} + (1 - \alpha_1) X_i$$

$$i = 18, 20 \quad (6.2.3)$$

$$\sigma_{i,new}^{H,A} = \sqrt{\alpha_1 \left[(\sigma_{i,old}^{H,A})^2 + (\bar{X}_{i,old}^{H,A} - \bar{X}_{i,new}^{H,A})^2 \right] + (1 - \alpha_1) (X_i - \bar{X}_{i,new}^{H,A})^2}$$

$$i = 18, 20 \quad (6.2.4)$$

其中， α_1 為適應性參數，由於第二群與第一群皆含有兩種情緒類別，因此，在資料庫特徵參數之修正上，第二群之適應性參數與第一群相同， $\bar{X}_{i,old}^{H,A}$ 為未調適前，MLBN 資料庫第一層第一種分群方式中，第二群快樂與生氣之特徵參數原始平均值， $\bar{X}_{i,new}^{H,A}$ 為調適後，MLBN 資料庫第一層第一種分群方式中，第二群快樂與生氣之特徵參數新平均值，而 $\sigma_{i,old}^{H,A}$ 為未調適前，MLBN 資料庫第一層第一種分群方式中，

第二群快樂與生氣之特徵參數原始標準差， $\sigma_{i,new}^{H,A}$ 則為調適後，MLBN 資料庫第一層第一種分群方式中，第二群快樂與生氣之特徵參數新標準差。

接著，考慮 MLBN 第一層第二種分群方式(如圖 6-3(b)所示)，假設未調適前，MLBN 第一層第二種分群方式資料庫中，各特徵參數各群之平均值與標準差分別為 $\bar{X}_{i,old}^S$ 、 $\bar{X}_{i,old}^{N,H,A}$ 、 $\sigma_{i,old}^S$ 以及 $\sigma_{i,old}^{N,H,A}$ ，其中， $i=4,7$ ，由於辨識結果與測試語音之情緒不符，因此，必須對第一層第二種分群方式之特徵參數資料庫作修正，當測試語音之情緒為悲傷，而 MLBN 辨識結果非悲傷時，MLBN 第一層第二種分群方式中，第一群悲傷之資料庫特徵參數修正式如式(6.2.5)與式(6.2.6)所示。

$$\bar{X}_{i,new}^S = \beta_1 \bar{X}_{i,old}^S + (1 - \beta_1) X_i$$

$$i = 4, 7 \tag{6.2.5}$$

$$\sigma_{i,new}^S = \sqrt{\beta_1 \left[(\sigma_{i,old}^S)^2 + (\bar{X}_{i,old}^S - \bar{X}_{i,new}^S)^2 \right] + (1 - \beta_1) (X_i - \bar{X}_{i,new}^S)^2}$$

$$i = 4, 7 \tag{6.2.6}$$

其中， $\bar{X}_{i,old}^S$ 為未調適前，MLBN 資料庫第一層第二種分群方式中，第一群悲傷之特徵參數原始平均值， β_1 為適應性參數， $\bar{X}_{i,new}^S$ 為調適後，MLBN 資料庫第一層第二種分群方式中，第一群悲傷之特徵參數新平均值，而 $\sigma_{i,old}^S$ 為未調適前，MLBN 資料庫第一層第二種分群方式中，第一群悲傷之特徵參數原始標準差， $\sigma_{i,new}^S$ 則為調適後，MLBN

資料庫第一層第二種分群方式中，第一群悲傷之特徵參數新標準差。

若測試語音之情緒為中性、快樂或生氣，而 MLBN 辨識結果非中性、快樂或生氣時，MLBN 第一層第二種分群方式中，第二群中性、快樂與生氣之資料庫特徵參數修正式如式(6.2.7)與式(6.2.8)所示。

$$\bar{X}_{i,new}^{N,H,A} = \gamma_1 \bar{X}_{i,old}^{N,H,A} + (1 - \gamma_1) X_i$$

$$i = 4, 7 \quad (6.2.7)$$

$$\sigma_{i,new}^{N,H,A} = \sqrt{\gamma_1 \left[(\sigma_{i,old}^{N,H,A})^2 + (\bar{X}_{i,old}^{N,H,A} - \bar{X}_{i,new}^{N,H,A})^2 \right] + (1 - \gamma_1) (X_i - \bar{X}_{i,new}^{N,H,A})^2}$$

$$i = 4, 7 \quad (6.2.8)$$

其中， γ_1 為適應性參數，由於第一層第二種分群方式中，第一群與第二群分別含有一種及三種情緒類別，與第一層第一種分群方式中，第一群與第二群各有兩種情緒類別不同，因此，在資料庫特徵參數之修正上，採用不同之適應性參數 β_1 及 γ_1 ， $\bar{X}_{i,old}^{N,H,A}$ 為未調適前，MLBN 資料庫第一層第二種分群方式中，第二群中性、快樂與生氣之特徵參數原始平均值， $\bar{X}_{i,new}^{N,H,A}$ 為調適後，MLBN 資料庫第一層第二種分群方式中，第二群中性、快樂與生氣之特徵參數新平均值，而 $\sigma_{i,old}^{N,H,A}$ 為未調適前，MLBN 資料庫第一層第二種分群方式中，第二群中性、快樂與生氣之特徵參數原始標準差， $\sigma_{i,new}^{N,H,A}$ 則為調適後，MLBN 資料庫第一層第二種分群方式中，第二群中性、快樂與生氣之特徵參數新標準差。

最後，考慮 MLBN 第一層第三種分群方式(如圖 6-3(c)所示)，假設未調適前，MLBN 第一層第三種分群方式資料庫中，資料庫中正規化後第一共振峰之平均值與標準差分別為 $\bar{X}_{5,old}^N$ 、 $\bar{X}_{5,old}^{S,H,A}$ 、 $\sigma_{5,old}^N$ 以及 $\sigma_{5,old}^{S,H,A}$ ，由於辨識結果與測試語音之情緒不符，因此，必須對第一層第三種分群方式之特徵參數資料庫作修正，當測試語音之情緒為中性，而 MLBN 辨識結果非中性時，MLBN 第一層第三種分群方式中，第一群中性之資料庫特徵參數修正式如式(6.2.9)與式(6.2.10)所示。

$$\bar{X}_{5,new}^N = \beta_1 \bar{X}_{5,old}^N + (1 - \beta_1) X_5 \quad (6.2.9)$$

$$\sigma_{5,new}^N = \sqrt{\beta_1 \left[(\sigma_{5,old}^N)^2 + (\bar{X}_{5,old}^N - \bar{X}_{5,new}^N)^2 \right] + (1 - \beta_1) (X_5 - \bar{X}_{5,new}^N)^2} \quad (6.2.10)$$

其中， $\bar{X}_{5,old}^N$ 為未調適前，MLBN 資料庫第一層第三種分群方式中，第一群中性之特徵參數原始平均值， $\bar{X}_{5,new}^N$ 為調適後，MLBN 資料庫第一層第三種分群方式中，第一群中性之特徵參數新平均值，而 $\sigma_{5,old}^N$ 為未調適前，MLBN 資料庫第一層第三種分群方式中，第一群中性之特徵參數原始標準差， $\sigma_{5,new}^N$ 則為調適後，MLBN 資料庫第一層第三種分群方式中，第一群中性之特徵參數新標準差。

若測試語音之情緒為悲傷、快樂或生氣，而 MLBN 辨識結果非悲傷、快樂或生氣時，第一層第三種分群方式中，第二群悲傷、快樂與生氣之資料庫特徵參數修正式如式(6.2.11)與式(6.2.12)所示。

$$\bar{X}_{5,new}^{S,H,A} = \gamma_1 \bar{X}_{5,old}^{S,H,A} + (1 - \gamma_1) X_5 \quad (6.2.11)$$

$$\sigma_{5,new}^{S,H,A} = \sqrt{\gamma_1 \left[(\sigma_{5,old}^{S,H,A})^2 + (\bar{X}_{5,old}^{S,H,A} - \bar{X}_{5,new}^{S,H,A})^2 \right] + (1-\gamma_1) (X_5 - \bar{X}_{5,new}^{S,H,A})^2} \quad (6.2.12)$$

其中，由於第一層第三種分群方式中，第一群與第二群分別含有一種及三種情緒類別，因此，在資料庫特徵參數之修正上，與第一層第二種分群方式採用相同之適應性參數 β_1 及 γ_1 ， $\bar{X}_{5,old}^{S,H,A}$ 為未調適前，MLBN 資料庫第一層第三種分群方式中，第二群悲傷、快樂與生氣之特徵參數原始平均值， $\bar{X}_{5,new}^{S,H,A}$ 為調適後，MLBN 資料庫第一層第三種分群方式中，第二群悲傷、快樂與生氣之特徵參數新平均值，而 $\sigma_{5,old}^{S,H,A}$ 為未調適前，MLBN 資料庫第一層第三種分群方式中，第二群悲傷、快樂與生氣之特徵參數原始標準差， $\sigma_{5,new}^{S,H,A}$ 則為調適後，MLBN 資料庫第一層第三種分群方式中，第二群悲傷、快樂與生氣之特徵參數新標準差。

接下來，考慮 MLBN 第二層，主要有兩種分群方式(如圖 5-11 所示)，首先，先考慮 MLBN 第二層第一種分群方式(如圖 6-4(a)所示)，假設未調適前，MLBN 第二層第一種分群方式資料庫中，各特徵參數各群之平均值與標準差分別為 $\bar{X}_{i,old}^S$ 、 $\bar{X}_{i,old}^N$ 、 $\bar{X}_{i,old}^{H,A}$ 、 $\sigma_{i,old}^S$ 、 $\sigma_{i,old}^N$ 以及 $\sigma_{i,old}^{H,A}$ ，其中， $i=1, 2, 6, 8, 9, 13, 15, 16, 17$ ，由於辨識結果與測試之語音情緒不符，因此，必須對第二層第一種分群之特徵資料庫作修正，若測試情緒為悲傷而 MLBN 辨識結果非悲傷時，MLBN 第二層第一種分群方式中，第一群悲傷之資料庫特徵參數修正式如式(6.2.13)與式

(6.2.14)所示。

$$\bar{X}_{i,new}^S = \beta_1 \bar{X}_{i,old}^S + (1 - \beta_1) X_i$$

$$i = 1, 2, 6, 8, 9, 13, 15, 16, 17 \quad (6.2.13)$$

$$\sigma_{i,new}^S = \sqrt{\beta_1 \left[(\sigma_{i,old}^S)^2 + (\bar{X}_{i,old}^S - \bar{X}_{i,new}^S)^2 \right] + (1 - \beta_1) (X_i - \bar{X}_{i,new}^S)^2}$$

$$i = 1, 2, 6, 8, 9, 13, 15, 16, 17 \quad (6.2.14)$$

其中， $\bar{X}_{i,old}^S$ 為未調適前，MLBN 資料庫第二層第一種分群方式中，第一群悲傷之特徵參數原始平均值， $\bar{X}_{i,new}^S$ 為調適後，MLBN 資料庫第二層第一種分群方式中，第一群悲傷之特徵參數新平均值，而 $\sigma_{i,old}^S$ 為未調適前，MLBN 資料庫第二層第一種分群方式中，第一群悲傷之特徵參數原始標準差， $\sigma_{i,new}^S$ 則為調適後，MLBN 資料庫第二層第一種分群方式中，第一群悲傷之特徵參數新標準差。

若測試情緒為中性而 MLBN 辨識結果非中性時，則 MLBN 第二層第一種分群方式中，第二群中性之資料庫特徵參數修正式如式 (6.2.15) 與式 (6.2.16) 所示。

$$\bar{X}_{i,new}^N = \beta_1 \bar{X}_{i,old}^N + (1 - \beta_1) X_i$$

$$i = 1, 2, 6, 8, 9, 13, 15, 16, 17 \quad (6.2.15)$$

$$\sigma_{i,new}^N = \sqrt{\beta_1 \left[(\sigma_{i,old}^N)^2 + (\bar{X}_{i,old}^N - \bar{X}_{i,new}^N)^2 \right] + (1 - \beta_1) (X_i - \bar{X}_{i,new}^N)^2}$$

$$i = 1, 2, 6, 8, 9, 13, 15, 16, 17 \quad (6.2.16)$$

其中， $\bar{X}_{i,old}^N$ 為未調適前，MLBN 資料庫第二層第一種分群方式

中，第二群中性之特徵參數原始平均值， $\bar{X}_{i,new}^N$ 為調適後，MLBN 資料庫第二層第一種分群方式中，第二群中性之特徵參數新平均值，而 $\sigma_{i,old}^N$ 為未調適前，MLBN 資料庫第二層第一種分群方式中，第二群中性之特徵參數原始標準差， $\sigma_{i,new}^N$ 為調適後，MLBN 資料庫第二層第一種分群方式中，第二群中性之特徵參數新標準差。

若測試情緒為快樂或生氣，而 MLBN 辨識結果非快樂與生氣時，則 MLBN 第二層第一種分群方式中，第三群快樂與生氣之資料庫特徵參數修正式如式(6.2.17)與式(6.2.18)所示。

$$\bar{X}_{i,new}^{H,A} = \alpha_1 \bar{X}_{i,old}^{H,A} + (1 - \alpha_1) X_i$$

$$i = 1, 2, 6, 8, 9, 13, 15, 16, 17 \quad (6.2.17)$$

$$\sigma_{i,new}^{H,A} = \sqrt{\alpha_1 \left[(\sigma_{i,old}^{H,A})^2 + (\bar{X}_{i,old}^{H,A} - \bar{X}_{i,new}^{H,A})^2 \right] + (1 - \alpha_1) (X_i - \bar{X}_{i,new}^{H,A})^2}$$

$$i = 1, 2, 6, 8, 9, 13, 15, 16, 17 \quad (6.2.18)$$

其中， $\bar{X}_{i,old}^{H,A}$ 為未調適前，MLBN 資料庫第二層第一種分群方式中，第三群快樂與生氣之特徵參數原始平均值， $\bar{X}_{i,new}^{H,A}$ 為調適後，MLBN 資料庫第二層第一種分群方式中，第三群快樂與生氣之特徵參數新平均值，而 $\sigma_{i,old}^{H,A}$ 為未調適前，MLBN 資料庫第二層第一種分群方式中，第三群快樂與生氣之特徵參數原始標準差， $\sigma_{i,new}^{H,A}$ 為調適後，MLBN 資料庫第二層第一種分群方式中，第三群快樂與生氣之特徵參數新標準差。

接著，考慮 MLBN 第二層第二種分群方式(如圖 6-4(b)所示)，假設未調適前，MLBN 資料庫中 MFCC C9 之平均值與標準差分別為 $\bar{X}_{19,old}^{S,N}$ 、 $\bar{X}_{19,old}^H$ 、 $\bar{X}_{19,old}^A$ 、 $\sigma_{19,old}^{S,N}$ 、 $\sigma_{19,old}^H$ 以及 $\sigma_{19,old}^A$ ，若測試情緒為悲傷或中性而 MLBN 辨識結果非悲傷與中性時，MLBN 第二層第二種分群方式中，第一群悲傷與中性之資料庫特徵參數修正式如式(6.2.19)與式(6.2.20)所示。

$$\bar{X}_{19,new}^{S,N} = \alpha_1 \bar{X}_{19,old}^{S,N} + (1 - \alpha_1) X_{19} \quad (6.2.19)$$

$$\sigma_{19,new}^{S,N} = \sqrt{\alpha_1 \left[(\sigma_{19,old}^{S,N})^2 + (\bar{X}_{19,old}^{S,N} - \bar{X}_{19,new}^{S,N})^2 \right] + (1 - \alpha_1) (X_{19} - \bar{X}_{19,new}^{S,N})^2} \quad (6.2.20)$$

其中， $\bar{X}_{19,old}^{S,N}$ 為未調適前，MLBN 資料庫中，悲傷與中性之 MFCC C9 原始平均值， $\bar{X}_{19,new}^{S,N}$ 為調適後，悲傷與中性之 MFCC C9 新平均值，而 $\sigma_{19,old}^{S,N}$ 為未調適前，悲傷與中性之 MFCC C9 原始標準差， $\sigma_{19,new}^{S,N}$ 則為調適後，悲傷與中性之 MFCC C9 新標準差。

若測試情緒為快樂而 MLBN 辨識結果非快樂時，MLBN 第二層第二種分群方式中，快樂之資料庫特徵參數修正式如式(6.2.21)與式(6.2.22)所示。

$$\bar{X}_{19,new}^H = \beta_1 \bar{X}_{19,old}^H + (1 - \beta_1) X_{19} \quad (6.2.21)$$

$$\sigma_{19,new}^H = \sqrt{\beta_1 \left[(\sigma_{19,old}^H)^2 + (\bar{X}_{19,old}^H - \bar{X}_{19,new}^H)^2 \right] + (1 - \beta_1) (X_{19} - \bar{X}_{19,new}^H)^2} \quad (6.2.22)$$

其中， $\bar{X}_{19,old}^H$ 為未調適前，MLBN 資料庫中，快樂之 MFCC C9 原

始平均值， $\bar{X}_{19,new}^H$ 為調適後，快樂之 MFCC C9 新平均值，而 $\sigma_{19,old}^H$ 為未調適前，快樂之 MFCC C9 原始標準差， $\sigma_{19,new}^H$ 為調適後，快樂之 MFCC C9 新標準差。

若測試情緒為生氣而 MLBN 辨識結果非生氣時，MLBN 第二層第二種分群方式中，生氣之資料庫特徵參數修正式如式(6.2.23)與式(6.2.24)所示。

$$\bar{X}_{19,new}^A = \beta_1 \bar{X}_{19,old}^A + (1 - \beta_1) X_{19} \quad (6.2.23)$$

$$\sigma_{19,new}^A = \sqrt{\beta_1 \left[(\sigma_{19,old}^A)^2 + (\bar{X}_{19,old}^A - \bar{X}_{19,new}^A)^2 \right] + (1 - \beta_1) (X_{19} - \bar{X}_{19,new}^A)^2} \quad (6.2.24)$$

其中， $\bar{X}_{19,old}^A$ 為未調適前，MLBN 資料庫中，生氣之 MFCC C9 原始平均值， $\bar{X}_{19,new}^A$ 為調適後，生氣之 MFCC C9 新平均值，而 $\sigma_{19,old}^A$ 為未調適前，MLBN 資料庫中，生氣之 MFCC C9 原始標準差， $\sigma_{19,new}^A$ 為調適後，生氣之 MFCC C9 新標準差。

最後，考慮 MLBN 第三層，如第五章之圖 5-13 所示，假設未調適前，MLBN 第三層資料庫中，各特徵參數之平均值與標準差分別為 $\bar{X}_{i,old}^S$ 、 $\bar{X}_{i,old}^N$ 、 $\bar{X}_{i,old}^H$ 、 $\bar{X}_{i,old}^A$ 、 $\sigma_{i,old}^S$ 、 $\sigma_{i,old}^N$ 、 $\sigma_{i,old}^H$ 以及 $\sigma_{i,old}^A$ ，其中， $i=3,10,11,12,14,21,22,23$ ，若測試情緒為悲傷而 MLBN 辨識結果非悲傷時，第三層中悲傷之資料庫特徵參數修正式如式(6.2.25)與式(6.2.26)所示。

$$\bar{X}_{i,new}^S = \beta_1 \bar{X}_{i,old}^S + (1 - \beta_1) X_i$$

$$i = 3, 10, 11, 12, 14, 21, 22, 23 \quad (6.2.25)$$

$$\sigma_{i,new}^S = \sqrt{\beta_1 \left[(\sigma_{i,old}^S)^2 + (\bar{X}_{i,old}^S - \bar{X}_{i,new}^S)^2 \right] + (1 - \beta_1) (X_i - \bar{X}_{i,new}^S)^2}$$

$$i = 3, 10, 11, 12, 14, 21, 22, 23 \quad (6.2.26)$$

其中， $\bar{X}_{i,old}^S$ 為未調適前，MLBN 資料庫第三層中，悲傷之特徵參數原始平均值， $\bar{X}_{i,new}^S$ 為調適後，悲傷之特徵參數新平均值，而 $\sigma_{i,old}^S$ 為未調適前，MLBN 資料庫第三層中，悲傷之特徵參數原始標準差， $\sigma_{i,new}^S$ 為調適後，悲傷之特徵參數新標準差。

若測試情緒為中性而 MLBN 辨識結果非中性時，MLBN 資料庫第三層中，中性之資料庫特徵參數修正式如式(6.2.27)與式(6.2.28)所示。

$$\bar{X}_{i,new}^N = \beta_1 \bar{X}_{i,old}^N + (1 - \beta_1) X_i$$

$$i = 3, 10, 11, 12, 14, 21, 22, 23 \quad (6.2.27)$$

$$\sigma_{i,new}^N = \sqrt{\beta_1 \left[(\sigma_{i,old}^N)^2 + (\bar{X}_{i,old}^N - \bar{X}_{i,new}^N)^2 \right] + (1 - \beta_1) (X_i - \bar{X}_{i,new}^N)^2}$$

$$i = 3, 10, 11, 12, 14, 21, 22, 23 \quad (6.2.28)$$

其中， $\bar{X}_{i,old}^N$ 為未調適前，MLBN 資料庫第三層中，中性之特徵參數原始平均值， $\bar{X}_{i,new}^N$ 為調適後，中性之特徵參數新平均值，而 $\sigma_{i,old}^N$ 為未調適前，MLBN 資料庫第三層中，中性之特徵參數原始標準差， $\sigma_{i,new}^N$ 為調適後，中性之特徵參數新標準差。

若測試情緒為快樂而 MLBN 辨識結果非快樂時，MLBN 資料庫

第三層中，快樂之資料庫特徵參數修正式如式(6.2.29)與式(6.2.30)所示。

$$\bar{X}_{i,new}^H = \beta_1 \bar{X}_{i,old}^H + (1 - \beta_1) X_i$$

$$i = 3, 10, 11, 12, 14, 21, 22, 23 \quad (6.2.29)$$

$$\sigma_{i,new}^H = \sqrt{\beta_1 \left[(\sigma_{i,old}^H)^2 + (\bar{X}_{i,old}^H - \bar{X}_{i,new}^H)^2 \right] + (1 - \beta_1) (X_i - \bar{X}_{i,new}^H)^2}$$

$$i = 3, 10, 11, 12, 14, 21, 22, 23 \quad (6.2.30)$$

其中， $\bar{X}_{i,old}^H$ 為未調適前，MLBN 資料庫第三層中，快樂之特徵參數原始平均值， $\bar{X}_{i,new}^H$ 為調適後，快樂之特徵參數新平均值，而 $\sigma_{i,old}^H$ 為未調適前，MLBN 資料庫第三層中，快樂之特徵參數原始標準差， $\sigma_{i,new}^H$ 為調適後，快樂之特徵參數新標準差。

若測試情緒為生氣而 MLBN 辨識結果非生氣時，MLBN 資料庫第三層中，生氣之資料庫特徵參數修正式如式(6.2.31)與式(6.2.32)所示。

$$\bar{X}_{i,new}^A = \beta_1 \bar{X}_{i,old}^A + (1 - \beta_1) X_i$$

$$i = 3, 10, 11, 12, 14, 21, 22, 23 \quad (6.2.31)$$

$$\sigma_{i,new}^A = \sqrt{\beta_1 \left[(\sigma_{i,old}^A)^2 + (\bar{X}_{i,old}^A - \bar{X}_{i,new}^A)^2 \right] + (1 - \beta_1) (X_i - \bar{X}_{i,new}^A)^2}$$

$$i = 3, 10, 11, 12, 14, 21, 22, 23 \quad (6.2.32)$$

其中， $\bar{X}_{i,old}^A$ 為未調適前，MLBN 資料庫第三層中，生氣之特徵參數原始平均值， $\bar{X}_{i,new}^A$ 為調適後，生氣之特徵參數新平均值，而 $\sigma_{i,old}^A$ 為

未調適前，MLBN 資料庫第三層中，生氣之特徵參數原始標準差， $\sigma_{i,new}^A$ 為調適後，生氣之特徵參數新標準差。

適應性 MLBN 資料庫修正式(6.2.1) ~ (6.2.32)中，有三個適應性參數，分別為 α_1 、 β_1 及 γ_1 ，此三個適應性參數分別對應 MLBN 資料庫各分群中含有之情緒類別數目，當此分群中含有兩個情緒類別數目時，適應性參數使用 α_1 ，當此分群中含有一個情緒類別數目時，適應性參數使用 β_1 ，而當此分群中含有三個情緒類別數目時，適應性參數則使用 γ_1 。 α_1 、 β_1 及 γ_1 的另外一個含意為資料庫中特徵參數之修正速度，當 α_1 、 β_1 及 γ_1 越小，資料庫之修正值越趨近於測試之特徵參數值，資料庫之修正速度越快，但由於個人之情緒反應計算所得之特徵參數並非固定值，其本身亦為一分佈狀態，因此，適當的適應性參數，將有助於適應性 MLBN 分類器的穩定性。舉例來說，假設原資料庫特徵參數之平均值與標準差分別為 155 及 30，當測試者有一百筆測試資料，其測試資料之平均值與標準差分別為 220 及 20 時，將一百筆測試資料分別以不同之適應性參數 α 對原資料庫作調適，圖 6-5 及圖 6-6 分別為 $\alpha = 0.5$ 及 $\alpha = 0.9$ 時，其資料庫平均值與標準差之變化，由圖中可以發現，當 $\alpha = 0.5$ 時，資料庫調適之速度較快達到測試者真實的狀況，但調適後之平均值與標準差較不穩定，相對的，當 $\alpha = 0.9$ 時，雖然資料庫調適修正之速度較慢，但調適後之平均直與標準差較

為穩定，因此，適應性參數的選擇，將對適應性 MLBN 分類器的效果有很大的影響。在後續之適應性 MLBN 實驗中，我們將固定適應性參數 $\alpha_1 = 0.9$ 、 $\beta_1 = 0.8$ 以及 $\gamma_1 = 0.95$ ，以利適應性 MLBN 之實驗進行與分析。

經由適應性 MLBN 修正後，MLBN 資料庫中各情緒之特徵參數平均值與標準差將會依照使用者之情緒狀態作調整，調適次數越多，則新 MLBN 資料庫將更貼近實際測試使用者的真實狀態，以下將就適應性 MLBN 分類器進行同語系與不同語系之適應性實驗，並就實驗結果分別加以說明。

6.2.2 適應性 MLBN 實驗結果與分析

由於適應性 MLBN 調適方式是依據測試者個人的情緒分佈狀態去做調適修正，因此，在本節適應性 MLBN 實驗中，將採用 MLBN Outside Test 方式進行辨識，首先，進行同語系適應性 MLBN 實驗，以德國語料庫當作適應性 MLBN 實驗語料，將德國語料庫 P1 語料當作測試語料，並以 P2 ~ P10 語料當作訓練語料，因此悲傷有 7 筆、中性有 11 筆、快樂有 7 筆、生氣有 14 筆，共 39 筆測試語料，而訓練語料悲傷有 55 筆、中性有 68 筆、快樂有 64 筆、生氣有 113 筆，接著，從 39 筆測試語料中隨機選取一筆資料並以 MLBN 語音情緒分類器進行辨識，若辨識結果正確，則從剩餘之測試資料中隨機選取另一

筆資料進行測試，若辨識結果錯誤，則根據上一節所提出之適應性 MLBN 資料庫修正方法，對 MLBN 資料庫進行調適修正，修正完後，再從剩餘之測試資料中隨機選取另一筆資料，並以修正過後之新 MLBN 資料庫進行測試，直到 39 筆測試語料測試完畢，最後，將測試完畢並修正過後之新 MLBN 分類器對 39 筆測試資料重新進行測試。

接著，以 P2 當作測試資料，另外 9 人(P1 以及 P3 ~ P10)則當作訓練資料，從 P2 測試資料中隨機選取一筆資料並以 MLBN 分類器進行辨識，與 P1 之調適過程相同，若辨識結果正確，則從剩餘之測試資料中隨機選取另一筆資料進行測試，若辨識結果錯誤，則同樣根據適應性 MLBN 資料庫修正方法，對 MLBN 資料庫進行調適修正，修正完後，再從剩餘之測試資料中隨機選取另一筆資料，並以修正過後之新 MLBN 資料庫進行測試，直到 P2 所有測試資料測試完畢，最後，將測試完畢並修正過後之新 MLBN 分類器對 P2 所有資料進行測試。以此類推，對 P3 ~ P10 做 MLBN Outside Test，並進行 MLBN 資料庫調適與測試。

表 6-3 為 MLBN 分類器德國語料庫 P1 ~ P10 同語系調適前後之鑑別矩陣(P1 ~ P10 調適前後各情緒之辨識正確結果詳見附錄 C)，從整體辨識結果來看，MLBN 分類器在調適過後，悲傷之辨識正確率從

84%提升到 99%，中性則維持極佳的辨識效果，從 97%提升到 99%，快樂之辨識效果則是大幅提升，辨識正確率從 54%提升到 83%，至於生氣，辨識正確率則是從 75.6%提升至 90.6%，整體之辨識正確率則是從調適前 77.6%提升到 92.3%。表 6-4 為 P1 ~ P10 以個人本身之語料為訓練與測試語料之 MLBN Dependent Test 鑑別矩陣，由調適後整體辨識正確率 92.3%與個人語料訓練與測試之整體辨識正確率 93.5%比較可以發現，兩者相當接近，因此，經由適應性 MLBN 調適過程，可以有效縮小資料庫與語者之間的差異，提升個人本身的情緒辨識效果。

接下來，考慮適應性 MLBN 應用於不同語系時，其調適過程對辨識結果之影響，同樣的，以德國語料庫所有資料當作訓練資料，並以工業技術研究院所錄製之語料當作測試語料，測試適應性 MLBN 在不同語系間之效果。在調適與測試的過程中，與 KNN 採用相同的調適與測試語料，以 P1 為例，各情緒各有 15 筆，共 60 筆語料為調適語料，首先，從 60 筆調適語料中隨機選取一筆語料並以 MLBN 分類器進行辨識，若辨識結果正確，則從剩餘調適語料中隨機選取另一筆調適語料進行測試，若辨識結果錯誤，則根據前面所提出之 MLBN 資料庫修正方式，對 MLBN 資料庫進行修正，當修正完後，再從剩餘調適語料中隨機選取另一筆調適語料，並以修正過後之新 MLBN

資料庫進行測試，直到 60 筆調適語料測試調適完畢，最後，將 P1 剩餘之四種情緒各 15 筆，共 60 筆測試語料，分別對未調適前及調適完後之 MLBN 分類器做測試。

接著，與工業技術研究院語料 P1 之調適與測試流程相同，並採用與 KNN 相同之工業技術研究院 P2 至 P20 調適與測試語料，對 P2 至 P20 進行 MLBN 資料庫調適與測試，表 6-5 為 MLBN 分類器於工業技術研究院語料 P1 ~ P20 調適前後整體統計之鑑別矩陣(P1 ~ P20 調適前後各情緒之辨識正確結果詳見附錄 C)，從鑑別矩陣來看，在調適前，悲傷與中性會被 MLBN 分類器視為同一情緒類別，悲傷與中性之情緒絕大部分會被 MLBN 辨識為中性，悲傷情緒中有一小部份會被正確辨識為悲傷，而快樂之辨識結果較為分散，有一部份被辨識為悲傷，一部分被辨識為快樂，另外，分別有一小部分被辨識為中性及生氣，生氣則視大部分會被 MLBN 分類器辨識為悲傷，一部分被辨識為中性及快樂，少數則被辨識為生氣，而在調適過後，悲傷之辨識率從 20.3% 大幅提升到 86.0%，中性之辨識正確率則是下降，從 95.0% 下降到 85.0%，而快樂與生氣之辨識效果在調適過後，均大幅提升，辨識率分別從 31.3% 提升到 83.3% 以及 4.7% 提升到 75.3%，整體之辨識正確率則是從調適前 37.83% 大幅提升至 82.42%，由鑑別矩陣結果也可以發現，經由適應性 MLBN 調適過程，可以使各情緒之

辨識正確率接近，各情緒類別之分類效果達到均衡狀態。

最後，我們增加適應性 MLBN 調適語料數量，用以確定調適過程增加時，是否對適應性 MLBN 之分類有所助益，首先，先以工研院 P1 所有語料當作調適語料，因此悲傷有 30 筆、中性有 30 筆、快樂有 30 筆、生氣有 30 筆，共 120 筆調適語料，接著，與前面之測試調適過程相同，從 120 筆調適語料中隨機選取一筆資料並以 MLBN 分類器進行辨識，若辨識結果正確，則再從剩餘測試資料中隨機選取另一筆資料進行測試，若辨識結果錯誤，則根據 6.2.1 所提出之 MLBN 資料庫修正方法，對 MLBN 資料庫進行修正，修正完後，再從剩餘測試資料中隨機選取另一筆資料，並以修正過後之新 MLBN 資料庫進行測試調適，直到 120 筆測試語料測試調適完畢，最後，將調適完畢並修正過後之新 MLBN 分類器對 P1 所有資料進行測試。

接著，分別對 P2 至 P20 個別語料進行調適與測試，表 6-6 為調適語料增加後，MLBN 分類器於工業技術研究院語料 P1 ~ P20 調適前後整體統計之鑑別矩陣(P1 ~ P20 調適前後各情緒之辨識正確結果詳見附錄 C)，從鑑別矩陣來看，在調適前，其辨識結果與前面取一半語料調適另一半語料測試之結果相近，悲傷與中性會被 MLBN 分類器視為同一情緒類別，悲傷與中性之情緒絕大部分會被 MLBN 辨識為中性，悲傷情緒中有一小部分會被正確辨識為悲傷，而快樂之辨識

結果較為分散，有一部份被辨識為悲傷，一部分被辨識為快樂，另外，分別有一小部分被辨識為中性及生氣，生氣則視大部分會被 MLBN 分類器辨識為悲傷，一部分被辨識為中性及快樂，少數則被辨識為生氣，而在調適過後，悲傷之辨識率從 20.7% 大幅提升到 91.0%，中性之辨識正確率則是些微下降，從 96.0% 下降到 91.8%，而快樂與生氣之辨識效果在調適過後，則是大幅提升，辨識率分別從 35.0% 提升到 88.7% 以及 5.7% 提升到 84.2%，整體之辨識正確率則是從調適前 39.33% 大幅提升至 88.92%，與表 6-5 比較可以發現，在調適語料增加後，各情緒之整體辨識正確率皆提升約 6~8%，顯示增加調適次數將有助於提升 MLBN 語音情緒之分類效果。

由以上之實驗結果，當 MLBN 資料庫訓練語料語系與使用者使用之語系不同時，經由本論文所提出之適應性 MLBN 調適過程，將可以有效提升各情緒之辨識效果，而增加調適次數，對於 MLBN 語音情緒之分類效果也是有所助益的，此外，在適應性 MLBN 資料庫之修正與計算上，也僅需修正 MLBN 資料庫中各特徵參數之平均值與標準差，不需再額外加入任何資料，增加實際應用時硬體上之需求，因此，適應性 MLBN 大大提升了語音情緒辨識在硬體架構及實際應用上的可行性。

6.3 適應性 MLBN 語音情緒辨識實驗與分析

6.3.1 適應性 MLBNC

適應性 MLBNC 之調適流程如圖 6-7 所示，主要之流程與適應性 MLBN 分類器之調適流程類似，首先，將測試語音經正規化後之特徵參數，利用 MLBNC 分類器計算各情緒發生之機率，當 MLBNC 分類器辨識出來之情緒與調適測試之語音情緒相同時，MLBNC 資料庫不做任何更改變動，而當辨識結果與調適測試之語音情緒不同時，前面所提之 MLBN 分類器的調適過程是將 MLBN 上各層各群之平均值與標準差做調適修正，而本節所使用之適應性 MLBNC 則是將各層各群之特徵參數向量平均值與共變異數做調適修正，以下將就 MLBNC 分類器資料庫各層各群之調適計算分別加以說明。

在適應性 MLBNC 分類器調適的過程中，首先，將測試之情緒語音訊號，經計算所得之特徵參數 X 透過 MLBNC 分類器分類後，當分類辨識之結果與測試語音之情緒不同時，則必須修正 MLBNC 資料庫中各層各群之資料，在 MLBNC 資料庫修正的過程中，首先，先考慮 MLBNC 第一層第一種分群方式(如圖 6-8(a)所示)，假設未調適前，MLBNC 資料庫中各群之特徵參數平均值向量與共變異數分別為

$\mu_{1,1,old}^1$ 、 $\mu_{1,2,old}^1$ 、 $\Sigma_{1,1,old}^1$ 以及 $\Sigma_{1,2,old}^1$ ，其中

$$\mu_{1,1,old}^1 = \begin{bmatrix} \bar{X}_{18,old}^{S,N} & \bar{X}_{20,old}^{S,N} \end{bmatrix}^T \quad (6.3.1)$$

$$\mu_{1,2,old}^1 = \begin{bmatrix} \bar{X}_{18,old}^{H,A} & \bar{X}_{20,old}^{H,A} \end{bmatrix}^T \quad (6.3.2)$$

而 $\Sigma_{1,1,old}^1$ 及 $\Sigma_{1,2,old}^1$ 則分別一 2×2 之陣列。

由於辨識結果與測試語音之情緒不符，因此，必須對 MLBNC 第一層第一種分群方式之資料庫作修正，根據 5.3.1 之 MLBNC 推導，第一層第一種分群方式所使用之測試特徵參數向量為 $\mathbf{R}_1 = [X_{18} \quad X_{20}]^T$ ，當測試語音之情緒為悲傷或中性，而 MLBNC 辨識結果非悲傷或中性時，MLBNC 第一層第一種分群方式中，第一群悲傷與中性之資料庫特徵參數平均值向量與共變異數之修正式如式 (6.3.3) 與式 (6.3.4) 所示。

$$\boldsymbol{\mu}_{1,1,new}^1 = \alpha_2 \boldsymbol{\mu}_{1,1,old}^1 + (1 - \alpha_2) \mathbf{R}_1 \quad (6.3.3)$$

$$\begin{aligned} \Sigma_{1,1,new}^1 = & \alpha_2 \left[\Sigma_{1,1,old}^1 + (\boldsymbol{\mu}_{1,1,new}^1 - \boldsymbol{\mu}_{1,1,old}^1)(\boldsymbol{\mu}_{1,1,new}^1 - \boldsymbol{\mu}_{1,1,old}^1)^T \right] \\ & + (1 - \alpha_2) (\boldsymbol{\mu}_{1,1,new}^1 - \mathbf{R}_1)(\boldsymbol{\mu}_{1,1,new}^1 - \mathbf{R}_1)^T \end{aligned} \quad (6.3.4)$$

其中， $\boldsymbol{\mu}_{1,1,old}^1$ 為未調適前，MLBNC 資料庫第一層第一種分群方式中，第一群悲傷與中性之特徵參數原始平均值向量， α_2 為適應性參數， $\boldsymbol{\mu}_{1,1,new}^1$ 為調適後，MLBNC 資料庫第一層第一種分群方式中，第一群悲傷與中性之特徵參數新平均值向量，而 $\Sigma_{1,1,old}^1$ 為未調適前，MLBNC 資料庫第一層第一種分群方式中，第一群悲傷與中性之特徵參數原始共變異數， $\Sigma_{1,1,new}^1$ 則為調適後，MLBNC 資料庫第一層第一種分群方式中，第一群悲傷與中性之特徵參數新共變異數。

若測試語音之情緒為快樂或生氣，而 MLBNC 辨識結果非快樂或生氣時，MLBNC 第一層第一種分群方式中，第二群快樂與生氣之資

料庫特徵參數平均值向量與共變異數之修正式如式(6.3.5)與式(6.3.6)

所示。

$$\boldsymbol{\mu}_{1,2,new}^1 = \alpha_2 \boldsymbol{\mu}_{1,2,old}^1 + (1 - \alpha_2) \mathbf{R}_1 \quad (6.3.5)$$

$$\begin{aligned} \Sigma_{1,2,new}^1 = & \alpha_2 \left[\Sigma_{1,2,old}^1 + (\boldsymbol{\mu}_{1,2,new}^1 - \boldsymbol{\mu}_{1,2,old}^1)(\boldsymbol{\mu}_{1,2,new}^1 - \boldsymbol{\mu}_{1,2,old}^1)^T \right] \\ & + (1 - \alpha_2) (\boldsymbol{\mu}_{1,2,new}^1 - \mathbf{R}_1)(\boldsymbol{\mu}_{1,2,new}^1 - \mathbf{R}_1)^T \end{aligned} \quad (6.3.6)$$

其中， $\boldsymbol{\mu}_{1,2,old}^1$ 為未調適前，MLBNC 資料庫第一層第一種分群方式中，第二群快樂與生氣之特徵參數原始平均值向量， $\boldsymbol{\mu}_{1,2,new}^1$ 為調適後，MLBNC 資料庫第一層第一種分群方式中，第二群快樂與生氣之特徵參數新平均值向量，而 $\Sigma_{1,2,old}^1$ 為未調適前，MLBNC 資料庫第一層第一種分群方式中，第二群快樂與生氣之特徵參數原始共變異數， $\Sigma_{1,2,new}^1$ 則為調適後，MLBNC 資料庫第一層第一種分群方式中，第二群快樂與生氣之特徵參數新共變異數。

接著，考慮 MLBNC 第一層第二種分群方式(如圖 6-8(b)所示)，假設未調適前，MLBNC 資料庫中各群之特徵參數平均值向量與共變異數分別為 $\boldsymbol{\mu}_{2,1,old}^1$ 、 $\boldsymbol{\mu}_{2,2,old}^1$ 、 $\Sigma_{2,1,old}^1$ 以及 $\Sigma_{2,2,old}^1$ ，其中

$$\boldsymbol{\mu}_{2,1,old}^1 = \left[\bar{X}_{4,old}^S \quad \bar{X}_{7,old}^S \right]^T \quad (6.3.7)$$

$$\boldsymbol{\mu}_{2,2,old}^1 = \left[\bar{X}_{4,old}^{N,H,A} \quad \bar{X}_{7,old}^{N,H,A} \right]^T \quad (6.3.8)$$

而 $\Sigma_{2,1,old}^1$ 及 $\Sigma_{2,2,old}^1$ 則分別一 2×2 之陣列。

由於辨識結果與測試語音之情緒不符，因此，必須對 MLBNC 第一層第二種分群方式之資料庫作修正，根據 5.3.1 之 MLBNC 推導，

第一層第二種分群方式所使用之測試特徵參數向量為 $\mathbf{R}_2 = [X_4 \ X_7]^T$ ，當測試語音之情緒為悲傷，而 MLBNC 辨識結果非悲傷時，MLBNC 第一層第二種分群方式中，第一群悲傷之資料庫特徵參數平均值向量與共變異數之修正式分別如式(6.3.9)與式(6.3.10)所示。

$$\boldsymbol{\mu}_{2,1,new}^1 = \beta_2 \boldsymbol{\mu}_{2,1,old}^1 + (1 - \beta_2) \mathbf{R}_2 \quad (6.3.9)$$

$$\begin{aligned} \Sigma_{2,1,new}^1 = & \beta_2 \left[\Sigma_{2,1,old}^1 + (\boldsymbol{\mu}_{2,1,new}^1 - \boldsymbol{\mu}_{2,1,old}^1)(\boldsymbol{\mu}_{2,1,new}^1 - \boldsymbol{\mu}_{2,1,old}^1)^T \right] \\ & + (1 - \beta_2) (\boldsymbol{\mu}_{2,1,new}^1 - \mathbf{R}_2)(\boldsymbol{\mu}_{2,1,new}^1 - \mathbf{R}_2)^T \end{aligned} \quad (6.3.10)$$

其中， $\boldsymbol{\mu}_{2,1,old}^1$ 為未調適前，MLBNC 資料庫第一層第二種分群方式中，第一群悲傷之特徵參數原始平均值向量， β_2 為適應性參數， $\boldsymbol{\mu}_{2,1,new}^1$ 為調適後，MLBNC 資料庫第一層第二種分群方式中，第一群悲傷之特徵參數新平均值向量，而 $\Sigma_{2,1,old}^1$ 為未調適前，MLBNC 資料庫第一層第二種分群方式中，第一群悲傷之特徵參數原始共變異數， $\Sigma_{2,1,new}^1$ 則為調適後，MLBNC 資料庫第一層第二種分群方式中，第一群悲傷之特徵參數新共變異數。

若測試語音之情緒為中性、快樂或生氣，而 MLBNC 辨識結果非中性、快樂或生氣時，MLBNC 第一層第二種分群方式中，第二群中性、快樂與生氣之資料庫特徵參數平均值向量與共變異數之修正式分別如式(6.3.11)與式(6.3.12)所示。

$$\boldsymbol{\mu}_{2,2,new}^1 = \gamma_2 \boldsymbol{\mu}_{2,2,old}^1 + (1 - \gamma_2) \mathbf{R}_2 \quad (6.3.11)$$

$$\begin{aligned} \Sigma_{2,2,new}^1 = & \gamma_2 \left[\Sigma_{2,2,old}^1 + (\boldsymbol{\mu}_{2,2,new}^1 - \boldsymbol{\mu}_{2,2,old}^1)(\boldsymbol{\mu}_{2,2,new}^1 - \boldsymbol{\mu}_{2,2,old}^1)^T \right] \\ & + (1 - \gamma_2)(\boldsymbol{\mu}_{2,2,new}^1 - \mathbf{R}_2)(\boldsymbol{\mu}_{2,2,new}^1 - \mathbf{R}_2)^T \end{aligned} \quad (6.3.12)$$

與調適性 MLBN 適應性參數相同，由於第一層第二種分群方式中，第一群與第二群分別含有一種及三種情緒類別，因此，在資料庫特徵參數之修正上，採用與第一層第一種分群方式不同之適應性參數 β_1 及 γ_1 ， $\boldsymbol{\mu}_{2,2,old}^1$ 為未調適前，MLBNC 資料庫第一層第二種分群方式中，第二群中性、快樂與生氣之特徵參數原始平均值向量， $\boldsymbol{\mu}_{2,2,new}^1$ 為調適後，MLBNC 資料庫第一層第二種分群方式中，第二群中性、快樂與生氣之特徵參數新平均值向量，而 $\Sigma_{2,2,old}^1$ 為未調適前，MLBNC 資料庫第一層第二種分群方式中，第二群中性、快樂與生氣之特徵參數原始共變異數， $\Sigma_{2,2,new}^1$ 則為調適後，MLBNC 資料庫第一層第二種分群方式中，第二群中性、快樂與生氣之特徵參數新共變異數。

最後，考慮 MLBNC 第一層第三種分群方式，由於只有一個特徵參數，與 MLBN 第一層第三種分群方式(圖 6-3(c))相同，因此，適應性 MLBNC 第一層第三種分群方式之修正與 MLBN 第一層第三種分群方式之修正相同，其資料庫特徵參數修正式分別如式(6.2.9) ~ (6.2.12)所示。

接下來，考慮 MLBNC 第二層，主要有兩種分群方式(如圖 5-16 所示)，首先，先考慮 MLBNC 第二層第一種分群方式(如圖 6-9 所示)，假設未調適前，MLBNC 資料庫中各群之特徵參數平均值向量與共變

異數分別為 $\mu_{1,1,old}^2$ 、 $\mu_{1,2,old}^2$ 、 $\mu_{1,3,old}^2$ 、 $\Sigma_{1,1,old}^2$ 、 $\Sigma_{1,2,old}^2$ 以及 $\Sigma_{1,3,old}^2$ ，其中 $\Sigma_{1,1,old}^2$ 、

$\Sigma_{1,2,old}^2$ 以及 $\Sigma_{1,3,old}^2$ 分別為 9×9 之陣列，

$$\mu_{1,1,old}^2 = \left[\bar{X}_{1,old}^S \quad \bar{X}_{2,old}^S \quad \bar{X}_{6,old}^S \quad \bar{X}_{8,old}^S \quad \bar{X}_{9,old}^S \quad \bar{X}_{13,old}^S \quad \bar{X}_{15,old}^S \quad \bar{X}_{16,old}^S \quad \bar{X}_{17,old}^S \right]^T \quad (6.3.13)$$

$$\mu_{1,2,old}^2 = \left[\bar{X}_{1,old}^N \quad \bar{X}_{2,old}^N \quad \bar{X}_{6,old}^N \quad \bar{X}_{8,old}^N \quad \bar{X}_{9,old}^N \quad \bar{X}_{13,old}^N \quad \bar{X}_{15,old}^N \quad \bar{X}_{16,old}^N \quad \bar{X}_{17,old}^N \right]^T \quad (6.3.14)$$

$$\mu_{1,3,old}^2 = \left[\bar{X}_{1,old}^{H,A} \quad \bar{X}_{2,old}^{H,A} \quad \bar{X}_{6,old}^{H,A} \quad \bar{X}_{8,old}^{H,A} \quad \bar{X}_{9,old}^{H,A} \quad \bar{X}_{13,old}^{H,A} \quad \bar{X}_{15,old}^{H,A} \quad \bar{X}_{16,old}^{H,A} \quad \bar{X}_{17,old}^{H,A} \right]^T \quad (6.3.15)$$

由於辨識結果與測試語音之情緒不符，因此，必須對 MLBNC 第二層第一種分群方式之資料庫作修正，根據 5.3.1 之 MLBNC 推導，第二層第一種分群方式所使用之測試特徵參數向量為 $\mathbf{U} = [X_1 \quad X_2 \quad X_6 \quad X_8 \quad X_9 \quad X_{13} \quad X_{15} \quad X_{16} \quad X_{17}]^T$ ，若測試語音之情緒為悲傷，而 MLBNC 辨識結果非悲傷時，MLBNC 第二層第一種分群方式中，第一群悲傷之資料庫特徵參數平均值向量與共變異數之修正式分別如式(6.3.16)與式(6.3.17)所示。

$$\mu_{1,1,new}^2 = \beta_2 \mu_{1,1,old}^2 + (1 - \beta_2) \mathbf{U} \quad (6.3.16)$$

$$\Sigma_{1,1,new}^2 = \beta_2 \left[\Sigma_{1,1,old}^2 + (\mu_{1,1,new}^2 - \mu_{1,1,old}^2)(\mu_{1,1,new}^2 - \mu_{1,1,old}^2)^T \right] + (1 - \beta_2)(\mu_{1,1,new}^2 - \mathbf{U})(\mu_{1,1,new}^2 - \mathbf{U})^T \quad (6.3.17)$$

其中， $\mu_{1,1,old}^2$ 為未調適前，MLBNC 資料庫第二層第一種分群方式中，第一群悲傷之特徵參數原始平均值向量， $\mu_{1,1,new}^2$ 為調適後，MLBNC 資料庫第二層第一種分群方式中，第一群悲傷之特徵參數新平均值向

量，而 $\Sigma_{1,1,old}^2$ 為未調適前，MLBNC 資料庫第二層第一種分群方式中，第一群悲傷之特徵參數原始共變異數， $\Sigma_{1,1,new}^2$ 則為調適後，MLBNC 資料庫第二層第一種分群方式中，第一群悲傷之特徵參數新共變異數。

若測試情緒為中性而 MLBNC 辨識結果非中性時，則 MLBNC 第二層第一種分群方式中，第二群中性之資料庫特徵參數修正式如式(6.3.18)與式(6.3.19)所示。

$$\mu_{1,2,new}^2 = \beta_2 \mu_{1,2,old}^2 + (1 - \beta_2) \mathbf{U} \quad (6.3.18)$$

$$\Sigma_{1,2,new}^2 = \beta_2 \left[\Sigma_{1,2,old}^2 + (\mu_{1,2,new}^2 - \mu_{1,2,old}^2)(\mu_{1,2,new}^2 - \mu_{1,2,old}^2)^T \right] + (1 - \beta_2)(\mu_{1,2,new}^2 - \mathbf{U})(\mu_{1,2,new}^2 - \mathbf{U})^T \quad (6.3.19)$$

其中， $\mu_{1,2,old}^2$ 為未調適前，MLBNC 資料庫第二層第一種分群方式中，第二群中性之特徵參數原始平均值向量， $\mu_{1,2,new}^2$ 為調適後，MLBNC 資料庫第二層第一種分群方式中，第二群中性之特徵參數新平均值向量，而 $\Sigma_{1,2,old}^2$ 為未調適前，MLBNC 資料庫第二層第一種分群方式中，第二群中性之特徵參數原始共變異數， $\Sigma_{1,2,new}^2$ 則為調適後，MLBNC 資料庫第二層第一種分群方式中，第二群中性之特徵參數新共變異數。

若測試情緒為快樂或生氣而 MLBNC 辨識結果非快樂或生氣時，則 MLBNC 第二層第一種分群方式中，第三群快樂與生氣之資料庫特徵參數修正式如式(6.3.20)與式(6.3.21)所示。

$$\boldsymbol{\mu}_{1,3,new}^2 = \alpha_2 \boldsymbol{\mu}_{1,3,old}^2 + (1 - \alpha_2) \mathbf{U} \quad (6.3.20)$$

$$\begin{aligned} \Sigma_{1,3,new}^2 = & \alpha_2 \left[\Sigma_{1,3,old}^2 + (\boldsymbol{\mu}_{1,3,new}^2 - \boldsymbol{\mu}_{1,3,old}^2)(\boldsymbol{\mu}_{1,3,new}^2 - \boldsymbol{\mu}_{1,3,old}^2)^T \right] \\ & + (1 - \alpha_2)(\boldsymbol{\mu}_{1,3,new}^2 - \mathbf{U})(\boldsymbol{\mu}_{1,3,new}^2 - \mathbf{U})^T \end{aligned} \quad (6.3.21)$$

其中， $\boldsymbol{\mu}_{1,3,old}^2$ 為未調適前，MLBNC 資料庫第二層第一種分群方式中，第三群快樂與生氣之特徵參數原始平均值向量， $\boldsymbol{\mu}_{1,3,new}^2$ 為調適後，MLBNC 資料庫第二層第一種分群方式中，第三群快樂與生氣之特徵參數新平均值向量，而 $\Sigma_{1,3,old}^2$ 為未調適前，MLBNC 資料庫第二層第一種分群方式中，第三群快樂與生氣之特徵參數原始共變異數， $\Sigma_{1,3,new}^2$ 則為調適後，MLBNC 資料庫第二層第一種分群方式中，第三群快樂與生氣之特徵參數新共變異數。

接著，考慮 MLBNC 第二層第二種分群方式，由於只有一個特徵參數 MFCC C9，與 MLBN 第二層第二種分群方式(圖 6-4(b))相同，因此，適應性 MLBNC 第二層第二種分群方式之修正與 MLBN 第二層第二種分群方式之修正相同，其資料庫特徵參數修正式分別如式 (6.2.19) ~ (6.2.24) 所示。

最後，考慮 MLBNC 第三層，如第五章之圖 5-17 所示，假設未調適前，MLBNC 第三層資料庫中，各特徵參數向量平均值與共變異數分別為 $\boldsymbol{\mu}_{1,old}^3$ 、 $\boldsymbol{\mu}_{2,old}^3$ 、 $\boldsymbol{\mu}_{3,old}^3$ 以及 $\boldsymbol{\mu}_{4,old}^3$ 、 $\Sigma_{1,old}^3$ 、 $\Sigma_{2,old}^3$ 、 $\Sigma_{3,old}^3$ 以及 $\Sigma_{4,old}^3$ ，其中 $\Sigma_{1,old}^3$ 、 $\Sigma_{2,old}^3$ 、 $\Sigma_{3,old}^3$ 以及 $\Sigma_{4,old}^3$ 分別為 8×8 之陣列，

$$\boldsymbol{\mu}_{1,old}^3 = \left[\bar{X}_{3,old}^S \quad \bar{X}_{10,old}^S \quad \bar{X}_{11,old}^S \quad \bar{X}_{12,old}^S \quad \bar{X}_{14,old}^S \quad \bar{X}_{21,old}^S \quad \bar{X}_{22,old}^S \quad \bar{X}_{23,old}^S \right]^T$$

(6.3.22)

$$\boldsymbol{\mu}_{2,old}^3 = [\bar{X}_{3,old}^N \quad \bar{X}_{10,old}^N \quad \bar{X}_{11,old}^N \quad \bar{X}_{12,old}^N \quad \bar{X}_{14,old}^N \quad \bar{X}_{21,old}^N \quad \bar{X}_{22,old}^N \quad \bar{X}_{23,old}^N]^T$$

(6.3.23)

$$\boldsymbol{\mu}_{3,old}^3 = [\bar{X}_{3,old}^H \quad \bar{X}_{10,old}^H \quad \bar{X}_{11,old}^H \quad \bar{X}_{12,old}^H \quad \bar{X}_{14,old}^H \quad \bar{X}_{21,old}^H \quad \bar{X}_{22,old}^H \quad \bar{X}_{23,old}^H]^T$$

(6.3.24)

$$\boldsymbol{\mu}_{4,old}^3 = [\bar{X}_{3,old}^A \quad \bar{X}_{10,old}^A \quad \bar{X}_{11,old}^A \quad \bar{X}_{12,old}^A \quad \bar{X}_{14,old}^A \quad \bar{X}_{21,old}^A \quad \bar{X}_{22,old}^A \quad \bar{X}_{23,old}^A]^T$$

(6.3.25)

由於 MLBNC 辨識結果與測試語音之情緒不符，因此，必須對 MLBNC 第三層之特徵參數平均值向量與共變異數資料庫作修正，根據 5.3.1 之 MLBNC 推導，第三層所使用之測試特徵參數向量為 $\mathbf{V} = [X_3 \quad X_{10} \quad X_{11} \quad X_{12} \quad X_{14} \quad X_{21} \quad X_{22} \quad X_{23}]^T$ ，若測試語音之情緒為悲傷，而 MLBNC 辨識結果非悲傷時，MLBNC 第三層中，第一群悲傷之資料庫特徵參數平均值向量與共變異數之修正式分別如式 (6.3.26) 與式 (6.3.27) 所示。

$$\boldsymbol{\mu}_{1,new}^3 = \beta_2 \boldsymbol{\mu}_{1,old}^3 + (1 - \beta_2) \mathbf{V} \quad (6.3.26)$$

$$\begin{aligned} \Sigma_{1,new}^3 = & \beta_2 \left[\Sigma_{1,old}^3 + (\boldsymbol{\mu}_{1,new}^3 - \boldsymbol{\mu}_{1,old}^3)(\boldsymbol{\mu}_{1,new}^3 - \boldsymbol{\mu}_{1,old}^3)^T \right] \\ & + (1 - \beta_2) (\boldsymbol{\mu}_{1,new}^3 - \mathbf{V})(\boldsymbol{\mu}_{1,new}^3 - \mathbf{V})^T \end{aligned} \quad (6.3.27)$$

其中， $\boldsymbol{\mu}_{1,old}^3$ 為未調適前，MLBNC 資料庫第三層中，悲傷之特徵參數原始平均值向量， $\boldsymbol{\mu}_{1,new}^3$ 為調適後，MLBNC 資料庫第三層中，悲傷之特徵參數新平均值向量，而 $\Sigma_{1,old}^3$ 為未調適前，MLBNC 資料庫第三層中，悲傷之特徵參數原始共變異數， $\Sigma_{1,new}^3$ 則為調適後，MLBNC 資料庫第三層中，悲傷之特徵參數新共變異數。

若測試語音之情緒為中性，而 MLBNC 辨識結果非中性時，MLBNC 第三層中，第二群中性之資料庫特徵參數平均值向量與共變異數之修正式分別如式(6.3.28)與式(6.3.29)所示。

$$\boldsymbol{\mu}_{2,new}^3 = \beta_2 \boldsymbol{\mu}_{2,old}^3 + (1 - \beta_2) \mathbf{V} \quad (6.3.28)$$

$$\begin{aligned} \Sigma_{2,new}^3 = \beta_2 & \left[\Sigma_{2,old}^3 + (\boldsymbol{\mu}_{2,new}^3 - \boldsymbol{\mu}_{2,old}^3)(\boldsymbol{\mu}_{2,new}^3 - \boldsymbol{\mu}_{2,old}^3)^T \right] \\ & + (1 - \beta_2)(\boldsymbol{\mu}_{2,new}^3 - \mathbf{V})(\boldsymbol{\mu}_{2,new}^3 - \mathbf{V})^T \end{aligned} \quad (6.3.29)$$

其中， $\boldsymbol{\mu}_{2,old}^3$ 為未調適前，MLBNC 資料庫第三層中，中性之特徵參數原始平均值向量， $\boldsymbol{\mu}_{2,new}^3$ 為調適後，MLBNC 資料庫第三層中，中性之特徵參數新平均值向量，而 $\Sigma_{2,old}^3$ 為未調適前，MLBNC 資料庫第三層中，中性之特徵參數原始共變異數， $\Sigma_{2,new}^3$ 則為調適後，MLBNC 資料庫第三層中，中性之特徵參數新共變異數。

若測試語音之情緒為快樂，而 MLBNC 辨識結果非快樂時，MLBNC 第三層中，第三群快樂之資料庫特徵參數平均值向量與共變異數之修正式分別如式(6.3.30)與式(6.3.31)所示。

$$\boldsymbol{\mu}_{3,new}^3 = \beta_2 \boldsymbol{\mu}_{3,old}^3 + (1 - \beta_2) \mathbf{V} \quad (6.3.30)$$

$$\begin{aligned} \Sigma_{3,new}^3 = \beta_2 & \left[\Sigma_{3,old}^3 + (\boldsymbol{\mu}_{3,new}^3 - \boldsymbol{\mu}_{3,old}^3)(\boldsymbol{\mu}_{3,new}^3 - \boldsymbol{\mu}_{3,old}^3)^T \right] \\ & + (1 - \beta_2)(\boldsymbol{\mu}_{3,new}^3 - \mathbf{V})(\boldsymbol{\mu}_{3,new}^3 - \mathbf{V})^T \end{aligned} \quad (6.3.31)$$

其中， $\boldsymbol{\mu}_{3,old}^3$ 為未調適前，MLBNC 資料庫第三層中，快樂之特徵參數原始平均值向量， $\boldsymbol{\mu}_{3,new}^3$ 為調適後，MLBNC 資料庫第三層中，快樂之特徵參數新平均值向量，而 $\Sigma_{3,old}^3$ 為未調適前，MLBNC 資料庫第

三層中，快樂之特徵參數原始共變異數， $\Sigma_{3,new}^3$ 則為調適後，MLBNC 資料庫第三層中，快樂之特徵參數新共變異數。

若測試語音之情緒為生氣，而 MLBNC 辨識結果非快樂時，MLBNC 第三層中，第四群生氣之資料庫特徵參數平均值向量與共變異數之修正式分別如式(6.3.32)與式(6.3.33)所示。

$$\boldsymbol{\mu}_{4,new}^3 = \beta_2 \boldsymbol{\mu}_{4,old}^3 + (1 - \beta_2) \mathbf{V} \quad (6.3.32)$$

$$\begin{aligned} \Sigma_{4,new}^3 = \beta_2 & \left[\Sigma_{4,old}^3 + (\boldsymbol{\mu}_{4,new}^3 - \boldsymbol{\mu}_{4,old}^3)(\boldsymbol{\mu}_{4,new}^3 - \boldsymbol{\mu}_{4,old}^3)^T \right] \\ & + (1 - \beta_2)(\boldsymbol{\mu}_{4,new}^3 - \mathbf{V})(\boldsymbol{\mu}_{4,new}^3 - \mathbf{V})^T \end{aligned} \quad (6.3.33)$$

其中， $\boldsymbol{\mu}_{4,old}^3$ 為未調適前，MLBNC 資料庫第三層中，生氣之特徵參數原始平均值向量， $\boldsymbol{\mu}_{4,new}^3$ 為調適後，MLBNC 資料庫第三層中，生氣之特徵參數新平均值向量，而 $\Sigma_{4,old}^3$ 為未調適前，MLBNC 資料庫第三層中，生氣之特徵參數原始共變異數， $\Sigma_{4,new}^3$ 則為調適後，MLBNC 資料庫第三層中，生氣之特徵參數新共變異數。

適應性 MLBNC 資料庫修正式(6.3.1) ~ (6.3.33)中，有三個適應性參數，分別為 α_2 、 β_2 及 γ_2 ，此三個適應性參數分別對應 MLBNC 資料庫之各分群中含有之情緒類別數目，當此分群中含有兩個情緒類別數目時，適應性參數使用 α_2 ，當此分群中含有一個情緒類別數目時，適應性參數使用 β_2 ，而當此分群中含有三個情緒類別數目時，適應性參數則使用 γ_2 ，與適應性 MLBN 之實驗相同，在適應性 MLBNC 之後續實驗中，我們將固定適應性參數 $\alpha_2 = 0.9$ 、 $\beta_2 = 0.8$ 以及

$\gamma_2 = 0.95$ ，以利適應性 MLBNC 之實驗進行與分析。

經由適應性 MLBNC 修正後，MLBNC 資料庫中各情緒之特徵參數平均值向量與共變異數將會依照使用者之情緒分佈狀態作調整，以下將就適應性 MLBNC 分類器進行同語系與不同語系之適應性實驗，並就實驗結果分別加以說明。

6.3.2 適應性 MLBNC 實驗結果與分析

適應性 MLBNC 與適應性 MLBN 之調適方式相似，主要是依據測試者個人的情緒分佈狀態去做調適修正，因此，在本節適應性 MLBNC 實驗中，與適應性 MLBN 實驗相同，將採用 MLBNC Outside Test 方式進行辨識，首先，先進行同語系適應性 MLBNC 實驗，以德國語料庫當作適應性 MLBNC 實驗語料，首先，將德國語料庫 P1 語料當作測試語料，並以 P2 ~ P10 語料當作 MLBNC 訓練語料，因此悲傷有 7 筆、中性有 11 筆、快樂有 7 筆、生氣有 14 筆，共 39 筆測試語料，而訓練語料悲傷有 55 筆、中性有 68 筆、快樂有 64 筆、生氣有 113 筆，接著，從 39 筆測試語料中隨機選取一筆資料並以 MLBNC 分類器進行辨識，若辨識結果正確，則從剩餘測試資料中隨機選取另一筆資料進行測試，若辨識結果錯誤，則根據前面所提出之適應性 MLBNC 資料庫修正方式，對 MLBNC 資料庫進行修正，修正完後，再從剩餘測試資料中隨機選取另一筆資料，並以修正過後之新

MLBNC 資料庫進行測試，直到 39 筆測試語料測試調適完畢，最後，將測試完畢並修正過後之新 MLBNC 分類器對 39 筆測試資料重新進行測試。

接著，與 P1 語料之調適與測試流程相同，對 P2 ~ P10 做 MLBNC Outside Test，並進行 MLBNC 資料庫調適與測試，表 6-7 為 P1 ~ P10 多層共變異數貝氏網路調適前後之整體鑑別矩陣(P1 ~ P10 調適前後各情緒之辨識正確結果詳見附錄 C)，從整體辨識結果來看，MLBNC 分類器在調適過後，悲傷之辨識正確率從 82% 大幅提升到 100%，中性則維持一樣的辨識正確率，快樂之辨識效果在調適過後則是大幅提升，辨識正確率從 62% 提升到 80%，至於生氣，辨識正確率則是從 85.8% 提升至 96.9%，四種情緒類別之辨識正確率皆有所提升，而整體之辨識正確率則是從調適前 83.5% 提升到調適後 94.7%，因此，經由適應性 MLBNC 調適過程，可以有效提升辨識效果。

接下來，考慮適應性 MLBNC 應用於不同語系時，其調適效果對辨識結果之影響，同樣的，以德國語料庫所有資料當作訓練資料，並以工業技術研究院所錄製之語料當作測試語料，測試適應性 MLBNC 在不同語系間之效果。在調適與測試的過程中，與 KNN 及適應性 MLBN 採用相同的調適與測試語料，以 P1 為例，各情緒各有 15 筆，共 60 筆語料為調適語料，其餘則為測試語料，首先，從 60 筆調適語

料中隨機選取一筆語料並以 MLBNC 分類器進行辨識，若辨識結果正確，則從剩餘調適語料中隨機選取另一筆調適語料進行測試，若辨識結果錯誤，則根據前面所提出之 MLBNC 資料庫修正方式，對 MLBNC 資料庫進行修正，當修正完後，再從剩餘調適語料中隨機選取另一筆調適語料，並以修正過後之新 MLBNC 資料庫進行測試，直到 60 筆調適語料測試調適完畢，最後，將 P1 剩餘之四種情緒各 15 筆，共 60 筆測試語料，分別對未調適前及調適完後之 MLBNC 分類器做測試。

接著，與工業技術研究院語料 P1 之調適與測試流程相同，對工業技術研究院語料 P2 ~ P20 做不同語系之 MLBNC Outside Test，並進行 MLBNC 資料庫調適與測試，表 6-8 為 MLBNC 分類器於工業技術研究院語料 P1 ~ P20 調適前後整體統計之鑑別矩陣(P1 ~ P20 調適前後各情緒之辨識正確結果詳見附錄 C)，從鑑別矩陣來看，在調適前，悲傷與中性會被視為同一群組，悲傷之情緒絕大部分會被 MLBNC 分類器辨識為中性，少數會被正確辨識為悲傷，中性情緒則是被辨識正確為中性居多，而快樂與生氣則會被歸類為另一個群組，快樂有 43.0% 會被辨識為快樂，49.0% 會被辨識為生氣，而生氣情緒則是分別有 17.0% 及 15.7% 被辨識為中性及快樂，63.3% 會被辨識正確為生氣，調適前之整體辨識率為 52.58%，適應性 MLBNC 在調適

過後，悲傷之辨識率從 17.0% 大幅提升到 88.7%，中性則與原先的辨識正確率下降 4%，辨識正確率為 83.0%，快樂之辨識效果在調適過後，也是有所提升，辨識率從 43.0% 提升到 64.7%，至於生氣，辨識效果則快樂相近，辨識正確率約提升 25.0%，辨識率從 63.3% 提升至 88.3%，整體之辨識正確率則是從調適前 52.58% 提升到 81.17%，因此，當 MLBNC 訓練語系與使用者使用之語系不同時，經由本論文所提出之適應性 MLBNC 調適過程，將可以有效提升辨識效果。

最後，與適應性 MLBN 相同，增加適應性 MLBNC 調適語料數量，用以確定調適過程增加時，是否對適應性 MLBNC 之分類效果有所助益，首先，先以工研院 P1 所有語料當作調適語料，因此悲傷有 30 筆、中性有 30 筆、快樂有 30 筆、生氣有 30 筆，共 120 筆調適語料，接著，與前面之 MLBNC 測試調適過程相同，從 120 筆調適語料中隨機選取一筆資料並以 MLBNC 分類器進行辨識，若辨識結果正確，則再從剩餘測試資料中隨機選取另一筆資料進行測試，若辨識結果錯誤，則根據 6.3.1 所提出之 MLBNC 資料庫修正方法，對 MLBNC 資料庫進行修正，修正完後，再從剩餘測試資料中隨機選取另一筆資料，並以修正過後之新 MLBNC 資料庫進行測試調適，直到 120 筆測試語料測試調適完畢，最後，將調適完畢並修正過後之新 MLBNC 分類器對 P1 所有資料進行測試。

接著，分別對 P2 至 P20 個別語料進行 MLBNC 調適與測試，表 6-9 為調適語料增加後，MLBNC 分類器於工業技術研究院語料 P1 ~ P20 調適前後整體統計之鑑別矩陣(P1 ~ P20 調適前後各情緒之辨識正確結果詳見附錄 C)，從鑑別矩陣來看，在調適前，其辨識結果與前面取一半語料調適另一半語料測試之結果相近，悲傷與中性會被視為同一群組，悲傷之情緒絕大部分會被 MLBNC 分類器辨識為中性，少數會被正確辨識為悲傷，中性情緒則是被辨識正確為中性居多，快樂有 39.7% 會被辨識為快樂，52.5% 會被辨識為生氣，而生氣情緒則是分別有 15.5% 及 18.3% 被辨識為中性及快樂，62.5% 會被辨識正確為生氣，而在調適過後，悲傷之辨識率從 18.0% 大幅提升到 93.3%，中性之辨識正確率則是些微上升，從 88.2% 上升到 93.2%，而快樂與生氣之辨識效果在調適過後，辨識正確率皆是大幅提升，辨識率分別從 39.7% 提升到 80.3% 以及 62.5% 提升到 93.0%，其中，快樂情緒在增加調適語料後，仍有一小部分測試語料被辨識為生氣，而就整體辨識效果來看，整體之辨識正確率則是從調適前 52.08% 大幅提升至 89.96%，與表 6-8 比較可以發現，在調適語料增加後，悲傷、中性以及生氣情緒之整體辨識正確率皆提升約 5% 左右，而快樂情緒則提升了約 16%，顯示增加調適次數將有助於提升 MLBNC 語音情緒之分類效果。

表 6-10 為各分類於不同測試條件下之整體辨識結果整理，從辨識結果可以發現，不論是 Inside Test 或 Outside Test，SVM 均有最佳的辨識效果，而當測試語系與訓練所使用之語系不同時，各分類器之辨識效果均不佳，除了 MLBNC 之辨識正確率超過五成外，其餘分類器之辨識正確率皆低於 50%。由蒙地卡羅模擬法對各分類器之辨識效果所進行之測試結果，SVM 以及 MLBNC 均提供了良好的辨識效果，而在適應性的實驗上，由於 SVM 本身必須先由訓練資料庫來找出最佳的分割模型，因此，SVM 在調適上較為複雜且耗時，在實際應用上，不容易實現，從 KNN、MLBN 以及 MLBNC 之適應型測試結果來看，在一半語料調適與一半語料測試的相同測試條件下，適應性 MLBN 及適應性 MLBNC 經調適過後，辨識正確率皆優於 KNN，當調適次數增加時，適應性 MLBN 及適應性 MLBNC 之辨識正確率皆有所提升，顯示增加調適次數有助於提升適應性 MLBN 及 MLBNC 語音情緒之分類效果，而在調適流程上，調適型 KNN 需增加資料庫之新資料，而 MLBN 之修正過程僅需修正資料庫中特徵參數平均值與標準差，MLBNC 之修正過程也僅需修正資料庫中特徵參數平均值向量與共變異數，因此，透過適應性 MLBN 與適應性 MLBNC 之分類與調適方式，除了可以減少資料庫之資料量外，更可以減少資料庫在硬體上之需求，對於未來語音情緒辨識在實際應用及硬體需求上將

有相當大的助益。

6.4 小結

首先，本章先使用適應性 KNN 針對不同語系的使用者做調適測試，實驗中以德國語料庫所有語料當作訓練資料，並以工研院語料中其個人語料一半語料當作調適語料，另一半語料當作測試語料，在調適的過程中，當辨識結果正確時，KNN 資料庫不做任何更改變動，而當辨識結果錯誤時，則是將辨識錯誤之測試資料，於標示類別後加入訓練資料中，藉以擴充訓練資料以符合測試者實際的特徵狀態。從 KNN 調適前後的辨識結果可以發現，在調適前，除了悲傷有 16% 被正確辨識為悲傷，快樂有 22.3% 被正確辨識為快樂外，絕大部分情緒皆被辨識為中性，調適前之整體辨識正確率為 34.17%，而調適過後，原先辨識結果集中在某一情緒的狀況有所改善，調適後主要分為兩個族群，悲傷與中性為一族群，快樂與生氣則為另一個族群，悲傷、快樂及生氣之辨識正確率分別由 16.0%、22.3%、1.0% 提升至 73.7%、75.0% 以及 71.0%，悲傷有 26.3% 被辨識為中性，快樂有 15.7% 被辨識為生氣，生氣則有 14.3% 被辨識為快樂，而中性之辨識正確率則是下降，辨識正確率由 97.3% 下降至 75.0%，其中，有一部份原先辨識正確之中性情緒，在新 KNN 資料庫下，被辨識為悲傷，而從鑑別矩陣來看，各情緒之辨識正確率在調適後趨於一致，顯示新資料庫對於各

情緒之分類效果較趨穩定，調適後之整體辨識正確率為 73.7%，調適前後，整體辨識率從 34.17% 提升到 73.67%，因此，透過調適過程增加訓練資料數量以符合實際使用者的狀況，可以有效提升 KNN 分類器之整體辨識效果。

接著，6.2 節則是介紹適應性 MLBN 的調適流程與資料庫修正方法，並進行適應性 MLBN 實驗與分析，在 MLBN 調適過程中，當 MLBN 分類器辨識出來之情緒與測試之語音情緒相同時，MLBN 資料庫不做任何更改變動，而當辨識結果與調適測試之語音情緒不同時，則是按照 6.2.1 所提出修正方法，將 MLBN 上各層各群之平均值與標準差做修正，從 6.2.2 適應性 MLBN 實驗結果來看，當訓練資料與使用者為相同語系時，MLBN 分類器在調適過後，悲傷之辨識正確率從 84% 提升到 99%，中性則維持極佳的辨識效果，從 97% 提升到 99%，快樂之辨識效果則是大幅提升，辨識正確率從 54% 提升到 83%，至於生氣，辨識正確率則是從 75.6% 提升至 90.6%，整體之辨識正確率則是從調適前 77.6% 提升到 92.3%。

當訓練資料與使用者為不同語系時，首先，與 KNN 調適與測試之語料相同，以德國語料庫所有語料當作訓練資料，並以工研院語料中其個人語料一半語料當作調適語料，另一半語料當作測試語料，從實驗結果，適應性 MLBN 在調適過後，悲傷之辨識正確率從 20.3%

大幅提升到 86.0%，中性之辨識正確率則是下降，從 95.0% 下降到 85.0%，而快樂與生氣之辨識效果在調適過後，均大幅提升，辨識率分別從 31.3% 提升到 83.3% 以及 4.7% 提升到 75.3%，整體之辨識正確率則是從調適前 37.8% 大幅提升至 82.4%。

接著，增加適應性 MLBN 調適語料數量，以德國語料庫所有語料當作訓練資料，並以工研院語料中其個人語料所有語料當作調適語料，並以所有語料在 MLBN 調適後進行測試，從實驗結果，適應性 MLBN 在調適過後，悲傷之辨識率從 20.7% 大幅提升到 91.0%，中性之辨識正確率則是些微下降，從 96.0% 下降到 91.8%，而快樂與生氣之辨識效果在調適過後，則是大幅提升，辨識率分別從 35.0% 提升到 88.7% 以及 5.7% 提升到 84.2%，整體之辨識正確率則是從調適前 39.33% 大幅提升至 88.92%，與工研院語料中其個人語料一半語料當作調適語料，另一半語料當作測試語料之 MLBN 實驗結果比較，在調適語料增加後，各情緒之整體辨識正確率皆提升約 6~8%，顯示增加調適次數將有助於提升 MLBN 語音情緒之分類效果。從 6.2.2 適應性 MLBN 結果得知，經由適應性 MLBN 調適過程，不論是同語系或不同語系，適應性 MLBN 均可以有效提升辨識效果。

最後，6.3 節則是介紹適應性 MLBNC 的調適流程與資料庫修正方法，並進行適應性 MLBNC 實驗與分析，在 MLBNC 調適過程中，

當 MLBNC 分類器辨識出來之情緒與測試之語音情緒相同時，MLBNC 資料庫不做任何更改變動，而當辨識結果與調適測試之語音情緒不同時，則是按照 6.3.1 所提出之 MLBNC 資料庫修正方法，將 MLBNC 上各層各群之特徵參數平均值向量與共變異數做修正，從 6.3.2 適應性 MLBNC 實驗結果來看，當訓練資料與使用者為相同語系時，MLBNC 分類器在調適過後，悲傷之辨識正確率從 82% 大幅提升到 100%，中性則維持一樣的辨識正確率，快樂之辨識效果在調適過後則是大幅提升，辨識正確率從 62% 提升到 80%，至於生氣，辨識正確率則是從 85.8% 提升至 96.9%，四種情緒類別之辨識正確率皆有所提升，而整體之辨識正確率則是從調適前 83.5% 提升到調適後 94.7%。

當訓練資料與使用者為不同語系時，與 KNN 及 MLBN 調適與測試之語料相同，以德國語料庫所有語料當作訓練資料，並以工研院語料中其個人語料一半語料當作調適語料，另一半語料當作測試語料，從實驗結果，適應性 MLBNC 在調適過後，悲傷之辨識率從 17.0% 大幅提升到 88.67%，中性則與原先的辨識正確率下降 4%，辨識正確率為 83.0%，快樂之辨識效果在調適過後，也是有所提升，辨識率從 43.0% 提升到 64.7%，至於生氣，辨識效果則快樂相近，辨識正確率約提升 25.0%，辨識率從 63.3% 提升至 88.3%，整體之辨識正確率則

是從調適前 52.58% 提升到 81.17%。

接著，與 MLBN 相同，增加適應性 MLBNC 調適語料數量，以德國語料庫所有語料當作訓練資料，並以工研院語料中其個人語料所有語料當作調適語料，並以所有語料在 MLBNC 調適後進行測試，從實驗結果，在 MLBNC 調適過後，悲傷之辨識率從 18.0% 大幅提升到 93.3%，中性之辨識正確率則是些微上升，從 88.2% 上升到 93.2%，而快樂與生氣之辨識效果在調適過後，辨識正確率皆是大幅提升，辨識率分別從 39.7% 提升到 80.3% 以及 62.5% 提升到 93.0%，其中，快樂情緒在增加調適語料後，仍有一小部分測試語料被辨識為生氣，而就整體辨識效果來看，整體之辨識正確率則是從調適前 52.08% 大幅提升至 89.96%，與工研院語料中其個人語料一半語料當作調適語料，另一半語料當作測試語料之 MLBNC 實驗結果比較，在調適語料增加後，悲傷、中性以及生氣情緒之整體辨識正確率皆提升約 5% 左右，而快樂情緒則提升了約 16%，顯示增加調適次數將有助於提升 MLBNC 語音情緒之分類效果。從 6.3.2 適應性 MLBNC 結果得知，經由適應性 MLBNC 調適過程，不論是同語系或不同語系，適應性 MLBNC 均可以有效提升辨識效果。

最後，從適應性 KNN、MLBN 以及 MLBNC 之比較來看，在整體辨識正確率上，適應性 MLBN 以及適應性 MLBNC 經調適過後，

辨識正確率皆優於 KNN，而在調適流程上，調適型 KNN 增加了資料庫之新資料，而 MLBN 之修正過程僅需修正資料庫中特徵參數平均值與標準差，MLBNC 之修正過程也僅需修正資料庫中特徵參數平均值向量與共變異數，因此，透過適應性 MLBN 與適應性 MLBNC 之分類與調適方式，除了可以減少資料庫之資料量外，更可以減少資料庫在硬體上之需求，對於未來語音情緒辨識在實際應用及硬體需求上將有相當大的助益。

最後，從各分類於不同測試條件下之整體辨識結果可以發現，不論是 Inside Test 或 Outside Test，SVM 均有最佳的辨識效果，而當測試語系與訓練所使用之語系不同時，各分類器之辨識效果均不佳，由蒙地卡羅模擬對各分類器之辨識效果所進行之測試結果，SVM 以及 MLBNC 則提供了良好的辨識效果，而在適應性的實驗上，從 KNN、MLBN 以及 MLBNC 之適應型測試結果來看，在一半語料調適與一半語料測試的相同測試條件下，適應性 MLBN 及適應性 MLBNC 經調適過後，辨識正確率皆優於 KNN，當調適次數增加時，適應性 MLBN 及適應性 MLBNC 之辨識正確率皆有所提升，顯示增加調適次數有助於提升適應性 MLBN 及 MLBNC 語音情緒之分類效果，而在調適流程上，調適型 KNN 需增加資料庫之新資料，而 MLBN 之修正過程僅需修正資料庫中特徵參數平均值與標準差，MLBNC 之修正

過程也僅需修正資料庫中特徵參數平均值向量與共變異數，因此，透過適應性 MLBN 與適應性 MLBNC 之分類與調適方式，除了可以減少資料庫之資料量外，更可以減少資料庫在硬體上之需求，對於未來語音情緒辨識在實際應用及硬體需求上將有相當大的助益。



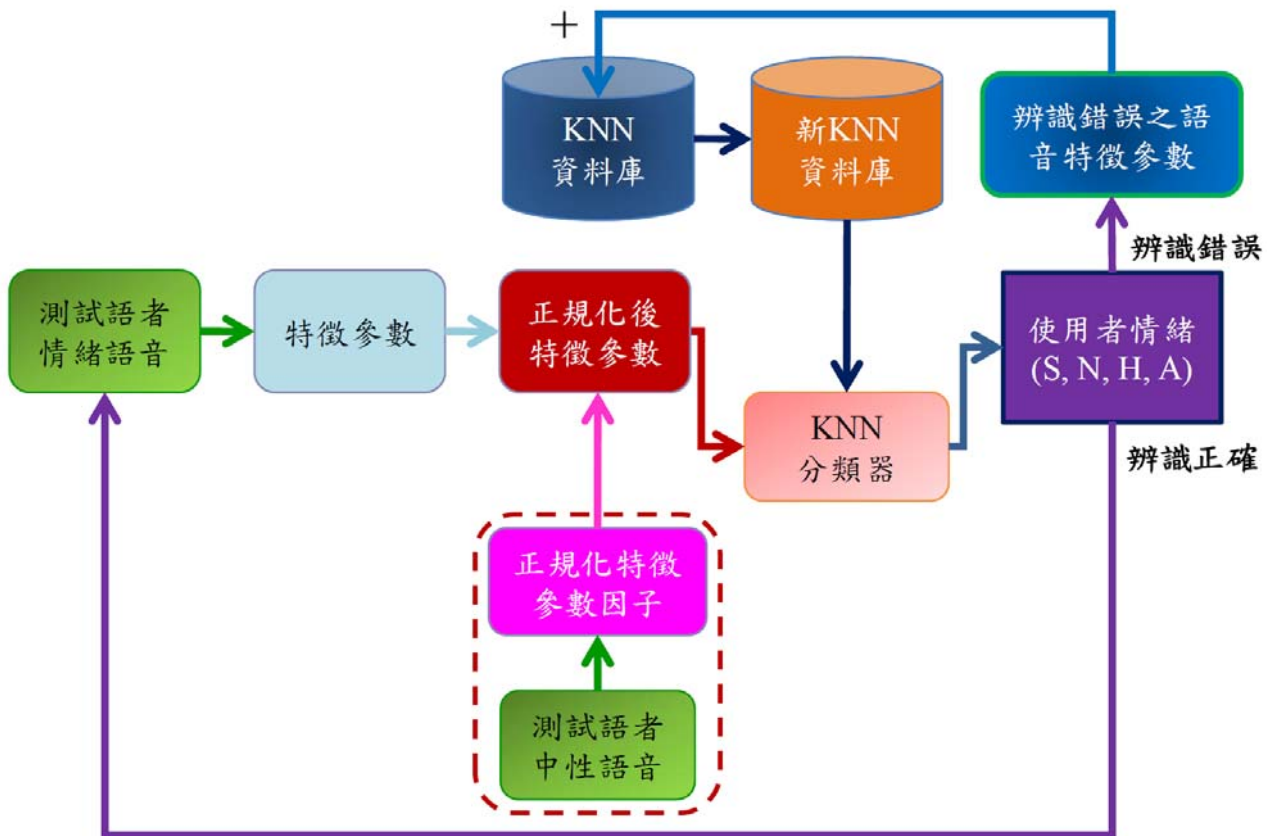


圖 6-1 適應性 KNN 調適流程

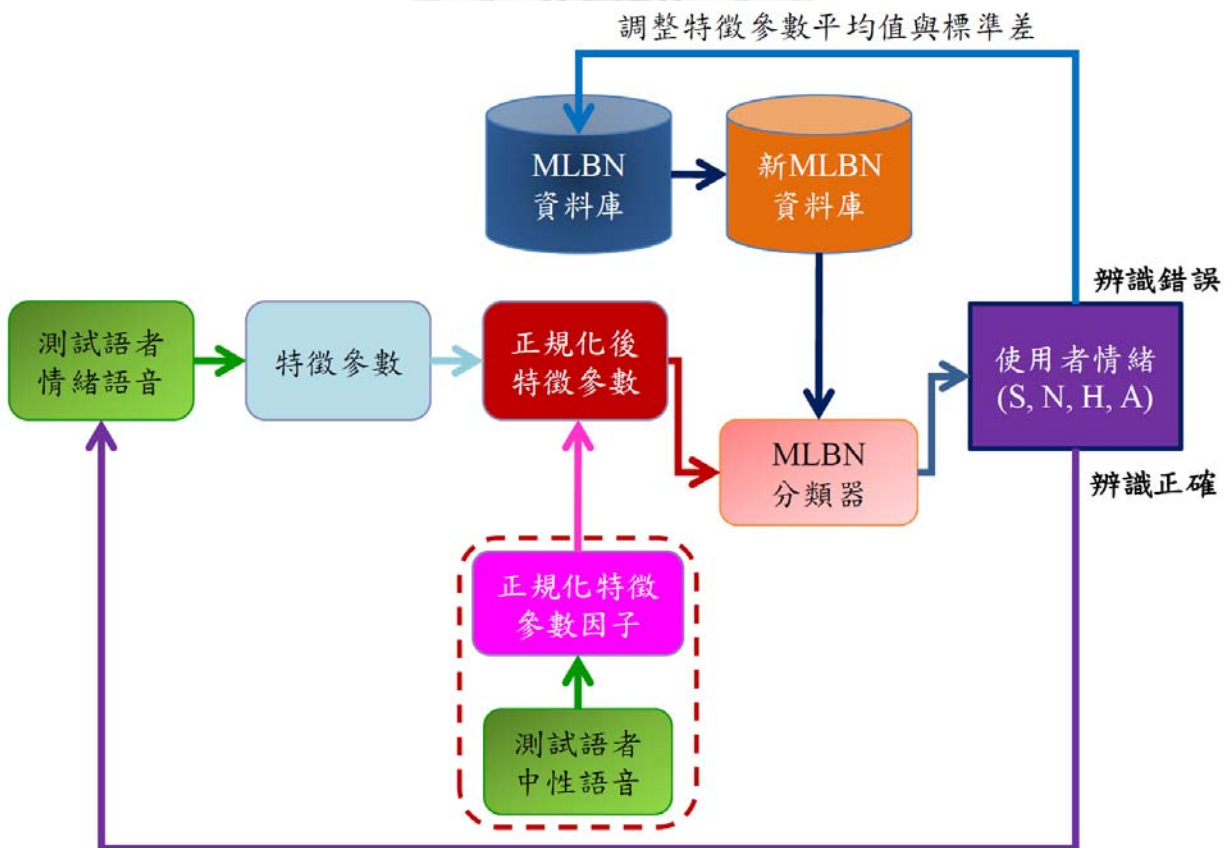


圖 6-2 適應性 MLBN 調適流程

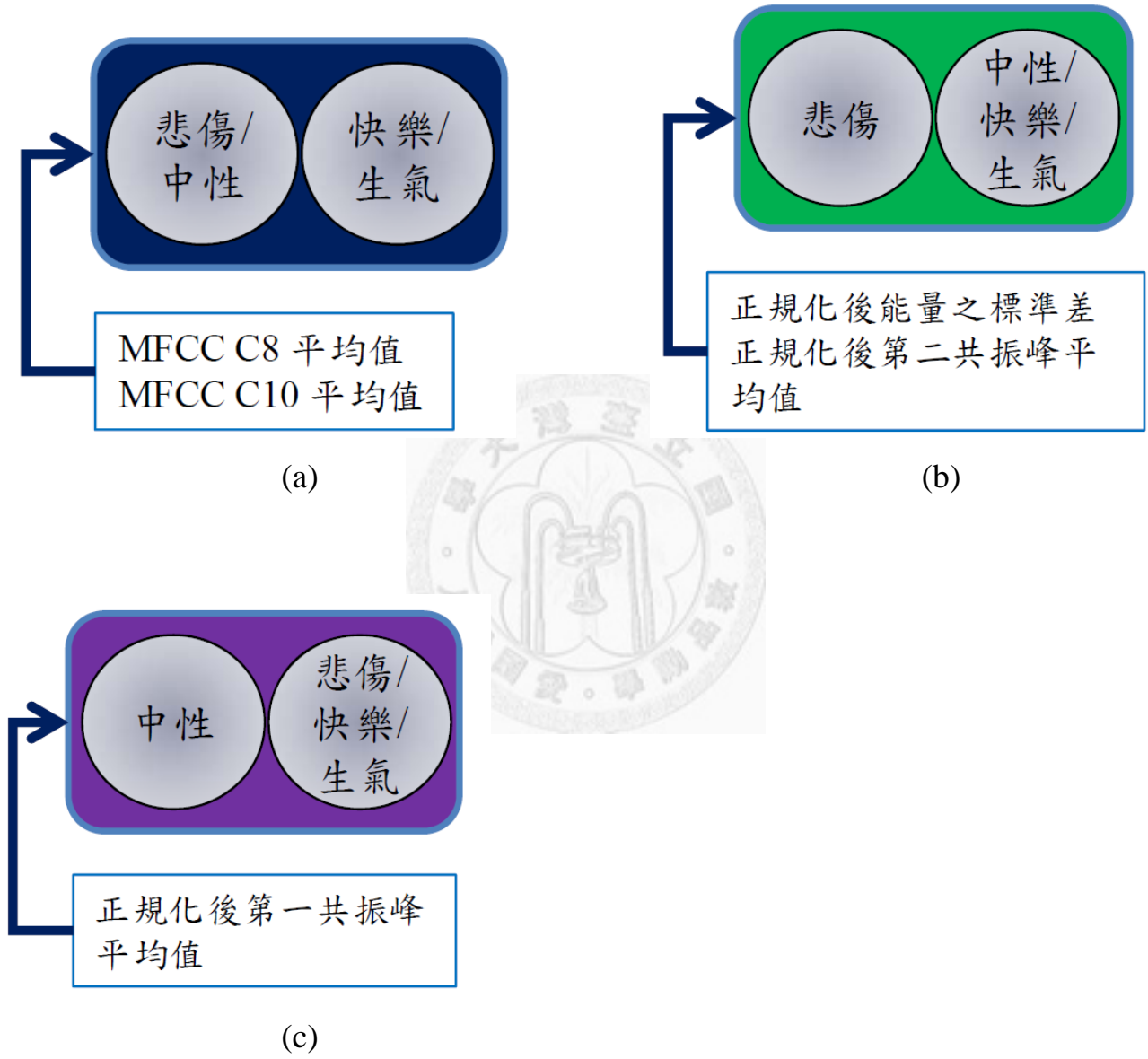


圖 6-3 MLBN 第一層三種分群方式

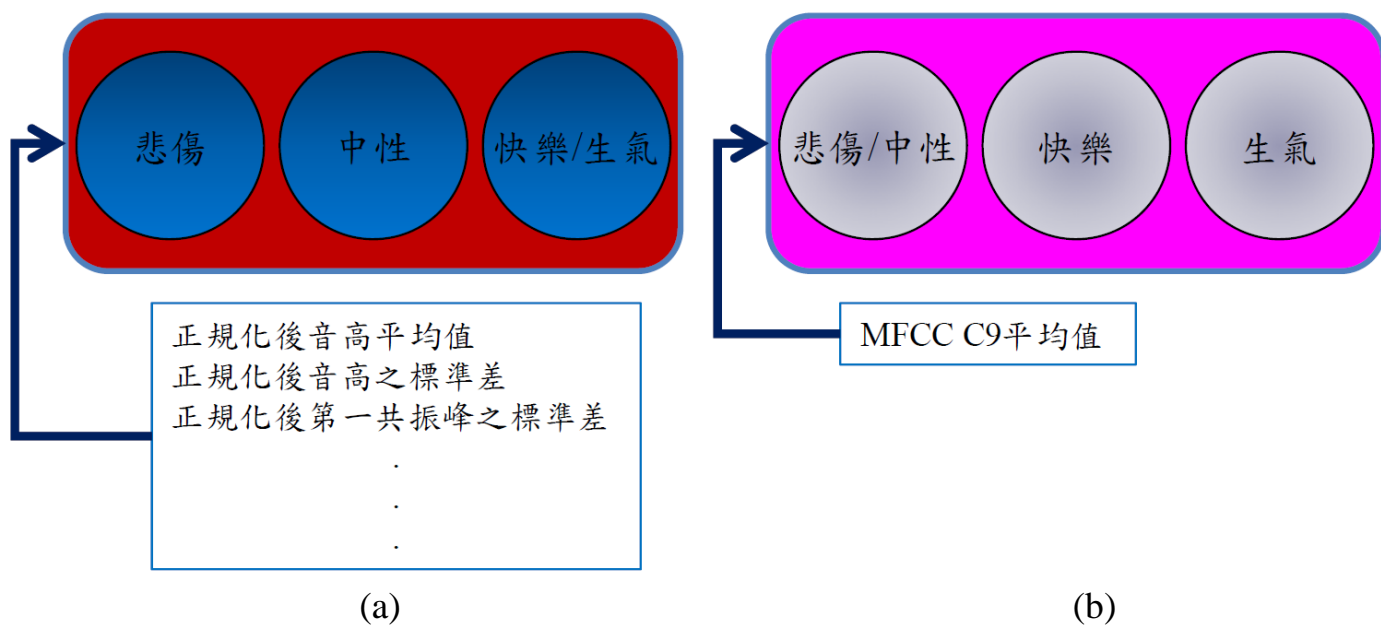


圖 6-4 MLBN 第二層兩種種分群方式

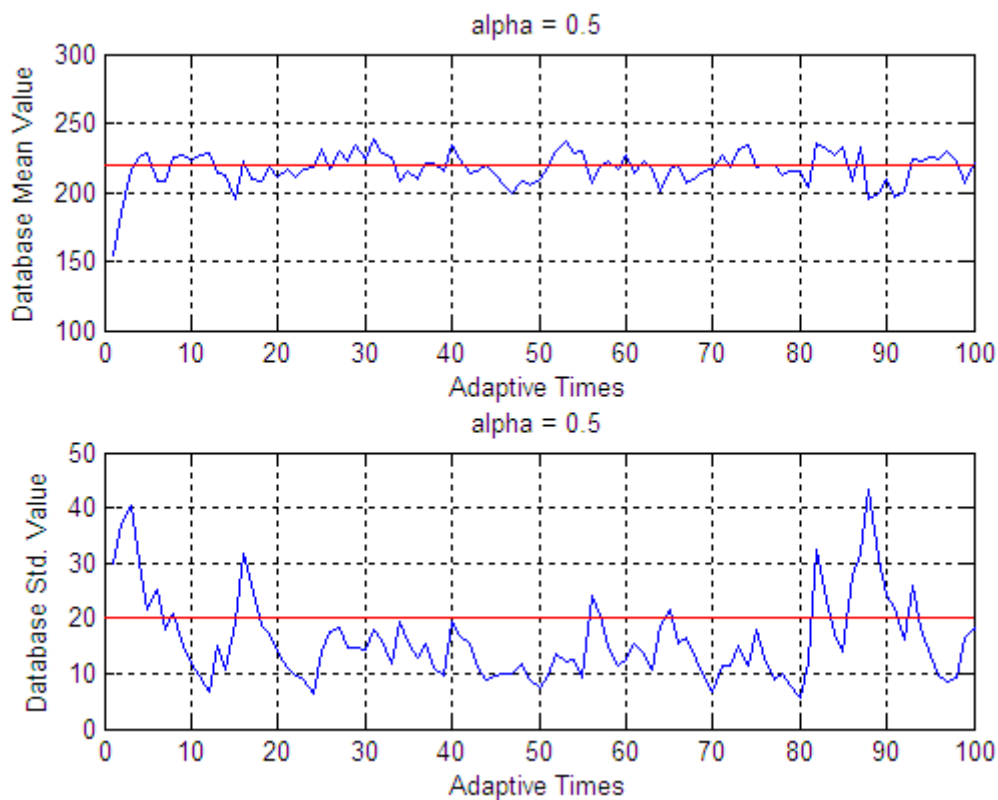


圖 6-5 $\alpha = 0.5$ 時，資料庫調適後變化圖

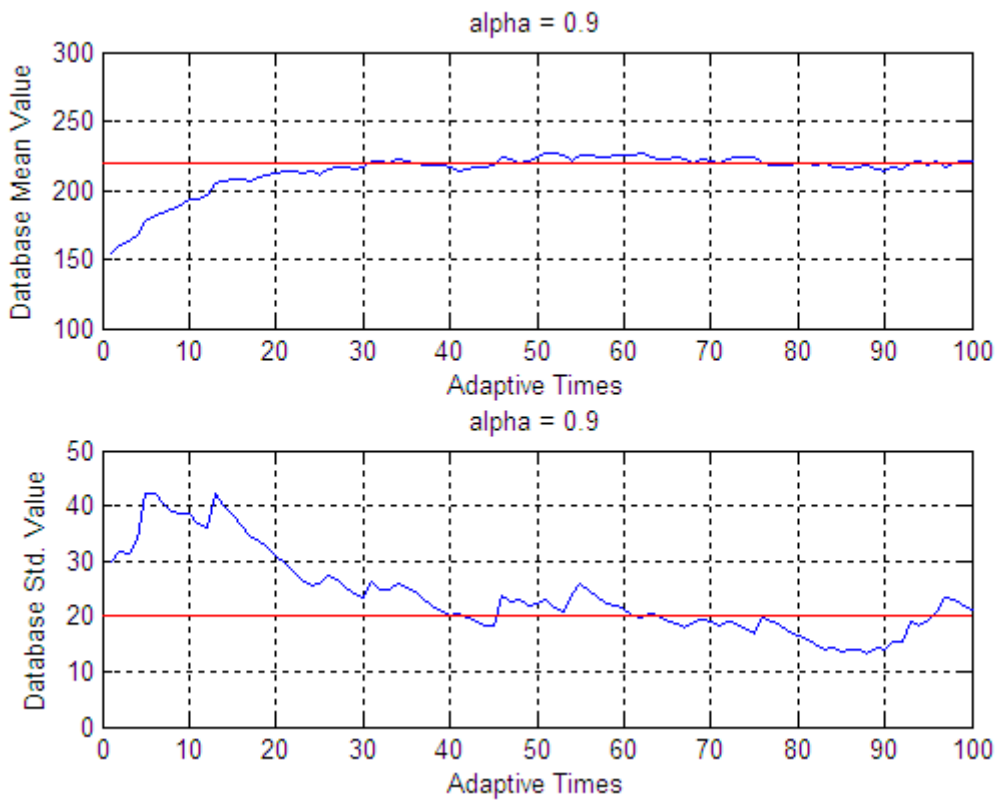


圖 6-6 $\alpha = 0.9$ 時，資料庫調適後變化圖



圖 6-7 適應性 MLBNC 調適流程

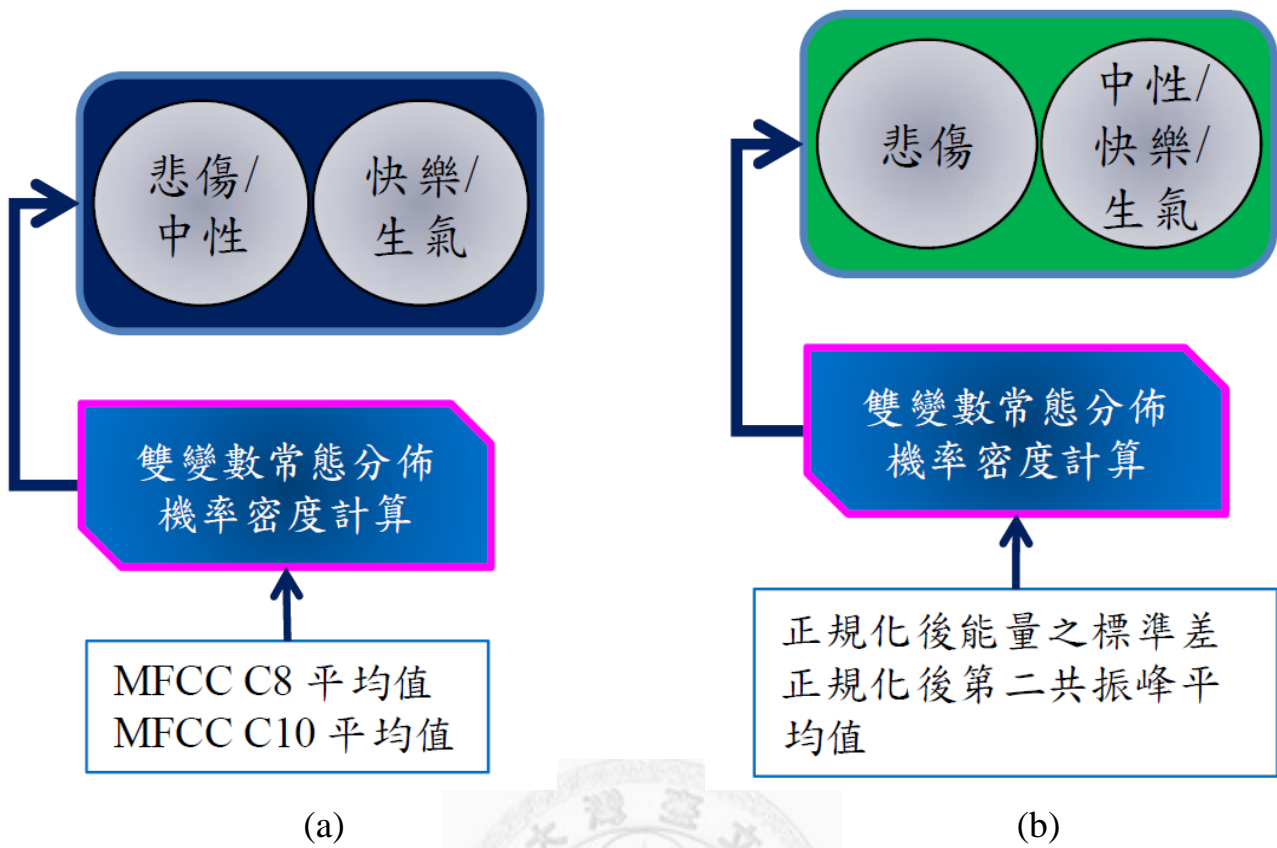


圖 6-8 MLBNC 第一層三種分群方式

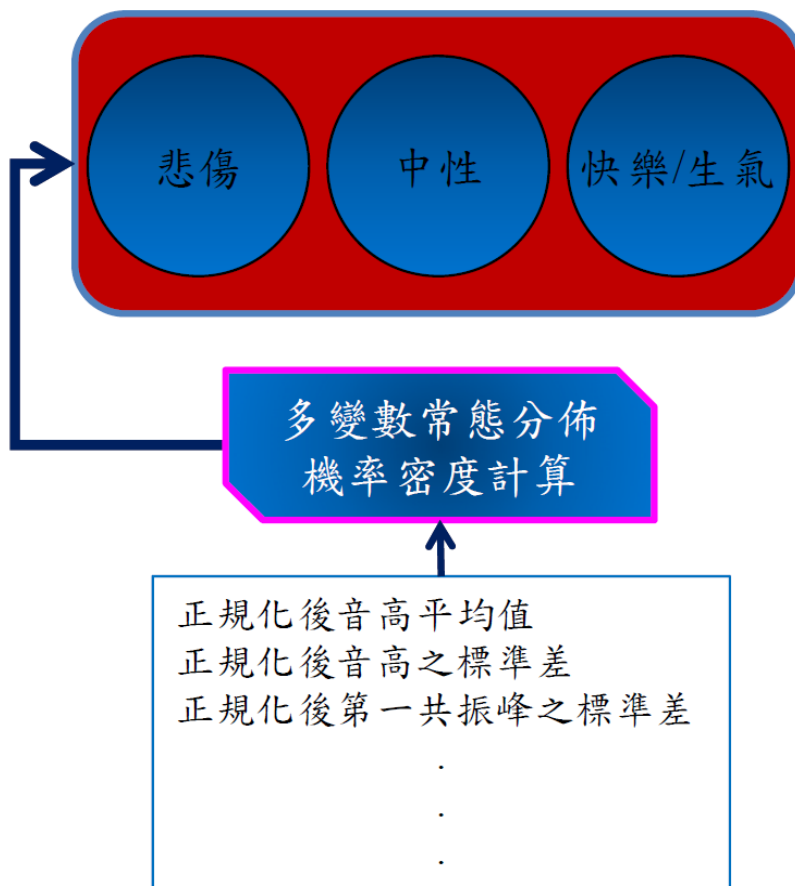


圖 6-9 MLBNC 第二層第一種分群方式

表 6-1 KNN 調適前後之鑑別矩陣(工研院 P1)

P1 調適前	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(15)	0	100% (15/15)	0	0	0 (0/15)
中性(15)	0	100% (15/15)	0	0	100% (15/15)
快樂(15)	7% (1/15)	93% (14/15)	0	0	0 (0/15)
生氣(15)	0	100% (15/15)	0	0	0 (0/15)
調適前 平均辨識率					25% (15/60)
調適次數	10	2	6	6	
P1 調適後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(15)	33% (5/15)	67% (10/15)	0	0	33% (5/15)
中性(15)	33% (5/15)	66% (10/15)	0	0	66% (10/15)
快樂(15)	7% (1/15)	13% (2/15)	47% (7/15)	33% (5/15)	47% (7/15)
生氣(15)	33% (5/15)	7% (1/15)	7% (1/15)	53% (8/15)	53% (8/15)
調適後 平均辨識率					50% (30/60)

表 6-2 KNN 不同語系調適前後之鑑別矩陣(一半調適，一半測試)

調適前	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(300)	16.0% (48/300)	83.3% (250/300)	0.7% (2/300)	0	16.0% (48/300)
中性(300)	2.3% (7/300)	97.3% (292/300)	0.3% (1/300)	0	97.3% (292/300)
快樂(300)	4.7% (14/300)	68.0% (204/300)	22.3% (67/300)	5.0% (15/300)	22.3% (67/300)
生氣(300)	5.3% (16/300)	92.7% (278/300)	1.0% (3/300)	1.0% (3/300)	1.0% (3/300)
調適前 平均辨識率					34.17% (410/1200)
調適後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(300)	73.7% (221/300)	26.3% (79/300)	0	0	73.7% (221/300)
中性(300)	25.0% (75/300)	75.0% (225/300)	0	0	75.0% (225/300)
快樂(300)	3.3% (10/300)	6.0% (18/300)	75.0% (225/300)	15.7% (47/300)	75.0% (225/300)
生氣(300)	5.7% (17/300)	9.0% (27/300)	14.3% (43/300)	71.0% (213/300)	71.00% (213/300)
調適後 平均辨識率					73.67% (884/1200)

表 6-3 MLBN 德國語料庫同語系調適前後之鑑別矩陣

調適前	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	84% (52/62)	16% (10/62)	0	0	84% (52/62)
中性(79)	3% (2/79)	97% (77/79)	0	0	97% (77/79)
快樂(71)	1% (1/71)	14% (10/71)	54% (38/71)	31% (22/71)	54% (38/71)
生氣(127)	0	0.8% (1/127)	23.6% (30/127)	75.6% (96/127)	75.6% (96/127)
調適前 平均辨識率					77.6% (263/339)
調適後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	98% (61/62)	2% (1/62)	0	0	98% (61/62)
中性(79)	1% (1/79)	99% (78/79)	0	0	99% (78/79)
快樂(71)	0	4% (3/71)	83% (59/71)	13% (9/71)	83% (59/71)
生氣(127)	0	0.8% (1/127)	8.7% (11/127)	90.6% (115/127)	90.6% (115/127)
調適後 平均辨識率					92.3% (313/339)

表 6-4 使用正規化特徵參數 MLBN Dependent Test 鑑別矩陣

辨識結果 測試情緒 (測試語料數量)	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	97% (60/62)	3% (2/62)	0	0	97% (60/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	0	89% (63/71)	11% (8/71)	89% (63/71)
生氣(127)	0	0.8% (1/127)	8.7% (11/127)	90.6% (115/127)	90.6% (115/127)
平均辨識率					93.5% (317/339)



表 6-5 MLBN 不同語系調適前後之鑑別矩陣(一半調適，一半測試)

調適前	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(300)	20.3% (61/300)	74.7% (224/300)	2.3% (7/300)	2.7% (8/300)	20.3% (61/300)
中性(300)	4.3% (13/300)	95.0% (285/300)	0	0.7% (2/300)	95.0% (285/300)
快樂(300)	46.3% (139/300)	8.0% (24/300)	31.3% (94/300)	14.3% (43/300)	31.3% (94/300)
生氣(300)	60.0% (180/300)	16.3% (49/300)	19.0% (57/300)	4.7% (14/300)	4.7% (14/300)
調適前 平均辨識率					37.83% (454/1200)
調適後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(300)	86.0% (258/300)	13.3% (40/300)	0	0.7% (2/300)	86.0% (258/300)
中性(300)	14.7% (44/300)	85.0% (255/300)	0	0.3% (1/300)	85.0% (255/300)
快樂(300)	0.3% (1/300)	0	83.3% (250/300)	16.3% (49/300)	83.3% (250/300)
生氣(300)	5.0% (15/300)	5.0% (15/300)	14.7% (44/300)	75.3% (226/300)	75.3% (226/300)
調適後 平均辨識率					82.42% (989/1200)

表 6-6 調適次數增加時，MLBN 不同語系調適前後之鑑別矩陣

調適前	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(600)	20.7% (124/600)	75.8% (455/600)	1.3% (8/600)	2.2% (13/600)	20.7% (124/600)
中性(600)	3.5% (21/600)	96.0% (576/600)	0.2% (1/600)	0.3% (2/600)	96.0% (576/600)
快樂(600)	42.8% (257/600)	8.2% (49/600)	35.0% (210/600)	14.0% (84/600)	35.0% (210/600)
生氣(600)	59.3% (356/600)	15.8% (95/600)	19.2% (115/600)	5.7% (34/600)	5.7% (34/600)
調適前 平均辨識率					39.33% (944/2400)
調適後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(600)	91.0% (546/600)	8.5% (51/600)	0	0.5% (3/600)	91.0% (546/600)
中性(600)	8.0% (48/600)	91.8% (551/600)	0	0.3% (1/600)	91.8% (551/600)
快樂(600)	0	0.3% (1/600)	88.7% (532/600)	11.2% (67/600)	88.7% (532/600)
生氣(600)	1.7% (10/600)	1.7% (10/600)	12.5% (75/600)	84.2% (505/600)	84.2% (505/600)
調適後 平均辨識率					88.92% (2134/2400)

表 6-7 MLBNC 德國語料庫同語系調適前後之鑑別矩陣

調適前	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	82% (51/62)	16% (10/62)	2% (1/62)	0	82% (51/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	3% (2/71)	62% (44/71)	35% (25/71)	62% (44/71)
生氣(127)	0	0	14.2% (18/127)	85.8% (109/127)	85.8% (109/127)
調適前 平均辨識率					83.5% (283/339)
調適後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(62)	100% (62/62)	0	0	0	100% (62/62)
中性(79)	0	100% (79/79)	0	0	100% (79/79)
快樂(71)	0	0	80% (57/71)	20% (14/71)	80% (57/71)
生氣(127)	0	0	3.1% (4/127)	96.9% (123/127)	96.9% (123/127)
調適後 平均辨識率					94.7% (321/339)

表 6-8 MLBNC 不同語系調適前後之鑑別矩陣(一半調適，一半測試)

調適前	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(300)	17.0% (51/300)	72.7% (218/300)	6.7% (20/300)	3.7% (11/300)	17.0% (51/300)
中性(300)	7.3% (22/300)	87.0% (261/300)	4.3% (13/300)	1.3% (4/300)	87.0% (261/300)
快樂(300)	4.0% (12/300)	4.0% (12/300)	43.0% (129/300)	49.0% (149/300)	43.0% (129/300)
生氣(300)	4.0% (12/300)	17.0% (51/300)	15.7% (47/300)	63.3% (190/300)	63.3% (190/300)
調適前 平均辨識率					52.58% (631/1200)
調適後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(300)	88.7% (266/300)	7.0% (21/300)	1.0% (3/300)	3.3% (10/300)	88.7% (266/300)
中性(300)	13.3% (40/300)	83.0% (249/300)	1.3% (4/300)	2.3% (7/300)	83.0% (249/300)
快樂(300)	0	0	64.7% (194/300)	35.3% (106/300)	64.7% (194/300)
生氣(300)	1.7% (5/300)	0.3% (1/300)	9.7% (29/300)	88.3% (265/300)	88.3% (265/300)
調適後 平均辨識率					81.17% (974/1200)

表 6-9 調適次數增加時，MLBN 不同語系調適前後之鑑別矩陣

調適前	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(600)	18.0% (108/600)	71.5% (429/600)	6.5% (39/600)	4.0% (24/600)	18.0% (108/600)
中性(600)	7.0% (42/600)	88.2% (529/600)	4.0% (24/600)	0.8% (5/600)	88.2% (529/600)
快樂(600)	2.8% (17/600)	5.0% (30/600)	39.7% (238/600)	52.5% (315/600)	39.7% (238/600)
生氣(600)	3.7% (22/600)	15.5% (93/600)	18.3% (110/600)	62.5% (375/600)	62.5% (375/600)
調適前 平均辨識率					52.08% (1250/2400)
調適後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
悲傷(600)	93.3% (560/600)	5.8% (35/600)	0.3% (1/600)	0.7% (4/600)	93.3% (560/600)
中性(600)	5.3% (32/600)	93.2% (559/600)	0	1.5% (9/600)	93.2% (559/600)
快樂(600)	0	0	80.3% (482/600)	19.7% (118/600)	80.3% (482/600)
生氣(600)	0.3% (2/600)	0.5% (3/600)	6.2% (37/600)	93.0% (558/600)	93.0% (558/600)
調適後 平均辨識率					89.96% (2159/2400)

表 6-10 各分類於不同測試條件下之整體辨識結果整理

測試條件		各分類器不同測試條件下之整體辨識正確率				
		KNN (總辨識個數/ 總測試個數)	SVM (總辨識個數/ 總測試個數)	MLBN (總辨識個數/ 總測試個數)	MLBNC (總辨識個數/ 總測試個數)	
Inside Test		88.8% (301/339)	100% (339/339)	81.1% (275/339)	88.8% (301/339)	
Outside Test		82.6% (280/339)	91.7% (311/339)	77.6% (263/339)	83.5% (283/339)	
不同語系之 Outside Test		34.21% (821/2400)	46.92% (1126/2400)	39.33% (944/2400)	52.08% (1250/2400)	
蒙地卡羅模擬		82.78% (33100/40k)	89.89% (35954/40k)	82.16% (32864/40k)	87.38% (34950/40k)	
同語系個人 調適	調適前	/		77.6% (263/339)	83.5% (283/339)	
	調適後			92.3% (313/339)	94.7% (321/339)	
不同語系個 人調適 (一半調適，一半 測試)	調適前	34.17% (410/1200)	/		37.83% (454/1200)	52.58% (631/1200)
	調適後	73.67% (884/1200)			82.42% (989/1200)	81.17% (974/1200)
不同語系個 人調適 (全部調適，全部 測試)	調適前	/		39.33% (944/2400)	52.08% (1250/2400)	
	調適後			88.92% (2134/2400)	89.96% (2159/2400)	

第七章 結論與未來展望

本研究主要的目的在於發展語音情緒辨識的方法，藉由比較使用者帶有情緒之語音訊號與資料庫間之音高、共振峰、音框能量以及梅爾倒頻譜係數之分佈狀態，分析判斷出此語音訊號之情緒反應。我們首先計算與情緒有關的語音特徵參數，並藉由統計與正規化之過程，縮小人與人之間在相關情緒特徵上的差異，以利後續相關之分類辨識使用，在研究中，我們利用現有之 KNN 及 SVM 分類方法，分別對原始與正規化特徵參數進行 Inside 及 Outside Test，由辨識結果顯示，使用正規化特徵參數可以有效縮小人與人之間在特徵參數上的差異，增加各情緒之間的分類效果，同時，我們針對 KNN 及 SVM 分類器在訓練語料與測試語料為不同語系之情況下進行辨識實驗，實驗中以德國語料庫所有資料當作訓練資料，並以工業技術研究院所錄製之語料當作測試語料，由實驗結果顯示，當 KNN 與 SVM 分類器在訓練語料與測試語料為不同語系時，兩者皆無法得到良好的分類效果。

接著，我們分別利用分類樹與貝氏網路的優點，將不同的特徵參數以分類樹分層解析方式，並結合貝氏網路之機率運算，發展出多層貝氏網路(MLBN)以及多層共變異數貝氏網路(MLBNC)語音情緒辨

識方法，在 MLBN 與 MLBNC 語音情緒辨識的過程中，首先，我們根據資料庫中各正規化後特徵參數之分佈狀態，以分類樹分層分群的方式，將使用之正規化後特徵參數分成三層，第一層為分兩群，主要有三種不同分群方式，第二層為分三群，主要有兩種不同分群方式，第三層則是悲傷、中性、快樂與生氣皆為獨立的群。接著，根據各特徵參數分層分群方式與貝氏決策原理，分別計算出測試語音之各情緒類別發生機率。在 MLBN 語音情緒辨識之計算過程中，各特徵參數之間彼此互為獨立參數，而 MLBNC 語音情緒辨識之計算過程，則是考慮各特徵參數彼此之間的相關性。同樣的，我們利用本文所發展之 MLBN 及 MLBNC 語音情緒辨識方法，分別進行 Inside、Outside 以及訓練語料與測試語料為不同語系之相關實驗，並利用蒙地卡羅模擬法對各分類器之辨識效果進行測試，根據各分類器之實驗結果，歸納出下列之結論：

1. 由 KNN、SVM、MLBN 之原始與正規化參數之 Outside Test 辨識結果顯示，使用正規化特徵參數可以有效縮小語者之間在特徵參數上的差異，增加各情緒之間的分類效果。
2. 由各分類器使用正規化後特徵參數 Inside Test 辨識結果比較，當測試資料在訓練資料中時，SVM 分類器之辨識結果明顯優於其他分類器之分類效果，KNN、MLBN 以及 MLBNC 對於中性情緒有較

佳的辨識效果，對於快樂情緒，辨識效果較差。

3. 由貝氏決策(BD)與多層貝氏網路(MLBN)之正規化後特徵參數 Inside Test 辨識結果，顯示透過分類樹與貝氏網路結合之多層貝氏網路語音情緒辨識方法，確實有助於改善單純使用貝氏決策之辨識效果，提高語音情緒之辨識正確率。
4. 根據各分類器正規化後特徵參數之 Outside Test 辨識結果，當測試資料不在訓練資料中時，由於快樂與生氣情緒之特徵參數分佈重疊性高，造成各分類器對於快樂與生氣之分類效果相對較差。
5. 根據蒙地卡羅模擬法對各分類器之辨識效果進行測試之結果，四種分類器對於中性均有極佳的辨識效果，而快樂則為最不容易辨識的情緒，各分類器對於快樂與生氣皆容易產生混淆的狀況。
6. 由各分類器可靠度測試辨識結果顯示，SVM 及 MLBNC 均提供了良好的辨識效果，由於 MLBNC 分類出來之結果為各情緒發生的機率，而非僅僅只是一個判斷結果，因此，在實際應用面上，MLBNC 分類器將比 SVM 分類器更可以提供較大且可靠的應用空間。
7. 由各分類器不同語系之辨識結果顯示，當訓練語料之語系與使用者語系不同時，由於發音方式及表達情緒方式上的不同，使得使用者各情緒之特徵參數與實際訓練之資料庫有所差異，造成各分

類器辨識效果均不佳之情況。

由 KNN、SVM、MLBN 以及 MLBNC 之 Inside 與 Outside Test 辨識結果以及各分類器於不同語系之測試辨識結果可以發現，當使用者資料在訓練資料中時，其辨識效果優於使用者資料不在訓練資料中，而當使用者之使用語系以及情緒表達方式與訓練語料庫有很大之差異時，四種分類器之均無法得到有效的分類效果，因此，如何透過適當的調適方式(Adaptive process)調整資料庫中的資料庫狀態，進而改善各情緒之分類效果將是語音情緒辨識的另一個課題所在。在本研究中，我們提出適應性 MLBN 與適應性 MLBNC 語音情緒辨識方法，並透過相關適應性實驗來驗證其可行性，由適應性實驗結果可以歸納出下列結論：

1. 由適應性 KNN、適應性 MLBN 以及適應性 MLBNC 語音情緒調適後辨識結果顯示，透過資料庫修正，可以有效提升辨識效果，提高正確辨識率。
2. 由適應性 KNN、適應性 MLBN 以及適應性 MLBNC 語音情緒之調適後辨識結果比較，適應性 MLBN 與適應性 MLBNC 語音情緒辨識方法於資料庫修正後，其辨識效果明顯優於適應性 KNN。
3. 由適應性 MLBN 以及適應性 MLBNC 語音情緒辨識實驗結果顯示，增加 MLBN 以及 MLBNC 調適次數將有助於提升 MLBN 及

MLBNC 之語音情緒分類效果。

4. 在 MLBN 及 MLBNC 之調適計算上，MLBN 之修正過程僅需計算資料庫中特徵參數平均值與標準差，而 MLBNC 之修正過程也僅需計算資料庫中特徵參數向量之平均值與共變異數，不需額外增加資料量與硬體需求，因此，透過適應性 MLBN 與適應性 MLBNC 之分類與調適方式，將有助於未來在硬體上實現語音情緒之相關辨識。

語音情緒辨識為未來智慧生活上人機介面的課題之一，如何使機器從使用者之語音訊號上得到相關的資訊，進而做出正確的辨識判斷結果，將是此研究未來仍須繼續努力的方向，對於後續之研究有幾個方向可以持續做深入探討。

1. 將語音情緒辨識所使用之特徵參數與語音辨識上所使用之特徵參數相結合，進而使得實際應用上，語音辨識的同時也能偵測使用者當時的情緒反應。
2. 目前所使用之特徵參數，對於各情緒類別並未能有明顯且關鍵的分辨效果，因此，找出能明顯區隔各情緒的特徵參數也是後續的研究課題之一。
3. 語音情緒辨識之使用環境通常包含雜訊或不同背景之噪音，如何從含雜訊或背景噪音之語音訊號中，得到語音情緒辨識所需之正

確情緒相關特徵資料，也是語音情緒辨識結果正確與否的重要關鍵之一。

4. 由於語音情緒辨識實際應用時所使用之訓練語料通常不包含使用者真實的情緒資料，因此，如何透過適當且快速的調適過程，使新資料庫更貼近使用者之情緒狀態，將是語音情緒辨識相關研究另一個發展的重點所在。
5. 在調適的過程中，使用者如何將真實的情緒反應回饋給辨識系統，也是適應性調適方法所必須面對的問題，因此，適當且正確的回饋機制將是調適結果優劣的重要關鍵之一。

而在語音情緒辨識之實際應用上，有幾個應用平台可以當作未來技術發展的參考。

1. 遠距醫療照護：在不久的將來，遠距醫療服務將是趨勢所在，在人機互動的過程中，如何偵測使用者的情緒狀態，並適時給於回饋，將是未來語音情緒辨識技術可以切入與發揮的所在。
2. Call Centers：目前客服系統已經廣泛應用語音辨識技術，在使用的過程中，消費者的情緒反應若可以同時被偵測，不僅可以避免消費者的生氣情況產生，也可以提供更好更適合的服務。
3. 安全監控系統：在飛航通關或安全人員檢測等相關安全監控系統上，若能將說話者語音情緒反應做及時監測，並同時結合影像辨

識，將有助於減少相關安全人員的負擔，提供更快速便利的安全環境。

4. 互動玩具與機器人：在科技越來越發達的年代，互動玩具及機器人將是未來產品的趨勢，因此，如何讓機器人或互動玩具具有辨識情緒的功能，也將是未來科技發展的重要方向之一。



參考文獻

1. M. Álvarez, R. Galán, F. Matía, D. Rodríguez-Losada, A. Jiménez, “An Emotional Model for a Guide Robot,” *IEEE Trans., Systems, Man and Cybernetics, Part A: Systems and Humans*, Vol. 40, No. 5, pp. 982-992, 2010.
2. J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke, “Prosody-based automatic detection of annoyance and frustration in human-computer dialog,” in *Proc. of the International Conference on Spoken Language Processing (ICSLP)*, pp. 2037-2040, 2002.
3. R. Banse and K. R. Scherer, “Acoustic profiles in vocal emotion expression,” *Journal of Personality and Social Psychology*, Vol. 70, pp. 614-6362, 1996.
4. F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss, “A Database of German Emotional Speech,” *Proc. INTERSPEECH (ISCA)*, pp.1517-1520, 2005.
5. C. Busso, S. Lee, and S. Narayanan, "Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 17, pp. 582-596, 2009.
6. S. Casale, A. Russo, G. Scebba, and S. Serrano, "Speech Emotion Classification Using Machine Learning Algorithms," *IEEE International Conference on Semantic Computing*, pp. 158-165, 2008.
7. S. Chandrakala, C. C. Sekhar, “Combination of generative models and SVM based classifier for speech emotion recognition,” *International Joint Conference on Neural Networks*, pp. 497-502, 2009.
8. C.-C. Chang and C.-J. Lin, LIBSVM: a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
9. Chui, Charles K., *An Introduction to Wavelets*, San Diego, Academic Press, 1992.
10. D. Chen, D. Jiang, I. Ravayse, H. Sahli, “Audio-Visual Emotion Recognition Based on a DBN Model with Constrained Asynchrony,” *International Conference on Image and Graphics*, pp 912-916, 2009
11. Z. J. Chuang and C. H. Wu, “Emotion recognition using acoustic features and textual content,” *International Conference on Multimedia and Expo*, Vol. 1, pp. 53-56, 2004

12. Cooley, James W., and John W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.* 19: 297-301, 1965.
13. C. Cortes and V. Vapnik, "Support-Vector Networks", *Machine Learning*, 1995.
14. R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, and W. Fellenz, "Emotion recognition in human-computer interaction," in *Proc. of the IEEE Signal Processing Magazine*, Vol. 18, pp. 32-80, 2001.
15. R. Cowie and R. R. Cornelius, "Describing the emotional states that are expressed in speech," *Speech Communication*, Vol. 40, pp. 5-32, 2003.
16. S. Das, A. Halder, P. Bhowmik, A. Chakraborty, A. Konar, R. Janarthanan, "A Support Vector Machine Classifier of Emotion from Voice and Facial Expression Data," *Nature & Biologically Inspired Computing*, pp. 1010-1015, 2009
17. W. Dongrui, D. P. Thomas, M. Emily, N. Shrikanth, "Speech Emotion Estimation in 3D Space," *IEEE International Conference on Multimedia & Expo*, pp. 737-742, 2010.
18. E. Douglas-Cowie, R. Cowie, and M. Schroeder, "A New Emotion Database: Considerations, Sources and Scope," in *Proc. ISCA Workshop Speech and Emotion : A conceptual framework for research*, pp. 39-44, 2000.
19. H. P. Espinosa, C. A. R. García, L. V. Pineda, "Features selection for primitives estimation on emotional speech," *IEEE International Conference on Acoustics Speech and Signal Processing*, pp. 5138-5141, 2010.
20. B. Fasel and J. Luetin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, Vol. 36, No.1, pp. 259-275, 2003.
21. N. Fragopanagos and J. Taylor, "Emotion recognition in human-computer interaction," *Neural Networks*, Special Issue, pp. 1-17, 2005.
22. M. Fujita, Y. Kuroki, T. Ishida and T.T. Doi, "A Small Humanoid Robot SDR-4X for Entertainment Applications," *International Conference on Advanced Intelligent Mechatronics*, pp. 938-943, 2003
23. M. Fujita, "On Activating Human Communications with Pet-type Robot AIBO," *Proc. of the IEEE*, Vol. 92, No. 11, pp. 1804-1813, 2004.

24. H. J. Go, K.C. Kwak, D.J. Lee and M.G. Chun, "Emotion recognition from the facial image and speech signal," *SICE Annual Conference*, Vol. 3, pp. 2890-2895, 2003.
25. F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, Vol. 66, pp. 51-83, 1978.
26. R. Huber, A. Batliner, J. Buckow, E. Noth, V. Warnke, and H. Niemann, "Recognition of emotion in a realistic dialogue scenario," in *Proc. of the International Conference on Spoken Language Processing (ICSLP)*, Vol. 1, pp. 665-668, 2000.
27. A. Iida, N. Campbell, S. Iga, F. Higuchi, and M. Yasumura, "A speech synthesis system with emotion for assisting communication", in *Proc. of the ISCA Workshop (ITRW) on Speech and Emotion: A conceptual framework for research*, pp. 167-172, 2002.
28. T. Iliou, C. N. Anagnostopoulos, "Comparison of Different Classifiers for Emotion Recognition," *PCI ' Panhellenic Conference on Informatics*, pp. 102-106, 2009.
29. D.N. Jiang and L.H. Cai, "Speech Emotion Classification with the Combination of Statistic Features and Temporal Features", *IEEE International Conference on Multimedia and Expo* , pp. 1967-1970, 2004
30. T. Johnstone, "Emotional speech elicited using computer games," In *Proc. of the International Conference on Spoken Language Processing*, pp. 1985-1988, 1996.
31. S. Kadambe and G. F. Boudreaux-Bartels, "Application of the wavelet transform for pitch detection of speech signals," *IEEE Trans., Information Theory*, Vol. 38, No. 2, pp. 917-924, 1992.
32. A. B. Kandali, A. Routray, T. K. Basu, "Comparison of Features Based on MFCCs and Eigen Values of Autocorrelation Matrix for Cross-Lingual Vocal Emotion Recognition in Five Languages of Assam," *Annual IEEE India Conference*, pp. 1-4, 2009.
33. O. W. Kwon, K. Chang, J. Hao, and T. W. Lee, "Emotion recognition by speech signals," in *Proc. of the 8th European Conference on Speech Communication and Technology*, pp. 125-128, 2003
34. C. M. Lee, S. Narayanan, and R. Pieraccini, "Classifying emotions in human-machine spoken dialogs," in *Proc. of the International Conference on Multimedia and Expo*, pp. 737-740, 2002.
35. C. M. Lee, S. Yildirim, M. Bulut, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee, and S. Narayanan, "Emotion recognition based on phoneme

- classes,” in *Proc. of the International Conference on Spoken Language Processing (ICSLP)*, pp. 889-892, 2004.
36. Y. L. Lin and G. Wei, “Speech emotion recognition based on HMM and SVM,” in *Proc. of the 2005 International Conference on Machine Learning and Cybernetics*, Vol. 8, pp. 4898-4901, 2005.
 37. M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with gabor wavelets,” in *Proc. third IEEE Int. C. on Automatic Face and Gesture Recognition*, pp. 200-205, 1998.
 38. V. Makarova and V. A. Petrushin, “RUSLANA: A database of Russian Emotional Utterances,” in *Proc. of the 2002 Int. Conf. Spoken Language Processing (ICSLP)*, pp. 2041-2044, 2002.
 39. A. P. Meshram, S. D. Shirbahadurkar, A. Kohok, S. Jadhav, “An Overview and Preparation for Recognition of Emotion from Speech Signal with Multi Modal Fusion,” *Computer and Automation Engineering (ICCAE)* , pp. 446-452, 2010
 40. A. Metallinou, S. Lee and S. Narayanan, “Audio-Visual Emotion Recognition using Gaussian Mixture Models for Face and Voice,” *Proceedings of IEEE International Symposium of Multimedia*, pp 250-257, 2008
 41. J. M. Montero, J. Gutierrez-Arriola, J. Colas, E. Enriquez, and J. M. Pardo, “Analysis and modelling of emotional speech in Spanish”, in *Proc. of the ICPhS'99*, pp. 957-960, 1999.
 42. E. Mower, M. J. Mataric, S. Narayanan, “Human Perception of Audio-Visual Synthetic Character Emotion Expression in the Presence of Ambiguous and Conflicting Information,” *IEEE Transactions on Multimedia*, Vol. 11, No. 5, pp. 843-855, 2009.
 43. C. Nadeu, J. Pascual, J. Hernando, “Pitch determination using the cepstrum of the one-sided autocorrelation sequence,” *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 5, pp. 3677 - 3680, 1991
 44. T. L. New, S. W. Foo, and L. C. De Silva, “Speech emotion recognition using hidden Markov models,” *Speech Communication*, Vol. 41, pp. 603-623, 2003.
 45. Y. Niimi, M. L. Kasamatu, T. Nishimoto, and M. Araki, “Synthesis of emotional speech using prosodically balanced VCV Segments,” in *Proc. of the 4th ISCA tutorial and Workshop on research synthesis*, paper 133, 2001.
 46. M. Pantic and J. M. Rothkrantz, “Toward an affect-sensitive multimodal human-computer interaction,” *Proc. IEEE*, Vol. 91, No.9, pp. 1370-1390, 2003.

47. T. L. Pao and Y. T. Chen, "Emotion recognition in Mandarin speech and its application in training of hearing impaired," *National Computer Symposium*, pp. 379-391, 2003.
48. T. L. Pao, Y. T. Chen, J. H. Yeh, and W. Y. Liao, "Combining acoustic features for improved emotion recognition in Mandarin speech," *Affective Computing & Intelligent Interaction (ACII), Lecture Notes in Computer Science*, Vol. 3784, pp. 279-285, 2005.
49. V. Petrushin, "Emotion in speech: Recognition and application to call centers," in *Proc. of Artificial Neural Networks in Engineering*, pp. 7-10, 1999.
50. V. Petrushin, "Emotion recognition in speech signal: Experimental study, development, and application," in *Proc. of the Sixth International Conference on Spoken Language Processing (ICSLP)*, pp. 222-225, 2000.
51. R. W. Picard, *Affective computing*, The MIT Press, United States, 1998
52. T. S. Polzin and A. Waibel, "Emotion-sensitive human-computer interfaces," in *Proc. of the ISCA Workshop on Speech and Emotion*, pp. 201-206, 2000.
53. L. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 25, No.1, pp. 24-33, 1977.
54. K. R. Scherer, "A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology," In *Proc. of the ICSLP*, pp. 379-382, 2000.
55. K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Communication*, Vol. 40, pp. 227-256, 2003.
56. F. Schiel, Steininger Silke, and Turk Ulrich, "The Smartkom Multimodal Corpus at BAS," in *Proc. of the Language Resources and Evaluation*, pp. 35-41, 2002.
57. M. Schroder and M. Grice, "Expressing vocal effort in concatenative synthesis", in *Proc. of the 15th Int. Conf. Phonetic Sciences*, pp. 2589-2592, 2003.
58. B. Schuller, G. Rigoll and M. Lang, "Hidden Markov Model-based Speech Emotion Recognition", *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp 1-4, 2003
59. B. Schuller, B. Vlasenko, F. Eyben, G. Rigoll, A. Wendemuth,

- “Acoustic Emotion Recognition-A Benchmark Comparison of Performances,” *IEEE Workshop, Automatic Speech Recognition & Understanding*, pp. 552-557, 2009
60. H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi, “Recognition of Six Basic Facial Expressions by Feature-Points Tracking using RBF Neural Network and Fuzzy Inference System,” *Proc. of the 2004 IEEE International Conference on Multimedia and Expo*, pp. 1219-1222, 2004
 61. Shakhnarovich, Darrell, and Indyk, ed, *Nearest-Neighbor Methods in Learning and Vision*, MIT Press, 2005.
 62. J. Tao, Y. Kang, A. Li, “Prosody conversion from neutral speech to emotional speech,” *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14 No. 4, pp.1145–1154, 2006.
 63. S. Wang, X. Ling, F. Zhang, J. Tong, “Speech Emotion Recognition Based on Principal Component Analysis and Back Propagation Neural Network,” *International Conference on Measuring Technology and Mechatronics Automation*, pp. 437–440, 2010.
 64. Y. Wang, L. Guan, “Recognizing Human Emotional State From Audiovisual Signals,” *IEEE Transactions on Multimedia*, Vol. 10, No. 4, pp. 659-668, 2008.
 65. T. Wilhelm, H.J. Bohme, H.M. Grofi and A. Backhaus, “Statistical and Neural Methods for Vision-based Analysis of Facial Expressions and Gender,” *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 3, pp. 2203-2208, 2004
 66. C. H. Wu, Z. J. Chuang, Y. C. Lin, “Emotion recognition from text using semantic labels and separable mixture models,” *ACM Transactions on Asian Language Information Processing*, Vol. 5, Issue 2, 2006.
 67. S. Wu, T. H. Falk, W.-Y. Chan, “Automatic Recognition of Speech Emotion using Long-term Spectro-temporal Features,” *16th International Conference on Digital Signal Processing*, pp. 1-6, 2009.
 68. T. Wu, Y. Yang, Z. Wu, and D. Li, “MASC: A speech corpus in Mandarin for emotion analysis and affective speaker recognition,” in *Proc. of the IEEE Workshop on Speaker and Language Recognition*, pp. 1-5, 2006.
 69. L. Xin, L. Xiang, "Novel Hilbert Energy Spectrum Based Features for Speech Emotion Recognition," *International Conference on Information Engineering*, Vol. 1, pp.189-193, 2010.
 70. S. Yacoub, S. Simske, X. Lin, and J. Burns, “Recognition of emotion

in interactive voice response systems,” in *Proc. of Eurospeech, 8th European Conference on Speech Communication and Technology*, 2003.

71. I. Yanushevskaya, C. Gobl, and A. Chasaide, “Voice quality and f0 cues for affect expression: implications for synthesis,” In *Proc. of the EUROSPEECH, Interspeech 2005*, pp. 1849-1852, 2005.
72. Z. Zeng, M. Pantic, G. I. Roisman, T. S. Huang, “A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 1, pp. 39-58, 2009.
73. S. Zhang, “Speech emotion recognition based on Fuzzy Least Squares Support Vector Machines,” *World Congress on Intelligent Control and Automation*, pp. 1299-1302, 2008.
74. 林宗勳, Support Vector Machines 簡介,
<http://www.cmlab.csie.ntu.edu.tw/~cyy/learning/tutorials/SVM2.pdf>



附錄 A

表 A-1 KNN Outside Test 各情緒辨識正確率(原始特徵參數)

各情緒辨識正確率 測試人員 (測試語料數量)	悲傷 (辨識正確個數/測試個數)	中性 (辨識正確個數/測試個數)	快樂 (辨識正確個數/測試個數)	生氣 (辨識正確個數/測試個數)	辨識正確率
P1(39)	0 (0/7)	100% (11/11)	70% (5/7)	79% (11/14)	69% (27/39)
P2(42)	70% (6/9)	100% (10/10)	82% (9/11)	100% (12/12)	88% (37/42)
P3(30)	75% (3/4)	100% (9/9)	50% (2/4)	92% (12/13)	87% (26/30)
P4(21)	67% (2/3)	100% (4/4)	100% (4/4)	100% (10/10)	95% (20/21)
P5(35)	43% (3/7)	100% (9/9)	25% (2/8)	82% (9/11)	66% (23/35)
P6(22)	0 (0/4)	50% (2/4)	0 (0/2)	100% (12/12)	64% (14/22)
P7(36)	100% (5/5)	100% (9/9)	50% (5/10)	100% (12/12)	86% (31/36)
P8(41)	60% (6/10)	100% (7/7)	13% (1/8)	100% (16/16)	73% (30/41)
P9(34)	75% (3/4)	100% (11/11)	50% (3/6)	100% (13/13)	88% (30/34)
P10(39)	78% (7/9)	100% (5/5)	9% (1/11)	100% (14/14)	69% (27/39)
平均辨識率	56% (35/62)	97% (77/79)	45% (32/71)	95.3% (121/127)	78.2% (265/339)

表 A-2 KNN Outside Test 各情緒辨識正確率(正規化特徵參數)

各情緒辨識正確率 測試人員 (測試語料數量)	悲傷 (辨識正確個數/測試個數)	中性 (辨識正確個數/測試個數)	快樂 (辨識正確個數/測試個數)	生氣 (辨識正確個數/測試個數)	辨識正確率
P1(39)	0 (0/7)	100% (11/11)	43% (3/7)	86% (12/14)	67% (26/39)
P2(42)	89% (8/9)	100% (10/10)	45% (5/11)	100% (12/12)	83% (35/42)
P3(30)	50% (2/4)	100% (9/9)	50% (2/4)	92% (12/13)	83% (25/30)
P4(21)	100% (3/3)	100% (4/4)	100% (4/4)	100% (10/10)	100% (21/21)
P5(35)	86% (6/7)	100% (9/9)	100% (8/8)	82% (9/11)	91% (32/35)
P6(22)	75% (3/4)	100% (4/4)	0 (0/2)	100% (12/12)	86% (19/22)
P7(36)	100% (5/5)	100% (9/9)	80% (8/10)	100% (12/12)	94% (34/36)
P8(41)	70% (7/10)	100% (7/7)	13% (1/8)	100% (16/16)	76% (31/41)
P9(34)	75% (3/4)	100% (11/11)	50% (3/6)	92% (12/13)	85% (29/34)
P10(39)	44% (4/9)	100% (5/5)	45% (5/11)	100% (14/14)	72% (28/39)
平均辨識率	66% (41/62)	100% (79/79)	55% (39/71)	95.3% (121/127)	82.6% (280/339)

表 A-3 SVM Outside Test 各情緒辨識正確率(原始特徵參數)

各情緒辨識正確率 測試人員 (測試語料數量)	悲傷 (辨識正確個數/測試個數)	中性 (辨識正確個數/測試個數)	快樂 (辨識正確個數/測試個數)	生氣 (辨識正確個數/測試個數)	辨識正確率
P1(39)	14% (1/7)	100% (11/11)	100% (7/7)	93% (13/14)	82% (32/39)
P2(42)	100% (9/9)	100% (10/10)	82% (9/11)	83% (10/12)	90% (38/42)
P3(30)	50% (2/4)	100% (9/9)	25% (1/4)	100% (13/13)	83% (25/30)
P4(21)	33% (1/3)	100% (4/4)	100% (4/4)	100% (10/10)	90% (19/21)
P5(35)	86% (6/7)	100% (9/9)	100% (8/8)	82% (9/11)	91% (32/35)
P6(22)	100% (4/4)	100% (4/4)	100% (2/2)	100% (12/12)	100% (22/22)
P7(36)	100% (5/5)	100% (9/9)	100% (10/10)	83% (10/12)	94% (34/36)
P8(41)	60% (6/10)	100% (7/7)	88% (7/8)	100% (16/16)	88% (36/41)
P9(34)	100% (4/4)	100% (11/11)	100% (6/6)	92% (12/13)	97% (33/34)
P10(39)	67% (6/9)	100% (5/5)	55% (6/11)	100% (14/14)	79% (31/39)
平均辨識率	71% (44/62)	100% (79/79)	85% (60/71)	93.7% (119/127)	89.1% (302/339)

表 A-4 SVM Outside Test 各情緒辨識正確率(正規化特徵參數)

各情緒辨識正確率 測試人員 (測試語料數量)	悲傷 (辨識正確個數/測試個數)	中性 (辨識正確個數/測試個數)	快樂 (辨識正確個數/測試個數)	生氣 (辨識正確個數/測試個數)	辨識正確率
P1(39)	29% (2/7)	100% (11/11)	43% (3/7)	93% (13/14)	74% (29/39)
P2(42)	100% (9/9)	100% (10/10)	91% (10/11)	100% (12/12)	98% (41/42)
P3(30)	75% (3/4)	100% (9/9)	50% (2/4)	100% (13/13)	90% (27/30)
P4(21)	100% (3/3)	100% (4/4)	100% (4/4)	100% (10/10)	100% (21/21)
P5(35)	86% (6/7)	100% (9/9)	100% (8/8)	82% (9/11)	91% (32/35)
P6(22)	75% (3/4)	100% (4/4)	100% (2/2)	100% (12/12)	95% (21/22)
P7(36)	100% (5/5)	100% (9/9)	80% (8/10)	100% (12/12)	94% (34/36)
P8(41)	90% (9/10)	100% (7/7)	50% (4/8)	100% (16/16)	88% (36/41)
P9(34)	75% (3/4)	100% (11/11)	100% (6/6)	100% (13/13)	97% (33/34)
P10(39)	89% (8/9)	100% (5/5)	91% (10/11)	100% (14/14)	95% (37/39)
平均辨識率	82% (51/62)	100% (79/79)	80% (57/71)	97.6% (124/127)	91.7% (311/339)

表 A-5 MLBN Outside Test 各情緒辨識正確率(原始特徵參數)

各情緒辨識正確率 測試人員 (測試語料數量)	悲傷 (辨識正確個數/測試個數)	中性 (辨識正確個數/測試個數)	快樂 (辨識正確個數/測試個數)	生氣 (辨識正確個數/測試個數)	辨識正確率
P1(39)	0 (0/7)	100% (11/11)	29% (2/7)	50% (7/14)	51% (20/39)
P2(42)	44% (4/9)	100% (10/10)	64% (7/11)	58% (7/12)	67% (28/42)
P3(30)	75% (3/4)	100% (9/9)	0 (0/4)	92% (12/13)	80% (24/30)
P4(21)	100% (3/3)	75% (3/4)	50% (2/4)	60% (6/10)	67% (14/21)
P5(35)	86% (6/7)	100% (9/9)	38% (3/8)	64% (7/11)	71% (25/35)
P6(22)	0 (0/4)	100% (4/4)	0 (0/2)	67% (8/12)	55% (12/22)
P7(36)	100% (5/5)	100% (9/9)	70% (7/10)	92% (11/12)	89% (32/36)
P8(41)	70% (7/10)	100% (7/7)	38% (3/8)	94% (15/16)	78% (32/41)
P9(34)	100% (4/4)	64% (7/11)	100% (6/6)	92% (12/13)	85% (29/34)
P10(39)	78% (7/9)	0 (0/5)	0 (0/11)	100% (14/14)	54% (21/39)
平均辨識率	63% (39/62)	87% (69/79)	42% (30/71)	78.0% (99/127)	69.9% (237/339)

表 A-6 MLBN Outside Test 各情緒辨識正確率(正規化特徵參數)

各情緒辨識正確率 測試人員 (測試語料數量)	悲傷 (辨識正確個數/測試個數)	中性 (辨識正確個數/測試個數)	快樂 (辨識正確個數/測試個數)	生氣 (辨識正確個數/測試個數)	辨識正確率
P1(39)	71% (5/7)	100% (11/11)	71% (5/7)	43% (6/14)	69% (27/39)
P2(42)	100% (9/9)	90% (9/10)	64% (7/11)	83% (10/12)	83% (35/42)
P3(30)	50% (2/4)	100% (9/9)	0 (0/4)	92% (12/13)	77% (23/30)
P4(21)	100% (3/3)	100% (4/4)	75% (3/4)	50% (5/10)	71% (15/21)
P5(35)	86% (6/7)	100% (9/9)	75% (6/8)	55% (6/11)	77% (27/35)
P6(22)	75% (3/4)	100% (4/4)	0 (0/2)	50% (6/12)	59% (13/22)
P7(36)	100% (5/5)	100% (9/9)	100% (10/10)	92% (11/12)	97% (35/36)
P8(41)	100% (10/10)	100% (7/7)	25% (2/8)	94% (15/16)	83% (34/41)
P9(34)	75% (3/4)	91% (10/11)	83% (5/6)	85% (11/13)	85% (29/34)
P10(39)	67% (6/9)	100% (5/5)	0 (0/11)	100% (14/14)	64% (25/39)
平均辨識率	84% (52/62)	97% (77/79)	54% (38/71)	75.6% (96/127)	77.6% (263/339)

表 A-7 MLBNC Outside Test 各情緒辨識正確率(正規化特徵參數)

各情緒辨識正確率 測試人員 (測試語料數量)	悲傷 (辨識正確個數/測試個數)	中性 (辨識正確個數/測試個數)	快樂 (辨識正確個數/測試個數)	生氣 (辨識正確個數/測試個數)	辨識正確率
P1(39)	57% (4/7)	100% (11/11)	71% (5/7)	79% (11/14)	79% (31/39)
P2(42)	89% (8/9)	100% (10/10)	73% (8/11)	42% (5/12)	74% (31/42)
P3(30)	25% (1/4)	100% (9/9)	25% (1/4)	100% (13/13)	80% (24/30)
P4(21)	67% (2/3)	100% (4/4)	50% (2/4)	90% (9/10)	81% (17/21)
P5(35)	100% (7/7)	100% (9/9)	63% (5/8)	91% (10/11)	89% (31/35)
P6(22)	50% (2/4)	100% (4/4)	50% (1/2)	100% (12/12)	86% (19/22)
P7(36)	100% (5/5)	100% (9/9)	100% (10/10)	67% (8/12)	89% (32/36)
P8(41)	100% (10/10)	100% (7/7)	38% (3/8)	94% (15/16)	85% (35/41)
P9(34)	100% (4/4)	100% (11/11)	83% (5/6)	92% (12/13)	94% (32/34)
P10(39)	89% (8/9)	100% (5/5)	36% (4/11)	100% (14/14)	79% (31/39)
平均辨識率	82% (51/62)	100% (79/79)	62% (44/71)	85.8% (109/127)	83.5% (283/339)

附錄 B

表 B-1 正規化後音高平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	70.07%	99.13%	99.53%
中性	87.77%	1	99.40%	99.69%
快樂	96.57%	96.08%	1	41.30%
生氣	98.36%	98.10%	67.89%	1

表 B-2 正規化後音高之標準差各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	62.30%	88.25%	85.35%
中性	89.49%	1	97.41%	96.96%
快樂	79.22%	86.33%	1	38.55%
生氣	80.70%	85.52%	73.59%	1

表 B-3 正規化後能量平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	56.30%	48.10%	72.08%
中性	86.44%	1	88.01%	94.84%
快樂	77.13%	68.67%	1	73.72%
生氣	78.86%	86.42%	63.61%	1

表 B-4 正規化後能量之標準差各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	74.72%	73.99%	68.76%
中性	90.83%	1	72.28%	76.62%
快樂	88.32%	35.97%	1	71.96%
生氣	83.36%	41.67%	41.18%	1

表 B-5 正規化後第一共振峰平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	62.22%	53.51%	70.80%
中性	81.42%	1	77.72%	83.65%
快樂	62.85%	50.30%	1	71.09%
生氣	34.37%	60.58%	46.93%	1

表 B-6 正規化後第一共振峰之標準差各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	49.11%	78.00%	76.76%
中性	75.18%	1	77.83%	80.56%
快樂	79.48%	67.53%	1	73.04%
生氣	69.64%	54.10%	38.83%	1

表 B-7 正規化後第二共振峰平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	62.27%	67.67%	70.36%
中性	83.75%	1	77.05%	80.56%
快樂	70.71%	41.73%	1	72.71%
生氣	58.19%	46.58%	38.07%	1

表 B-8 正規化後第二共振峰之標準差各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	76.83%	86.95%	86.44%
中性	83.69%	1	82.36%	78.92%
快樂	78.54%	50.71%	1	36.61%
生氣	83.70%	51.48%	72.84%	1

表 B-9 正規化後第三共振峰平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	60.34%	79.64%	76.05%
中性	85.39%	1	85.77%	84.20%
快樂	62.20%	55.03%	1	47.47%
生氣	56.00%	52.27%	63.31%	1

表 B-10 正規化後第三共振峰之標準差各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	51.12%	64.65%	69.89%
中性	79.59%	1	76.24%	79.35%
快樂	44.82%	43.36%	1	67.21%
生氣	47.18%	44.85%	40.61%	1

表 B-11 MFCC C1 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	59.80%	91.33%	94.98%
中性	62.48%	1	93.23%	96.46%
快樂	74.00%	80.67%	1	61.61%
生氣	87.07%	91.02%	62.62%	1

表 B-12 MFCC C2 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	82.94%	97.01%	99.16%
中性	60.55%	1	89.53%	94.60%
快樂	90.42%	77.50%	1	54.51%
生氣	98.02%	91.77%	72.80%	1

表 B-13 MFCC C3 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	69.75%	92.55%	93.38%
中性	55.06%	1	85.16%	87.66%
快樂	90.45%	83.68%	1	73.44%
生氣	89.06%	81.95%	38.41%	1

表 B-14 MFCC C4 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	82.08%	93.29%	96.45%
中性	84.87%	1	87.23%	90.87%
快樂	85.82%	62.06%	1	57.12%
生氣	93.42%	77.51%	64.60%	1

表 B-15 MFCC C5 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	75.92%	94.80%	92.27%
中性	66.74%	1	80.08%	74.71%
快樂	95.12%	85.53%	1	63.13%
生氣	92.21%	80.53%	51.28%	1

表 B-16 MFCC C6 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	75.63%	84.54%	87.09%
中性	51.27%	1	69.22%	78.36%
快樂	75.28%	66.07%	1	74.45%
生氣	67.04%	53.68%	38.48%	1

表 B-17 MFCC C7 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	68.45%	86.09%	84.92%
中性	66.63%	1	78.54%	74.57%
快樂	77.83%	61.27%	1	34.96%
生氣	80.29%	64.69%	71.33%	1

表 B-18 MFCC C8 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	33.31%	69.07%	76.29%
中性	69.03%	1	69.50%	76.96%
快樂	80.07%	78.94%	1	74.65%
生氣	76.44%	75.43%	40.68%	1

表 B-19 MFCC C9 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	58.96%	76.26%	82.65%
中性	48.22%	1	73.59%	79.33%
快樂	64.96%	61.41%	1	51.42%
生氣	84.55%	82.44%	76.35%	1

表 B-20 MFCC C10 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	36.65%	55.45%	52.66%
中性	72.58%	1	65.69%	62.61%
快樂	80.07%	76.55%	1	34.92%
生氣	82.78%	79.17%	70.78%	1

表 B-21 MFCC C11 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	73.72%	57.90%	71.25%
中性	43.17%	1	37.43%	39.18%
快樂	61.61%	73.54%	1	72.15%
生氣	38.77%	67.78%	39.44%	1

表 B-22 MFCC C12 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	50.36%	80.97%	81.41%
中性	69.91%	1	84.33%	82.88%
快樂	47.17%	52.64%	1	54.84%
生氣	47.96%	50.10%	59.24%	1

表 B-23 MFCC C13 平均值各情緒之正確辨識率矩陣

次辨識類別 主辨識類別	悲傷	中性	快樂	生氣
悲傷	1	76.98%	73.77%	77.82%
中性	41.64%	1	41.99%	51.24%
快樂	40.37%	73.68%	1	72.01%
生氣	43.35%	55.81%	50.62%	1

附錄 C

表 C-1 KNN 不同語系 P1 ~ P20 調適前後之辨識結果(一半調適，一半測試)

	調適前後	悲傷(15)	中性(15)	快樂(15)	生氣(15)	辨識正確率 (60)
P1(60)	調適前	0	100%	0	0	25%
	調適後	33%	67%	47%	53%	50%
	調適次數	10	2	6	6	
P2(60)	調適前	0	100%	40%	0	35%
	調適後	80%	80%	80%	67%	77%
	調適次數	8	1	8	7	
P3(60)	調適前	0	93%	27%	0	30%
	調適後	67%	47%	67%	80%	65%
	調適次數	12	5	8	7	
P4(60)	調適前	0	100%	33%	7%	35%
	調適後	60%	67%	80%	73%	70%
	調適次數	8	1	10	8	
P5(60)	調適前	0	100%	67%	0	42%
	調適後	93%	60%	73%	67%	73%
	調適次數	10	3	5	6	
P6(60)	調適前	0	100%	13%	0	28%
	調適後	73%	80%	73%	60%	72%
	調適次數	9	1	6	8	
P7(60)	調適前	0	93%	27%	0	30%
	調適後	27%	67%	80%	73%	62%
	調適次數	10	8	8	6	
P8(60)	調適前	0	100%	13%	0	28%
	調適後	87%	80%	87%	60%	78%
	調適次數	9	5	6	8	
P9(60)	調適前	0	100%	40%	0	35%
	調適後	80%	93%	87%	73%	83%
	調適次數	8	3	6	6	
P10(60)	調適前	0	100%	40%	0	35%
	調適後	53%	87%	87%	93%	80%
	調適次數	9	1	9	9	

	調適前後	悲傷(15)	中性(15)	快樂(15)	生氣(15)	辨識正確率 (60)
P11(60)	調適前	80%	100%	0	7%	47%
	調適後	100%	100%	93%	60%	88%
	調適次數	3	0	6	6	
P12(60)	調適前	73%	93%	13%	0	45%
	調適後	93%	93%	67%	100%	88%
	調適次數	3	0	7	10	
P13(60)	調適前	27%	100%	7%	0	33%
	調適後	67%	80%	73%	67%	72%
	調適次數	6	3	7	6	
P14(60)	調適前	40%	93%	0	0	33%
	調適後	87%	87%	53%	80%	77%
	調適次數	5	3	8	11	
P15(60)	調適前	53%	100%	0	0	38%
	調適後	87%	87%	67%	47%	72%
	調適次數	7	4	10	11	
P16(60)	調適前	0	87%	13%	0	25%
	調適後	53%	60%	80%	47%	60%
	調適次數	9	7	12	11	
P17(60)	調適前	33%	93%	40%	0	42%
	調適後	73%	53%	73%	93%	73%
	調適次數	10	5	5	6	
P18(60)	調適前	0	100%	73%	7%	45%
	調適後	87%	100%	87%	80%	88%
	調適次數	6	1	8	8	
P19(60)	調適前	0	93%	0	0	23%
	調適後	80%	33%	87%	87%	72%
	調適次數	9	4	8	8	
P20(60)	調適前	13%	100%	0	0	28%
	調適後	93%	80%	60%	60%	73%
	調適次數	9	1	12	8	

表 C-2 MLBN 德國語料庫同語系 P1 ~ P20 調適前後之辨識結果

	調適前後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
P1(39)	調適前	71% (5/7)	100% (11/11)	71% (5/7)	43% (6/14)	69% (27/39)
	調適後	86% (6/7)	100% (11/11)	71% (5/7)	93% (13/14)	90% (35/39)
	調適次數	2	0	2	4	
P2(42)	調適前	100% (9/9)	90% (9/10)	64% (7/11)	83% (10/12)	83% (35/42)
	調適後	100% (9/9)	100% (10/10)	73% (8/11)	100% (12/12)	93% (39/42)
	調適次數	0	1	6	1	
P3(30)	調適前	50% (2/4)	100% (9/9)	0 (0/4)	92% (12/13)	77% (23/30)
	調適後	100% (4/4)	100% (9/9)	75% (3/4)	92% (12/13)	93% (28/30)
	調適次數	2	0	3	3	
P4(21)	調適前	100% (3/3)	100% (4/4)	75% (3/4)	50% (5/10)	71% (15/21)
	調適後	100% (3/3)	100% (4/4)	100% (4/4)	90% (9/10)	95% (20/21)
	調適次數	0	0	1	2	
P5(35)	調適前	86% (6/7)	100% (9/9)	75% (6/8)	55% (6/11)	77% (27/35)
	調適後	100% (7/7)	89% (8/9)	100% (8/8)	91% (10/11)	94% (33/35)
	調適次數	1	0	2	3	
P6(22)	調適前	75% (3/4)	100% (4/4)	0 (0/2)	50% (6/12)	59% (13/22)
	調適後	100% (4/4)	100% (4/4)	100% (2/2)	92% (11/12)	95% (21/22)
	調適次數	1	0	2	2	
P7(36)	調適前	100% (5/5)	100% (9/9)	100% (10/10)	92% (11/12)	97% (35/36)
	調適後	100% (5/5)	100% (9/9)	70% (7/10)	100% (12/12)	92% (33/36)
	調適次數	0	0	0	1	

	調適前後	悲傷	中性	快樂	生氣	辨識正確率
P8(41)	調適前	100% (10/10)	100% (7/7)	25% (2/8)	94% (15/16)	83% (34/41)
	調適後	100% (10/10)	100% (7/7)	88% (7/8)	88% (14/16)	93% (38/41)
	調適次數	0	0	5	2	
P9(34)	調適前	75% (3/4)	91% (10/11)	83% (5/6)	85% (11/13)	85% (29/34)
	調適後	100% (4/4)	100% (11/11)	67% (4/6)	92% (12/13)	91% (31/34)
	調適次數	1	1	1	2	
P10(39)	調適前	67% (6/9)	100% (5/5)	0 (0/11)	100% (14/14)	64% (25/39)
	調適後	100% (9/9)	100% (5/5)	100% (11/11)	71% (10/14)	90% (35/39)
	調適次數	2	0	8	3	



表 C-3 MLBN 不同語系 P1 ~ P20 調適前後之辨識結果(一半調適，一半測試)

	調適前後	悲傷(15)	中性(15)	快樂(15)	生氣(15)	辨識正確率 (60)
P1(60)	調適前	0	100%	20%	20%	35%
	調適後	87%	93%	60%	60%	75%
	調適次數	7	4	5	6	
P2(60)	調適前	7%	100%	73%	7%	47%
	調適後	67%	100%	67%	67%	75%
	調適次數	4	2	7	6	
P3(60)	調適前	20%	80%	53%	0	38%
	調適後	80%	53%	73%	73%	70%
	調適次數	9	6	7	7	
P4(60)	調適前	13%	87%	40%	0	35%
	調適後	87%	73%	93%	93%	87%
	調適次數	5	2	2	6	
P5(60)	調適前	0	100%	93%	0	48%
	調適後	93%	87%	100%	67%	87%
	調適次數	6	3	1	2	
P6(60)	調適前	7%	87%	47%	27%	42%
	調適後	87%	73%	80%	87%	82%
	調適次數	6	3	3	7	
P7(60)	調適前	0	100%	73%	7%	45%
	調適後	80.00%	73.33%	100%	87%	85%
	調適次數	6	3	5	5	
P8(60)	調適前	0	93%	40%	7%	35%
	調適後	80%	93%	93%	73%	85%
	調適次數	8	4	3	8	
P9(60)	調適前	7%	100%	67%	20%	48%
	調適後	87%	100%	93%	87%	92%
	調適次數	2	2	5	5	
P10(60)	調適前	7%	93%	53%	0	38%
	調適後	80%	87%	93%	100%	90%
	調適次數	4	1	5	5	

	調適前後	悲傷(15)	中性(15)	快樂(15)	生氣(15)	辨識正確率 (60)
P11(60)	調適前	87%	100%	7%	0	48%
	調適後	93%	93%	93%	93%	93%
	調適次數	0	1	6	7	
P12(60)	調適前	27%	100%	0	0	32%
	調適後	100%	93%	80%	100%	93%
	調適次數	2	0	3	3	
P13(60)	調適前	13%	100%	0	0	28%
	調適後	100%	80%	93%	87%	90%
	調適次數	6	2	4	4	
P14(60)	調適前	87%	93%	13%	0	48%
	調適後	100%	87%	80%	80%	87%
	調適次數	3	2	5	5	
P15(60)	調適前	53%	100%	0	0	38%
	調適後	73%	80%	80%	20%	63%
	調適次數	6	4	9	12	
P16(60)	調適前	0	93%	0	0	23%
	調適後	93%	93%	80%	40%	77%
	調適次數	7	4	8	11	
P17(60)	調適前	13%	100%	0	0	28%
	調適後	80%	80%	100%	100%	90%
	調適次數	6	3	4	2	
P18(60)	調適前	27%	87%	47%	0	40%
	調適後	93%	93%	67%	53%	77%
	調適次數	2	1	7	8	
P19(60)	調適前	7%	93%	0	7%	27%
	調適後	80%	80%	87%	73%	80%
	調適次數	8	3	8	8	
P20(60)	調適前	33%	93%	0	0	32%
	調適後	80%	87%	53%	67%	72%
	調適次數	5	3	7	7	

表 C-4 調適次數增加時，MLBN 不同語系 P1 ~ P20 調適前後之辨識結果

	調適前後	悲傷(30)	中性(30)	快樂(30)	生氣(30)	辨識正確率 (120)
P1(120)	調適前	0	100%	27%	27%	38.3%
	調適後	70%	80%	57%	80%	71.7%
	調適次數	11	5	10	14	
P2(120)	調適前	7%	100%	80%	10%	49.2%
	調適後	93%	97%	70%	83%	85.8%
	調適次數	7	1	11	11	
P3(120)	調適前	17%	87%	60%	7%	42.5%
	調適後	83%	80%	93%	87%	85.8%
	調適次數	11	13	9	8	
P4(120)	調適前	10%	90%	50%	3%	38.3%
	調適後	97%	93%	93%	100%	95.8%
	調適次數	7	4	3	8	
P5(120)	調適前	0	97%	97%	0	48.3%
	調適後	100%	100%	100%	87%	96.7%
	調適次數	5	2	3	6	
P6(120)	調適前	7%	93%	53%	27%	45.0%
	調適後	100%	83%	100%	80%	90.8%
	調適次數	9	5	6	10	
P7(120)	調適前	0	100%	80%	3%	45.8%
	調適後	70%	90%	93%	90%	85.8%
	調適次數	11	7	6	7	
P8(120)	調適前	0	97%	47%	10%	38.3%
	調適後	93%	97%	100%	73%	90.8%
	調適次數	9	4	3	11	
P9(120)	調適前	7%	97%	70%	23%	49.2%
	調適後	100%	100%	90%	90%	95.0%
	調適次數	5	2	9	10	
P10(120)	調適前	3%	97%	60%	0	40.0%
	調適後	83%	93%	100%	93%	92.5%
	調適次數	6	2	5	6	

	調適前後	悲傷(30)	中性(30)	快樂(30)	生氣(30)	辨識正確率 (120)
P11(120)	調適前	90%	100%	7%	0	49.2%
	調適後	100%	100%	100%	93%	98.3%
	調適次數	1	2	7	7	
P12(120)	調適前	37%	100%	0	0	34.2%
	調適後	100%	100%	100%	97%	99.2%
	調適次數	2	0	5	4	
P13(120)	調適前	13%	100%	7%	0	30.0%
	調適後	70%	100%	93%	100%	90.8%
	調適次數	7	6	8	7	
P14(120)	調適前	80%	93%	10%	0	45.8%
	調適後	97%	97%	93%	73%	90.0%
	調適次數	4	3	10	10	
P15(120)	調適前	47%	97%	0	0	35.8%
	調適後	97%	80%	77%	77%	82.5%
	調適次數	6	2	13	17	
P16(120)	調適前	3%	97%	0	0	25.0%
	調適後	97%	100%	87%	57%	85.0%
	調適次數	9	6	10	19	
P17(120)	調適前	13%	93%	0	0	26.7%
	調適後	80%	93%	100%	100%	93.3%
	調適次數	9	3	3	1	
P18(120)	調適前	30%	93%	50%	0	43.3%
	調適後	100%	97%	70%	60%	81.7%
	調適次數	4	1	15	16	
P19(120)	調適前	13%	93%	0	3%	27.5%
	調適後	97%	63%	87%	93%	85.0%
	調適次數	11	4	12	11	
P20(120)	調適前	37%	97%	3%	0	34.2%
	調適後	93%	93%	70%	70%	81.7%
	調適次數	9	4	15	12	

表 C-5 MLBNC 德國語料庫同語系 P1 ~ P20 調適前後之辨識結果

	調適前後	悲傷 (辨識個數/ 測試個數)	中性 (辨識個數/ 測試個數)	快樂 (辨識個數/ 測試個數)	生氣 (辨識個數/ 測試個數)	辨識正確率
P1(39)	調適前	43% (3/7)	100% (11/11)	57% (4/7)	79% (11/14)	74% (29/39)
	調適後	100% (7/7)	100% (11/11)	86% (6/7)	93% (13/14)	95% (37/39)
	調適次數	2	0	3	3	
P2(42)	調適前	67% (6/9)	100% (10/10)	73% (8/11)	92% (11/12)	83% (35/42)
	調適後	100% (9/9)	100% (10/10)	82% (9/11)	100% (12/12)	95% (40/42)
	調適次數	1	0	2	3	
P3(30)	調適前	25% (1/4)	100% (9/9)	25% (1/4)	100% (13/13)	80% (24/30)
	調適後	100% (4/4)	100% (9/9)	75% (3/4)	100% (13/13)	97% (29/30)
	調適次數	2	0	2	1	
P4(21)	調適前	67% (2/3)	100% (4/4)	75% (3/4)	80% (8/10)	81% (17/21)
	調適後	100% (3/3)	100% (4/4)	100% (4/4)	100% (10/10)	100% (21/21)
	調適次數	1	0	2	1	
P5(35)	調適前	100% (7/7)	100% (9/9)	50% (4/8)	91% (10/11)	86% (30/35)
	調適後	100% (7/7)	100% (9/9)	63% (5/8)	100% (11/11)	91% (32/35)
	調適次數	0	0	4	4	
P6(22)	調適前	25% (1/4)	100% (4/4)	50% (1/2)	100% (12/12)	82% (18/22)
	調適後	100% (4/4)	100% (4/4)	100% (2/2)	92% (11/12)	95% (21/22)
	調適次數	1	0	1	0	
P7(36)	調適前	100% (5/5)	100% (9/9)	100% (10/10)	83% (10/12)	94% (34/36)
	調適後	100% (5/5)	100% (9/9)	90% (9/10)	92% (11/12)	94% (34/36)
	調適次數	0	0	1	5	

	調適前後	悲傷	中性	快樂	生氣	辨識正確率
P8(41)	調適前	100% (10/10)	100% (7/7)	25% (2/8)	94% (15/16)	83% (34/41)
	調適後	100% (10/10)	100% (7/7)	50% (4/8)	100% (16/16)	90% (37/41)
	調適次數	0	0	5	5	
P9(34)	調適前	100% (4/4)	100% (11/11)	83% (5/6)	92% (12/13)	94% (32/34)
	調適後	100% (4/4)	100% (11/11)	100% (6/6)	100% (13/13)	100% (34/34)
	調適次數	0	0	1	1	
P10(39)	調適前	89% (8/9)	100% (5/5)	36% (4/11)	100% (14/14)	79% (31/39)
	調適後	100% (9/9)	100% (5/5)	82% (9/11)	93% (13/14)	92% (36/39)
	調適次數	1	0	6	2	



表 C-6 MLBNC 不同語系 P1 ~ P20 調適前後之辨識結果(一半調適，一半測試)

	調適前後	悲傷(15)	中性(15)	快樂(15)	生氣(15)	辨識正確率 (60)
P1(60)	調適前	13%	93%	13%	80%	50%
	調適後	80%	73%	40%	100%	73%
	調適次數	6	4	7	8	
P2(60)	調適前	0	100%	7%	100%	52%
	調適後	80%	100%	47%	93%	80%
	調適次數	4	5	7	5	
P3(60)	調適前	0	73%	13%	100%	47%
	調適後	80%	53%	73%	93%	75%
	調適次數	9	5	5	6	
P4(60)	調適前	13%	80%	47%	13%	38%
	調適後	87%	67%	67%	93%	78%
	調適次數	3	2	7	5	
P5(60)	調適前	0	80%	87%	27%	48%
	調適後	87%	87%	67%	100%	85%
	調適次數	6	5	4	2	
P6(60)	調適前	0	93%	7%	93%	48%
	調適後	93%	53%	87%	87%	80%
	調適次數	6	3	4	6	
P7(60)	調適前	0	93%	27%	93%	53%
	調適後	100%	67%	53%	100%	80%
	調適次數	6	5	8	7	
P8(60)	調適前	0	100%	27%	33%	40%
	調適後	73%	87%	73%	80%	78%
	調適次數	7	4	3	3	
P9(60)	調適前	0	100%	33%	93%	57%
	調適後	100%	100%	47%	100%	87%
	調適次數	4	3	5	3	
P10(60)	調適前	0	100%	60%	87%	62%
	調適後	80%	93%	60%	93%	82%
	調適次數	3	3	5	4	

	調適前後	悲傷(15)	中性(15)	快樂(15)	生氣(15)	辨識正確率 (60)
P11(60)	調適前	67%	100%	0	100%	67%
	調適後	87%	100%	67%	80%	83%
	調適次數	1	0	4	4	
P12(60)	調適前	53%	93%	33%	100%	70%
	調適後	93%	100%	93%	100%	97%
	調適次數	1	1	3	2	
P13(60)	調適前	20%	80%	100%	20%	55%
	調適後	100%	87%	93%	93%	93%
	調適次數	5	3	5	2	
P14(60)	調適前	53%	73%	7%	93%	57%
	調適後	100%	93%	27%	73%	73%
	調適次數	3	3	8	6	
P15(60)	調適前	47%	80%	73%	0	50%
	調適後	60%	80%	67%	73%	70%
	調適次數	5	7	4	10	
P16(60)	調適前	7%	87%	100%	0	48%
	調適後	93%	100%	67%	67%	82%
	調適次數	4	5	5	7	
P17(60)	調適前	27%	80%	93%	53%	63%
	調適後	100%	80%	60%	100%	85%
	調適次數	4	6	2	1	
P18(60)	調適前	20%	87%	73%	33%	53%
	調適後	100%	80%	53%	87%	80%
	調適次數	2	4	9	7	
P19(60)	調適前	13%	73%	0	100%	47%
	調適後	93%	73%	93%	87%	87%
	調適次數	5	4	4	3	
P20(60)	調適前	7%	73%	60%	47%	47%
	調適後	87%	87%	60%	67%	75%
	調適次數	2	2	5	5	

表 C-7 調適次數增加時，MLBN 不同語系 P1 ~ P20 調適前後之辨識結果

	調適前後	悲傷(30)	中性(30)	快樂(30)	生氣(30)	辨識正確率 (120)
P1(120)	調適前	7%	97%	7%	77%	46.7%
	調適後	93%	93%	57%	67%	77.5%
	調適次數	14	10	14	13	
P2(120)	調適前	0	100%	3%	100%	50.8%
	調適後	90%	100%	80%	97%	91.7%
	調適次數	7	5	11	12	
P3(120)	調適前	0	83%	10%	100%	48.3%
	調適後	80%	77%	87%	100%	85.8%
	調適次數	13	14	11	9	
P4(120)	調適前	13%	80%	60%	17%	42.5%
	調適後	87%	93%	97%	93%	92.5%
	調適次數	8	8	11	9	
P5(120)	調適前	3%	73%	90%	17%	45.8%
	調適後	97%	97%	97%	100%	97.5%
	調適次數	6	6	4	3	
P6(120)	調適前	0	97%	3%	87%	46.7%
	調適後	97%	100%	87%	90%	93.3%
	調適次數	6	5	7	5	
P7(120)	調適前	0	93%	20%	93%	51.7%
	調適後	90%	90%	57%	97%	83.3%
	調適次數	9	10	12	9	
P8(120)	調適前	0	100%	13%	23%	34.2%
	調適後	93%	77%	100%	97%	91.7%
	調適次數	13	10	5	9	
P9(120)	調適前	0	100%	37%	97%	58.3%
	調適後	100%	93%	80%	93%	91.7%
	調適次數	5	4	13	8	
P10(120)	調適前	0	100%	43%	83%	56.7%
	調適後	90%	93%	83%	97%	90.8%
	調適次數	5	4	15	7	

	調適前後	悲傷(30)	中性(30)	快樂(30)	生氣(30)	辨識正確率 (120)
P11(120)	調適前	73%	100%	0	100%	68.3%
	調適後	100%	100%	73%	97%	92.5%
	調適次數	1	0	9	7	
P12(120)	調適前	60%	90%	33%	100%	70.8%
	調適後	93%	100%	100%	100%	98.3%
	調適次數	2	2	5	1	
P13(120)	調適前	30%	90%	97%	23%	60.0%
	調適後	97%	100%	90%	100%	96.7%
	調適次數	5	6	7	4	
P14(120)	調適前	50%	80%	3%	90%	55.8%
	調適後	100%	100%	60%	87%	86.7%
	調適次數	3	7	14	16	
P15(120)	調適前	40%	80%	70%	0	47.5%
	調適後	100%	87%	90%	90%	91.7%
	調適次數	9	9	11	16	
P16(120)	調適前	7%	90%	93%	3%	48.3%
	調適後	100%	90%	80%	80%	87.5%
	調適次數	4	8	15	3	
P17(120)	調適前	23%	67%	93%	57%	60.0%
	調適後	83%	93%	97%	100%	93.3%
	調適次數	10	6	4	1	
P18(120)	調適前	23%	93%	60%	37%	53.3%
	調適後	97%	100%	40%	90%	81.7%
	調適次數	2	8	18	12	
P19(120)	調適前	17%	70%	0	100%	46.7%
	調適後	90%	100%	67%	97%	88.3%
	調適次數	8	10	5	7	
P20(120)	調適前	13%	80%	57%	47%	49.2%
	調適後	90%	80%	87%	90%	86.7%
	調適次數	7	13	15	12	