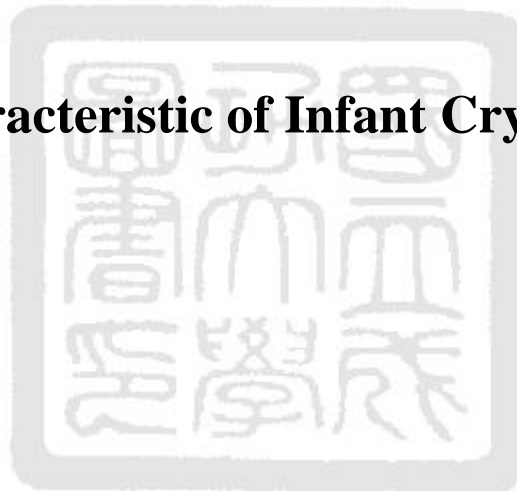


國立成功大學  
工程科學研究所  
碩士論文

新生嬰兒哭聲情緒之辨識

**Acoustic Characteristic of Infant Cry Vocalizations**



研 究 生：游祿勳

指導教授：周榮華

中華民國九十六年七月

國立成功大學

碩士論文

新生嬰兒哭聲情緒之辨識

研究生：游祿勳

本論文業經審查及口試合格特此證明

論文考試委員

王明翠 周學華

指導教授：周學華

系(所)主管：黃悅民

中華民國 96 年 7 月 6 日

## 中文摘要

新生的嬰兒由於其發聲器官及心智不夠成熟，無法利用語言及肢體動作表達自己本身的生理需求及情緒，因此哭聲為其唯一與外界表達溝通之途徑。在早期的研究中，可以發現嬰兒的哭聲為一個強而有力的表達工具，可以藉由嬰兒哭聲判別情緒、生理疾病和生理需求。

本文提出了一個嬰兒情緒哭聲辨識系統，分辨五種的嬰兒情緒哭聲，即：A：秤重（Scale） B：洗澡(Bath) C：飢餓(Hunger) D：打針（Injection）和 E：酒精（Alcohol）。

本研究探討利用不同的語音特徵參數，在五種嬰兒哭聲的辨識情形，並使用類神經網路和 K-最近鄰居法則為辨識機制。其所使用之特徵為：倒頻譜(Cepstrum)，梅爾倒頻譜(MFCC)，線性預估參數(LPC) 和線性預估導出之倒頻譜參數(LPCC)。

實驗中，藉由改變聲學特徵向量的維數，可發現使用適合的維數，將會提高辨識率，另外，實驗分析，LPCC 為最佳辨識特徵，LPCC 不同特徵維度中，以 LPCC96 為最佳 LPCC 維度，平均辨識率約為 83.3%。

關鍵詞：倒頻譜、梅爾倒頻譜、線性預估參數、線性預估導出之倒頻譜參數、嬰兒哭聲、辨識

## **Abstract**

An infant's crying is the only communication method in the first month of life. It can be used to describe the needs of the babies or for making medical diagnoses of pathologies at very early stages of life.

This work presents the development of an automatic recognition system of infant cry, with the objective to classify five types of crying due to: scale ,bath , hunger ,injection ,and alcohol .

Four kinds of acoustic characteristics are used to classify five types of infant cry. Both artificial neural networks (ANN) and K-nearest neighbor rule(KNNR) are examined as potential classifiers.

The acoustic characteristics used to distinguish the infant cries are Cepstrum Coefficients (Cepstrum) , Mel-Frequency Cepstrum Coefficients(MFCC) , Linear Predict Coefficients ( LPC) , LPC derived cepstrum coefficients (LPCC).

Different numbers of acoustic characteristic coefficients are used to classify infant cry in the experiments. Better classification results can be obtained with a proper number of coefficients.

The best acoustic characteristic is LPCC, and the best overall accuracy obtained with LPCC96 is about 83% .

**Keywords:** cepstrum, MFCC, LPC,LPCC,Infant Cry ,Recognition

## 誌謝

歲月如梭，時光飛逝，在成大讀書的這六年，轉眼間也就結束，在最後的兩年研究生生活，經歷了許多的歷練和學習，對我的人生有很大的影響。

在學生生活中，首先要誠摯的感謝指導教授 周榮華 老師這六年的諄諄的教誨和指導，讓我學業上能自由的發揮，使我受益良多。在此，要感謝構裝實驗室的小綿羊學長、柳金學姐、小馨學長和佩璇學姐的照顧和指導，尤其要特別感激柳金學姐，柳金學姐教導了我許多與人相處之道，也開闊了我的眼界，受益匪淺。也感激實驗室的同學和學弟妹，尤其是小天學妹，我們一起打拼、扶持和吃喝玩樂的快樂時光，會是我生命中最美好的回憶，也感激系統實驗室的全部人員，謝謝你們帶給我許多的歡樂和笑聲。

要感激的人實在是太多了，在我求學的過程中，感激有大家的幫助，讓我能夠順利的完成學業，謝謝大家。

## 目錄

中文摘要.....	I
英文摘要.....	II
致謝.....	III
目錄.....	IV
表目錄.....	VII
圖目錄.....	IX
符號說明.....	XII
第一章 緒論.....	1
1.1 簡介.....	1
1.2 文獻回顧.....	2
1.2.1 心智發展成熟度和生理疾病判斷.....	2
1.2.2 情緒及生理需求辨識.....	5
1.3 研究動機與目的.....	9
1.4 章節概要.....	10
第二章 嬰兒哭泣情緒模型之建立.....	11
2.1 嬰兒哭泣情緒辨識系統.....	11

2.2 語音信號前處理.....	13
2.2.1 數位取樣.....	13
2.2.2 邊緣端點偵測.....	14
2.2.3 切割音框.....	15
2.2.4 預強調.....	16
2.2.5 視窗化.....	16
2.3 小結.....	17
第三章 嬰兒哭泣情緒特徵參數的擷取.....	18
3.1 倒頻譜.....	19
3.2 梅爾倒頻譜係數.....	21
3.3 線性預估編碼.....	24
3.4 線性預估模型所導出的倒頻譜係數.....	27
3.5 基頻.....	28
3.6 降低特徵維度-主要分量分析.....	29
第四章 辨識機制.....	33
4.1 類神經網路簡述.....	33
4.2 倒傳遞類神經網路.....	36
4.3 K-最近鄰居分類法.....	38
第五章 嬰兒哭泣情緒辨識實驗與結果.....	40

5.1 嬰兒哭聲情緒資料庫介紹.....	40
5.2 實驗介紹.....	42
5.2.1 實驗一.....	43
5.2.2 實驗二.....	44
5.2.3 實驗三.....	47
5.2.4 實驗四.....	50
5.2.5 實驗五.....	52
5.2.6 實驗六.....	63
5.2.7 實驗七.....	64
5.3 綜合分析與討論.....	66
第六章 結論與未來展望.....	68
6.1 結論.....	68
6.2 未來展望.....	69
參考文獻.....	70



## 表目錄

表 1.1 不同 ANN 類型之辨識率 (特徵參數為 Mel-Cepstrum Coefficients)	
[16] .....	7
表 1.2 不同 ANN 類型之辨識率(特徵參數為 Mel Filter-Energy Coefficients)	
[16] .....	8
表 1.3 不同 PCA 維度之辨識率(特徵參數為 MFCC 辨識機制為 ANN)	
[17].....	8
表 1.4 不同 PCA 維度之辨識率(特徵參數為 MFCC 辨識機制為 SVM)	
[17].....	8
表 1.5 不同 Module 之辨識率[18].....	8
表 5.1 嬰兒情緒哭聲資料庫.....	41
表 5.2 哭聲情緒辨識系統訓練與測試數量.....	42
表 5.3 LPCC96 不同降維方式辨識率比較.....	43
表 5.4 不同維度的 CEP 辨識率比較(ANN).....	45
表5.5不同維度的CEP辨識率比較(KNNR).....	46
表5.6不同維度的MFCC辨識率比較(ANN).....	48
表5.7不同維度的MFCC辨識率比較(KNNR).....	48
表 5.8 不同維度的 LPC 辨識率比較(ANN).....	50
表 5.9 不同維度的 LPC 辨識率比較(KNNR).....	51

表 5.10 不同維度的 LPCC 辨識率比較(ANN).....	53
表 5.11 不同維度的 LPCC 辨識率比較(KNNR).....	54
表 5.12 四種不同特徵最佳維度辨識率比較(ANN).....	55
表 5.13 四種不同特徵最佳維度辨識率比較(KNNR).....	56
表 5.14 LPCC96 各類別辨識音檔數目比較(個) (ANN).....	58
表 5.15 LPCC96 各類別辨識音檔數目比較(%) (ANN).....	59
表 5.16 LPCC96 各類別辨識音檔數目比較(個) (KNNR).....	59
表 5.17 LPCC96 各類別辨識數目比較(%) (KNNR).....	60
表 5.18 MFCC72 各類別辨識音檔數目比較(個) (ANN).....	60
表 5.19 MFCC72 各類別辨識數目比較(%) (ANN).....	61
表 5.20 MFCC72 各類別辨識音檔數目比較(個) (KNNR).....	61
表 5.21 MFCC72 各類別辨識數目比較(%) (KNNR).....	62
表 5.22 LPCC96 組合特徵辨識率比較(ANN) .....	63
表 5.23 MFCC72 組合特徵辨識率比較(ANN) .....	65

## 圖目錄

圖 2.1 嬰兒哭聲情緒辨識系統架構之流程.....	12
圖 2.2 嬰兒哭聲情緒辨識系統架構.....	13
圖 2.3 嬰兒哭聲情緒訊號前處理.....	13
圖 3.1 發音模型的系統圖[19].....	19
圖 3.2 利用倒頻譜將激發源訊和聲道脈衝訊號分離[19].....	21
圖 3.3 倒頻譜擷取流程圖.....	21
圖 3.4 梅爾刻度之帶通濾波器.....	23
圖 3.5 梅爾倒頻譜擷取流程圖.....	24
圖 3.6 線性預估編碼流程圖.....	26
圖 3.7 LPCC 流程圖.....	27
圖 3.8 Pitch 流程圖.....	28
圖 3.9 特徵參數 PCA 降維流程圖.....	30
圖 3.10 A-10-4032 cep48 全部音框疊合正規化平均值（不經 PCA 處理）.....	31
圖 3.11 A-10-4032 cep48 PCA1 及 PCA2(未經正規化處理).....	31
圖 3.12 A-10-4032 cep48 PCA1+PCA2(未經正規化處理）.....	31

圖 3.13 A-10-4032 cep48 PCA1+PCA2 (經正規化處理).....	32
圖 4.1 Sigmoid transfer function.....	37
圖 4.2 Linear Transfer Function.....	37
圖 4.3 Log-Sigmoid Transfer Function.....	37
圖 5.1 LPCC96 不同降維方式辨識率比較.....	44
圖 5.2 不同維度的 CEP 辨識率比較(ANN).....	45
圖 5.3 不同維度的 CEP 辨識率比較(ANN).....	46
圖 5.4 不同維度的 MFCC 辨識率比較(ANN).....	48
圖 5.5 不同維度的 MFCC 辨識率比較(KNNR).....	49
圖 5.6 不同維度的 LPC 辨識率比較(ANN).....	51
圖 5.7 不同維度的 LPC 辨識率比較(KNNR).....	51
圖 5.8 不同維度的 LPCC 辨識率比較(ANN).....	53
圖 5.9 不同維度的 LPCC 辨識率比較(KNNR).....	54
圖 5.10 四種不同特徵最佳維度辨識率比較(ANN).....	56
圖 5.11 四種不同特徵最佳維度辨識率比較(KNNR).....	56
圖 5.12 LPCC96 各類別辨識音檔數目比較(%) (ANN).....	59
圖 5.13 LPCC96 各類別辨識數目比較(%) (KNNR).....	60
圖 5.14 MFCC72 各類別辨識數目比較(%) (ANN).....	61
圖 5.15 MFCC72 各類別辨識數目比較(%) (KNNR).....	62

圖 5.16 LPCC96 組合特徵辨識率比較(ANN) .....63

圖 5.17 MFCC72 組合特徵辨識率比較(ANN).....65



## 符號說明

$E(n)$  : 能量

$W(n)$  : 視窗函數

$E_p(Z)$  : 激發源訊號(頻域)

$\theta(Z)$  : 發音的模型(頻域)

$H(Z)$  : 發音腔道模型(頻域)

$G(Z)$  : 聲道脈波模型(頻域)

$R(Z)$  : 嘴唇端的輻射模型(頻域)

$s(n)$  : 語音訊號

$\hat{s}(n)$  : 預估的語音訊號

$B_m(k)$  : 第  $m$  個頻帶的三角濾波器

$f_m$  : 第  $m$  個頻帶之中心頻率

$C_m(t)$  : 第幾個音框的梅爾倒頻譜係數

$\Delta C_m(t)$  : 第幾個音框的一次差量梅爾倒頻譜係數

$\eta$  : 學習速率

$W$  : 權重值

# 第一章

---

## 緒論

### 1.1 簡介 (Introduction)

生命的創造是世界上最美妙的事情，兒女是父母愛情的結晶，是家庭快樂的泉源，迎接新生命是令人喜悅，新生嬰兒就好像初生的嫩芽，需要父母細心的灌溉、撫育，才能平安、健康的成長。然而嬰幼兒的早期的身心發展奠定了未來的身心發展的基礎，對未來的發展影響甚巨，因此如何妥善的照顧新生兒，就成為大眾關注的課題。

隨著時代的進步，社會結構的變遷，從農業時代的三代同堂家庭，轉變為現今核心小家庭，父母大都皆需出外工作維持生活所需，繁忙的工作剝奪了與嬰兒相處的時間，使得父母無法精確的掌控辨讀嬰兒的情緒和生理狀況，於是希望能藉由機器輔助判別嬰兒的情緒及身體狀況，幫助父母妥善的照顧新生兒。

新生的嬰兒，由於其發聲器官及心智不夠成熟，無法利用語言及肢體動作，表達自己本身的生理需求及情緒，因此哭聲為其唯一與外界表達溝通之途徑，在新生兒出生後的第一個月，新生兒只能發出哭泣

聲，無法發出其他的聲音，隨著新生兒的成長，發聲器官的成熟，慢慢的會綻放笑容，發出模糊不清的話語；在第八個月，嬰兒能夠重複某些音節，例如爸爸媽媽，此時，嬰兒已能夠簡略的使用語言和表情表達情緒。

在這短暫的一至八個月中，新生嬰兒只能利用哭聲表達溝通自身的情緒和感覺，此時，需要使用嬰兒情緒辨識系統，完整的辨識哭聲所代表的情緒，而一個情緒辨識系統，最重要的是特徵參數的擷取，如何選取一個具有鑑別性的特徵參數，則是此次研究最主要的目標。

## 1.2 文獻回顧 (Literature Review)

在早期的關於嬰兒哭聲研究中，可以發現嬰兒的哭聲為一個強而有力的表達工具，可以藉由嬰兒哭聲判別情緒、生理疾病、身心發展成熟度和生理需求，相關的文獻研究相當豐富，在此將文獻資料依新生嬰兒哭聲原因分成下述幾項：(1) 心智發展成熟度和生理疾病判斷；(2) 情緒及生理需求辨識。

### 1.2.1 心智發展成熟度和生理疾病判斷

自從 1960 年代小兒科醫生發現利用新生嬰兒的哭泣聲，可以有效的評估鑑別新生嬰兒的腦部發育情形和生理疾病之後，許多的研究者開始



這方面的相關研究，希望能夠盡早發現生理疾病，及時予以治療。

在台灣，每年約八百分之一到千分之一新生兒，在出生時有嚴重和重度的聽力喪失，另外有千分之三的新生兒有程度略輕微的聽力喪失，包括輕至中度的雙側性或單側性聽力喪失，然而國內新生兒接受聽力篩檢的比率僅 5%~10%，導致許多聽障兒錯失最佳的治療時機。新生兒一歲以前是為語言發育的關鍵時期，在這時期，新生兒接受不同的語言刺激，學習表達語言，因此聽力損害的診斷必須盡早，才能使聽力喪失的嬰兒，通過訓練而擁有完整語言的發育，一歲以前為最佳的治癒時期。

新生兒哭泣的聲學特徵 (Acoustic Feature) 常用以分析辨識新生兒聽力障礙，最常被擷取的聲學特徵包含：基頻的輪廓 (Fundamental Frequency Curves)、基頻的平均值、標準差和變動範圍、第一共振峰 (F1)、第二共振峰 (F2)、第三共振峰 (F3) 的輪廓和平均值及標準差、哭泣的持續時間、哭泣的間隔時間和新生兒的哭泣的旋律 (Melody) 等。

但是基頻和共振峰擷取容易受到外界雜訊的干擾和影響，因 Petroni 和 Malowany 等人[1][2][3]，對共振峰擷取方式，提出具有強健性的擷取方法，並觀察共振峰的分佈情形。於 Baeck 和 Souza[4]與 Varallyay 等人[5]的文章中，對第一共振峰 (F1)、第二共振峰 (F2)、第三共振峰 (F3) 的輪廓和平均值及標準差提出新的擷取方式，歸納統計出共振峰的輪廓曲線、分佈範圍頻率的差異。

Fort 和 Manfredi [6][7]的文章中，對正常的哭聲加入不同的噪音參數，針對噪音對基頻和共振峰擷取的影響，提出對噪音具有強健性的擷取方式。Harada 等人[8]、Perez 等人[9]以及 Wermke 等人[10]的文章中，提出新生兒隨著腦部神經細胞的發展，聽力器官的成熟，開始接受外界聲音的刺激，逐漸發展出複雜的哭泣旋律，因此新生兒的哭泣旋律的難易複雜度代表著新生嬰兒腦部神經細胞的發展連結程度，同時也可以偵測新生兒聽力是否正常。

Garcia 和 Reyes Garcia [11][12]針對聽力正常嬰兒和聽力受創的嬰兒，提出利用MFCC(梅爾倒頻譜係數，Mel frequency Cepstrum coefficients)和LPC(線性預估編碼係數，linear prediction coefficients)這兩種不同的語音特徵，使用類神經網路(Neural Network)辨識機制，辨識其生理狀況，在文中藉由改變類神經網路的學習收斂機制與係數，達到不同的類神經網路學習訓練方式，比較的學習收斂機制有Backpropagation、Gradient Descent Backpropagation、Gradient Descent with Adaptive Learning Rate Backpropagation 和Scaled Conjugate Gradient Method等這六種機制，其中以Scaled Conjugate Gradient Method (SCG)的學習方式最好，其辨識率可以高達97.43%，此外使用同樣的類神經網路 (SCG) 為其辨識機制，LPC為其辨識特徵，所得之辨識率為94.3%，比選用MFCC的語音特徵的辨識率 (97.43%) 低。

Lederman 等人[13]在其文中，講述了何謂 Respiratory Distress Syndrome RDS(呼吸窘迫症候群)及 RDS 對新生兒所造成的死亡威脅，提出如能及早發現治療，則可以大幅降低新生兒的死亡率，因此文中提出利用 Mel-Cepstrum Coefficients 為主要的辨識特徵，搭配五種狀態，三個混合模型的隱藏式馬可夫模型 HMM 和混合式高斯模型 GMM 為其辨識機制，探討其辨識結果之優劣。

### 1.2.2 情緒及生理需求辨識

頻譜的特徵是最普遍被選用的特徵參數，被選用的頻譜特徵參數有線性預估參數 LPC、梅爾倒頻譜參數 MFCC、梅爾能量參數 Mel Filter-Band Energy Coefficients、倒頻譜係數以及頻率的能量參數等。

Petroni 等人[14][15][16]的論文中，對於使用不同類神經網路模型為辨識機制有著深入的研究，文中主要使用兩種頻譜特徵 MFCCs 和 Mel Filter-Band Energy Coefficients 為分析的特徵，在類神經網路模型上，將模型分成五種，分別為 FF ( feed-forward neural network with full connections )、FT ( feed-forward neural network with tessellated connections )、RNN (recurrent neural network )、TDNN (time-delay neural network) 和 CC(cascade correlation network)，前兩種類神經網路 FF 和 FT 並不包含著時間的資訊，RNN 和 TDNN 則為跟時間有密切相關的類神經

網路。

使用這五種不同類神經網路的訓練模型，對下列這三種的語音情緒：Anger、Fear 和 Pain 予以辨識，比較其辨識結果之優劣的，可發現利用 FF 和 FT 為辨識機制，其辨識率高於使用 RNN 和 TDNN 的辨識率，此結果顯示包含時間資訊的類神經網路 RNN 和 TDNN 並沒有增進辨識率，反而因為包含了時間資訊，使得其結構受到限制，導致無法訓練出複雜合適的網路結構，使錯誤率提高。

在論文的後段，結合 MFCCs 、Mel Filter-Band Energy Coefficients 為其分析的特徵和五種的網路模型，分析其辨識結果，發現使用 MFCC 為特徵參數，搭配 FF 為其辨識網路，可求得情緒哭泣辨識之最佳結果，使用此組合，其辨識率可高達 83.59% ，其辨識率如表 1.1 和表 1.2 所示。

由於在錄製嬰兒情緒哭聲時，無法精準的控制錄音時間，導致錄音的資料量太龐大，使得所需計算時間太長，Barajas-Montiel和 Reyes Garcia[17]就提出利用PCA（主要分量分析法 Principle Component Analysis）將MFCC的特徵維度降低，只擷取最重要的MFCC係數，可以節省計算時間和儲存空間，文章中將MFCC係數從304個減少至2-16個，並且使用兩種不同的辨識機制:Boosting Ensemble of Artificial Neural Networks and Boosting Ensemble of Support Vector Machines，用以分辨兩種嬰兒哭聲: Pain-Nonpain 和Hungry-Nohungry，結果顯示Artificial Neural

Networks 的辨識率高於 Support Vector Machines，且用以辨識 Pain-Nonpain 的辨識率較佳，其辨識率如表1.3和表1.4所示。

目前在情緒辨識研究方面，架構上可分成單一模組和雙模組辨識兩大類型，單一模組主要是採取語音特徵或臉部表情特徵，而雙模組是將臉部和語音特徵互相結合辨識，然而嬰兒哭聲辨識發展至今，在架構上大都為採取語音特徵的單一模組，但Pal 等人[18]在2006年發表雙模組辨識嬰兒情緒語音，臉部的特徵主要擷取嘴巴的開合、眼睛的睜閉、眉毛的位置，哭泣的情緒特徵則使用基頻、和第一、二、三共振峰，所辨識之情緒有 anger、pain、sadness、hunger、fear，單一使用臉部表情模組的辨識率平均為64%，使用哭泣情緒模組的辨識率平均為74.2%，結合兩者的雙模組平均辨識率有75.2%，相較於哭泣情緒單模組有小幅的增進，增進幅度有1%，五個所辨識的哭聲情緒中，以pain情緒辨識率最高，為93%，Anger情緒辨識率最低，僅62%，辨識率結果如表1.5。

表 1.1 不同 ANN 類型之辨識率

(特徵參數為 Mel-Cepstrum Coefficients [16])

不同 ANN 類型之辨識率 (特徵參數為 Mel-Cepstrum Coefficients)					
ANN	FF	FT	RNN	TDNN	CC
Anger	79.5%	62.9%	53.7%	64.8%	44.4%
Fear	43.7%	12.5%	12.5%	0.0%	0.0%
Pain	92.0%	84.0%	84.8%	70.4%	50.4%
Total	83.5%	73.3%	70.2%	63.0%	40.0%

表 1.2 不同 ANN 類型之辨識率

(特徵參數為 Mel Filter-Energy Coefficients) [16]

不同 ANN 類型之辨識率 (特徵參數為 Mel Filter-Energy Coefficients)					
ANN	FF	FT	RNN	TDNN	CC
Anger	79.5%	62.9%	54.7%	18.5%	11.1%
Fear	12.5%	0.0%	0.0%	0.0%	0.0%
Pain	87.2%	85.6%	84.8%	88.0%	80.8%
Total	78.9%	73.3%	70.2%	61.5%	59.4%

表 1.3 不同 PCA 維度之辨識率

(特徵參數為 MFCC 辨識機制為 ANN) [17]

不同 PCA 維度之辨識率 (特徵參數為 MFCC 辨識機制為 ANN)				
PCA 維度	PCA2	PCA3	PCA10	PCA16
Pain-Nonpain	88.2%	92.1%	94.0%	96.4%
Hungry-Nohungry	67.7%	70.0%	79.9%	87.6%

表 1.4 不同 PCA 維度之辨識率

(特徵參數為 MFCC 辨識機制為 SVM) [17]

不同 PCA 維度之辨識率 (特徵參數為 MFCC 辨識機制為 SVM)				
PCA 維度	PCA2	PCA3	PCA10	PCA16
Pain-Nonpain	85.2%	85.4%	83.9%	85.1%
Hungry-Nohungry	51.3%	53.0%	74.5%	76.4%

表 1.5 不同 Module 之辨識率[18]

不同 Module 之辨識率			
	Image Module	Sound Module	Decision Fusion
Pain	76.0%	91.0%	93.0%
Hunger	56.0%	72.0%	73.0%
Fear	71.0%	71.0%	72.0%
sadness	62.0%	79.0%	76.0%
Anger	53.0%	58.0%	62.0%
Total	64.0%	74.2%	75.2%

### 1.3 研究動機與目的

本研究利用新生嬰兒的哭泣聲，辨識新生嬰兒的情緒和生理需求，嘗試瞭解新生嬰兒為何哭泣，有效找出哭泣之原因。讓電腦正確的辨識新生嬰兒的情緒和生理需求，最重要的因素就是特徵參數的擷取。在本論文中，探討不同的語音特徵參數，在五種不同情緒的嬰兒哭聲的辨識情形。此外，藉由改變不同的頻譜特徵參數的擷取數目，討論辨識率的改變情形，希望能夠找出一組較適當的特徵參數，以代表五種嬰兒情緒哭聲的特性，且不易受環境干擾並能具有高鑑別性。

在嬰兒哭泣情緒辨識方面，頻譜參數主要是採用MFCC和LPC參數，聲學參數則多用基頻和共振峰，本論文主要比較LPC、LPC所導出的倒頻譜參數（LPC derived cepstrum coefficients-LPCC）、倒頻譜和MFCC四種不同的頻譜特徵參數在嬰兒情緒哭聲辨識上的影響。

在聲學特徵方面，則選取了基頻和能量，並討論這些參數是否對嬰兒情緒哭聲辨識系統有所幫助。

選取了最重要的特徵參數後，開始進行嬰兒情緒哭泣的辨識。本論文主要使用的是 ANN（Artificial Neural Network）類神經網路為辨識的機制，但因資料量太大，所以在擷取特徵參數上，採用了 PCA 降低資料

維度。

然而，一個具有鑑別性的特徵參數，應適用於多種的辨識機制，因此本研究中亦使用 KNNR (K-nearest neighbor rule) 演算法做實驗，輔助 ANN 模型探討不同特徵參數的優劣。

## 1.4 章節概要

本論文章節概述如下：

第一章：『緒論』：說明本論文的研究動機、目的、前人所做之相關研究和本論文的研究方向。第二章：『嬰兒哭泣情緒模型之建立』：介紹整個哭泣辨識系統的架構和流程，對特徵分析中前置處理的部分，作詳細的敘述。第三章：『嬰兒哭泣情緒特徵參數的擷取』：介紹基本的特徵參數擷取方式和整個流程，主要使用的是 MFCC、LPC、LPCC、CEP 等語音特徵。第四章：『辨識機制』：介紹辨識機制的演算法，文中主要介紹的是 ANN 和 KNNR。第五章：『嬰兒哭泣情緒辨識實驗與結果』：針對不同的特徵參數擷取方法和兩種不同的分類法作嬰兒哭泣情緒辨識實驗，並比較其結果。第六章：『結論與建議』：簡單歸納各辨識特徵的特性及優劣，並提出研究之建議。



## 第二章

---

### 嬰兒哭泣情緒模型之建立

#### 2.1 嬰兒哭泣情緒辨識系統

本文的辨識系統架構之流程，如圖 2.1 所示。

整個情緒辨識系統，主要可分為三部份：

第一部份：訊號的前處理和特徵參數之擷取。

第二部份：輸入特徵參數，訓練辨識模型。

第三部份：輸入測試訊號於辨識模型，得知辨識結果。

以下分別說明之：

##### (1) 訊號的前處理和特徵參數之擷取

嬰兒哭聲訊號，經由數位錄音裝置，將連續語音信號，以特定編碼率（bit rate）與取樣率（sample rate）轉換為離散時間序列。在進行特徵擷取時，必須先進行訊號的前處理，以過濾雜訊和有效的擷取所需之資訊，訊號的前處理主要包含：數位取樣、邊緣端點偵測、切割音框、預強調和視窗化，訊號經過前處理之後，再對處理過後的語音訊號擷取特

徵參數，特徵參數的萃取是為了能找出代表語音資料特性的特徵參數，取代龐大的語音訊號。

## (2) 輸入特徵參數，訓練辨識模型

在訓練階段，由於無法將所有之處理後的訓練特徵參數直接加以比對，因此必須建立情緒辨識模型，將所建立之情緒辨識模型，作為測試時的樣本模型。

## (3) 輸入測試訊號於辨識模型，得知辨識結果

在辨識階段，輸入待測之語音，經過相同的前處理步驟，擷取所需之特徵參數，將其特徵參數與之前所訓練之情緒辨識模型予以樣本比對，找尋相似度最高的情緒類型，得知其辨識結果。

本論文的整個辨識系統架構如圖 2.2 所示，架構中的每個階段和步驟，將會在之後的章節有更詳盡的解說。

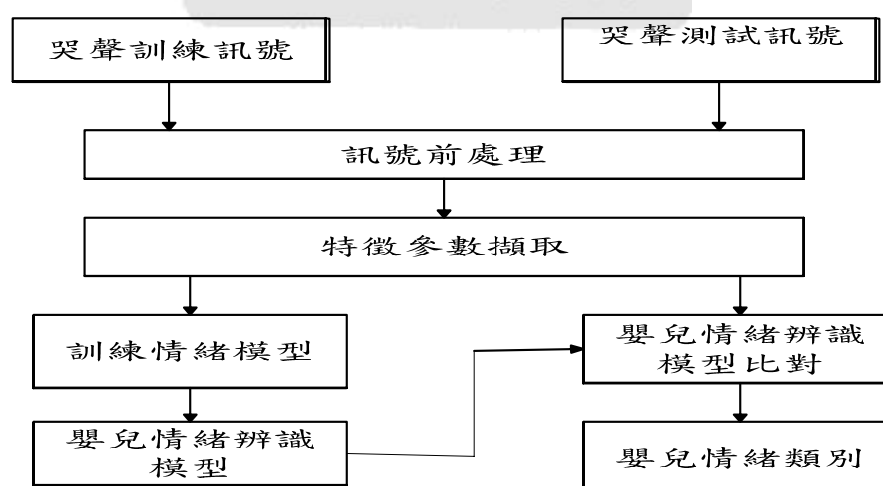


圖 2.1 嬰兒哭聲情緒辨識系統架構之流程

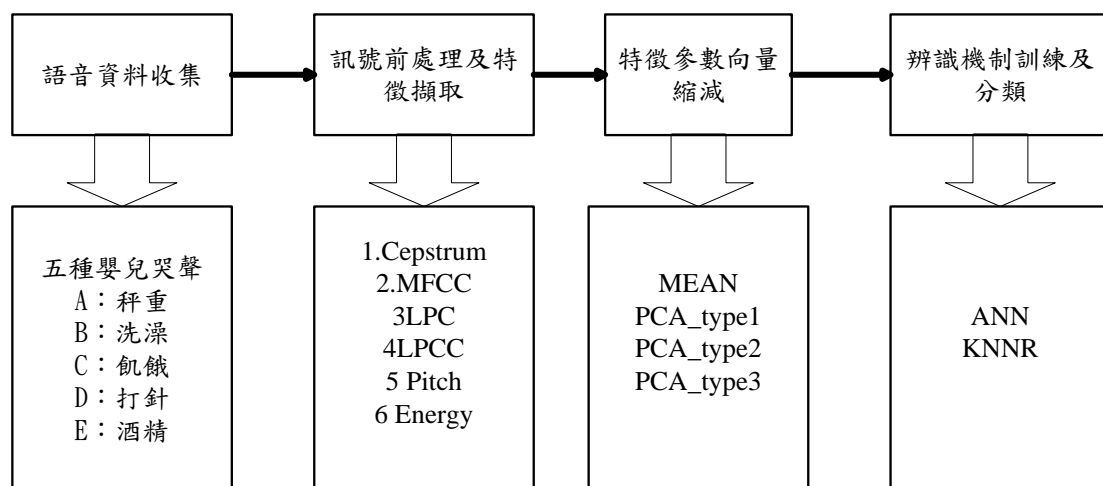


圖 2.2 嬰兒哭聲情緒辨識系統架構

## 2.2 語音信號前處理

對所蒐集的嬰兒哭聲訊號求取特徵參數之前，首先將音頻訊號作前處理，在前處理的過程中，依序分為數位取樣、邊緣端點偵測、切割音框、預強調和視窗化等階段，流程如圖 2.3 所示。

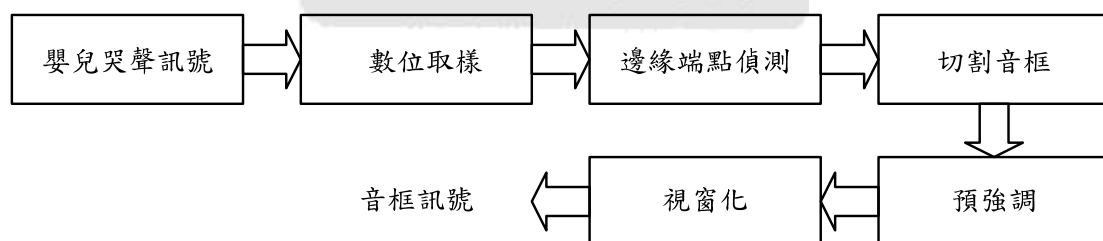


圖 2.3 嬰兒哭聲情緒訊號前處理

### 2.2.1 數位取樣

聲帶振動所產生之氣流，在發音腔道內共振，產生聲波，這就是聲

音的產生。聲音是一種波形，隨著時間作連續的變化，是一種連續的類比訊號，然而，眾所皆知，電腦無法處理類比訊號，因此必須將類比訊號取樣，轉換為數位訊號，這就是數位取樣。

在研究中，嬰兒哭聲訊號經由數位 MD（MiniDisc）錄音裝置，將連續語音信號轉換為取樣率為 44KHZ、單聲道及取樣解析度為八位元的 WAV 訊號。

### 2.2.2 邊緣端點偵測

當錄製一段嬰兒哭聲時，由於嬰兒哭聲的起始點和結束點，不易事先預測和控制，因此一段嬰兒哭聲前後常包含著靜音。

靜音部份可能包含著錄音環境中的背景雜訊或錄音設備的噪音，這些並非所需之哭泣語音資訊，所以靜音部份對情緒的辨識並無幫助，反而會因處理資料太多，而降低處理速度，因此需偵測出嬰兒哭聲訊號的邊緣端點，將前後無用之靜音切除，得到有效的哭聲範圍。然而必須正確的偵測出嬰兒哭聲訊號的邊緣端點，避免有效的哭聲被切除，進而影響辨識結果。

在本文中，使用的是音框能量（energy）偵測邊緣端點，在語音特徵中，聲音能量的變化與波形振幅有關，振幅越大，能量越強，將音框能量值定義為一個音框長度內訊號值的平方總和，將能量值以對數值表示，因為在聽覺上人耳對聲音的強度感知，並非線性的。其公式如下：

$$E(n) = \log\left[\sum_{m=0}^{N-1} (|x(n+m)|)^2\right]$$

上式中，N 為此音框語音樣本的取樣點個數，E (n) 為能量值，x(n) 為取樣之訊號。

音框能量越小，雜訊的可能性就越大，因此能量偵測會根據音框能量的大小是否超過所設的門檻值，判斷是否為靜音雜訊部分。若為靜音雜訊部份，則加以刪除。

### 2.2.3 切割音框

由於處理信號時，信號的分析必須建立在穩定及有限的時間上，然而語音訊號是隨著時間變化的，此種特性並無法以線性非時變的方法分析擷取整段完整的語音，但語音的變化是緩慢的，因此可以假設在短時間之內，訊號的波形會呈現固定週期的趨勢，訊號具有半穩態的特性，在此短時間內的語音訊號可視為非時變訊號，此短時間常被稱為一個音框的時間，音框變成處理語音的最小單位。

語音訊號的特徵，通常無法以一個音框的特徵資料做為代表，因此常會將整段訊號切割為多個音框訊號，再針對每個音框訊號以短時距分析擷取特徵。音框長度通常固定，包含相同的取樣點，但音框的長度並無標準，有許多的選擇，唯若音框長度太長，則此音框的振幅變化劇烈，不符合半穩態特性，若音框太短，則音框訊號不具有足夠的語音特性，音框長度通常取 10~40ms，這樣的音框長度適合計算取得語音的特

徵參數，在本文中，根據使用 32ms 的音框長度，重疊 16ms 的音框長度。

#### 2.2.4 預強調

在語音信號處理當中，進行預強調，主要是為了補償語音在被壓抑的高頻部份，當聲音經過嘴唇之後，其聲音有如經過一個低通濾波器，訊號高頻會產生衰減，即高頻部份被壓抑，但人類的聽覺系統會自動將聲波高頻部份放大，為了讓之後的聲音訊號的強度值，更能接近人類聽覺系統，可將每個音框內的語音訊號，通過一個高頻濾波器，調整高頻成分的權重，以還原語音訊號。

#### 2.2.5 視窗化

由於在處理語音訊號時，將整段語音切割為多個連續的短時域音框，以符合短時距分析，然而，音框切割會造成音框訊號端點之間不連續，語音訊號邊緣產生劇烈變化，為了使每個音框在頻譜上的能量能更集中、特徵益加明顯，利用視窗化使其不連續性降低。

而視窗化函數如下：

$$Y(n) = X(n) \times W(n)$$

其中  $X(n)$  為輸入之音框語音訊號， $W(n)$  為所使用之視窗函數， $Y(n)$  為經過視窗化處理之音框語音訊號。

語音處理中，最常被使用的視窗函數為下列三種：

##### 1. 矩形視窗（Rectangular Window）：

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

2. 漢明窗 (Hamming Window) :

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

3. 漢尼窗 (Hanning Window) :

$$w(n) = \begin{cases} 0.5 - 0.5 \cos\left(\frac{2\pi n}{N-1}\right) & 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases}$$

在本文中所使用之視窗函數為漢明窗 (Hamming window)，因為漢明窗具有壓抑兩端訊號且保持中央部份的訊號的優點。

## 2.3 小結

本章之內容，主要闡述在特徵擷取前，需事先對原始的語音訊號做前處理的相關過程。人類的聲音為連續的類比訊號，隨時間而變化，具有半穩態特性，為了符合人類的聽覺系統對聲音音強和頻率的感知並非線性之關係，因此對原始訊號的高頻部份予以放大補償。

在本研究中，嬰兒的哭聲情緒辨別，與母親及家人本身的主觀認知有很大的關係，且由於人類聽覺系統對音調頻率的非線性關係，之後所擷取的頻率特徵將會利用非線性轉換機制取得，使特徵參數能更貼合人類的聽覺認知系統。關於特徵參數的擷取，在下一章節將有詳盡的解說。

## 第三章

---

### 嬰兒哭泣情緒特徵參數的擷取

由於聲音訊號的數位資料龐大，不易以完整未經處理的聲音訊號予以辨識，需對原始聲音訊號擷取其特徵參數，特徵參數是從高維度原始訊號向量空間  $X$  轉換成低維度的特徵空間  $Y$ ，此轉換必須保存原始訊號的大量資訊、降低維度和減少計算的複雜度。對情緒辨識系統而言，語音特徵參數的選擇是非常重要的，最佳的語音特徵參數是一組可以完整呈現每個聲音訊號特性的參數，且此參數必須具備著鑑別性。

情緒辨識系統的正确率，與特徵參數的擷取，有著密不可分之關係，選擇一個強健且具有鑑別性的特徵參數，對辨識系統而言是事半功倍，適合的特徵參數，不但可以減少所使用辨識機制的複雜度，降低系統整體的計算量，並可大幅的提升辨識率。

情緒辨識模型中，常被使用為的特徵參數有 LPC、MFCC 和 CEP 等，本章將一一的介紹。



### 3.1 倒頻譜(Cepstrum)

在語音訊號分析當中，常將時域的波形，經過 Fourier 轉換為頻域中的頻譜，在頻譜中觀察此訊號的特性。

在發音時，聲帶造成空氣的震動造成氣流，然後此氣流向發音腔道傳播，最後經由嘴唇向外界送出，聲帶為發音的音源，常以激發源訊號  $E_p(Z)$  表示，聲道則是假設為線性非時變系統，可用數位濾波器代表聲道發音之模型，發音的模型  $\theta(Z)$  包含了發音腔道模型  $H(z)$ ，聲道脈波模型  $G(z)$ ，以及嘴唇端的輻射模型  $R(Z)$ 。在時域上，語音的產生是由激發源訊號  $e_p(n)$  與聲道脈衝響應  $\theta(n)$  兩者褶積 (Convolution) 的結果，因此一個發音模型之系統方塊圖，包含了一激發訊號之不同表示方式，根據聲帶振動與否，決定不同之激發方式，其圖 3.1 所示[19]

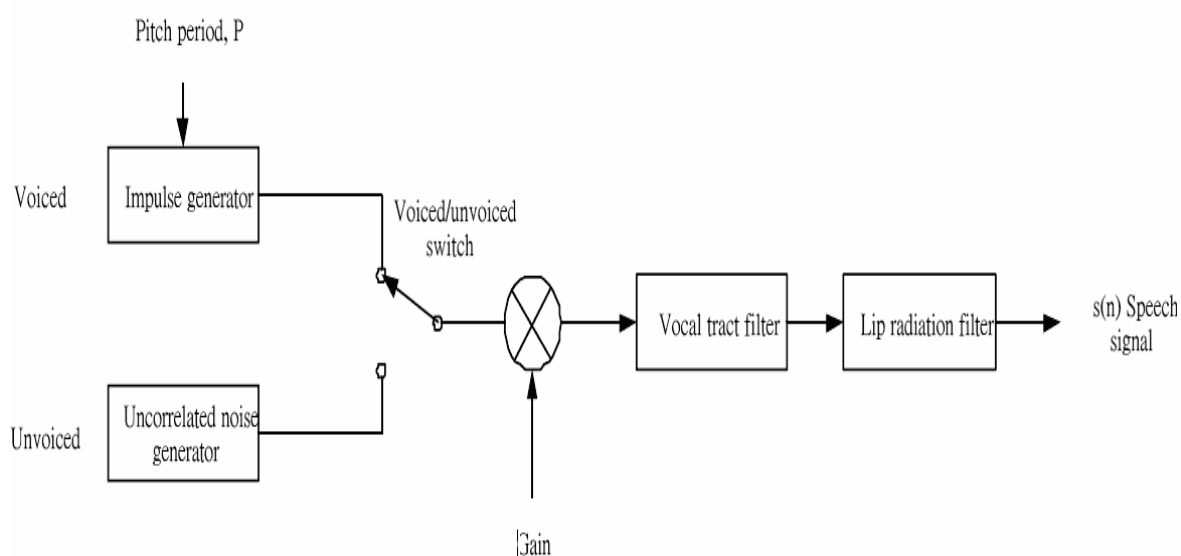


圖 3.1 發音模型的系統圖[19]

其數學式為

$$S(z) = E_p(z) \times \theta(z)$$
$$s(n) = e_p(n) * \theta(n)$$

$s(n)$  語音訊號， $e_p(n)$  激發源訊號， $\theta(n)$  聲道脈衝響應。

在情緒辨識系統當中，假設當同種情緒產生時，其聲道模型相似，不同情緒之間，其聲道模型有所不同，因此聲道特徵具備著鑑別性。

然而，如何擷取聲道特徵，則是一大問題，由於 $\theta(n)$ 和 $e_p(n)$ 之間為褶積，並非線性關係，無法直接線性分離。

由於在時域上，聲道特徵無法直接擷取，於是將語音訊號經過傅利葉轉換為頻域中的頻譜，將頻譜取對數值，此時語音訊號可視為激發源訊號與聲道脈衝在倒頻域上相加，然在再作逆向之 Fourier 轉換，回到時域中，將得到一組新參數，稱為倒頻譜係數，圖 3.2 可表示利用倒頻譜的運算可將激發源訊和聲道脈衝訊號分離。

由於因聲道震動時，激發訊號為脈衝訊號，變化劇烈，為頻譜中的高頻部份，聲道模型為低頻部份，因此可利用濾波器濾除不需要之部分。

倒頻譜係數具有分離發音腔道模型與激發訊號的特性，可以更準確計算發音腔道參數，進而掌握語音頻譜特性，在情緒辨識系統當中，被視為有效之參數，廣為使用，其倒頻譜參數的擷取流程如圖圖 3.3 所示。

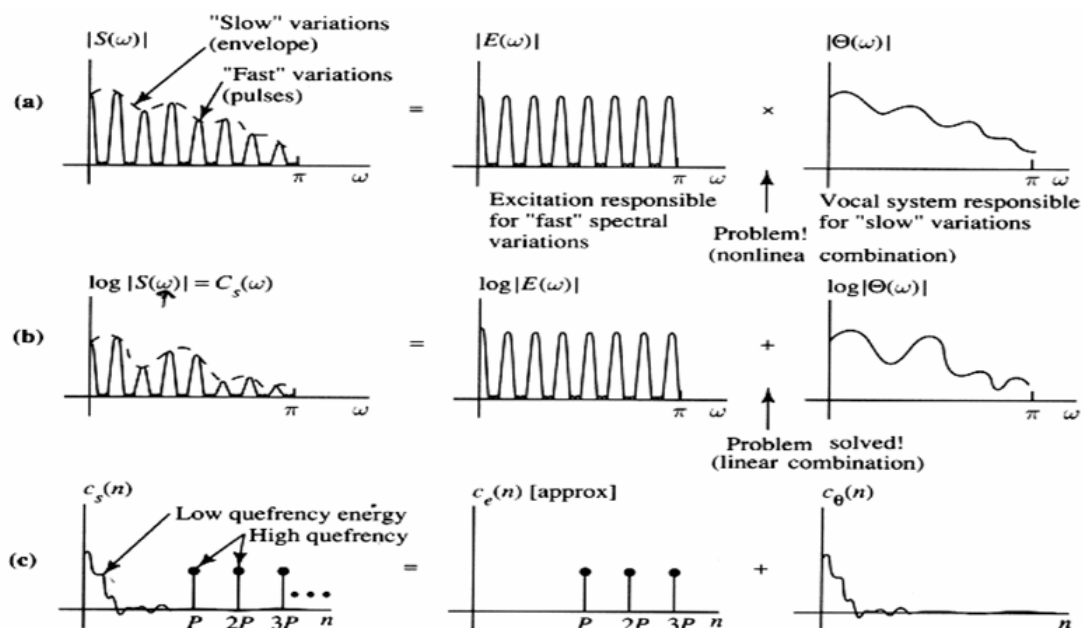


圖 3.2 利用倒頻譜將激發源訊和聲道脈衝訊號分離[19]

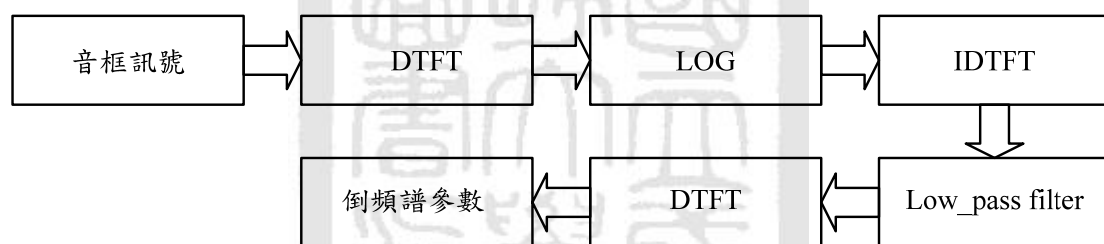


圖 3.3 倒頻譜擷取流程圖

### 3.2 梅爾倒頻譜係數

在倒頻譜中，可發現倒頻譜對每個頻帶都以相同的重要性看待。但人類的聽覺系統對頻率的感知能力，與頻率並非成線性之關係，人耳在對頻域的感受上，並非對全頻域皆有相同之敏感度。人耳對低頻率的訊

號有較高的解析度，在低頻域時可分辨較小之頻率差異。此外，人耳對單一音調的知覺感受會受到該音高頻率周圍特定頻寬內的頻譜能量所影響，該特定頻寬被稱為臨界頻寬。

在線性頻率低於 1KHz 以下，臨界頻寬約為 100Hz，在此頻域中，人耳可清楚感受到頻率的變化量，但若頻率超過 1KHz，其臨界頻寬大小成對數之關係，人耳對頻率之變化變的較不靈敏。

人耳的感知頻率與實際頻率之間關係可表示為：

$$\text{Bark} = 13 \tan^{-1}(0.00076f) + 3.5 \tan^{-1}\left(\frac{f^2}{7500^2}\right)$$

其中，Bark 為感知頻率之單位，f 為實際頻率(HZ)

與 Bark 刻度相似的另一個量度法為梅爾刻度 (Mel scale)，在情緒辨識系統中最常被使用，梅爾刻度是將 1KHz 以上聲帶頻率投射為對數，1KHz 以下為線性分佈，其公式如下：

$$\text{Mel} = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

針對人類聽覺系統的特性，對低頻部份攫取較多特徵，高頻部份攫取較少特徵，配合人耳聽覺特性，在頻域中以梅爾刻度劃分頻帶，將個別頻帶的頻率成分，結合在一起，視為能量強度，再將此頻帶能量強度轉換成倒頻譜，此倒頻譜係數稱為梅爾倒頻譜 (MFCC)。

根據臨界頻帶及遮蔽效應，將帶通濾波器設計為跨越相鄰頻帶之三角形濾波器如圖 3.4 所示，三角形中心頻率為臨界頻帶中心頻率，中心頻

率以數學式表示如下：

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

其中  $B_m(k)$  為第  $m$  個頻帶的三角濾波器， $f_m$  為第  $m$  個頻帶之中心頻率， $M$  為全部之頻帶。

將各頻率能量與三角濾波器相乘，然後累加取對數值，得到濾波器輸出的對數能量，其公式如下：

$$Y(m) = \log \left\{ \sum_{k=f_{m-1}}^{f_{m+1}} |X(k)|^2 B_m(k) \right\}$$

將對數能量轉為時域，得到梅爾倒頻譜（MFCC） $C_m$ 。

$$C_m = \frac{1}{M} \sum_{m=1}^M Y(m) \cos\left(\frac{\pi n(m-\frac{1}{2})}{M}\right)$$

其梅爾倒頻譜（MFCC）的擷取流程如圖 3.5 所示。

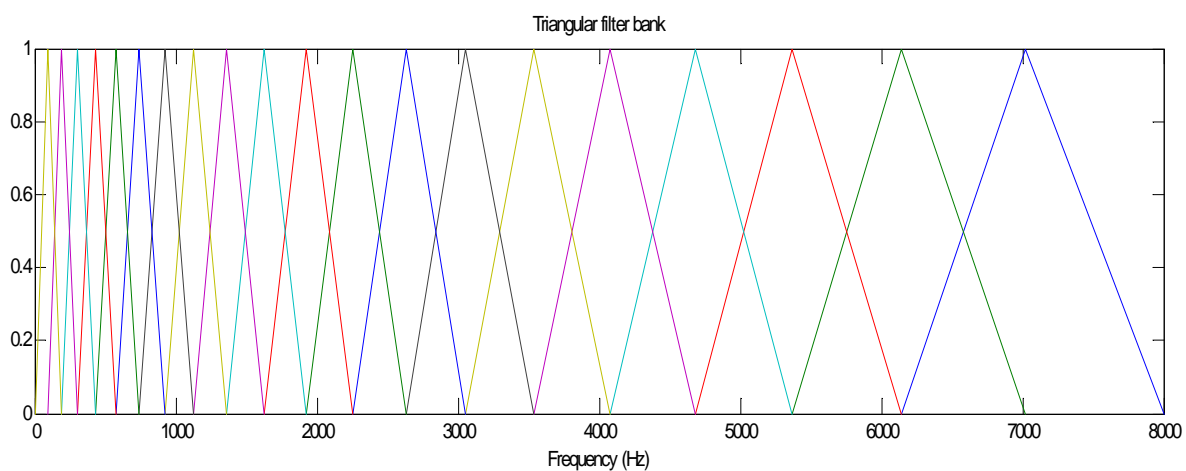


圖 3.4 梅爾刻度之帶通濾波器

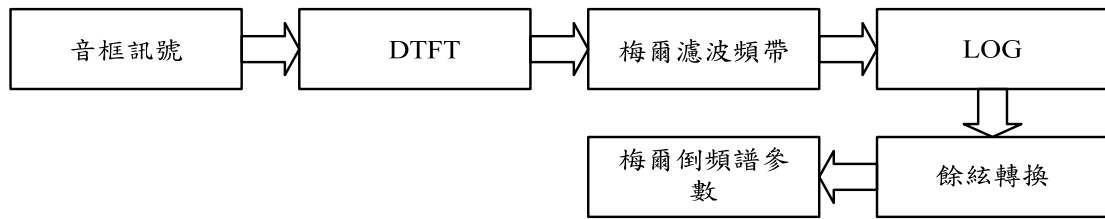


圖 3.5 梅爾倒頻譜擷取流程圖

然而，為使語音特徵更能代表語音特性，增加了差量倒頻譜，差量倒頻譜代表著倒頻譜係數對時間的變化性，利用倒頻譜係數相對於時間之斜率，觀察倒頻譜係數在時間的動態變化。

一次差量倒頻譜公式如下：

$$\Delta C_m(t) = \frac{\sum_{\tau=-M}^M C_m(t+\tau)\tau}{\sum_{\tau=-M}^M \tau^2}, m=1, \dots, M$$

$C_m(t)$  代表第幾個音框的梅爾倒頻譜係數， $\Delta C_m(t)$  代表第幾個音框的一次差量梅爾倒頻譜係數

### 3.3 線性預估編碼（LPC）

發音模型包含了激發訊號和聲帶脈波訊號。在時域上，由於激發訊號和聲帶脈波訊號無法以線性直接分離，故倒頻譜使用 log 方式，分離激發訊號和聲道脈波訊號。

在一個音框時間之內，由於語音具有半穩態性，因此假設此時發音腔道為靜止之一線性非時變系統，此時發音模型可用線性預估模型描述。

在語音訊號處理中，線性預估模型為一重要表示式，可利用線性預估模型擷取重要的特徵參數，此特徵參數為線性預估編碼（LPC），用此參數可描述語音的頻域特性。

線性預估編碼（LPC）為一重要之特徵，其優點於可用較少的參數，有效的表示該發音模型的訊號特徵，計算簡易。

LPC 主要的概念為一個取樣點的振幅，與鄰近取樣的振幅有相關性，因此假設目前的訊號樣本值，可由前 P 個訊號樣本值之線性組合予以估算預測，其公式如下：

$$s(n) = \sum_{k=1}^p \alpha_k s(n-k) + \Theta_0 e_p(n)$$

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

$s(n)$  為實際的語音訊號， $\hat{s}(n)$  為預估的語音訊號， $e_p(n)$  為激發訊號，假設激發訊號為週期性脈衝訊號，每隔一時間，出現一脈衝，在此一間隔時間，激發訊號可視為零， $\alpha_1, \alpha_2, \dots, \alpha_p$  為 LPC 參數。

實際訊號與預估訊號之誤差為

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

$$= \sum_{k=1}^p \alpha_k s(n-k) + \Theta_0 e_p(n) - \sum_{k=1}^p \alpha_k s(n-k)$$

希望能夠尋找一組 LPC 參數，使實際訊號與預估訊號之誤差和為最小，利用下列最小平方法公式找尋誤差和最小值。

$$\begin{aligned}\xi &= \frac{1}{N} \sum_{n=0}^{N-1} \{ \varepsilon(n) \}^2 = \frac{1}{N} \sum_{n=0}^{N-1} \left\{ [s(n) - \sum_{k=1}^p \alpha_k s(n-k)]^2 \right\} \\ \frac{\partial \xi}{\partial \alpha_l} &= \frac{1}{N} \sum_{n=0}^{N-1} \left\{ 2[s(n) - \sum_{k=1}^p \alpha_k s(n-k)] s(n-l) \right\} = 0 \\ \Rightarrow \frac{1}{N} \sum_{n=0}^{N-1} s(n)s(n-l) - \sum_{k=1}^p \left\{ \frac{1}{N} \sum_{n=1}^N \alpha_k s(n-k)s(n-l) \right\} \\ &= r_x(l) - \sum_{k=1}^p \alpha_k r_k(|l-k|) = 0, \quad l=1,2,\dots,p\end{aligned}$$

線性預估方程式，解法有許多種，計算最迅速為自相關係數解法。

自相關係數定義為：

$$r_x(l) = \frac{1}{N} \sum_{n=0}^{N-1} s(n)s(n-l), \quad l=1,2,\dots,p$$

$$r_x(l) - \sum_{k=1}^p \alpha_k r_k(|l-k|) = 0$$

$$\Rightarrow r_x(l) = \sum_{k=1}^p \alpha_k r_k(|l-k|)$$

利用杜賓遞迴演算法(Levinson-Durbin recursion)，解出自相關函數係數，便可得到所需之 LPC 係數。

其線性預估編碼(LPC)的擷取流程如圖 3.6 所示。

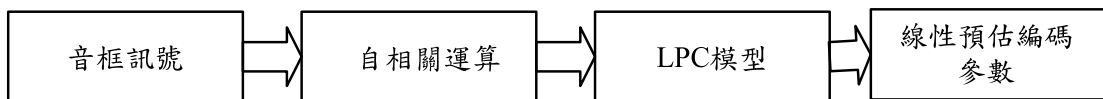


圖 3.6 線性預估編碼流程圖



### 3.4 線性預估模型所導出的倒頻譜係數 (LPCC)

線性預估模型是為一個數學線性模型，然而，發音系統並非是個完全的線性模型，LPC 特徵只能表示發音系統的線性部份的特性，無法表示其非線性部分之特性。使用線性預估模型所導出的倒頻譜參數 (LPCC) 可同時具備發音系統中線性和非線性特性，彌補 LPC 之不足部份。

首先依照章節 3.3 LPC 參數擷取步驟，訊號前處理取得音框，將音框作自相關運算，求得 P 階的線性預估參數，得到線性預估模型，再利用倒頻譜擷取方式，從線性預估模型求得倒頻譜係數，此係數稱為 LPCC 如圖 3.7 所示。

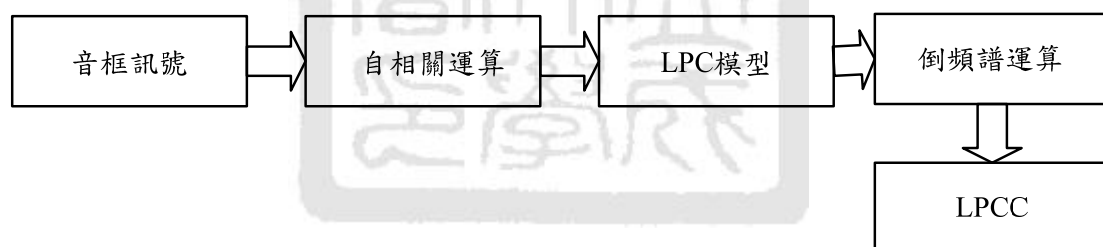


圖 3.7 LPCC 流程圖

### 3.5 基頻 (Pitch)

語音訊號的形成，必須要由聲帶音源發出激發訊號，經過聲道脈衝模型，調變聲帶音源訊號，就可發出不同之聲音。所謂之基頻，意指聲帶音源訊號原始的震動頻率，因此基頻可代表聲音頻率的高低。

基頻擷取流程如圖 3.8 所示，簡述如下：

1. 將整段原始訊號經前處理，切割多個互相重疊之音框訊號。
2. 針對每個音框予以自相關運算，並經由中央濾波器 (center clipping)，濾除中央原點附近之雜訊。
3. 找尋整段音框訊號中，相鄰峰點在時間軸之距離，其倒數為基頻。
4. 計算所有音框之基頻，即可得到整個訊號的基頻分佈圖。

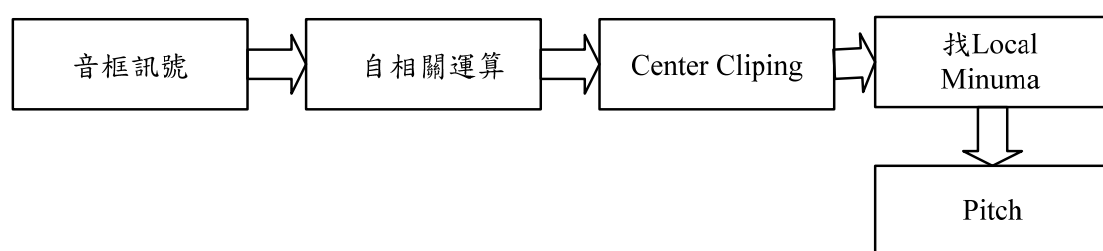


圖 3.8 Pitch 流程圖

### 3.6 降低特徵維度-主要分量分析 (PCA)

在錄製嬰兒哭聲時，由於無法控制嬰兒哭聲之時間長短，導致錄音時間長短不同，在將哭聲訊號切割為重疊之音框訊號時，發現音框數目不一，且音框數目龐大，然而過大的資料量並非皆有助於情緒的辨識，因並非所有特徵皆具有貢獻性，不良特徵會導致分析時的誤差，尤其當不良特徵數量過大時，會大幅的影響辨識的結果。

因此對高維度的資料必須進行降維之處理，本節所使用之主要分量分析法，是一種常見的降維演算法，此演算法不需事前知悉此特徵的資料類別，屬於不包含類別資訊的資料擷取，其目標只是降低資料的維度，將高維度空間中的資料投影至低維度空間，並找出最佳之投影方式，使投影後資料點能盡量分散，希望能保持資料在高維度空間分佈的特性。

本研究將不同的 PCA 參數互相組合，提出三種 PCA 降維方式，圖 3.9 為其擷取流程圖，根據此流程圖作一個詳細的說明。

當哭聲訊號輸入時，經過前處理得到  $M$  組的音框訊號，每一組音框訊號經過特徵參數之擷取，可得到  $D$  維度的特徵參數，此時可得到  $M$  組的  $D$  維度的特徵參數向量，組成為  $M \times D$  的矩陣。

但此維度依然太大，因此對  $M \times D$  的矩陣予以 PCA 的數學運算，本研究中只使用 PCA1 (第一主要分量)、PCA2 (第二主要分量)、PCA3 (第

三主要分量)這三部份,PCA1、PCA2、PCA3 皆為  $D$  維的特徵向量,PCA1 保有著  $M \times D$  的矩陣的最重要的空間向量分佈,擁有大部份的重要資訊,PCA2 擁有少許的重要資訊,PCA3 擁有比 PCA2 更少許的資訊。

為了觀察 PCA1、PCA2、PCA3 所攜帶資訊的重要性,因此對這三種係數加以組合,PCA\_type1 是直接只使用 PCA1 為特徵參數,PCA\_type1 為  $D$  維的向量,PCA\_type2 是結合 PCA1 和 PCA2 這兩種  $D$  維特徵向量,因此 PCA\_type2 為 2 個  $D$  維的特徵向量,結合方式如圖 3.12 所示,依此類推,PCA\_type3 使用相同的結合方式,為 3 個  $D$  維的特徵向量。

圖 3.10 為 A-10-4032 的哭聲訊號的平均值的表示圖,圖 3.11 為對 A-10-4032 的哭聲訊號予以 PCA 運算,取得 PCA1(第一主要分量)、PCA2 (第二主要分量)這兩種參數。圖 3.12 和圖 3.13 則為結合 PCA1 和 PCA2 這兩種  $D$  維特徵向量,而產生 PCA\_type2 的特徵參數。

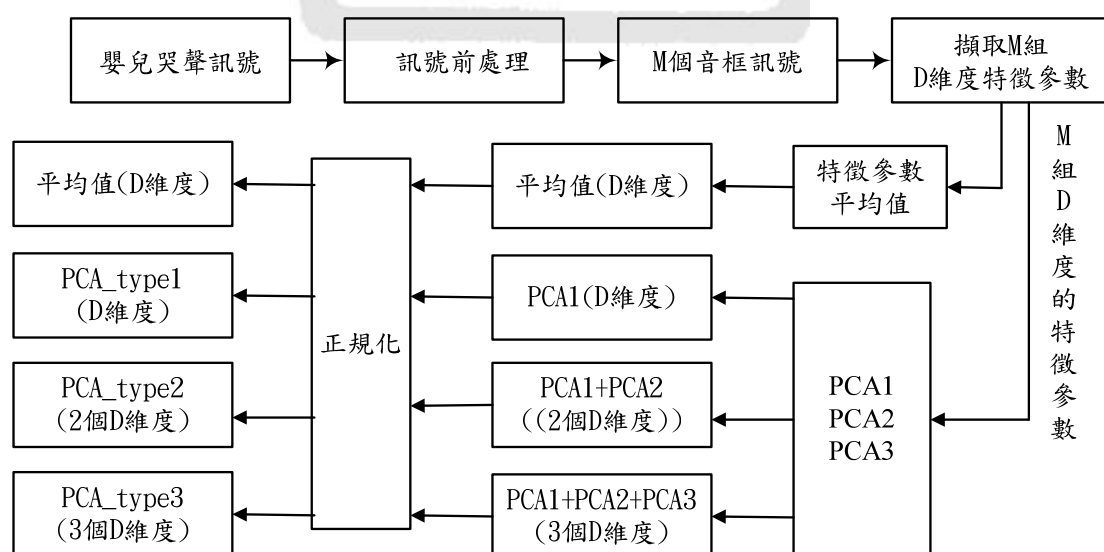


圖 3.9 特徵參數 PCA 降維流程圖

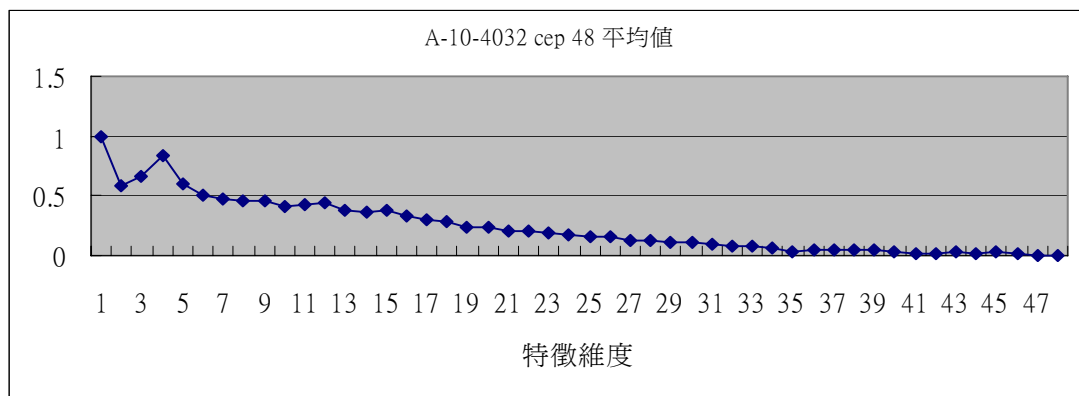


圖 3.10 A-10-4032 cep48 全部音框疊合正規化平均值  
(不經 PCA 處理)

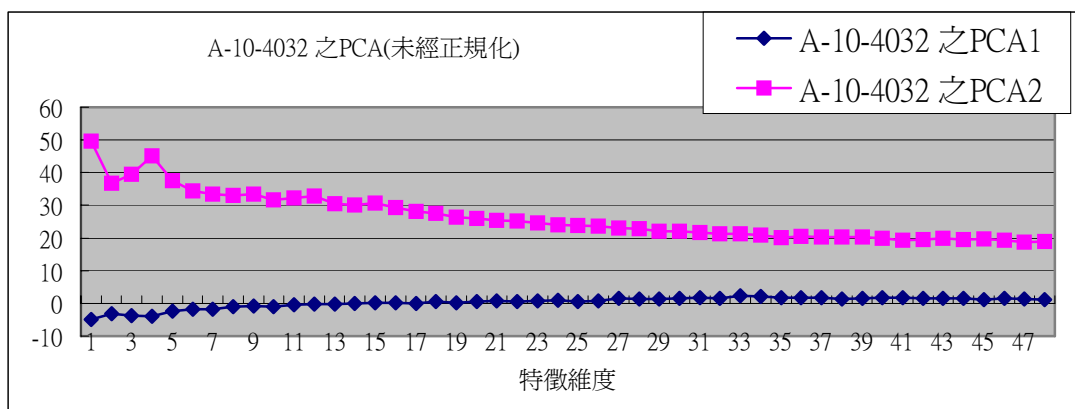


圖 3.11 A-10-4032 cep48 PCA1 及 PCA2(未經正規化處理)

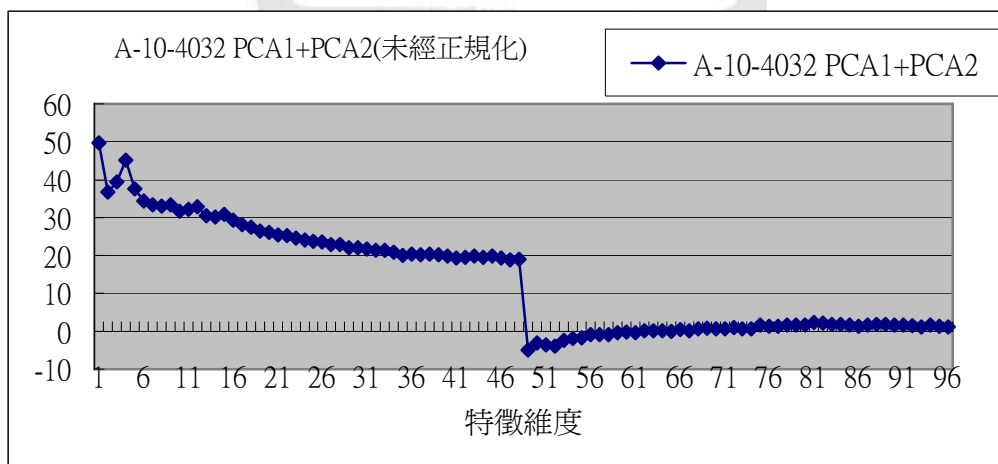


圖 3.12 A-10-4032 cep48 PCA1+PCA2(未經正規化處理)

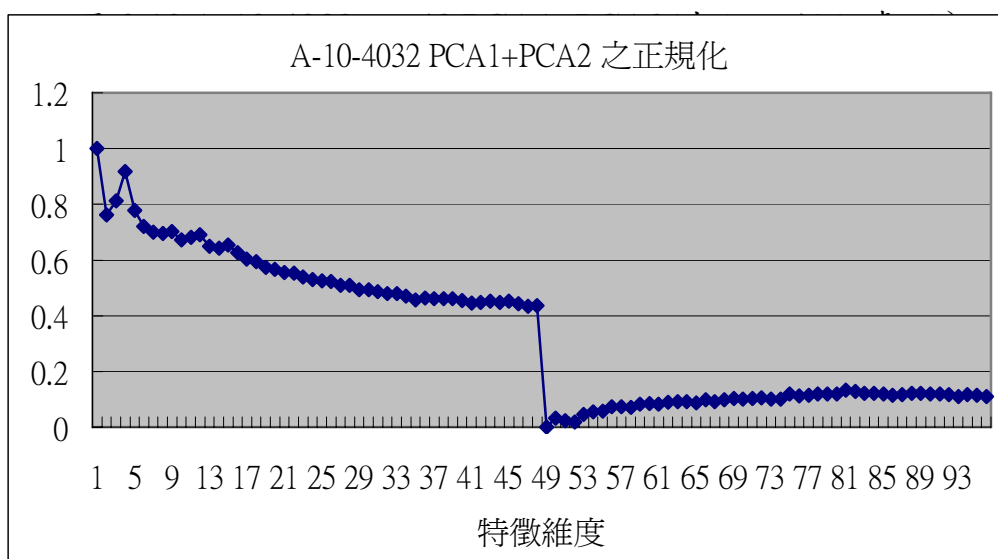


圖 3.13 A-10-4032 cep48 PCA1+PCA2 (經正規化處理)



## 第四章

---

### 辨識機制

本章節將闡述如何利用所擷取的特徵參數，使用辨識機制演算法，訓練情緒辨識模型。將模型訓練完畢之後，即可輸入欲辨識之情緒哭聲訊號，與事先訓練好的情緒辨識模型的資料庫進行比對辨識，尋找相似度最高之情緒類型語音，以完成辨識工作。在本文中，選擇類神經網路及 K-最近鄰居（k-nearest neighbor K-NNR）分類法。下面將會分別介紹這兩種方法

#### 4.1 類神經網路簡述

類神經網路是模擬人類神經網路的一個計算系統，神經網路是由神經細胞加上軸突與突觸所構成，類神經網路在架構上是由眾多相連的人工神經元及連接其間的權重值所組成，經過不斷的學習訓練模擬人類神經網路的學習能力。

人工神經元透過神經鍊結與其他神經元互相的連結，形成一個高度

密集的網路，此網路具備著高度的分辨能力，同時也具有記憶性，然後可以藉由輸入訊號的刺激抑制或增強神經元之間的神經鍊結，使得神經網路具有強大的學習能力。

神經網路的架構包含著輸入層(Input Layer)、隱藏層(Hidden Layer)、輸出層(Output Layer)和激發函數(Activation function)等。輸入所提供之激發訊號，經過其間運算單元之計算，輸出系統之響應，運算單元之間訊號傳遞之強弱由權重值決定。學習過程中，利用學習法則調變權重值，使網路達到最佳狀態。

神經網路的基本架構，可分成三個部分予以說明：

(1) 運算單元：為類神經網路之基本單元，常稱為神經元，當接受激發訊號輸入後，經函數轉換處理完之後，傳送至其他運算單元或輸出層。

(2) 神經層：將具備相同功能之神經元集合為一神經層，神經網路通常具有輸入層、輸出層和隱藏層這三部份。

(3) 神經網路：所有神經層之集合稱為神經網路，可利用學習法則加以調變權重值，使神經網路具有強健性。

根據學習型態，則可分為兩種類型：

(1) 監督式學習(supervised learning)

監督式學習是以疊代方式修正神經網路的權重值，由於希望輸出



結果能符合預期之結果，故在整個訓練過程中，將不斷修正權重值，使其神經網路輸出值與期望輸出值之間的誤差降低，直至誤差小於臨界值或疊代之次數超過限制之次數，訓練才會停止。

## （2）非監督式學習（unsupervised learning）

非監督式學習僅需要提供輸入訊號，不需提供期望輸出值，因為網路不需要使用誤差訊息去改善神經網路的輸出，僅需要依照輸入資料便可以判斷其類別。

一般而言，非監督式學習神經網路通常是被使用在分類的問題上，而監督式神經網路則是用在模型訓練方面。

綜合言之，神經網路其優點可歸納如下：

（1）高度的平行運算能力：神經網路利用大量的神經元，模擬生物神經系統，使其具備高度平行處理能力。

（2）學習與記憶能力：神經網路利用輸入訊號的刺激或抑制調變權重值，學習並記憶輸入與輸出之間的相互關係。

（3）容錯強健度：神經網路是採取分散記憶式儲存資訊，因此當輸入訊號不完整或訊號夾雜少量雜訊時，神經網路可作適當的處理，具備模糊推論之能力，且網路具有大量的鍊結，當部份鍊結失效時，網路依然可提供正確之輸出。

（4）非線性處理能力：神經網路可利用回授能力，克服系統中高

度的非線性關係。

由於類神經網路的優點與特性，非常適合應用在情緒辨識模型，下節將對所選用之類神經網路類型與架構予以說明。

## 4.2 倒傳遞類神經網路

傳遞神經網路（Back-Propagation Neural Network, BPN），是目前類神經網路學習模式中最具代表性，應用實務最普遍的模式。本研究所採用的類神經網路為倒傳遞類神經網路。倒傳遞類神經網路為多層前饋式網路，可以處理輸入輸出之間非線性關係之辨識，用監督式學習方式學習。網路模組是由輸入層、隱藏層和輸出層所構成，網路可包含多層的隱藏層，輸入層和輸出層則皆為單層。每一個神經元的輸出，都乘上其相對應的權重值再加總，再透過激發函數的計算產生輸出訊號。在訓練階段，藉由輸入訊號於網路架構，獲得實際輸出值，計算實際輸出值與期望值的平方誤差，將此誤差訊號回饋至網路系統，疊代調整權重值，重複此步驟計算，直至誤差達到最小平方誤差。

在倒傳遞網路中，常用之激發函數有

### (1) Sigmoid Transfer Function

$$f(net) = \frac{2}{1 + \exp(-2net)} - 1$$

(2) Linear Transfer Function

$$f(net) = net$$

(3) Log-Sigmoid Transfer Function

$$f(net) = \frac{1}{1 + \exp(-net)}$$

其圖形如圖 4.1 至 4.3 所示。

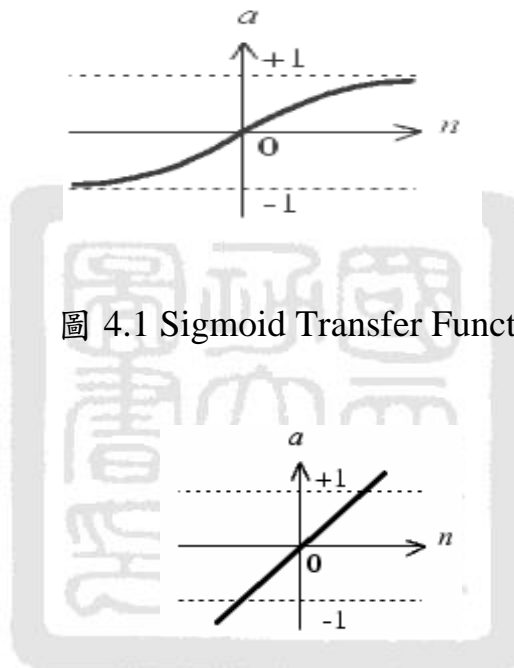


圖 4.1 Sigmoid Transfer Function

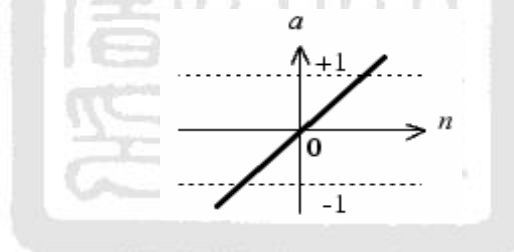


圖 4.2 Linear Transfer Function

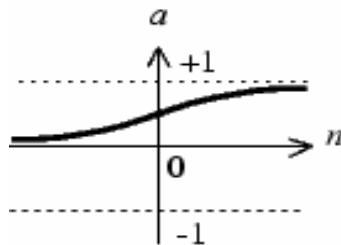


圖 4.3 Log-Sigmoid Transfer Function

倒傳遞神經網路選用的學習法則為梯度下降演算法 (gradient descent algorithm)，定義誤差函數為

$$E_k = \frac{1}{2} \sum_j (d_j - o_j)^2 \quad E = \sum_k E_k$$

$E_k$  為第  $K$  個訓練神經元之誤差， $E$  為所有訓練神經元的總訓練誤差

$d_j$  為第  $j$  個輸出節點之期望輸出值， $O_j$  為實際運算之輸出值

定義調變權重值的通式為

$$\Delta W = -\eta \cdot \nabla_w E = -\eta \frac{\partial E}{\partial W} \quad \eta \text{ 為學習速率 } W \text{ 為權重值}$$

倒傳遞的學習過程流程，分為下列幾步驟：

步驟 1：決定網路架構、網路層數和各層之間的神經元數目。

步驟 2：設定網路的權重值和偏權值的初始值，可隨機設定。

步驟 3：開始輸入新的訓練訊號，計算隱藏層和輸出層的輸出。

步驟 4：計算實際輸出值和期望值之誤差，調整權重值和偏權值。

步驟 5：計算誤差如未達到門檻值，回到步驟三，直至網路收斂為止

### 4.3 K-最近鄰居分類法

K-最近鄰居分類法被稱為 KNNR，是一種傳統的辨識方法，其辨識原理因為擁有相似特徵的資料，在特徵空間中，將會聚集在一起，因此將要欲辨識之特徵以高度空間來表示，則同一類之資料彼此之間的距離

應該會最短。當未知的特徵輸入時，投射至高度特徵空間，計算與訓練資料特徵的距離，尋找離未知特徵訊號最近的訊號點，判斷此未知特徵訊號與最近的訊號點為相同類別。



## 第五章

---

### 嬰兒哭泣情緒辨識實驗與結果

#### 5.1 嬰兒哭聲情緒資料庫介紹

本系統所處理之訊號，為事先錄製好的語音信號。錄音設備為 SONY HD player MZ-NH1，其所錄製訊號取樣頻率為44KHz、單聲道及取樣解析度為8位元的PCM語音訊號，透過USB傳輸介面傳至電腦，再利用軟體轉換為wave檔案格式。所收集之嬰兒哭聲情緒語音資料庫的來源，皆為成大醫院的新生嬰兒，其出生之年齡從出生一天至六天之內，嬰兒並無生理疾病及先天生理缺陷，其語音資料庫可以分成下列五種：

A 秤重 (scale)：此資料庫中所錄製之嬰兒哭聲，是錄製新生嬰兒在早上全身赤裸量測體重，因溫度改變所造成之哭聲。

B 洗澡 (bath)：此資料庫中所錄製之嬰兒哭聲，是錄製新生嬰兒因洗澡所造成之哭聲。

C 飢餓 (hunger)：此資料庫中所錄製之嬰兒哭聲，是錄製新生嬰兒因肚子餓所造成之哭聲。

D 打針 (injection)：此資料庫中所錄製之嬰兒哭聲，是錄製新生嬰兒因打針疼痛所造成之哭聲。

E 酒精 (alcohol)：此資料庫中所錄製之嬰兒哭聲，是錄製新生嬰兒塗抹酒精於臍帶，而感受刺激所造成之哭聲。

將所收集之嬰兒哭聲訊號建立成一嬰兒哭聲資料庫，資料庫整理方式如表5.1所示：

表5.1嬰兒情緒哭聲資料庫

編號	類別	推測原因	身分	出生日期	性別	GROUP
A-63-5057	秤重	不安全感	63	10月13日	女	5057
E-59-5065	酒精	溫度改變	59	10月12日	男	5065
C-60-5066	飢餓	肚子餓	60	10月11日	男	5066

在資料庫中，每個嬰兒哭聲訊號皆有個別之編號，編號代表此嬰兒哭聲訊號的相關資訊，例如編號A-63-5057，A代表此哭聲訊號屬於A類別，是由量測體重，因溫度改變所造成之哭聲，63代表是由代號63的嬰兒發出此嬰兒哭聲，5057是原始的編號，代表是第五次錄音時，所錄製的第57個訊號。此資料庫所收集的嬰兒哭聲是由253位嬰兒所發出，資料庫中收集了618個哭聲訊號，將哭聲訊號分為兩類，一類為測試訊號，另一類則為訓練訊號，其分類如表5.2所示。

表5.2哭聲情緒辨識系統訓練與測試數量

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

類別	訓練數量	測試數量	總和數量
A	70	40	110
B	81	72	153
C	121	73	194
D	34	43	77
E	46	38	84

## 5.2 實驗介紹

本論文有七個實驗，各實驗目標如下：

實驗一：比較不同降維方式對辨識率之影響，降維方式有 MEAN、PCA\_type1、PCA\_type2、PCA\_type3 等四種。

實驗二：探討不同維度之倒頻譜特徵參數對辨識率之影響。

實驗三：探討不同維度之梅爾倒頻譜特徵參數對辨識率之影響。

實驗四：探討不同維度之線性預估編碼參數對辨識率之影響。

實驗五：探討不同維度之 LPCC 對辨識率之影響。

實驗六：擷取嬰兒哭聲的其他特徵，與 LPCC 作特徵之組合，比較其辨識率的優劣。

實驗七：擷取嬰兒哭聲的其他特徵，與梅爾倒頻譜作特徵之組合，



比較其辨識率的優劣。

使用辨識率為判別嬰兒情緒辨識系統好壞之準則，辨識率以哭聲訊號音檔為單位，辨識正確之音檔佔測試音檔之比率。

數學式如下：

$$\text{辨識率} = \text{辨識正確音檔數} / \text{測試音檔數} * 100\%$$

### 5.2.1 實驗一

以最佳的維度特徵 LPCC96 維為特徵參數，ANN 為辨識模型，辨識五種嬰兒情緒哭聲。

實驗目的為選擇最適當之特徵參數降維方式，因此對 MEAN、PCA\_type1、PCA\_type2 和 PCA\_type3 這四種降維方式予以比較，期能找到最適合的降維方式，其結果如表 5.3 和圖 5.1 所示。

表 5.3 LPCC96 不同降維方式辨識率比較

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

	LPCC96 不同降維方式辨識率(ANN)			
類別	MEAN	PCA_type1	PCA_type2	PCA_type3
A	52.5%	62.0%	90.0%	62.5%
B	61.1%	65.3%	80.6%	70.8%
C	50.7%	60.3%	82.2%	53.4%
D	62.8%	46.5%	88.4%	69.8%
E	44.4%	50.0%	81.6%	61.1%

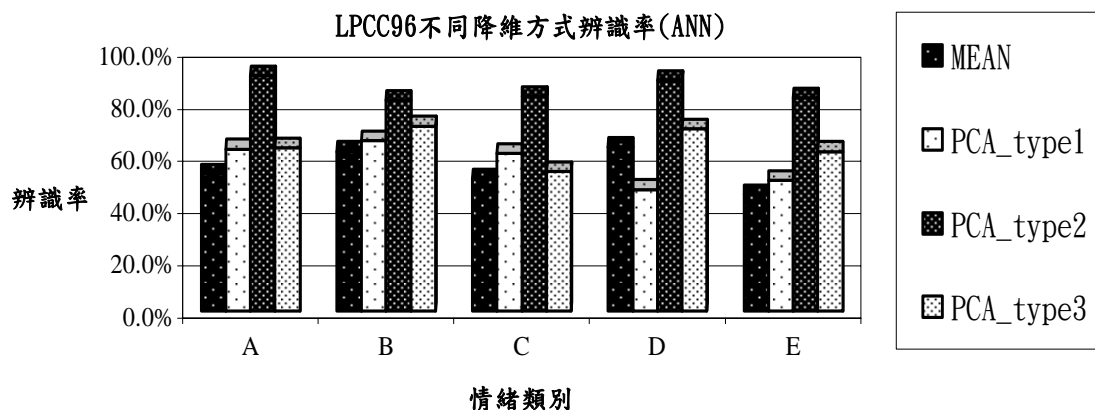


圖 5.1 LPCC96 不同降維方式辨識率比較

### 討論分析

1 針對 MEAN、PCA\_type1、PCA\_type2、PCA\_type3 四種降維方式予以比較，發現 PCA\_type2 的辨識率最高，次之為 PCA\_type3，MEAN、PCA\_type1 這兩種則是相差無幾。

2 同樣為利用 PCA 方式降低維度，比較 PCA\_type2、PCA\_type3，其差異為 PCA\_type3 的維度高於 PCA\_type2，但其辨識率並無高於 PCA\_type2，代表高維度的特徵不一定會導致辨識率增加，太高維度的特徵會使計算複雜且其特徵中易包含雜訊，影響辨識率。

3 在之後的情緒辨識實驗中，特徵將採用 PCA\_type2 的降維方式。

### 5.2.2 實驗二

以倒頻譜為其特徵，用 ANN 和 KNNR 這兩種辨識機制，利用 PCA\_type2 方法降低維度，其結果如表 5.4、表 5.5、圖 5.2 和圖 5.3 所示。

### 實驗目的

比較倒頻譜參數的不同維度，對辨識率之影響，挑選一最佳辨識維度，實驗之維度有 CEP24、CEP48、CEP60、CEP72、CEP96 和 CEP120 六種不同維度。

#### ANN 測試之實驗結果

表 5.4 不同維度的 CEP 辨識率比較(ANN)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

不同維度的 CEP 辨識率(ANN)						
類別	CEP24	CEP48	CEP60	CEP72	CEP96	CEP120
A	47.5%	52.5%	60.0%	65.0%	72.5%	75.0%
B	69.4%	62.5%	58.3%	72.2%	66.7%	62.5%
C	63.0%	65.8%	61.6%	52.1%	61.6%	60.3%
D	53.5%	60.5%	72.1%	67.4%	69.8%	72.1%
E	52.6%	52.6%	65.8%	65.8%	44.7%	57.9%

不同CEP維度之辨識率(ANN)

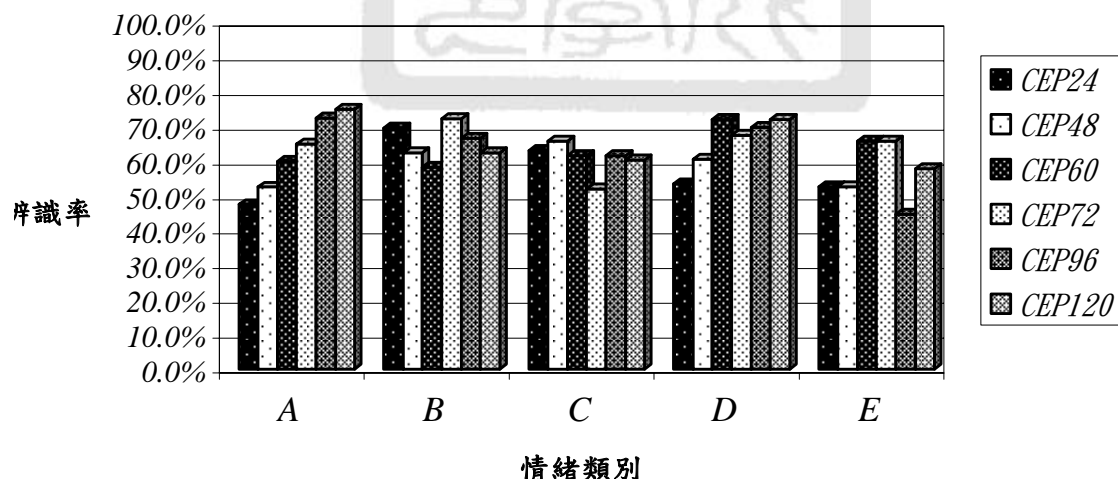


圖 5.2 不同維度的 CEP 辨識率比較(ANN)

## KNNR 測試之實驗結果

表5.5不同維度的CEP辨識率比較(KNNR)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

	不同維度的 CEP 辨識率(KNNR)					
類別	CEP24	CEP48	CEP60	CEP72	CEP96	CEP120
A	15.0%	15.0%	15.0%	25.0%	35.0%	27.5%
B	30.6%	33.3%	38.9%	40.3%	45.8%	50.0%
C	31.5%	38.4%	35.6%	37.0%	45.8%	35.6%
D	18.6%	30.2%	27.9%	32.6%	14.0%	18.6%
E	26.3%	34.2%	13.2%	26.3%	15.8%	21.1%

不同CEP維度之辨識率(KNNR)

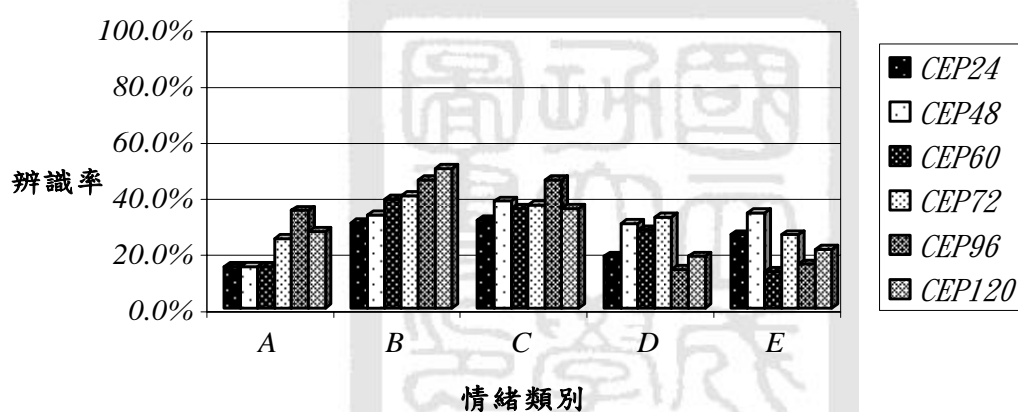


圖 5.3 不同維度的 CEP 辨識率比較(ANN)

## 討論分析

- 從表 5.4 與表 5.5 中可發現，在相同維度的倒頻譜特徵中，ANN 辨識機制的辨識率高於使用 KNNR 為辨識機制，兩者差異頗大，KNNR 之辨識率皆低於 50%，ANN 則皆高於 50%，代表倒頻譜特徵在特徵空間中，不同類別的差異性不大，同類別之間的相似性不高。

2. 從圖 5.2 與圖 5.3 中可發現，隨著維度的增加，其辨識率並無明顯的提升，ANN 辨識機制與 KNNR 辨識機制皆有相同情形，代表維度的增加對辨識率影響不大。
3. 比較 CEP24、CEP48、CEP60、CEP72、CEP96、CEP120 這六種不同維度的辨識率，發現 CEP96 擁有最高的辨識率，因此挑選 CEP96 為倒頻譜最佳維度特徵。

### 5.2.3 實驗三

以梅爾倒頻譜為其特徵，使用 ANN 和 KNNR 這兩種辨識機制，利用 PCA\_type2 方法降低維度，其結果如表 5.6、表 5.7、圖 5.4 和圖 5.5 所示。

#### 實驗目的

比較梅爾倒頻譜參數的不同維度，對辨識率之影響，挑選一最佳辨識維度，實驗之維度有 MFCC24、MFCC48、MFCC60、MFCC72 和 MFCC96 五種不同維度。

## ANN 測試之實驗結果

表 5.6 不同維度的 MFCC 辨識率比較(ANN)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

不同維度的 MFCC 辨識率(ANN)					
類別	MFCC24	MFC48	MFCC60	MFCC72	MFCC96
A	37.5%	67.5%	45.0%	87.5%	70.0%
B	50.0%	62.5%	65.3%	76.4%	75.0%
C	53.4%	46.6%	49.3%	75.3%	47.9%
D	67.4%	86.0%	72.1%	79.1%	65.1%
E	78.9%	73.7%	94.7%	89.5%	38.1%

不同MFCC維度之辨識率(ANN)

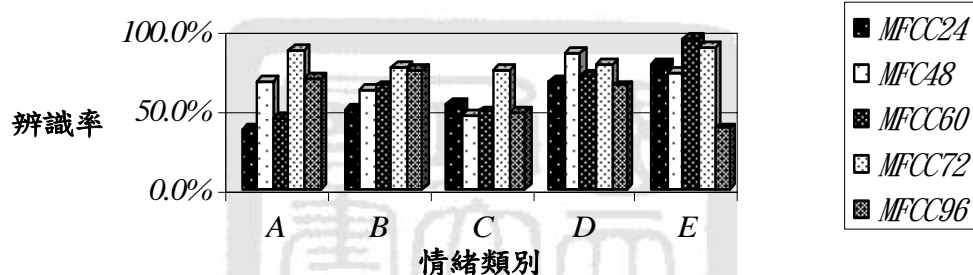


圖 5.4 不同維度的 MFCC 辨識率比較(ANN)

## KNNR 測試之實驗結果

表5.7不同維度的MFCC辨識率比較(KNNR)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

不同維度的 MFCC 辨識率(KNNR)					
類別	MFCC24	MFC48	MFCC60	MFCC72	MFCC96
A	35.0%	22.5%	25.0%	60.0%	15.0%
B	47.2%	43.1%	43.1%	25.0%	40.3%
C	35.6%	41.1%	47.9%	56.2%	34.2%
D	30.2%	32.6%	34.9%	86.0%	32.6%
E	63.2%	44.7%	60.5%	92.1%	52.4%

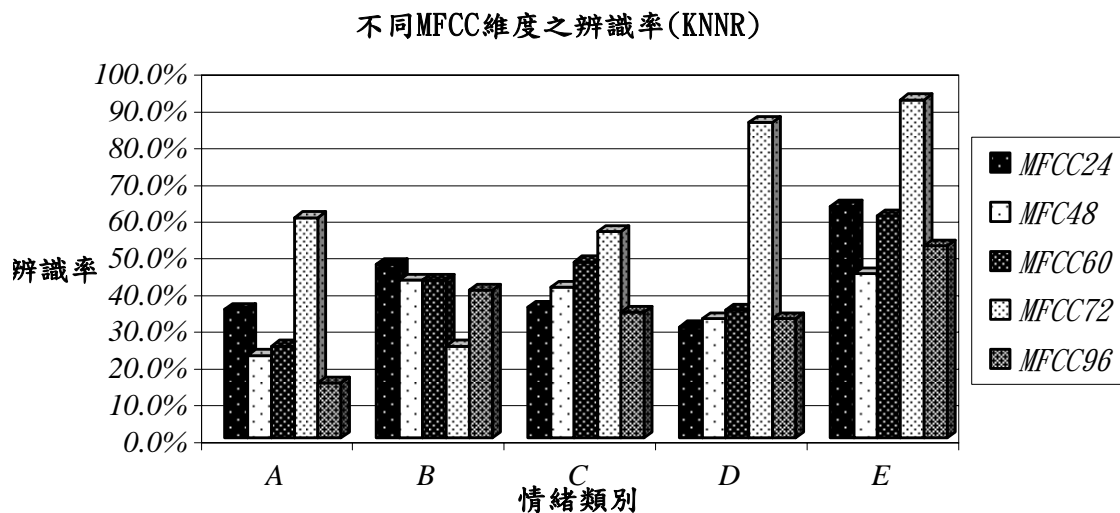


圖 5.5 不同維度的 MFCC 辨識率比較(KNNR)

#### 討論分析

1. 從表 5.6 與表 5.7 中可發現，在相同維度的梅爾倒頻譜特徵中，ANN 辨識機制中，每個類別的辨識率並非皆高於 KNNR 辨識模組，在 MFCC72 特徵中，KNNR 辨識模組中的 D、E 類別辨識率高於 ANN 辨識模組，但對總體辨識率而言，ANN 辨識模組的辨識率還是高於 KNNR 辨識模組。
2. 由圖 5.4 與圖 5.5 中可發現，隨著維度的增加，其辨識率有明顯的提升，ANN 辨識機制可明顯觀察到此情形，代表維度的變化對辨識率影響很大，在 ANN 辨識模組中，從 MFCC24 到 MFCC72，辨識率變化最大，漲幅為最大，觀察五個類別的辨識率，發現 E 類別辨識率為最高。
3. 比較 MFCC24、MFCC48、MFCC60、MFCC72、MFCC96 五種不同維度的辨識率，發現 MFCC72 擁有最高的辨識率，因此挑選 MFCC72 為

梅爾倒頻譜最佳維度特徵。

#### 5.2.4 實驗四

以線性預估編碼為其特徵，使用 ANN 和 KNNR 這兩種辨識機制，利用 PCA\_type2 方法降低維度，其結果如表 5.8、表 5.9、圖 5.6 和圖 5.7 所示。

##### 實驗目的

比較線性預估編碼參數的不同維度，對辨識率之影響，挑選一最佳辨識維度，實驗之維度有 LPC24、LPC48、LPC60、LPC72 四種不同維度。

##### ANN 測試之實驗結果

表 5.8 不同維度的 LPC 辨識率比較(ANN)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

不同維度的 LPC 辨識率(ANN)				
類別	LPC24	LPC48	LPC60	LPC72
A	55.0%	65.0%	62.5%	60.0%
B	68.1%	63.9%	68.1%	62.5%
C	65.8%	60.3%	61.6%	53.4%
D	67.4%	65.1%	67.4%	55.8%
E	63.2%	63.2%	63.2%	63.2%



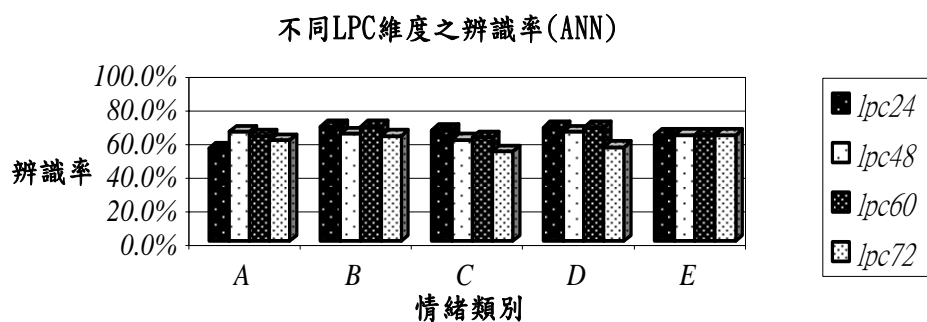


圖 5.6 不同維度的 LPC 辨識率比較(ANN)

## KNNR 測試之實驗結果

表 5.9 不同維度的 LPC 辨識率比較(KNNR)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

不同維度的 LPC 辨識率(KNNR)				
類別	LPC24	LPC48	LPC60	LPC72
A	15.0%	17.5%	32.5%	15.0%
B	31.9%	34.7%	40.3%	43.1%
C	28.8%	31.5%	27.4%	26.0%
D	14.0%	14.0%	14.0%	18.6%
E	57.9%	52.6%	50.0%	65.8%

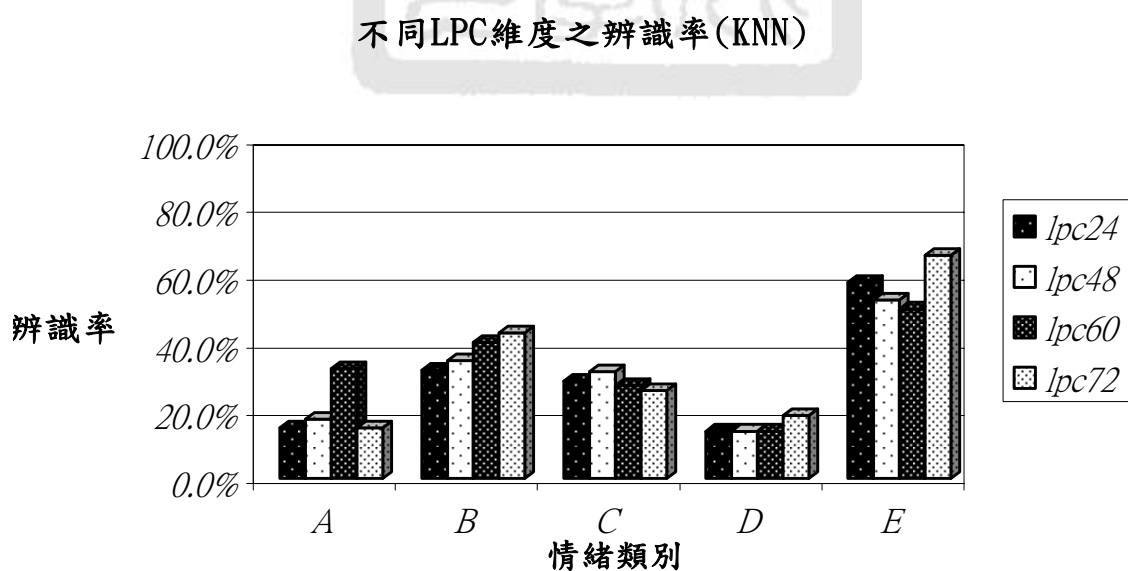


圖 5.7 不同維度的 LPC 辨識率比較(KNNR)

## 討論分析

1. 由表 5.8 與表 5.9 中可發現，在相同維度的線性預估編碼特徵中，ANN 辨識機制的辨識率，高於使用 KNNR 為辨識機制者，兩者差異非常大，KNNR 之辨識率大部分低於 50%，唯有 E 類別之辨識率較高，超過 50%，但 ANN 則皆高於 50%。
2. 從圖 5.6 與圖 5.7 中可發現，隨著維度的增加，其辨識率並無明顯的提升，ANN 辨識機制與 KNNR 辨識機制皆有相同情形，代表維度的增加對辨識率影響不大，觀察五個類別的辨識率，發現在 ANN 模組中，不同類別其辨識率差異不大，約莫在 60%到 70%之間，但 KNNR 模組中，不同的類別，其辨識率有明顯之差異，E 類別辨識率最高，皆高於 50%，D 類別辨識率最低，不超過 20%，D 和 E 類別之間，有 30%的辨識率差異。
3. 比較 LPC24、LPC48、LPC60、LPC72 四種不同維度的辨識率，發現 LPC60 擁有最高的辨識率，因此挑選 LPC60 為線性預估編碼最佳維度特徵。

### 5.2.5 實驗五

以 LPCC 為其特徵，使用 ANN 和 KNNR 這兩種辨識機制，利用 PCA\_type2 方法降低維度，其結果如表 5.10、表 5.11、圖 5.8 和圖 5.9 所示。

## 實驗目的

比較 LPCC 的不同維度，對辨識率之影響，挑選一最佳辨識維度，實驗之維度有 LPCC24、LPCC48、LPCC60、LPCC72、LPCC96 和 LPCC120 六種。

## 實驗結果（ANN 測試）

表 5.10 不同維度的 LPCC 辨識率比較(ANN)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

類別	不同維度的 LPCC 辨識率(ANN)					
	LPCC24	LPCC48	LPCC60	LPCC72	LPCC96	LPCC120
A	7.5%	47.5%	62.5%	62.5%	90.0%	65.0%
B	62.5%	62.5%	63.9%	62.5%	80.6%	69.4%
C	52.1%	65.8%	63.0%	60.3%	82.2%	56.2%
D	69.8%	67.4%	55.8%	67.4%	88.4%	79.1%
E	59.5%	75.7%	70.3%	78.4%	81.6%	86.5%

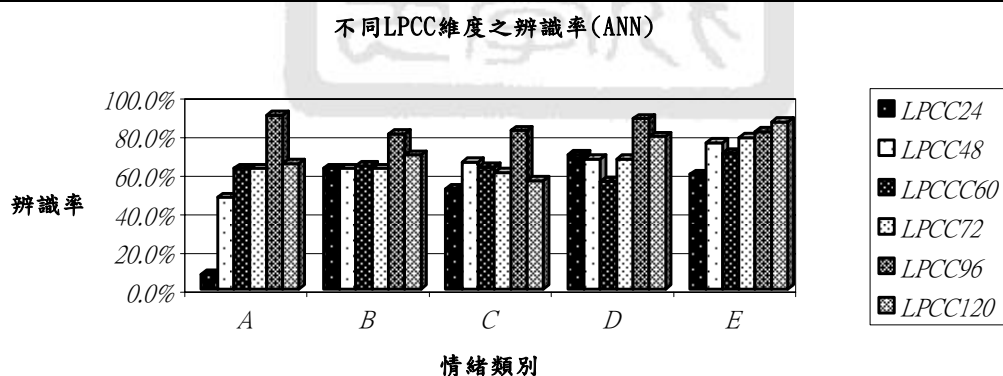


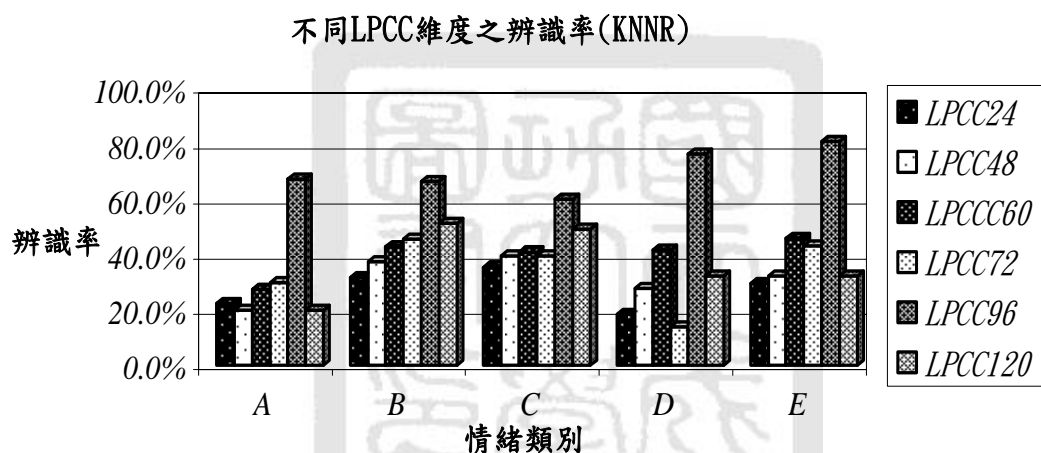
圖 5.8 不同維度的 LPCC 辨識率比較(ANN)

## 實驗結果 (KNNR 測試)

表 5.11 不同維度的 LPCC 辨識率比較(KNNR)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

	不同維度的 LPCC 辨識率(KNNR)					
類別	LPCC24	LPCC48	LPCCC60	LPCC72	LPCC96	LPCC120
A	22.5%	20.0%	27.5%	30.0%	67.5%	20.0%
B	31.9%	37.5%	43.1%	45.8%	66.7%	51.4%
C	35.6%	39.7%	41.1%	39.7%	60.3%	49.3%
D	18.6%	27.9%	41.9%	14.0%	76.7%	32.6%
E	29.7%	32.4%	45.9%	43.2%	81.1%	32.4%



圖

5.9 不同維度的 LPCC 辨識率比較(KNNR)

## 討論分析

- 1.由表 5.10 與表 5.11 中可發現，在相同維度的 LPCC 特徵中，ANN 辨識機制的辨識率高於使用 KNNR 為辨識機制者，兩者差異非常大，KNNR 之辨識率大部分低於 50%，唯有 LPCC96 之辨識率較高，超過 60%。
2. 從圖 5.8 與圖 5.9 中可發現，隨著維度的增加，其辨識率有明顯的提升，ANN 辨識機制與 KNNR 辨識機制皆有相同情形，尤其是當 LPCC72 增

加為 LPCC96，其辨識率的提升最為明顯。

3. 比較 LPCC24、LPCC48、LPCC60、LPCC72、LPCC96、LPCC120 六種不同維度的辨識率，發現 LPCC96 擁有最高的辨識率，因此挑選 LPCC96 為 LPCC 最佳維度特徵。

#### 四種不同特徵的比較

從實驗二至實驗五，比較倒頻譜、梅爾倒頻譜、線性預估編碼和 LPCC 四種特徵，擷取不同維度的特徵參數，比較其辨識率變化，接下來，針對四種不同特徵的最佳維度，對其辨識率加以比較，找出最佳之辨識特徵，可得倒頻譜的最佳維度特徵為 CEP96，梅爾倒頻譜最佳維度特徵為 MFCC72，線性預估編碼最佳維度特徵為 LPC60，LPCC 最佳維度特徵為 LPCC96，其結果如表 5.12、表 5.13、圖 5.10 和圖 5.11 所示。

表 5.12 四種不同特徵最佳維度辨識率比較(ANN)

(A:秤重 B：洗澡 C：飢餓 D：打針 E：酒精)

	四種不同特徵最佳維度辨識率(ANN)			
類別	CEP96	MFCC72	LPC60	LPCC96
A	60.0%	87.5%	72.5%	90.0%
B	66.7%	76.4%	66.7%	80.6%
C	54.8%	75.3%	61.6%	82.2%
D	55.8%	79.1%	69.8%	88.4%
E	63.2%	89.5%	44.7%	81.6%

四種不同特徵最佳維度辨識率(ANN)

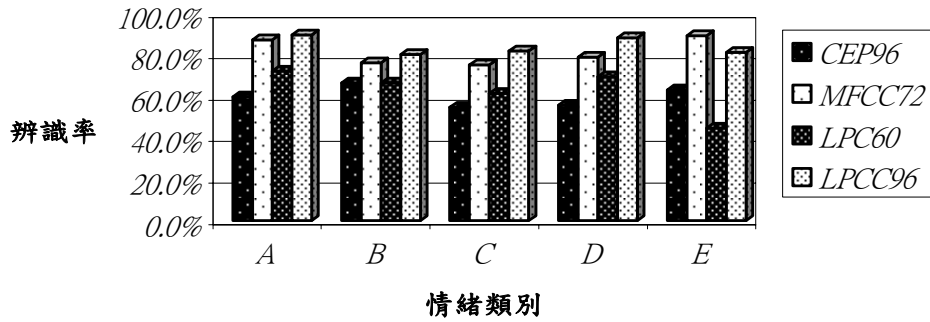


圖 5.10 四種不同特徵最佳維度辨識率比較(ANN)

表 5.13 四種不同特徵最佳維度辨識率比較(KNNR)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

四種不同特徵最佳維度辨識率(KNNR)				
類別	CEP96	MFCC72	LPC60	LPCC96
A	35.0%	60.0%	32.5%	67.5%
B	45.8%	25.0%	40.3%	66.7%
C	43.8%	56.2%	27.4%	60.3%
D	14.0%	86.0%	14.0%	76.7%
E	15.8%	92.1%	50.0%	81.1%

四種不同特徵最佳維度辨識率(KNNR)

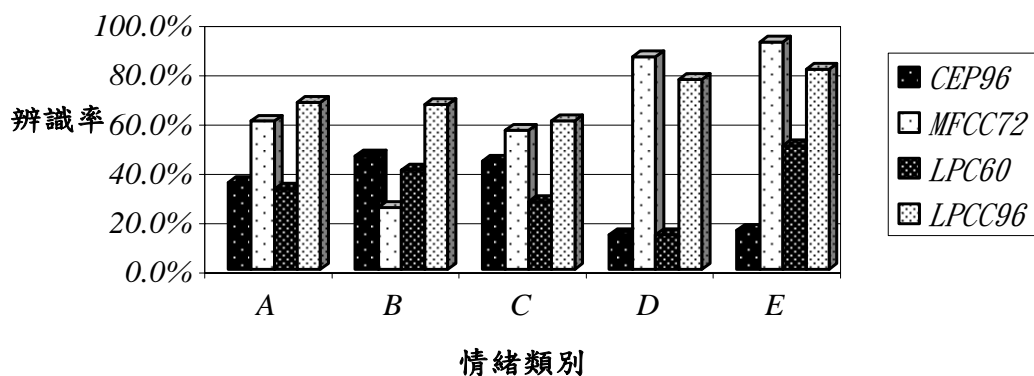


圖 5.11 四種不同特徵最佳維度辨識率比較(KNNR)

## 討論分析

1. 由表 5.12 與表 5.13 中可發現，使用同樣的辨識特徵，ANN 辨識機制的辨識率高於使用 KNNR 為辨識機制者，因此 ANN 辨識機制優於 KNNR 辨識機制。

2. 從圖 5.10 與圖 5.11 中可發現，四種特徵中，以 LPCC96 辨識率最高，MFCC72 次之，在 ABC 類別中，CEP96 辨識率高於 LPC60，在 DE 類別中 CEP96 辨識率低於 LPC60，CEP96 和 LPC60 會隨著不同的類別，其辨識率比較有不同情形，無法斷定 CEP96 和 LPC60 何者較為優秀。

3. 觀察四種特徵之辨識率，發現 LPCC96 和 MFCC72 的辨識率遠高於 CEP96 和 LPC60，約有 10% 至 20% 的差異，因此可斷定 LPCC 和 MFCC 特徵較 CEP 和 LPC 特徵佳，具有鑑別性和較高的辨識率。雖然 LPCC 總體辨識率高於 MFCC 的總體辨識率，但 MFCC 在 DE 類別的辨識上，有較佳的成效，因此在接下來的實驗中，不再考慮 CEP 和 LPC，只針對 LPCC 和 MFCC。

4. CEP 和 MFCC 特徵雖然皆為倒頻譜特徵，但因 MFCC 特徵考慮了人耳的聽覺特性，增加了梅爾濾波器，導致 MFCC 辨識率遠大於 CEP。

5. LPC 和 LPCC 的特徵擷取過程十分相似，但因 LPC 為線性系統，不似 LPCC 除了具有線性特徵之外，也包含了非線性之特徵，故 LPCC 辨識成效優於 LPC。



6. 由總體辨識率上判別，LPCC 為最佳的情緒辨識特徵參數。

#### LPCC 和 MFCC 對五種嬰兒哭聲的辨識情形

觀察 LPCC 和 MFCC 對五種嬰兒哭聲的辨識情形，以瞭解 LPCC 和 MFCC

特徵對五種嬰兒哭聲的辨識成效，及五種情緒哭聲之間的相互關係。

其採用特徵為 LPCC96 和 MFCC72，辨識機制為 ANN 與 KNNR。

表 5.14 至表 5.21 及圖 5.12 至圖 5.15 皆為五種情緒哭聲的訊號輸入，經

辨識機制運算，所得之實際輸出與正確預期輸出之關係，A 類別使用了

40 個測試音檔，B 類別使用了 72 個測試音檔，C 類別使用了 73 個測試

音檔，D 類別使用了 43 個測試音檔，E 類別使用了 38 個測試音檔。

表 5.14 LPCC96 各類別辨識音檔數目比較(個)(ANN)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

		LPCC96 各類別辨識音檔數目 (個)(ANN)				
		辨識類別				
		A	B	C	D	E
測試類別	A(40)	36	2	1	1	0
	B(72)	8	58	5	1	0
	C(73)	4	8	60	1	0
	D(43)	0	0	3	38	2
	E(38)	3	1	0	2	31



表 5.15 LPCC96 各類別辨識音檔數目比較(%) (ANN)

		LPCC96 各類別辨識音檔數目 (%) (ANN)				
		辨識類別				
		A	B	C	D	E
測試類別	A(40)	90.0%	5.0%	2.5%	2.5%	0.0%
	B(72)	11.1%	80.6%	6.9%	1.4%	0.0%
	C(73)	5.5%	11.0%	82.2%	1.4%	0.0%
	D(43)	0.0%	0.0%	7.0%	88.4%	4.7%
	E(38)	7.9%	2.6%	0.0%	5.3%	81.6%

LPCC96各類別辨識數目 (%) (ANN)

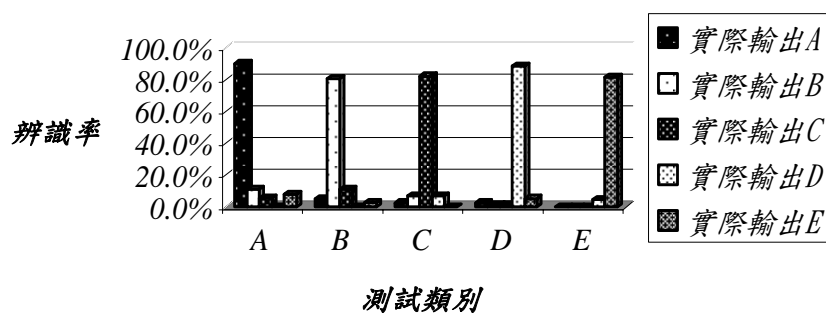


圖 5.12 LPCC96 各類別辨識音檔數目比較(%) (ANN)

表 5.16 LPCC96 各類別辨識音檔數目比較(個) (KNNR)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

		LPCC96 各類別辨識音檔數目 (個) (KNNR)				
		辨識類別				
		A	B	C	D	E
測試類別	A(40)	27	4	5	1	3
	B(72)	11	48	10	1	2
	C(73)	11	12	44	3	3
	D(43)	2	1	2	33	5
	E(38)	0	3	3	1	30

表 5.17 LPCC96 各類別辨識數目比較(%) (KNNR)

		LPCC96 各類別辨識數目 ( % ) ( KNNR )				
		辨識類別				
		A	B	C	D	E
測試類別	A(40)	67.5%	10.0%	12.5%	2.5%	7.5%
	B(72)	15.3%	66.7%	13.9%	1.4%	2.8%
	C(73)	15.1%	16.4%	60.3%	4.1%	4.1%
	D(43)	4.7%	2.3%	4.7%	76.7%	11.6%
	E(38)	0.0%	8.1%	8.1%	2.7%	81.1%

LPCC96各類別辨識數目 ( % ) ( kNNR )

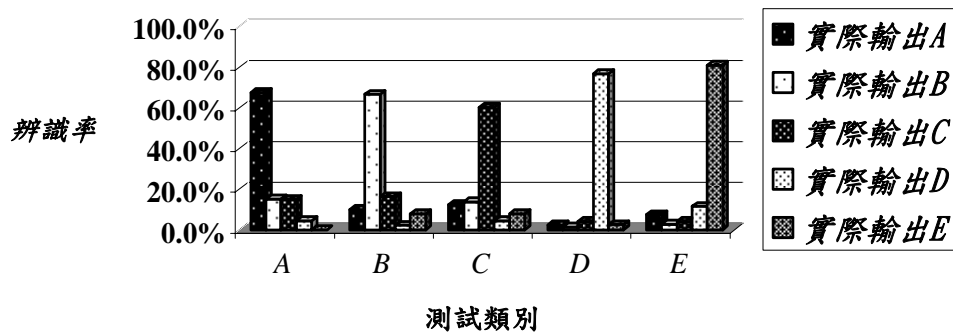


圖 5.13 LPCC96 各類別辨識數目比較(%) (KNNR)

表 5.18 MFCC72 各類別辨識音檔數目比較(個) (ANN)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

		MFCC72 各類別辨識數目 ( 個 ) ( ANN )				
		辨識類別				
		A	B	C	D	E
測試類別	A(40)	35	3	2	0	0
	B(72)	6	55	6	0	5
	C(73)	6	9	55	1	2
	D(43)	3	1	1	34	4
	E(38)	1	0	3	0	34

表 5.19 MFCC72 各類別辨識數目比較(%) (ANN)

		MFCC72 各類別辨識數目 ( % ) ( ANN )				
		辨識類別				
		A	B	C	D	E
測試類別	A(40)	87.5%	7.5%	5.0%	0.0%	0.0%
	B(72)	8.3%	76.4%	8.3%	0.0%	6.9%
	C(73)	8.2%	12.3%	75.3%	1.4%	2.7%
	D(43)	7.0%	2.3%	2.3%	79.1%	9.3%
	E(38)	2.6%	0.0%	7.9%	0.0%	89.5%

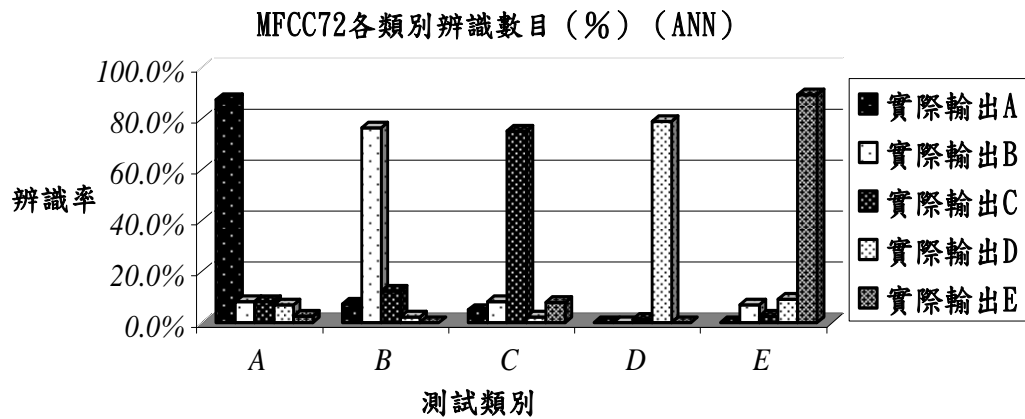


圖 5.14 MFCC72 各類別辨識數目比較(%) (ANN)

表 5.20 MFCC72 各類別辨識音檔數目比較(個) (KNNR)

		MFCC72 各類別辨識音檔數目 ( 個 ) ( KNNR )				
		辨識類別				
		A	B	C	D	E
測試類別	A(40)	24	11	5	0	0
	B(72)	18	18	14	14	8
	C(73)	10	16	41	3	3
	D(43)	2	0	1	37	3
	E(38)	0	0	2	1	35

表 5.21 MFCC72 各類別辨識數目比較(%) (KNNR)

		MFCC72 各類別辨識數目 ( %) ( KNNR )				
		辨識類別				
		A	B	C	D	E
測試類別	A(40)	60.0%	27.5%	12.5%	0.0%	0.0%
	B(72)	25.0%	25.0%	19.4%	19.4%	11.1%
	C(73)	13.7%	21.9%	56.2%	4.1%	4.1%
	D(43)	4.7%	0.0%	2.3%	86.0%	7.0%
	E(38)	0.0%	0.0%	5.3%	2.6%	92.1%

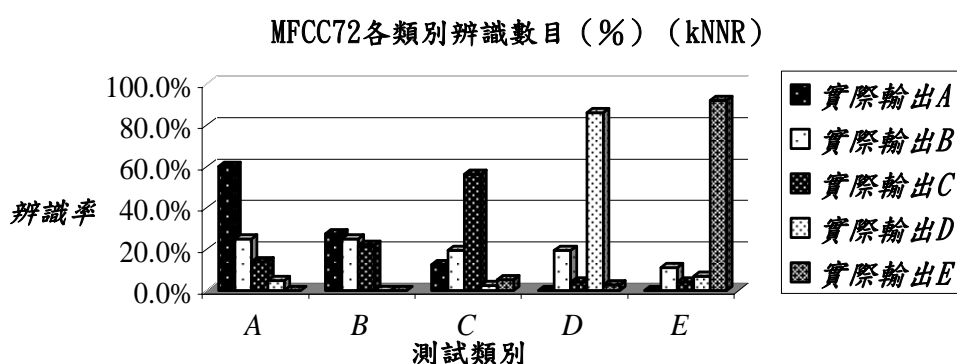


圖 5.15 MFCC72 各類別辨識數目比較(%) (KNNR)

#### 討論分析

1. 從上面的圖表中可發現，ABC 三類別的相似性較高，容易互相辨識錯誤，DE 類別的相似性較低。

2. 從嬰兒哭聲情緒的所發生之原因探討，可發現 A 類別為秤重時，因赤裸造成不安全感，B 類別因洗澡的溫度變化感到不舒服，C 類別為肚子餓，D 類別為打針疼痛，E 類別為酒精消毒臍帶，造成之刺激疼痛。

從這些原因中，可將 ABC 三類別分類於生理的不舒服，DE 類別分類於生理的疼痛，因此 ABC 三類別的相似性較高。

## 5.2.6 實驗六

以 LPCC96 為其特徵，使用 ANN 辨識機制，利用 PCA\_type2 方法降低維度，其結果如表 5.22 和圖 5.16 所示。

### 實驗目的

將 LPCC96 特徵與其他特徵互相組合，期能找出比 LPCC96 特徵更佳之特徵組合參數，使用特徵有 LPCC96 的一維差量係數、基頻 (Pitch) 和對數能量 (Energy)。

### 實驗結果 (ANN 測試)

表 5.22 LPCC96 組合特徵辨識率比較(ANN)

(A:秤重 B:洗澡 C:飢餓 D:打針 E:酒精)

類別	LPCC96 組合特徵辨識率(ANN)			
	LPCC+Energy	LPCC+Pitch	LPCC+LPCC 差量	LPCC
A	60.0%	20.0%	75.0%	90.0%
B	63.9%	59.7%	70.8%	80.6%
C	69.9%	52.1%	58.9%	82.2%
D	60.5%	48.8%	74.4%	88.4%
E	76.3%	18.4%	70.3%	81.6%

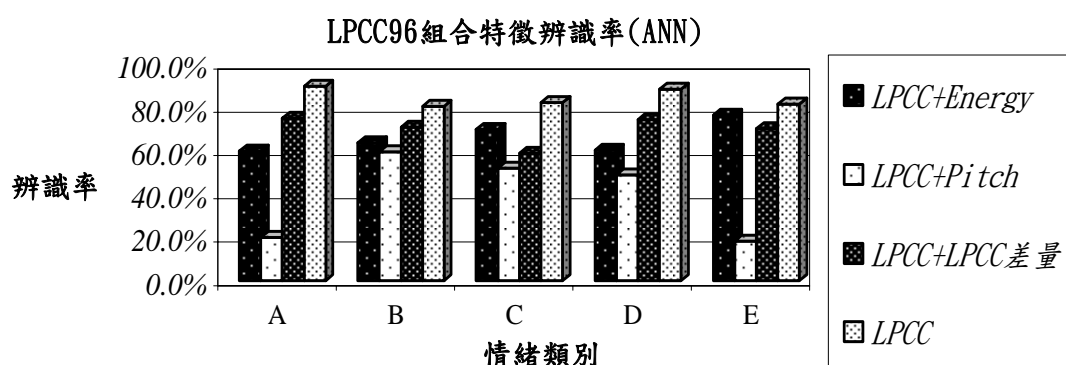


圖 5.16 LPCC96 組合特徵辨識率比較(ANN)

## 討論分析

1.從表 5.22 中得知，單獨 LPCC96 特徵的辨識率為最高，LPCC+Energy 、LPCC+Pitch、LPCC+LPCC 差量這三組特徵組合之辨識率，並未比單獨 LPCC96 的辨識率好，代表 Energy、Pitch、LPCC 差量這三項特徵不具有良好的鑑別性，與 LPCC96 特徵組合，不但未將辨識率提升，反而降低辨識率。

### 5.2.7 實驗七

以 MFCC72 為其特徵，使用 ANN 辨識機制，利用 PCA\_type2 方法降低維度，其結果如表 5.23 和圖 5.17 所示。

## 實驗目的

將 MFCC72 特徵與其他特徵互相組合，期能找出比 MFCC72 特徵更佳之特徵組合參數，使用特徵有 MFCC72 的一維差量係數、基頻（Pitch）和對數能量（Energy）。

## 實驗結果（ANN 測試）

表 5.23 MFCC72 組合特徵辨識率比較(ANN)

(A:秤重 B：洗澡 C：飢餓 D：打針 E：酒精)

MFCC72 組合特徵辨識率(ANN)				
類別	MFCC+Energy	MFCC+Pitch	MFCC+MFCC 差量	MFCC
A	47.5%	7.5%	7.5%	72.5%
B	20.8%	20.8%	20.8%	66.7%
C	56.2%	45.2%	45.2%	61.6%
D	58.1%	20.9%	20.9%	69.8%
E	97.4%	86.8%	86.8%	44.7%

MFCC72組合特徵辨識率(ANN)

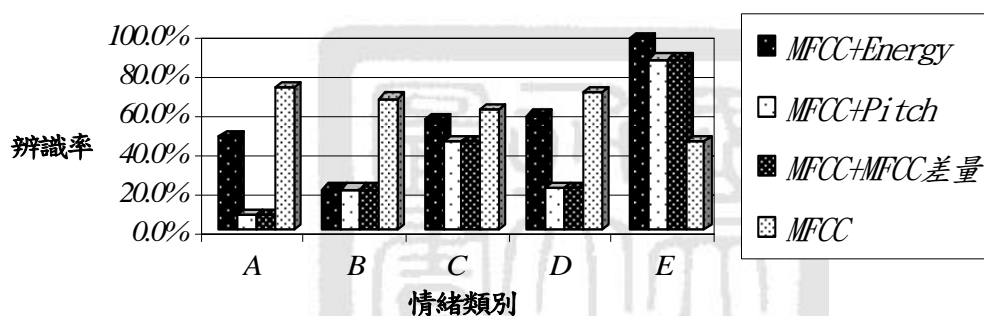


圖 5.17 MFCC72 組合特徵辨識率比較(ANN)

## 討論分析

從表 5.23 中得知，單獨 MFCC72 特徵的辨識率為最高，MFCC+Energy、MFCC+Pitch 和 MFCC+MFCC 差量這三組特徵組合之辨識率，並未比單獨 MFCC72 的辨識率高。

將實驗六和實驗七互相比較，發現其實驗結果相似，利用 Energy、Pitch、特徵一階差量，這三組特徵與原本的特徵 LPCC 或 MFCC 組合之辨識率，並未比單獨特徵的辨識率高，代表 Energy、Pitch、特徵一階差

量這三項特徵不具有良好的鑑別性，Energy 和 Pitch 為音源激發訊號特徵，由此可發現，音源訊號特徵對嬰兒哭聲情緒辨識系統無良好之辨識功效，五種嬰兒情緒哭聲的音源訊號特徵無明顯的差異。

特徵一階差量，代表著特徵係數對時間的變化性，利用特徵係數相對於時間之斜率，觀察特徵係數在時間的動態變化。然而特徵一階差量並非為高鑑別性特徵，代表特徵係數在時間的動態變化，跟嬰兒哭聲情緒辨識系統無明顯的關係存在。

### 5.3 綜合分析與討論

根據實驗一至實驗七的分析與討論，可得到下列的綜合討論結果

- 1.就 MEAN、PCA\_type1、PCA\_type2 和 PCA\_type3 這四種降維方式，以 PCA\_type2 的辨識結果最佳。
2. 比較四種特徵之辨識率，LPCC 和 MFCC 的辨識率優於 CEP 和 LPC，從此結果得知，一個語音特徵訊號需符合人類的聽覺系統，且聽覺系統並非完全的線性系統，故特徵參數也需具有非線性特徵的特性，這樣的特徵參數才能夠具有良好鑑別性和強健性。
3. 音源訊號特徵，對嬰兒哭聲情緒辨識系統，無良好之辨識功效，特徵一階差量代表著特徵係數對時間的變化性，然而特徵係數在時間的動態



變化，跟嬰兒哭聲情緒辨識系統無明顯的關係存在，對辨識率提升沒有任何幫助，在選擇兩個特徵參數互相結合時，必須選擇兩者皆為良好之辨識特徵，否則將會互相影響，導致辨識率下降。



## 第六章

---

### 結論與未來展望

#### 6.1 結論

本論文主要探討不同的情緒辨識特徵參數，對嬰兒哭聲情緒辨識系統之辨識率影響。研究著重在如何選取一較佳特徵參數，增進嬰兒哭聲情緒辨識系統的辨識率。特徵參數的選取，對整個辨識系統的影響為最大，一個好的特徵參數，可以使用簡單的辨識機制如 KNN，就可得到良好的辨識效果，如一開始使用不好的特徵參數，即使使用複雜的辨識機制如 ANN，其辨識率的提升也是有限的。本文比較了前人研究所使用的倒頻譜參數、梅爾倒頻譜、線性預估參數這三種特徵參數，並另外使用 LPCC 參數，此參數鮮少被使用在嬰兒哭聲情緒辨識系統，除了這四種特徵參數外，也比較了音源訊號和差量訊號的特徵參數組合。

實驗中顯示，符合人類聽覺系統的特徵參數才會具有良好鑑別性和強健性，選取良好的特徵參數 LPCC96 為辨識特徵，選用最直觀簡單的 KNN 辨識機制，可得到平均 70% 左右的辨識率，使用複雜的辨識機制 ANN，則五種類別的辨識率皆在 80% 以上，平均辨識率為 83%。

## 6.2 未來展望

由於本研究所探討的是嬰兒情緒辨識系統中，其特徵擷取參數的比較與辨識率的關係，然而嬰兒辨識系統涵蓋之範圍很廣，未來研究可朝下列幾點進行：

- 1.在辨識模型中，本論文只針對嬰兒情緒哭聲作辨識，未來可將其推廣至嬰兒疾病哭聲辨識系統。
- 2.由於嬰兒哭聲訊號非常難以取得和收集，使得訊號資料庫資料太少，導致辨識率不佳，期能發展出良好的辨識系統，只需少量資訊，就可完美辨識。
- 3.由於錄製哭聲訊號時，周遭盡量是保持安靜狀態，但一般生活中，並非時時保持安靜狀態，希能夠克服雜訊的影響，將系統應用在平常的生活環境。

## 參考文獻

- [1] M. Petroni, A. S. Malowany, C. C. Johnston, and B. J. Stevens, “A New, *Robust Vocal Fundamental Frequency (F0) Determination Method for the Analysis of Infant Cries*”, Proceedings of the IEEE Symposium on Computer-Based Medical Systems, Winston-Salem, NC, USA, pp.223-228, 1994.
- [2] M. Petroni, A. S. Malowany, C. C. Johnston, and B. J. Stevens, “A *Robust and Accurate Cross-Correlation-Based Fundamental Frequency (F0) Determination Method for the Improved Analysis of Infant Cries*”, Annual International Conference of the IEEE Engineering in Medicine and Biology – Proceedings, Montreal, Can, vol. 2, pp. 975-976, 1995.
- [3] M. Petroni, A. S. Malowany, C. C. Johnston, and B. J. Stevens, “A *Crosscorrelation-Based Method for Improved Visualization of Infant Cry Vocalizations*”, Canadian Conference on Electrical and Computer Engineering, Halifax, Can, vol. 2, pp. 453-456, 1994.
- [4] H. E. Baeck and M. N. Souza, “*Study of Acoustic Features of Newborn Cries that Correlate with the Context*”, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Kyoto University, vol. 3, pp. 2174-2177, 2001.
- [5] G. Jr. Varallyay, Z. Benyo, A. Illenyi, Z. Farkas and L. Kovacs, “*Acoustic Analysis of The Infant Cry: Classical and New Methods*”, Annual International Conference of the IEEE Engineering in Medicine and Biology – Proceedings, San Francisco, CA, United States, IEEE International Conference, vol. 1, pp. 313-316, 2004.
- [6] Ada Fort and Claudia Manfredi, “*Acoustic Analysis of Newborn Infant Cry Signals*”, Medical Engineering & Physics, vol.20, pp. 432 - 442, 1998.

- [7] A. Fort, A. Ismaelli, C. Manfredi and P. Brusaglioni, “*Parametric and Non-Parametric Estimation of Speech Formants: Application to Infant Cry*”, Medical Engineering & Physics, vol.18, pp. 677-691, 1996.
- [8] T. Harada, A. Saito, T. Sato and T. Mori, “*Infant Behavior Recognition System Based On Pressure Distribution Image*”, Proceedings - IEEE International Conference on Robotics and Automation, San Francisco, CA, USA, vol.4, pp. 4082 - 4088, 2000.
- [9] A. M. Perez, A. Gutierrez, M. Sanchez, J. Remolina and O. Aguilera, “*Determination of A Mathematical Indicator From the Acoustical Analysis of Primal Crying of Newborns to Evaluate Their Well-being*”, Annual International Conference of the IEEE Engineering in Medicine and Biology – Proceedings, Chicago, IL, USA, vol.3, pp. 1073 - 1075, 1997.
- [10] K. Wermke, W. Mende, C. Manfredi and P. Brusaglioni, “*Developmental Aspects of Infant’s Cry Melody and Formants*”, Medical Engineering & Physics, vol.24, pp. 501 - 514, 2002.
- [11] J. O. Garcia and C. A. Reyes Garcia, “*Mel-Frequency Cepstrum Coefficients Extraction from Infant Cry for Classification of Normal and Pathological Infant with Feed-Forward Neural Networks*”, Proceedings of the International Joint Conference on Neural Networks, Portland, vol. 4, pp. 3140-3145, 2003.
- [12] J. O. Garcia and C. A. Reyes Garcia, “*Implementation and Analysis of Training Algorithms for the Classification of Infant Cry with Feed-Forward Neural Networks*”, IEEE International Symposium on Intelligent Signal Processing , pp. 271-276, 2003
- [13] D. Lederman, A. Cohen , E. Zmora , K. Wermke, S. Hauschildt and A. Stellzig-Eisenhauer, “*On the Use of Hidden Markov Models in Infants' Cry Classification*”, The 22nd Convention of Electrical and Electronics Engineers, Israel, pp. 350-352, 2002.

- [14] M. Petroni, A. S. Malowany, C. C. Johnston, and B. J. Stevens, “A *Comparison of Neural Network Architectures for the Classification of Three Types of Infant Cry Vocalizations*”, Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, Montreal, Can, vol. 1, pp. 821-822, 1995.
- [15] M. Petroni, A. S. Malowany, C. C. Johnston, and B. J. Stevens, “*Classification of Infant Cry Vocalizations Using Artificial Neural Networks (ANNs)*”, ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Detroit, MI, USA ,vol. 5, pp. 3475-3478, 1995.
- [16] M. Petroni, A. S. Malowany, C. C. Johnston, and B. J. Stevens, “*On the Use of Artificial Neural Networks (ANNs) for the Classification of Three Types of Infant Cries*”, IEEE Pacific RIM Conference on Communications, Computers, and Signal Processing – Proceedings, Victoria, BC, Can, pp. 501-504, 1995.
- [17] S. E. Barajas-Montiel and C. A. Reyes Garcia, “*Identifying Pain and Hunger in Infant Cry with Classifiers Ensembles*”, Proceedings - International Conference on Computational Intelligence for Modelling, Control and Automation, CIMCA 2005 and International Conference on Intelligent Agents, Web Technologies and Internet, Vienna, Austria, vol. 2, pp. 770-775, 2005.
- [18] P. Pal , A .N. Iyer and R. E. Yantorno, “*Emotion Detection From Infant Facial Expressions and Cries*”, ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings, Toulouse, France, vol.2, pp. II - II, 2006.
- [19] J. R. Deller, J. H. L. Hansen and J. G. Proakis, “*Discrete Time Processing of Speech Signals* ”, Prentice Hall, New Jersey ,1987
- [20] O. Wasz-Hockert, T. Partanen, V. Vuorenkoski, E. Valanne and K. Michelsson, “*The Identification of Some Specific Meanings in Infant Vocalizaton* ”, Experiencia, vol.20, pp. 154-156,1964

- [21] A. Ismaelli, G. Rapisardi, G. Donzelli, R. Moroni and P. Brusaglioni, “*A New Device for Computerized Infant Cry Analysis in The NICU*”, Annual International Conference of the IEEE Engineering in Medicine and Biology – Proceedings, Baltimore, MD, USA, vol. 2, pp. 854-855, 1994.
- [22] X. Qiaobing, R. K. Ward and C. A. Laszlo, “*Determining Normal Infants' Level-of-Distress from Cry Sounds*”, Canadian Conference on Electrical and Computer Engineering, Vancouver, BC, Canada, vol. 2, pp. 1094-1096, 1993.
- [23] C. Manfredi, V. Tocchioni, L. Bocchi and L. Kovacs, “*A Robust Tool for Newborn Infant Cry Analysis*”, Annual International Conference of the IEEE Engineering in Medicine and Biology – Proceedings, New York, NY, United States, pp. 509-512, 2006.
- [24] C. J. Clement, F. J. Koopmans-van Beinum and L. C. W. Pols, “*Acoustical Characteristics of Sound Production of Deaf and Normally Hearing Infants*”, International Conference on Spoken Language Processing, ICSLP, Proceedings, Philadelphia, PA, USA, vol.3, pp. 1073 - 1075, 1997.
- [25] 王小川，語音訊號處理，全華科技圖書公司，民國94年。
- [26] 張柏雄著，“中文語音情緒之自動辨識”，國立成功大學碩士論文，民國90年。
- [27] 陳雅菁著，“類神經網路在兒童構音異常診斷上之應用”，國立成功大學碩士論文，民國84年。
- [28] 謝依蘭，語音訊號數位處理，松岡圖書公司，民國81年。

[29]魏渭堂，彩色圖說嬰幼兒發展保育學—學齡前兒童的照顧與發展，  
合計圖書出版社，民國95年。

[30]盧素碧，嬰幼兒保育，文景出版社，民國77年。

[31]Jyh-Shing Roger Jang，線上中文教材：音訊處理與辨識

Home page : <http://www.cs.nthu.edu.tw/~jang>.

[32]Jyh-Shing Roger Jang，線中上文教材：資料群聚與樣式辨識

Home page : <http://www.cs.nthu.edu.tw/~jang>

[33]張銘豐著，“以時頻域特徵參數為主之嬰兒哭聲辨識系統研製”，逢  
甲大學碩士論文,民國93年。

