

國立中央大學

數學研究所

碩士論文

神經元電位訊號雜訊  
之組成分析

An Analysis of the Components of Noise in the  
Recording of Neuronal Potential Signals

研究生：吳蕙稜

指導教授：單維彰博士

中華民國九十七年六月



# 國立中央大學圖書館 碩博士論文電子檔授權書

(95年7月最新修正版)

本授權書所授權之論文全文電子檔(不包含紙本、詳備註1說明)，為本人於國立中央大學，撰寫之碩/博士學位論文。(以下請擇一勾選)

- (  )同意(立即開放)  
(  )同意(一年後開放)，原因是：\_\_\_\_\_。  
(  )同意(二年後開放)，原因是：\_\_\_\_\_。  
(  )不同意，原因是：\_\_\_\_\_。

以非專屬、無償授權國立中央大學圖書館與國家圖書館，基於推動「資源共享、互惠合作」之理念，於回饋社會與學術研究之目的，得不限地域、時間與次數，以紙本、微縮、光碟及其它各種方法將上列論文收錄、重製、公開陳列、與發行，或再授權他人以各種方法重製與利用，並得將數位化之上列論文與論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

研究生簽名： 吳蕙稜 學號：952201010

論文名稱：神經元電位訊號雜訊之組成分析

指導教授姓名：單維彰 博士

系所：數學所 博士班 碩士班

日期：民國97年6月11日

備註：

1. 本授權書之授權範圍僅限電子檔，紙本論文部分依著作權法第15條第3款之規定，採推定原則即預設同意圖書館得公開上架閱覽，如您有申請專利或投稿等考量，不同意紙本上架陳列，須另行加填聲明書，詳細說明與紙本聲明書請至 <http://blog.lib.ncu.edu.tw/plog/> 碩博士論文專區查閱下載。
2. 本授權書請填寫並親筆簽名後，裝訂於各紙本論文封面後之次頁(全文電子檔內之授權書簽名，可用電腦打字代替)。
3. 請加印一份單張之授權書，填寫並親筆簽名後，於辦理離校時交圖書館(以統一代轉寄給國家圖書館)。
4. 讀者基於個人非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。



國立中央大學碩士班研究生

論文指導教授推薦書

數 學 學系/研究所 吳蕙稜 研究生

所提之論文

神經元電位訊號雜訊之組成分析

係由本人指導撰述，同意提付審查。

指導教授 吳 唯 彭 (簽章)

97 年 5 月 21 日



國立中央大學碩士班研究生  
論文口試委員審定書

數學 學系/研究所 吳蕙稜 研究生

所提之論文

神經元電位訊號雜訊之組成分析

經本委員會審議，認定符合碩士資格標準。

學位考試委員會召集人

蔣孟利

委

員

嚴健彰

單維彰

中華民國 97 年 6 月 2 日



# 摘要

本論文探討雜訊的性質，是否符合一般分類演算法的基本假設「正規分布」，並且從此筆雜訊中擷取的雜訊，是否符合獨立。若雜訊較符合正規分布，則我們使用的分類方法較有效能。我們可經由雜訊直方圖的觀察，探討其正規性。並從雜訊的直方圖中發現異常突起的特性，且此異常突起大比例均勻散佈於雜訊中。由其他實驗雜訊的比較，發現異常的雜訊可能為不同儀器設備所致，或是由於電極接收到其他微弱訊號。

觀察雜訊在頻率域上的表現，發現頻譜上的異常突起，就是造成直方圖異常突起的原因。另外，我們發現雜訊在頻率域上的能量呈現穩定起伏，於是本論文提出一種新的濾波方式，使得雜訊更符合正規分布，獨立性也更好。另一方面，在真實的神經訊號分類中，利用雜訊在頻率域上能量的穩定起伏，對神經訊號做濾波，發現濾波後會使得分類程式對於主成份分析散佈圖有部分重疊的情形分類較佳。

關鍵字：雜訊、正規分布、動作電位、主成份分析

# Abstract

This paper probes into property of noise and discusses whether its distribution is the 'normal distribution' which is generally the basic assumption of algorithm of classification. And getting noise from it, whether its distribution is independent and identically distributed. If the distribution of noise is normal, the classifications are relatively efficient. In addition, by observing the histogram of noise and its normalization model, we find abnormality, uniformly scattering with large percentage near the center. By comparing with noise of other experiments, we conclude that abnormal noise in the histogram may be resulted from the instruments or just the other weak signals received by the electrode .

When observing the representation of noise in the frequency domain, we find irregularity in the frequency domain which causes abnormality in the histogram in the time domain. Moreover, we find the power of noise shows steady wave in the frequency domain. Accordingly, we propose a kind of new filtering method, to make the distribution of noise conform to the normal distribution and make it more independent than the original one. On the other hand, in real nerve signals, the classification will be improved by a filtering based on the fact that the power of noise shows steady wave in the frequency domain. After filtering, the classification is better than that by PCA scattering plot.

Keywords: noise, normal distribution, action potential, PCA

# 目錄

中文摘要 . . . . .	i
英文摘要 . . . . .	ii
目錄 . . . . .	iii
圖目錄 . . . . .	v
表目錄 . . . . .	vii
<b>1 緒論</b>	<b>1</b>
1.1 背景知識 . . . . .	1
1.2 實驗訊號的製作 . . . . .	5
1.3 研究動機 . . . . .	6
<b>2 理論與方法</b>	<b>7</b>
2.1 取得雜訊的程序 . . . . .	7
2.2 正規分布的探討 . . . . .	7
2.3 正規分布的白雜訊 . . . . .	11
2.4 獨立且來自相同分布 . . . . .	12
<b>3 異常分布的探討</b>	<b>16</b>
3.1 異常訊號的均勻散佈 . . . . .	16
3.2 異常訊號的探討 . . . . .	19
3.2.1 其他實驗雜訊的比較 . . . . .	19
3.2.2 四捨五入的影響 . . . . .	20
3.2.3 問題與討論 . . . . .	22
3.3 異常訊號在頻率域上的探討 . . . . .	26
3.3.1 濾波對訊號分布的影響 . . . . .	26
3.3.2 逐次濾波對訊號分布的影響 . . . . .	33

3.3.3 濾波對 iid 的影響 . . . . .	34
3.4 異常訊號的濾波方法 . . . . .	36
<b>4 雜訊對神經訊號分類的影響</b>	<b>45</b>
4.1 主成份分析 . . . . .	45
4.1.1 最佳投影向量 . . . . .	45
4.1.2 PCA 簡介 . . . . .	47
4.1.3 PCA 分類動作電位 . . . . .	48
4.2 K-Means 分群法 . . . . .	54
4.2.1 K-Means 簡介 . . . . .	54
4.2.2 K-Means 的數值驗證 . . . . .	57
4.2.3 應用 K-Means 於 PCA . . . . .	61
4.3 神經訊號分類 . . . . .	63
4.3.1 濾波前後對分類演算法的比較 . . . . .	63
4.3.2 雜訊三的低頻探討 . . . . .	71
4.4 實際應用的限制 . . . . .	74
 參考文獻	 <b>75</b>
 附錄	 <b>76</b>



## 圖目錄

1	神經訊號	4
2	動作電位原始波形	5
3	訊號一，真實的神經訊號波形	10
4	雜訊一的直方圖及常態分布曲線	10
5	隨機雜訊取法示意圖	10
6	matlab 模擬高斯白雜訊	11
7	四種等高線圖形	13
8	雜訊一於兩種區段內的直方圖	18
9	標記分析圖	18
10	雜訊一的均勻散佈圖	18
11	雜訊二與訊號三解析度四捨五入至小數第三位的直方圖變化	21
12	訊號三放大後直方分布圖以及正規分布模型	25
13	(I)雜訊一 (II)雜訊二 (III)雜訊三 的頻譜圖	25
14	雜訊一、二的濾波分析	30
15	雜訊三、四的濾波分析	31
16	雜訊五、六的濾波分析	32
17	雜訊一在二維空間的函數圖形與等高線圖	35
18	雜訊一中八段 0.1 秒的雜訊頻譜圖	40
19	雜訊一中八段 1 秒的雜訊頻譜圖	41
20	雜訊一的濾波準則曲線	42
21	曲線擬合的數據	42
22	濾波前後的正規模型	42
23	雜訊二中八段 1 秒的雜訊頻譜圖	43
24	雜訊六中八段 1 秒的雜訊頻譜圖	44

25	最佳投影向量的數值驗證 . . . . .	47
26	spike1 和 spike2 PCA 分析動作電位 . . . . .	51
27	spike1 和 spike3 PCA 分析動作電位 . . . . .	51
28	spike3 和 spike4 PCA 分析動作電位 . . . . .	52
29	spike2 和 spike4 PCA 分析動作電位 . . . . .	52
30	動作電位平移的 PCA 分析動作電位 . . . . .	53
31	K-Means 中心迭代圖 . . . . .	56
32	離群值 . . . . .	56
33	K-Means 的散布圖分析 I . . . . .	58
34	K-Means 的散布圖分析 II . . . . .	59
35	$\sigma$ 與 $2\sigma$ 中心距的收斂情形 . . . . .	60
36	應用 K-Means 於 PCA 的範例 . . . . .	62
37	神經訊號於頻率域上的表現 . . . . .	68
38	動作電位的原始波形 . . . . .	68
39	神經訊號於不同 $\alpha$ 值的濾波曲線 b . . . . .	69
40	分類錯誤率變化直方圖 . . . . .	69
41	spike 的頻譜圖 . . . . .	70
42	雜訊三與其模擬神經訊號的頻譜分析 . . . . .	73
43	雜訊一與其模擬神經訊號的頻譜分析 . . . . .	73
44	六筆真實的神經訊號 . . . . .	77

# 表目錄

1	隨機雜訊的規律 . . . . .	8
2	雜訊一的隨機雜訊之相對平均距離 . . . . .	9
3	共變異數對等高線圖形分析 . . . . .	13
4	matlab 內建隨機訊號的共變異數絕對最大值與平均值 . . . . .	14
5	雜訊一、二、三的濾波分析 . . . . .	27
6	雜訊四、五、六的濾波分析 . . . . .	29
7	逐次濾波後的相對誤差 . . . . .	33
8	濾波分析比較 . . . . .	38
9	第一組成份及第二組成份的累積變異與變異差距百分比 . . . . .	50
10	K-Means 的數值分析 . . . . .	57
11	K-Means 的平均分類錯誤率 . . . . .	60
12	應用 K-Means 於 PCA 點圖的分類錯誤率 . . . . .	61
13	雜訊一濾波前後分類錯誤率 . . . . .	65
14	雜訊三濾波前後分類錯誤率 . . . . .	66
15	雜訊濾波前後分類錯誤率 . . . . .	72
16	訊號的檔案來源 . . . . .	76
17	雜訊的檔案來源 . . . . .	76



# 第 1 章 緒論

此章節介紹本論文所需具備的背景知識，以及實驗訊號的製作方式，最後探討此論文的研究動機。

## 1.1 背景知識

神經系統是藉由產生及傳遞電化學衝動的神經元 (*Neurons*)，以及可幫助神經元功能的支持細胞所組成的，包含了神經元和他們之間的連結，大致上可以分成中央神經系統 (*CNS : Central Nerve System*) 和周邊神經系統 (*PNS : Peripheral Nerve System*)。中央神經系統包含了大腦和脊椎神經，其他地方就是周邊神經系統。神經系統含兩種主要的細胞—神經元及支持細胞。神經元是神經系統的基本構造及功能單位，可傳遞電化學衝動，及釋放特殊化學控制因子；藉由這些活動，神經元有感覺、學習、記憶及控制肌肉和腺體的作用。

雖然各種神經元的大小和形狀不同，但通常有三種主要部分：細胞體、樹突以及軸突。細胞體 (*cell body*) 為神經元膨大的部分，樹突 (*dendrite*) 為細胞體向外延伸出去像樹枝形狀的細小突起，負責接收訊號，為電衝動的接受區並將所接受的電衝動傳至細胞體。軸突 (*Axon*) 是較長的突起，負責把訊號傳遞出去。兩個神經元的軸突和樹突之間，會有神經元突觸 (*synapse*)，通常為神經元軸突連接到另一個神經元的樹突。「神經細胞 A」軸突末端的突觸，與「神經細胞 B」樹突前端的突觸相連結；兩個神經細胞藉由突觸彼此傳遞訊息，「神經細胞 A」在軸突末端也就是突觸前 (*presynaptic*) 細胞藉由釋放的神經傳導物質 (*neurotransmitters*) 把訊息傳出去，以刺激在「神經細胞 B」樹突前端的突觸後 (*postsynaptic*) 細胞接收訊息。

當神經元將訊息從細胞本體發送到軸突之時，膜電位之差異稱之為動作電位 (*Action potential*)，可將訊號從某個神經細胞傳給下一個神經細胞，達成傳

遞訊息的任務。靜止膜電位表示神經元在靜止狀態時膜電位的差異，神經元平均為  $-70\text{ mV}$ ；動作電位則是處於靜止膜電位狀態的細胞膜受到適當刺激而產生的特殊波形。當神經元處於靜止狀態時，細胞膜內側比細胞膜外側所帶負電荷較多，此電位差主要為膜的通透性所造成的。

當軸突的細胞膜被刺激而去極化至臨界值時，鈉離子通道打開，增加膜對鈉離子的通透性，造成細胞外鈉離子流到細胞膜內，導致細胞膜內比起細胞膜外有著比較高的電位，所以電位上升，這種電位變化稱為去極化。後來因為大量的鈉離子擴散且快速去極化至  $0$  甚至到  $35\text{ mV}$ ，使得細胞內相較於細胞外帶正電。然後當鉀離子通道打開，造成細胞內的鉀離子流到細胞膜外面，讓細胞膜的電位下降，同時鈉離子的通透性大大減低，使得電位下降到  $-80\text{ mV}$ ，這種電位變化則稱為再極化。直到鉀離子通道關閉並且回歸靜止膜電位。由於鈉離子及鉀離子擴散的改變所造成的膜電位變化，即為動作電位。動作電位可以沿著細胞膜一直傳到到細胞的末端，再透過突觸將刺激傳給下一個神經元。

當突觸後細胞接收到神經傳導物質，又會打開特定的離子通道。鈉離子通道打開後，又依照剛才提到的動作電位形成的原理，使得細胞電位從靜態電位變成正的電位，細胞電位超過一個閾值又可以造成動作電位，讓神經傳遞的訊息繼續傳遞到下一個細胞。動作電位的產生遵循全有全無律 (*All or None Law*)，由於某些刺激所產生去極化電流，去極化小於臨界值時，受電位控制的門檻是關閉的，但去極化達臨界值時稱之為「閾值」，便產生最大的電位變化(即動作電位)，在示波器上看起來只是一個尖峰，所以動作電位有時稱為尖峰電位 (*spike potentials*)。假如神經元沒有達到閾值，動作電位就無從產生。當達到閾值之時，一個固定大小的動作電位始得以產生。某一種特定神經元的動作電位大小都是固定的。在神經細胞中不會有或大的或小的動作電位，即所有的動作電位都是一樣大小。因此，神經元要不就是沒有達到閾值，要不就是激發一個完整的動作電位，這是所謂的「全有全無定律」。假設一個強度固定的刺激持續地作用，使得產生固定頻率的動作電位。假如刺激強度增加，動作電位

產生的頻率也因此增加。但動作電位的產生頻率有上限，所以當刺激大到某種程度，動作電位產生的頻率不會隨著刺激強度的增加而增加。

在大腦裡，大多數的神經元透過動作電位來傳遞訊息，這些神經活動可以用微電極偵測，並且以電訊號的形式紀錄下來。由於動作電位是膜電位的變化所造成的，所以我們可以在膜外紀錄電位的變化。雖然神經元產生動作電位時，微電極可以探測到細胞膜外的電位變化，隨著微電極與神經元的相對位置不同，微電極探測到的同一個神經元動作電位波形也不一樣，並且一根微電極通常會探測到多個神經元發出的動作電位，所以我們需要分類動作電位。以下為了描述方便，當我們說神經元所發出的動作電位，指的是神經元發出動作電位時，微電極所收到的電位變化。

隨著研究目的不同，需要區分不同神經元所發出的動作電位。這個做法來自兩個假設。第一個假設，在短時間內，我們紀錄到每一根神經的動作電位的波形是不一樣的；也就是說，一個波形可以代表一根神經訊號。第二個假設，同一根神經訊號所發出的動作電位波形不會隨著時間改變。由於微電極所探測到的動作電位波形，會隨著電極與神經元的距離和相對位置而有所不同；也就是說即使同一個神經元，只要與微電極的相對位置發生變化，還是會記錄到不同的波形。所以第一個假設總是會成立，但第二個假設卻不一定成立；有些神經元發出的動作電位波形，會隨著時間而改變。在這種狀況下，我們很難將神經元的動作電位分類，甚至可能將同一個神經元的動作電位誤判為兩個以上神經元的動作電位。在我們的研究中，假設訊號是在短時間內記錄的，因此維持第二個假設。

在分類動作電位時，我們還要考慮雜訊的干擾。雜訊的干擾有兩種，一種是環境雜訊，另一種為生物體內各種活動的干擾。由於神經細胞產生的電位相當微弱，因此對環境雜訊的干擾會較敏感。環境雜訊大致來源有三，第一為外在電磁波干擾，強度並不固定，藉由濾波可減弱一部分干擾；第二為熱運動所產生；第三則由接地的不穩定所產生，也就是紀錄設備所產生的雜訊，這不



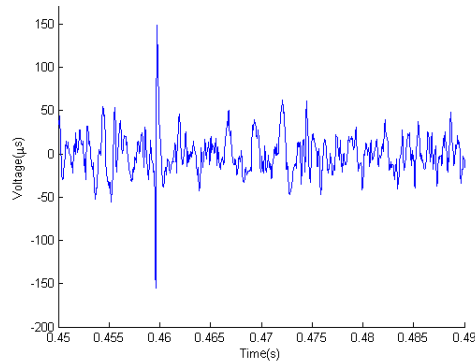


圖 1: 神經訊號

一定會出現。當雜訊較大時，分類就會比較困難。生物體內的干擾主要是神經元彼此的干擾：通常微電極會探測到多個神經元的動作電位，如果有多個神經元同時產生動作電位，則它們重疊的可能會比較高，妨礙彼此之間的分類。另外，生物體內心臟、肌肉所產生的電訊號也會造成一部分干擾。

神經元的動作電位通常產生一個特別的波形，不同神經元的動作電位在被微電極探測到時，其特徵也就會有差異。造成這差異的主要原因是微電極與神經元的相對距離。較靠近微電極的神經元，探測到的動作電位振幅會比較大。測量神經元的一種簡單的方法為 Thresholding 探測，是一種將小動作電位當成雜訊，只分類大動作電位的方法。首先，決定一個電壓門檻作為這個波形是否該視為動作電位的波形的門檻，低於門檻即視為雜訊。如圖 1 所示，如果認為振幅超過  $100 \mu\text{V}$  才是動作電位，圖中只有 0.46 s 附近才有一個動作電位，剩下的都被視為雜訊，即使其中可能包含了兩個以上神經元的動作電位。門檻的大小決定了我們要處理的動作電位數目，如果門檻設定的比較高，預處理的動作電位就會比較少，相對的就會遺失許多動作電位，減少有利的分析資訊；如果門檻設定的比較低，預處理的動作電位就會比較多，分類也會變的比較困難。我們必須觀察探測到的電訊號，根據經驗決定門檻大小。通常動作電位從發生到結束不超過 1.6 ms，因此我們取出來的波形應該比這時間短，如果發現波形持續超過 1.6 ms，那這個波形可能是數個動作電位重疊的結果。

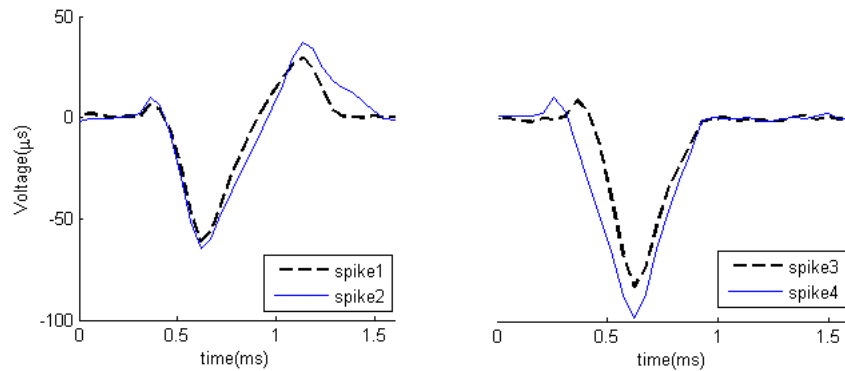


圖 2: 動作電位原始波形

## 1.2 實驗訊號的製作

本論文的訊號並非真實的訊號，而是由人工的動作電位加上真實的雜訊，也就是由實驗訊號中取得的雜訊，真正的神經雜訊很難用簡單的數學模型描述。本論文的實驗雜訊由國立宜蘭大學生物機電學系蔡孟利教授提供，取得的方式應用到 thresholding 探測，首先用此法找到大動作電位，將這些大動作電位拿掉，剩下的就是雜訊了，在本論文第二章會詳細敘述。雖然利用人工訊號得到的結果比不上真實訊號來得有說服力，但人工訊號較容易控制，使得我們可以避開一些問題。例如，我們可以控制動作電位的波形，使得上述的兩個假設一定會成立。使用人工訊號的另一個好處是我們可以預先得知正確的分類結果，使得我們能夠驗證本論文的分類法是否可靠。

圖 2 為本論文模擬的動作電位波形。為了實驗分析方便，此四個波形為兩對相似的波形，通常往上的振幅不超過往下的振幅。除了形狀外，波形的持續時間也有限制，1.1 有提到，通常一個動作電位的發生不超過 1.6 ms，每張圖的橫軸都是 32 個樣本點。人工的動作電位製造出來後，加上真實的雜訊，就是本文所採用的實驗訊號。而實驗中所採用的人工動作電位最大振幅約從 80  $\mu\text{V}$  到 120  $\mu\text{V}$ 。

### 1.3 研究動機

由於很多分類方法都假設雜訊的性質，最普通的假設就是正規分布，要測試一個分類方法有沒有效果，首先要觀察所使用雜訊有多符合這個假設，所以我們開始研究使用的雜訊是否符合正規分布，因此發現了訊號直方圖的異常突起，也就是有一微弱訊號大比例分布於此，我們想要探討這是什麼？結果發現可能為機器有問題，或者是接收訊號時可能接收到某些特定的低頻，這些都是事前無法預測，但是都可以在訓練雜訊的時期發現，我們的論文就是在探討如何處理訊號使得符合演算法上的數學假設，而符合假設與否，對分類的演算法有何影響？



## 第 2 章 理論與方法

此章節我們探討本論文所使用的雜訊是否符合一般的統計假設：獨立且來自相同分布 ( independent and identically distributed ; iid ) ; 首先說明雜訊取得的步驟，接著討論此雜訊是否符合正規分布，並且從中擷取的雜訊是否也來自相同分布，然後探討何謂高斯白雜訊，最後討論擷取出的雜訊是否獨立，則我們就能夠說明這筆雜訊符合 iid 。

### 2.1 取得雜訊的程序

此節探討由實驗中截取出真實神經訊號「訊號一」人工擷取雜訊的方式：取得的雜訊的方式應用到 thresholding 探測，利用此法找到大動作電位，將這些大動作電位拿掉，剩下的就是雜訊了；如圖 3 為訊號一其中的四段神經訊號波形。我們將取一段沒有較大動作電位的雜訊，如圖 3 中曲線標記所示，依此步驟，可以擷取出更多的雜訊，作為實驗所使用的「雜訊一」。本論文主要使用訊號為雜訊一，訊號一及其他雜訊訊號請見附錄，於第三章分析時將說明。

### 2.2 正規分布的探討

這一節我們針對雜訊一做直方圖，測試此雜訊是否符合常態分布模型。製作直方圖的方式為：將雜訊一所對應的振幅範圍等分 (在這裡切割成 300 等份)，記錄雜訊有多少個樣本點屬於此等份，以及每一等份對應的百分比值，也就是如圖 4 所示。橫軸為雜訊一所對應值域的等份，在  $[-1.356, 1.298]$  中等分成 300 等份， $-1.356$  為雜訊的最小值， $1.298$  為雜訊的最大值；縱軸為每一等份所佔的百分比。雜訊一的期望值約為  $-0.0032$ 、標準差約為  $0.2516$ ，我們可以繪製常態分布曲線，如圖 4 中的曲線所標示。

本論文中的實驗爲了模擬真實的神經訊號，並不會重複使用雜訊，而且不同筆雜訊均來自相同分布並且都遵從正規分布。所以嘗試從雜訊一中隨機找 C 組有次序並且不重複的雜訊，驗證是否均來自相同分布，並且遵從正規分布。隨機有次序的雜訊找法爲：以連續 k 點爲一筆資料，新的 C 組雜訊將有 n 筆連續 k 點的資料(分別爲 noise 1 至 noise C，每組樣本點爲  $N=k \times n$ )；當然，我們找到的 C 組雜訊是不重複選取的，由於雜訊一的樣本點爲 302155，所以限定  $N \times C \leq 302155$ ，否則就必需使用更多的雜訊。

本論文在模擬實驗中使用了四種 spike，並且設定一個動作電位的發生爲 32 個樣本點，即將分類的訊號爲四組 1000 筆疊加了雜訊的動作電位，所以設定  $C=4$ 、 $n=1000$ 、 $k=32$ ，所以新的四組雜訊，分別爲 noise 1 至 noise 4 將有 1000 筆連續 32 個樣本點的資料。

現在從雜訊一中隨機產生一個數，以此數作爲新雜訊的起始點，接下來有次序的取每一筆雜訊，不足的回原雜訊起始點開始取，也就是將雜訊一重排，目的爲隨機並且有次序的找四組的雜訊。如圖 5 示意圖，程序如下：我們找這四筆雜訊的方法是將重排後的雜訊一均分爲 1000 等份，每一等份爲 302 個樣本點(不足的不計一等份)，再將每等份 302 個點均分成 4 等份，每一等份爲 75 個樣本點(同樣的，不足的不計一等份)，分別以每 1000 筆資料的 4 等份，擷取每等份 75 點的前 32 個連續的樣本點，一共有四組 1000 筆連續 32 點的資料；如表 1 所示，無論我們需要多少組 32 點的 1000 筆資料，都可以由此方式找出，但前提必須符合  $N=k \times n$ ， $N \times C \leq 302155$ ，其中， $C=4$ 、 $n=1000$ 、 $k=32$ 。

第m筆資料	
noise1	$302 \times (m-1)+0 \times 75 +1 \sim 302 \times (m-1)+0 \times 75 +32$
noise2	$302 \times (m-1)+1 \times 75 +1 \sim 302 \times (m-1)+1 \times 75 +32$
noise3	$302 \times (m-1)+2 \times 75 +1 \sim 302 \times (m-1)+2 \times 75 +32$
noise4	$302 \times (m-1)+3 \times 75 +1 \sim 302 \times (m-1)+3 \times 75 +32$
noiseC	$302 \times (m-1)+(C-1) \times 75 +1 \sim 302 \times (m-1)+(C-1) \times 75 +32$

表 1: 隨機雜訊的規律

接下來驗證上述程序找出來的四段雜訊是否來自相同分布並且遵從正規分布。分別造了四次四組的雜訊，找出每一次取出來四段雜訊的期望值、標準差與雜訊一期望值、標準差的相對平均距離；期望值(標準差)的相對平均距離為新造四組雜訊的期望值(標準差)分別與雜訊一的期望值(標準差)距離取平均。觀察期望值與標準差的偏移情形，偏移的越小代表找出的新雜訊與雜訊一的分布越接近。如下表 2，可以發現標準差相對誤差最多 3%，期望值偏移不超過  $\pm 0.003$ ，雖然我們得到的新雜訊無法完全來自相同分布，但是可靠度已經相當高了，因為真實的神經訊號雜訊無法 100% 來自完全相同的分布。所以可以假設我們隨機有次序找到的雜訊來自相同分布，並且遵從正規分布。

		第一組	第二組	第三組	第四組	相對平均距離
第一次	期望值	-0.0007	-0.0033	-0.0031	-0.0038	0.0008
	標準差	0.2513	0.2497	0.2493	0.2547	0.0019
第二次	期望值	-0.0040	-0.0019	-0.0022	-0.0033	0.0008
	標準差	0.2519	0.2518	0.2501	0.2499	0.0009
第三次	期望值	-0.0026	-0.0019	-0.0047	-0.0029	0.0009
	標準差	0.2528	0.2504	0.2510	0.2500	0.0012
第四次	期望值	-0.0031	-0.0030	-0.0041	0.0024	0.0017
	標準差	0.2525	0.2580	0.2517	0.2509	0.0020

表 2: 雜訊一的隨機雜訊之相對平均距離

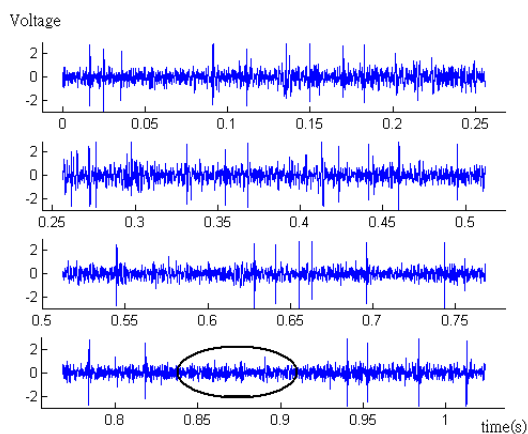


圖 3: 訊號一，真實的神經訊號波形

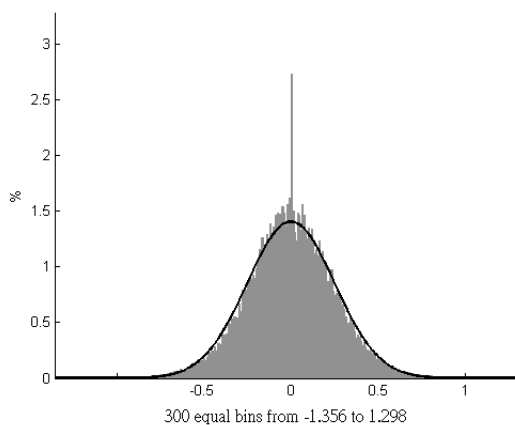
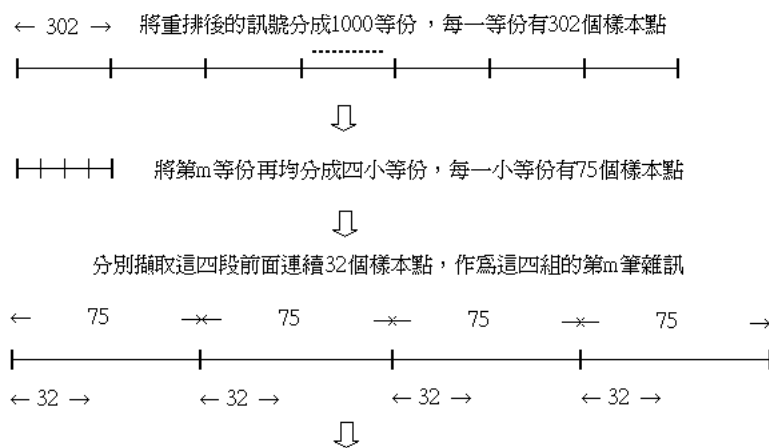


圖 4: 雜訊一的直方圖及常態分布曲線



這1000等份皆可用此方法，由訊號中取出四組1000筆連續32個樣本點的雜訊。

圖 5: 隨機雜訊取法示意圖

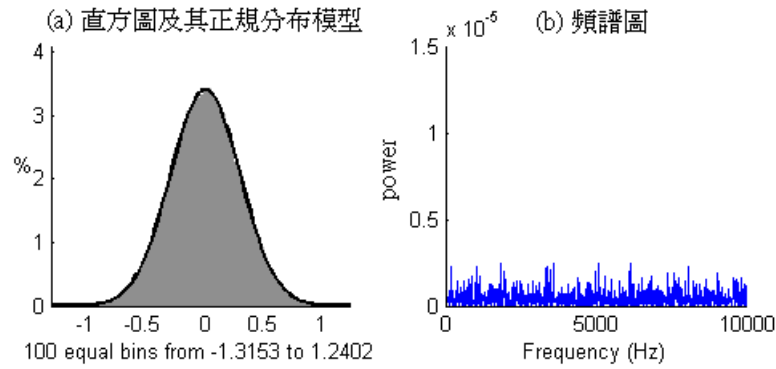


圖 6: matlab 模擬高斯白雜訊

### 2.3 正規分布的白雜訊

上一節我們討論雜訊一中隨機有次序產生的雜訊是否來自正規分布，在這裡即將探討何謂高斯白雜訊(Gaussian white noise)，也就是正規分布的白雜訊。白雜訊指的是訊號做了傅立葉轉換後取絕對值(平方)的強度大約相同，也就是說白雜訊在頻率域上的振幅都幾乎一樣大，呈現均勻分布。高斯白雜訊即為雜訊在時間域上呈現正規分布，在頻率域上呈現均勻分布。所以我們由 matlab 中我們創造的一筆隨機正規分布的雜訊，設定取樣頻率為 20k，持續時間為 0.1 秒，觀察是否為高斯白雜訊。如圖 6 所示，圖 6(a) 為直方圖及其正規分布模型，圖 6(b) 為頻譜圖。

傅立葉轉換就是將訊號在時間域上的函數轉換為頻率域上的函數。頻譜圖即為訊號傅立葉轉換後取絕對值平方的表現，若訊號為  $x$ ，則傅立葉級數為：

$$f(x) = a_0 + \sum_{k=1}^{\infty} \{a_k \cos(2k\pi x) + b_k \sin(2k\pi x)\} \quad (2.1)$$

傅立葉轉換就是  $f(x)$  用複數表現，表示成  $c_k = a_k + b_k i$ ，所以頻譜圖的縱軸為  $|c_k|^2$ ， $|c_k|$  表示頻率強度，強度的平方讓特徵更加明顯，單位則為能量，橫軸則為頻率。所以由圖 6 中 matlab 創造符合高斯分布的訊號可以發現，由於在頻率域上的振幅強度大約相同，可以說 matlab 內建所造的訊號是符合正規分布的白雜訊。



## 2.4 獨立且來自相同分布

在 2.2 節中，可以從雜訊一中隨機有次序找到來自相同分布並且遵從正規分布的雜訊。現在我們即將探討雜訊是否獨立 (independent)，如果獨立又符合雜訊來自相同分布，就可以說我們的雜訊符合一般統計假設 iid 並且遵從正規分布。在這裡利用當變數均來自正規分布時，共變異數為零若且唯若變數獨立來佐證；由於真實的神經訊號雜訊無法 100% 來自完全相同的正規分布，所以我們只要能夠找出共變異數夠小，即可說明變數為獨立。

首先，在  $d$  維度上多變數正規分布函數為：

$$p(x) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \quad (2.2)$$

其中， $x \sim N_d(\mu, \Sigma)$ 。在這裡，我們討論二維空間的情況，也就是  $d = 2$ ， $x = [x_1, x_2]^T$  為樣本點， $\mu = [\mu_1, \mu_2]^T$  為期望向量， $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$  為共變異數矩陣，其中， $\sigma_1$  為樣本  $x_1$  的標準差， $\sigma_2$  為樣本  $x_2$  的標準差， $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$  為相關係數， $\sigma_{12}$  為共變異數。因為  $|\Sigma| = (1 - \rho^2)\sigma_1^2\sigma_2^2$  以及  $z_i = \frac{x_i - \mu_i}{\sigma_i}$ ， $\forall i=1, 2$ ，得到  $(x - \mu)^T\Sigma^{-1}(x - \mu) = z_1^2 - 2\rho z_1 z_2 + z_2^2 = c^2$ ， $c$  為常數，所以 (2.2) 式可寫成：

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)} \quad (2.3)$$

變數  $x$  向量分別以  $x_1$  為橫軸， $x_2$  為縱軸，對應函數  $p(x_1, x_2)$  在  $x_1 - x_2$  平面的投影，會呈現如等高線圖的橢圓形，也就是函數  $p(x_1, x_2)$  圖形如山峰狀起伏，山峰高度對應於  $x_1 - x_2$  平面的等高線。當變數  $x_1$ 、 $x_2$  為獨立，也就是  $\sigma_{12} = E(x_1 x_2) - E(x_1)E(x_2) = 0$ ，等高線圖的主軸也就是橢圓的長短軸會平行於  $x_1 - x_2$  軸。當  $|\rho| < 1$ ，等高線為橢圓狀， $\rho > 0$  也就是  $\sigma_{12} > 0$ ，表長軸有正斜率， $\rho < 0$  也就是  $\sigma_{12} < 0$ ，表長軸有負斜率。橢圓以  $\mu$  為中心，軸的方向為  $\pm c\lambda_i^{\frac{1}{2}}u_i$ ， $i=1, 2$ ，其中， $(\lambda_i, u_i)$  為共變異數矩陣  $\Sigma$  的特徵值與特徵向量，

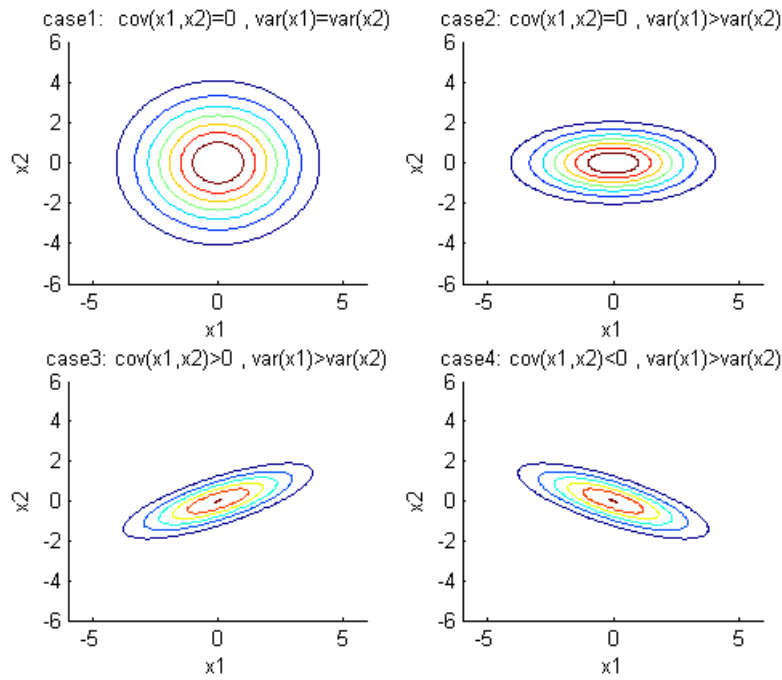


圖 7: 四種等高線圖形

我們可以舉下面四個例子來驗證，例一： $\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ ，例二： $\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ ，  
 例三： $\Sigma = \begin{bmatrix} 4 & 1.5 \\ 1.5 & 1 \end{bmatrix}$ ，例四： $\Sigma = \begin{bmatrix} 4 & -1.5 \\ -1.5 & 1 \end{bmatrix}$ ，期望向量均以 0 向量為中心，如圖 7，可以看到例一及例二共變異數  $\sigma_{12} = 0$ ，所以等高線圖形長短軸平行於  $x_1 - x_2$  軸，但例三及例四則以共變異數  $\sigma_{12}$  的正負號決定傾斜方向，長短軸則由變異數  $\sigma_1$  及  $\sigma_2$  大小決定，如表 3。

共變異數	代表圖形
$\sigma_{12} = 0$	$\sigma_1 = \sigma_2$ 為圓。
$\sigma_{12} = 0$	$\sigma_1 \neq \sigma_2$ ，為主軸平行於 $x_1$ 、 $x_2$ 的橢圓(長軸平行於變異數較大一方)
$\sigma_{12} > 0$	表長軸為正斜率的橢圓。
$\sigma_{12} < 0$	表長軸為負斜率的橢圓。

表 3: 共變異數對等高線圖形分析

在這裡，我們從 matlab 中隨機找出 A、B 兩群，此兩群各有 1000 個樣本點，對於「A、B兩群是否互相獨立？」做一數值測試，也就是說藉由 matlab 的內建隨機產生的高斯白雜訊驗證兩群是否相互獨立。A、B 兩群分別視為前述  $x_1$ 、 $x_2$  樣本點，並且均來自正規分布，我們可以藉由以下定理：當  $x_1$ 、 $x_2$  獨立，我們就可以說共變異數等於零；反之，不成立。但是當  $x_1$ 、 $x_2$  都來自正規分布時，此定理逆敘述成立，也就是說共變異數為零若且唯若  $x_1$ 、 $x_2$  獨立。藉由 matlab 對內建隨機訊號的觀察，觀測真實神經訊號的共變異數是否可靠。由此，我們可以隨機測試 1000 次，找出其共變異數矩陣非對角線項  $\sigma_{12}$  的絕對最大值以及絕對平均值。如下表 4，找出十次隨機測試 1000 次的絕對最大值以及絕對平均值。觀察  $|\sigma_{12}|$  約小於 0.08 都在可靠接受範圍；表示  $|\sigma_{12}|$  在此範圍內我們即接受此樣本為獨立，但是  $|\sigma_{12}|$  越接近零表示此樣本獨立性越強。

絕對最大值				
0.0792	0.0760	0.0756	0.0712	0.0828
0.0648	0.0721	0.0742	0.0656	0.0739
絕對平均值				
0.0179	0.0172	0.0180	0.0175	0.0183
0.0183	0.0173	0.0177	0.0177	0.0175

表 4: matlab 內建隨機訊號的共變異數絕對最大值與平均值

另外，我們可以由一必要條件來檢查雜訊是否符合 iid：假設  $X, Y$  均來自正規分布，並且符合 iid，也就是  $X, Y \sim N(\mu, \sigma)$ ，若  $\alpha, \beta \in \mathfrak{R}^+$  並且  $\alpha^2 + \beta^2 = 1$ ，令  $Z = \alpha X + \beta Y$ ，則  $\sigma_Z = \sigma_X = \sigma_Y$ 。接下來驗證由 *matlab* 擷取出來的高斯白雜訊是否符合此必要條件，程序如下：首先，令  $\alpha = \cos \theta$  以及  $\beta = \sin \theta$ ，其中， $\theta \in [0, \frac{\pi}{2}]$ ，使得  $\alpha^2 + \beta^2 = 1$  並且  $\alpha, \beta$  均為正數；第二，由 *matlab* 隨機產生雜訊，並且設定  $\mu = 0$ ， $\sigma = 1$ ，持續時間設定 5 秒，取樣頻率為 20k，由於取樣個數必須夠大，才會符合假設條件： $X, Y$  皆來自相同分布，所以產生兩筆不同的雜訊  $X, Y$ ，使得  $X, Y \sim N(0, 1)$ ；第三，檢驗  $Z = \alpha X + \beta Y$  是否符合  $\sigma_Z \approx \sigma_X \approx \sigma_Y$ 。經由此程序，我們找出來的  $\sigma_X = 0.9971$ ， $\sigma_Y = 1.0012$ ， $\sigma_Z = 0.9992$ ，由於雜訊為隨機產生，所以  $\sigma_X$  與  $\sigma_Y$  會與原先設定的  $\sigma = 1$  有所偏移，所以找出全部雜訊樣本的標準差，也就是  $X$  與  $Y$  的總標準差  $\hat{\sigma}$ ，計算相對誤差  $|\frac{\sigma_Z - \hat{\sigma}}{\hat{\sigma}}| = 0.006\%$ 。相對誤差很小，也就是說  $\sigma_Z \approx \sigma_X \approx \sigma_Y$  成立，表示由 *matlab* 內建擷取出來的高斯白雜訊非常符合 iid。

## 第 3 章 異常分布的探討

第二章探討一般統計假設 iid 的理論與方法，並且由直方圖中可以發現雜訊有中央突起，本章節即將探討此異常突起的特性，首先發現此異常突起為均勻散佈，並且由其他實驗雜訊的比較中，探討此異常突起是否為四捨五入造成，或者是不同儀器設備所致，還是電極接收到其他微弱訊號。並且觀察雜訊在頻率域上的表現呈現穩定起伏，於是我們提出一種新的濾波方式，使得雜訊更加符合正規分布，獨立性也更好。

### 3.1 異常訊號的均勻散佈

從第二章的討論中，可以發現直方圖中都有一中央突起，如圖 4。我們想要知道那一段中央突起是什麼或者是有什麼特性？觀察雜訊一的直方圖，將值域分割成 300 等份，每一等份  $h$  約為 0.0088，如圖 4，最高百分比 2.7337% 為中央突起，落於第 154 等份，此等份落於  $[-0.0025, -0.0025+h] = [-0.0025, 0.0064]$  區間內；因為訊號解析度以 0.001 為單位，所以值域可以化簡成  $[-0.003, 0.006]$ 。

首先，觀察值域  $[-0.025, 0.025]$  之間對應雜訊一共有 28760 樣本點，如圖 8(a) 為其直方圖，其中  $x$  軸：以 0.001 為單位並且落於  $[-0.025, 0.025]$  之間的資料， $y$  軸：這些資料的個數百分比；其次，發現於  $[-0.005, 0.009]$  的分布比例似乎比其他區段高，因此細看以 0.001 為單位的 15 區段，對應於雜訊一共有 11273 樣本點，如圖 8(b) 為其直方圖， $x$  軸：以 0.001 為單位並且落於  $[-0.005, 0.009]$  的 15 區段， $y$  軸：這 15 筆資料的個數百分比。觀察圖 8(b) 可以清楚看到有三根特別突起(為  $x$  軸等於 0.003、0.004、0.005，共有 4125 樣本點)，甚至前後幾根(如， $x$  軸等於 0.002、0.006 共有 1735 樣本點)，都是造成圖 4 中央突起特別因素。



我們想要探討是什麼因素造成這種現象？第一：可能為四捨五入所造成的結果。由於我們的資料來自真實的生物訊號，但是我們在取資料時為了模擬，所有的資料都會在小數第四位四捨五入，得到全部的資料都是小數點以下三位，可能會有截去誤差。但是，由於我們的資料數夠大，且每一資料四捨五入的機率都是相同的，所以並不會有如此顯著的影響。第二：可能為機器所致。第三：可能為某潛在的生物活動，或接收到其他訊號及組織所產生的電訊號。

由於中央突起表示有一微弱訊號大比例的散佈於雜訊中。我們想要觀察此微弱訊號是否有特別分布或其特徵？所以嘗試找出其微弱訊號於雜訊一的散佈圖，也就是將這些於  $[-0.025, 0.025]$  之間的訊號在雜訊一之間取出來，雜訊一的樣本點落於  $[-0.025, 0.025]$  區段內即標記為 1，否則即標記為 0，如圖 9，為雜訊一其中一段 0.001 秒的標記圖。將此標記依序排列之，標記以 548 個記號為單位，排列成 548 列，也就是標記 0.0247 秒為一列，可表示成最大方陣，元素不是 0 就是 1，若方陣元素對應於 1 則畫點，如圖 10(a) 即為散布圖。圖 10(b) 則於  $[-0.005, 0.009]$  區段間，同上述方式找出其散佈圖。如果上述散佈圖能夠觀察出一些訊號與時間有關，則猜測此微弱訊號有一規律 spike，但圖 10 的散佈圖如同雜訊般均勻散佈。因此，造成大分布比例的微弱訊號，可以說是均勻散佈在所有實驗的時間內。

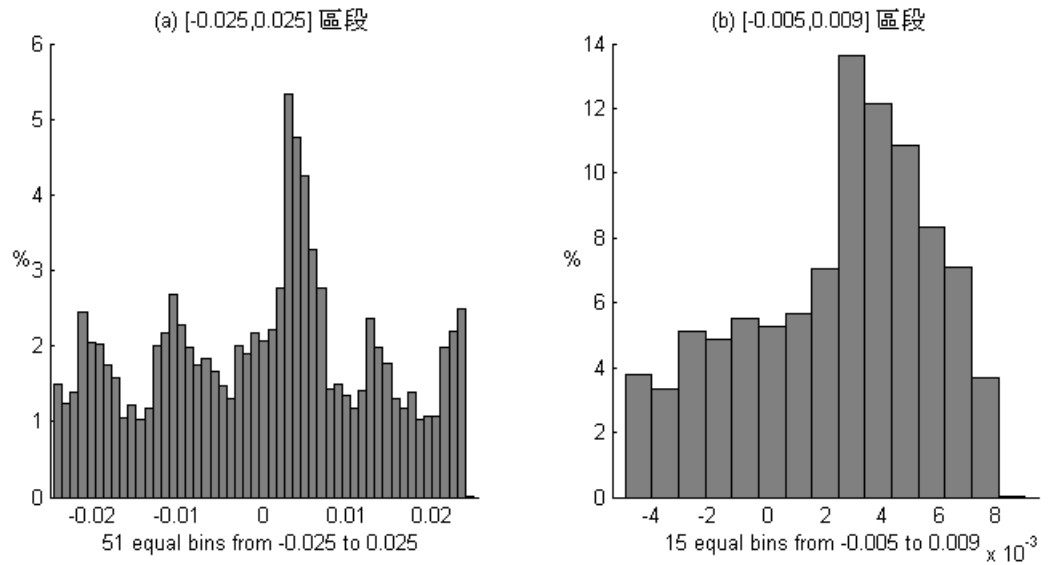
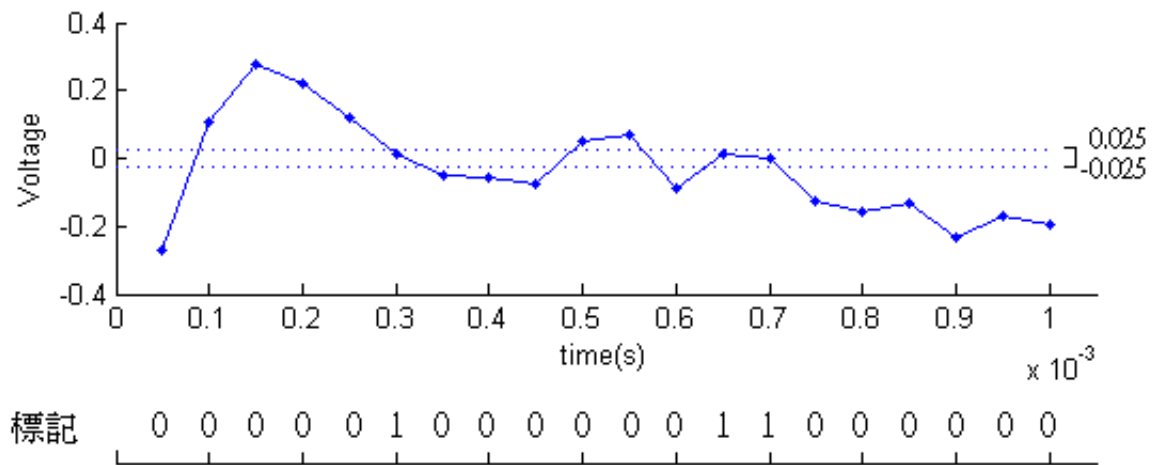


圖 8: 雜訊一於兩種區段內的直方圖



若在  $[-0.025, 0.025]$  區段內，則標記 1，否則標記 0

圖 9: 標記分析圖

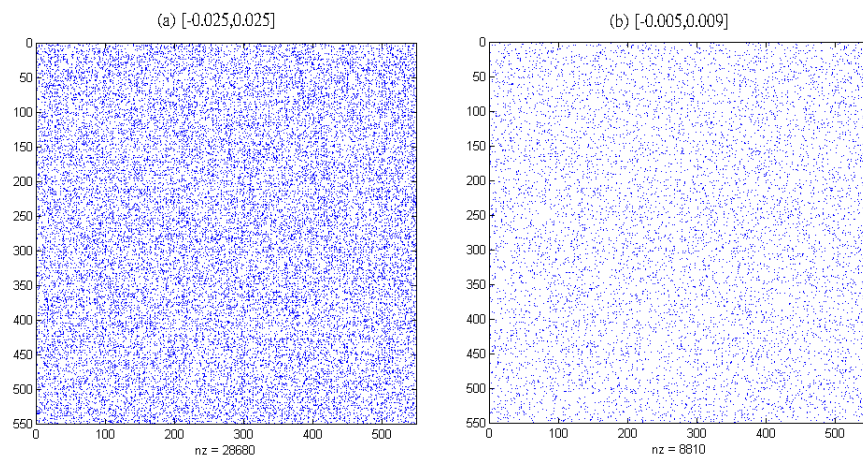


圖 10: 雜訊一的均勻散佈圖

## 3.2 異常訊號的探討

我們想觀察其他生物訊號資料是否由均勻散佈的異常訊號所造成的中央突起，會不會因為不同機器所擷取、解析度不同或者是電極接收其它微弱訊號所致。所以在此節比較其他實驗雜訊，取樣頻率皆為每秒四萬點，解析度為六位小數，並且觀察四捨五入是否造成影響，最後針對上述異常訊號的實驗測試作一討論。在此，關於訊號和雜訊的來源以及相關資訊請見附錄。

### 3.2.1 其他實驗雜訊的比較

第一筆雜訊實驗測試所使用的訊號命名為訊號二，類似訊號一；我們使用 2.1 節人工選取雜訊的程序，找出一筆 349360 個樣本點的雜訊，此訊號命名為雜訊二。關於訊號二與雜訊二的相關資訊請見附錄。

由於雜訊二大部分值介於  $[-1,1]$  之間，我們將其大於 1 及小於 -1 的值忽略。解析度為六位小數，並且為了與雜訊一比較，將  $[-1,1]$  等分為 740 等份，每一等份為 0.0027，期望值為 -0.0050，標準差為 0.2161。如圖 11(a) 為雜訊二的直方圖。這張圖的製作，與圖 4 性質相同，前者每個區段 ( $h=0.0088$ ) 內大約有八或九種可能的數值，因為雜訊一只有三位有效數字，解析度是 0.001。現在每個區段內 ( $h=0.0027$ ) 內約有八或九種可能的數值，因為雜訊二有六位有效數字，解析度是  $0.000312x$ ，而  $x$  為進位或捨去。將其排序，重覆的數據點視為一個數據點，並且計算差異，也就是將其第二項減第一項，第三項減第二項，依此類推；觀察此程序「排序-去除重覆-計算差異」擷取出來的資料點，可以發現數值大多為 0.000312、0.000313，共有 555 個資料點，其他少數約為 0.000312 (或 0.000313) 的整數倍。由此可見，獲得這些訊號的軟體或硬體，最低解析度大約就是  $0.000312x$ ，而  $x$  為進位或捨去。而且有「量子」現象，即所有數據的差距都是最低解析度的整數倍。

第二筆雜訊實驗測試所使用的訊號命名為訊號三，值域落在  $[-0.2163, 0.0928]$  之間，內容全為雜訊，關於訊號三的相關資訊請見附錄。多數的資料介於  $[-0.1, 0.1]$  之間，將其資料大於 0.1 及小於 -0.1 的值忽略，直方圖分成 75 等份如圖 11(b)，每一等份為 0.0027，大約有八或九種不同數值落在一個  $h=0.0027$  的區段內。取  $[-0.1, 0.1]$  之內的雜訊，仍採用前述「排序-去除重複-計算差異」的檢查程序，發現數值大多為  $0.000312x$  的整數倍，顯示紀錄這筆訊號的軟體或硬體，同前述的量子現象，最低解析度似乎就是 0.000312。

### 3.2.2 四捨五入的影響

由於在第二章所使用的雜訊一解析度為三位小數，現在我們來測試解析度六位小數四捨五入至三位小數後，對於雜訊二及訊號三的影響。如圖 11(c) 與圖 11(d) 為雜訊二及訊號三四捨五入至解析度為三位小數後的直方圖。雜訊二在圖 11(a) 與圖 11(d) 比較之，由於多數資料落於  $\pm 1$  區段間，將其分成 225 等份，每一等份約 0.0089，發現每段約有八至九種可能的數值落在其中，因為其解析度為 0.001。訊號三在圖 11(b) 與圖 11(d) 比較之，由於多數資料落於  $[-0.1, 0.1]$  區段間，將其分成 75 等份，每一等份約 0.0027，發現每段約有二至三種可能的數值落在其中，因為其解析度為 0.001；若等分成 23 等份，每一等份約 0.0087，每段約有八至九種可能的數值，但在這裡不考慮此等分數，因為等份太少無法由直方圖看出其特色。

我們可以發現訊號三解析度三位小數與六位小數同樣有異常突起。由於雜訊二在解析度六位小數直方圖分析時，本來就沒有很明顯的中央突起，再加上等份  $h$  放大並且解析度變大所致，所以在解析度三位小數直方圖分析的中央突起更不明顯，但同樣表示有一微弱訊號均勻散布於此。也就是說，對於前述所提到的：解析度四捨五入為三位小數對於直方圖並無太多影響，所謂的「中央突起」現象不是因為將數據做四捨五入造成的。

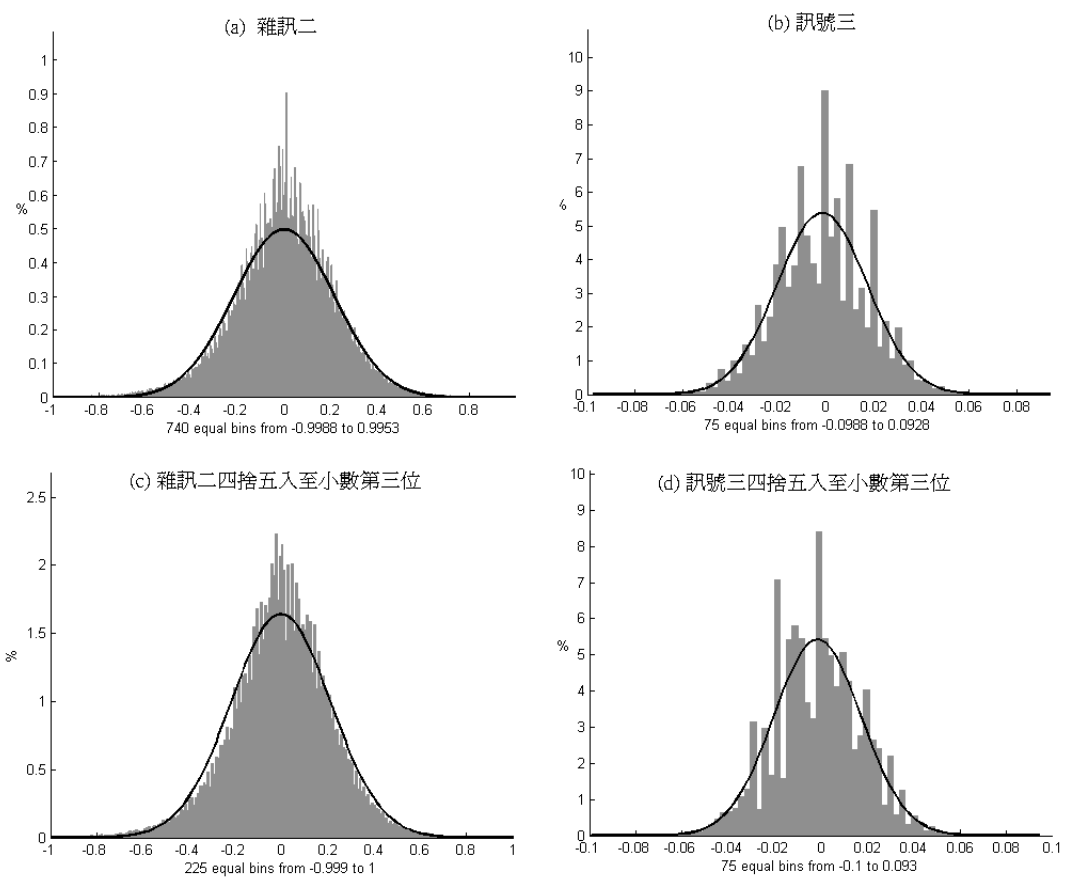


圖 11: 雜訊二與訊號三解析度四捨五入至小數第三位的直方圖變化



### 3.2.3 問題與討論

針對上述實驗提出了幾個問題，並且觀察與討論之。問題一：為什麼會有最低解析度的量子現象？這是實驗中預期的嗎？這就是機器的特性，我們叫它 AD 轉換，也就是從類比到數位的轉換會有最低解析度，類比就是連續訊號，在數學上就是所有的實數都有可能發生，可是機器不可能精確到任意小數點的實數。因為這樣的特性，對於我們的雜訊必需做一些欲處理的動作，例如，四捨五入。問題二：觀察圖 11(b)，在直方圖上觀察到訊號三有太多異常突起，而且理論模型比起其他雜訊的實驗更不吻合直方圖。我們強烈懷疑此雜訊的正規性 (normality)？這組被當作雜訊的實驗訊號，可能並不理想。問題三：觀察圖 11(a)，雜訊二雖然不像以前的實驗雜訊一，表現那麼明顯的「中央突起」，這樣子還算是有「中央突起」現象嗎？如果可以接受了，不算它有，那麼前次的雜訊一的擷取或預處理過程就是有缺陷的，該要檢討。但是，前次雜訊一只要「濾掉」突起部分，其實比現在這筆雜訊二更接近正規分布模型，應該更加可靠。

針對以上的實驗，首先，在這裡包括所使用電極、線路及放大器都不一樣，所以我們發現訊號三值域明顯較小，約落在  $[-0.1, 0.1]$  之間，可以用一數值實驗來驗證，也就是將訊號三先放大 20 倍，再四捨五入轉換解析度為三位小數，並且取奇數點，發現大部分訊號落於  $[-1.5, 1.5]$  之間，並且觀察其直方圖 (圖 12)。「取奇數點」是因為取樣頻率不同，訊號三為 40k，雜訊一為 20k，我們要使得取樣頻率一樣，所以取奇數點；「放大 20 倍」使得訊號三與雜訊一的振幅差不多。可以發現訊號三放大後，與原始訊號雷同，同樣有異常突起。爾後，我們將訊號三轉換解析度為三位小數，放大 20 倍並且取樣頻率轉換成 20k 的訊號命名為雜訊三；同樣，雜訊二指的是解析度轉換成三位小數，取樣頻率轉換成 20k。關於雜訊二與雜訊三的來源與相關資訊，請見附錄。

其次，雜訊二及訊號三同樣有所謂的量子現象，但我們可以確定兩筆是完全不同的，訊號三是作實驗以前擷取的雜訊，另一個雜訊二與雜訊一雷同，是作實驗中獲取的雜訊，所以有比較大的 spike，需要人工擷取雜訊。但這兩筆精確到了六位小數，可以發現兩兩之間的差是一樣的，同為  $0.000312x$  或者是為其倍數 ( $x$  為捨去或進位)，但雜訊一為精確到  $0.001$ 。

第三，雜訊的來源，一個是環境雜訊，包含 white noise；一個是生物體內各種活動的干擾，也就是探針接收到遠方 spike 傳來相互疊合的微弱訊號。雜訊二和訊號三都是雜訊；雜訊二像是肌電位或心電圖干擾的情況，比較不像是 white noise，雖然沒有很突出的中央突起，但是可以把他認定與雜訊一為同樣一種訊號。訊號三像是在麻醉的狀況下，或老鼠較安靜的時候，像是低頻電波，或是像心電圖這種比較低頻的波，或是儀器有濾不掉的雜訊(我們的分析發現訊號三不太符合正規分布，有好幾根異常突起，第一為來自遠方的微弱訊號，第二為儀器的關係，可能品質不好)。

第四，經常有人錯誤的假設高斯分布的雜訊就是 white noise，這兩者不一定相等，高斯分布指的是雜訊在時間軸上原始的訊號的數值分布，也就是直方圖符合正規分布；white noise 指的是做了傅立葉轉換後取絕對值平方之後能量大約相同，也就是說 white noise 在頻率域上的振幅都幾乎一樣大，呈現均勻分布。如果我們的測試正確，應該在時間域上做頻次圖，像個正規分布，但是把雜訊做傅立葉轉換取絕對值平方後，能量應該要大約相同。我們已測試過這幾筆訊號是否符合高斯分布，接下來測試是否為 white noise，也就是做了傅立葉轉換取絕對值平方之後是不是平緩的。如圖 13，時間設定為持續 0.1 秒，圖 13(a) 為三筆訊號頻譜圖，圖 13(b)、圖 13(c) 為觀察局部放大的頻譜圖。使用雜訊一、雜訊二以及雜訊三做頻率域上分析，取樣頻率同為 20k。由於雜訊三放大了 20 倍，使得中心偏移，因為機器的特性，我們可以將此雜訊期望值平移到 0。

由上面的實驗分析，可以知道我們的三筆雜訊只能說是正規分布，而且並不是非常符合正規分佈的數學模型，並且不能說是 white noise。但由圖 13 可以發現，約在頻率大於 7000 的部份，我們可以猜測主要為 white noise、遠方來的雜訊以及機器的特性所構成。所以我們可以知道高頻主要為 white noise，低頻有其他來源的雜訊，我們通稱為生物的造成的雜訊。在這裡驗證了高斯分布不一定是 white noise，我們只能說此雜訊符合高斯分布。

---

註：在 matlab 中可以使用 randn 此指令，隨機創造一筆來自高斯分布  $N(0,1)$  的雜訊，可以放大或縮小 a 倍，使得此雜訊與雜訊一的範圍雷同，以利實驗分析。

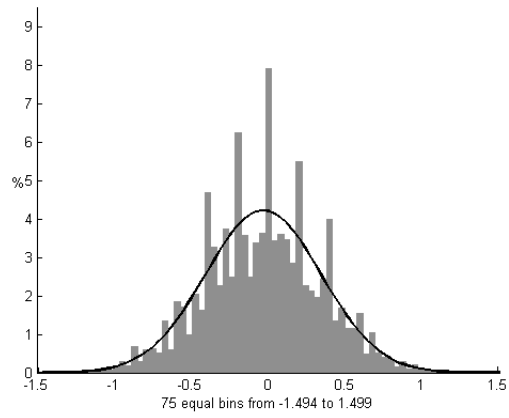


圖 12: 訊號三放大後直方分布圖以及正規分布模型

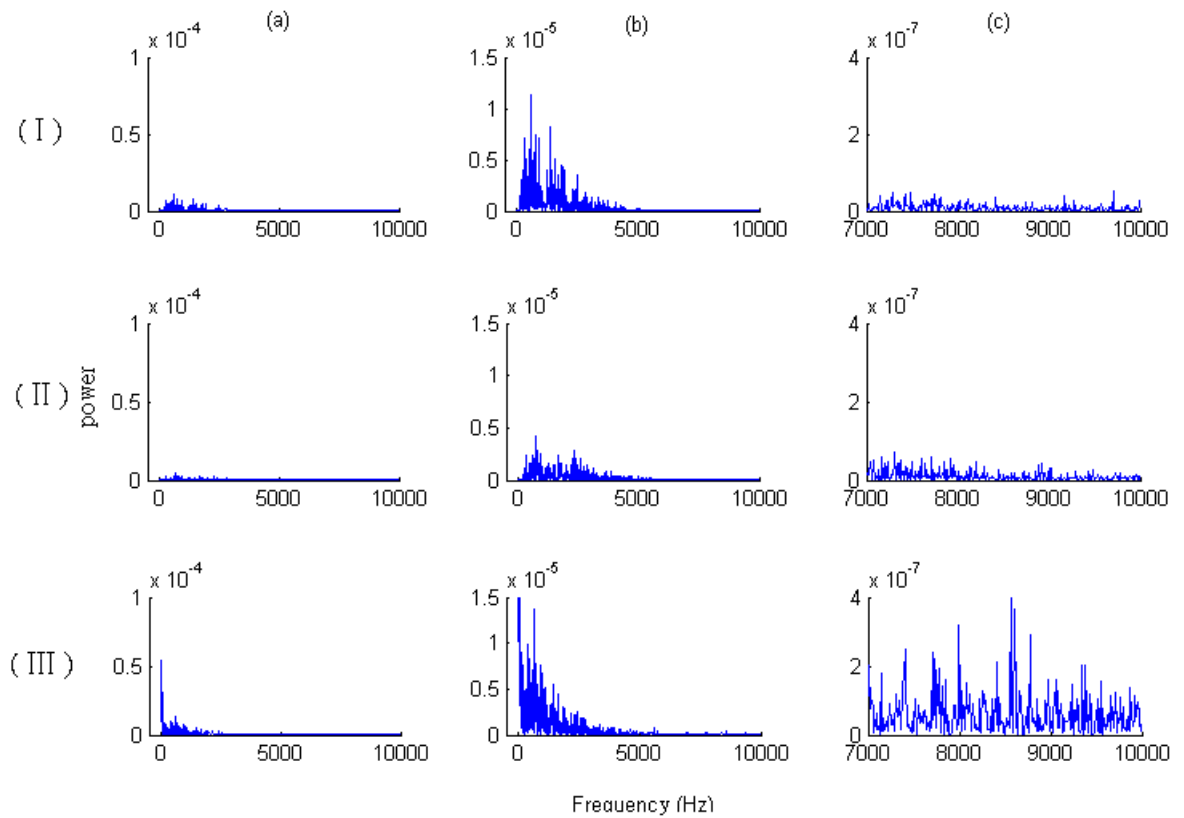


圖 13: (I)雜訊一 (II)雜訊二 (III)雜訊三 的頻譜圖

### 3.3 異常訊號在頻率域上的探討

觀察雜訊在頻率域上的表現，發現有一能量極大的頻率，並且雜訊一及雜訊三較雜訊二明顯。所以觀察濾波對訊號分布的影響，發現濾波後訊號更加正規，接著想要知道的是「多濾掉幾的頻率，是否會使得訊號更加正規」，並且利用第二章的理論與方法驗證我們的雜訊是否符合 iid。另外，我們可以從頻率域上的觀察，發現頻譜呈現穩定起伏，於是提出一種新的濾波方式，使得訊號更符合正規性與 iid。

#### 3.3.1 濾波對訊號分布的影響

由於在 3.2 節觀察到有些雜訊在頻率域上有一能量極大的頻率，首先觀察雜訊一擷取 0.1 秒的訊號來分析，取樣頻率 20k。

觀察圖 14(a)-1 雜訊一的頻譜圖，由圖中發現有一能量較大的頻率，針對此頻率，由以下實驗步驟對雜訊一探討。

**step1** 頻率域的觀察：可利用訊號傅立葉轉換後，觀察頻率域上的能量大小，

如圖 14(a)-1 為其頻譜圖。圖中發現有一個較大的值，這表示所對應頻率的能量特別大，並且可以發現此強度大的頻率即為突起的頻率。

**step2** 濾波前後直方圖與正規分布模型：利用逆向的傅立葉轉換將此頻率濾

掉，比較濾波前後雜訊一的直方圖與正規分布模型，如圖 14(a)-2 與 14(a)-3。在這裡，濾掉的頻率即為能量最大的頻率，也就是明顯突起的頻率 570 Hz。

**step3** 誤差波形：找出的新訊號與原訊號的誤差圖形類似一正弦波，並且此波

於  $[-0.07, 0.07]$  之間震盪，這就是我們所濾掉頻率為 570 Hz 的波。發現誤差波形為一相位角的正弦波，將此正弦波濾掉之後發現雜訊一直方圖中央突起的微弱訊號似乎不明顯了，會使得雜訊一直方圖變的更加正規化。由此實驗可發現雜訊一減弱一個頻率後使得雜訊更加正規化，並且

先前探討異常訊號的中央突起似乎不明顯了。

**step4** 相對誤差的比較：我們想知道濾掉一個強度較大的頻率後，直方圖有多符合正規分布模型，也就是在濾波後直方圖對於高斯模型曲線的相對誤差是否有變好？並且使用  $l_2$ -norm 計算其相對誤差，得雜訊一濾波前相對誤差為 29.05%，濾波後相對誤差為 24.13%。

同上述四步驟分析雜訊二以及雜訊三(圖 14(b) 與圖 15(a))，濾掉頻率、誤差波形振幅、雜訊值域範圍以及濾波前後相對誤差比較如表 5，並且發現雜訊三有一能量極大的頻率，比雜訊一、二大很多。由於雜訊二的中央突起本來就沒有如此明顯，濾波前後中央突起與直方圖似乎沒有太大改變。但雜訊三濾波後直方圖的異常突起都變少了，使得雜訊更加正規化；在此呼應先前 3.2.3 小節的問題二：雜訊三的直方圖有太多異常突起，所以我們強烈懷疑此雜訊的正規性，由此亦可驗證此觀測。

濾波分析	雜訊一	雜訊二	雜訊三
濾掉頻率	570 Hz	700 Hz	30 Hz
誤差波形振幅	[-0.07,0.07]	[-0.04,0.04]	[-0.15,0.15]
雜訊值域範圍	[-1,1]	[-1,1]	[-1,1]
濾波前相對誤差	29.05%	28.24%	46.85%
濾波後相對誤差	24.13%	25.48%	15.11%

表 5: 雜訊一、二、三的濾波分析

---

註：相對誤差 =  $\sqrt{\frac{\sum(\hat{x}-x)^2}{\sum(x)^2}} \times 100\%$ ；估計值  $\hat{x}$  為直方圖所對應的值，實際值  $x$  為所找出正規分布模型曲線相對應的值。

此三筆訊號都可以發現一有相位角的正弦波，將此正弦波濾掉之後使得原訊號分布圖變的更加正規化。雖然雜訊一與雜訊二極為相似，此三筆雜訊大部分落於  $[-1, 1]$  之間，並且雜訊一的正弦波震盪於  $[-0.07, 0.07]$ ，雜訊二的正弦波震盪於  $[-0.04, 0.04]$ ，雜訊三的正弦波震盪於  $[-0.15, 0.15]$ ，可以發現雜訊三「中央突起」較明顯，雜訊二「中央突起」較不明顯，這是因為雜訊三的正弦波振幅比例相較於  $[-1, 1]$  之間來的大，所以有很多異常突起；其次為雜訊一，再者為雜訊二，所以雜訊一中央突起比雜訊二明顯。

觀察濾波前後相對誤差，發現濾波後相對誤差變小了。雜訊三濾波前後誤差相差最大，也就是說此減弱此低頻強烈影響雜訊三的正規性；濾波前雜訊二比雜訊一相對誤差較小，但濾波後雜訊二比雜訊一相對誤差大，並且雜訊二無太大改變，這也呼應了 3.2.3 小節問題三的觀察，濾波後雜訊一比雜訊二更符合高斯模型。

擷取另外三筆訊號取樣頻率均為 40k，分別為訊號四、訊號五、訊號六，均在生物體麻醉的狀況下取得，事先以機器做 0.5 Hz 至 3000 Hz 的帶通濾波，均放大 20000 倍。此三筆訊號擷取方式與訊號三同，為人工判讀不含有 spike 的訊號。為了實驗的一致性，將三筆訊號均取奇數點並且取到小數第 3 位(第 4 位四捨五入)使得解析度以及取樣頻率一致，分別產生雜訊四、雜訊五、雜訊六，取樣頻率均為 20k。(關於訊號以及雜訊來源，請見附錄。)

同前述四步驟分析雜訊四、雜訊五以及雜訊六(圖 15(b) 與圖 16)，誤差波形振幅、雜訊值域範圍以及濾波前後相對誤差比較如表 6，雜訊四以及雜訊五同樣突起的頻率為 360 Hz，誤差波形為類似震盪於  $[-0.04, 0.04]$  一相位角的正弦波。雜訊六突起頻率為 370 Hz，誤差波型震盪於  $[-0.015, 0.015]$ 。此三筆雜訊同樣沒有明顯的中央突起，並且頻率所對應的能量同樣沒有先前討論的雜訊三頻率的能量大。



濾波分析	雜訊四	雜訊五	雜訊六
濾掉頻率	360 Hz	360 Hz	370 Hz
誤差波形振幅	[-0.04,0.04]	[-0.04,0.04]	[-0.015,0.015]
雜訊值域範圍	[-0.35,0.32]	[-0.35,0.27]	[-0.335,0.29]
濾波前相對誤差	17.97%	21.67%	20.23%
濾波後相對誤差	15.18%	18.51%	17.92%

表 6: 雜訊四、五、六的濾波分析

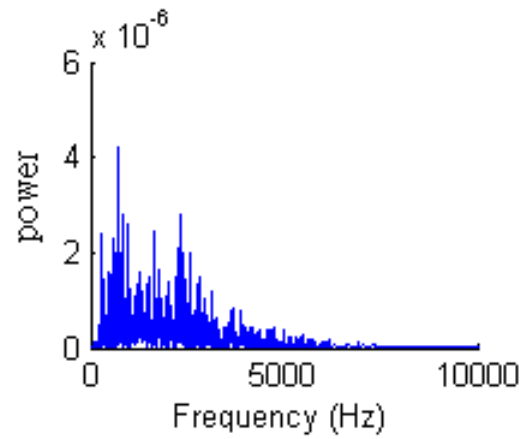
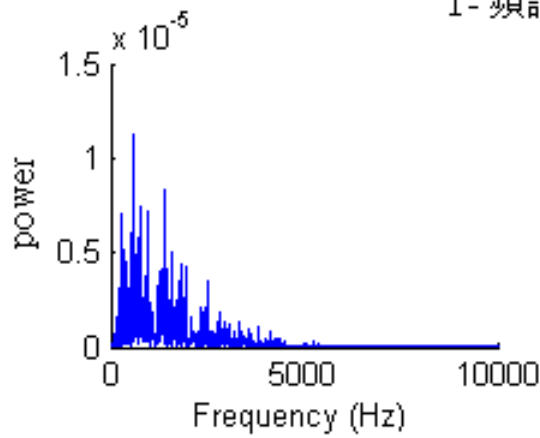
由於此三筆雜訊為同一隻老鼠不同電極的紀錄，所以紀錄的條件是一樣的，均在老鼠麻醉的狀況下取得，會使得訊號接收到遠方神經元產生的雜訊減弱，肌電圖所產生的雜訊也沒有了。我們發現這一部份的訊號已經比雜訊一、雜訊二以及雜訊三相對誤差小很多。

另外，我們由圖 14 至圖 16 的直方圖中可以發現，有一些不落於中央的異常突起，由於在此段討論之實驗雜訊資料只是任取一段 0.1 秒的雜訊一，只有 2000 點。故不合模型的「異常」分布，看似較多。此段討論的目的是比較濾波前後對於分布造成的影響，所以其直方圖與 P.10 之圖 4 不同(使用全部雜訊約 30 萬點製圖)。就圖 4 觀察，並無不再中央的異常分布。

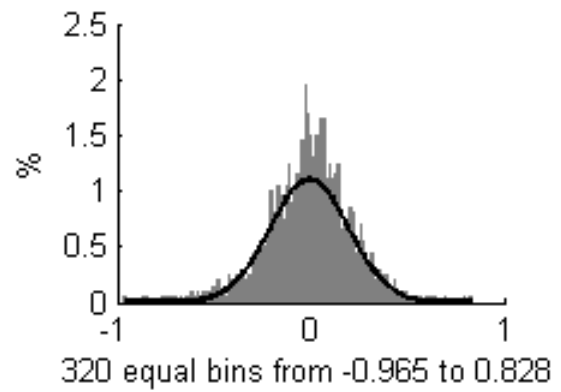
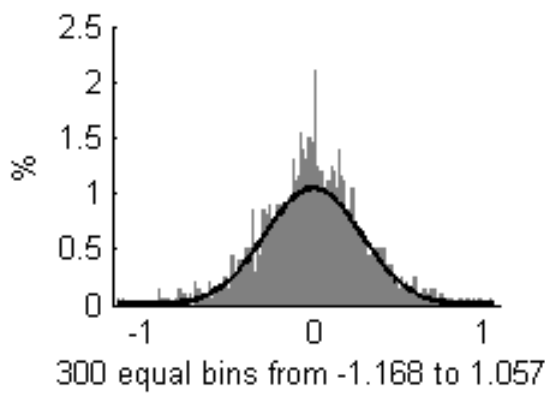
(a) 雜訊一

(b) 雜訊二

1- 頻譜圖



2-原始訊號直方圖與其正規模型



3-1<sup>st</sup>濾波後直方圖與其正規模型

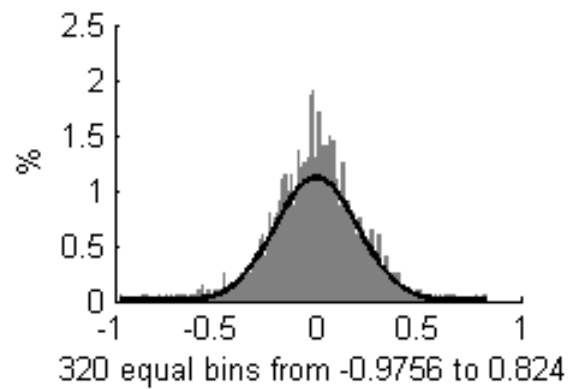
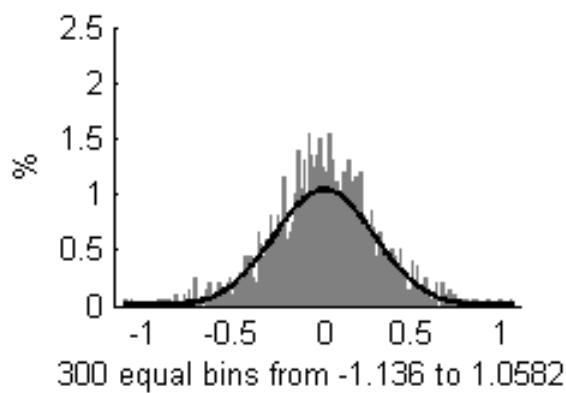
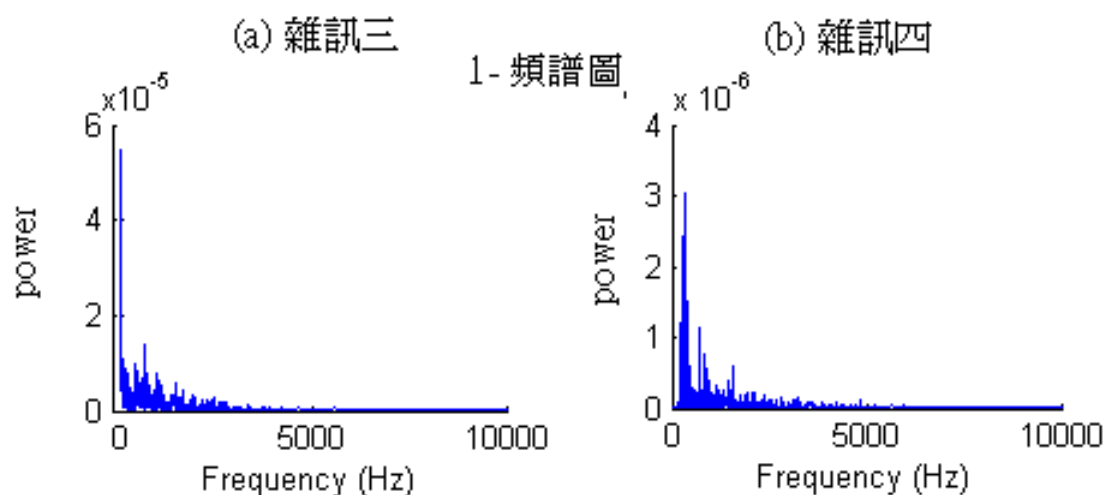
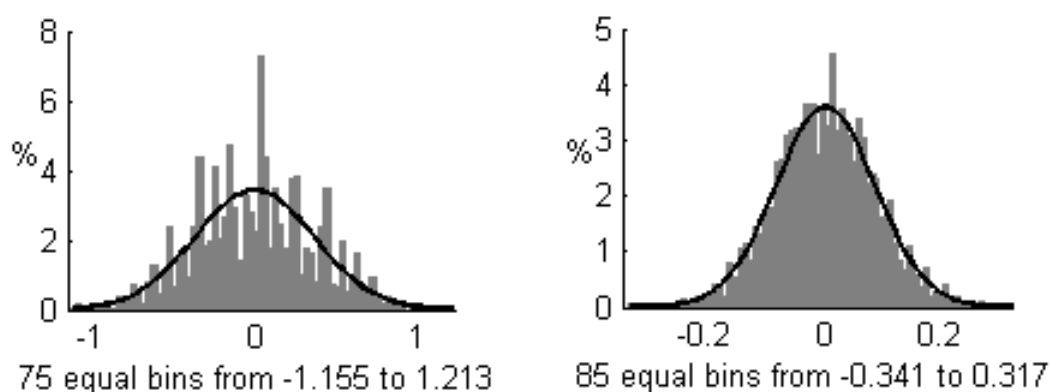


圖 14: 雜訊一、二的濾波分析



2-原始訊號直方圖與其正規模型



3-1<sup>st</sup>濾波後直方圖與其正規模型

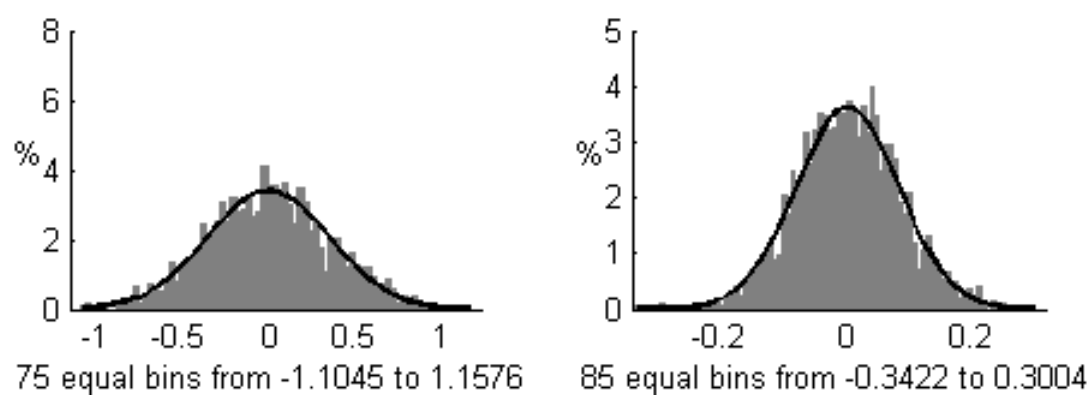
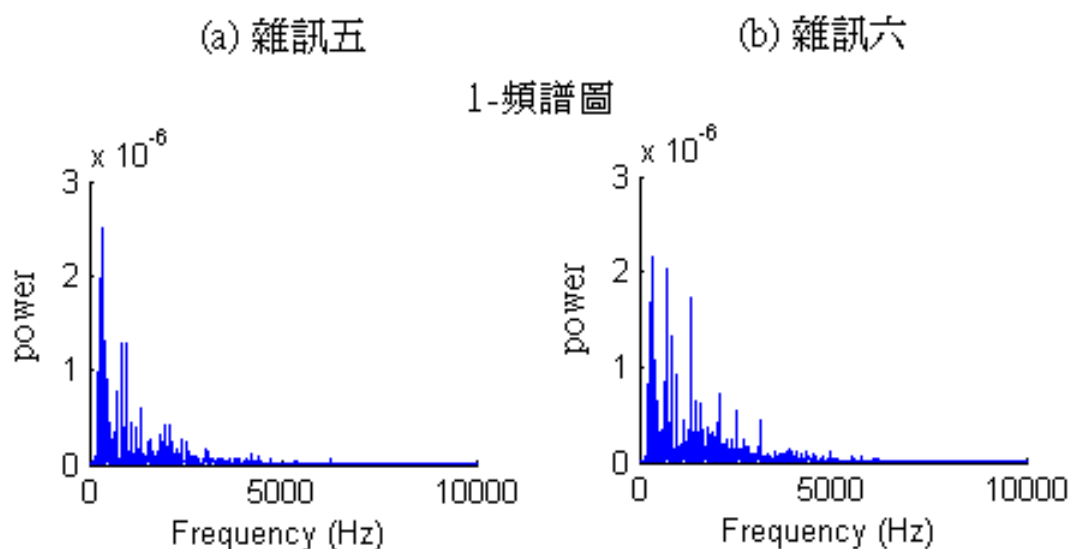
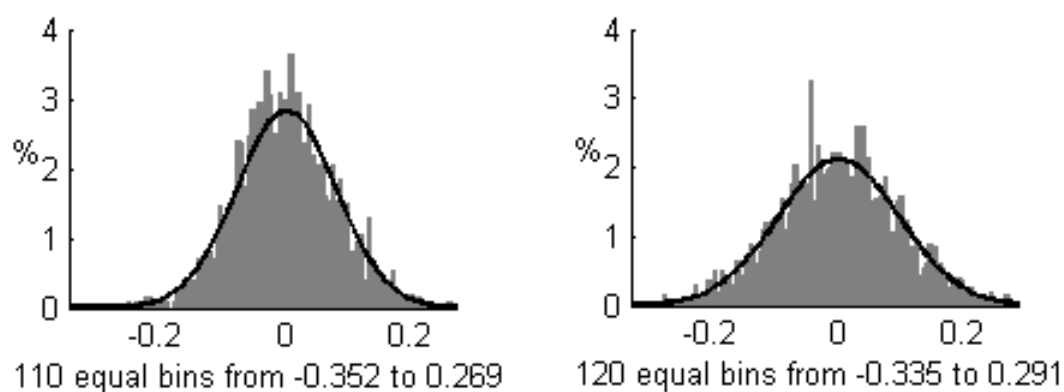


圖 15: 雜訊三、四的濾波分析



2-原始訊號直方圖與其正規模型



3-1<sup>st</sup>濾波後直方圖與其正規模型

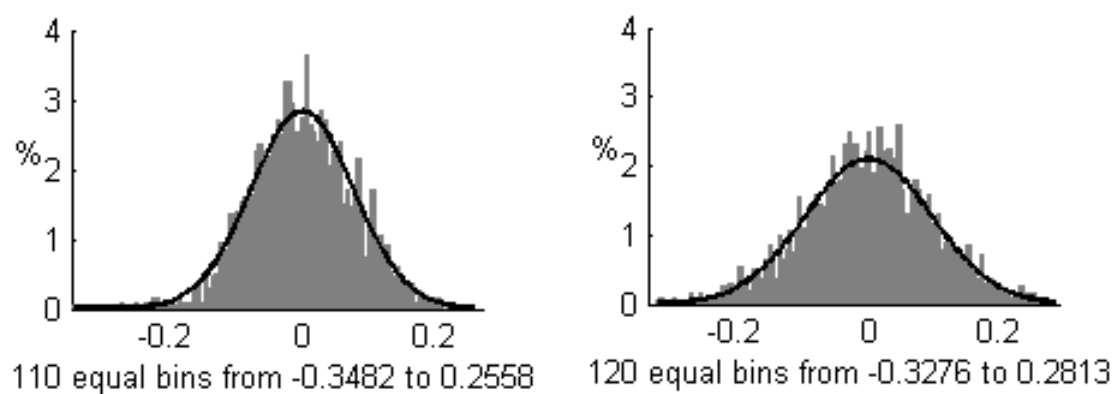


圖 16: 雜訊五、六的濾波分析

### 3.3.2 逐次濾波對訊號分布的影響

在上一節中，我們只濾掉一個頻率就會使得訊號更正規化；我們的問題是「多濾掉幾個頻率，是否會使得訊號更加正規？」如果我們要讓此濾波方式適用一般情況的雜訊，我們不應該只特定濾掉一個頻率，因為在實驗過程中難保頻率會做些微改變。

藉由以下實驗來探討「濾掉頻率與相對誤差」的關係，觀察逐次濾掉能量最大的頻率後，相對誤差的變化關係。測試的訊號為雜訊一及雜訊三，由於此兩訊號都有明顯的「中央突起」，並且雜訊三比較不符合正規分布模型，我們希望藉由此探討，將減弱一些頻率，使得訊號更加正規化。表 7 為濾波後的相對誤差，我們可以觀察出雜訊一經過第一次濾波後，濾掉 570 Hz 後相對誤差變小了，但第二次濾波後相對誤差都比較大；雜訊三第一次濾掉 30 Hz 後相對誤差就小很多，並且第二次濾掉 10 Hz 後相對誤差又更小一些，但其餘的濾波相對誤差都顯得不穩定，所以並不是濾掉越多的頻率相對誤差就越小。

雜訊一	原始訊號	570 Hz	1410 Hz	770 Hz	580 Hz	940 Hz	300 Hz
		29.05%	24.13%	26.49%	28.83%	31.37%	28.75%
雜訊三	原始訊號	30 Hz	10 Hz	630 Hz	90 Hz	20 Hz	380 Hz
		46.86%	15.11%	14.09%	15.48%	14.94%	12.18%

表 7: 逐次濾波後的相對誤差

### 3.3.3 濾波對 iid 的影響

在第二章，我們探討本論文所使用的雜訊一是否符合正規模型，但是發現此雜訊卻有異常突起，並且濾波能夠改善此雜訊的正規性，所以在這裡探討濾波是否能改善此雜訊的獨立性，以致於讓此雜訊更符合 iid。

利用 2.4 的理論與方法，討論濾波前後對於雜訊一獨立性的影響？藉由以下定理：當兩樣本均來自正規分布時，共變異數為零若且唯若兩樣本獨立。首先，觀察未濾波時雜訊一的共變異數，利用雜訊一創造兩組實驗數據 C 群以及 D 群，其中，此兩群設定持續時間 0.05 秒，並且每一群各有 1000 個樣本點，C 群為 c 疊加上第一筆雜訊，D 群為 d 疊加上第二筆雜訊，雜訊由雜訊一產生，取法則由 2.1 闡述隨機有次序的產生 2 群持續 0.05 秒的雜訊。在這裡設  $c=0$ 、 $d=0.75$ ，所以可以得到期望向量  $\mu \approx [-0.0015, 0.77515]^T$ ，標準差分別為

$$\sigma_C \approx 0.2542、\sigma_D \approx 0.2453，共變異數矩陣 \Sigma = \begin{bmatrix} 0.0805 & 0.0030 \\ 0.0030 & 0.0866 \end{bmatrix}。$$

C、D 兩群分別視為 (2.3) 式中的  $x_1$ 、 $x_2$  樣本點，也就是  $x_1$ 、 $x_2$  均來自正規分布，並且由 (2.3) 式計算其對應函數  $p(x_1, x_2)$ ，我們可以找到投影於 x y 軸的等高線圖，如圖 17 為雜訊一的函數圖形以及等高線圖，其中，x 軸對應樣本點  $x_1$ ，y 軸對應樣本點  $x_2$ ，z 軸為  $p(x_1, x_2)$ 。並且圖形以  $\mu$  為中心，發現其等高線圖近似圓形，這是因為變異數幾乎相等並且共變異數接近零。在這裡共變異數不為零，但是共變異數接近零，所以我們藉由 matlab 對內建隨機訊號的觀察，如表 4，觀測我們的共變異數是否可靠。由於 matlab 內建產生隨機雜訊的  $|\sigma_{12}|$  平均約在 0.018，但這裡  $|\sigma_{12}| = 0.003$  比 0.018 小很多，所以我們能夠接受此雜訊為獨立，並且符合統計假設 iid。

其次，觀察濾波後雜訊一的共變異數；為了比較濾波前後的共變異數差異，在這裡由雜訊一中擷取出一筆雜訊，必須注意與上述測試使用同一筆雜訊以利分析。濾波後共變異數為 0.0019，比濾波前共變異數 0.003 小很多，結果

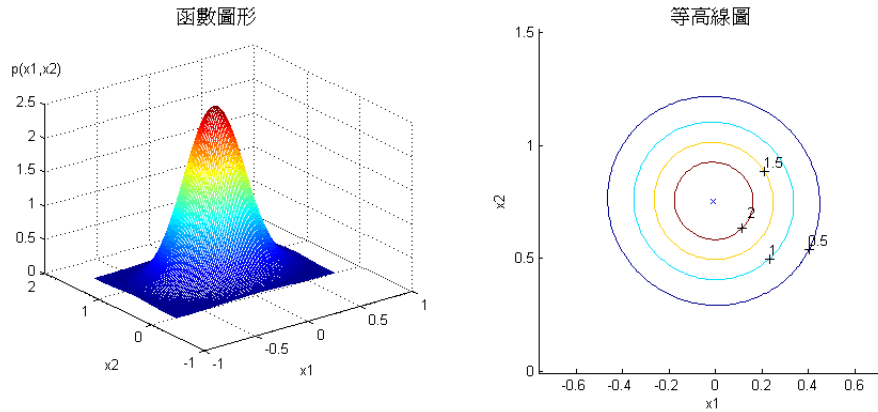


圖 17: 雜訊一在二維空間的函數圖形與等高線圖

發現濾波後獨立性更強，並且更加符合統計假設 iid。

另外，可用必要條件檢驗雜訊一是否符合 iid：假設  $X, Y \sim N(\mu, \sigma)$  並且符合 iid，若  $\alpha, \beta \in \mathfrak{R}^+$  並且  $\alpha^2 + \beta^2 = 1$ ，令  $Z = \alpha X + \beta Y$ ，則  $\sigma_Z = \sigma_X = \sigma_Y$ 。在這裡驗證雜訊一是否符合此必要條件，並且濾波後標準差的相對誤差是否變小；此為控制實驗，所以濾波前後必須使用同一雜訊以比較相對誤差。假設  $X, Y$  均由雜訊一中擷取，持續時間設定 5 秒，可得  $\sigma_X = 0.2525$ 、 $\sigma_Y = 0.2525$ ，並且求得  $\sigma_Z = 0.2518$ ，利用母體標準差  $\sigma = 0.2516$  計算標準差的相對誤差為 0.09%；但雜訊濾波後  $\sigma_X = 0.2524$ 、 $\sigma_Y = 0.2524$ ，並且  $\sigma_Z = 0.2517$ ，計算其相對誤差： $|\frac{\sigma_Z - \sigma}{\sigma}| \times 100\% = 0.02\%$ ；發現雜訊濾波後標準差的相對誤差減少，結果同樣驗證濾波後獨立性更強。



### 3.4 異常訊號的濾波方法

對於時間域上雜訊所呈現的異常突起，我們可以觀察頻率域上的表現；若將造成異常突起的頻率濾掉，會使得時間域上所呈現的分布，更加符合 iid (我們稱此為濾波方法一)。但面臨的問題是：第一，擷取不同的雜訊樣本，異常突起的頻率會有所偏移，皆影響雜訊的分布，針對隨機的雜訊，如何選擇欲濾波的頻率？第二，濾波時，若是將此頻率全濾掉，會擾亂頻率的組成，並且不符合真實情況，所以該減弱此頻率多少能量，何者為減弱頻率能量的標準？是我們這一節所要探討的。

在探討這兩個問題之前，必須先觀察雜訊的組成，首先觀察持續時間設定 0.1 秒，並且隨機選取的八段雜訊(圖 18)，發現 0.1 秒的雜訊段呈現一特殊起伏，但此曲線在隨機的八段雜訊頻譜上，並沒有呈現一致性，異常突起仍特別多。於是，觀察持續時間設定 1 秒的八段雜訊(圖 19)，發現 1 秒的雜訊段呈現一特殊起伏，並且起伏趨勢呈現穩定；由於此雜訊取樣頻率為 20k，若取樣個數越多，表示此雜訊在頻譜的表現越佳。所以在此節討論中，我們皆使用持續 1 秒的雜訊段作討論。

由以上討論中，此雜訊在頻率域上呈現穩定特殊起伏(如圖 20(a) 的曲線 a)，並且對頻譜圖數據點的縱軸取對數，繪製散佈圖如圖 20(b)，發現點圖呈現帶狀散佈，表示此雜訊組成帶有資訊。我們可藉由頻譜上的穩定起伏作為減弱能量的準則，不僅可以減弱異常突起的頻率，也不至於擾亂雜訊的組成，使得雜訊不符合真實訊號；找出濾波的準則曲線時，必須擷取適當的數據點使得曲線擬合較佳，並且使用一函數曲線找出回歸線。濾波準則的曲線即此回歸線，此曲線對應的數值即橫軸對應此頻譜圖的頻率，縱軸對應能量，若訊號頻譜圖部分頻率能量大於此濾波準則，則減弱此頻率的能量。我們稱此濾波程序為濾波方法二，找出濾波準則程序如下：

**step1** 擷取數據點—由於頻譜的特殊起伏，在頻率約 300 Hz 以下大幅上升，然後開始緩慢下降，並且開始有異常突起，我們希望將特殊起伏大幅下降部份，異常突起的頻率濾掉，觀察頻譜在低頻 100 Hz 以下完全沒有異常突起，5000 Hz 以上漸趨平緩，所以數據點擷取範圍設定為 100 Hz ~ 5000 Hz。在其範圍內，將頻率等分 10 等份， $(\mu_i, \sigma_i)$  為落於每等份內的期望值與標準差， $i=1 \dots 10$ ，由於每等份的能量分布為右偏分配，並且避免擷取到能量太大的頻率，所以將能量落於  $[\mu_i+3\sigma_i, \mu_i+4\sigma_i]$  之間的數據點擷取出來，作為曲線擬合的數據點(如圖 21(a))。

**step2** 曲線擬合—我們以兩種函數模擬濾波的準則曲線，第一，以雙曲線的一葉作為濾波的準則：

$$y = \frac{b}{x - a} \quad (3.1)$$

第二，以指數函數作為濾波的準則：

$$y = ae^{-bx} \quad (3.2)$$

其中， $x$  與  $y$  分別為數據點的座標；由 step1 擷取出的數據點，使用上述兩種模型做回歸分析，即由 (3.1) 式得  $a = 290.61$  與  $b = 0.0313$  繪製迴歸模型 1 以及 (3.2) 式得  $a = 0.1084 \times 10^{-3}$  與  $b = 0.7215 \times 10^{-3}$  繪製迴歸模型 2，如圖 21(a) 所示，可以發現 (3.2) 式找出的迴歸模型 2 曲線擬合較佳。另外，觀察擷取出來的數據點並且函數值  $y$  取其對數，可以發現轉換後的數據，可用一直線來模擬趨勢(圖 21(b)，迴歸模型 3)，表示原資料為指數曲線，並且此雜訊帶有資訊，如同圖 20(b) 點圖呈現帶狀。所以在此驗證使用指數函數模擬濾波曲線較為恰當。

由上述程序找出濾波準則後，令找出欲濾波的頻率為  $f_j$  及其能量  $d_j$ ， $j = 1 \cdots N$ ， $N$  為欲濾波的頻率個數；並且令  $\hat{d}_j$  為濾波準則找出的頻率能量；我們建議減弱能量： $(1 - \sqrt{\frac{\hat{d}_j}{d_j}}) \times 100\%$ ， $j = 1 \cdots N$ ；由於本論文討論的頻譜皆在訊號作傅立葉轉換後取絕對值平方，所以上述減弱能量會有根號產生。此方法主要是讓頻率能量減弱後，還能保留適當的能量，並不會因為減弱的能量過大或過低以致擾亂雜訊組成。

觀察濾波前後的直方圖(圖 22)，並且求得正規模型相對誤差以及標準差如表 8，發現使用濾波方法二後，相對誤差下降了 39.83%，卻只比濾波方法一好一些，但標準差下降了 1.52%，比濾波方法一的標準差下降 0.19% 還要多。所以我們可以得知使用濾波方法二後，會使得訊號更加正規，甚至比濾波方法一還要正規，並且標準差下降幅度較大。接下來檢驗訊號使用濾波方法二濾波後獨立性是否更好，從雜訊一中取兩筆持續時間 1 秒的雜訊，取得雜訊程序如 2.1 節闡述；由於濾波後標準差變小，所以我們觀察原訊號以及兩種濾波方法所得的共變異數，發現濾波方法二的共變異數確實更接近零，表示此濾波方式會讓訊號獨立性更好。

濾波比較	原訊號	濾波方法一	濾波方法二
相對誤差	18.10%	11.74%	10.89%
標準差	0.2627	0.2622	0.2587
共變異數	-0.0005	-0.0005	-0.0004

表 8: 濾波分析比較

以上從訊號中擷取一段雜訊，持續時間設定 1 秒，可以發現呈現一致性的特殊起伏，使用濾波方法二將其異常突起的頻率濾掉；若觀察 10 筆隨機選取 1 秒的雜訊，利用此濾波程序，可以找到 10 筆參數  $a$ 、 $b$ ，求其平均值，並繪製濾波準則曲線(圖 20(a) 的曲線 b)，在持續時間為 1 秒中的雜訊段，作為此訊號濾波準則的表現。

另外觀察其他雜訊是否也適用此濾波方式，所以觀察雜訊二在 1 秒的八段雜訊頻譜圖(圖 23)。雜訊一與雜訊二頻譜同樣呈現穩定起伏，由於取得程序如 2.1 節所述，所以我們認定為同一類型的訊號，因此可使用上述濾波方法二來過濾。但針對雜訊三，由於 3.3.1 小節討論此雜訊在時間域與頻率域上的異常突起特別明顯，並不適合作為實驗訊號，所以在此並不針對雜訊三作濾波。由於雜訊四、五、六的紀錄條件是一樣的，並且雜訊四、五的取樣個數太少，所以只觀察雜訊六在頻率域上的表現(圖 24)，但不如雜訊一、二呈現穩定起伏。這可能是因為此雜訊取得方式為生物體麻醉的狀況下獲得，以致接收到雜訊的部份訊號減弱(如遠方神經元以及肌電圖產生的雜訊)，所以並沒有呈現穩定起伏。這表示由此擷取方式(麻醉)獲得的雜訊並不適用濾波方法二。所以，我們可以從訊號的頻率域發現潛藏的資訊，藉由此特性，找出其濾波準則，濾掉頻率域上的異常突起。

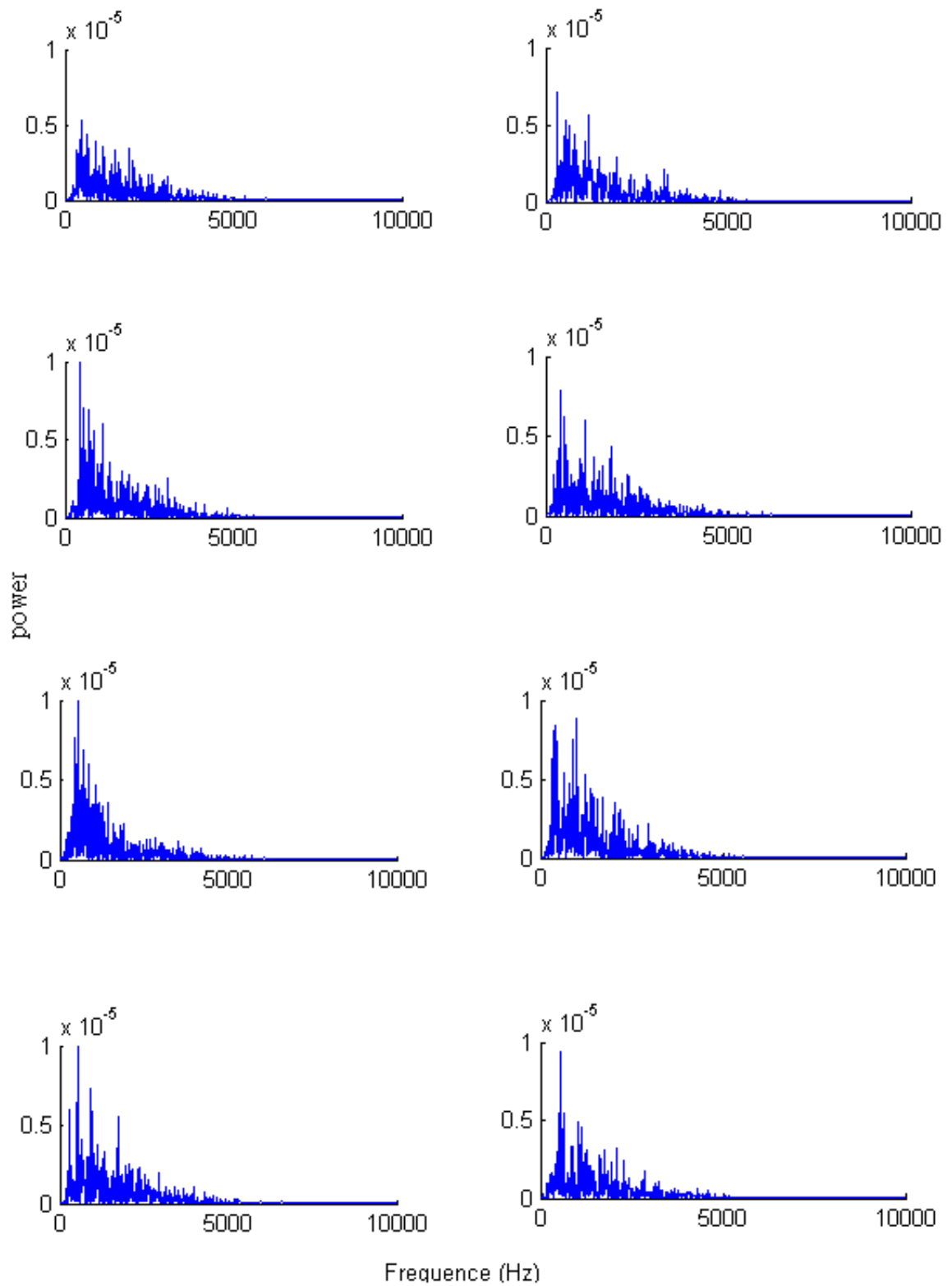


圖 18: 雜訊一中八段 0.1 秒的雜訊頻譜圖

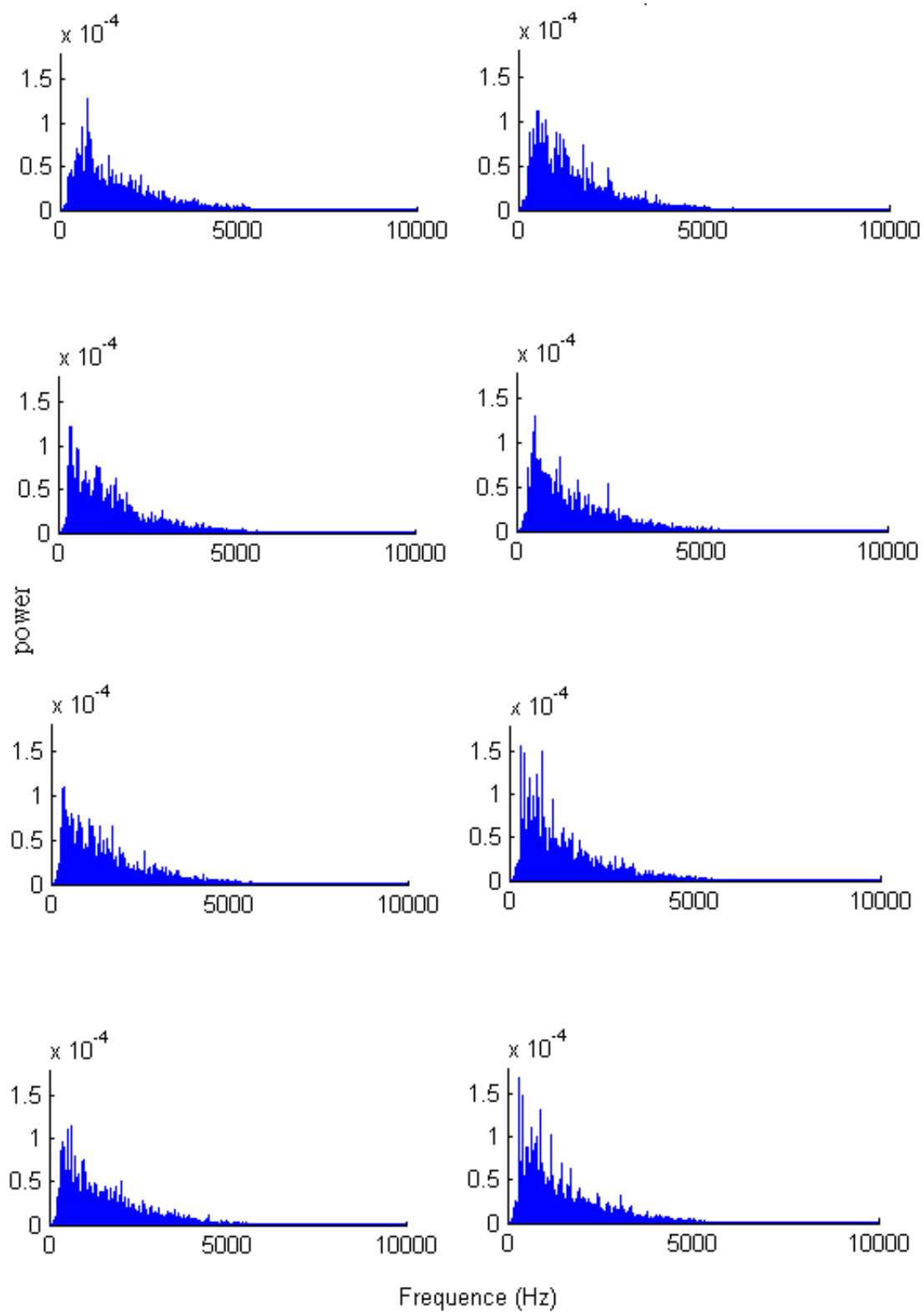


圖 19: 雜訊一中八段 1 秒的雜訊頻譜圖

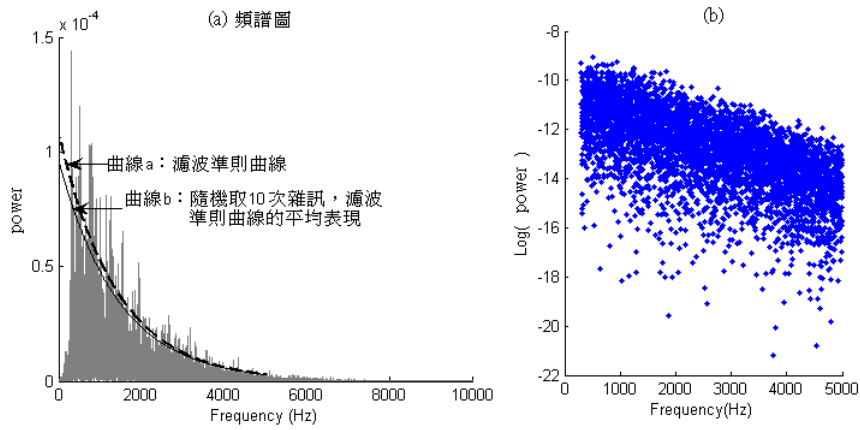


圖 20: 雜訊一的濾波準則曲線

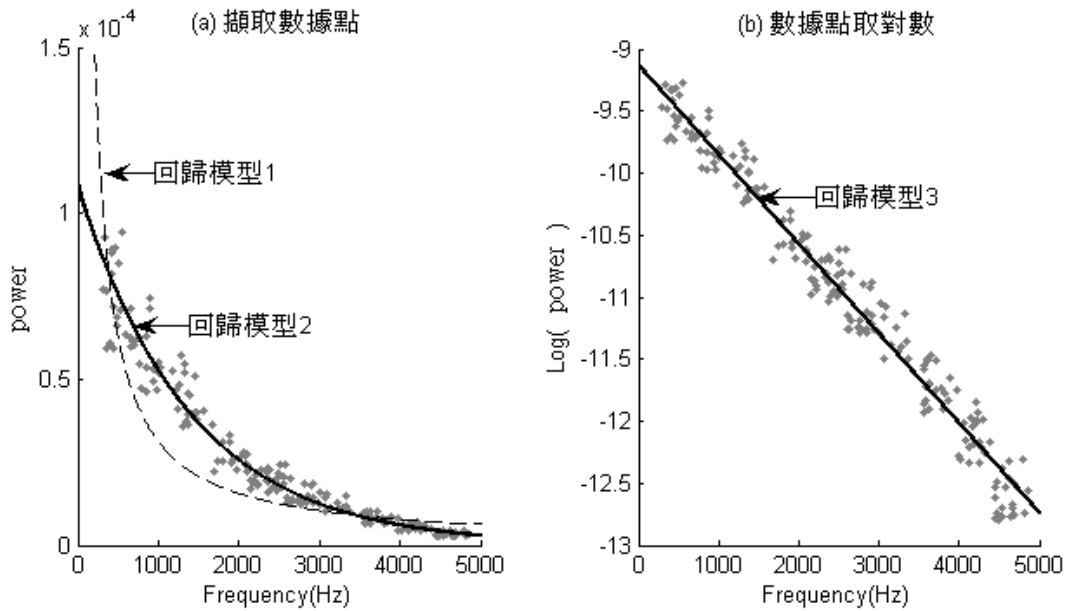


圖 21: 曲線擬合的數據

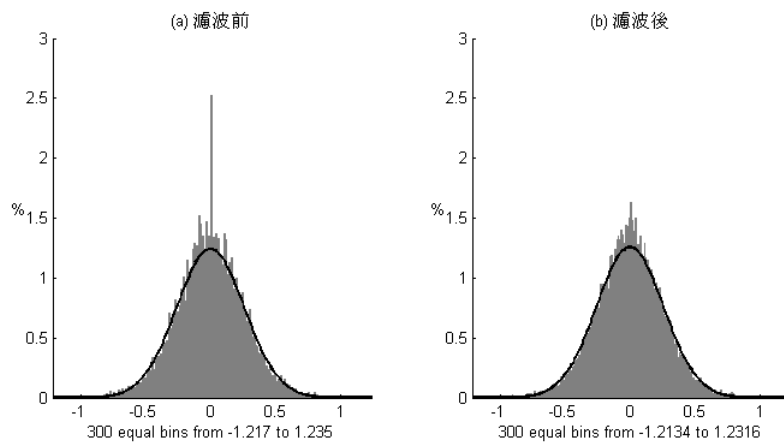


圖 22: 濾波前後的正規模型



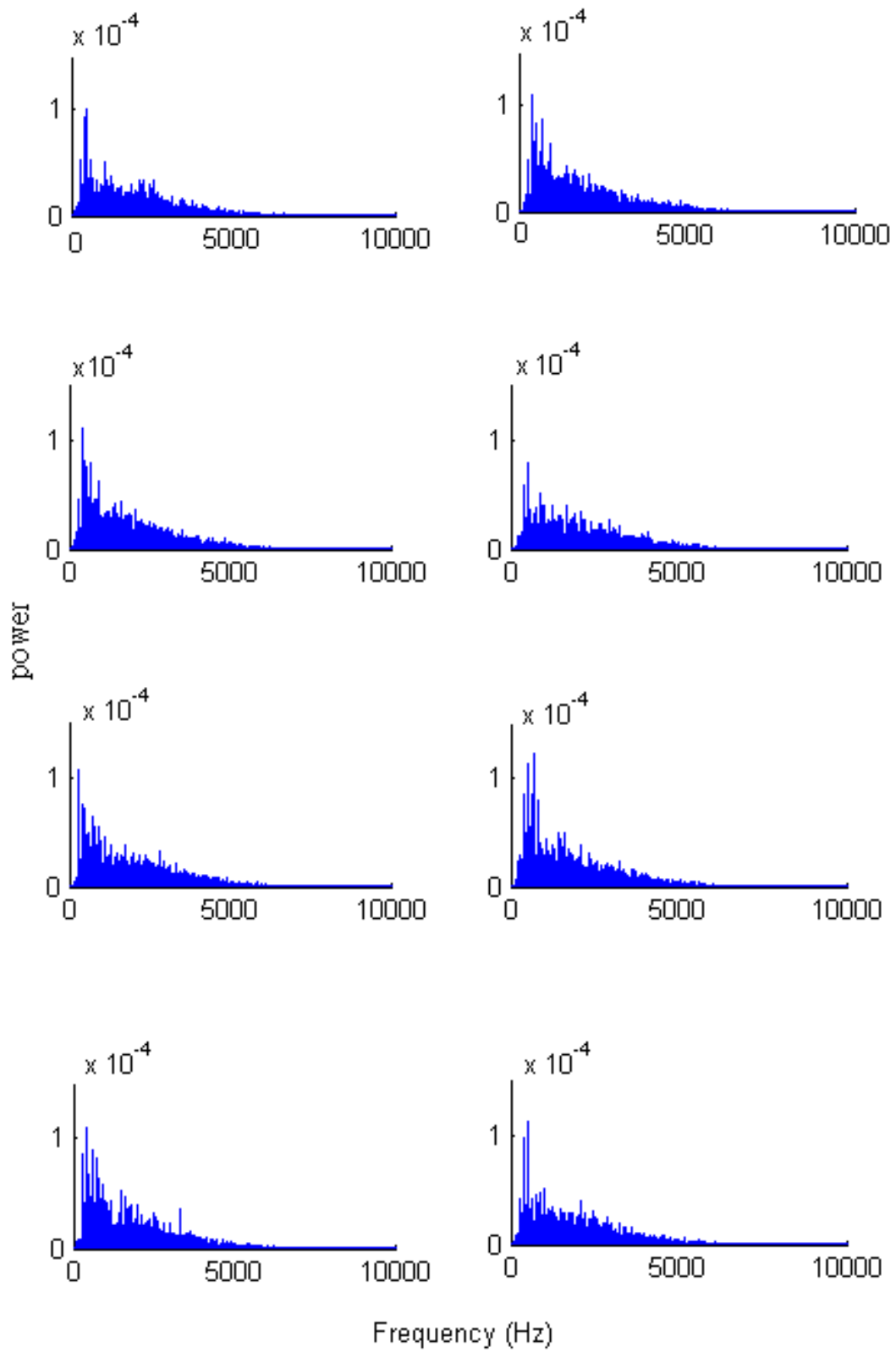


圖 23: 雜訊二中八段 1 秒的雜訊頻譜圖

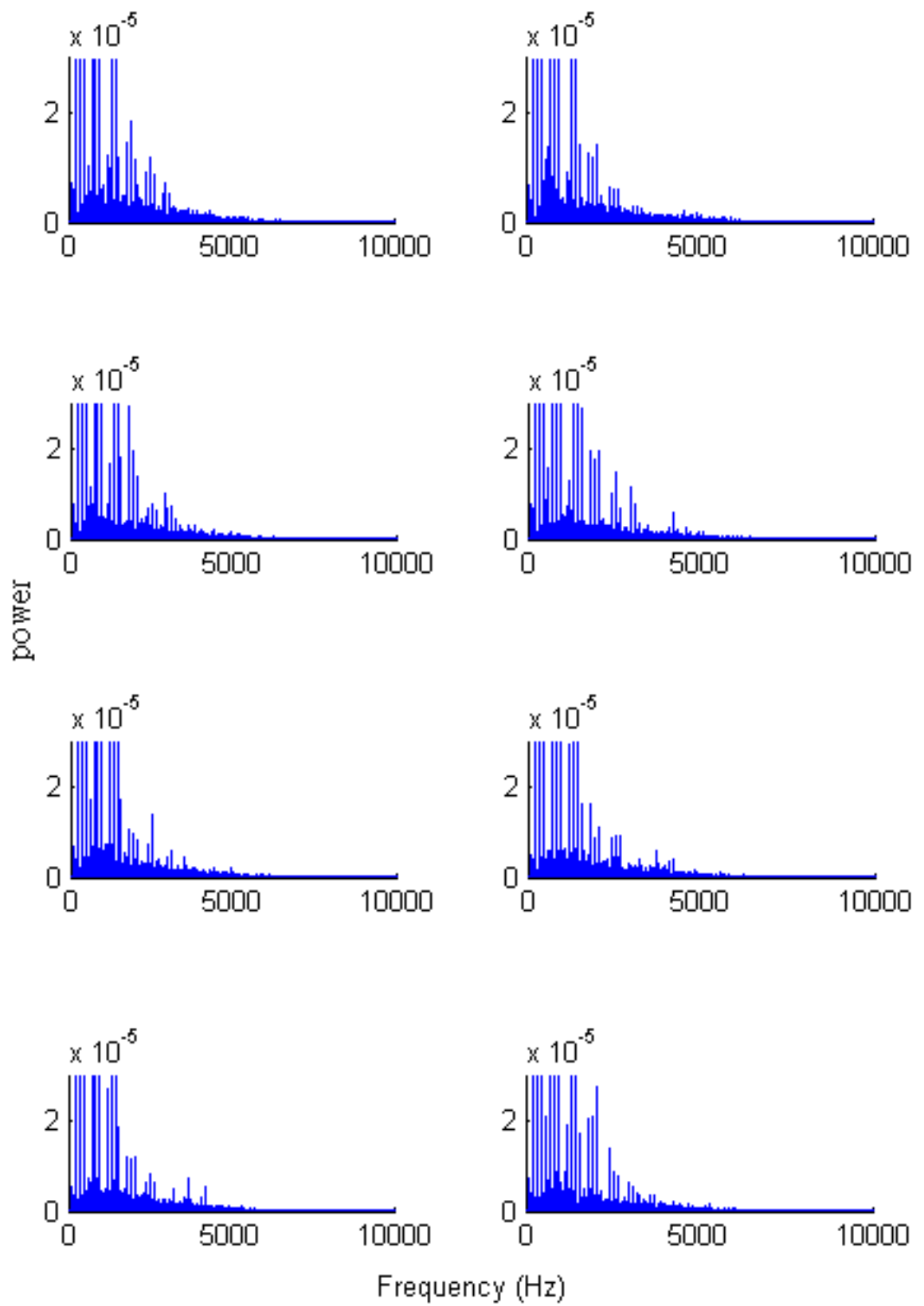


圖 24: 雜訊六中八段 1 秒的雜訊頻譜圖

## 第 4 章 雜訊對神經訊號分類的影響

在第三章中，發現雜訊可藉由濾波使得異常分布更符合正規分布，並且獨立性更好；這一章即將要探討的是雜訊對於神經訊號分類的影響。首先觀察雜訊對動作電位分類的影響，在這裡使用主成分分析觀察其表現，並且利用 K-Means 分群法分類，計算分類錯誤率。利用雜訊在頻率域上的穩定起伏，對神經訊號做濾波，觀察濾波後對分類造成的影響。

### 4.1 主成份分析

在此節中，藉由維度減少的概念，如將樣本投影到最佳投影向量上，來降低分類的複雜度，引發主成份分析的概念介紹之，進而觀察主成份分析在動作電位分類上的表現。

#### 4.1.1 最佳投影向量

假設樣本  $X$  為  $d$  維度空間，樣本數為  $N$ ，並且符合  $X \sim N_d(\mu, \Sigma)$ ；若能夠找到一個向量  $w$  使得  $d$  維度空間的樣本  $X$  投影到向量  $w$  上，使得維度減少為 1 維，並且此向量最能夠解釋此樣本的資訊，那這個向量  $w$  就是最佳投影向量。

首先，必須找到一個  $w$  向量使得  $w^T X \sim N_d(w^T \mu, w^T \Sigma w)$  會符合以下兩點，第一， $\|w\|=1$ ；第二， $w^T \Sigma w = \max_i (w_i^T \Sigma w_i)$ ， $\forall i = 1 \cdots d$ 。其中， $\mu$  為此樣本的期望值向量， $\Sigma$  為共變異數矩陣， $(\lambda_i, w_i)$ ：為  $\Sigma$  的特徵值與特徵向量，因為  $\Sigma = U \Lambda U^T$  為 SVD 矩陣 (Singular Value Decomposition) 且  $U U^T = 1$ ， $\Lambda$  為對角矩陣，對角元素為  $\Sigma$  的特徵值  $\lambda_i$  由大到小的排序，也就是  $\lambda_1 \geq$

$\lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_d$ ， $U$  為所對應的特徵向量矩陣，特徵向量  $w_i$  為其行向量。所以我們可知最佳投影向量  $w$  為  $w_1$ ， $w^T \Sigma w = (w^T U) \Lambda (w^T U)^T = e_1^T \Lambda e_1 = \lambda_1$ ，其中， $w_1^T U = e_1^T$ 、 $e_i = (0, 0, \dots, 1^{i^{th}}, \dots, 0, 0)^T$ 。

假設  $\Sigma$  為對角矩陣，對於對角矩陣而言，SVD 只是重排而已，也就是  $U$  為排列 (permutation) 矩陣。舉例來說， $\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 9 \end{bmatrix}$ ，可以得到  $U =$

$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ ， $U$  為排列矩陣， $\Lambda = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ 。對於特徵向量矩陣  $U$ ，我們可以

選取少數變異大的向量  $w_i$  並且將原樣本  $X$  投影於  $w_i$  上，使得原變數的變異大都集中在少數的主成份向量上，這就是主成份分析的主要想法，也就是說第一組成分為變異最大的特徵向量  $w_1$ ，其次，第二主成分為變異第二大的特徵向量  $w_2$ 。我們於下一節討論主成份分析。

在這裡，我們討論的最佳投影向量  $w$  也就是第一主成分  $w_1$ 。舉例來說， $\Sigma = \begin{bmatrix} 4 & 1.5 \\ 1.5 & 1 \end{bmatrix}$  計算出來的  $U \approx \begin{bmatrix} -0.9239 & -0.3827 \\ -0.3827 & 0.9239 \end{bmatrix}$ ， $\Lambda \approx \begin{bmatrix} 4.6213 & 0 \\ 0 & 0.3787 \end{bmatrix}$ ，也就是所對應的第一主成分  $w = [-0.9239, -0.3827]^T$ 。在這裡，我們可以來做一

數值驗證，也就是驗證  $w$  是否符合  $w^T \Sigma w = \max_i (w_i^T \Sigma w_i)$ ， $\forall i = 1 \dots d$ 。將一單位圓沿著圓弧切割成 720 等分，可以得到方向向量  $v_j = [\cos(\phi_j), \sin(\phi_j)]^T$ ， $\phi_j = \frac{2\pi}{720} \times j$ ， $j = 1 \dots 720$ ，觀察  $v_j^T \Sigma v_j$  的最大值，並且找到對應的方向向量  $v \approx [-0.9239, -0.3827]^T$ 。我們可以發現找到的  $v$  與最佳投影向量接近，也就是說方向向量  $v$  相當於第一主成分  $w_1$ ，就是我們找到的最佳投影向量  $w$ 。如圖 25，橫軸為角度  $\phi$  的變化，縱軸為  $v^T \Sigma v$  的變化，我們可以發現在  $\phi$  為 3.5343 時，最大值約為 4.6213，同時，可以發現此圖形呈現週期函數圖，另外有一角度的值與最大值幾乎一樣， $\phi$  為 0.3927，其值約為 4.6213，其方向向量  $\hat{v} \approx$

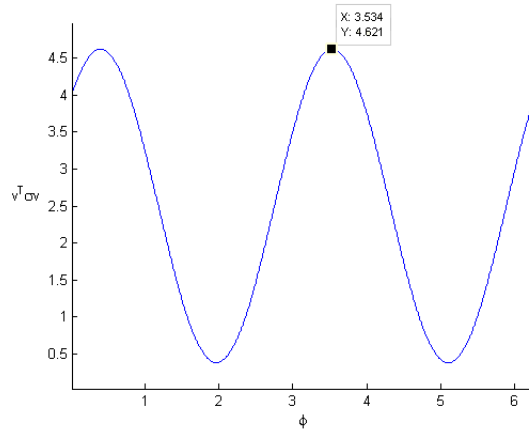


圖 25: 最佳投影向量的數值驗證

$[0.9239, 0.3827]^T$  與向量  $v$  相反，而且其  $\phi$  值相差了  $\pi$ ，這解釋了所計算的最大組成分向量有正負兩個方向性的選擇，也就是  $\Sigma$  的特徵值所對應的特徵向量有正負兩個方向性的選擇。

在 [5] 中，要在 32 維的向量裡，找到標準差最大的那一維，也就是 32 維的訊號投影到一個特殊的  $w$  方向上，將 32 維度減少成 1 維。所以，在 [5] 中，由於 32 維彼此互相獨立，必須找到其中一個維度使得  $\sigma_k$  最大， $k = 1 \dots 32$ ，並且我們發現  $\sigma_{21}$  最大，相當於  $w = e_{21}$  (對於對角矩陣而言，SVD 只是重排)，則可以說  $w = e_{21}$  為最佳投影向量。

### 4.1.2 PCA 簡介

我們使用維度減少來降低欲分類的複雜度，在這裡討論主成份分析 (Principal Components Analysis; PCA)。PCA 是將一組有相關的變數轉成無相關的獨立變數，新的獨立變數為原變數的線性組合，使得原變數的變異大都集中在少數的新變數上，選取少數變異大的新變數，捨棄變異小的變數，以精簡複雜的變數。進一步可用於因素分析 (Factor Analysis; FA) 及群集分析 (Clustering) 等。也就是將原來有相關的變數轉成少數無相關的變數，配合如因素分析、群集分析，對於預測的準確性有很大的幫助。相關理論可參考 [5]。

如何選擇選取少數的變異大的新變數，可以使用解釋變異的百分比來考量。也就是假設對應主成分的特徵值為  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_i \geq \dots \geq \lambda_d$ ，所以第  $i$  主成份解釋變異的百分比為  $\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_i + \dots + \lambda_d} \times 100\%$  以及累積至第  $i$  主成份解釋變異的百分比為  $\frac{\lambda_1 + \lambda_2 + \dots + \lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_i + \dots + \lambda_d} \times 100\%$ ，其中  $d$  為原變數的維度。以下三點為決定主成份個數的方法：

1. 保留比平均特徵值大的成份。
2. 在所有特徵值中，保留 80% 或 90% 以上解釋變異的特徵值所對應的特徵向量。
3. 做一特徵值大小與特徵值個數的圖解，以及特徵值的累積解釋變異圖，觀察並決定適當的主成份個數。

但在本論文中，為了資料的描述方便，使用第一主成分以及第二主成分繪製 PCA 點圖以利分析。

#### 4.1.3 PCA 分類動作電位

針對主成份分析，可用 PCA 點圖來重新描述原資料。因為前兩個主成份解釋了變異數的最大百分比，我們可以將其兩主成份用 PCA 點圖來表示，也就是第一主成份作橫軸，第二組成份作縱軸，將數據在二維空間做圖。假如前兩個主成份的解釋能力夠好，那這張 PCA 點圖比原始資料給了樣本更好的描述。實驗中所使用的波型為圖 2 的動作電位原始波形，在這裡我們設定訊雜比為 1.5，也就是說，波形訊號標準差為 0.2630，雜訊標準差為 0.2516，我們的雜訊乘上 0.7 會符合此訊雜比。兩種不同波型的動作電位疊加雜訊一千次各有一千筆動作電位，所得兩千筆動作電位即為欲分類的動作電位。計算這些動作電位的主成份後，使用兩主成份(通常用第一主成份及第二主成份)用於分類。所以一個動作電位表示成一個二維的向量，因此我們有兩千個二維的向量，這兩

千筆動作電位分別表示平面上兩千個點，也就是用 PCA 點圖來表示。

將使用 PCA 點圖測試動作電位分類結果。我們測試以下四個例子，並且由表 9 中觀察第一主成份與第二主成份的累積解釋變異百分比。例一：測試 spike1 及 spike2 的分類情形(圖 26)，圖 26(a) 為 spike1 及 spike2 動作電位各疊加雜訊一千次的結果，圖 26(b) 為 PCA 點圖分類情形，圖 26(c) 為每一主成份向量的解釋變異百分比，圖 26(d) 為累積至第  $i$  主成份解釋變異的百分比，發現這兩類動作電位幾乎看成一類；由於 PCA 點圖使用了前兩主成份，計算累積至前兩主成份的解釋變異百分比為 32%，表示此 PCA 點圖解釋了原資料的 32%。例二：測試 spike1 及 spike3 的分類情形(圖 27)；例三：測試 spike3 及 spike4 的分類情形(圖 28)。例二與例三雷同，雖然 PCA 點圖分類可以大致分類出兩類動作電位，但這兩類動作電位有一部分是重疊不容易分清楚的，遇到這種情況我們通常會將圖畫上一條線，明顯的將點圖分成兩類，這兩類就是不同的動作電位，但這操作並不客觀，分類結果隨著操作者不同而不同。例四：測試 spike2 及 spike4 的分類情形(圖 29)，這兩類動作電位只有少部分重疊，可以很清楚分類出兩類動作電位。

另外，我們發現例二 PCA 點圖解釋了原資料的 33% 的變異，只比例一解釋了原資料的 32% 的變異稍微好一些，但 PCA 點圖呈現出來卻與例三解釋了原資料 35% 變異的 PCA 點圖雷同，為什麼有如此差異？由於 4.2.1 小節提到樣本投影到向量  $w$  上，降低維度後觀察散布情況，也就是  $w^T X \sim N_d(w^T \mu, w^T \Sigma w)$ ，其中變異大小為  $w^T \Sigma w$ ，假如投影到  $w_1$  上，變異大小則為  $w_1^T \Sigma w_1 = (w_1^T U) \Lambda (w_1^T U)^T = e_1^T \Lambda e_1 = \lambda_1$ ；假如投影到  $w_2$  上，變異大小則為  $w_2^T \Sigma w_2 = (w_2^T U) \Lambda (w_2^T U)^T = e_2^T \Lambda e_2 = \lambda_2$ ；所以，假如第一主成份  $w_1$  以及第二主成份  $w_2$  的變異相差越大，呈現在 PCA 點圖的分類情況會越好。可由表 9 中觀察上述四個例子的第一組成份及第二組成份的變異差距百分比  $(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \dots + \lambda_i + \dots + \lambda_d} \times 100\%)$ ，發現例一變異差距百分比約 2.32% 為最小，所以兩類動作電位幾乎視成一類；例二及例三變異差距百分比約為 7% ~ 8% 左右，



所以兩類動作電位有部份重疊，但還是可以分成兩類；例四變異差距百分比接近 20%，所以非常明顯分類出兩類動作電位。

第一與第二主成份	例一	例二	例三	例四
累積解釋變異百分比	32 %	33 %	35 %	44 %
變異差距百分比	2.32 %	7.21 %	8.39 %	19.84 %

表 9: 第一組成份及第二組成份的累積變異與變異差距百分比

我們觀察動作電位平移對 PCA 圖的影響，如例五：我們將一樣的波形完全平移，也就是最低點沒有對齊，可以發現一樣的波形平移後，在 PCA 點圖上分類會明顯分出兩群(圖 30)，解釋變異高達 42%。所以 PCA 對時間的平移非常敏感。一般來說，我們在做動作電位分類時，會將最低點對齊，以避免此現象發生。

另外計算出累積變異百分比高達 80% 時，所對應有幾個主成份，也就是需要幾個主成份可以重新描述原資料的 80%，如例一需要前九個主成份，例二、例三與例四需要八個主成份。發現可以將原本 32 維度的資料減少為 10 維度以下的資料，並且對原本的資料解釋了 80% 的變異，我們可以將此應用於集群分析 (Clustering)，以利分類的準確性。

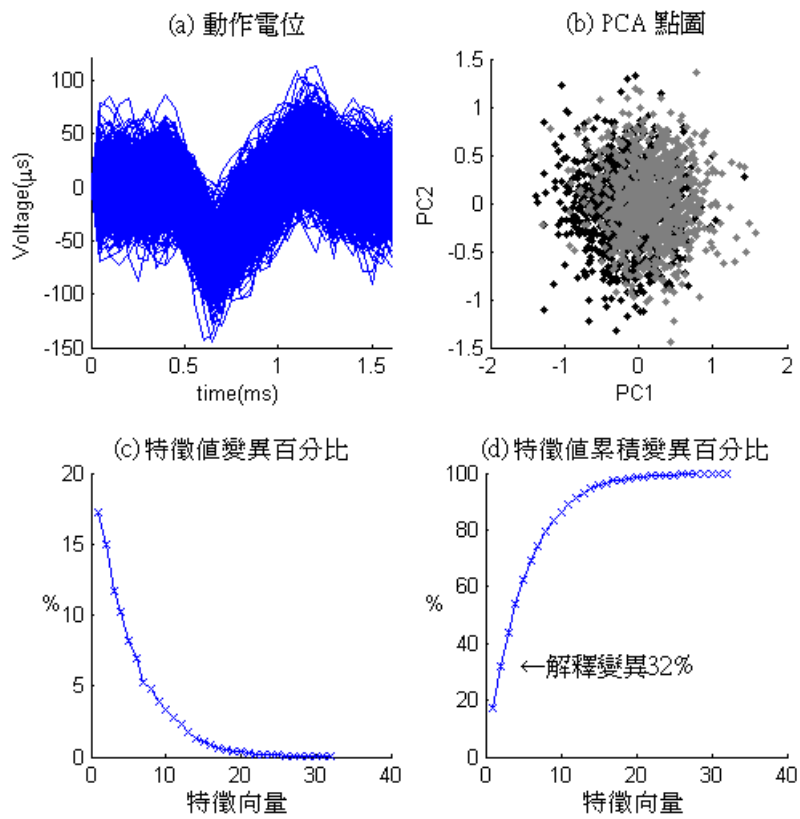


圖 26: spike1 和 spike2 PCA 分析動作電位

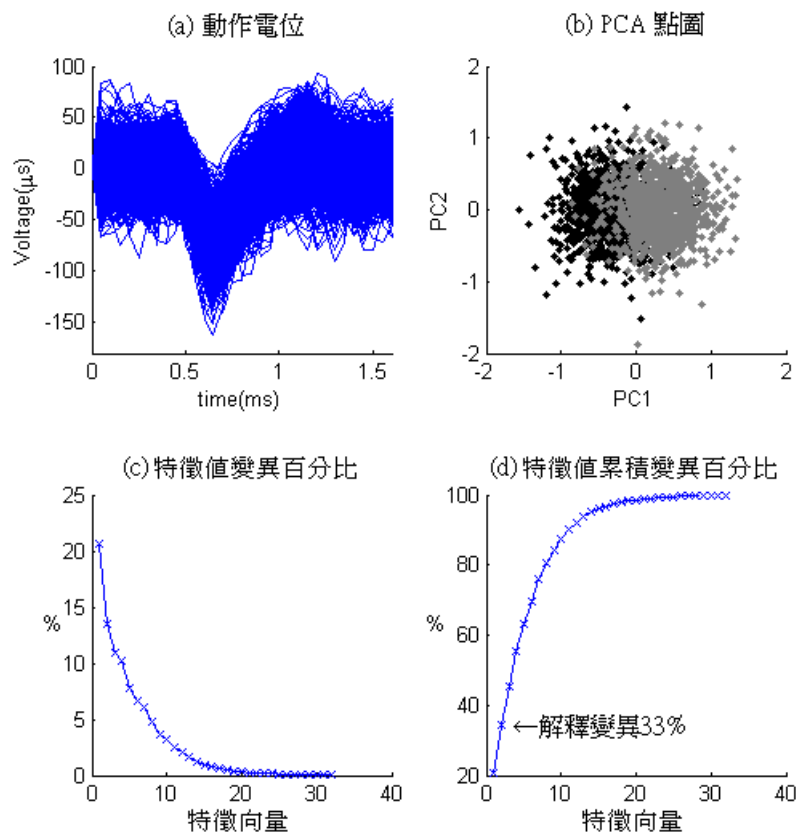


圖 27: spike1 和 spike3 PCA 分析動作電位

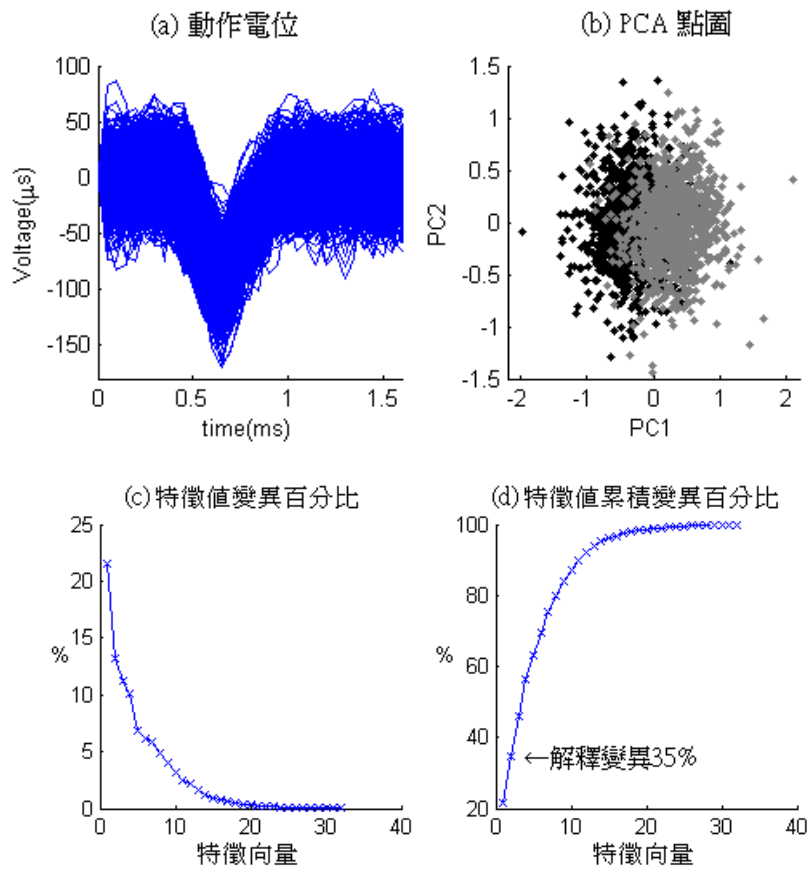


圖 28: spike3 和 spike4 PCA 分析動作電位

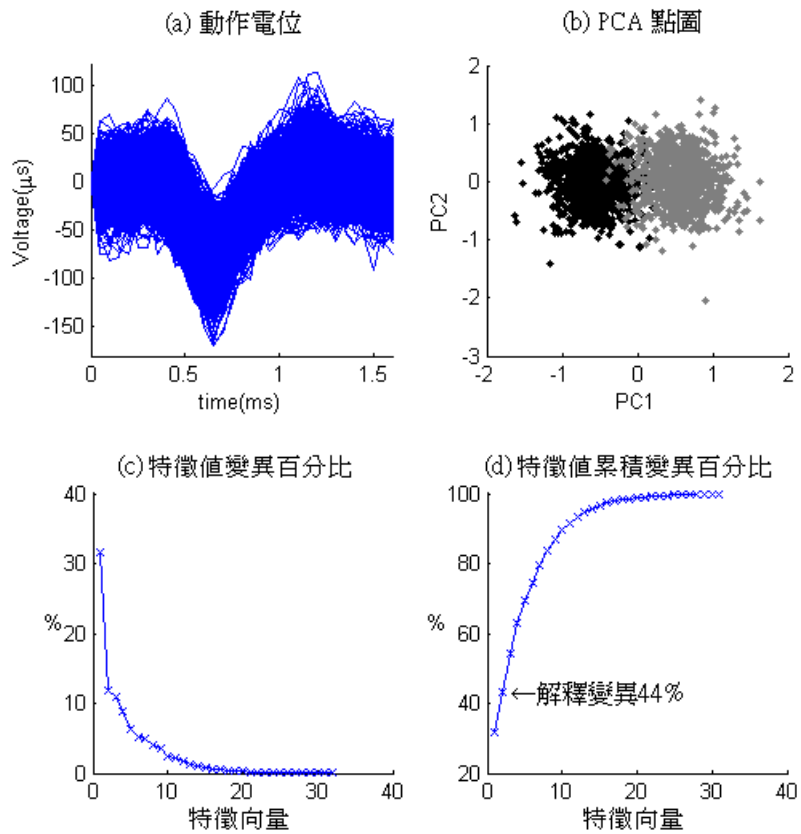


圖 29: spike2 和 spike4 PCA 分析動作電位

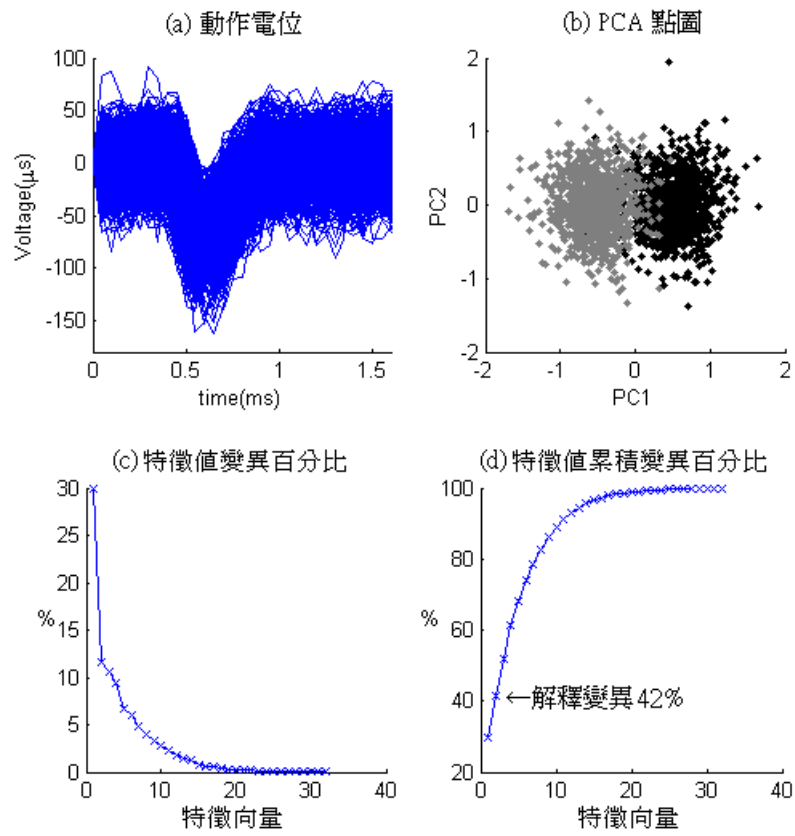


圖 30: 動作電位平移的 PCA 分析動作電位

## 4.2 K-Means 分群法

這一節即將探討 PCA 在集群分析上的應用。其中，我們要介紹的是 K-Means 分群法，並且使用一數值驗證來觀察此演算法的特性，以及 PCA 分類動作電位上的應用。

### 4.2.1 K-Means 簡介

我們接收資料資料，並不是單純的常態分配，而是混合的高斯分布模型，如 PCA 演算法所呈現，第一主成分與第二主成分所得到的資料即為混合資料高斯分布模型，所以單單使用高斯機率分布是不夠的。於是我們可以利用 EM 演算法(Expectation Maximization Algorithm)，使用混合高斯分布模型，對樣本進行分類，其中 K-Means 為 EM 演算法的特例。EM 演算法使用機率方法來決定資料點存在於群集內的機率。但針對 K-Means 演算法會使用距離來指派資料點至它最接近的群集，也就是資料點存在於群集內的機率不是 0 就是 1。但此章節並不討論 EM 演算法，我們針對 K-Means 演算法進行討論。

K-Means 分群法為分割式分群法的一種，必須先指定群聚的數目  $K$ ，經過反覆的迭代運算，逐一降低誤差函數的值。假設給定  $K$  個初始中心  $m_i$ ， $i = 1 \dots K$ ，以及給定樣本  $X = \{x_t\}_{t=1}^N$ ， $N$  為樣本數，對於任意的樣本  $x_t$  可以找到與此樣本點最接近的初始中心  $m_j$ ， $j \in 1 \dots K$ ，使得  $x_t$  與  $m_j$  的距離比  $x_t$  與其他中心的距離更接近，也就是  $\|x_t - m_j\| = \min_{i=1}^K \|x_t - m_i\|$ ，我們就說樣本點  $x_t$  屬於第  $j$  群；首先隨機選定初始中心(圖 31(a) □ 標記所示)，針對每一樣本點可以找到距離最近的初始中心並且找到歸屬的群聚，由此可藉由定義標籤  $b_i^t$  來決定第  $x_t$  個樣本點是否屬於第  $i$  群。假如第  $x_t$  個樣本點屬於第  $i$  群，則  $b_i^t=1$ ，假如第  $x_t$  個樣本點不屬於第  $i$  群，則  $b_i^t = 0$ ；也就是說假如  $\|x_t - m_j\| = \min_{i=1}^K \|x_t - m_i\|$ ，則  $b_i^t = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ 。定義誤差函數為  $E(\{m_i\}_{i=1}^K | X) = \sum_t^N \sum_{i=1}^K b_i^t \|x_t - m_i\|^2$ ，最好的中心向量為誤差函數的最小

化，所以得  $m_i = \frac{\sum_{t=1}^N b_i^t x_t}{\sum_{t=1}^N b_i^t}$  表示為下一步迭代中心(圖 31，□ 標記每次迭代的中心)。由於  $m_i$  改變所以標籤  $b_i^t$  也隨之改變，在此並不能解最佳化的問題，必須用迭代程序來解最佳化問題，使目標函數逐次變小。對於  $i = 1 \dots K$ ，迭代程序為

1. 隨機選擇初始向量  $m_i$ ，對於全部的樣本  $X = \{x_t\}_{t=1}^N$  估計標籤  $b_i^t$ 。
2. 估計完標籤就可以計算新的  $m_i$ ，標籤  $b_i^t$  就會改變，需要重新計算  $m_i$ 。

重覆這兩步迭代程序直到中心向量穩定為止，也就是迭代程序找出來的中心向量會逐次收斂。

此演算法最後中心向量  $m_i$  與初始中心有高度的相關，針對初始中心有以下幾個選擇，第一，隨機選擇初始向量  $m_i$ ；第二，先估計全部樣本的中心，由中心加上  $K$  個小的隨機向量，作為初始中心  $m_i$ ；第三，將樣本所對應的區域切割成  $K$  群，然後計算這  $K$  群的中心做為初始值。在此論文中，選擇第二個方法做為初始中心的選擇。由於 K-Means 使用平均值來計算中心點，很容易受到離群值 (outliers) 的影響，使得中心偏移。所以我們可以先忽略離群值，決定群數以及執行演算法時，不考慮離群值的狀況下做分類，分類後再決定離群值的群組。

離群值決定方式：由於我們知道正規分布樣本有 99.7% 落在  $[\mu - 3\sigma, \mu + 3\sigma]$  之間，計算樣本  $X$  的期望值  $\mu$  以及標準差  $\sigma$ ，將  $[\mu - 3.5\sigma, \mu + 3.5\sigma]$  以外的樣本都視為離群值，也就是距中心  $\pm 3.5\sigma$  以外的樣本即視為離群值。所以在執行 K-Means 演算法時，分別計算樣本點投影於  $x$  軸及  $y$  軸的離群值，用上述尋找離群值的方式(圖 32，○ 即為離群值)，先剔除離群值然後應用 K-Means 演算法，再尋找離群值的歸屬群組，發現原本的演算法分類錯誤率為 13.5% 降低為 12%，可見 K-Means 演算法容易受到離群值的影響。爾後，我們提到的 K-Means 演算法：先遵循上述找出離群值的方式忽略離群值，然後執行此演算法，再尋找離群值的歸屬群組。

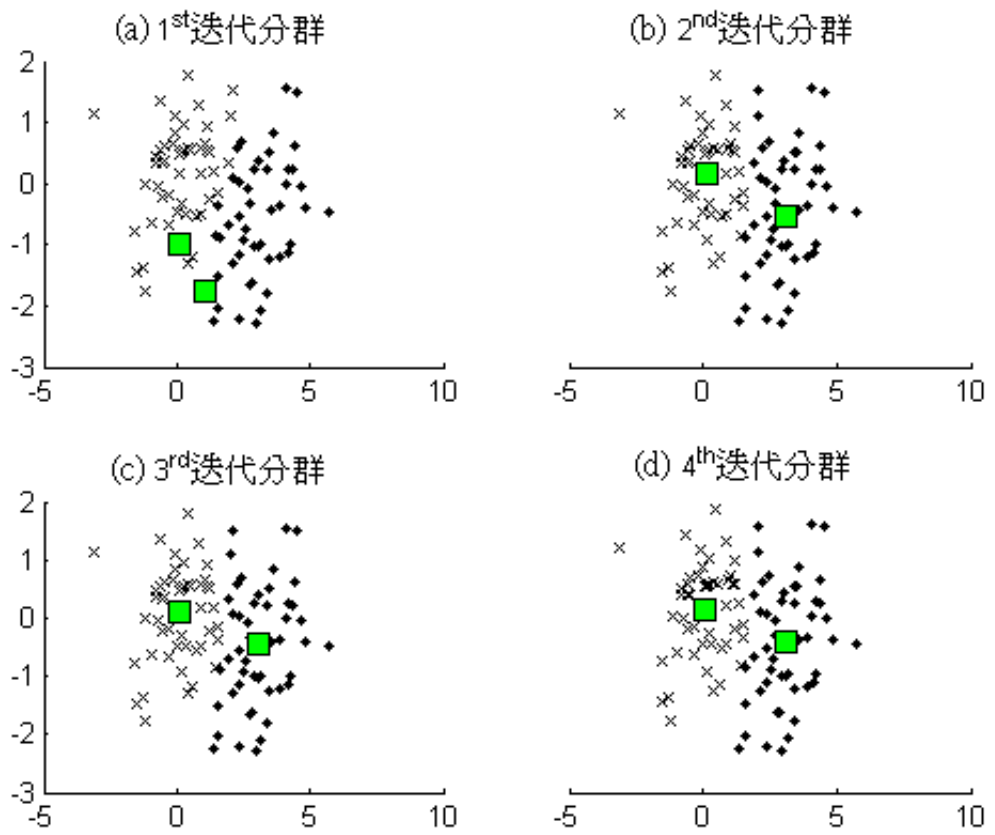


圖 31: K-Means 中心迭代圖

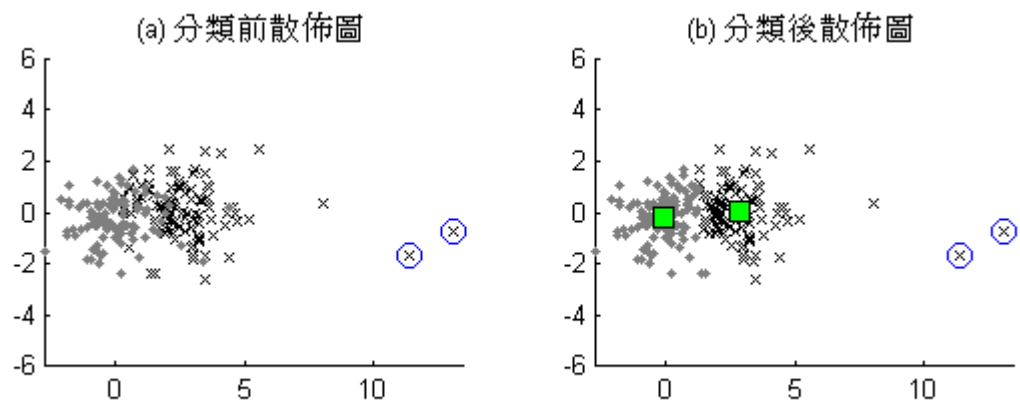


圖 32: 離群值

### 4.2.2 K-Means 的數值驗證

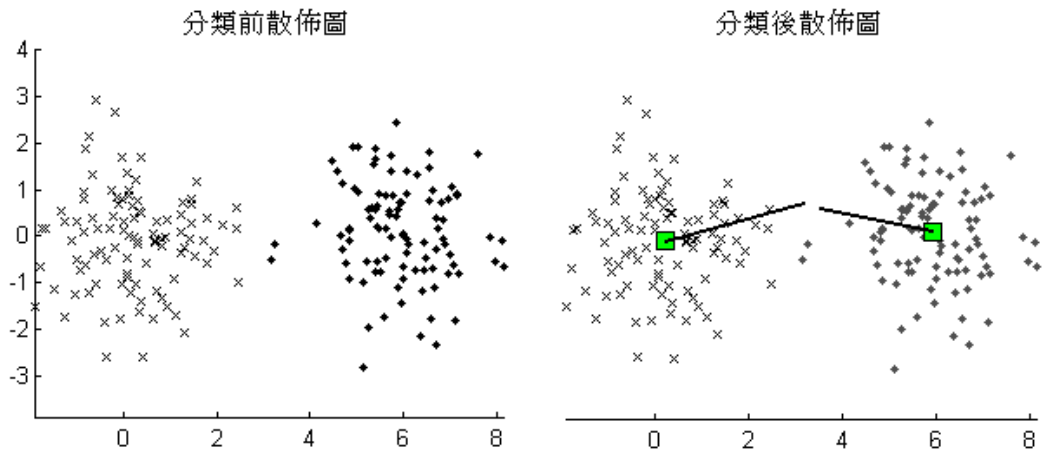
針對 K-Means 做一數值模擬，給定已知兩群，測試當兩群中心接近時，觀察其分類情形。初始中心設定為比  $r$  小的隨機向量，其中  $r$  為樣本中心與全部樣本點最大距離的 0.75 倍。我們假設兩群樣本 A、B 均來自正規分布  $N(\mu_i, \sigma)$ ， $\sigma = 1$ ；觀察兩組樣本中心越接近時，所分類的狀況。首先觀察中心距離為  $6\sigma$  的分類狀況，以  $\sigma$  為一間距，逐次觀察中心距離為  $5\sigma$ 、 $4\sigma$ 、 $3\sigma$ 、 $2\sigma$  以及  $\sigma$  的分類狀況(圖 33 與圖 34，分類後散佈圖迭代路徑如圖直線 '—' 所示)。並且計算其迭代次數、迭代中心以及分類錯誤率如下表 10。

兩群中心距	$6\sigma$	$5\sigma$	$4\sigma$
兩群中心	(0,0) (6,0)	(0,0) (5,0)	(0,0) (4,0)
迭代中心	(0.20,-0.08) (5.92,0.12)	(-0.01,-0.08) (5.02,0.01)	(-0.04,-0.07) (4.12,-0.04)
迭代次數	2	2	4
分類錯誤率	0	0.5%	2.5%
兩群中心距	$3\sigma$	$2\sigma$	$\sigma$
兩群中心	(0,0) (3,0)	(0,0) (2,0)	(0,0) (1,0)
迭代中心	(0.04,-0.02) (3.11,0.17)	(-0.52,0.05) (1.93,-0.04)	(-0.36,0.37) (1.34,-0.39)
迭代次數	6	6	7
分類錯誤率	9%	17%	35%

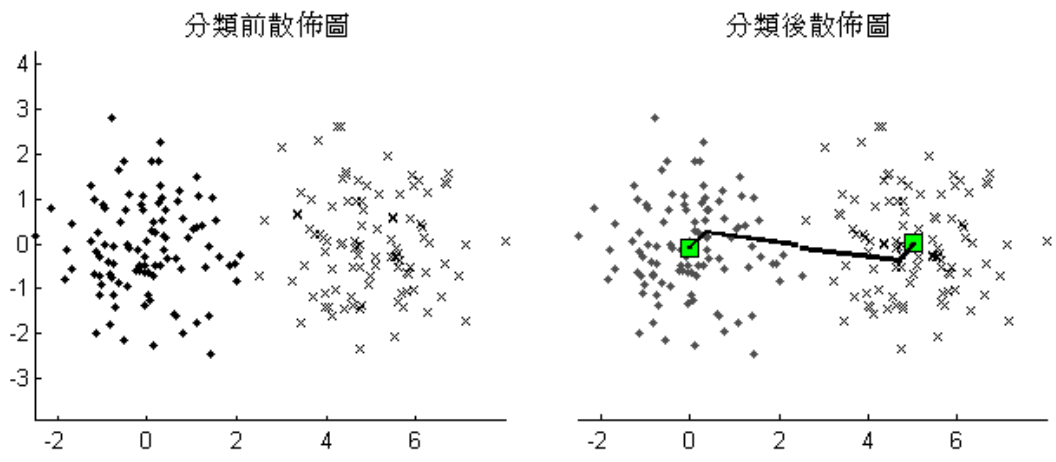
表 10: K-Means 的數值分析



(a) 中心距  $6\sigma$



(b) 中心距  $5\sigma$



(c) 中心距  $4\sigma$

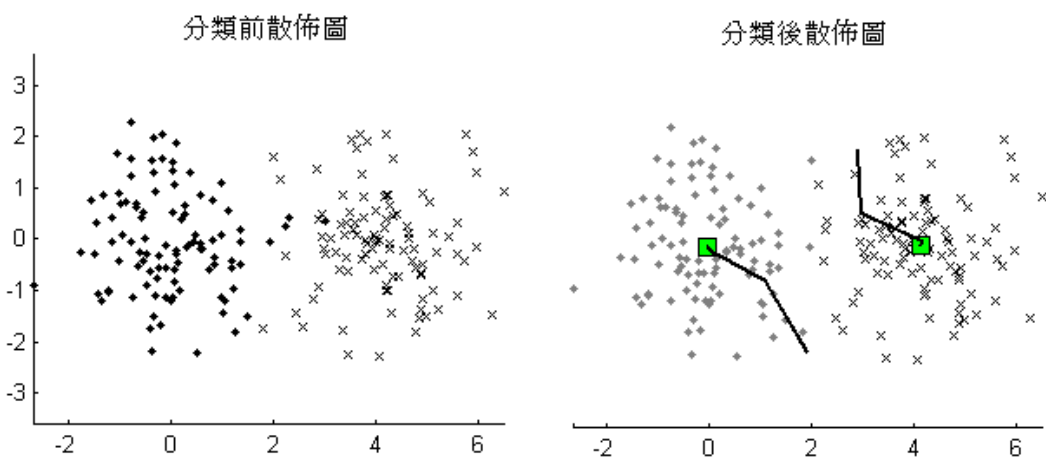
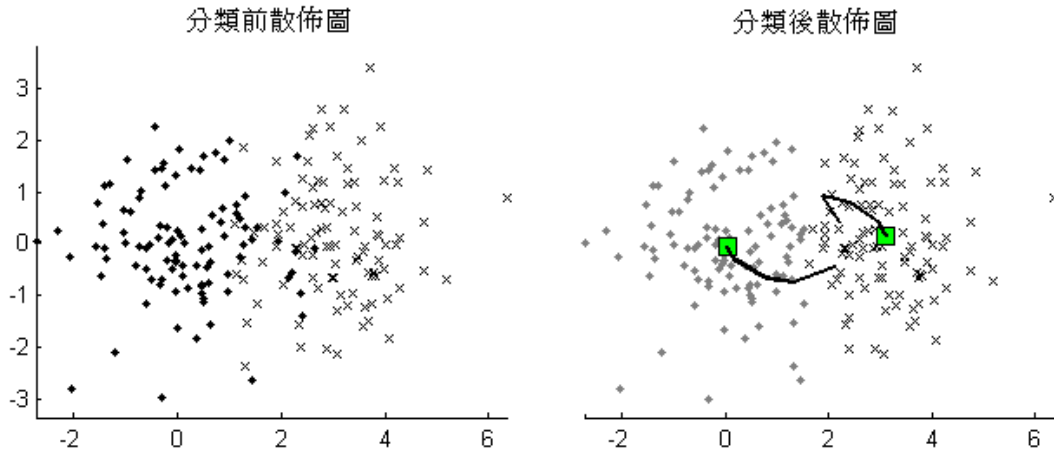
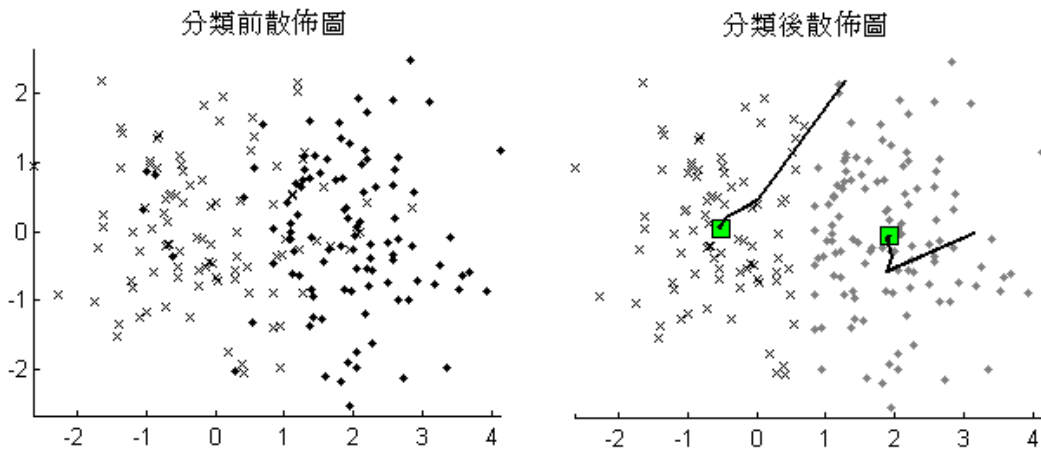


圖 33: K-Means 的散佈圖分析 I

(d) 中心距  $3\sigma$



(e) 中心距  $2\sigma$



(f) 中心距  $\sigma$

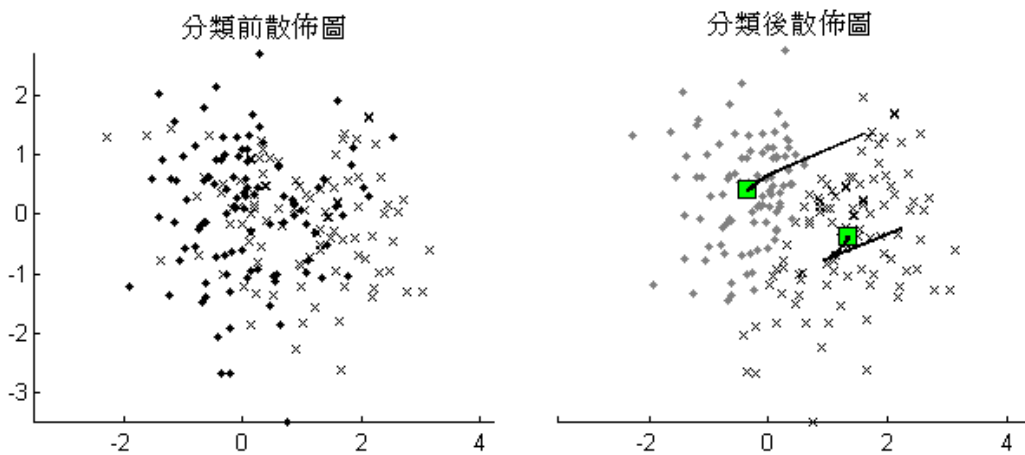


圖 34: K-Means 的散佈圖分析 II

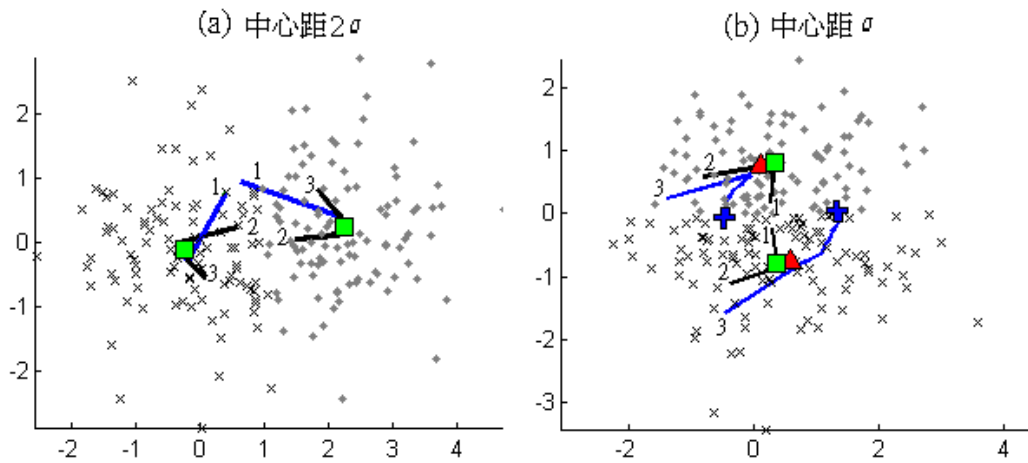


圖 35:  $\sigma$  與  $2\sigma$  中心距的收斂情形

重複上述實驗測試 50 次，計算其平均分類錯誤率，如表 11。發現在中心距離為  $2\sigma$  以上時，每次執行演算法的最後迭代中心幾乎都會收斂至同一點附近；如圖 35，為相同的樣本點，分群數為 2，執行三次演算法的收斂情形，編號 1、2、3 表示執行三次演算法的迭代路徑。圖 35(a) 設定兩群中心距為  $2\sigma$ ，發現兩群樣本執行三次演算法都收斂到同一收斂中心；圖 35(b) 設定兩群中心距為  $\sigma$ ，發現在中心距離為  $\sigma$  時，執行三次演算法的最後迭代中心不收斂。所以，我們可以在  $[\sigma, 2\sigma]$  區內，找出此演算法不收斂的範圍，發現在中心距約在  $1.6\sigma$  以下，收斂中心就會開始不穩定，也就是每次執行演算法的最後迭代中心不一定會收斂至同一點附近。

由以上兩樣本來自「正規分布」的數值實驗，可以觀察出 K-Means 演算法對於重疊度很高的樣本，是非常不穩定的，並且發現在中心距為  $1.6\sigma$  時，收斂中心開始不穩定。

兩群中心距	$6\sigma$	$5\sigma$	$4\sigma$	$3\sigma$	$2\sigma$	$\sigma$
平均分類錯誤率	0.15%	0.5%	2.47%	6.82%	16.41%	34%

表 11: K-Means 的平均分類錯誤率

### 4.2.3 應用 K-Means 於 PCA

我們將 K-Means 演算法應用於 PCA 上，探討利用 PCA 分類動作電位時，分類是否可靠？在這裡我們模擬兩種不同波型隨機出現的兩千筆 spike 疊加雜訊後，使用 PCA 分類動作電位，再應用 K-Means 於 PCA 點圖做分群(圖 36)，這裡的實驗波形對應 4.1.3 的例一：spike1 及 spike2，例二：spike1 及 spike3，例三：spike3 及 spike4，例四：spike2 及 spike4，並且計算分類錯誤率(表 12)，觀察分類是否可靠。

範例	例一	例二	例三	例四
分類錯誤率	33.05 %	19.25 %	15.85 %	3.35 %

表 12: 應用 K-Means 於 PCA 點圖的分類錯誤率

將此實驗與上一節的數值模擬比較，發現例一的分類錯誤率高達 31.35%，與 K-Means 數值模擬兩群中心距相差  $\sigma$  的分類錯誤率 34% 接近，代表例一的實驗顯示：利用 K-Means 執行此試驗，演算法已經相當不穩定；在 PCA 演算法中，此兩波型 spike1 及 spike2 被視為非常相似。另外，例三的分類錯誤率 16.50%，與 K-Means 數值模擬兩群中心距相差  $2\sigma$  的分類錯誤率 16.41% 接近，表示此演算法相當穩定，並且 PCA 演算法中能夠區分 spike3 以及 spike4。同樣的，例三與例二雷同，但例二的分類錯誤率稍高，同樣也可藉由 PCA 演算法來分類。在例四中的分類錯誤率很小為 3.05%，表示此兩個波型特徵非常不同，很容易藉由 PCA 分類出來。在此，可以呼應先前 4.2.3 提到：第一主成分及第二主成分若變異相差越大，則分類越明顯，如表 9 為此兩主成分的變異差距百分比。同樣發現，例一：變異相差最小，所以分類越不明顯；例四：變異相差最大，所以最明顯可以區分出兩群；例二及例三雷同，但例三變異相差比較大，所以例三比例二明顯區分出兩群。

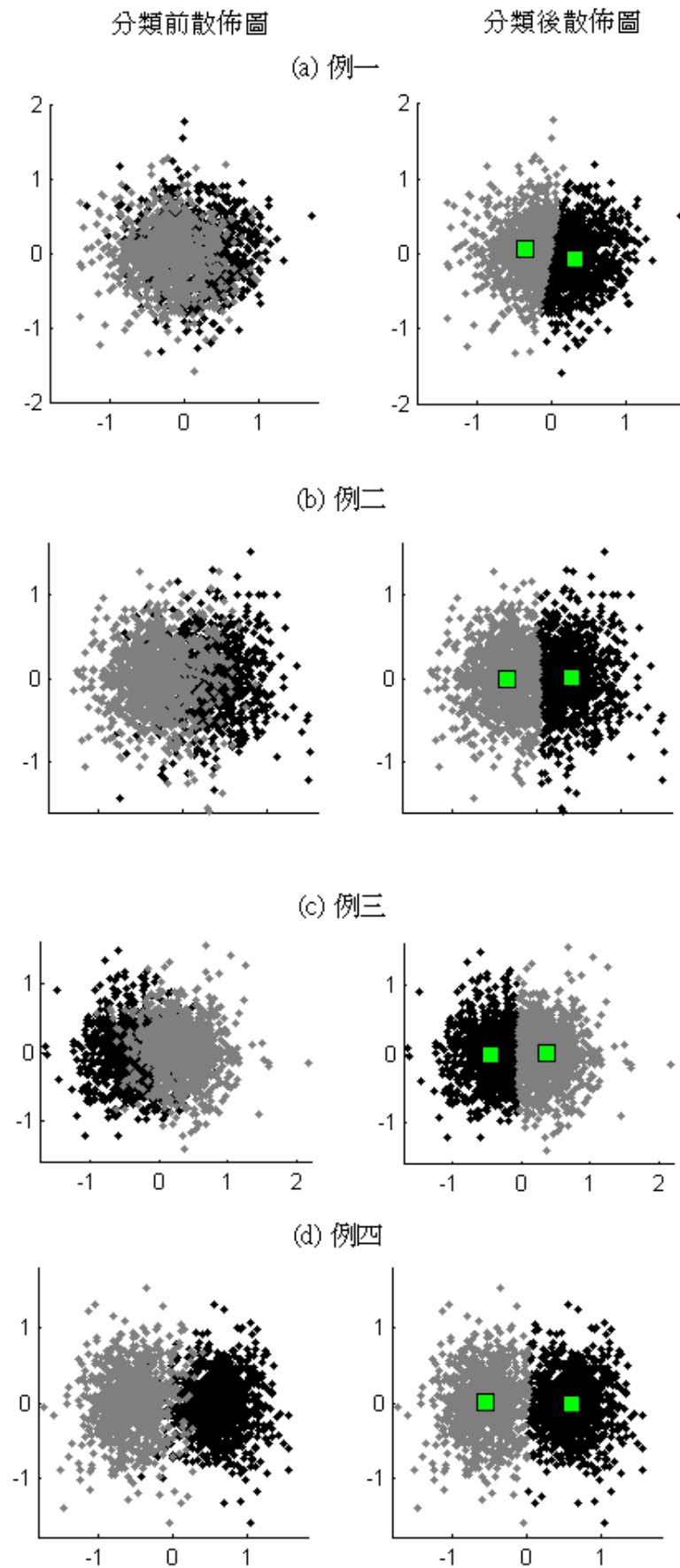


圖 36: 應用 K-Means 於 PCA 的範例

## 4.3 神經訊號分類

此節我們針對神經訊號提出新的濾波方式，觀察濾波前後對於神經訊號分類演算法的影響，比較其分類錯誤率。若以雜訊三作為模擬實驗訊號，發現此雜訊接受到不尋常的低頻，影響分類演算法的效能。

### 4.3.1 濾波前後對分類演算法的比較

在第二章，我們在時間域上發現有些雜訊的分布有「中央突起」，於是第三章我們觀察到這些異常突起，都是在頻率域上某部份頻率的異常突起所造成，若將這些頻率能量減弱，會使得雜訊的分布正規性更好。但是我們想知道的是「減弱雜訊組成的頻率」是否會造成分類演算法的失效？或者是讓分類演算法效能更好，降低分類錯誤率。

在動作電位分類實驗中，經由探測動作電位，挑選出 spike 於分類實驗中工作；每一筆 spike 均為 32 個資料點。將挑選出來的每一筆 spike 依序排列，作為分類程式輸入的數值。我們的工作就是將挑選出的 spike 濾波，觀察濾波對分類演算法的影響。但真實的神經訊號為雜訊疊加上 spike，當然不能只對雜訊做濾波，也就是雜訊連同 spike 一起做濾波，會不會使得分類程式看不出有 spike？於是我們觀察動作電位在頻率域上的表現。

觀察動作電位在頻率域上的表現(圖 37，在這裡只展現 spike1 與 spike2 兩種動作電位在頻率域上的表現)，為了模擬真實的神經訊號動作電位分類，此圖製作方式由兩種動作電位隨機產生，為 2000 筆 32 維的動作電位在頻率域上的表現，圖 37(a) 為 2000 筆動作電位的頻譜圖，圖 37(b) 為局部放大頻譜圖，圖 37(c) 為產生此動作電位的雜訊在頻率域上的表現，由於真實的神經訊號疊加了 spike，所以圖 37(a) 與 (b) 強烈突起的頻率，表示造成 spike 特徵的頻率，以及圖 37(b) 雜訊部份的起伏略高，所以在圖 37(d) 觀察曲線 a (模擬神經訊號的雜訊趨勢)以及曲線 b (模擬雜訊的趨勢)。曲線 a 與曲線 b 可視為濾

波準則的輔助，可以用來幫助濾波以及分類。

濾波程序如下：首先，找出濾波準則曲線 a 以及曲線 b。濾波準則如 3.4 節闡述的濾波方法二，不同的是 step1 — 擷取數據點。我們針對此段的神經訊號雜訊做學習，擷取雜訊程序如 2.1 節闡述，觀察雜訊的在時間域以及頻率域上的趨勢。曲線 a 為模擬神經訊號的雜訊趨勢，由於神經訊號疊加了 spike，以至於神經訊號雜訊趨勢會比原本雜訊趨勢略高(曲線 b)，所以擷取數據點時，在欲濾波的頻率段分割成十等分，若取每等分中的最大值作為模擬曲線 a 的數據點；若將每等分中  $[\mu_i + \alpha\sigma_i, \mu_i + \beta\sigma_i]$  之間的數據點擷取出來，作為模擬曲線 b 的數據點，其中， $(\mu_i, \sigma_i)$  為落於每等份內的期望值與標準差， $\beta = \alpha + 1$ ， $\alpha$  可由使用者設定，但  $\alpha$  值設定越小，雜訊濾波減弱頻率越多。本論文在 3.4 探討的濾波方法二，就是  $\alpha = 3$  的曲線 b 作為濾波的準則。step2 — 曲線擬合，與 3.4 節闡述同，並且同樣可找多次曲線的平均參數，作為濾波準則的曲線 a 以及曲線 b。其次，將落於曲線 a 與曲線 b 之間的頻率減弱至曲線 b。我們稱此濾波程序為濾波方法三，並且觀察濾波後是否提高分類演算法的效能。

在這裡，我們延用上一節的例子，以雜訊一作為實驗雜訊測試，模擬 2000 筆 32 維的動作電位，例一：spike1 及 spike2，例二：spike1 及 spike3 及，例三：spike3 及 spike4，例四：spike2 及 spike4，應用 PCA 分類動作電位，執行 K-Means 分群法，觀察濾波前後分類錯誤率。首先繪製上述例一所模擬真實神經訊號的頻譜圖(圖 37)，以便觀察濾波對波形特徵(圖 38)的影響。模擬時，針對隨機產生的 spike 並且疊加雜訊，並繪製其濾波曲線 a 以及不同  $\alpha$  值的曲線 b (圖 39)。觀察比較濾波前後不同  $\alpha$  值的分類錯誤率(表 13)，以及將其濾波前後分類錯誤率變化以直方圖 40(a) 來表示，此圖橫軸顯示四個例子，以及每個例子不同  $\alpha$  值(依序排列為  $\alpha = 3、2、1、0、-1、-2$ )，縱軸顯示為濾波前後分類錯誤率變化，也就是濾波前與濾波後分類錯誤率變化之差。

由於  $\alpha$  越小表示減弱雜訊的組成頻率能量越多，並且發現此  $\alpha = -1$ ，雜訊幾乎呈現 white noise 時，分類較佳。由於  $\alpha = -2$  時，雜訊幾乎濾完了，但發現分類錯誤率並沒有降低反而增加，所以並不是減弱越多雜訊( $\alpha$  越小)，分類就會較好。我們通常設定  $\alpha = -1 \sim 3$ ，但隨著擷取自不同生物體的雜訊，頻譜的能量起伏不同，相對的  $\alpha$  的設定也有所不同，使用此濾波方式須先觀察雜訊的頻譜。另外，發現濾波後例二與例三分類錯誤率明顯下降，表示此濾波方法三對於 PCA 點圖上有部分重疊的動作電位，有明顯的成效。但因為 K-Means 對重疊度很高的樣本辨識度很低，所以像例一分類錯誤率很高，濾波後雖然有改善，但也效果不大。其次，觀察濾波後例二的分類錯誤率反而比例三低，表示濾波後例二波型特徵更加明顯；反觀動作電位波形(圖 38)，發現例二波形特徵差異性比例三大，亦可驗證上述發現。

雜訊一	分類錯誤率	例一	例二	例三	例四
$\alpha = 3$	濾波前	31.45 %	18.95 %	17.15 %	3.15 %
	濾波後	31.05 %	18.40 %	16.85 %	2.85 %
$\alpha = 2$	濾波前	32.80 %	17.90 %	16.95 %	2.85 %
	濾波後	33.85 %	16.50 %	16.95 %	2.40 %
$\alpha = 1$	濾波前	33.00 %	19.60 %	17.10 %	3.25 %
	濾波後	31.75 %	16.90 %	16.25 %	2.75 %
$\alpha = 0$	濾波前	32.40 %	19.00 %	17.30 %	3.35 %
	濾波後	31.95 %	13.35 %	14.45 %	2.75 %
$\alpha = -1$	濾波前	33.90 %	18.35 %	16.50 %	3.25 %
	濾波後	30.60 %	8.7 %	11.25 %	3.85 %
$\alpha = -2$	濾波前	32.35 %	18.25 %	16.80 %	3.60 %
	濾波後	43.15 %	43.50 %	47.80 %	32.10 %

表 13: 雜訊一濾波前後分類錯誤率

接著測試以雜訊三作為實驗訊號，同樣測試了  $\alpha = -2 \sim 3$  的濾波前後分類錯誤率(表 14 與圖 40(b))。發現  $\alpha = 2$  時，濾波後分類錯誤率稍低，但錯誤率明顯比雜訊一的分類實驗大，並且其他  $\alpha$  的分類狀況似乎都不好。同樣回應了 3.2.3 小節的問題二，這組被當作雜訊的實驗訊號，可能並不理想。雖然濾波會使得  $\alpha = 2$  雜訊三分類實驗錯誤率下降一點，但同樣也不盡理想。另外，



由於此濾波曲線的擬合方式為擷取數據點的回歸曲線，若是雜訊在頻率域上有能量很大的頻率(像雜訊三)，只要此頻率能量高於濾波準則的範圍內，此濾波方式就不會減弱此頻率能量。像上述雜訊三的分類實驗，由於 3.3.1 小節發現此雜訊有一能量極大的低頻，所以我們在濾波時並不會濾到此頻率，以致於濾波對雜訊三分類實驗，並不會造成明顯影響。

雜訊三	分類錯誤率	例一	例二	例三	例四
$\alpha = 3$	濾波前	47.90 %	32.20 %	32.45 %	15.30 %
	濾波後	47.55 %	32.55 %	31.65 %	15.30 %
$\alpha = 2$	濾波前	48.40 %	32.85 %	34.05 %	14.90 %
	濾波後	47.45 %	32.60 %	33.90 %	14.50 %
$\alpha = 1$	濾波前	49.80 %	30.65 %	33.65 %	13.60 %
	濾波後	50.20 %	31.10 %	33.20 %	14.35 %
$\alpha = 0$	濾波前	46.80 %	31.45 %	33.05 %	13.10 %
	濾波後	48.15 %	32.65 %	33.00 %	17.10 %
$\alpha = -1$	濾波前	50.40 %	32.45 %	33.30 %	13.65 %
	濾波後	50.75 %	39.90 %	41.20 %	30.80 %
$\alpha = -2$	濾波前	49.05 %	30.20 %	31.80 %	13.95 %
	濾波後	50.30 %	44.10 %	45.15 %	41.45 %

表 14: 雜訊三濾波前後分類錯誤率

在上述兩個雜訊實驗中，都可以發現並不是雜訊濾掉越多( $\alpha$ 越小)越好。若雜訊濾掉越多，應該更突顯 spike 的特色，實際上不然。所以我們觀察隨機選取 2000 筆兩種 spike (均不疊加雜訊)，在頻率域上的表現(圖 41)。在這裡使用例一作爲實驗測試，發現 spike 的特徵，除了許多強烈突起的頻率外，還有一些能量較低的頻率(如圖 41 曲線標記所示)。這些能量較低的頻率，在圖 37 中容易被視爲雜訊。所以若是雜訊率掉太多(如 $\alpha=-2$ )，可能會將影響 spike 的特徵濾掉，以致影響到分類的結果。

於是，我們得知在做此類生物實驗的時候，對於擷取出來的神經訊號，必須將訊號分成兩部分，第一部份針對神經訊號擷取雜訊，觀察雜訊的分布是否有異常突起，並且可以在頻率域上觀察頻譜圖是否有特別突出的頻率，若此雜訊有異常發現，則表示可能擷取訊號有問題或者是剛好接收到異常的電磁波，此筆訊號用於分類演算法可能遺失一些資訊。若模擬實驗中，有使用到雜訊的需求，亦可使用濾波方法二以致雜訊分布更加符合正規模型。第二部份則是經由雜訊在第一部分的學習階段利用雜訊在頻率域上呈現的穩定性，使用濾波方法三將神經訊號組成雜訊的頻率減弱，再進行分類演算法，使得有部分重疊的動作電位分的較開。

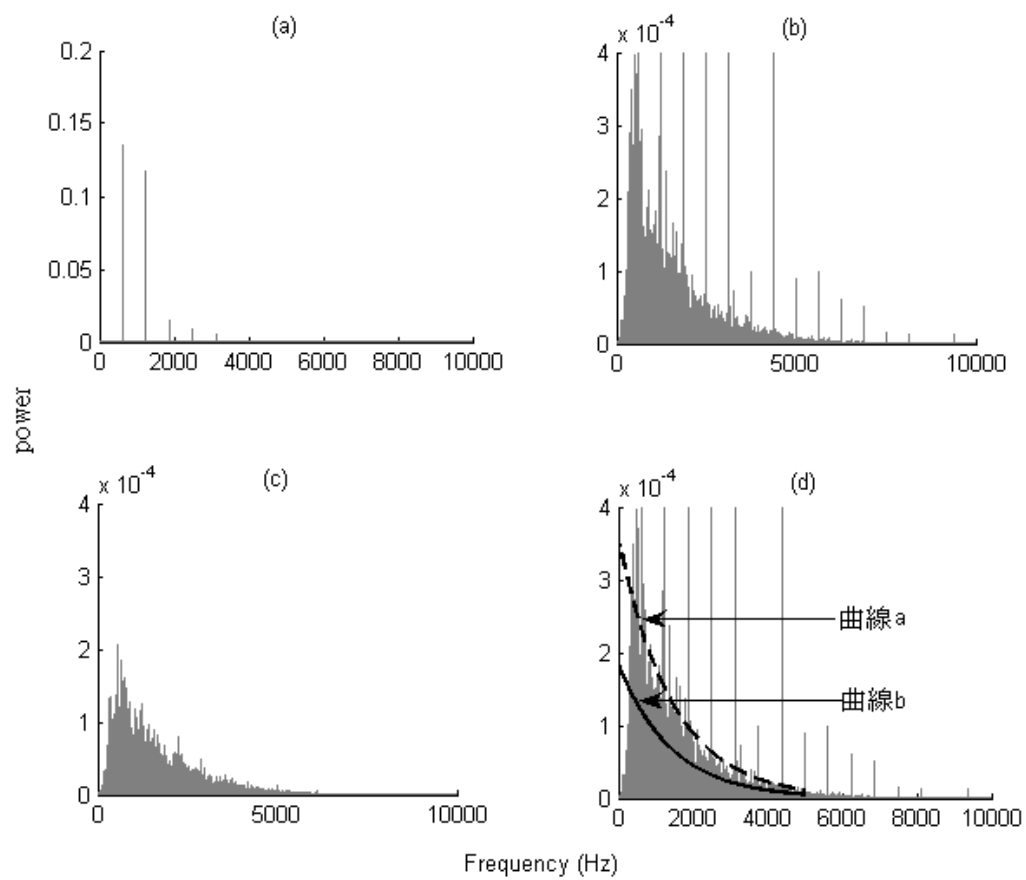


圖 37: 神經訊號於頻率域上的表現

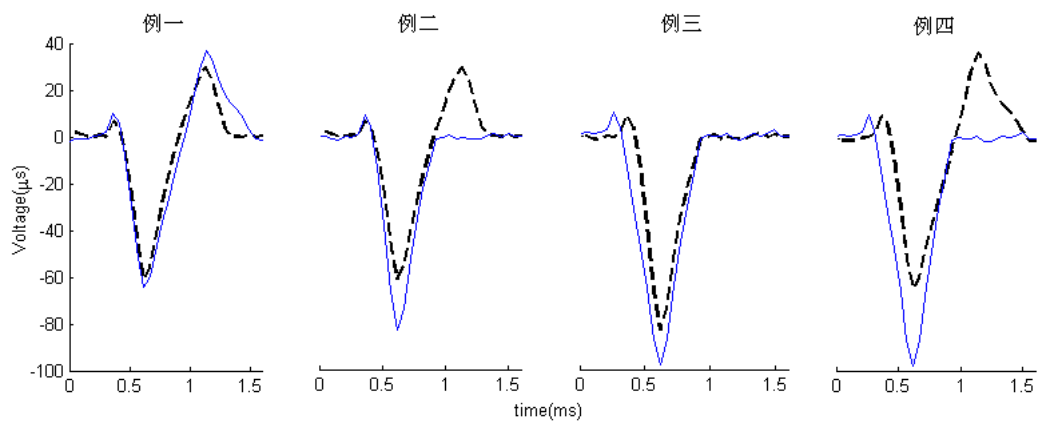


圖 38: 動作電位的原始波形

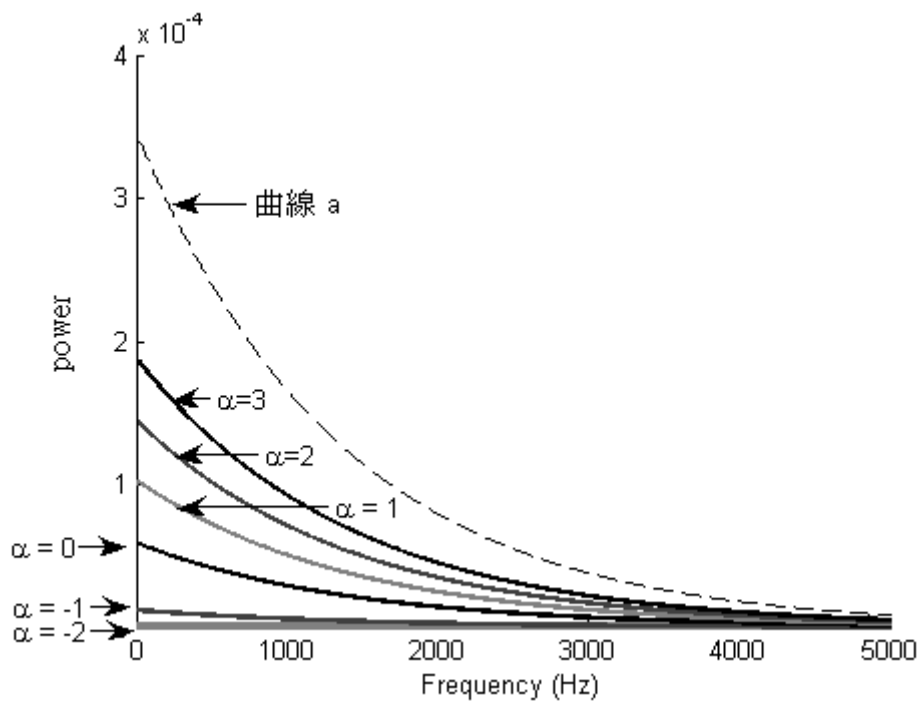


圖 39: 神經訊號於不同  $\alpha$  值的濾波曲線 b

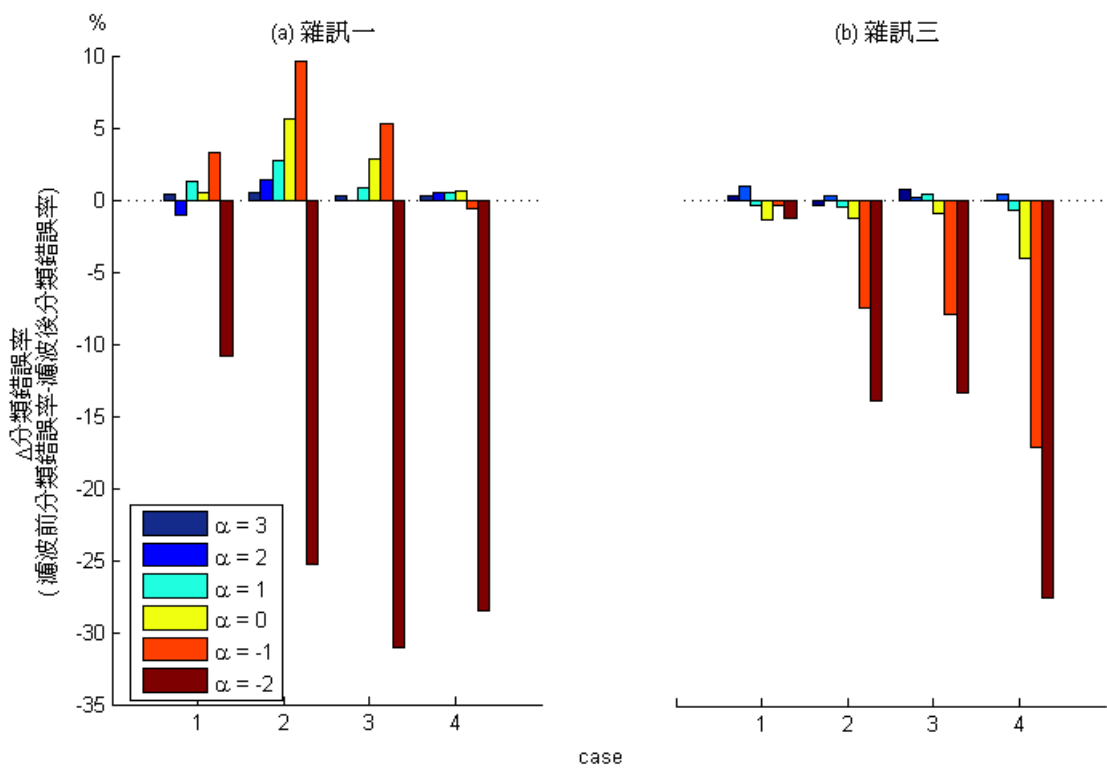


圖 40: 分類錯誤率變化直方圖

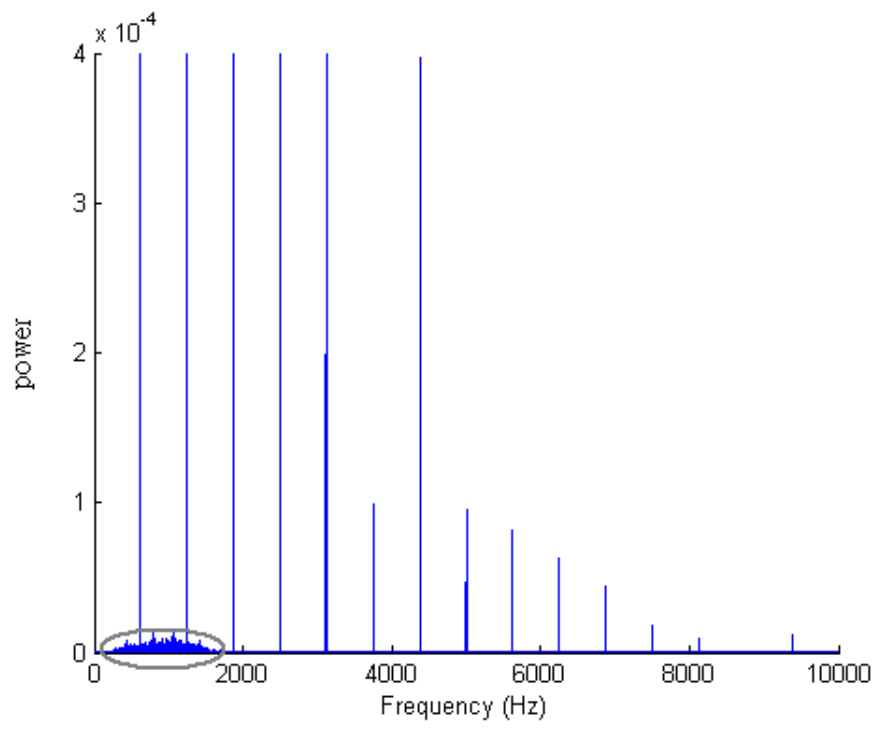


圖 41: spike 的頻譜圖

### 4.3.2 雜訊三的低頻探討

由於我們在 4.3.1 小節發現，由於濾波方法三無法濾掉雜訊三在頻譜上呈現的異常突起，使得此分類方法對於以雜訊三作為實驗雜訊的分類實驗效果不彰。並且，我們可以發現雜訊三在頻譜上的異常突起約落於 30 Hz (請見 3.3.1 以及表 5 之探討)，於是我們觀察雜訊三的頻譜圖(圖 42)。此張圖為隨機的 2000 筆 32 維動作電位，圖 42(a) 是以雜訊三作為實驗雜訊，並且由例一的兩個 spike 隨機模擬神經訊號的頻譜圖，圖 42(b) 則為圖 42(a) 於 0 至 200 Hz 的局部放大圖，圖 42(c) 為雜訊三的頻譜圖，圖 42(d) 為圖 42(c) 於 0 至 200 Hz 的局部放大圖。發現雜訊低頻呈現的能量與神經訊號低頻呈現的能量雷同，表示 spike 並不會對低頻造成影響。

發現雜訊三在低頻呈現不尋常的異常突起，於是我們以雜訊一作為實驗雜訊(圖 43)比較之。比較此兩雜訊的低頻段，發現雜訊一在低頻段的能量非常小(圖 43(d))，呈現白雜訊，並且雜訊低頻呈現的能量與神經訊號低頻呈現的能量雷同，同樣驗證 spike 並不會對低頻造成影響。反觀雜訊三在低頻卻有不尋常的起伏(圖 42(d))，以至於影響了分類實驗的效果。由於雜訊三的來源為訊號三，所以也呼應了 3.2.3 小節中的第三點探討，訊號三可能接收到其他的低頻電波，或是儀器有濾不掉的雜訊。

此外，我們可以由模擬神經訊號的頻譜發現，神經訊號頻譜低頻的部份(圖 42(b) 與 43(b))幾乎與雜訊頻譜的低頻(圖 42(d) 與 43(d))雷同，也就是說 spike 的組成對於神經訊號頻譜低頻的影響不大。觀察雜訊三的頻譜發現 0 至 60 Hz 有不尋常的能量起伏，於是我們將以此兩雜訊作為實驗模擬的神經訊號之 0 至 60 Hz 內的低頻皆濾掉，並且配合濾波方法三，將此濾波方式命名為濾波方法四。由於 4.3.1 觀察到，若以雜訊一為實驗雜訊，則  $\alpha = -1$  時，分類較佳；以及若以雜訊三為實驗雜訊，則  $\alpha = 2$  時，分類亦較佳。所以我們在此節討論若以雜訊一為實驗雜訊，則設定  $\alpha = -1$ ，若以雜訊三為實驗雜訊，則

設定  $\alpha = 2$ ，觀察其濾波方法四的分類效能(表 15)，並且實驗時同樣隨機選取 spike，以模擬真實的神經訊號。

雜訊一	分類錯誤率	例一	例二	例三	例四
	濾波前	29.75 %	18.75 %	16.10 %	3.75 %
$\alpha = -1$	濾波方法三	30.00 %	12.15 %	12.50 %	4.15 %
$\alpha = -1$	濾波方法四	28.75 %	14.90 %	13.55 %	4.80 %
雜訊三	分類錯誤率	例一	例二	例三	例四
	濾波前	50.40 %	32.30 %	32.75 %	13.45 %
$\alpha = 2$	濾波方法三	50.10 %	30.90 %	31.85 %	13.35 %
$\alpha = 2$	濾波方法四	47.40 %	25.20 %	20.20 %	7.85 %

表 15: 雜訊濾波前後分類錯誤率

由表 15 可以發現，若以雜訊一作為實驗雜訊來說，濾波方法四對於神經訊號分類反而會比濾波方法三的分類還要差；分類錯誤率相較於濾波方法三提高約 1%~3%，但對於例一這類 spike 非常相似的例子，卻些微降低約 2%。因為 60 Hz 以下的雜訊組成能量本來就很低並且呈現均勻分布，濾掉 60 Hz 以下能量反而會讓原本 spike 波形稍微不明顯，但似乎影響不大。對於以雜訊三作為實驗雜訊來說，濾波方法四對於神經訊號分類卻比濾波方法三的分類還要好；這是因為雜訊三在低頻接收了其他低頻電波，所以有不尋常的起伏以致於影響分類，所以若將 60 Hz 以下的雜訊組成濾掉，發現確實會讓分類結果有顯著的改善。

由以上實驗我們可以發現雜訊三確實接收到了其他的低頻訊號，雖然我們嘗試將影響雜訊三的低頻濾掉，但卻還是無法將接收到的其他訊號完全濾掉，以致執行濾波方法四後，雖然分類錯誤率降低，但還是沒有以雜訊一作為實驗雜訊的分類效果好。也就是說，對於雜訊三這類接收到其他異常訊號的神經訊號，仍然無法對其做有效的處理方式，但至少可以使得分類結果有顯著的改善。

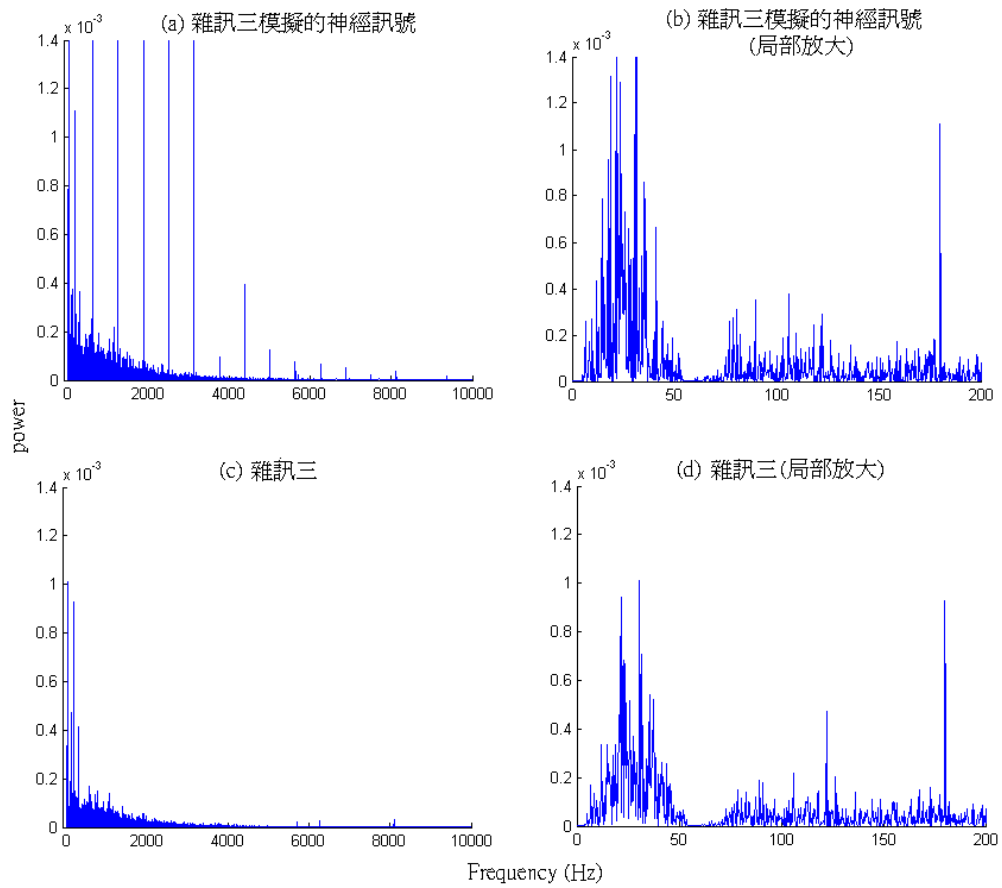


圖 42: 雜訊三與其模擬神經訊號的頻譜分析

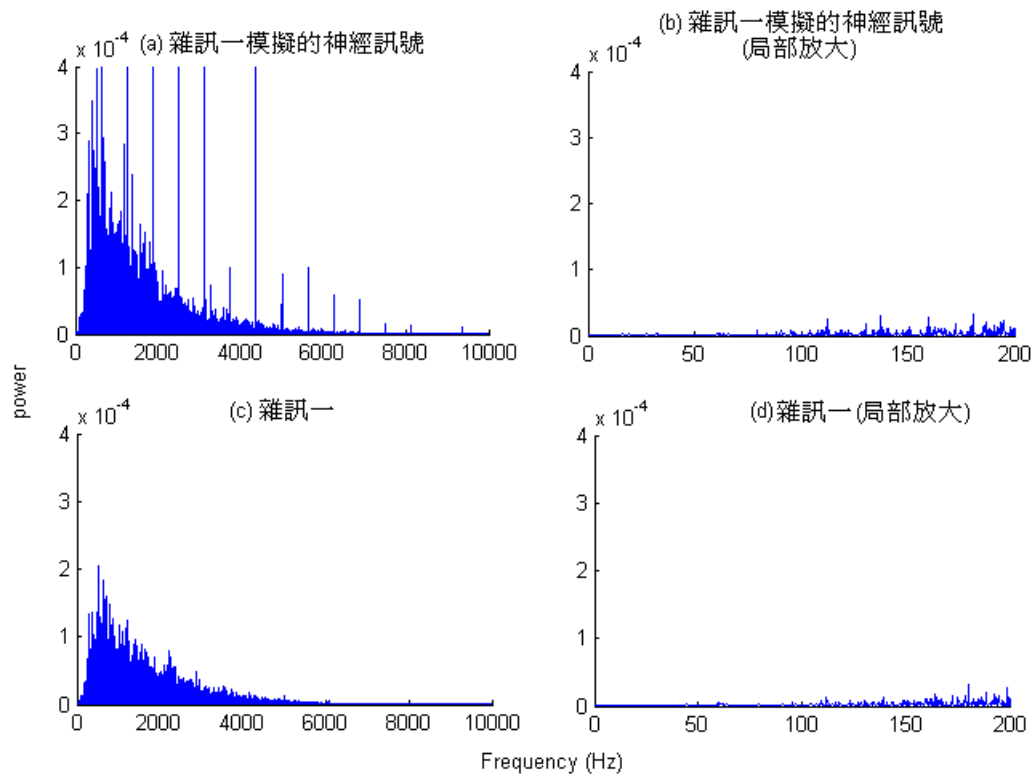


圖 43: 雜訊一與其模擬神經訊號的頻譜分析



#### 4.4 實際應用的限制

本論文中探討雜訊的性質，是否符合一般的統計假設「正規分布」以及 iid。基於這些基本的統計假設演算法像 EM、K-Means、貝氏分類法以及 [5]，若我們的雜訊較符合這些統計假設，則分類演算法較有效能。

並且於頻率域上，發現雜訊呈現穩定起伏，表示此雜訊帶有潛藏的資訊，藉由此一致性的表現，我們可以應用在濾波上，也就是 3.4 提出的濾波方法二，使得雜訊更符合正規分布以及 iid，並且可作為模擬實驗的實驗雜訊。

爾後，在做分類動作電位這類實驗時，我們可以將神經訊號分成兩部份：第一部分為學習階段，必須從中擷取出雜訊，藉由對雜訊的學習，可觀察雜訊在時間域以及頻率域上是否有異常突起；若有太多異常突起，表示分類動作電位的演算法不夠可靠，可能遺失一些資訊。第二部分為訓練階段，可使用本論文提出的濾波方法三針對神經訊號濾波後，進行分類；可將 PCA 點圖上有部分重疊的動作電位分的較開。但此方式對重疊度很高的樣本仍不適用。並且濾波方式可能有所缺失，若在學習階段發現像雜訊三這類的雜訊，也就是在頻率域上異常突起特別明顯的雜訊，可能就會失效，或是效果不彰。但我們可以使用濾波方法四，於學習階段觀察雜訊在低頻是否有異常突起(像雜訊三)，針對這類低頻有異常突起的神經訊號濾波後，會使得分類結果有顯著改善。

## 參考文獻

- [1] WIKIPEDIA The Free Encyclopedia. [http://en.wikipedia.org/wiki/White\\_noise](http://en.wikipedia.org/wiki/White_noise).
- [2] Ethem,Alpaydim.(2004).”Introduction to Machine Learning.”  
Londdon,England : The MIT Press.PP.85~150
- [3] 王錫崗等譯，”人體生理學”，新文京開發出版股份有限公司，2004。頁  
156~191。
- [4] 林震岩著，”多變量分析 SPSS 的操作與應用”，智勝文化出版，2006。頁  
358~447。
- [5] 李易霖，”神經細胞訊號的一種分類方法”，國立中央大學，碩士論文，  
2005。

## 附錄

本論文的實驗訊號由國立宜蘭大學生物機電學系蔡孟利教授所提供，下表 16 訊號一至訊號六均是由實驗紀錄的訊號，為真實的神經訊號。訊號一及訊號二為 spike 加上雜訊，其他訊號為人工判讀不含有 spike 的訊號。訊號四、訊號五以及訊號六為同一隻老鼠不同電極的紀錄，紀錄的條件是一樣的(麻醉)。我們可以由圖 44 觀察這六筆真實的神經訊號，持續時間均為 0.1 秒。

訊號	檔案來源	解析度	樣本點個數	取樣頻率
訊號一	signal9603.txt	小數點下三位小數	2437630	20 k
訊號二	R_0418_40KSel.txt	小數點下六位小數	2337024	40 k
訊號三	noise_070728Sel.txt	小數點下六位小數	2400001	40 k
訊號四	0923Sel.txt	小數點下六位小數	37900	40 k
訊號五	0923Sel-2.txt	小數點下六位小數	71000	40 k
訊號六	0923Sel-3.txt	小數點下六位小數	1432000	40 k

表 16: 訊號的檔案來源

以下雜訊一至雜訊六皆由訊號一至訊號六產生，雜訊一及雜訊二由訊號一及訊號二產生，為人工擷取雜訊，擷取程序請見第二章。由於雜訊一的取樣頻率為 20k，解析度為三位小數，為了雜訊一與其他雜訊的一致性，對於雜訊二至雜訊六皆做了「取奇數點」以及「四捨五入」的處理。取奇數點是為了將取樣頻率由 40k 轉換成 20k，四捨五入則讓解析度從六位小數轉換為三位小數，處理後的雜訊如表 17。

雜訊	訊號來源	解析度	樣本點個數	取樣頻率
雜訊一	訊號一	小數點下三位小數	301255	20 k
雜訊二	訊號二	小數點下三位小數	174680	20 k
雜訊三	訊號三	小數點下三位小數	1200001	20 k
雜訊四	訊號四	小數點下三位小數	18950	20 k
雜訊五	訊號五	小數點下三位小數	35500	20 k
雜訊六	訊號六	小數點下三位小數	716000	20 k

表 17: 雜訊的檔案來源

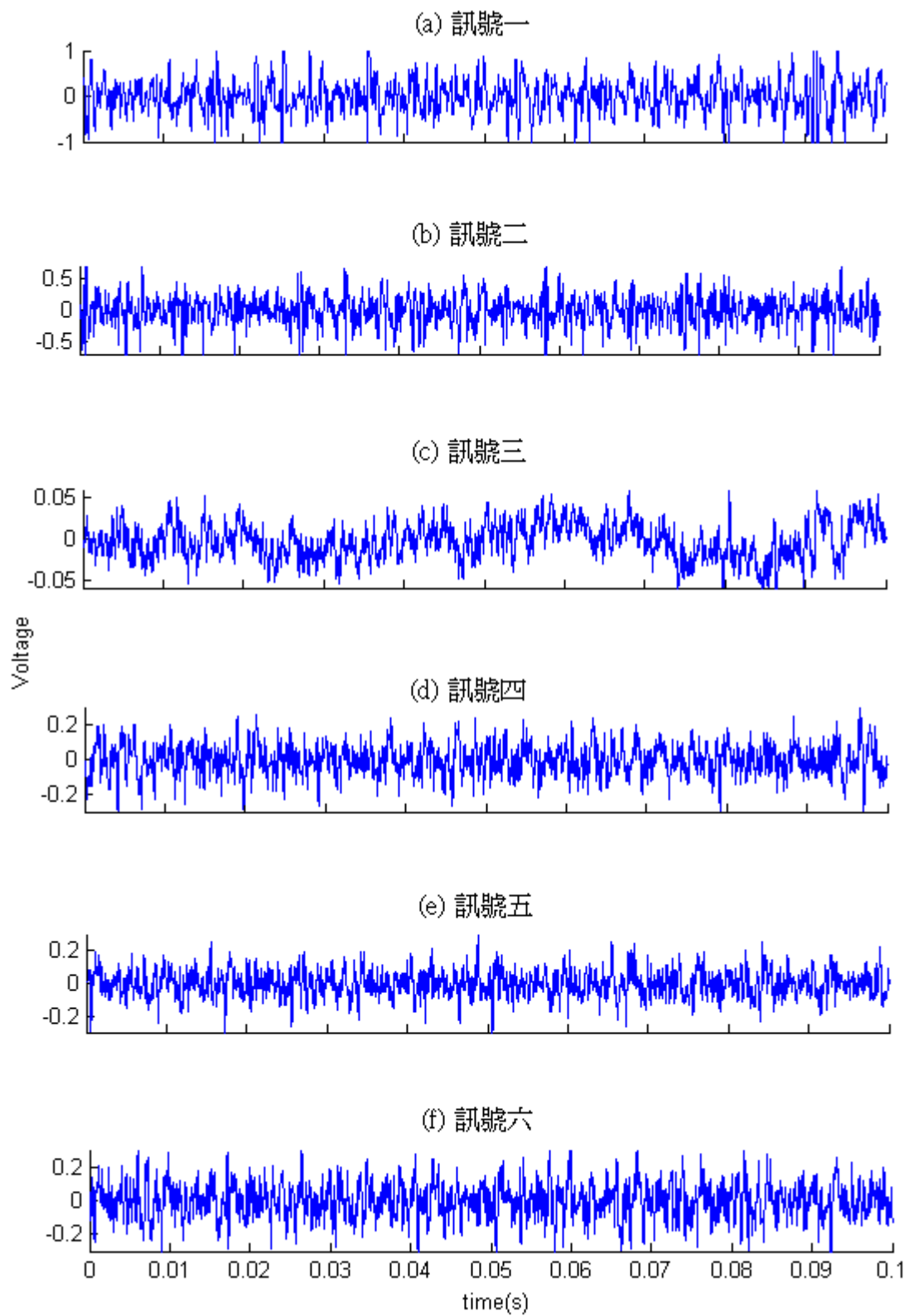


圖 44: 六筆真實的神經訊號