第四章 音框對數能量正規化

針對語音能量參數受到環境噪音干擾的改變現象，首先會在 4.1 節針對音框對數能量特徵做噪音干擾前後的觀察與進一步的探討，並且根據觀察結果提出本論文的音框對數能量正規化作法，其次討論正規化之實驗結果，最後則試驗音框對數能量正規化和倒頻譜正規化法之加成性。

4.1 音框對數能量特徵

根據觀察對數能量特徵值發現自動語音辨識結果會明顯的受到噪音干擾而辨識率變差，因此我們特別觀察語音對數能量特徵在不同噪音環境下的變化情形。在時間軸上，觀察對數能量由非語音到語音段落的改變過程，發現通常在一段乾淨語句中有語音出現的段落其對數能量特徵值會較高；反之若無語音出現的段落其對數能量特徵值則會接近於一穩定的低能量值。此外，當一段語句受到嚴重的噪音干擾前與噪音干擾後，語句中的能量特徵可以明顯看出在原本能量較低的部分會提高許多。

噪音環境干擾對於語句中對數能量特徵影響的變化程度可用圖 4.1.1 來做說明(圖示(a)為 20dB 噪音干擾環境對應乾淨環境情形，而圖示(b)為 10dB 噪音干擾環境對應乾淨環境情形)，語料來源是從 Aurora-2.0 訓練語料庫中乾淨訓練模式對應於的複合情境訓練模式的部分語料，圖示(a)中黑色點是以乾淨語句的每音框(Frame)對數能量同時以橫軸與縱軸座標值所繪出的參考點；而紅色叉則是以噪音干擾的語音對數能量為縱軸座標值對應乾淨語句的橫軸座標值所繪出的參考點；圖示(b)的參考點則與圖示(a)設定相同。由圖 4.1.1 可得知，當受到噪音影響時將會使得對數能量產生非線性的失真：在對數能量較高的音框僅有輕微的影響；但是相反地，在對數能量較低的音框上則會有嚴重的影響，故噪音干擾會讓

大部分的對數能量值被提高甚多，反而整體縮小對數能量的值域範圍，使得語音段落與非語音段落在對數能量的區隔上愈顯不易。

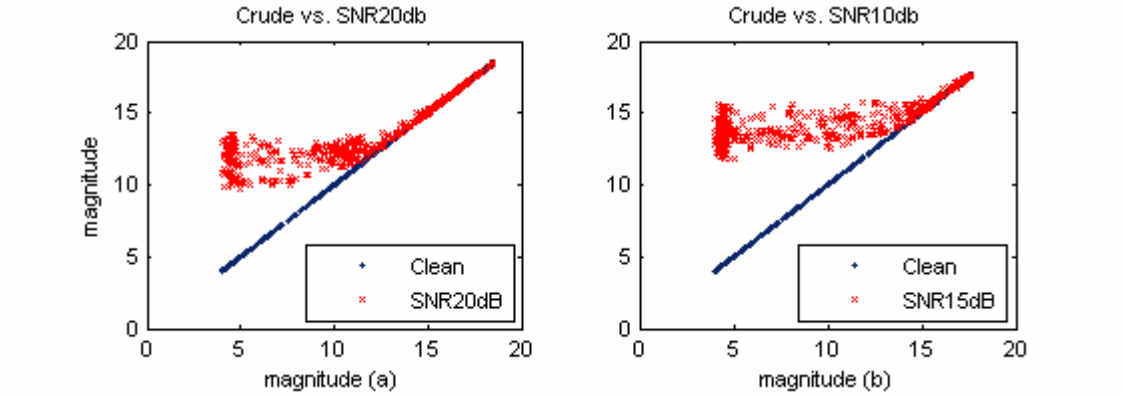

圖 4.1.1 加成性噪音影響語音對數能量特徵示意圖

在時間軸上對數能量的表現亦可用圖 4.1.2 來說明，語料來源是從Aurora-2.0 語料庫中的測試組A內挑選(語音內容為數字：139O)，環境狀況設定在乾淨環境、訊噪比(SNR)20dB噪音干擾和訊噪比(SNR)10dB噪音干擾。噪音環境對於語句對數能量特徵的影響表示：圖中橫座標表示連續的語音音框；縱軸座標代表在每一音框的對數能量；黑色實線代表原始乾淨語音的對數能量；藍色點線與紅色虛線分別代表訊噪比為 20dB與 10dB的噪音語音的對數能量。由圖亦可看出對數能量在乾淨環境下較低的音框，受噪音的影響程度較為嚴重。從圖 4.1.2 中對數能量在 5 附近的值域區間，原本是屬於非語音部分，但接受到噪音的干擾後，使得其對數能量相對提升。吾人認為上述情形是主要造成乾淨環境和噪音環境語音二者間在對數能量表現不匹配(Mismatch)的主要原因。

在這裡根據上述現象的觀察，因此我們假設若對數能量受噪音干擾的部分能重建出乾淨的語音還原成不受干擾時的情況，此時的辨識結果應該可以有效提升。所以在此特別針對對數能量維度，利用 Aurora-2.0 語料庫的語料對應特性，在相同內容的語料情形下可以利用乾淨環境的對數能量值將所有相對應語料之受噪音干擾的對數能量值作取代實驗。經過語音辨識步驟，最後可以取得一組將對數能量取代後的上限數據結果，如表 4.1.1。由結果顯示乾淨環境訓練模式平均正確率的上限可以高達 83.26%，而複合情境訓練模式可以達到 91.09%，可以

發現單獨的對數能量取代即可以有效影響辨識率結果。因此基於上述所觀察到的現象與結論，以下小節將提出對數能量尺度重刻法於強健性語音特徵處理的技術。

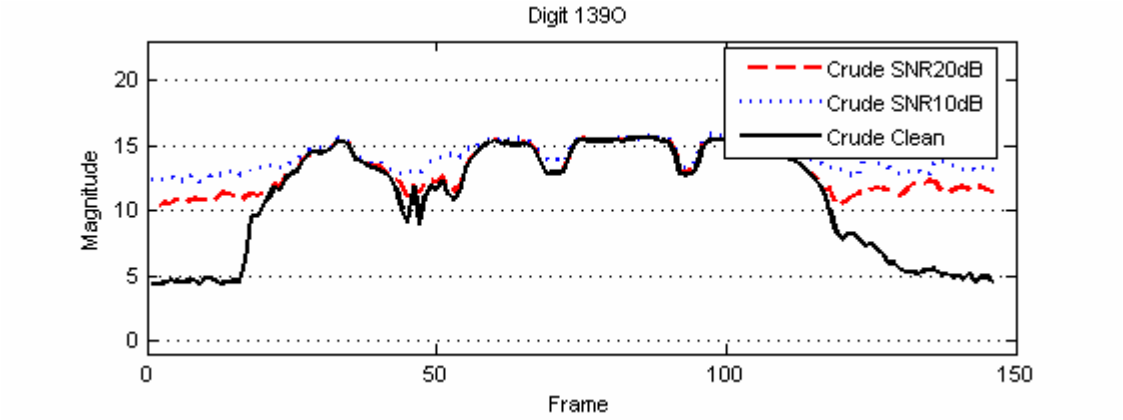

圖 4.1.2 數字語音之對數能量特徵示意圖(語音內容為：139O)

SNR 0~20dB 平均	測試組 A	測試組 B	測試組 C	總平均	基礎實驗總平均
乾淨環境訓練模式	84.05	85.8	76.58	83.26	58.96
複合情境訓練模式	91.45	93.07	86.41	91.09	83.82

表 4.1.1 無干擾之對數能量值取代干擾後的正確率上限

4.1.1 對數能量尺度重刻法 I (Log Energy Rescaling Normalization, LERN I)

在本節首先提出一個創新的語音對數能量特徵正規化技術－對數能量尺度重刻法 I，目的希望利用對數轉換函數方式來對語音對數能量參數作正規化，主要是考慮語句本身的音框對數能量特徵在不同噪音環境下的變化，試圖重建出乾淨的語音對數能量特徵。

根據上述所觀察到的現象顯示在一段乾淨語句中有語音出現的段落其對數能量特徵值會呈現較高值；反之若無語音出現的段落其對數能量特徵值則會趨近於一穩定的低能量值。另一方面，從觀察的現象可以發現受噪音干擾的語句，當

噪音能量大小沒有超過原本語音能量的情況下，該語句中受噪音干擾的音框對數能量會與相同語句在乾淨環境下的音框對數能量部分相近，因此我們認為語句中噪音能量沒有超過語音能量大小的情況下，對於大過噪音能量的音框部分在辨識系統上有相對的可靠性，而噪音所能干擾的範圍內之對數能量則容易造成錯誤的辨識結果。

因為上述的噪音干擾現象，對數能量尺度重刻的基本原理，希望將原特徵能量值乘上該特徵值所對應的等份區間之對數轉換函數值，改變特徵值尺度使處理後的特徵值能充分表現出乾淨環境下語音片段與非語音片段的對數能量差距，然而分別對每一語句劃分為均勻的相同等份，目的在於使語句和語句間互不相作用影響，僅考慮該語句中所有音框的等份數，最後利用各音框等份位置的對數轉換函數值做正規化。在這裡我們設定正規化對數函數輸出值介於 0 到 1 之間，使轉換後音框對數能量與原本對數能量的差異量自小到大有遞減的現象，因此在經過對數能量尺度重刻處理過後的音框對數能量，將可以得到原來對數能量值較低的語音段落其對數能量值越低，以及對數能量值較高的語音段落其對數能量值可以儘量維持不變，讓噪音語句在經過正規化後的對數能量可以趨近出乾淨環境下的對數能量特徵。

對數能量尺度重刻具體作法如下面步驟。首先，自每一語句(包含測試及訓練語句)的所有音框中找出最大對數能量值 $LE\_\max$ 以及最小對數能量值 $LE\_\min$ ：

$$\begin{aligned} LE\_max &= \underset{1\le i\le T}{Max}\, LogE[i], \\ LE\_min &= \underset{1\le i\le T}{Min}\, LogE[i] \end{aligned} \tag{4.1.1}$$

式中 T 為每一語句音框數，其次根據 $LE\_\max$ 及 $LE\_\min$ 決定對數能量值域範圍，並將此一範圍劃分成 M 個等份，因此每個等份區間的寬度為 L 可表示如下：

$$L = \frac{LE\_max - LE\_min}{M} \tag{4.1.2}$$

在本論文我們初步將等份個數 M 設為 100，而等份 m 所對應的對數轉換函數值如圖 4.1.3 和式(4.1.3)：

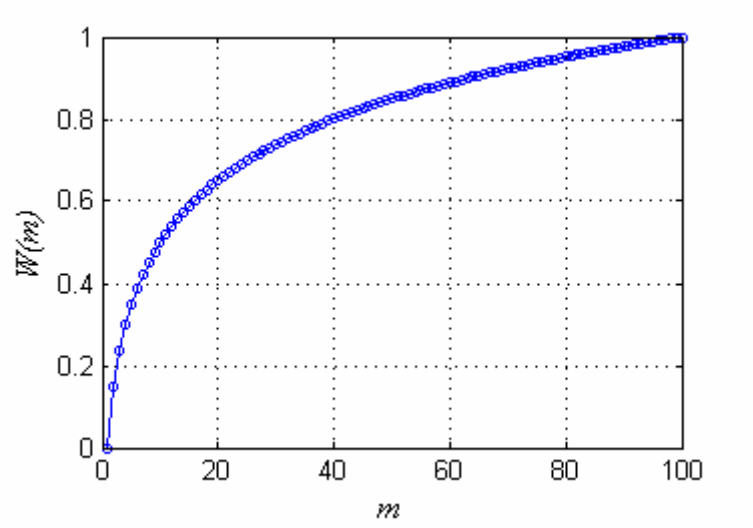

圖 4.1.3 對數轉換函數 $W(m)$ 圖示

$$W(m)=\frac{\log(m)}{\log(M)} \tag{4.1.3}$$

另一方面，每個音框對數能量所落至該等份的索引值可表示成：

$$Index_i=\left\lfloor \frac{LogE[i]-LE\_min}{L} \right\rfloor \tag{4.1.4}$$

因此音框 i 經正規化後的對數能量 $Log\hat{E}[i]$ 可以表示成：

$$Log\hat{E}[i]=LogE[i]\times W(Index_i) \tag{4.1.5}$$

從圖 4.1.2 看出，未經處理的噪音語音對數能量值在不同訊噪比狀況下(Clean, 15dB, 5dB)其音框能量值的曲線相對提昇許多，尤其以 5dB 噪音干擾的情況下，語句的對數能量的值域被嚴重壓縮。而在經過對數能量尺度重刻處理後，不同訊噪比情況下，於非語音段落的對數能量值由 10 大小降低至 5 以下，我們可以容易地由圖 4.1.4 的非語音區間察覺我們所提出對數能量尺度重刻方法能將噪音環境的對數能量曲線靠近乾淨語音對數能量曲線，但是轉換函數會讓轉換後的對數

能量過小造成對數能量曲線不連續的現象，如圖 4.1.4 橢圓區域範圍中的對數能量曲線，會有突然的波谷(Valley)出現。此外我們再一次將經過對數能量尺度重刻法處理後的音框對數能量利用圖 4.1.1 的方式畫出表示圖 4.1.5，由圖中可以看出，處理後的對數能量點雖然在低能量區間會呈現散亂的情況，並且因為對數轉換函數的關係，對數能量點在噪音環境下會比在乾淨環境降低的更多，如圖中橢圓區塊的值，但實際上大部分的能量點會接近於乾淨語句的每音框(Frame)對數能量參考點。另一方面，圖 4.1.4 與圖 4.1.5 的乾淨環境下之對數能量參考點是有經過對數能量尺度重刻法處理的結果，主要原因參考節後的實驗數據比較中得知有較佳的辨識效果，故圖中的對數能量參考點皆有採用對數能量尺度重刻法處理。

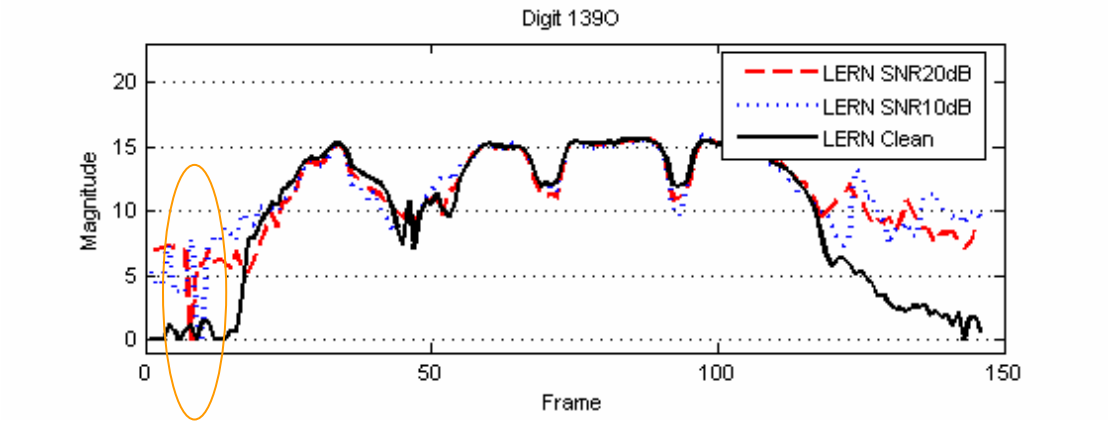

圖 4.1.4 對數能量尺度重刻法於語音對數能量特徵之作用結果圖示(1)

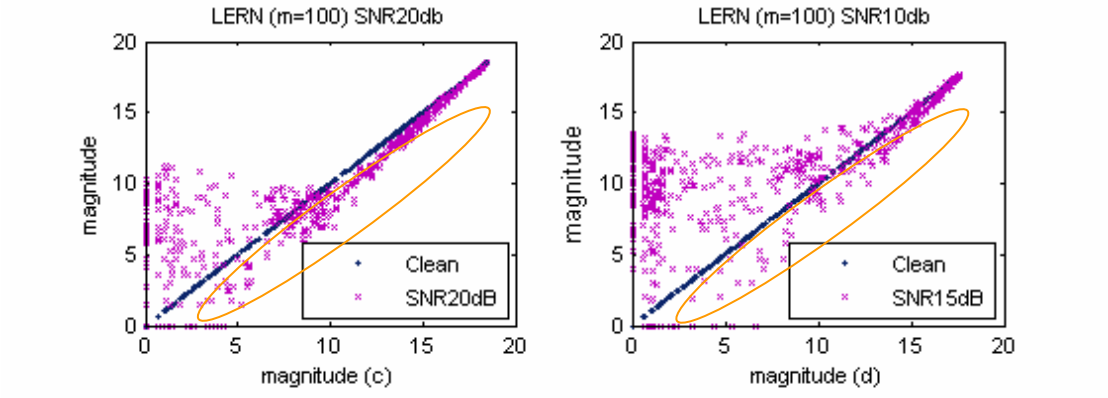

圖 4.1.5 對數能量尺度重刻法於語音對數能量特徵之作用結果圖示(2)

4.1.2 對數能量尺度重刻法 II (LERN II)

對數能量尺度重刻法 II，同方法 I 根據上述所觀察到的現象顯示在一段乾淨語句中有語音出現的段落其對數能量特徵值會呈現較高值；反之若無語音出現的段落其對數能量特徵值則會趨近於零。因此對數能量尺度重刻 II 的目標希望將原特徵能量值乘上該特徵值所對應的權重值。在此我們定義權重值函數如下式：

$$W(i)=\left(\frac{\left(\log E[i]-\alpha\cdot LE\_\min\right)}{LE\max-\alpha\cdot LE\min}\right)^{\beta} \tag{4.1.6}$$

式中 $LE\_\max$ 和 $LE\_\min$ 為最大對數能量值以及最小對數能量值，是由每一獨立測試語句中的所有音框 T 找出，其中 α 與 β 為控制權重曲線差異的參數，α 表示調整其值域範圍大小，而 β 指數使其成一非線性函數，權重曲線可用下圖 4.1.6 和圖 4.1.7 表示，範例圖 4.1.6 控制 α 為變數 β 為常數 1，另一圖 4.1.7 控制 β 為變數 α 為常數 1，圖中我們設定橫軸的對數能量值 $\log E[i]$ 由 0 值開始到最大值 100，縱軸則為權重值 $W(i)$ 的值域從 0 到 1。

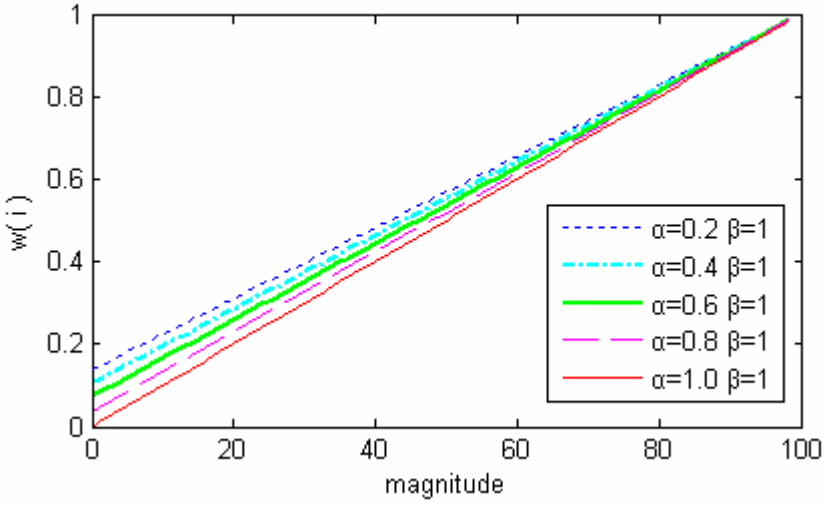

圖 4.1.6 在 β 值固定的情況下(β=1)，不同的 α 值對 $W(i)$ 的影響

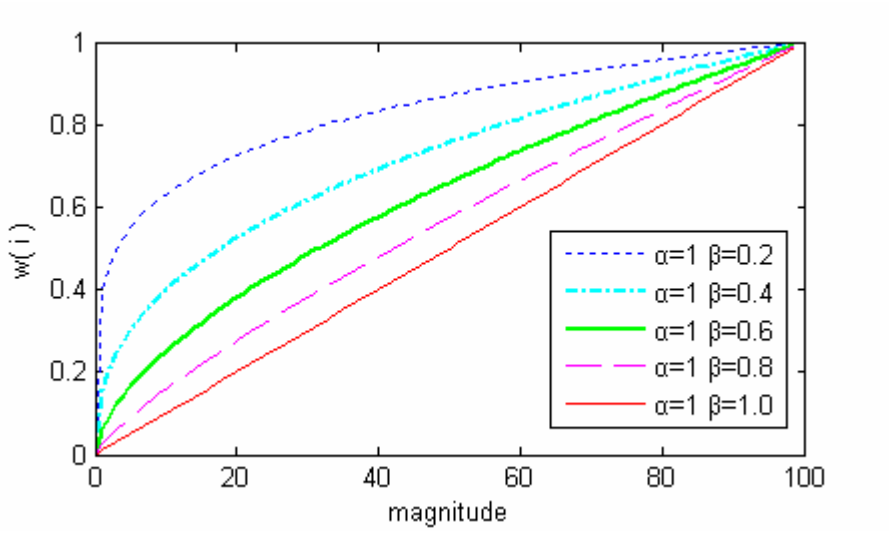

圖 4.1.7 在α值固定的情況下(α=1)，不同的β值對$W(i)$的影響

由圖 4.1.6 可以了解當α為變數時允許對能量值範圍做出微調的改變，而圖 4.1.7 則當β為變數時可以得到一非線性曲線，從曲線清楚看出在相對於對數能量較低的音框時可以有較低的權重值。因此當我們對原始對數能量乘上該音框的權重值後，最終能夠達到如觀察乾淨語句時有語音段落其對數能量特徵值較高；非語音段落其對數能量特徵值趨於零的目的，表示如下式：

$$\log \hat{E}[i] = \log E[i] \times \left(\frac{\left(\log E[i] - \alpha \cdot LE\_\min\right)}{LE\max - \alpha \cdot LE\min} \right)^{\beta} \tag{4.1.7}$$

至於權重函數中的α與β雙變數，在此為求取最佳值，我們進一步的利用線性迴歸的曲線擬合方法來取得α與β適當解。「曲線擬合」的概念是當給定一些資料點數(u_i, v_i)，若要以一個函數描述反應變數v_i與解釋變數u_i關係，通常可使用迴歸模型(Regression Models)來表示。換句話說，迴歸模型可用來解釋給定u_i的情況下，預測v_i的值為何。通常迴歸公式$G(u_i)$可依參數(Coefficients)組合不同表示成線性(linear)或非線性(nonlinear)型式，並且$G(u_i)$參數的選擇影響預測值$\tilde{v}_i$的準確性甚鉅，一般利用誤差平方和最小化(Minimization of the Sum of Squares Error)求得，亦即將所有u_i分別代入迴歸公式所求得的預測值$\tilde{v}_i$和實際觀測值v_i的誤差值平方合必須最小，其意謂著經由迴歸模型所預測出的值會跟實際的值較相似，

此法又可稱最小平方迴歸法(Least Squares Regression)。因此，對權重函數於誤差平方和E^2可以定義成式(4.1.8)：

$$E(\alpha,\beta)=\sum_{i=1}^{N}\left(v_i-\left(\log E[i]\times\left(\frac{\left(\log E[i]-\alpha\cdot LE\_\min\right)}{LE\max-\alpha\cdot LE\min}\right)^{\beta}\right)\right)^2 \tag{4.1.8}$$

但因為函數是非線性型式無法對參數α、β的導式為零求解，所以最後可利用梯度下降法(Gradient Descent)來求得適當解。

4.2 音框對數能量尺度重刻法實驗結果

本節討論音框對數能量正規化之強健技術。實驗結果將比較三組不同噪音設定下之乾淨環境結果與不同訊噪比-5dB~20dB 干擾下的結果，表格中之三組噪音環境(測試組 A、測試組 B 與測試組 C)所表示的正確率為 0dB 至 20dB 的算數平均數結果，最後總平均值(Average)計算方式為測試組 A 四份加上測試組 B 四份和測試組 C 兩份的平均值，如下：

$$\frac{\text{測試組 A}\times 4+\text{測試組 B}\times 4+\text{測試組 C}\times 2}{10} \tag{4.2.1}$$

4.2.1 對數能量尺度重刻法 I 實驗

對數能量尺度重刻法 I，在此將探討三組實驗數據。第一組實驗：在對數能量尺度重刻法上我們使用不同刻度作測試，主要針對對數表的 *M* 個等份各別設定為 50、70 到 500 與 1000 多種尺度做觀察。最後產生的結果如表 4.2.1。

從實驗一表格中得知，以 *M*=100 的尺度設定在乾淨語料訓練模式下會有最佳的正確率，而複合情境訓練模式下則是以 500 等份與 1000 等份的尺度設定會有好的效果。因此在第二組實驗：我們針對 100 的尺度設定，比較結果如表 4.2.2 與表 4.2.3。

	乾淨環境訓練模式				複合情境訓練模式			
M 等份	測試組 A	測試組 B	測試組 C	總平均	測試組 A	測試組 B	測試組 C	總平均
基礎實驗	58.94	58.48	59.97	58.96	85.22	83.99	80.67	83.82
M = 50	74.10	76.71	63.07	72.94	86.33	86.25	81.04	85.24
M = 60	74.16	76.71	63.30	73.01	86.32	86.24	80.97	85.22
M = 70	74.32	76.79	63.46	73.13	86.34	86.28	81.05	85.26
M = 80	74.35	76.76	63.62	73.17	86.36	86.31	81.14	85.29
M = 90	74.35	76.70	63.67	73.15	86.33	86.25	81.22	85.27
M = 100	74.36	76.72	63.83	73.20	86.31	86.27	81.22	85.28
M = 110	74.34	76.65	63.83	73.17	86.37	86.28	81.25	85.31
M = 120	74.28	76.60	63.85	73.12	86.38	86.25	81.30	85.31
M = 130	74.28	76.56	63.90	73.11	86.38	86.20	81.31	85.29
M = 140	74.23	76.47	63.90	73.06	86.37	86.16	81.37	85.28
M = 150	74.23	76.43	63.90	73.05	86.39	86.19	81.38	85.31
M = 500	73.47	75.21	63.33	72.14	86.75	85.85	81.62	85.36
M = 1000	73.24	74.90	63.74	72.00	86.67	85.89	81.68	85.36

表 4.2.1 對數能量尺度重刻法 I 於不同尺度等份的實驗結果

		乾淨環境訓練模式							
	訊噪比	Clean	20dB	15dB	10dB	5dB	0dB	-5dB	0~20dB 平均
測試組 A	地下鐵	99.08	97.45	94.44	84.03	58.27	25.88	10.25	72.01
	人聲	98.97	98.19	95.86	85.64	60.70	25.27	2.96	73.13
	汽車	98.93	98.00	96.21	88.49	69.25	39.61	14.67	78.31
	展覽會館	99.20	97.01	93.67	83.12	61.71	34.31	15.80	73.96
	平均	99.05	97.66	95.05	85.32	62.48	31.27	10.92	74.36
測試組 B	餐廳	99.08	98.16	95.36	86.06	62.97	29.66	7.61	74.44
	街道	98.97	97.70	94.77	85.67	65.72	38.42	15.54	76.46
	機場	98.93	98.27	96.33	89.74	70.12	38.38	12.11	78.57
	火車站	99.20	98.12	95.53	88.18	67.02	38.20	14.63	77.41
	平均	99.05	98.06	95.50	87.41	66.46	36.17	12.47	76.72
測試組 C	地下鐵	99.39	93.61	85.51	68.19	41.23	17.56	10.01	61.22
	街道	99.09	94.80	87.82	72.34	50.33	26.90	15.81	66.44
	平均	99.24	94.21	86.67	70.27	45.78	22.23	12.91	63.83

表 4.2.2 對數能量尺度重刻法 I 於乾淨環境訓練模式(100 等份)實驗結果

		複合情境訓練模式							
	訊噪比	Clean	20dB	15dB	10dB	5dB	0dB	-5dB	0~20dB 平均
測試組 A	地下鐵	98.43	96.90	96.07	91.83	82.74	59.87	23.73	85.48
	人聲	98.31	97.31	96.04	93.11	83.77	57.83	22.04	85.61
	汽車	98.18	97.26	96.72	94.33	86.43	59.47	21.09	86.84
	展覽會館	98.46	97.38	96.73	93.30	86.45	62.70	23.11	87.31
	平均	98.35	97.21	96.39	93.14	84.85	59.97	22.49	86.31
測試組 B	餐廳	98.43	96.96	95.12	89.56	80.84	57.08	21.28	83.91
	街道	98.31	97.40	96.13	93.11	84.07	59.95	24.94	86.13
	機場	98.18	97.73	96.63	94.12	87.09	67.91	31.82	88.70
	火車站	98.46	97.35	95.77	93.24	83.77	61.52	24.81	86.33
	平均	98.35	97.36	95.91	92.51	83.94	61.62	25.71	86.27
測試組 C	地下鐵	98.65	96.78	95.21	91.13	76.91	39.64	12.77	79.93
	街道	98.16	96.80	95.89	91.72	79.56	48.52	21.16	82.50
	平均	98.41	96.79	95.55	91.43	78.24	44.08	16.97	81.22

表 4.2.3 對數能量尺度重刻法 I 於複合情境訓練模式(100 等份)實驗結果

表 4.2.2 與表 4.2.3 中 0-20dB 平均為訊噪比 0~20dB 干擾下的平均值結果。綜合乾淨環境訓練模式與複合情境訓練模式的實驗結果於各組別之正確率比較，我們得知在三類測試組的正確率都有提升的效果，但對於測試組 C 中 MIRS 通道加成性噪音效果的提升卻比較小。其次，對於測試組 B 的非穩定性(Nonstationary)噪音則有最佳的提升效果。

實驗三我們觀察對數能量尺度重刻法I若僅使用在測試語料狀況下與同時使用在訓練語料和測試語料的不同，實驗結果如表 4.2.4(使用 M =100 的等份設定)。結果中發現，若同時針對在訓練語料和測試語料的情況下使用對數能量尺度重刻法，會有較高的正確率。在此我們分析表 4.2.4 結果並參考圖 4.2.1，圖(a)與表示只對測試語料處理，圖(b)則是同時對訓練語料和測試語料處理，對於實驗數據結果認為是因為訓練語料在低能量部分並沒有靠近靜音(能量值近似於零)的情況，而是在低能量部分仍然會有小能量的噪音值產生。所以實驗設定為訓練語料和測試語料同時經過我們的方法處理後會使對數能量的值域範圍較為相似，因此辨識率可以相對提高許多。

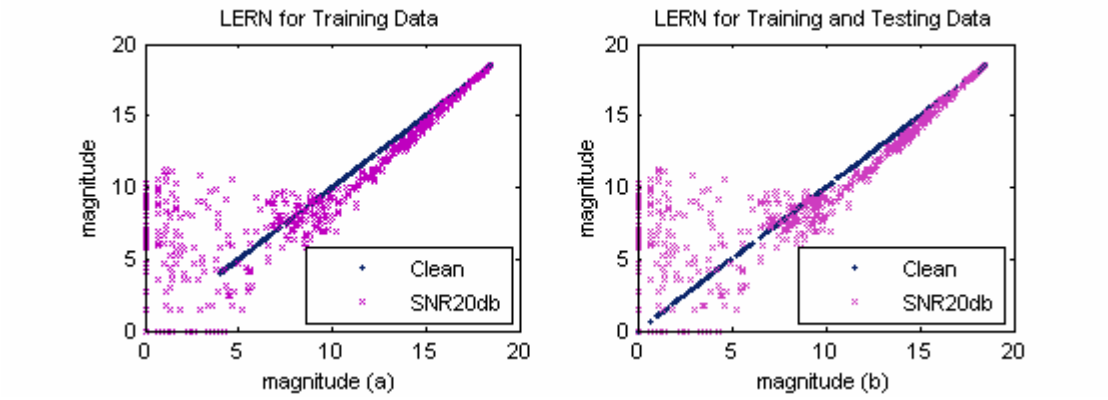

圖 4.2.1 對數能量尺度重刻法 I 於訓練語料和測試語料的差異圖示

(100 等份)	LERN 處理對象	測試組 A	測試組 B	測試組 C	總平均
乾淨環境	訓練語料＋測試語料	74.36	76.72	63.83	73.20
訓練模式	測試語料	72.20	75.45	60.58	71.18
複合情境	訓練語料＋測試語料	86.31	86.27	81.22	85.28
訓練模式	測試語料	77.76	81.98	69.86	77.87

表 4.2.4 對數能量尺度重刻法 I 於訓練語料和測試語料的差異結果

4.2.2 對數能量尺度重刻法 II 實驗

對數能量尺度重刻法 II，在此分為人工設定(Hand Set)參數與線性迴歸的曲線擬合法訓練參數的兩類實驗。實驗一：對於方法 II 我們以人工設定之α或β值為控制權重曲線差異的參數，使用不同的參數值只針對乾淨環境訓練模式的測試語料做處理，結果如表 4.2.5(實驗 1)。數據中以α為 1.0 和β為 0.2 之參數設定的組合最佳，平均正確率可達 72.02%，但是可以明顯發現在不同參數設定下的測試組 C，其結果大多呈現正確率下降的效果，並且在此發現當α參數為 1.0 和β為 1.0 的時候，對數能量尺度重刻法 II 為一線性調整，如式 4.2.1，由結果看出其效果極差。

$$Log\hat{E}[i] = LogE[i] \times \frac{LogE[i] - LE\_\min}{LE\_\max - LE\_\min} \tag{4.2.1}$$

其次，我們仍然使用人工設定之α或β值為控制權重曲線差異的參數，但同時針

對乾淨環境訓練模式和複合情境訓練模式的訓練語料與測試語料作處理，結果如表 4.2.6(實驗 2)。在乾淨環境訓練模式下的數據以α為 1.0 和β為 0.4 之參數設定的組合最佳，平均正確率可達 73%，而複合情境訓練模式下的結果則平均可以有 75%左右。最後比較表 4.2.5 與表 4.2.6，我們得知對數能量尺度重刻法 II 在使用人工設定參數值的情況下，當處理對象有同時對訓練語料和測試語料的時候會有比較好的結果。

乾淨環境訓練模式				
參數值	測試組 A	測試組 B	測試組 C	總平均
α: 0.2 β: 1.0	70.81	72.86	58.38	69.15
α: 0.4 β: 1.0	70.89	73.13	57.10	69.03
α: 0.6 β: 1.0	69.63	71.86	54.26	67.45
α: 0.8 β: 1.0	55.89	65.66	47.73	58.16
α: 1.0 β: 0.2	73.53	75.85	61.35	72.02
α: 1.0 β: 0.4	67.29	71.52	55.37	66.60
α: 1.0 β: 0.6	57.16	62.89	46.19	57.26
α: 1.0 β: 0.8	47.88	54.27	37.65	48.39
α: 1.0 β: 1.0	40.49	46.95	30.96	41.17

表 4.2.5 對數能量尺度重刻法 II 於人工設定參數值之測試(實驗 1)

	乾淨環境訓練模式				複合情境訓練模式			
參數值	測試組 A	測試組 B	測試組 C	總平均	測試組 A	測試組 B	測試組 C	總平均
α: 0.2 β: 1.0	57.66	59.49	57.29	58.32	85.66	84.60	80.47	84.20
α: 0.4 β: 1.0	60.34	62.52	57.71	60.69	85.80	84.64	80.46	84.27
α: 0.6 β: 1.0	64.33	66.96	58.40	64.20	85.94	85.15	80.89	84.61
α: 0.8 β: 1.0	69.59	72.99	58.89	68.81	87.00	87.35	80.91	85.92
α: 1.0 β: 0.2	73.25	75.31	62.83	71.99	86.71	86.05	81.67	85.44
α: 1.0 β: 0.4	74.25	77.25	61.98	73.00	86.51	85.80	80.95	85.11
α: 1.0 β: 0.6	73.30	77.16	60.39	72.26	87.17	87.10	80.32	85.77
α: 1.0 β: 0.8	72.40	76.63	59.44	71.50	86.46	87.69	80.68	85.80
α: 1.0 β: 1.0	71.71	76.45	58.76	71.02	86.26	87.13	80.41	85.44

表 4.2.6 對數能量尺度重刻法 II 於人工設定參數值之測試(實驗 2)

實驗二：我們利用線性迴歸的曲線擬合方法來取得α與β適當解。實驗上主要根據 Aurora-2.0 實驗語料庫中的兩組訓練語料，分別是乾淨環境(Clean)下之訓練語料和複合情境(Multi)下之訓練語料，由於此兩組訓練語料在語料內容上存在相對應的關係，因此吾人將此對應的關係利用曲線擬合法求得α與β解。此外在情境下之訓練語料共分為 5dB、10dB、15dB、20dB 與混合(Multi)四種不同訊噪比程度的噪音干擾情況，所以α與β參數可以分別求出在五組不同噪音干擾下的適當解與一組混和所有噪音干擾情況的適當解，其次如上所述的α與β解必須同時使用到兩組訓練語料，故實驗結果中表 4.2.7 於乾淨環境訓練模式，此時只針對測試語料做調整，而表 4.2.8 複合情境訓練模式則將α與β參數同時對訓練語料和測試語料做處理。另一方面如表 4.2.9 與表 4.2.10 我們比較在不同情況下的α與β參數對於不同噪音干擾程度的正確率效果，我們期望當參數取得的噪音干擾程度與測試環境相同的情形下可以有較好的辨識率，但很可惜的在此相同情況下的辨識率並不沒有因為噪音干擾程度一樣而特別提升，然而我們發現當 05dB 平均值的正確率在乾淨環境訓練模式下明顯變差，而複合情境訓練模式下的正確率則差異不大。最後結果顯示乾淨環境訓練模式如表 4.2.7，當訓練環境設定為 20dB 情境下所求得的α與β參數效果最佳，而複合情境訓練模式為 5dB 情境下所求得的參數效果最佳。

乾淨環境訓練模式					
訊噪比	參數值	測試組 A	測試組 B	測試組 C	總平均
05 dB	α: 0.90 β: 0.71	64.23	67.69	50.27	62.82
10 dB	α: 0.95 β: 0.43	70.11	73.17	56.23	68.56
15 dB	α: 0.98 β: 0.32	72.19	75.22	58.70	70.70
20 dB	α: 0.98 β: 0.25	73.26	75.85	60.11	71.66
Multi	α: 0.98 β: 0.34	71.24	74.36	57.91	69.82

表 4.2.7 對數能量尺度重刻法 II 於曲線擬合法之參數實驗(乾淨環境訓練模式)

複合情境訓練模式					
訊噪比	參數值	測試組 A	測試組 B	測試組 C	總平均
05 dB	α: 0.90 β: 0.71	86.61	86.38	80.98	85.39
10 dB	α: 0.95 β: 0.43	86.03	85.49	80.93	84.79
15 dB	α: 0.98 β: 0.32	86.42	85.71	81.18	85.09
20 dB	α: 0.98 β: 0.25	86.47	85.72	81.15	85.11
Multi	α: 0.98 β: 0.34	86.26	85.61	81.16	84.98

表 4.2.8 對數能量尺度重刻法 II 於曲線擬合法之參數實驗(複合情境訓練模式)

乾淨環境訓練模式								
訊噪比	參數值	Clean	20dB	15dB	10dB	05dB	00dB	0~20dB 平均
05 dB	α: 0.90 β: 0.71	98.42	89.98	81.82	68.07	47.70	26.55	62.82
10 dB	α: 0.95 β: 0.43	98.86	94.57	88.57	76.07	54.47	29.12	68.56
15 dB	α: 0.98 β: 0.32	98.91	95.96	90.89	79.15	57.34	30.17	70.70
20 dB	α: 0.98 β: 0.25	98.98	96.54	92.05	80.95	58.67	30.11	71.66
Multi	α: 0.98 β: 0.34	98.90	95.56	90.02	77.98	55.91	29.63	69.82

表 4.2.9 對數能量尺度重刻法 II 於不同訊噪比干擾下實驗(乾淨環境訓練模式)

複合情境訓練模式								
訊噪比	參數值	Clean	20dB	15dB	10dB	05dB	00dB	0~20dB 平均
05 dB	α: 0.90 β: 0.71	98.41	97.68	96.64	93.52	83.60	55.52	85.39
10 dB	α: 0.95 β: 0.43	98.43	97.30	96.00	92.35	82.33	56.00	84.79
15 dB	α: 0.98 β: 0.32	98.46	97.37	96.02	92.47	82.65	56.93	85.09
20 dB	α: 0.98 β: 0.25	98.42	97.33	96.01	92.52	82.62	57.05	85.11
Multi	α: 0.98 β: 0.34	98.40	97.22	95.90	92.36	82.65	56.78	84.98

表 4.2.10 對數能量尺度重刻法 II 於不同訊噪比干擾下實驗(複合情境訓練模式)

綜合比較：

本小節將綜合比較文獻參考的方法與吾人所提出的對數能量尺度重刻法，並且比較多項式擬合統計圖等化法 (Polynomial-Fit Histogram Equalization, PHEQ) [Lin et al. 2006]於能量維度的效果，多項式擬合統計圖等化法主要精神可以視為是利用一個轉換函數(Transformation Function)，此函數能將測試語句的語音特徵向量

每一維的統計分佈分別轉換至先前已從訓練語句中定義好的對應參考分佈。綜合實驗結果如下表，表 4.2.11 為乾淨環境訓練模式，表中得知所有的方法與基礎實驗比較都有明顯的進步。但各組分別觀察，發現結果當中的測試組 C 的效果表現有好有壞，且好的進步效果有限。總平均結果則是以對數能量尺度重刻法 I 的進步率為最高到 34.7%，其他方法則在 30%左右或更低一些。在表 4.2.12 則為複合情境訓練模式，結果顯示對數能量尺度重刻法 I 仍然有比較好的進步率，可以高達 9.0%。最後綜觀兩種環境的訓練模式來看，實驗中測試組 C 的 MIRS 通道加成性噪音干擾，在結果表現上明顯是所有方法都無法有效提升辨識率。

乾淨環境訓練模式					
0~20dB 平均	測試組 A	測試組 B	測試組 C	總平均	進步率
Baseline	58.94	58.48	59.97	58.96	
FES	70.60	71.20	60.90	68.90	24.23
LEDRN Linear	73.11	75.47	58.65	71.16	29.73
LEDRN Non-Linear	74.18	76.45	62.52	72.75	33.61
SLEN Ⅱ	69.93	74.59	55.85	68.98	24.41
PHEQ on Energy	72.45	76.03	61.50	71.69	31.03
LERN Ⅰ (M = 100)	74.36	76.72	63.83	73.20	34.70
LERN Ⅱ (α: 0.98 β: 0.25)	73.26	75.85	60.11	71.66	30.95

表 4.2.11 乾淨環境訓練模式下綜合實驗結果

複合情境訓練模式					
0~20dB 平均	測試組 A	測試組 B	測試組 C	總平均	進步率
Baseline	85.22	83.99	80.67	83.82	
FES	81.93	82.74	75.12	80.89	-18.10
SLEN Ⅱ	82.95	84.34	75.53	82.02	-11.11
PHEQ on Energy	86.65	86.03	79.27	84.92	6.82
LERN Ⅰ (M = 100)	86.31	86.27	81.22	85.28	9.00
LERN Ⅱ (α: 0.98 β: 0.25)	86.47	85.72	81.15	85.11	7.95

表 4.2.12 複合情境訓練模式下綜合實驗結果

4.3 音框對數能量正規化與倒頻譜正規化法之加成性

4.3.1 倒頻譜正規化法(Cepstral Mean and Variance Normalization, CMVN)

倒頻譜正規化法[Viikki and Laurila 1998]主要是減去倒頻譜特徵參數的平均值並針對特徵向量的標準差做正規化。假設一句話經特徵擷取後為一連串倒頻譜特徵向量 $C=\{C_1,C_2,\cdots C_t,\cdots,C_T\}$,$t=1,\cdots,T$，C_t 代表這語句的第 t 個特徵向量，T 為這語句的總特徵向量個數。最後在倒頻譜上經過倒頻譜正規化法處理後，可以適度的減少由不同的通道所造成不匹配影響。則倒頻譜正規化法求得的新特徵向量為 $\hat{C}_t$ ，如式(4.3.1)：

$$\hat{C}_t=\frac{C_t[n]-\mu[n]}{S[n]} \quad ,t=1,\cdots,T \tag{4.3.1}$$

其中

$$\mu[n]=\frac{1}{T}\sum_{t=1}^{T}C_t[n] \tag{4.3.2}$$

$$S[n]=\sqrt{\sum_{t=1}^{T}(C_t[n]-\mu[n])^2/T} \tag{4.3.3}$$

倒頻譜正規化法除了減少通道效應所造成的干擾外，同時也正規化語音特徵的機率分布，使各維度的語音特徵機率分布能夠標準化。也因為倒頻譜正規化法在語音辨識技術中效果佳並已被廣泛使用，所以特別將音框對數能量正規化於倒頻譜正規化法實驗其加成性的結果。

4.3.2 實驗結果

實驗結果如表 4.3.1 與表 4.3.2，表中我們比較基礎實驗(Baseline)結果、對數能量尺度重刻法 I (LERNI)結果、倒頻譜正規化法(CMVN) 結果和對數能量尺度重刻法 I 加上倒頻譜正規化法(LERNI+CMVN)的結果，從總平均正確率來看，在乾淨環境訓練模式下我們得知當倒頻譜正規化法加上對數能量尺度重刻法 I 的結果會有最佳的正確率，並且從各組別中可以發現對數能量尺度重刻法 I 對於不同的噪音環境的干擾加上倒頻譜正規化法都可以有正確率加成的作用，並沒有因為不同的噪音而下降辨識率。其次比較在複合情境訓練模式下的總平均正確率，隨然沒有特別提高辨識率，但正確率同樣高達 90.01%，顯示在複合情境訓練模式下對數能量尺度重刻法 I 亦不會嚴重干擾倒頻譜正規化法的效果。

乾淨環境訓練模式				
方法	測試組 A	測試組 B	測試組 C	總平均
Baseline	58.94	58.48	59.97	58.96
LERN	74.36	76.72	63.83	73.20
CMVN	77.27	80.40	72.83	77.64
LERNI+CMVN	80.41	82.98	76.63	80.68

表 4.3.1 對數能量尺度重刻法 I 與倒頻譜正規化法之加成性實驗(乾淨環境訓練模式)

複合情境訓練模式				
方法	測試組 A	測試組 B	測試組 C	總平均
Baseline	85.22	83.99	80.67	83.82
LERN	86.31	86.27	81.22	85.28
CMVN	90.30	90.50	88.48	90.01
LERNI+CMVN	90.46	90.42	88.33	90.01

表 4.3.2 對數能量尺度重刻法 I 與倒頻譜正規化法之加成性實驗(複合情境訓練模式)