

國立交通大學

電信工程研究所

碩士論文

以聲道差及回授作語音分離之研究

Source Separation Based On Binaural Cues And
FeedBack

研究生： 劉俊麟

指導教授： 謝世福 博士

中華民國 一百零三年 一月

摘要

雙聲道語音分離技術中是利用聲道間的強度差與相位差資料經統計分群後達到將混合聲音中分離出各別聲源的目的，而本篇論文內容針對雙麥克風接收的資料做時頻分析與機率統計並加入回授機制。經時頻分析後截取左右聲道強度平面以及聲道間的強度差、相位差平面，兩平面的資料以內部相關外部不相關的方式建立結合高斯機率模型，也就是獨立結合做的預分群演算法，而這些結合高斯機率模型的統計參數，可經由Expectation-Maximization EM 演算法估計，最後以機率遮蔽模式從混合聲音分離出不同聲源。回授部分是藉此分離的聲源進一步幫助先前結合高斯機率模型的參數估計，以達更好的分離效果。最後以電腦模擬驗證吾人提出之方法相較於其他演算法在 Signal-To-Distortion ratio(SDR)測試標準上皆有 2dB 以上的改善，且在錄音筆實錄測試上也有 1dB 以上的改善，證明此方法的實用性。

ABSTRACT

Binaural source separation aims to isolate individual sound source from mixture by clustering interaural phase and level differences data. In this thesis, we perform statistical analysis of interaural spectrogram and incorporate feedback mechanism. A joint Gaussian Mixture Model (GMM) is built for binaural cues and the microphone intensities with various degrees of correlations. The GMM parameters can be estimated by Expectation-Maximization algorithm. Probabilistic masking follows GMM to separate sound sources. These estimated sound sources can be fed back to enhance EM estimation. Computer simulations show that our algorithm has at least 2dB improvement in signal-to-distortion ratio (SDR). In real tests, 1dB SDR improvement can be attained.

目錄

第一章 簡介	1
2.1 1.1 研究背景	1
2.2 1.2 雙聲道語音分離演算法比較	2
2.3 1.3 論文貢獻	5
第二章 資料模型與演算法建立	7
2.4 2.1 雙聲道接收模型	7
2.5 2.2 資料擷取	7
2.6 2.3 資料相關性	9
2.6.1 2.3.1 聲道差平面資料關連性	9
2.6.2 2.3.2 左右聲道平面資料關連性	12
2.7 2.4 機率模型	13
2.7.1 2.4.1 聲道差平面機率模型	14
2.7.2 2.4.2 左右聲道平面機率模型	15
2.8 2.5 聲道差與左右聲道結合演算法	17
2.8.1 2.5.1 非獨立結合	17
2.8.2 2.5.2 獨立結合(預分群)	20
2.9 2.6 結合高斯模型參數估計	22
2.9.1 2.6.1 E-step	22
2.9.2 2.6.2 M-step	22
2.10 2.7 分離訊號回授	25
第三章 電腦模擬	27
2.11 3.1 模擬環境設定	27
2.12 3.2 左右聲道及聲道差資料分布情況	28

2.12.1 3.2.1 左右聲道差(ILD 部分)	28
2.12.2 3.2.2 左右聲道差(IPD 部分)	30
2.12.3 3.2.3 Phase ambiguous 問題	31
2.12.4 3.2.4 左右聲道強度	33
2.13 3.3 以往演算法效能比較	34
2.14 3.4 機率模型測試	34
2.14.1 3.4.1 聲道差平面分頻獨立性模擬	35
2.14.2 3.4.2 左右聲道平面獨立性模擬	36
2.15 3.5 聲道差與左右聲道結合演算法比較	37
2.15.1 3.5.1 左右聲道平面機率模型	38
2.15.2 3.5.2 聲道差平面機率模型	40
2.16 3.6 加入回授機制後成效與估計性能	41
2.16.1 3.6.1 MESSL-SP 加入回授機制效能比較	41
FB - ILD 參數估計比較	42
IPD 參數估計比較	44
3.6.2 條件結合加入 Feed Back 機制效能比較	45
2.17 3.7 三聲源分離情況下討論	47
2.17.1 3.7.1 三聲源下演算法成效比較	47
2.17.2 3.7.2 機率模型比較	48
2.17.3 3.7.3 三聲源下加入回授機制	50
2.18 3.8 實測效能	52
第四章 結論	54
Chapter 3 參考資料	55

圖目錄

圖. 1	雙聲道語音分離概念圖.....	1
圖. 2	2D-histogram of (ILD,ITD) , 左圖單一聲源；右圖六組聲源	1
圖. 3	前半部雙聲道語音分離架構圖.....	2
圖. 4	後半部雙聲道語音分離概念圖.....	2
圖. 5	二維 GMM 分法流程圖.....	3
圖. 6	HRTF 分法流程圖	4
圖. 7	MESSL 演算法流成圖.....	4
圖. 8	MESSL-SP 演算法流程圖	5
圖. 9	雙聲道接收示意圖	9
圖. 10	ILD 與 IPD 理論曲線(500Hz).....	10
圖. 11	$[-\pi, \pi]$ 區間下 ILD 與 IPD 理論曲線(500Hz).....	10
圖. 12	ILD 與 IPD 相關係數圖.....	11
圖. 13	單聲源下 ILD 與 IPD 資料分布比對圖(500Hz).....	11
圖. 14	單聲源下 ILD 與 IPD 資料分布比對圖(1500Hz).....	11
圖. 15	單聲源下 ILD 與 IPD 資料分布比對圖(5000Hz).....	11
圖. 16	L 與 R 理論曲線(ILD=10)	12
圖. 17	左右聲道相關係數圖.....	13
圖. 18	單聲源 L 、 R 資料分布(500Hz)	13
圖. 19	單聲源 L 、 R 資料分布(7500Hz)	13
圖. 20	實際聲道差平面資料統計圖(500Hz)	15
圖. 21	實際聲道差平面資料分布圖(500Hz)	15
圖. 22	實際左右聲道平面資料統計圖(500Hz)	16
圖. 23	實際左右聲道平面資料分布圖(500Hz)	16
圖. 24	ILD , IPD , L 、 R 彼此關係圖	17
圖. 25	ILD 與左聲道(L)相關係數圖.....	18
圖. 26	ILD 與右聲道(R)相關係數圖	18
圖. 27	IPD 與左聲道(L)相關係數圖	19
圖. 28	IPD 與右聲道(R)相關係數對頻率做圖	19
圖. 29	非獨立結合演算法流程圖	19
圖. 30	ILD , IPD , L , R彼此關係圖(修正後).....	20
圖. 31	獨立結合(預分群)流程圖	21
圖. 32	引進回授概念的 MESSL-SP 流程圖	25
圖. 33	雙聲源混合前 ILD 資料分布	28
圖. 34	雙聲源混合後 ILD 資料灰階圖.....	29
圖. 35	三聲源混合前 ILD 資料分布	29

圖. 36 三聲源混合後 ILD 資料灰階圖.....	29
圖. 37 雙聲源混合前 IPD 資料分布.....	30
圖. 38 雙聲源混合後 IPD 資料灰階圖.....	30
圖. 39 三聲源混合前 IPD 資料分布.....	31
圖. 40 三聲源混合後 IPD 資料灰階圖.....	31
圖. 41 正常情況下混合前 α, ϕ 平面資料分布	31
圖. 42 正常情況下混合後 α, ϕ 平面資料分布	32
圖. 43 <i>phase ambiguous</i> 下混合前 α, ϕ 平面資料分布	32
圖. 44 <i>phase ambiguous</i> 下混合後 α, ϕ 平面資料分布	32
圖. 45 雙聲源混合前 L, R 平面資料分布(500 Hz).....	33
圖. 46 雙聲源混合後 L, R 平面資料分布(500 Hz) , 兩條直線為理想分布曲線	33
圖. 47 以往演算法左聲道成效圖.....	34
圖. 48 以往演算法右聲道成效圖	34
圖. 49 聲道差平面獨立性成效比對圖(左聲道) , x 線為 MESSL(全頻獨立)；o 線為全頻非獨立；+線為 1000Hz 分頻處理； \diamond 為 500Hz 分頻處理	35
圖. 50 聲道差平面獨立性成效比對圖(右聲道) , x 線為 MESSL(全頻獨立)；o 線為全頻非獨立；+線為 1000Hz 分頻處理； \diamond 為 500Hz 分頻處理	35
圖. 51 $P_{\alpha, \phi, dep}$ 機率等高線圖(1500Hz) , 綠點為 Data	36
圖. 52 $P_{\alpha, \phi, indep}$ 機率等高線圖(1500Hz) , 綠點為 Data	36
圖. 53 PL, R, dep 機率等高線圖 (1500Hz) , 綠點為 Data	37
圖. 54 $PL, R, indep$ 機率等高線圖 (1500Hz) , 綠點為 Data	37
圖. 55 兩平面結合演算法成效圖(左聲道).....	38
圖. 56 兩平面結合演算法成效圖(右聲道).....	38
圖. 57 非獨立結合左右聲道邊際機率等高線圖.....	39
圖. 58 獨立結合(預分群)左右聲道邊際機率等高線圖.....	39
圖. 59 $P_{\alpha, \phi, dep}$ 機率等高線圖	40
圖. 60 $P_{\alpha, \phi, Cdep}$ 機率等高線圖	41
圖. 61 回授機制成效圖(左聲道)	41
圖. 62 回授機制成效圖(右聲道)	42
圖. 63 上圖為 SDR - $\Delta\mu\alpha$ 做圖；下圖為 $\Delta\mu\alpha$ - 角度做圖.....	43
圖. 64 $\mu\alpha, 1, error$ 曲線圖 , MESSL-SP(黑)、FB-ILD(紅)	43
圖. 65 $\mu\alpha, 2, error$ 曲線圖 , MESSL-SP(黑)、FB-ILD(紅)	43
圖. 66 $\Delta\mu\alpha$ 曲線圖 , MESSL-SP(黑)、FB-ILD(紅)	44
圖. 67 $\mu\phi, 1, error$ 曲線圖 , MESSL-SP(黑)、FB-IPD(紅)	44
圖. 68 $\mu\phi, 2, error$ 曲線圖 , MESSL-SP(黑)、FB-IPD(紅)	45
圖. 69 $\Delta\mu\phi$ 曲線圖 , MESSL-SP(黑)、FB-IPD(紅)	45
圖. 70 雙聲源獨立結合(預分群)回授演算法成效(左聲道)	46

圖. 71	雙聲源獨立結合(預分群)回授演算法成效(右聲道)	46
圖. 72	μ_1, error 曲線圖，獨立結合(黑)、獨立結合+回授(紅)	46
圖. 73	μ_2, error 曲線圖，獨立結合(黑)、獨立結合+回授(紅)	47
圖. 74	$\Delta\mu$ 曲線圖，獨立結合(黑)、獨立結合+回授(紅)	47
圖. 75	三聲源演算法成效圖(左聲道).....	48
圖. 76	三聲源演算法成效圖(右聲道).....	48
圖. 77	三聲源 MESSL-SP 聲道差平面機率等高線圖(1500Hz)，綠點為 Data ..	49
圖. 78	三聲源 MESSL 修正型聲道差平面機率等高線圖(1500Hz)，綠點為 Data ..	49
圖. 79	三聲源獨立結合(預分離)聲道差平面機率等高線圖(1500Hz)	49
圖. 80	MESSL-SP 左右聲道平面機率等高線圖(1500Hz)	50
圖. 81	獨立結合(預分群)左右聲道平面機率等高線圖(1500Hz).....	50
圖. 82	三聲源獨立結合(預分群)回授成效(左聲道).....	51
圖. 83	三聲源獨立結合(預分群)回授成效(右聲道).....	51
圖. 84	μ_1, error 曲線圖，獨立結合(黑)、獨立結合+回授(紅)	51
圖. 85	μ_1, error 曲線圖，獨立結合(黑)、獨立結合+回授(紅)	51
圖. 86	μ_1, error 曲線圖，獨立結合(黑)、獨立結合+回授(紅)	52
圖. 87	綠音設備：(1)麥克風 (2)綠音筆 (3)組合	52
圖. 88	實測擺設	53

表目錄

表. 1	以往演算法差別.....	3
表. 2	實測左邊音源分離 SDR 成果	53
表. 3	實測右邊音源分離 SDR 成果	53

第一章 簡介

1.1 1.1 研究背景

語音分離技術可以運用在助聽器、機器人音訊處理、語音辨識...等方面。在語音分離技術中，起始採用多麥克風接收分離系統[10],[11]，他們是藉由矩陣運算的方式來分離語音(但只能分離出與麥克風同數目的聲源)，由於麥克風數目的限制，近十年來的研究改以雙聲道語音分離如下圖.1 所示，目的是由雙耳接收的訊號 l 、 r 分出原聲音 s_1 及 s_2 。其截取雙聲道間的資料，也就是 *interaural level differences (ILD)*、*interaural time differences (ITD)* 及 *interaural phase differences (IPD)*，利用它們與各聲源間的關係，做到語音分離的效果是常見的手法，這種手法在過去的幾年中已有不少人提出，但依然存在著問題，我們先探討過去幾年的研究。

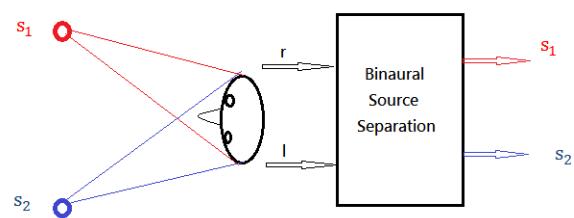


圖. 1 雙聲道語音分離概念圖

DUET(degenerate unmixing estimation technique)[1]便是這類研究的起始者，它採用 Spectrogram 的方式截取每個 T-F frame 中的 ILD 及 ITD，並放下聲音獨佔假設 WDO(Window Disjoint orthogonal)，接著用 2D-histogram 的統計，發現不同於每個聲源都有相應的一個頂點(如下圖圖.2 所示)，接著他用 ML rule 概念畫分(ILD,ITD)平面區塊，做到二分法的分離。DUET 依然有 2 個缺點，第一個是對於 ITD 截取技術的不純熟 (phase ambiguity 未完全解決)，使得它的系統在分離 1500Hz 以上的聲源時成效下降取多；第二個是二分法過於太硬的不合理處。

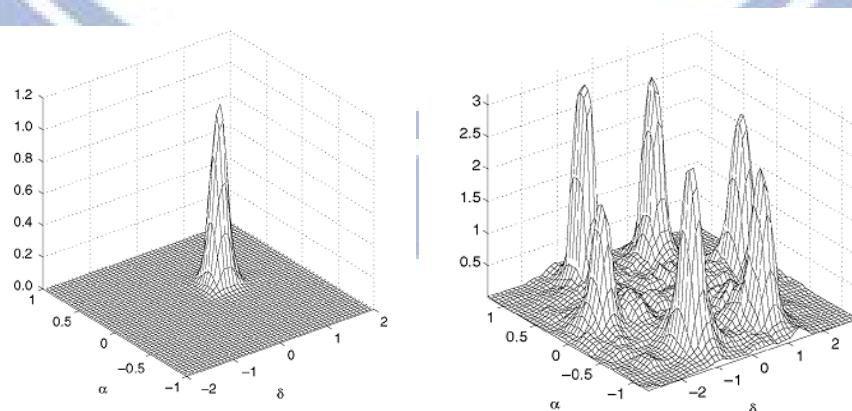


圖. 2 2D-histogram of (ILD,ITD) , 左圖單一聲源；右圖六組聲源

對於 DUET 的不完美處，新的研究，改用新的 ILD、ITD、IPD 截取技術和改用機率分法取代二分法的概念，引進結合高斯模型(GMM)，的想法出現，例如：[2]~[7]篇論文所研討的，但基於各方法對 GMM 模型的假設方式不同，下段落用架構圖及表格的方式說明他們之間差別。

1.2 1.2 雙聲道語音分離演算法比較

雙聲道語音分離技術的架構圖可分為前半部與後半部兩部分，前半部(圖.3)是資料劫取的部分大致整理為下列幾個方塊所組成，首先進入的 s_1, s_2 聲源經過他們各自的 channel 後混合以雙聲道接收 $l(t)、r(t)$ ，接著盡到 Spectrogram 中找出相應的 T – F domain，再以 cue estimate 截取出 ILD、IPD、ITD 等資料，這是一套固定的先處理動作。

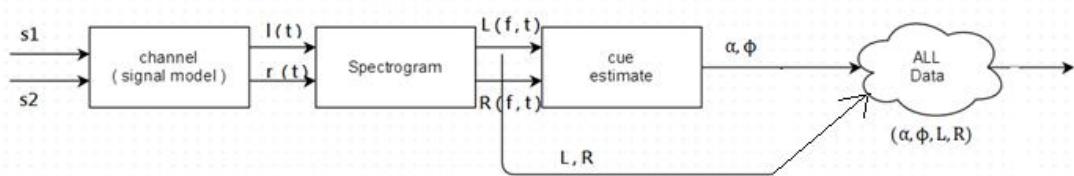


圖. 3 前半部雙聲道語音分離架構圖

至於後半部(圖.4)，就是個各類研究的演算法核心部分，大致可以區分為兩塊

(1) Mask：利用 decision rule 將聲音資料分開，做到語音分離的最後步驟。

(2) 演算法：找出最佳的 decision rule，以及建立 Decision rule 所需要的模型。

所有演算法的想法出發點都是由 Mask 所決定，例如原始的[1]DUET 中採用二分法 ML rule，因此只需由資料做統計，而不需要資料模型的架構，反之[2]~[7]中採用機率 Mask，需要機率模型的架構。

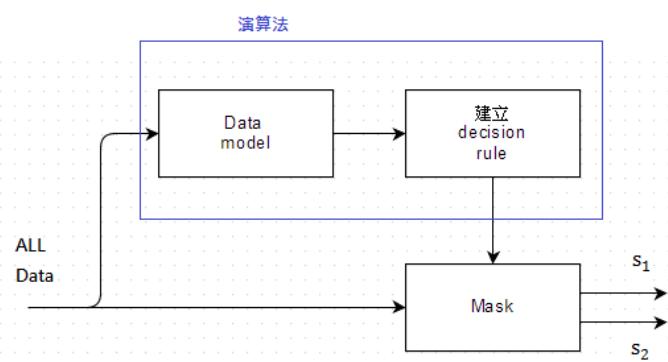


圖. 4 後半部雙聲道語音分離概念圖

我們先以表格 (Table.1)的方式說明他們在想法上的不同處之後一一以架構圖的方式說明各種演算法。

表.1 以往演算法差別

方法	Mask	機率模型間的關係	Total likelihood (GMM)	需要參數	decision
DUET	Binary Mask	---	---	Peak (mean)	ML rule
二維 GMM 分法	Probability Mask	相關	$p_{\alpha, \tau}$	Mean variance	Z_i (資料權重)
HRTF 分法	Probability Mask	相關	p_{θ}	Mean Variance	Z_i
MESSL	Probability Mask	不相關	$p_{\alpha} \cdot p_{\phi}$	Mean Variance	Z_i
MESSL-SP	Probability Mask	不相關 (S.I)	$p_{\alpha} \cdot p_{\phi} \cdot p_L \cdot p_R$	Mean Variance	Z_i
MESSL-EV	Probability Mask	不相關 (S.D)	$p_{\alpha} \cdot p_{\phi} \cdot p_L \cdot p_R$	Mean variance	Z_i

二維 GMM 分法[2]：如圖.5，使用機率 Mask，採用(ILD,ITD)所建立的二維 GMM 模型為基礎達到分離效果。架構圖如下，預先採用 ILD 來做到預先分離的效果，之後以 (ILD,ITD) 建立二維的機率模型，找出分離依據達到聲音分離。此演算法也在房間位子上做訓練提高了成效，但基於對房間的特定性，是個問題。

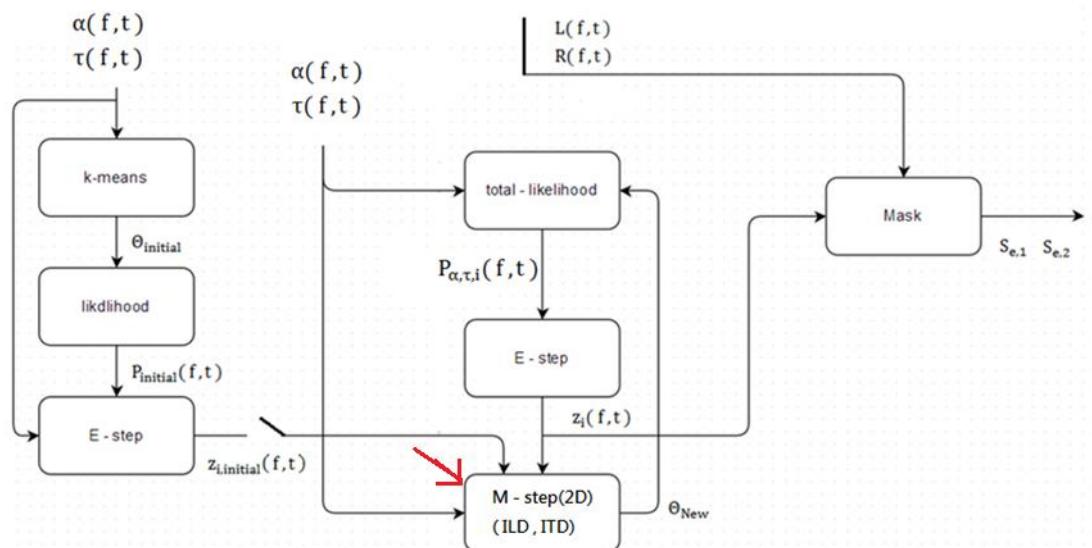


圖.5 二維 GMM 分法流程圖

HRTF 分法[4]：如圖.6，使用機率 Mask，模型方面先將 ILD、ITD 利用了 Head-Relation-Transform-function (HRTF) 資料庫之中”方位角”與 ILD 及 ITD 間的關係，將

ILD、與 ITD 經由查表的方式轉成兩組 azimuth_L 及 azimuth_T 後共同決定出一組 azimuth (降維)，再建立機率一維 GMM 模型。

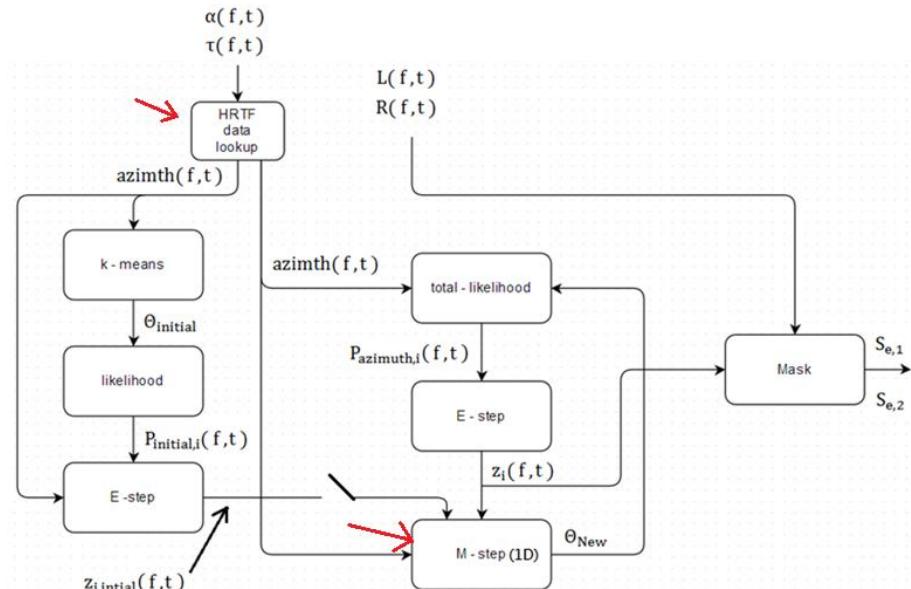


圖. 6 HRTF 分法流程圖

MESSL[5]：如圖.7，使用機率 Mask，模型方面建立兩個一維 GMM 並以獨立的假設結合在一起。如圖先以 ILD 預先分離後，在分別建立了 ILD 及 IPD 一維的機率模型，並以"獨立"的方式將它結合。

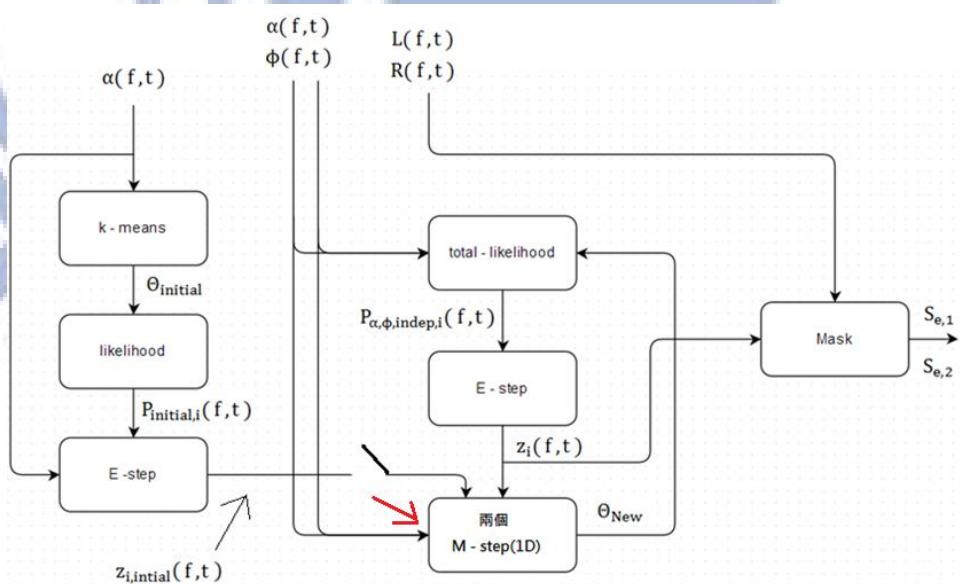


圖. 7 MESSL 演算法流成圖

MESSL-SP[6]：如圖.8，是基於 MESSL 的方法上，除了 ILD 及 IPD 還加上語音辨識的概念建立左右聲道強度模型，並共同達到分離(一樣採獨立方式)。(所給定的 source

prior 採用 Speaker-independent 的方式訓練)

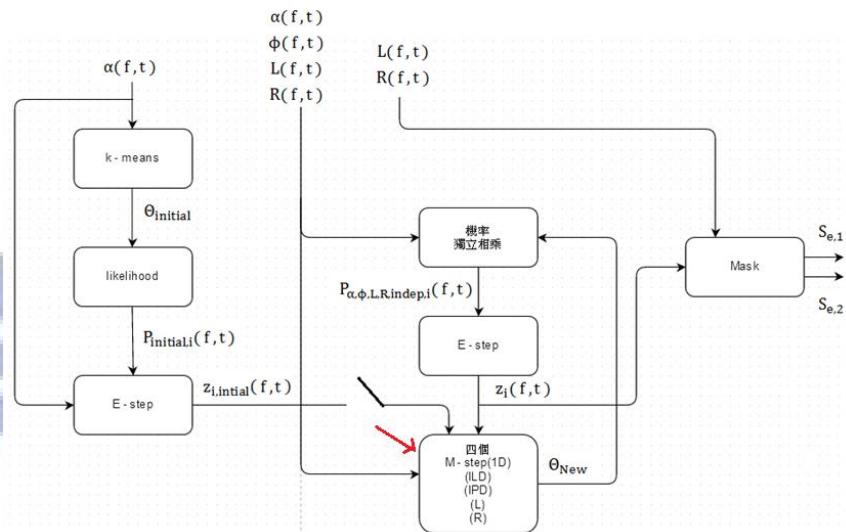


圖. 8 MESSL-SP 演算法流程圖

MESSL-EV [7] 則是 MESSL-SP 的改良版，他引進了語音辨識中 Eigen voice 的概念在左右聲道模型上做加強。(所給定的 source prior 採用 Speaker-dependent 的方式訓練)

以上的方法會在 3.3 節中先模擬並予以比較，比較結果最出色的是二維 GMM 分法，而值得一提的是引進”左右聲道資料”的 MESSL-SP 確實有比 MESSL 好，這代表加入”左右聲道資料”是有益的。

1.3 1.3 論文貢獻

本篇論文主要目的為雙聲道語音分離技術，而在過程中主要的貢獻可分為 5 點，如下：

(1) 非獨立結合 ILD、IPD 機率模型：

我們深入研究 ILD 與 IPD 的相關性與機率模型間的獨立性問題，從實際的資料分佈及資料間相關性我們發現，ILD 與 IPD 間機率模型關係並非獨立假設這麼容易，而是有相當大相關的非獨立關係，但若有 phase ambiguous 發生時，可以區別為高相關頻段(低頻)以及低相關頻段(高頻段)，但經由模擬測試，發現全頻段上採用一律非獨立的方式結合 ILD 及 IPD 兩模型最好。

(2) 以二維 GMM 處理左右聲道平面：

除了原本 ILD、IPD 資料外，我們引進近年提出的 ”左右聲道 L、R” 資料一並用於語音分離。近年的研究是採用語音辨識的方式來幫助語音分離，而我們的做法是採用強度方面，且用左右聲道強度平面建立二維機率模型，這麼做的原因是語音辨識系統中單聲源就採用 GMM 來做辨識，當有複數聲源時就難以看到區別

性。

(3) 獨立結合(預分群)方式結合聲道差與左右聲道兩平面：

假設兩平面獨立，避免左右聲道難分性的問題牽連聲道差平面，當然這不是完全否定兩平面間的關聯性，其中以聲道差平面資訊幫助左右聲道平面的機率正確估計，連結兩平面，此方法稱“獨立結合(預分群)”，在成效模擬實，它確實解決難分性問題，且比 MESSL 演算法(近年地位較大的演算法)好上 2~3dB。

(4) 加入分離訊號能回授機制：

對原本 EM 演算法做改善，將每次輪迴分離好的資料回授取代原本混合資料以求更精準的參數估計。

(5) 實測：

最後我們用錄音筆做實際的測試，測試結果證實我們所提出的演算法比上述別人的演算法都來的好。

論文章節簡述如下：在第二章中，大致可分為兩部分，第一部分為雙聲道語音分離前半部的細部說明，第二部分為我們所提出的演算法(詳細介紹可於第二章開頭中看到)。在第三章是電腦模擬的部分，其中包括了以往演算法的成效比較、我提出演算法的成效比較、實測部分。最後第四章為本論文的結論。

第二章 資料模型與演算法建立

本段落為雙聲道語音分離所需全部的流程，其中包括雙聲道接收模型(2.1 節)、資料擷取與資料類型間關係討論(2.2~2.3 節)、機率模型建立(2.4 節)、我所提出的演算法(2.4~2.7 節)。

1.4 2.1 雙聲道接收模型

如圖.2 所示，signal model 部分我們模擬各聲源經由各自的 delay 及 HRTF 響應後由雙麥克風接收的情形，如下(1)式。

$$l(t) = \sum_i s_i(t - \tau_i^l) * h_i^l(t) \quad \dots(1.1)$$

$$r(t) = \sum_i s_i(t - \tau_i^r) * h_i^r(t) \quad \dots(1.2)$$

而接收到的左右聲道經 spectrogram 並在 Window Disjoint Orthogonal (WDO 聲音獨佔)假設下，可寫成下兩式

$$L(t, f) = \sum_i S_i(f, t) H_i^l(f, t) e^{-j(\omega \tau_i^l + \angle H_i^l(f))}$$
$$\xrightarrow{WDO} S_i(f, t) H_i^l(f) e^{-j(\omega \tau_i^l + \angle H_i^l(f))} \quad \dots(2.1)$$

$$R(t, f) = \sum_i S_i(f, t) H_i^r(f, t) e^{-j(\omega \tau_i^r + \angle H_i^r(f))}$$
$$\xrightarrow{WDO} S_i(f, t) H_i^r(f) e^{-j(\omega \tau_i^r + \angle H_i^r(f))} \quad \dots(2.2)$$

1.5 2.2 資料擷取

我們先建立 interaural spectrogram，其中(3.1)式為在實際場景下的觀察(沒有 WDO 假設)，而(3.2)式為 signal model 所建立的模型推導結果。

Interaural Spectrogram :

$$X_{IS}(f, t) = \frac{L(f, t)}{R(f, t)} = 10^{\alpha(f, t)/20} \cdot e^{-j\phi(f, t)} \quad \dots \quad (3.1)$$

$$\xrightarrow{\text{WDO}} \frac{H_i^l(f)}{H_i^r(f)} e^{-j\omega(t_i^l + t_i^r + \arg\left(\frac{H_i^l(f)}{H_i^r(f)}\right))} \quad \dots \quad (3.2)$$

Let $\left(\tau_i^l - \tau_i^r + \arg\left(\frac{H_i^l(f)}{H_i^r(f)}\right) \right) = \tau$

經由(3.1)式和(3.2)式的對照，可以從中找出我們用於雙聲道語音分離所需要的資料，其中可分為兩種資料群組，分別為聲道差群組包含：強度差、相位差、時間差；以及左右聲道群組包含：左聲道強度、右聲道強度，如下方塊。

聲道差群組：

強度差 ILD :

$$\alpha(f, t) = 20 \log_{10}(|X_{IS}(f, t)|) \xrightarrow{\text{WDO}} 20 \log_{10} \left| \frac{H_i^l(f)}{H_i^r(f)} \right| \quad \dots (4.1)$$

相位差 IPD :

$$\phi(f, t) = \arg(X_{IS}(f, t)) \xrightarrow{WDO} \arg(\omega(\tau_i^l + \tau_i^r + \arg(\frac{H_i^l(f)}{H_i^r(f)}))) \quad \dots(4.2)$$

時間差 ITD :

$$\tau(f, t) = \frac{\phi(f, t)}{2\pi f} \xrightarrow{\text{WDO}} \left(\tau_i^l - \tau_i^r + \arg\left(\frac{H_i^l(f)}{H_i^r(f)}\right) \right) \quad \dots(4.3)$$

左右聲道群組：

左聲道強度：

$$\hat{L} = 20 \log_{10}(L) \xrightarrow{\text{WDO}} 20 \log_{10}(S_i(f, t) H_i^1(f)) \quad \dots(5.1)$$

右聲道強度：

$$\widehat{R} = 20 \log_{10}(R) \xrightarrow{\text{WDO}} 20 \log_{10}(S_i(f, t) H_i^r(f)) \quad \dots(5.2)$$

1.6 2.3 資料相關性

由於雙聲道語音分離領域中佔有大分量的 MESSL 演算法中，它以獨立的簡化假設來結合 ILD 與 IPD 的機率模型，而我們相信 ILD 與 IPD 之間應該有不斐的關連性在；而在左右聲道平面 \hat{L} 與 \hat{R} 之間也應存在著關連性。因此，我們這章節的目的在於找出聲道差平面和左右聲道平面它們各自內資料形正確關連性。

1.6.1 2.3.1 聲道差平面資料關連性

理論上由點聲源衰減公式及相位差公式，ILD 和 IPD 都與雙麥克風距離 d_l 、 d_r 有關，而且當 d_l 比 d_r 越大時 ILD 及 IPD 都越大，換一個想法，當 ILD 變大這代表 d_l 比 d_r 的差越大，而當 d_l 比 d_r 的差越大時 IPD 也越大；ILD 會隨著 IPD 變大而變大。

點聲源衰減公式：

$$\kappa \cdot 20 \log_{10}(\text{距離}) \quad \dots(6.1)$$

衰減差(α)：

$$\kappa \cdot 20 \log_{10}\left(\frac{d_l}{d_r}\right) \quad \dots(6.2)$$

相位差理論公式：

$$\phi \approx \frac{d_l - d_r}{c} \cdot 2\pi f \quad \dots(6.3)$$

κ ：介質衰減係數

d_l ：聲源與左麥克風距離

c ：聲速

d_r ：聲源與右麥克風距離

f ：頻率

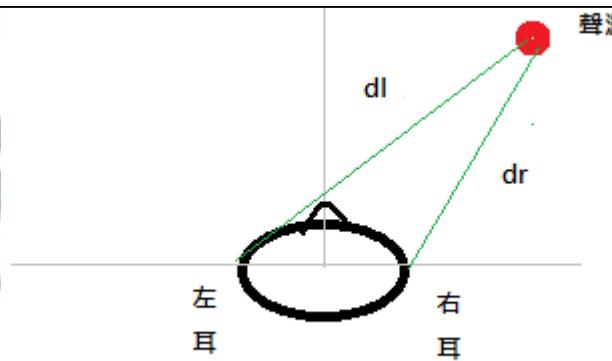


圖. 9 雙聲道接收示意圖

接著我們進一步推導出 ILD α 與 IPD ϕ 間的關係方程式，並畫出他們關係的曲線圖.10，可以看到 ILD 與 IPD 間的成長關係是類似 e^x 的成長曲線，而取 $[-\pi, \pi]$ 區間後

為圖.11。

$$\phi = \frac{2\pi f}{c} \cdot d_r (10^{\frac{\alpha}{K-20}} - 1) \quad \dots(7)$$

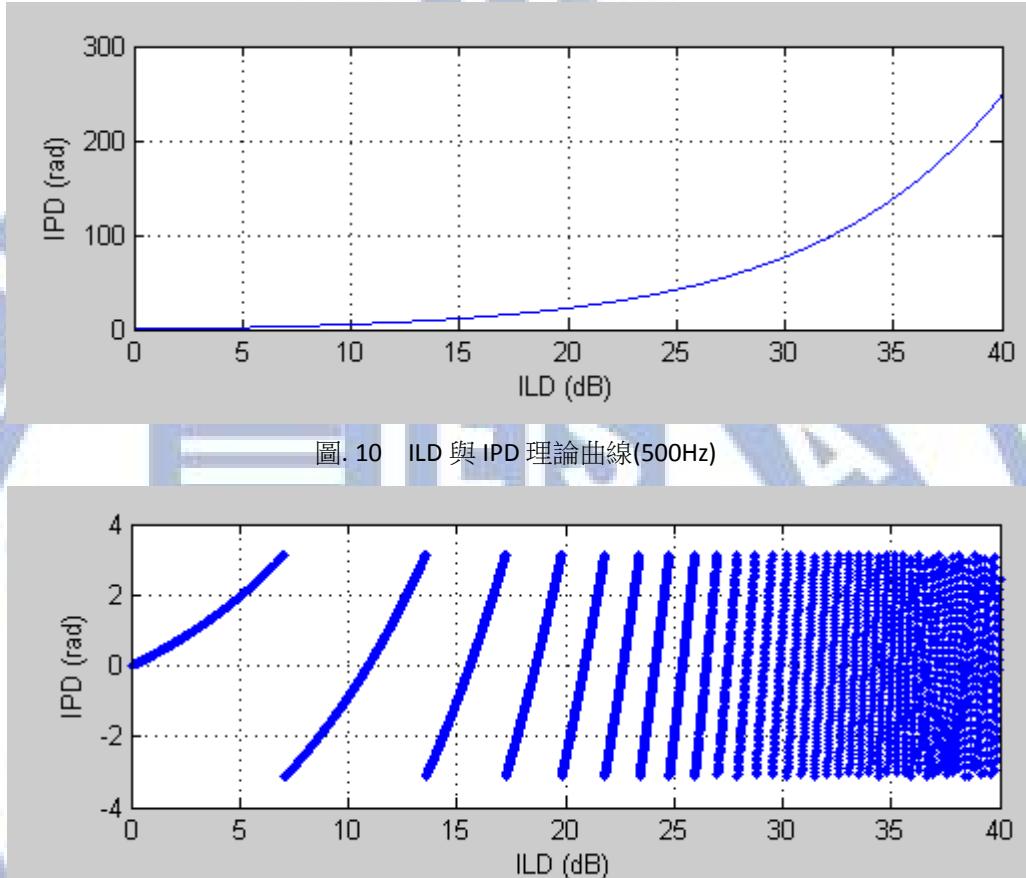


圖. 10 ILD 與 IPD 理論曲線(500Hz)

而從實際單聲源所截取出的 ILD 及 IPD 資料做相關係數的觀察看到(圖.12), ILD 及 IPD 間的相關係數在 200~1000 頻段上成線強烈的正相關性，可以和圖.13 做比對，隨著 ILD 增長 IPD 亦增長；1000~3000 頻段為負相關，這是由於 phase ambiguous 發生所導致(發生一次)，可於圖.14 中看到在 ILD 大於 10dB 的地方發生相位反轉；3000~7500 頻段為較低相關的頻段這是因為高頻段 e^x 成長曲線成長速度過大導致 phase ambiguous 的現象多次發生，使得相關係數變低，可於圖.15 中看到。

相關係數：

$$r_{\alpha, \phi} = \frac{E\{(\alpha - \bar{\alpha})^T (\phi - \bar{\phi})\}}{\sigma_{\alpha} \sigma_{\phi}} \quad \dots(8)$$

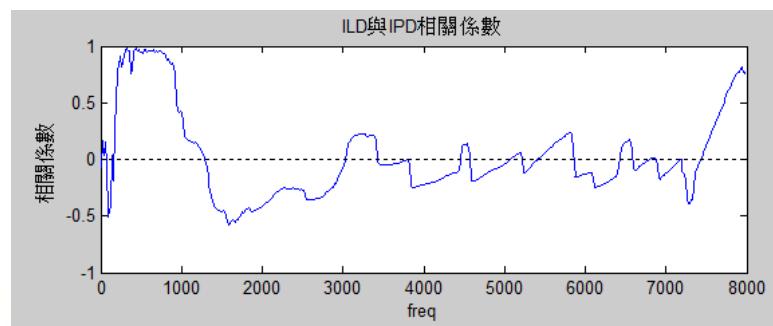


圖. 12 ILD 與 IPD 相關係數圖

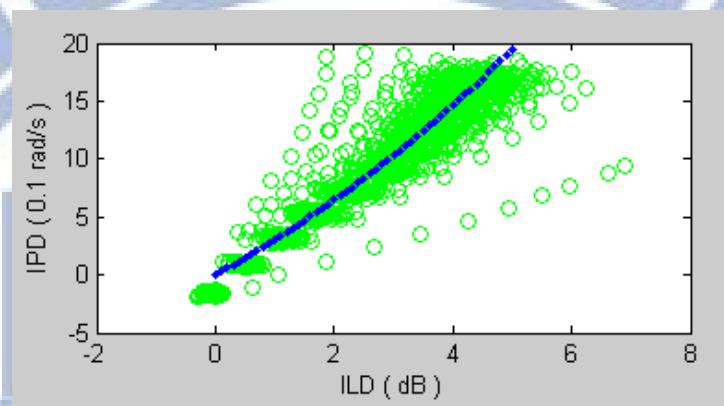


圖. 13 單聲源下 ILD 與 IPD 資料分布比對圖(500Hz)

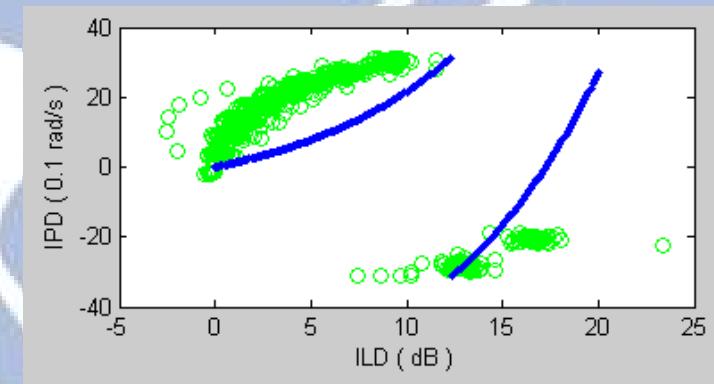


圖. 14 單聲源下 ILD 與 IPD 資料分布比對圖(1500Hz)

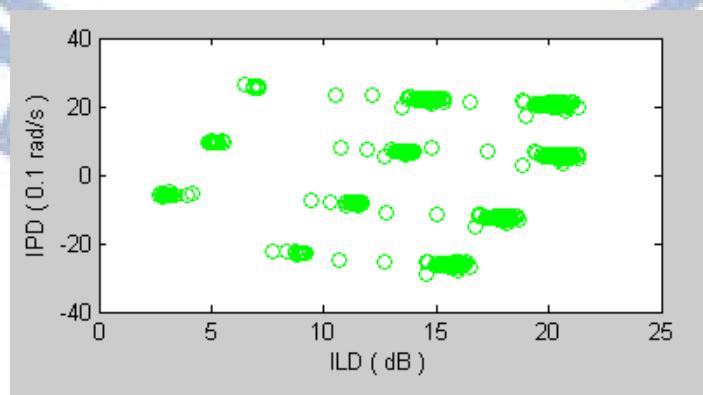


圖. 15 單聲源下 ILD 與 IPD 資料分布比對圖(5000Hz)

經過理論及實際驗證後，我們確定 ILD 與 IPD 在低頻段(1000Hz 以下)是有相當關聯性的非獨立關係，而高頻段本應有的相關性由於 phase ambiguous 的問題而無法突顯出來，整體下來並不像 MESSL 中獨立簡化假設這麼容易，這代表如果我們改掉獨立的假設，將會對整體效能會有不錯的改善。而在我們圖.12 上顯示低頻段有著高相關性，高頻段則為低相關性，這裡可以得知分頻處理應獨立性的問題是最好的方式，但由於獨立性的分頻點要取多少將會是個問題，從理論推斷，分頻點應座落於(7)式中 ϕ 第一次超越 $[-\pi, \pi]$ 區間的頻率點，而從(7)式中也看到，影響的參數除了頻率 f 外，還有聲速 c 、聲源距離 d_r 、介質衰減系數 κ 等影響，所以難以找出正確的分頻點。雖說如此，我們嘗試性的對獨立性分頻的測試(於 3.4.1 節中)，測試的結果為全頻非獨立的方式與分頻獨立的方式成效差不多，但兩種方式都比全頻獨立來的好上許多，因此最後我們還是採用全頻非獨立的方式做，避免找尋分頻點的問題。

1.6.2 2.3.2 左右聲道平面資料關連性

以理論聲音衰減公式看的話，左右聲道各自接收音量為音源強度減去各自的通道衰減，而理論上各自通道衰減理論上是固定的，所以可以說左右聲道音量都是由同一音源音量所決定，且從圖.16 中看到理論上他們資料的分佈是斜值線，因此彼此間的關聯性應該是很大。

理論左右聲道強度公式：

$$\text{左聲道接收音量} = \text{聲源音量} - \text{左聲道距離衰減}$$

$$\text{右聲道接收音量} = \text{聲源音量} - \text{右聲道距離衰減}$$

$$\text{左聲道音量} = \text{右聲道音量} - (\text{左聲道衰減} - \text{右聲道衰減})$$

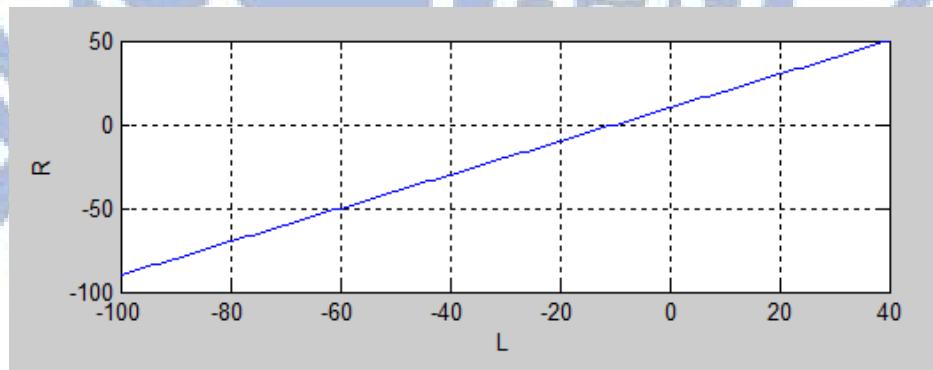


圖. 16 L 與 R 理論曲線(ILD=10)

而實際雙聲源接收下左右聲道資料的相關係數圖(圖.17)顯示：在 0~2000Hz 都為 0.8 以上的正相關，比對圖.18 也可以看到隨著 \hat{L} 上升 \hat{R} 也上升；至於高頻段相關係數

的下降是由於高頻段的聲音資料會有空白地方(能量低於-100dB)，於圖.19 中可以看到。但總歸來說 \hat{L} 、 \hat{R} 間的關聯性相當大，而不是 MESSL-SP 中所說獨立的。

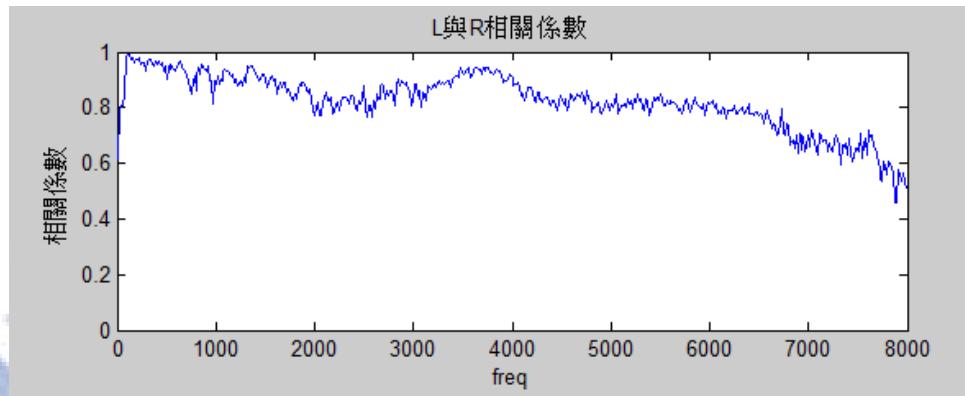


圖. 17 左右聲道相關係數圖

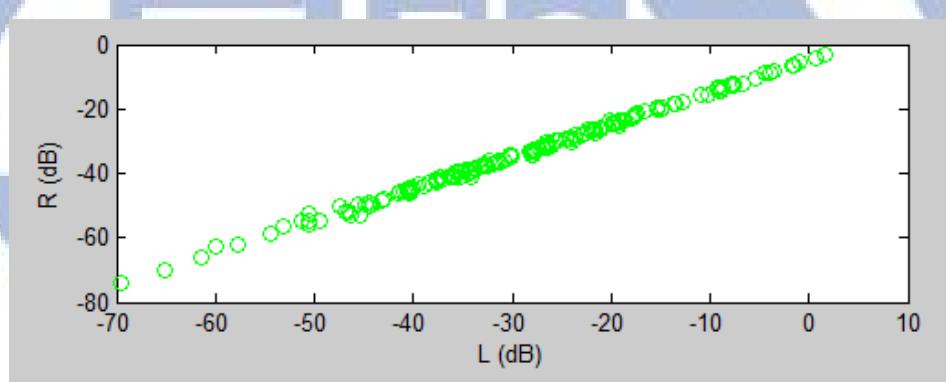


圖. 18 單聲源 \hat{L} 、 \hat{R} 資料分布(500Hz)

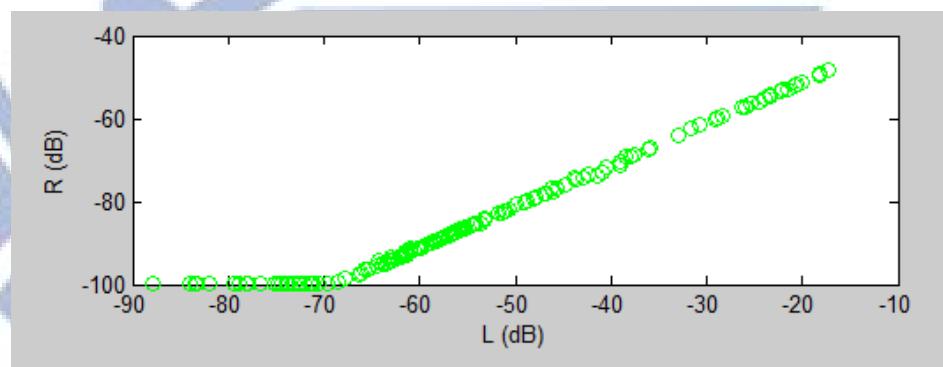


圖. 19 單聲源 \hat{L} 、 \hat{R} 資料分布(7500Hz)

在此我們也得知左右聲道平面的 \hat{L} 、 \hat{R} 之間也是有相當大關連性的非獨立關係。

1.7 2.4 機率模型

在討論完資料間彼此的關連性後，我們得知在聲道差及左右聲道內部都應該非獨

立(且相關性不小)，因此我們建立在這個基礎上開始為這兩個平面建立機率模型。

1.7.1 2.4.1 聲道差平面機率模型

以 ILD 為例：由各方位角所來的聲音在雙麥克風的接收下，因為角度的不同到達不同距離兩個麥克風的衰減也不同，但如果從麥克風角度來看來自同一聲源的衰減差為定值，而形成機率的源由是因環境和互相干擾所造成。對於每個聲源我們以高斯模型模擬(在很多文獻中都有提到且正確性也已被證實，此處不多說明)。

高斯機率模式：

$$\begin{aligned} (\alpha, \phi)_{i,f} &\sim N(\mu_{\alpha\phi,i}(f), \Sigma_{\alpha\phi,i}(f)) \\ &= \frac{1}{2\pi|\Sigma_{\alpha\phi,i}|^{1/2}} e^{-\frac{1}{2}((\alpha, \phi) - \mu_{\alpha\phi,i})^T \Sigma_{\alpha\phi,i}^{-1} ((\alpha, \phi) - \mu_{\alpha\phi,i})} \\ &\triangleq P_{\alpha, \phi, i} \end{aligned} \quad \dots(9)$$

$\mu_{\alpha\phi,i}$ ：平均值

$\Sigma_{\alpha\phi,i}$ ：共變數矩正

而各聲源間採用 Gaussian Mixture Model (GMM)模擬，如下方塊。

GMM：

$$P_{\alpha, \phi}(\alpha(f, t), \phi(f, t) | \Theta) = \sum_i \psi_i \cdot P_{\alpha, \phi, i} \quad \dots(10)$$

ψ_i ：GMM 權重(由 EM 決定)

在實際上我們觀察一組由不同方位角接收到的雙聲源中 ILD 及 IPD 的資料統計圖(圖.20)及分布圖(圖.21)，圖上可以看到有(0,0)和(-5,-20)為中心的兩個類高斯分布組成，在實際面可以看到結合高斯模型適用於此。

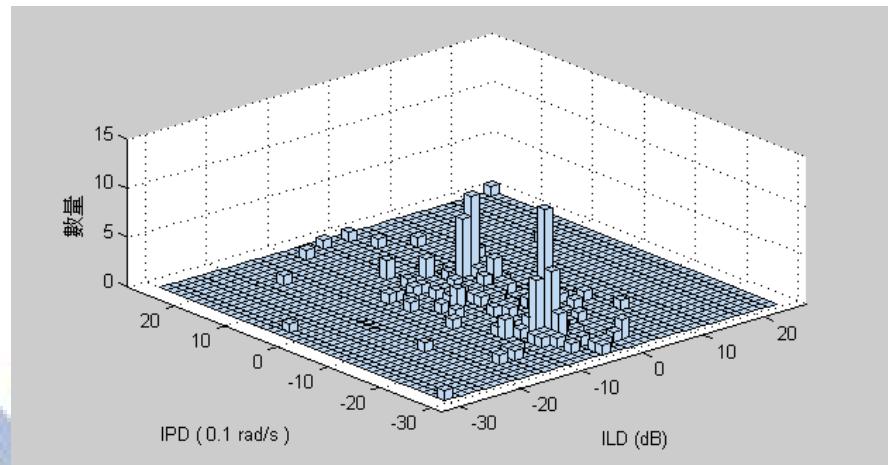


圖. 20 實際聲道差平面資料統計圖(500Hz)

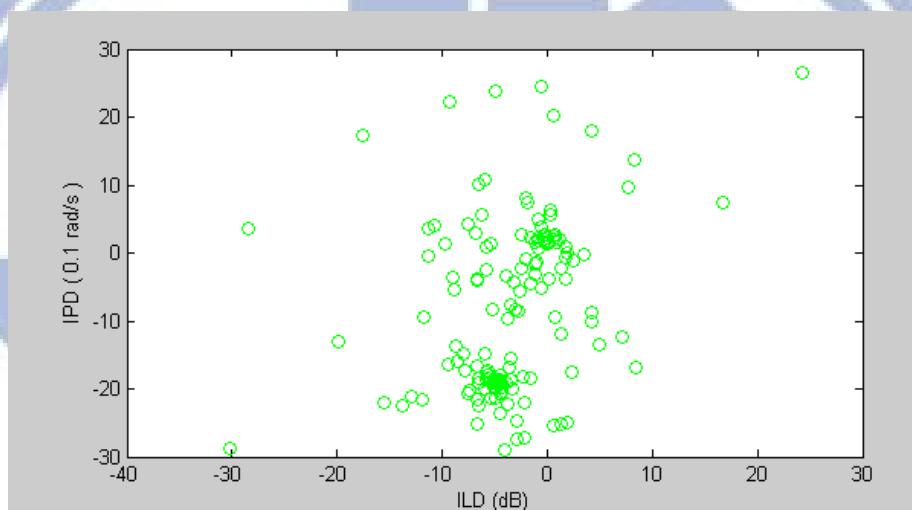


圖. 21 實際聲道差平面資料分布圖(500Hz)

1.7.2 2.4.2 左右聲道平面機率模型

由於此平面的資料是新的技術，所以我們對於模型的建立是由對在實際資料而來。左右聲道強度資料統計圖(圖.22)中看到的是成一直線很窄的山脈，看不出可分性；實際上雙聲源所擷取左右聲道強度資料分布圖(圖. 23)，圖中顯示左右聲道平面的資料分布是由兩條 ILD 不同的直線像左右散開的分布情況(詳細說明於 3.2.4)。[註]左右聲道平面並非像聲道差平面有著較久的研究歷史，但基於嘗試我們還是採用二維結合高斯模型模擬之。

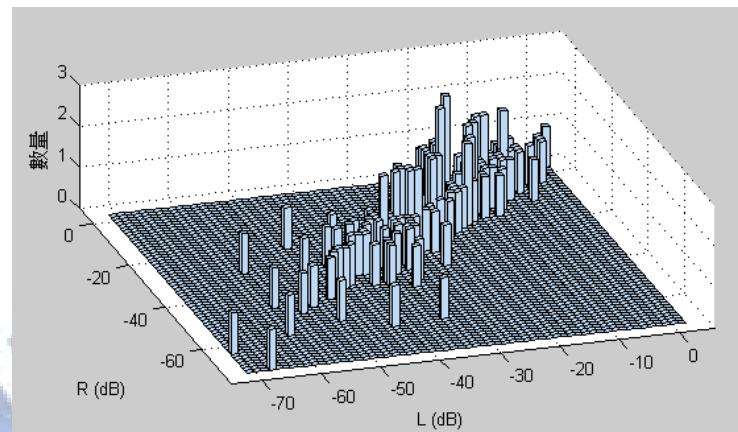


圖. 22 實際左右聲道平面資料統計圖(500Hz)

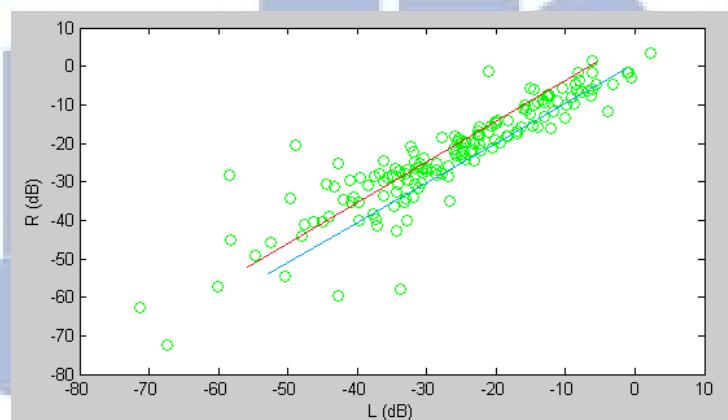


圖. 23 實際左右聲道平面資料分布圖(500Hz)

每個聲源的左右聲道以二維高斯分布模擬，標示如下：

高斯機率分布：

$$(\hat{L}, \hat{R})_{i,f} \sim N(\mu_{\hat{L}\hat{R},i}(f), \Sigma_{\hat{L}\hat{R},i}(f)) \triangleq P_{\hat{L},\hat{R},i} \quad \dots(11)$$

$\mu_{\hat{L}\hat{R},i}$ ：平均值

$\Sigma_{\hat{L}\hat{R},i}$ ：共變數矩正

GMM：

$$P_{\hat{L},\hat{R}}(\hat{L}(f,t), \hat{R}(f,t) | \theta) = \sum_i \psi_i \cdot P_{\hat{L},\hat{R},i} \quad \dots(12)$$

ψ_i ：GMM 權重

1.8 2.5 聲道差與左右聲道結合演算法

經過前兩節所討論的，我們知道 MESSL 中聲道差平面並非獨立假設那麼簡單，以及 MESSL-SP 中左右聲道平面的獨立性假設錯誤問題，因此我們先修正這兩平面的獨立性假設，接著需要找出適合的方式來結合兩個平面，而不是用簡單的獨立假設結合他們彼此。

在此之前我們先觀察這兩平面上的資料分布情況，在 3.2 章節中，我們可以很清楚的看到雙聲道語音分離的兩個困難，第一個是在聲道差平面 IPD 的 phase ambiguous 問題；第二個是左右聲道平面的難分性問題。

在得知難處後我嘗試兩種 Case，Case.1 採用聲道差與左右聲道非獨立的方式結合做分離，但此做法依然有問題在，而模擬的成效相當的差(於 3.3.3 節)。

Case.2 為假設聲道差與左右聲道兩個平面獨立，並在左右聲道的估計方式中加入硬式的預先分群機制解決難分性問題，最後在建立演算法價購後於 3.3.3 節中模擬，整體的成效比原 MESSL-SP 好上 2~3dB 不等。

1.8.1 2.5.1 非獨立結合

我們機於”嘗試”出發介紹第一種 case 的演算法，並於 3.5 節中模擬，當然模擬的成效結果是很糟糕的，可以於 3.5 節中看到。

1. 想法：

從 ILD 與左右聲道強度 \hat{L} 、 \hat{R} 的關係式中，我們可以清楚知道他們間的關聯性不低，是可以不需要用獨立假設強制簡化的；而 ILD 與 IPD 又有一定量的相關，因此我假設 IPD 與 \hat{L} 、 \hat{R} 之間會有些關連在。最後我以 $[ILD, IPD, \hat{L}, \hat{R}]$ 這四種類型的資料彼此非獨立假設建立第一種 case 的演算法(關係如圖.24)。

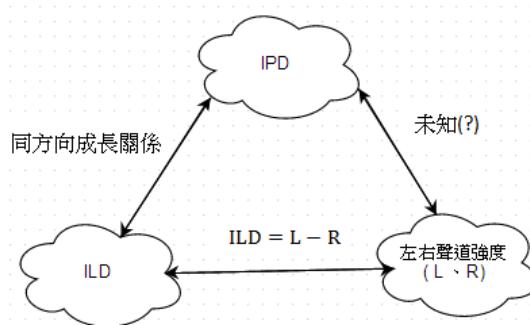


圖. 24 ILD, IPD, \hat{L} 、 \hat{R} 彼此關係圖

2. 實際資料驗證相關性

在實際的資料上我們看 \hat{L} 、 \hat{R} 與 ILD、IPD 間的相關係數。

(1) ILD 與 \hat{L} 、 \hat{R} 相關係數部分：

看到 ILD 與 \hat{L} 的相關係數圖(圖.25)可以看到，跟理論上雷同 ILD 與 \hat{L} 成正向關；ILD 與 \hat{R} 的相關係數圖(圖.26)可以看到，跟理論上雷同 ILD 與 $-\hat{R}$ 成正向關，這些敘述在 4kHz 以上的高頻段較明顯。

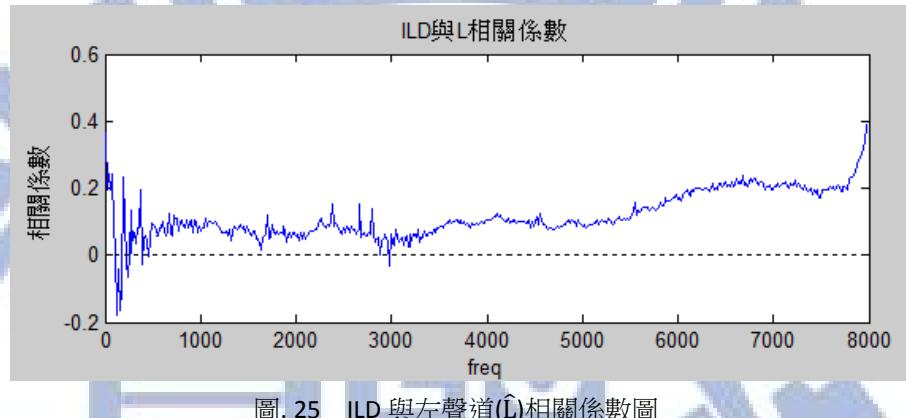


圖. 25 ILD 與左聲道(\hat{L})相關係數圖

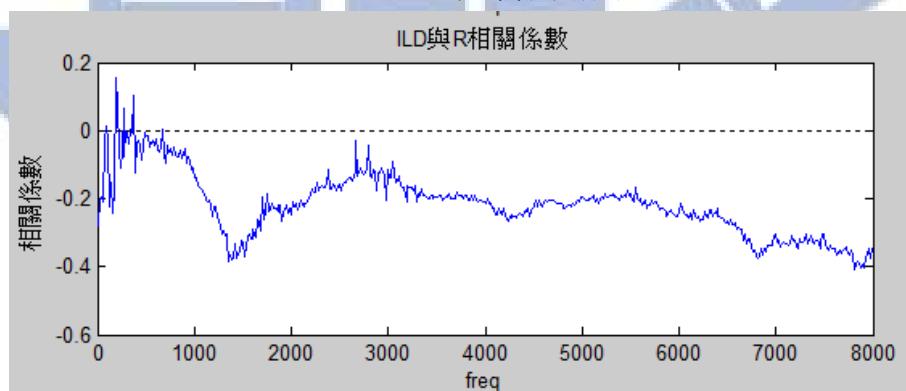


圖. 26 ILD 與右聲道(\hat{R})相關係數圖

(2) IPD 與 \hat{L} 及 \hat{R} 相關係數部分：

在 IPD 與 \hat{L} 相關係數圖(圖.27)中，我們看到 IPD 和 \hat{L} 之間的相關性在 0~7500Hz 都是小於 0.2 的，雖然在 7500~8000 頻段有升高，但依然小於 0.5，且高頻的資料空白處較多較不可信；至於 IPD 與 \hat{R} 的方面也差不多，此處就不多說明。總之 IPD 與 \hat{L} 及 \hat{R} 的關係是獨立的，若我繼續以非獨立假設套用於此，可能會有問題出現，但基於嘗試的階段我們硬做。(當然結果是不好的)

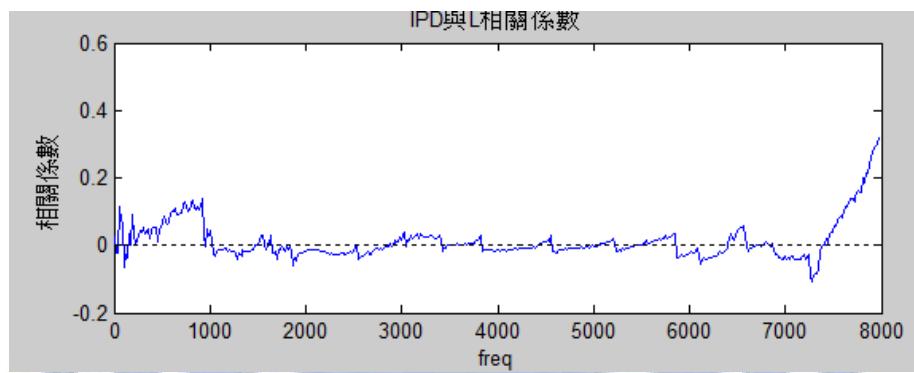


圖. 27 IPD 與左聲道(\hat{L})相關係數圖

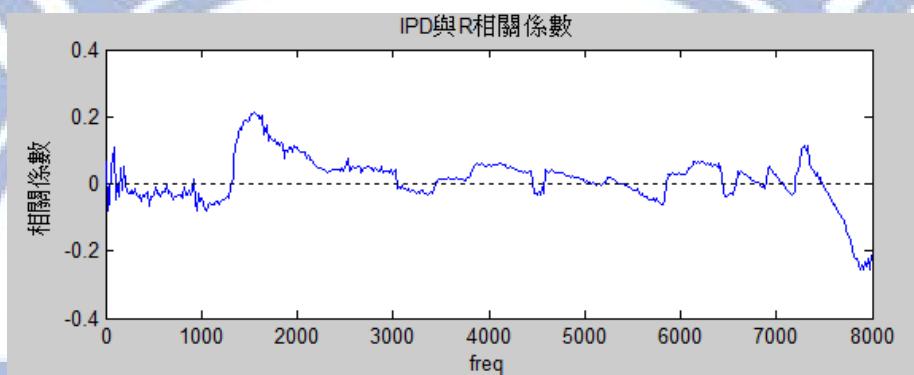


圖. 28 IPD 與右聲道(\hat{R})相關係數對頻率做圖

3. 非獨立結合演算法架構

從圖.29 中可以看到，我用全部類型的資料在四維的 EM 中解 MLE 問題。

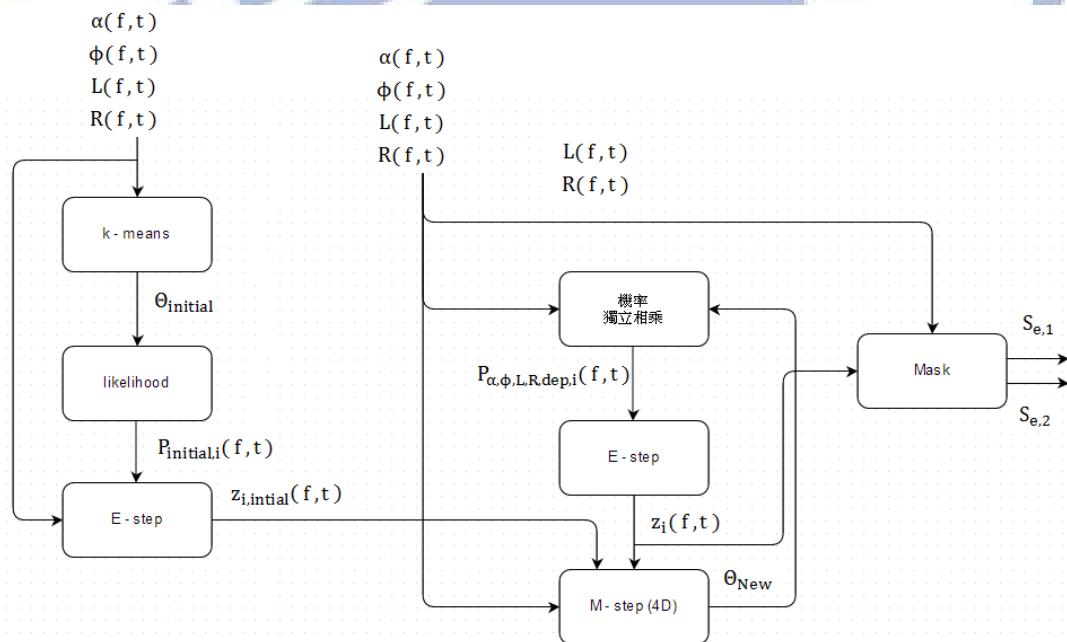


圖. 29 非獨立結合演算法流程圖

演算法中會運用到的式子會標於下方。

非獨立結合演算法：

(1) EM 解 MLE 所需 likelihood 的式子：

$$P_{\alpha, \phi, L, R, \text{dep}, i} = \psi_i \cdot P_{\alpha, \phi, \hat{L}, \hat{R}}(\alpha, \phi, \hat{L}, \hat{R} | i, \theta) \quad \dots \dots \dots (13.1)$$

(2) Mask：

左聲道分離：

$$S_{e,i,L} = L(f, t) \cdot z_i(f, t) \quad \dots \dots \dots (14.1)$$

右聲道分離：

$$S_{e,i,R} = R(f, t) \cdot z_i(f, t) \quad \dots \dots \dots (14.2)$$

1.8.2 2.5.2 獨立結合(預分群)

在經歷 case.1(非獨立結合聲道差與左右聲道)後，我們找到新的問題並做出修正，如下：

- (1) 左右聲道平面不同聲源資料太相近所導致的難分性未解決
- (2) IPD 資料與左右聲道平面間的關係應為獨立

修正的方式為拆開兩平面假設彼此獨立(解決(2))，並利用 ILD 與 \hat{L} 、 \hat{R} 的關係式加入硬式的資料預分群器，用於幫助左右聲道平面的機率估計(解決(1))，也連結兩平面。(修正後關係如圖.30)

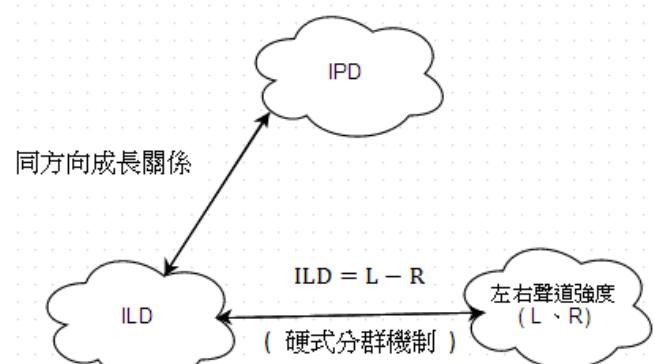


圖. 30 ILD, IPD, L, R彼此關係圖(修正後)

獨立結合(預分群)演算法架構：

圖.31 圖中與前面不同的地方為 L 、 R 資料會先經過“預先分群”產出兩群 $(\tilde{L}, \tilde{R})_1$ 及 $(\tilde{L}, \tilde{R})_2$ 在進入 EM 做機率模型的估計，而這個預分群器的指標值是不斷的再做更新的。

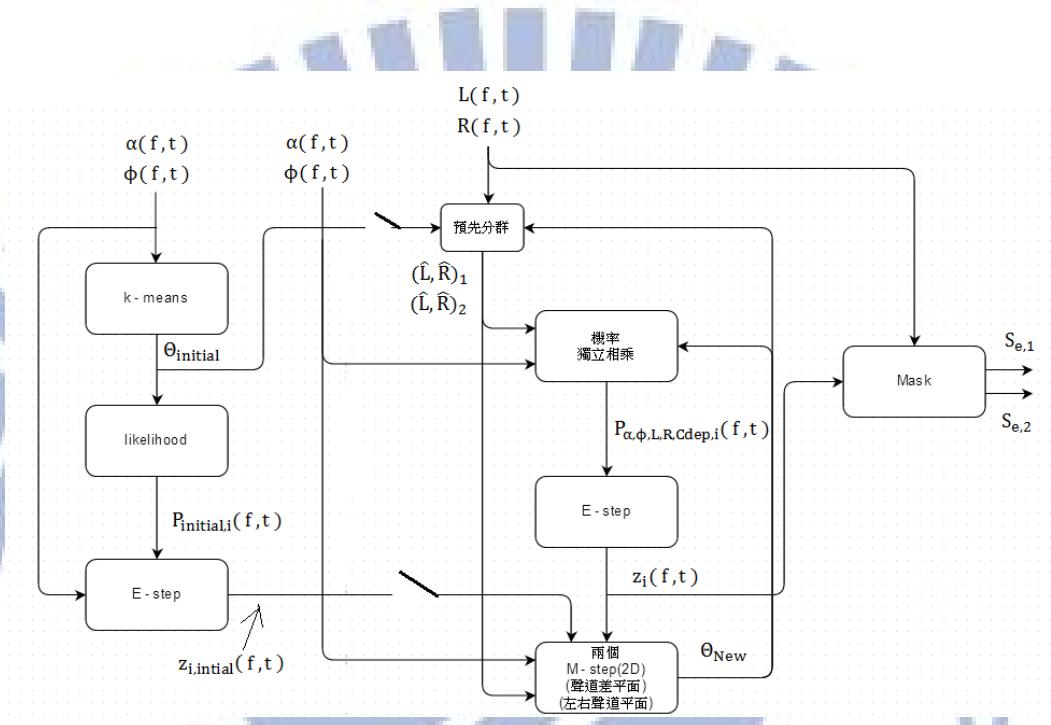


圖. 31 獨立結合(預分群)流程圖

演算法中會運用到的核心式子，標於下方塊

獨立結合(預分群)演算法：

(1)EM 解 MLE 所需 likelihood 的式子：

$$P_{\alpha, \phi, L, R, Cdep, i} = \psi_i \cdot P_{\alpha, \phi}(\alpha, \phi | i, \Theta) \cdot P_{\tilde{L}, \tilde{R} | \mu_{\alpha, i}}(\tilde{L}, \tilde{R} | i, \Theta) \quad \dots \dots \dots (15.1)$$

(2)預分群機制：

$$(\tilde{L}, \tilde{R})_i = \{ (\tilde{L}, \tilde{R}) \mid |\tilde{L} - \tilde{R} - \mu_{\alpha, i}| < |\tilde{L} - \tilde{R} - \mu_{\alpha, j}|, \text{for } j \neq i \} \quad \dots \dots \dots (15.2)$$

(3)Mask :

同(14.1)、(14.2)式

1.9 2.6 結合高斯模型參數估計

在我們利用聲倒差語左右聲道等資料加上 WDO 假設的情況下要做分離的時候，這個就是一個典型的資料分類問題，且在我們機率模型中，變成 maximum likelihood estimates (MLE)的問題，在此我們引進 EM algorithm 來群找這組 Θ 。

MLE :

$$\Theta_{\text{New}} = \arg \max_{\Theta_{\text{New}}} (P_{\text{total},i} | \Theta_{\text{old}}) \quad \dots(16)$$

$$\text{參數群 } \Theta = \{\psi_i, \mu_{\alpha\phi,i}, \Sigma_{\alpha\phi,i}, \mu_{LR,i}, \Sigma_{LR,i}\}$$

1.9.1 2.6.1 E-step

在 E-step 中，我們利用給定的 Θ_{initial} 算出各 T-F frame 的 Total likelihood 以及權重係數 z_i ； z_i 的意義是它在某個 T-F frame 上它歸屬 source i 的機率(權重)。

Total likelihood

$$P_{\text{total},i}(\alpha, \phi, \hat{L}, \hat{R} | \Theta)$$

權重係數

$$z_i(f, t) = \frac{P_{\text{total},i}(\alpha, \phi, \hat{L}, \hat{R} | \Theta)}{\sum_i P_{\text{total},i}(\alpha, \phi, \hat{L}, \hat{R} | \Theta))} \quad \dots(17)$$

1.9.2 2.6.2 M-step

在 M-step 中我們用以目前給定的 Θ_{old} 以及 E-step 算 Maximum log-likelihood 式子，以更新參數群 Θ_{new} ，再回到 E-step，成一個迴圈。

$$Q(\Theta_{\text{new}} | \Theta_{\text{old}}) = \sum_{f,t} z_i(f, t) \cdot \log (P_{\text{total},i}(\alpha, \phi, i, \hat{L}, \hat{R} | \Theta_{\text{old}})) \quad \dots(18)$$

參數更新公式如下方塊：於每個 source i 在每個頻段 f 中算出 (19.1)式 GMM 權重 ψ_i 為所有資料權重 z_i 所佔的比例；(19.2)式 $\mu_{\alpha\phi,i}$ 為所有資料經資料權重後取平均；(19.3)式 $\Sigma_{\alpha\phi,i}^2$ 為資料經資料權重後取變異數。

聲道差平面更新公式

GMM 權重：

$$\Psi_i(f) = \frac{1}{T} \cdot \sum_{t=1}^T z_i(f, t) \quad \dots(19.1)$$

Mean update :

$$\mu_{\alpha\phi,i}(f) = \frac{\sum_t z_i(f, t) \cdot (\alpha(f, t), \phi(f, t))}{\sum_t z_i(f, t)} \quad \dots(19.2)$$

Covariance update :

$$\Sigma_{\alpha\phi,i}^2(f) = \frac{\sum_t z_i(f, t) \cdot [(\alpha(f, t), \phi(f, t)) - \mu_{\alpha\phi,i}(f)]^2}{\sum_t z_i(f, t)} \quad \dots(19.3)$$

這邊列出原本 MESSL 中對兩聲道差平面的公式予以比較，可以看到原 MESSL 是因獨立簡化的假設下而將原本只有兩項的平均值和共變數矩陣拆開成 4 個 (ILD 與 IPD 各自的平均值與共變數)。

MESSL 參數群更新公式 (引用自[5]) :

GMM 權重：

$$\Psi_i(f) = \frac{1}{T} \cdot \sum_t z_i(f, t) \quad \dots(20.1)$$

ILD Mean :

$$\mu_{\alpha,i}(f) = \frac{\sum_t z_i(f, t) \cdot \alpha(f, t)}{\sum_t z_i(f, t)} \quad \dots(20.2)$$

ILD Variance :

$$\sigma_{\alpha,i}^2(f) = \frac{\sum_t z_i(f, t) \cdot (\alpha(f, t) - \mu_{\alpha,i}(f))^2}{\sum_t z_i(f, t)} \quad \dots(20.3)$$

IPD Mean :

$$\mu_{\phi,i}(f) = \frac{\sum_t z_i(f, t) \cdot \phi(f, t)}{\sum_t z_i(f, t)} \quad \dots(20.4)$$

IPD Variance :

$$\sigma_{\phi,i}^2(f) = \frac{\sum_t z_i(f, t) \cdot (\phi(f, t) - \mu_{\phi,i}(f))^2}{\sum_t z_i(f, t)} \quad \dots(20.5)$$

對於左右聲道平面的更新方式與聲道差平面雷同，所以不再多說明。

左右聲道平面更新公式

Mean update :

$$\mu_{\widehat{L}\widehat{R},i}(f) = \frac{\sum_t z_i(f, t) \cdot (\widehat{L}(f, t), \widehat{R}(f, t))}{\sum_t z_i(f, t)} \quad \dots(21.1)$$

Covariance update :

$$\Sigma_{\widehat{L}\widehat{R},i}^2(f) = \frac{\sum_t z_i(f, t) \cdot [(\widehat{L}(f, t), \widehat{R}(f, t)) - \mu_{\alpha\phi,i}(f)]^2}{\sum_t z_i(f, t)} \quad \dots(21.2)$$

此處一樣附上 MESSL-SP 對左右聲道平面的模型以及參數更新公式與以比較。(21)式與(25)相比，可以看到在我們方法中是直接對 \widehat{L} 、 \widehat{R} 做估計，而 MESSL-SP 是對通道 H_i^l 、 H_i^r 做估計，不一樣的原因在於我們對左右聲道平面的模型假設不同，看到 MESSL-SP 對左右聲道的基本模型是以聲源音量 \widehat{S}_i (變數)以及通道 \widehat{H}_i^l (定數)組成 \widehat{L} (變數)，於(22)式中看到，它採用由數據庫所給定的聲源機率(23)式簡化對 \widehat{L} 機率模型的估計難度，變為只需估計通道 \widehat{H}_i^l 即可，而 \widehat{H}_i^l 的更新公式於下方塊中 (25)式。

MESSL-SP 左右聲道更新公式(引自[6])

左右聲道模型 :

$$\widehat{L}(t, f) \approx \widehat{S}_i(t, f) + \widehat{H}_i^l(f) , \quad \text{if source } i \text{ dominate} \quad \dots(22.1)$$

$$\widehat{R}(t, f) \approx \widehat{S}_i(t, f) + \widehat{H}_i^r(f) , \quad \text{if source } i \text{ dominate} \quad \dots(22.2)$$

給定聲源機率(從數據庫) :

$$P(\widehat{S}_i(\cdot, t)) = N\left(\mu_{sp,i}(f), \sigma_{sp,i}^2(f)\right) \quad \dots(23)$$

機率模型 :

$$\widehat{L}_{i,f} \sim N(\mu_{sp,i}(f) + \widehat{H}_i^l(f), \sigma_{sp,i}^2(f)) \quad \dots(24.1)$$

$$\widehat{R}_{i,f} \sim N(\mu_{sp,i}(f) + \widehat{H}_i^r(f), \sigma_{sp,i}^2(f)) \quad \dots(24.2)$$

參數更新公式 :

$$H_i^l = \frac{\sum_t z_i(f, t) \cdot [\widehat{L}(f, t) - \mu_{sp,i}(f)]}{\sum_t z_i(f, t)} \quad \dots(25.1)$$

$$H_i^r = \frac{\sum_t z_i(f, t) \cdot [\widehat{R}(f, t) - \mu_{sp,i}(f)]}{\sum_t z_i(f, t)} \quad \dots(25.2)$$

1.10 2.7 分離訊號回授

在現今的演算法中，截取出 ILD、ITD、IPD 等資料後，套用 GMM 模型並以 EM 演算法來解 ML 的問題是一貫的作風，但在使用 EM 的過程中都是不斷的訓練之後得出一組最佳參數群 Θ 後，算出權重係數 z_i ，在分離出個別的聲源；由於回圈的連貫性，如果我們找出更正確的參數群 Θ 那麼我們有更正確的權重係數 z 用以分離訊號，因此，我打破 EM 演算法中訓練資料不變的概念，將每次回圈後預先分離出 $S_{e,i}$ (也就是 L_i 、 R_i)並取代 EM 演算法中所需要 L 、 R ，尋求更正確的參數群 Θ (架構圖在圖.7)，Feed Back 的概念可以分成三小塊做，分別為 FB-ILD (Feed Back ILD)、FB-IPD (Feed Back IPD)、FB-SP (Feed Back Source prior) 即 L 、 R 三種，在下頁列出演算法架構後於 4.5 節中模擬，模擬結果為 FB-ILD 使整體成效上升 1dB 不等；FB-IPD 則些許上升 0.2dB；FB-LR 部分反而是讓整體成效下降。

Feed Back 演算架構：

圖.32 圖中可以看到，由 Mask 輸出端拉了一條回授到 M-step 中取代原本的混合資料試圖幫助估計正確的參數群。而回授的方式是以開關的方式可選擇的。

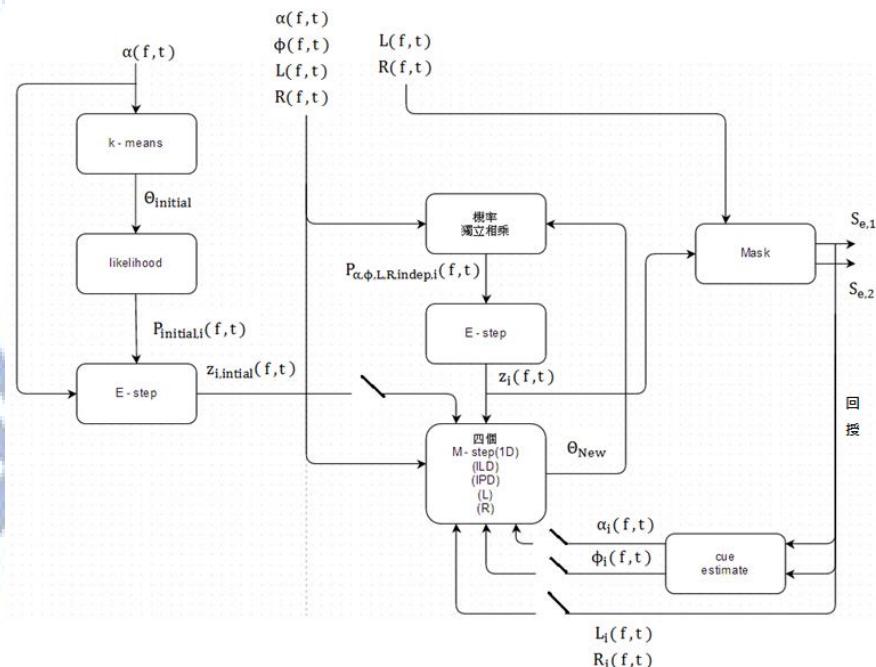


圖. 32 引進回授概念的 MESSL-SP 流程圖

擷取公式在聲道差平面將(4.1)、(4.2)式以(26.1)及(26.2)式取代；左右聲道平面將(5.1)、(5.2)以(27.1)及(27.2)式取代，資料取代公式如下標於下方塊中。

各別聲源回授取代公式：

(1) 聲道差 Feed Back :

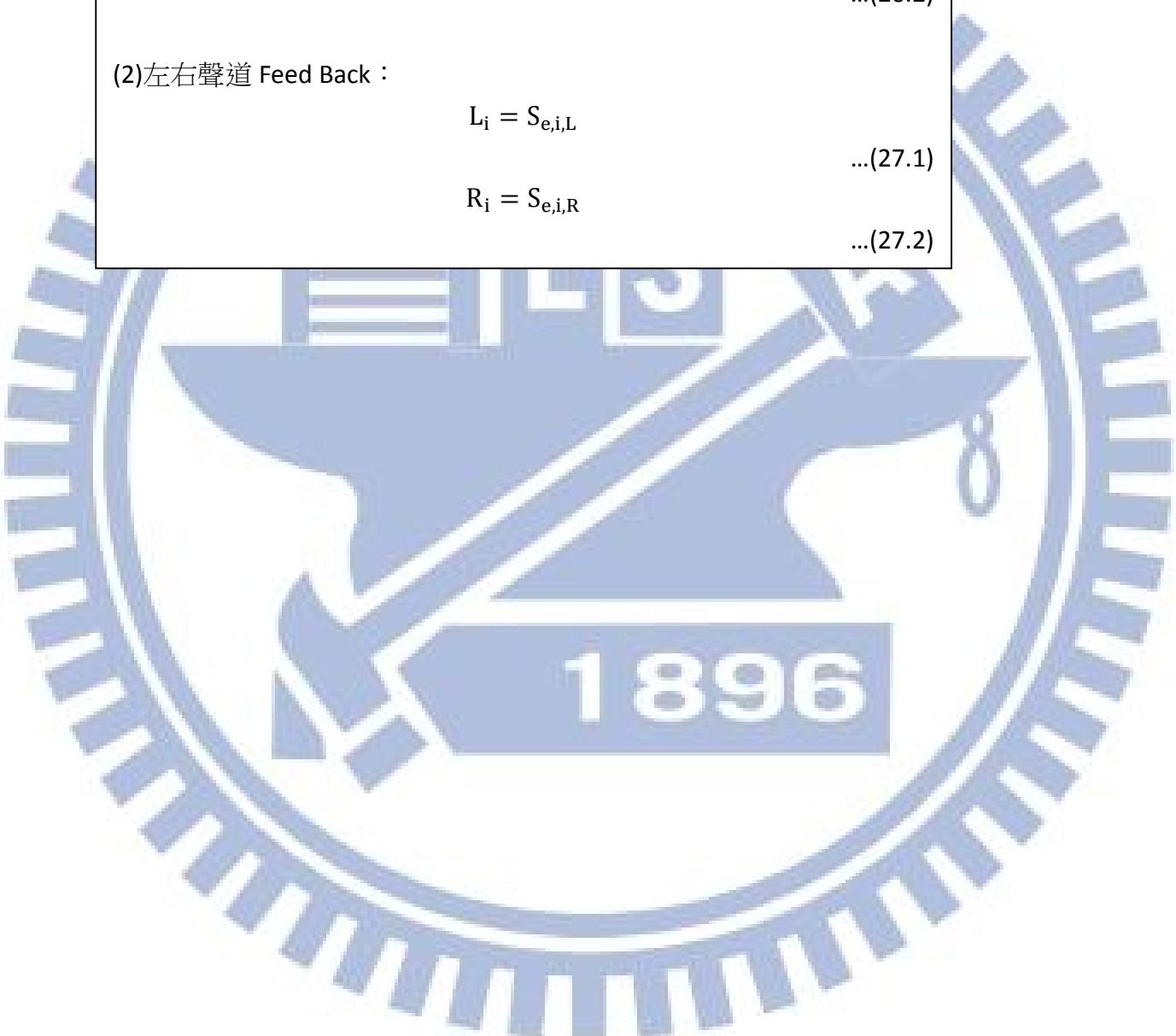
$$\alpha_i(f, t) = 20 \cdot \log_{10} \left(\frac{L_i}{R_i} \right) \quad \dots (26.1)$$

$$\phi_i(f, t) = \text{angle} \left(\frac{L_i}{R_i} \right) \quad \dots (26.2)$$

(2) 左右聲道 Feed Back :

$$L_i = S_{e,i,L} \quad \dots (27.1)$$

$$R_i = S_{e,i,R} \quad \dots (27.2)$$



第三章 電腦模擬

前幾節中了解目前雙聲道語音分離演算法以及我們所改善的演算法後，我們接著對這些演算法做 Performance 的測試，我們先制定測試標準，以及測試環境，於 3.1 節；接著在 3.2 節中觀察資料的分布情況並看出語音分離的難處。3.3~3.6 節中式開始一連串的演算法測試，3.7 節則是將聲源數增加至三個，最後 3.8 節中是我所做的實測。

1.11 3.1 模擬環境設定

測試標準

在測試的標準方面，我們採用文獻[12]所提到的 SDR (Signal Distortion Ratio)

$$s_e = s_{target} + error \quad \dots(28.1)$$

$$s_{target} = \frac{\langle s_e, s \rangle \cdot s}{\|s\|^2} \quad \dots(28.2)$$

我們先把 s_e 拆解成 s_{target} 及 $error$ 兩項，於(28.1)式，其中 s_{target} 為估計聲音 s_e 投影到正確聲音 s 上的向量(signal part)，於(28.2)式； $error$ 則為其餘部分 (Distortion part)，接著取能量比值 SDR，定義如下(28.3)式。

$$SDR = 10 \cdot \log_{10} \left(\frac{\|s_{target}\|^2}{\|error\|^2} \right) \quad \dots(28.3)$$

模擬環境.1 簡單測試

雙聲源：固定一組男女的聲音擺放在角度=[0 , 30]，其中 0 度角為主要目標。

三聲源：固定一組三位男性聲音在[-30 , 0 , 30]的擺放，其中 0 度角為主要目標。

模擬環境.2 角度與 database 測試

雙聲源：

將主要訊號放置在 0 度角的地方(正前方)，干擾訊號分別放置在 [10 20 30 40 50 60 70 80] 等角度，分別分離出的主訊號 SDR。模擬大小為從 50 個聲源大小的 TIMIT database 中(平均為 2~3 秒的聲源)，隨機選擇 2 聲源為一組做測試，且在每個角度擺放下各做 10 次後平均。

三聲源：

將主要訊號放置在 0 度角的地方(正前方)，兩個干擾訊號分別放置在 $\pm [10 20 30 40 50 60 70 80]$ 等角度，分別分離出的主訊號 SDR。

1.12 3.2 左右聲道及聲道差資料分布情況

在這章節中，我們明確的展示出雙聲源及三聲源混合前後聲道差及左右聲道的資料分布，而雙聲道語音分離的難點在這部分可以看到。

在 3.2.1 節 ILD 的部分，可以看得出可分性的跡象；在 3.2.2 節 IPD 部分雖然有可分性的跡象，但會遇到 phase ambiguous，而這個現象我們放在 3.2.3 章節中做觀察，最後在 3.2.4 節中我們觀測左右聲道的資料分布，看到了他的不可分性的問題。

1.12.1 3.2.1 左右聲道差(ILD 部分)

從圖.33 可看到雙聲源下混合前他們資料的區別性是很大的；而在圖.34 混合後灰階圖，可以看到兩條曲線的中間有資料參雜在一起的情況出現。

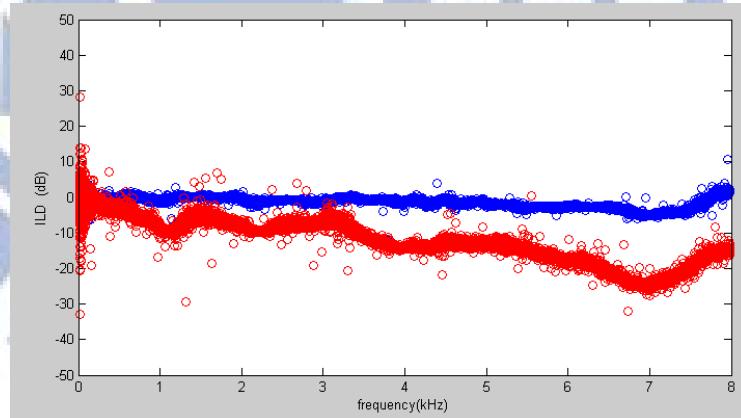


圖. 33 雙聲源混合前 ILD 資料分布

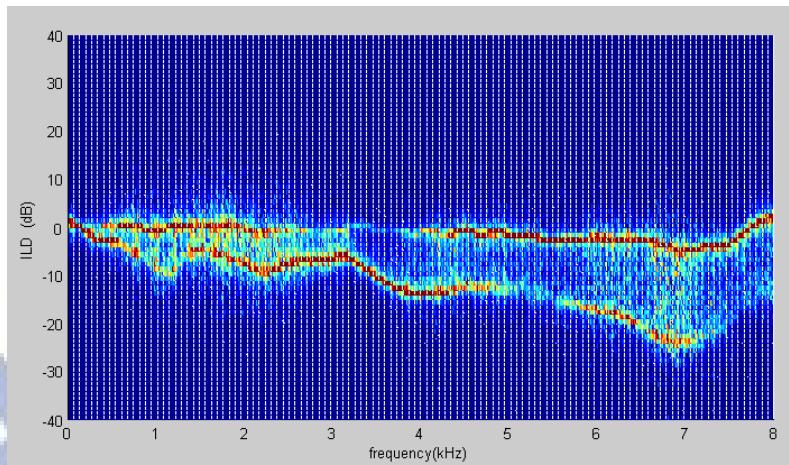


圖. 34 雙聲源混合後 ILD 資料灰階圖

從圖.35 可看到三聲源下混合前他們資料的區別性相較於雙聲源小了些，不過還是可以看出區別性；而在圖.36 混合後，可以看到曲線間參雜在一起的情況較為嚴重，例如在 3kHz 的地方；且由於聲元數的增多，彼此的干擾性也變大，例如在 3.5kHz~4kHz 的地方中下兩線消失。

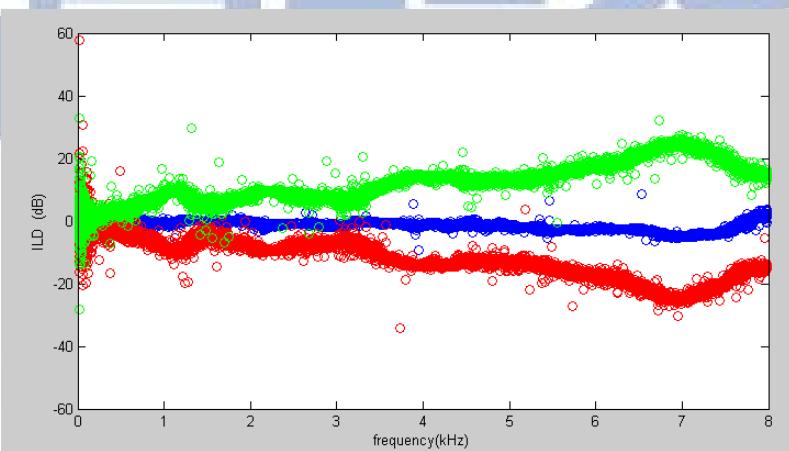


圖. 35 三聲源混合前 ILD 資料分布

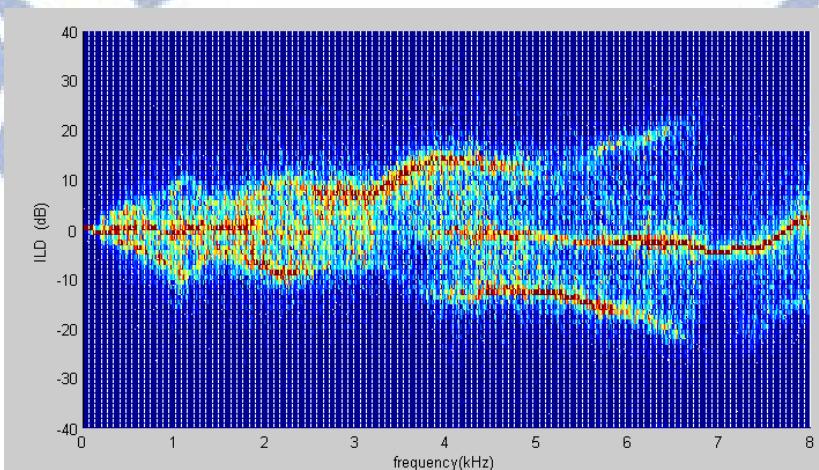


圖. 36 三聲源混合後 ILD 資料灰階圖

1.12.2 3.2.2 左右聲道差(IPD 部分)

圖.37 雙聲源中兩組資料群的分布情況足以區別，這些可以看到斜率也許是個不錯的分離依據(本論文沒討論)；而在圖.38 混合過後資料混散開，有參雜的情況；再者，由於我們相位的定義為 $[-\pi, \pi]$ ，在 *phase ambiguous* 發生的頻段(Ex：1800Hz...)，會讓 IPD 資料會有難以分離的困擾，此現象我們會在 3.2.3 節中做觀察。

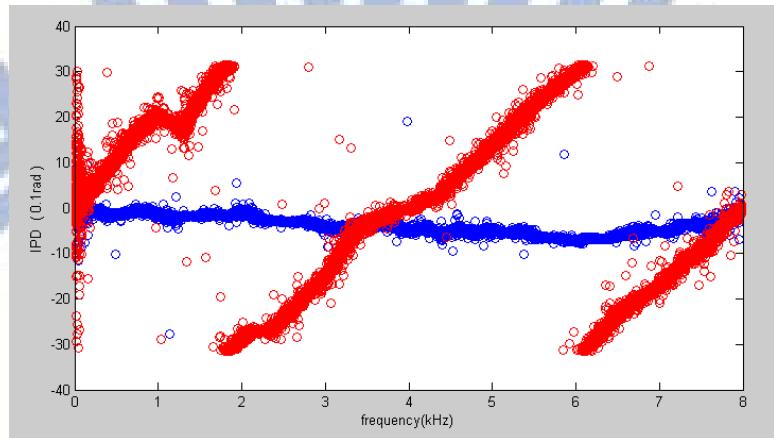


圖. 37 雙聲源混合前 IPD 資料分布

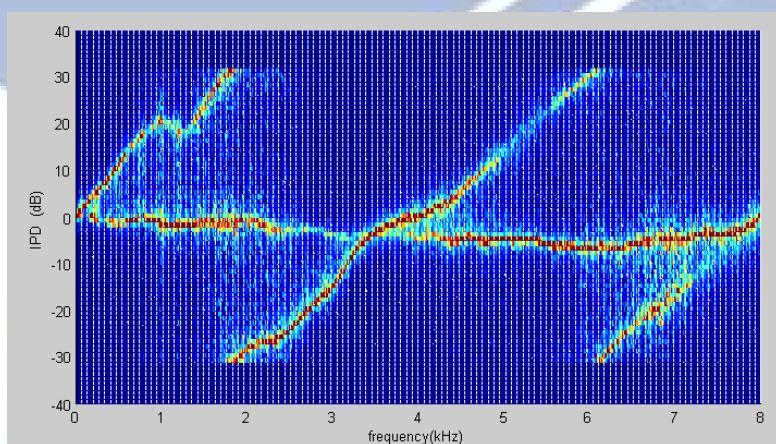


圖. 38 雙聲源混合後 IPD 資料灰階圖

圖.39 三聲源中兩組資料群的分布情況也看的到區別性，；而在圖.40 中上下兩線於 1.5~2kHz 消失，且 *phase ambiguous* 的發生次數也較多。

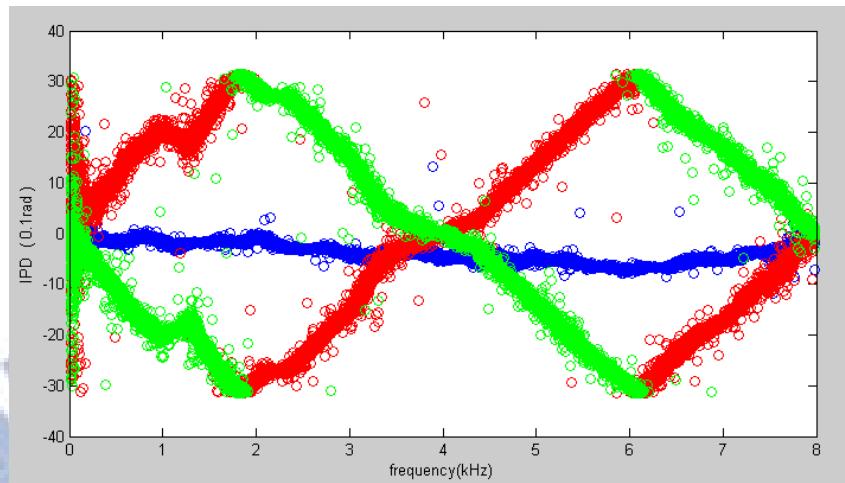


圖. 39 三聲源混合前 IPD 資料分布

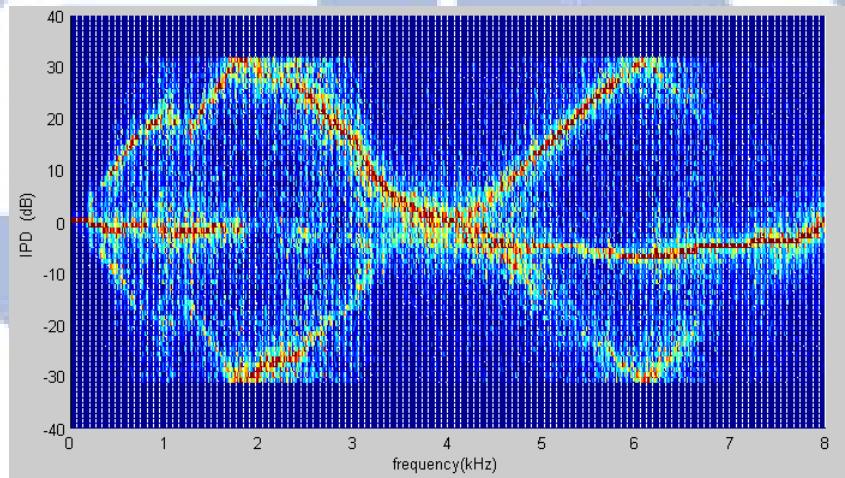


圖. 40 三聲源混合後 IPD 資料灰階圖

1.12.3 3.2.3 Phase ambiguous 問題

在圖.41 混合前圖中可以看到，正常情況來自不同聲源的資料是各別群聚的；而圖.42 混合後圖中資料是從各個聲源(黑色 X)的中心散開，而我們分離好壞的依據，就是看演算法是否能正確的劃分出圖.42 圖中屬於不同聲源的區塊。

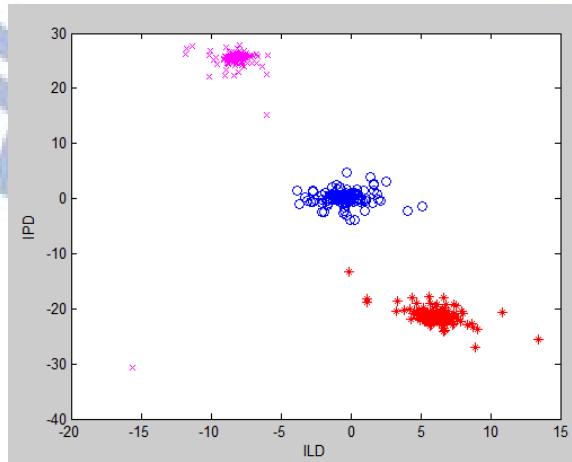


圖. 41 正常情況下混合前 (α, ϕ) 平面資料分布

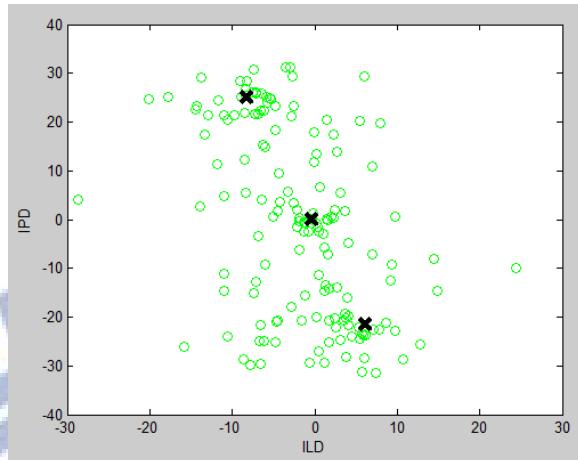


圖. 42 正常情況下混合後 (α, ϕ) 平面資料分布

反觀在圖.43 中可以看到，phase *ambiguous* 發生後 x 的群組不再是群聚在一群，而是分為上下兩群，這個會對 GMM 的估計帶來困擾，例如圖.44 圖中就看到 EM 初步估計出的平均值位置，左邊群組的中心點明顯偏離。

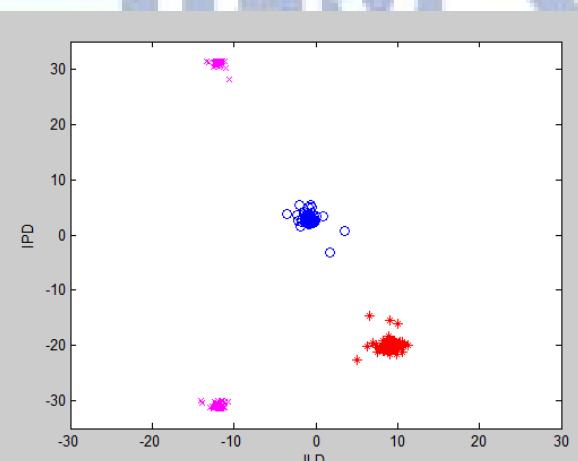


圖. 43 phase *ambiguous* 下混合前 (α, ϕ) 平面資料分布

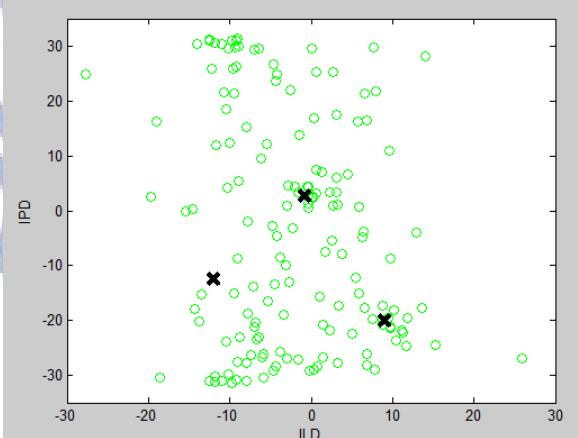


圖. 44 phase *ambiguous* 下混合後 (α, ϕ) 平面資料分布

1.12.4 3.2.4 左右聲道強度

在左右聲道的部分我們只採用它的強度資料。強度的資料分布，從圖.45 混和前可以看到，各聲援左右聲道強度的資料都沿著斜直線分布。

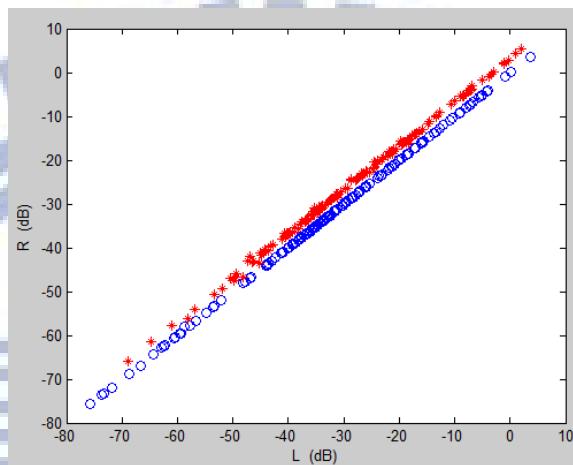


圖. 45 雙聲源混合前 (L, R) 平面資料分布(500 Hz)

接著看到圖.46 混合(L, R) 資料分布圖，發現四點：

1. 理論上(\hat{L} 、 \hat{R}) 資料的趨勢是斜值線，那是因為關係式 $ILD = \hat{L} - \hat{R}$ 之因素(理想中各方位的 ILD 為定值)，這在圖.44 中已經解釋。
2. 資料是沿著直線兩側量開的原因是受到彼此的干擾因素。
3. 在圖中右上方量開的程度較小，是因為能量大受干擾的程度相對較小因素。
4. \hat{L} 及 \hat{R} 的分布程度範圍為[-60,30]，相較之下 ILD 的分布範圍為[-10,10]較小，導致兩筆資料再圖中非常相近，難以分離。(問題)

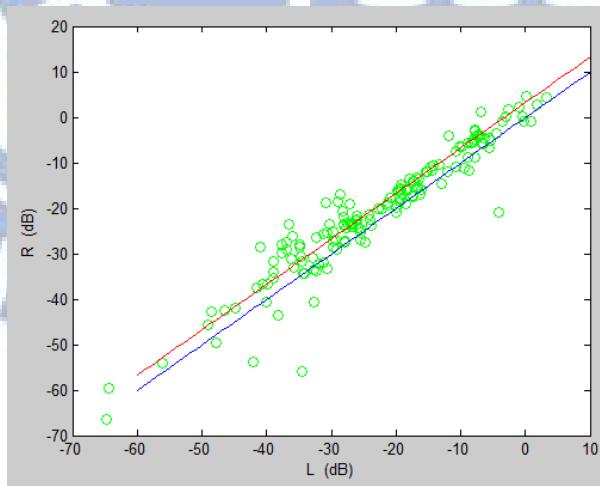


圖. 46 雙聲源混合後 (L, R) 平面資料分布(500 Hz)，兩條直線為理想分布曲線

1.13 3.3 以往演算法效能比較

在節我們先用 3.1 節提到的**模擬環境.2** 比較歷年的 5 種演算法效能。直接看到成果圖(圖.47 及圖.48)，在左聲道平均上 DUET 表現的最差只有 2.37dB；HRTF 分法有 4.8 離 MESSL 的 5.2 很相近，但由於 HRTF 的特定性，會有不少的問題；加入左右聲道資料的 MESSL-SP 有 5.56dB(比 MESSL 高一些)，而最高的是二維 GMM 分法有個平均 7.5dB。右聲道方面 DUET 為 2.37dB、HRTF 為 1.98dB、MESSL 為 1.4dB、MESSL-SP 為 2.38dB、二維 GMM 分法為 4.8dB。比較的結果錯出色的是二維 GMM 分法，值得一提的是 MESSL-SP 加入做右聲道資料後比原 MESSL 增長了 0.3dB，因此我們會已者兩個觀點出發建立我的新演算法。而從整體測試結果中可以看出一個相同趨勢的，角度越大分離的越好，這是由於角度大，各聲源間的 ILD、IPD 等資料區別性越大，越好分離之因素(會在圖.63 的解釋中證實)。

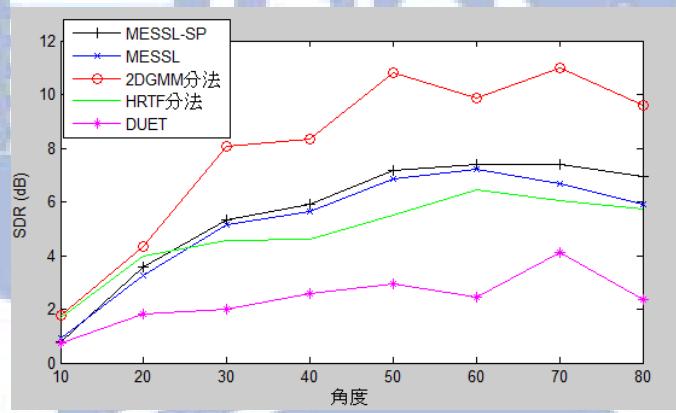


圖. 47 以往演算法左聲道成效圖

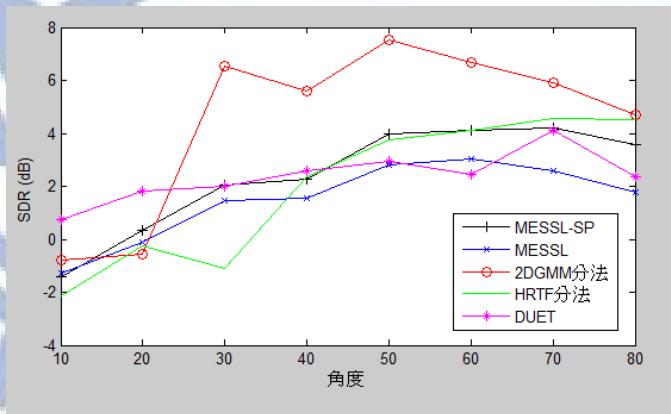


圖. 48 以往演算法右聲道成效圖

1.14 3.4 機率模型測試

在本章節中，我們先模擬 MESSL 在獨立性的比較(於 3.4.1 節)，接著模擬 MESSL-SP 中所加入的 L,R 資料的獨立性及其比較(於 3.4.2 節)，最後我們所提出的修

正演算法（於 3.3.3 節）。

1.14.1 3.4.1 聲道差平面分頻獨立性模擬

這節中我們模擬的是聲道差平面獨立性與否對整體演算法帶來的成果改變，我們直接看到成果(圖.49)，先看到 MESSL 的成果(聲道差獨立)，整體成果都是最差的；在看到分頻處理的部分(以分頻點為準，低頻採用非獨立，高頻採用獨立)，可以看到分頻點 500Hz 的比原 MESSL 好，分頻點為 1000Hz 資比前兩者都好，這是由於理論結果“低頻高相關因素”，而分頻點 1000Hz 更勝一籌是因對高相關性頻段的覆蓋率較多的因素(可與圖.12 做比對)；接著看到分頻點 1000Hz 與聲道差非獨立(即全頻獨立)，可看到兩方法彼此差不多，這裡可以看到只要對於高相關頻段的覆蓋率到一定程度，整體所帶來的改善就差不多。且平均上顯示聲道差非獨立的方式比獨立假設的方式在左聲道平均好讓 4.3dB；右聲道擇好上 3.5dB。

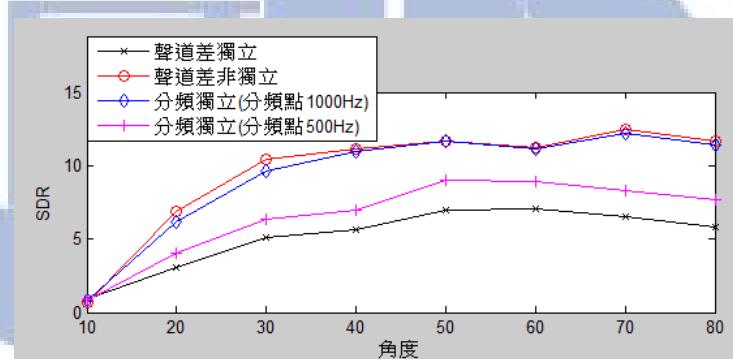


圖. 49 聲道差平面獨立性效果比對圖(左聲道)，x 線為 MESSL(全頻獨立)；o 線為全頻非獨立；+線為 1000Hz 分頻處理；diamond 為 500Hz 分頻處理

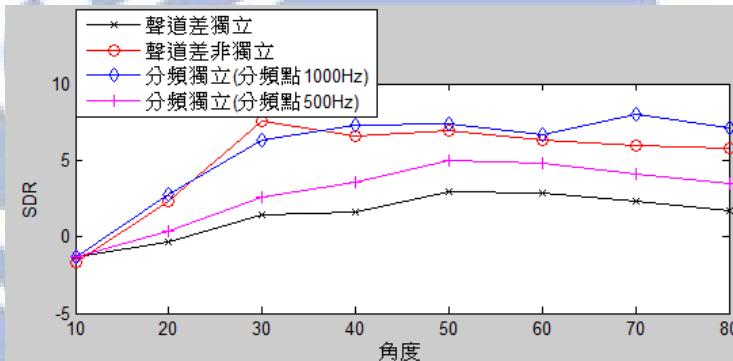


圖. 50 聲道差平面獨立性效果比對圖(右聲道)，x 線為 MESSL(全頻獨立)；o 線為全頻非獨立；+線為 1000Hz 分頻處理；diamond 為 500Hz 分頻處理

接著我們透過**模擬環境.1** 測試並觀看獨立與非獨立兩種方法所估計出的機率等高線分布圖。如下圖.51($P_{\alpha, \phi, dep}$ 機率等高線圖)所示 GMM 模型，在 Mean 方面，中間的群組抓的有點偏，右下角群組中抓得也有點偏，但大致上還算可以。而在獨立假設的簡化(圖.52)，右下角群組的 Mean 明顯錯誤，Variance 明顯偏大。從機率圖上可以明顯看出在獨立簡化的假設下是不好的。

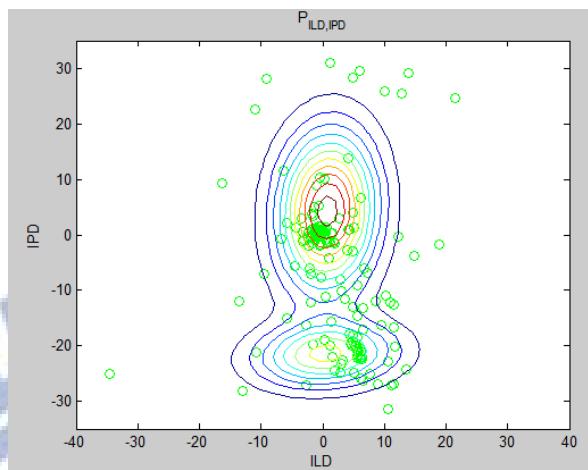


圖. 51 $P_{\alpha,\phi,dep}$ 機率等高線圖(1500Hz)，綠點為 Data

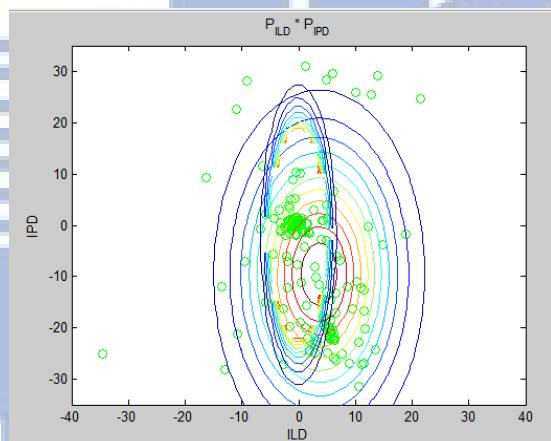


圖. 52 $P_{\alpha,\phi,indep}$ 機率等高線圖(1500Hz)，綠點為 Data

1.14.2 3.4.2 左右聲道平面獨立性模擬

我們用**模擬環境.1** 測試並觀看獨立與非獨立兩種方法(如下)所估計出的機率等高線分布圖，由於 4.2.4 節中提到的難以分離的危險，所以我們這部分直接進入機率等高線圖做比較。

非獨立情況下(圖.53)，機率估計的結果是右上角一群，左下角一群，這與預期的不同。獨立假設下估計的結果 (圖.54) 由於獨立簡化的關係(可視為將資料投影到各軸，各自估計)，估計出的結果明顯不符合資料分布，這代表獨立的假設在 $P_{L,R}$ 不適用。因此我們想嘗試用 ILD 與 L、R的關係建立 $P_{L,R}$ 機率模型。

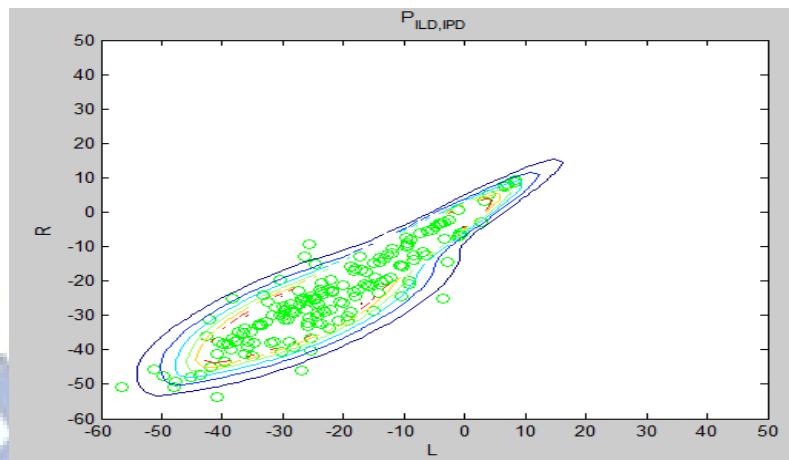


圖. 53 $P_{L,R,dep}$ 機率等高線圖 (1500Hz) , 綠點為 Data

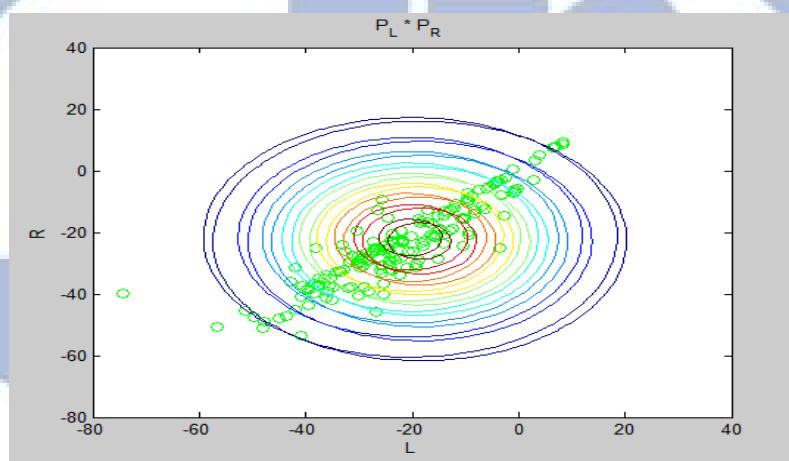


圖. 54 $P_{L,R,indep}$ 機率等高線圖 (1500Hz) , 綠點為 Data

1.15 3.5 聲道差與左右聲道結合演算法比較

這張節中是對應到 2.5 節的兩種 case 做模擬，我們直接看到成果圖.55，MESSL(修正獨立性)的方式在 3.4.1 節中已說明，而獨立結合(預分群)的方式則更上一階，在左聲道的部分比 MESSL-SP 平均好了 3.3dB，右聲道則好了 2dB 以上；反之採用左右聲道及聲道差非獨立的方式在成效上很低，這是因為平面的難分性若未解決問題。

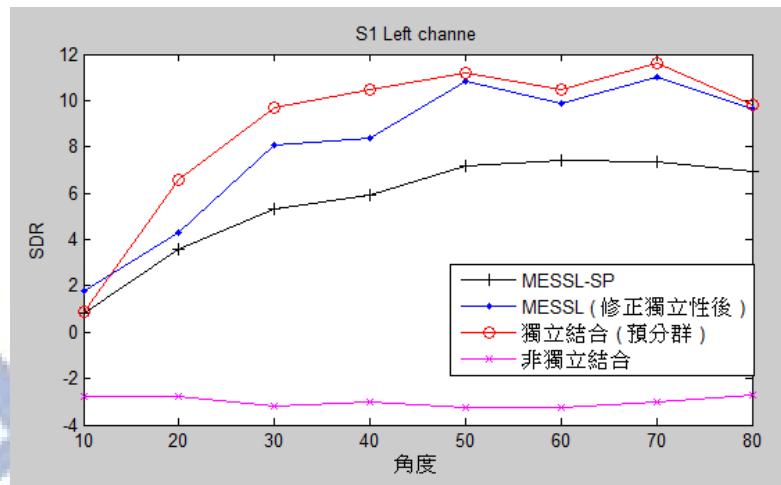


圖. 55 兩平面結合演算法成效圖(左聲道)

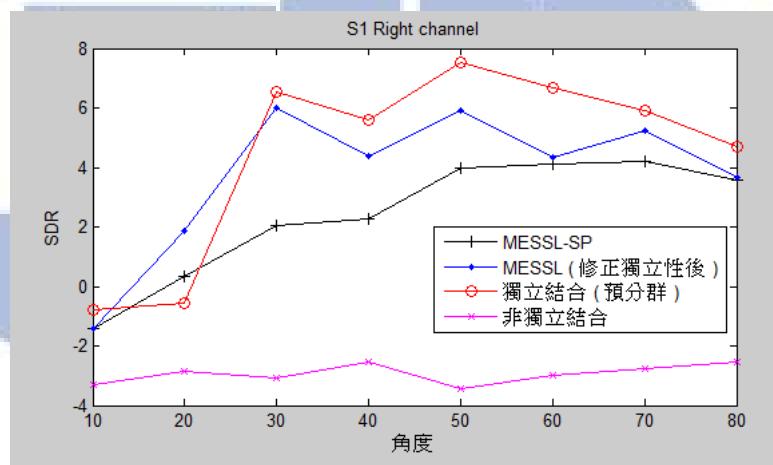


圖. 56 兩平面結合演算法成效圖(右聲道)

接著我們會以 3.1 中提到的**模擬環境.1** 來看兩平面的機率模型估計成效。

1.15.1 3.5.1 左右聲道平面機率模型

我們借由觀察左右聲道平面邊際機率的方式做討論，邊際機率如下。

右聲道平面邊際機率：

非獨立結合：

$$\tilde{P}_{L,R,dep} = \iint P_{\alpha,\phi,L,R,dep} d\alpha d\phi \quad \dots(29.1)$$

獨立結合(預分群)：

$$\tilde{P}_{L,R,Cdep} = \iint P_{\alpha,\phi,L,R,Cdep} d\alpha d\phi \quad \dots(29.2)$$

非獨立結合結果

圖.57 可以看到，機率估計結果是單峰 GMM，這但表左右聲道平面的資料難分性問題並未解決。

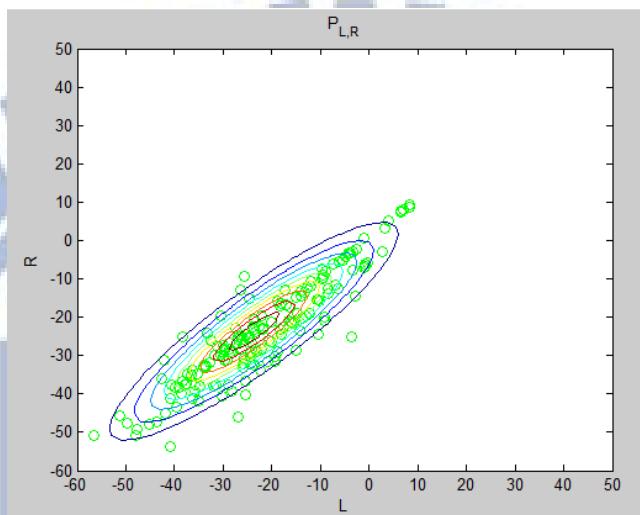


圖. 57 非獨立結合左右聲道邊際機率等高線圖

獨立結合(預分群)

相較於圖.57，圖.58 中有了群組的區別性出來，大大的改善難分性問題。

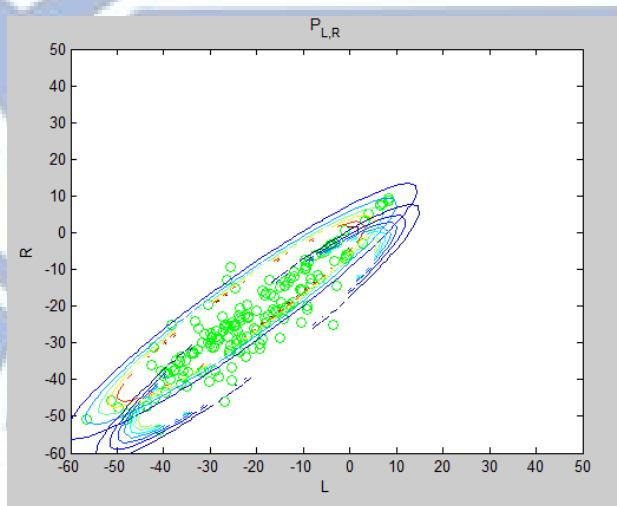


圖. 58 獨立結合(預分群)左右聲道邊際機率等高線圖

1.15.2 3.5.2 聲道差平面機率模型

我們借由觀察聲道差平面邊際機率的方式做討論，邊際機率如下。

聲道差平面邊際機率：

非獨立結合：

$$\tilde{P}_{\alpha, \phi, \text{dep}} = \iint P_{\alpha, \phi, L, R, \text{indep}} dL dR \quad \dots(32.1)$$

獨立結合(預分群)：

$$\tilde{P}_{\alpha, \phi, \text{Cdep}} = \iint P_{\alpha, \phi, L, R, \text{Cdep}} dL dR \quad \dots(32.2)$$

非獨立結合結果

從圖.59 中可以看到估計出的機率模型是單峰的 GMM，造成此原因是因為 \hat{L} 、 \hat{R} 平面資料難分性未解決而牽連之因素。

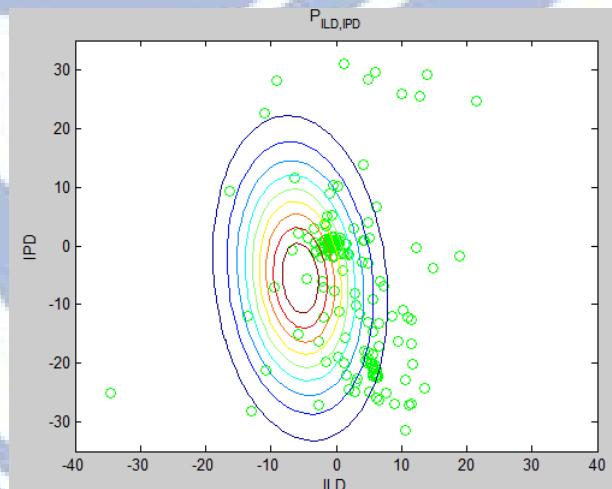


圖. 59 $\tilde{P}_{\alpha, \phi, \text{dep}}$ 機率等高線圖

獨立結合(預分群)

與圖.59 相比，圖.60 就分明了很多，因為是在兩平面獨立的情況下所做的，不會受左右聲道平面牽連。

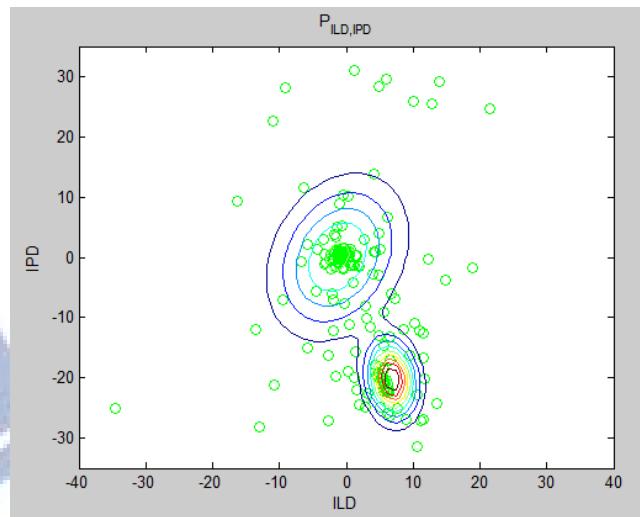


圖. 60 $\tilde{P}_{\alpha,\phi,Cdep}$ 機率等高線圖

1.16 3.6 加入回授機制後成效與估計性能

本章節為加入 Feed Back 機制後的改善結果，會在 3.6.1 節中比較各種 Feed Back 為 MESSL-SP 所帶來的效能改善；在 3.6.2 節中比較 Feed Back 為獨立結合(預分群)的方法所帶來的效能改善。

1.16.1 3.6.1 MESSL-SP 加入回授機制效能比較

看到圖.61 與圖.62 中，FB-ILD 平均比原 MESSL-SP 在左右聲道分別有 2 dB 左右的提升、1dB 左右的提升；FB-IPD 平均比原 MESSL-SP 在左右聲道都有 0.2dB 些微的提升；FB-SP case 平均則比原 MESSL-SP 在左右聲道都有 1dB 左右的下降。整體看來 FB-ILD 、FB-IPD 適合做維修正的方法，FB-SP 則否，其原因不明。

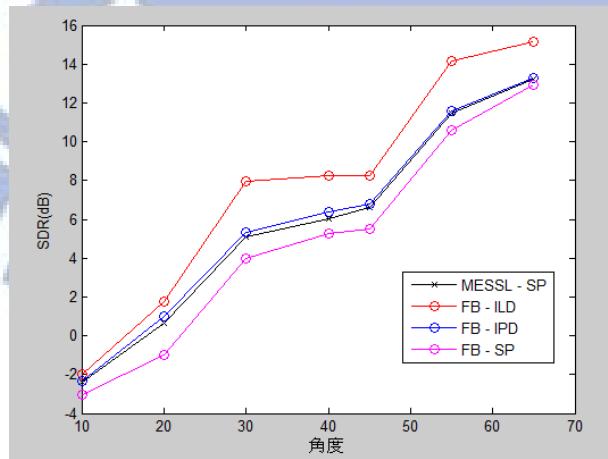


圖. 61 回授機制成效圖(左聲道)

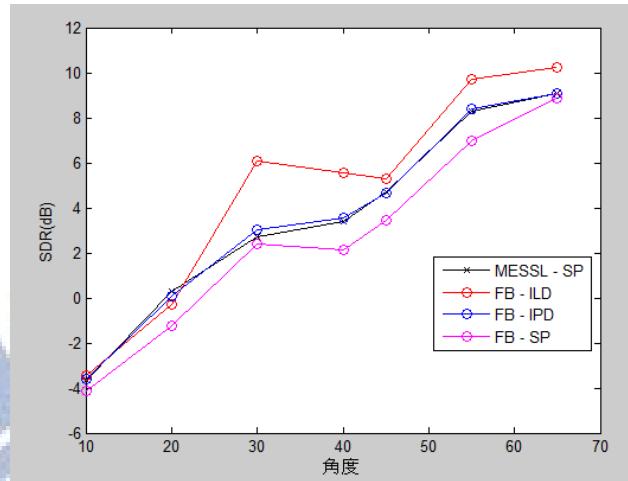


圖. 62 回授機制成效圖(右聲道)

接著我們要探討這些 Feed Back 機制到底改善了些甚麼。首先我們要了解各種 Feed Back 機制直接影響到的是估計參數群 Θ 的能力，而 Θ 估計的越準確代表它將會擁有一個更好的 P_{total} ，以及有更好的權重係數 Z_i 用以分離。所以我們直接討論各方法下估計參數群 Θ 的能力。我們取出 $[0,30]$ 度擺放的其中一個樣本做為比較依據，分別比較各種 Feed Back 機制的參數估計。

FB - ILD 參數估計比較

FB-ILD 主要改變是對 $\mu_{\alpha,i}$ 估計的修正，為了方便評比我們先設定兩個評分依據，分別為 $\mu_{\alpha,i,error}$ 及 $\Delta\mu_\alpha$ ，定義如下。

$\mu_{\alpha,i,error}$ ：估計 $\mu_{\alpha,i}$ 與真實 $\mu_{\alpha,i}$ 的差距，也就是 $\mu_{\alpha,i}$ 的估計誤差，估計誤差越小代表估計越準確，可以用以評斷估計的效能。

$$\mu_{\alpha,i,error} = |\mu_{\alpha,i,estimate} - \mu_{\alpha,i,real}|$$

$\Delta\mu_\alpha$ ： $\mu_{\alpha,1}$ 與 $\mu_{\alpha,2}$ 的差距，用以評估分離的難易。(距離越大越容易分離，且整體效能越好，在下面會證實)。

$$\Delta\mu_\alpha = |\mu_{\alpha,1,estimate} - \mu_{\alpha,2,estimate}|$$

為了了解 $\Delta\mu_\alpha$ 的用處，我們做了理想實驗下 $\Delta\mu_\alpha$ 與 SDR 和角度間的關係(圖.35)。我們看到圖.62 中上圖為 SDR - $\Delta\mu_\alpha$ 做圖，可以看的出來隨著 $\Delta\mu_\alpha$ 變大就有更好的 SDR；下圖 $\Delta\mu_\alpha$ - 角度做圖，圖中顯示 $\Delta\mu_\alpha$ 隨著角度變大，因此在各種模擬中 SDR 是隨著角度變大而上升的。

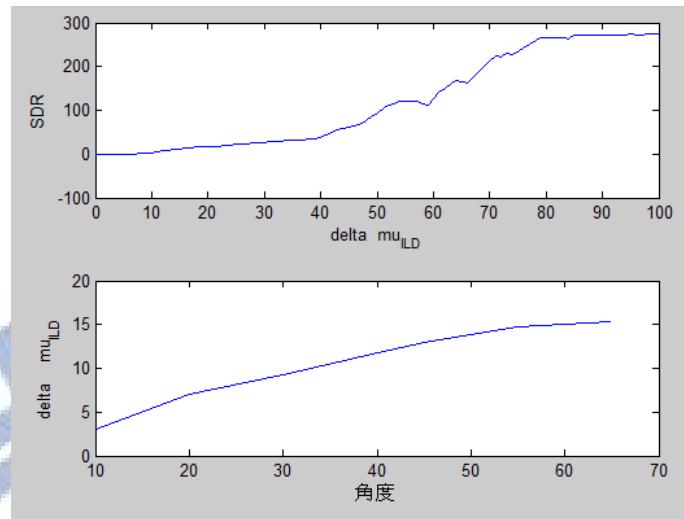


圖. 63 上圖為 SDR - $\Delta\mu_\alpha$ 做圖；下圖為 $\Delta\mu_\alpha$ - 角度做圖

在定義兩種評分標準後，我開始看 FB-ILD 的修正。

* $\mu_{\alpha,i,error}$

在圖.64($\mu_{\alpha,1,error}$ 曲線圖)中，FB - ILD 的方法大部分頻段都比原本 MESSL-SP 來的小，平均 $\mu_{\alpha,1,error}$ 比原 MESSL-SP 小了 0.2dB，這代表估計能力得到改善。

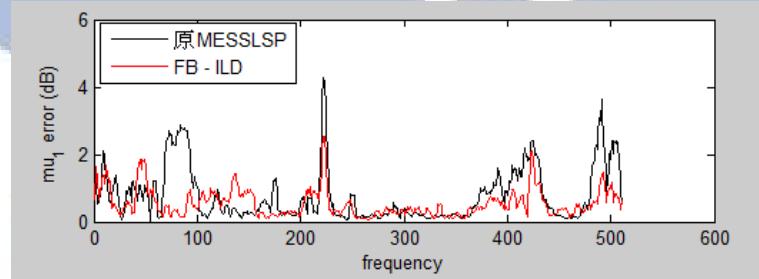


圖. 64 $\mu_{\alpha,1,error}$ 曲線圖，MESSL-SP(黑)、FB-ILD(紅)

[註]橫軸每點約等於 32Hz

在圖.65($\mu_{\alpha,2,error}$ 曲線圖)中，FB - ILD 在全頻帶的估計誤差都比原 MESSL-SP 來的小，平均 $\mu_{\alpha,2,error}$ 小了 2dB，估計能力得到改善。

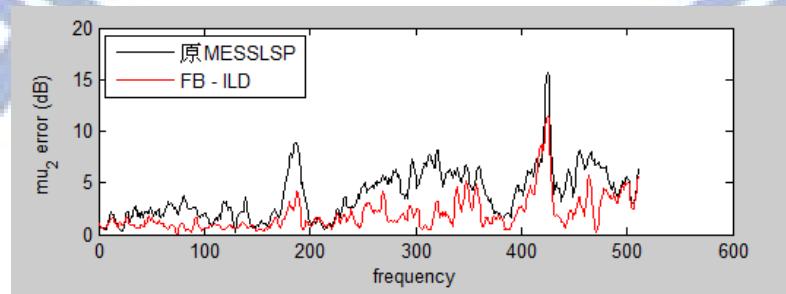


圖. 65 $\mu_{\alpha,2,error}$ 曲線圖，MESSL-SP(黑)、FB-ILD(紅)

[註]橫軸每點約等於 32Hz

* $\Delta\mu_\alpha$

在圖.66($\Delta\mu_\alpha$ 曲線圖)中顯示 FB-ILD(紅)在 $\Delta\mu_\alpha$ 的表現上比 MESSL-SP 方法來的大，代表更好分離。

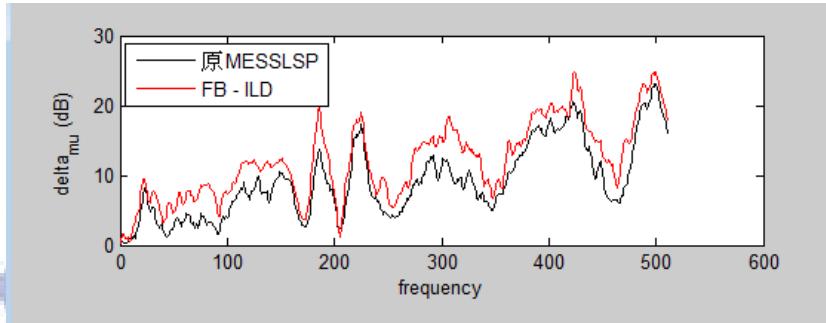


圖. 66 $\Delta\mu_\alpha$ 曲線圖，MESSL-SP(黑)、FB-ILD(紅)
[註]橫軸每點約等於 32Hz

總歸而言，FB-ILD 的方法估計 $\mu_{\alpha,i}$ 比較準確，且 $\Delta\mu_\alpha$ 的表現也較好，所以使得整體效能上升 2dB。

IPD 參數估計比較

FB-IPD 主要改變是對 $\mu_{\phi,i}$ 估計的修正，為了方便評比我們先設定兩個評分依據，分別為 $\mu_{\phi,i,error}$ 及 $\Delta\mu_\phi$ ，定義如下。

$\mu_{\phi,i,error}$ ：估計 $\mu_{\phi,i}$ 與真實 $\mu_{\phi,i}$ 的差距，也就是 $\mu_{\phi,i}$ 的估計誤差，估計誤差越小代表估計越準確，可以用以評斷估計的效能。

$$\mu_{\phi,i,error} = |\mu_{\phi,i,estimate} - \mu_{\phi,i,real}|$$

$\Delta\mu_\phi$ ： $\mu_{\phi,1}$ 與 $\mu_{\phi,2}$ 的差距，用以評估分離的難易。(距離越大越容易分離，且整體效能越好，在下面會證實)。

$$\Delta\mu_\phi = |\mu_{\phi,1,estimate} - \mu_{\phi,2,estimate}|$$

接著一樣觀察它們的效應。

* $\mu_{\phi,i,error}$

在圖.67($\mu_{\phi,1,error}$ 曲線圖)中，FB-IPD 的方法大部分頻段都稍稍比原本 MESSL-SP 來的小，平均 $\mu_{\phi,1,error}$ 比原 MESSL-SP 小了 0.1dB，估計能力稍稍得到改善。

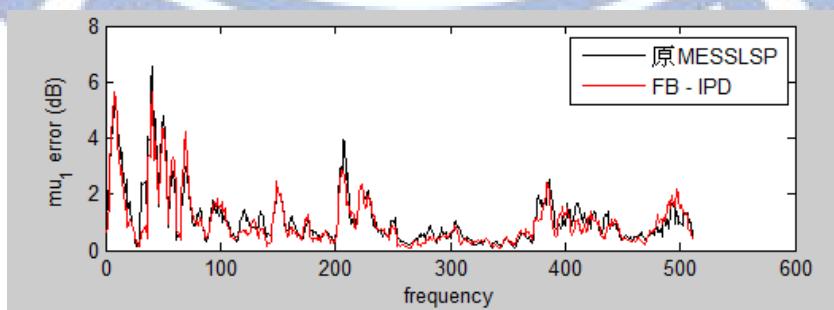


圖. 67 $\mu_{\phi,1,error}$ 曲線圖，MESSL-SP(黑)、FB-IPD(紅)
[註]橫軸每點約等於 32Hz

在圖.68 ($\mu_{\phi,2,error}$ 曲線圖) 中，FB – ILD 在全頻帶的估計誤差都比原 MESSL-SP 稍稍的小了一些(小了 0.5 dB)，得到些微的改善。

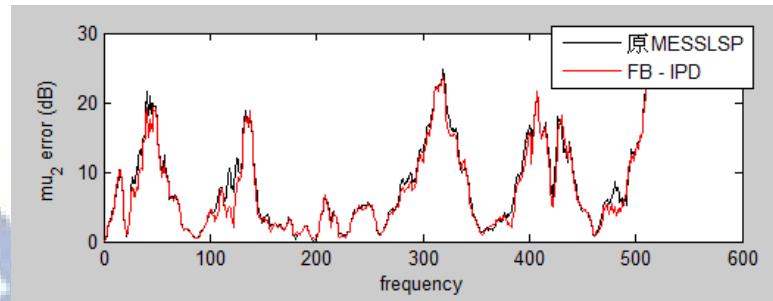


圖. 68 $\mu_{\phi,2,error}$ 曲線圖，MESSL-SP(黑)、FB-IPD(紅)

[註]橫軸每點約等於 32Hz

* $\Delta\mu_\phi$

在圖.69($\Delta\mu_\phi$ 曲線圖) 中顯示 FB-IPD(紅)在 $\Delta\mu_\alpha$ 的表現上比 MESSL-SP 方法來的稍稍大了些，這代表比較容易一些分離。

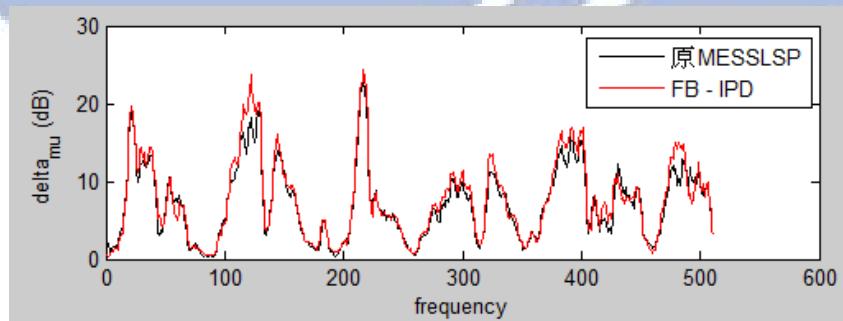


圖. 69 $\Delta\mu_\phi$ 曲線圖，MESSL-SP(黑)、FB-IPD(紅)

[註]橫軸每點約等於 32Hz

總歸而言，FB-IPD 的方法在 $\mu_{\phi,i}$ 估計能力及 $\Delta\mu_\alpha$ 的表現都只有些微的改善，使得整體效能稍稍上升 0.2 dB。

3.6.2 條件結合加入 Feed Back 機制效能比較

這節中我們模擬回授機制加入我們的演算法獨立結合(預分群)的改善，看到成果圖.70 及圖.71，可以看到回授機制在左聲道(干擾較小聲道)較有改善；在平均上加入 FB 機制後再左聲道比原本好了 0.8dB；右聲道則是差了 1dB 左右(干擾較大聲道)。

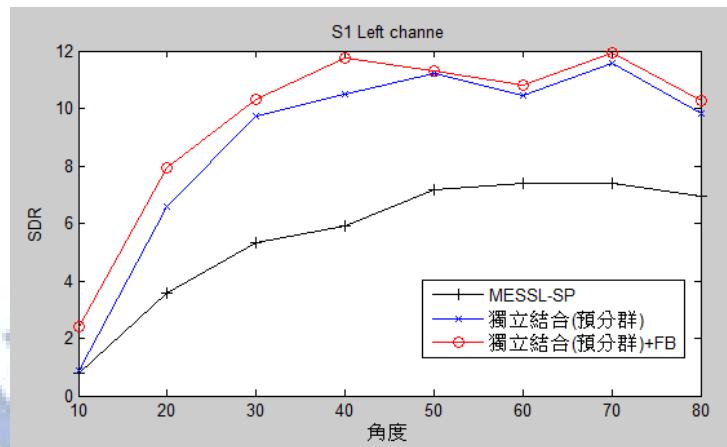


圖. 70 雙聲源獨立結合(預分群)回授演算法成效(左聲道)

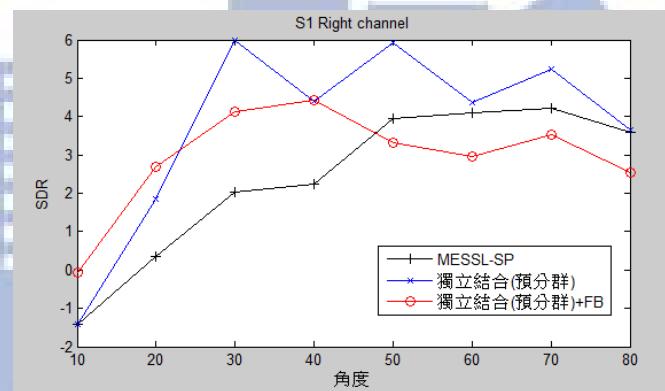


圖. 71 雙聲源獨立結合(預分群)回授演算法成效(右聲道)

我們仿造 3.6.1 節設立檢測標準，如下

$\mu_{i,error}$ ：估計 μ_i 與真實 μ_i 的差距，也就是 μ_i 的估計誤差，估計誤差越小代表估計越準確，可以用以評斷估計的效能。

$$\mu_{i,error} = |\mu_{i,estimate} - \mu_{i,real}|$$

$\Delta\mu$ ： $\mu_{\alpha,1}$ 與 $\mu_{\alpha,2}$ 的差距，用以評估分離的難易。(距離越大越容易分離，且整體效能越好，在下面會證實)。

$$\Delta\mu = |\mu_{1,estimate} - \mu_{2,estimate}|$$

從圖.72 中，可以看到加了回授後 $\mu_{1,error}$ 比原本來的平均小 0.8；圖.73 中，可以看到加了回授後 $\mu_{2,error}$ 比原本來的平均小 0.6dB。

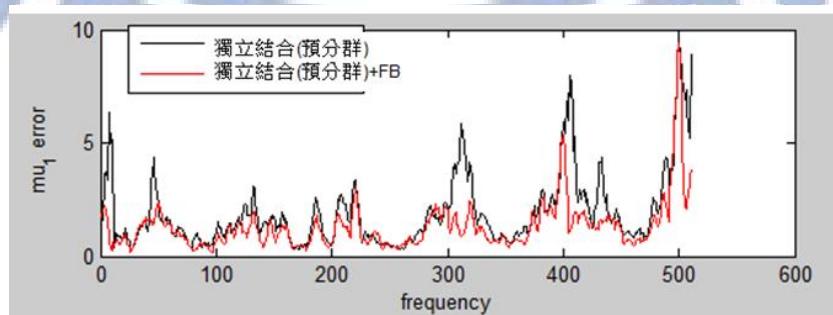


圖. 72 $\mu_{1,error}$ 曲線圖，獨立結合(黑)、獨立結合+回授(紅)

[註]橫軸每點約等於 32Hz

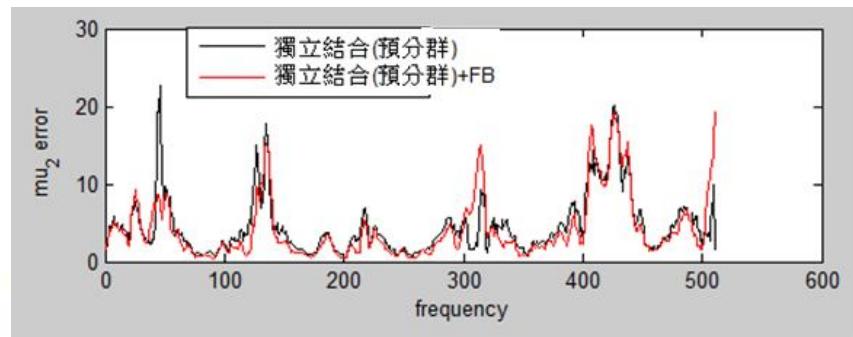


圖. 73 μ_2 ,error 曲線圖，獨立結合(黑)、獨立結合+回授(紅)

[註]橫軸每點約等於 32Hz

在 $\Delta\mu$ (圖.74)方面，有好有壞，難以分別好壞。

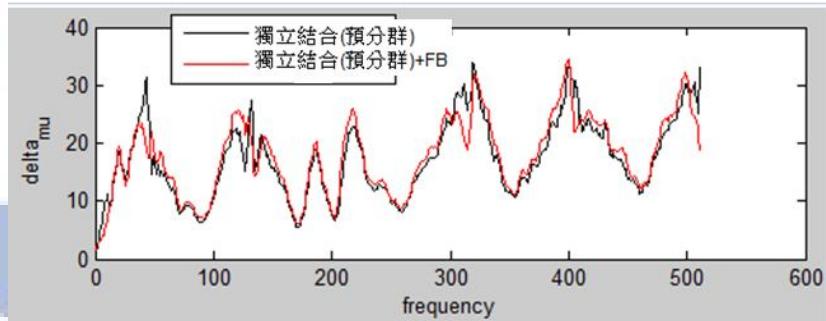


圖. 74 $\Delta\mu$ 曲線圖，獨立結合(黑)、獨立結合+回授(紅)

[註]橫軸每點約等於 32Hz

總而言之，由於對參數估計的準確度上升，對整體效能也會有上升。

1.17 3.7 三聲源分離情況下討論

此段落中，我們討論 MESSL-SP、聲道差非獨立(MESSL 修正型)與獨立結合(預分群)三種方法在三聲源分離的表現。我們會再 3.6.1 節中先測試這三種方法的 SDR 對角度測試，而我們發現我們改善的兩種方法，在三聲源的情況下並沒有大幅的改善 SDR，只有約 0~0.5dB 不等的改善；接著在 3.6.2 節中我們切入各方法在機率模型上的表現，可以看出我們提出的兩種方法不管在聲道差平面或是左右聲道強度平面上表現都比原 MESSL-SP 來的好。

1.17.1 3.7.1 三聲源下演算法成效比較

從圖.75 及圖.76 中我們看到 MESSL 修正型與獨立結合(預分群)的 case 分別與

MESSL-SP 相比，若從平均成效上看的話，MESSL 修正型與 MESSL-SP 相比：左聲道平均 SDR 上升 0.4 dB；右聲道上升 0.6，而 MESSL 修正型在低角度有分離錯誤的發生(未分離前 SDR 基準線為 3~5dB 不等)；獨立結合(預分群)與 MESSL-SP 相比：左聲道平均 SDR 上升 4.2 dB；右聲道上升 3.1dB。

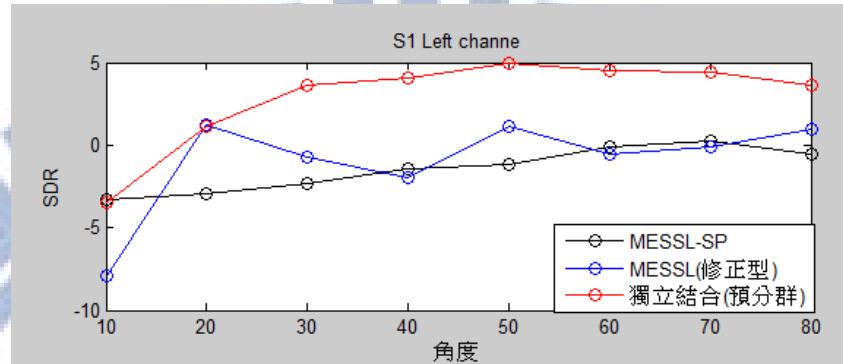


圖. 75 三聲源演算法成效圖(左聲道)

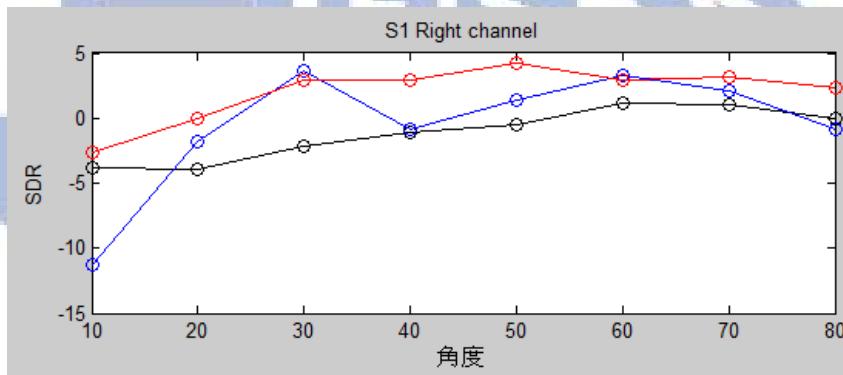


圖. 76 三聲源演算法成效圖(右聲道)

1.17.2 3.7.2 機率模型比較

1. 30 度樣本下，比較聲道差機率估計

(1) 聲道差邊際機率估計部分

首先，先看到圖.77 及圖.78，可以看到在修正獨立性後，不管再 Mean 的方面還是 Variance 的方面都比原本好上許多。

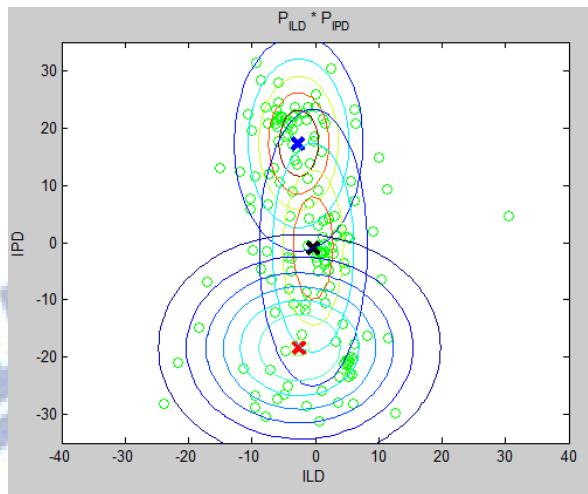


圖. 77 三聲源 MESSL-SP 聲道差平面機率等高線圖(1500Hz)，綠點為 Data

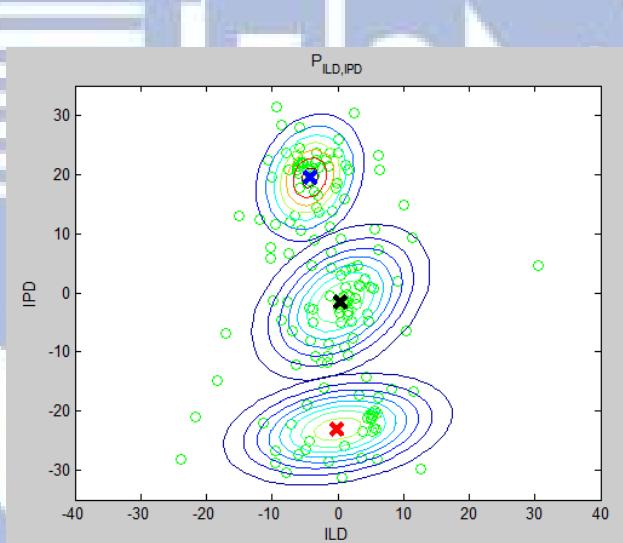


圖. 78 三聲源 MESSL 修正型聲道差平面機率等高線圖(1500Hz)，綠點為 Data

而 MESSL 修正型和獨立結合(預分群)比起來(圖.78 與圖.79)，基本上是相同的。

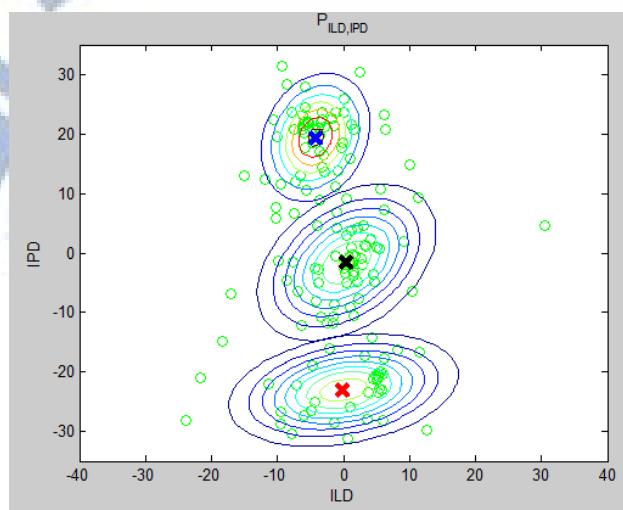


圖. 79 三聲源獨立結合(預分離)聲道差平面機率等高線圖(1500Hz)

(2) 左右聲道邊際機率估計部分

比較圖.80 和圖.81 後，可以很明顯的發現我們提出的方法在 (\hat{L}, \hat{R}) 資料的估計上確實有比源 MESSL-SP 來的好，但式依然有缺陷，例如可以看到在圖.81 中，中間(黑 x)的高斯分布太過細長，這是由於在 3.1 章節中提到的 (\hat{L}, \hat{R}) 資料分群機制採用硬式二分法，二分法會使得夾在中間的資料群被兩旁壓縮，若兩旁的數目過多會導致中間的群組完全被壓榨，可能會造成分離錯誤發生。

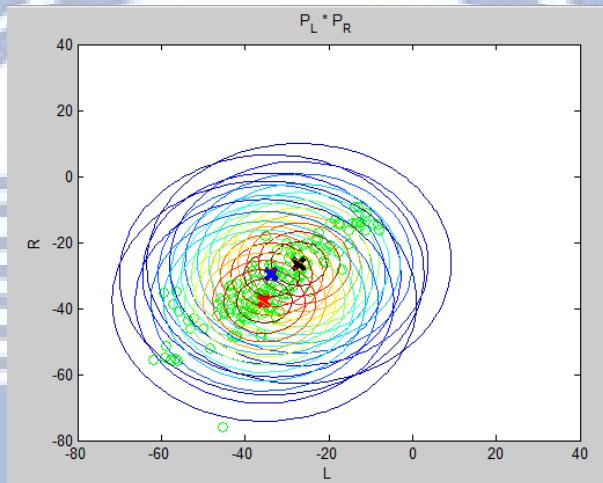


圖. 80 MESSL-SP 左右聲道平面機率等高線圖(1500Hz)

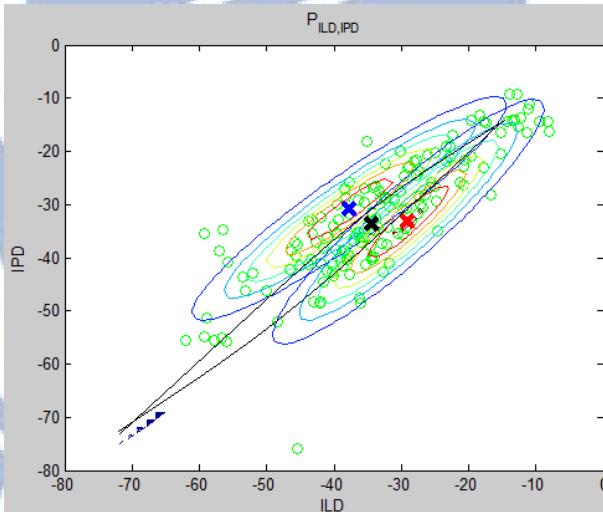


圖. 81 獨立結合(預分群)左右聲道平面機率等高線圖(1500Hz)

1.17.3 3.7.3 三聲源下加入回授機制

我們測試在獨立結合(預分群)演算法下加入回授機制會如何，成果如下圖.82 及圖.83，可以看到平均上加了 FB 的結果都比沒加前來的差。

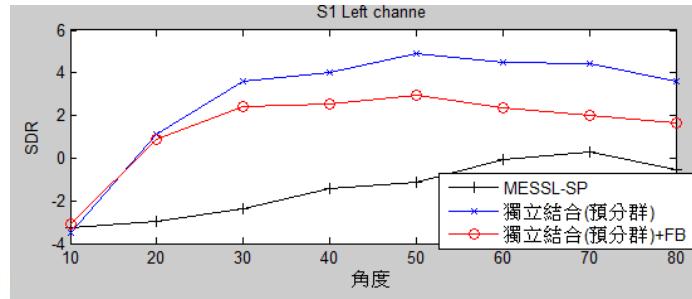


圖. 82 三聲源獨立結合(預分群)回授成效(左聲道)

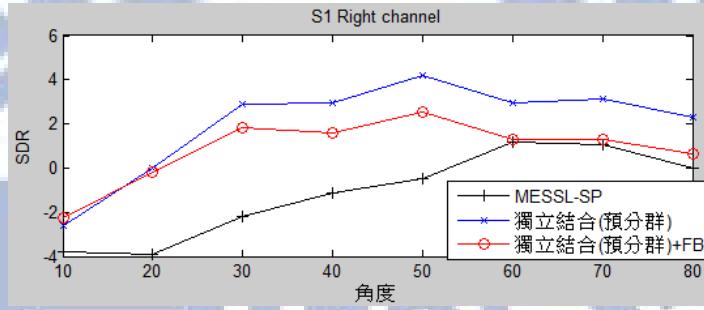


圖. 83 三聲源獨立結合(預分群)回授成效(右聲道)

接著我們採用 3.6.1 節中的檢視方法：

在 μ_{error} 方面，從圖.84 我們發現在主要目標參數 μ_1 的估計上準確性變差(平均下降 1.2)；圖.85 及圖.86 部分，其餘目標 μ_2 、 μ_3 的估計都有變準確，可以推論回授的機制再干擾較大的聲源估計上部會得以改善。(可以與圖.70 呼應)

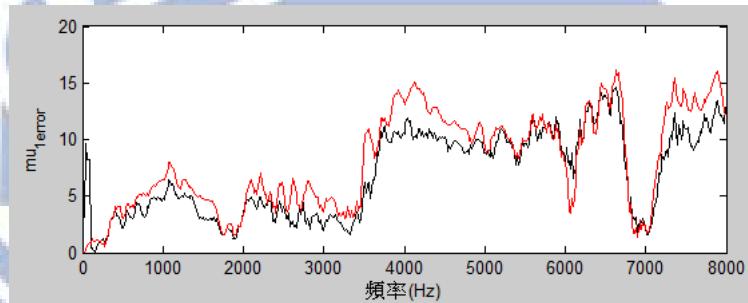


圖. 84 $\mu_{1,\text{error}}$ 曲線圖，獨立結合(黑)、獨立結合+回授(紅)

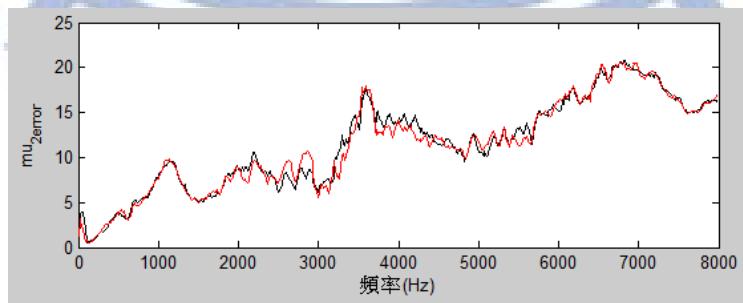


圖. 85 $\mu_{1,\text{error}}$ 曲線圖，獨立結合(黑)、獨立結合+回授(紅)

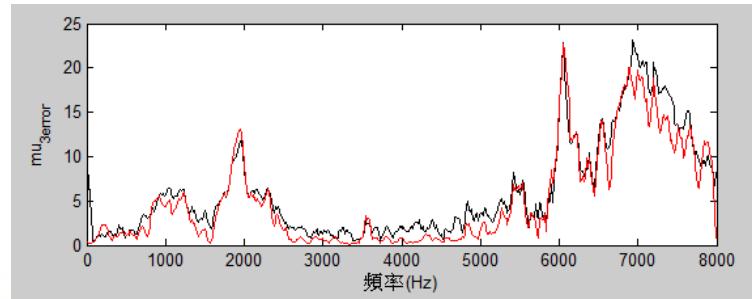


圖. 86 $\mu_{1,\text{error}}$ 曲線圖，獨立結合(黑)、獨立結合+回授(紅)

1.18 3.8 實測效能

本段落我們用下圖.87 中的 Sony 的 ICD-UX543F 錄音筆和 ECM-DS70P 外接立體聲麥克風(兩麥克風間距 3cm)於大小為 10(m)X7.6(m)的房間中間對雙聲源分離測試兩種設置：如圖.88 大角度(兩聲源分別放置在距麥克風 1 公尺 \pm 35度角)、小角度(兩聲源分別放置在距麥克風 1 公尺 \pm 15度角)。並由大小為 100 的 TIMIT 語音庫中任選兩聲源為一組，每種設置各測試 10 組平均之結果。



圖. 87 綠音設備：(1)麥克風 (2)錄音筆 (3)組合

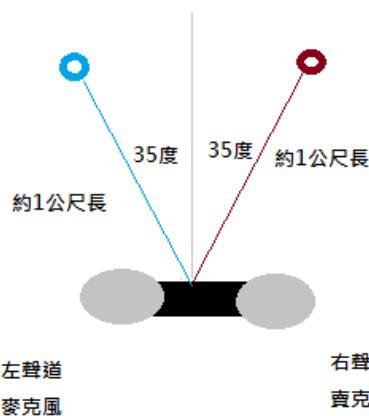


圖. 88 實測擺設

從下表.2 左邊音源成果，整體上看到不管大角度或小角度我們提出的方法都比 MESSL 來的好 1dB 以上的改善。

表. 2 實測左邊音源分離 SDR 成果

		獨立結合 (預分群)	MESSL	效能改善
小角度	左聲道 SDR	5.5 dB	4.3 dB	1.2 dB
	右聲道 SDR	2.9 dB	1.8 dB	1.1 dB
大角度	左聲道 SDR	7.6 dB	5.4 dB	2.2 dB
	右聲道 SDR	3.1 dB	2.3 dB	0.8 dB

而在表.3 右邊音源成果方面，整體平均約有 0.5dB 的改善，但在大角度右聲道 SDR 的部分測出結果是-0.3dB 的成長，是基於何原因我們並沒有深入討論。

表. 3 實測右邊音源分離 SDR 成果

		獨立結合 (預分群)	MESSL	效能改善
小角度	左聲道 SDR	-1.3 dB	-2.1 dB	0.8 dB
	右聲道 SDR	5.1 dB	4.6 dB	0.5 dB
大角度	左聲道 SDR	-0.5 dB	-1.5 dB	1 dB
	右聲道 SDR	6.4 dB	6.7 dB	-0.3 dB

第四章 結論

經由 MESSL 與二維 GMM 分法的比較，我們注意到各資料模型間正確關係的重要性，因此我們討論在不同頻率上 ILD 與 IPD 資料間的相關性，經實際模擬後我們採用全頻非獨立的方式處理聲道差平面。接著我們關注到 MESSL-SP 所引進的”左右聲道資料”是新的技術，不過由於它的模型架構不妥而導致成長有限，而我們經由對左右聲道平面資料分布的觀察，提出二維 GMM 的模型來處理。對於兩平面結合的部分，我們架設一個較可信的演算法架構，稱為獨立結合(預分群)演算法，它在成效上比以往各方法都好上許多。除此之外，我們也加入新的機制稱為”分離訊虔回授機制”，在理論的目的上它是用來幫助估計正確的參數用，但在測試結果顯示回授的機制在受干擾程度較小的情況下可以發揮其用處，而對於干擾較大的情況則是會使成效比原本來的差，其內部的原因我們也沒得出有用的結論，不過我們所提出的獨立結合(預分群)演算法不管在雙聲源或是三聲源的情況下都是目前演算法中最好的。

在未來的部分，我們希望可以進一步的討論回授機制與干擾程度的關係，並正確的改進回授機制；再者，我們希望可以從水平平面的聲音分離推廣到 3D 平面的分離(方位角範圍 $[-180, 180]$ 度，仰角範圍 $[-90, 90]$ 度)；以及做到聲源定位的功能(可以引用[3]、[4]的方法)。

Chapter 2 參考資料

- [1] Özgür Yilmaz and Scott Rickard, "Blind Separation of Speech Mixtures via Time-Frequency Masking." *IEEE Transactions ,Signal Processing* (Vol.52, Issue: 7) p1830 – 1847 , July 2004
- [2] Tobias May, Steven van de Par, and Armin Kohlrausch , "A Probabilistic Model for Robust Localization Based on a Binaural Auditory Front-End." *IEEE Transactions , Audio, Speech, Language Processing,* (Vol.19 , Issue: 1) ,p1 – 13 , Jan. 2011
- [3] Martin Raspaud, Harald Viste, and Gianpaolo Evangelista ,Senior Member , "Binaural Source Localization by Joint Estimation of ILD and ITD." , *IEEE Transactions Audio, Speech, Language Processing* (Vol.18 , Issue: 1) , p68 – 77 , Jan. 2010
- [4] joan Mouba and Sylvain Marchand , "A Source Loclization/Separation/Respatialization System Based On Unsupervised Classification Of Interaural Cues." , *Proc. of the 9 th Int. Conference on Digital Audio Effects* (p233-238), Montreal, Canada, September 18-20, 2006
- [5] Mandel, M .I., Weiss, R.J. and Ellis, D.P.W., 2010. "Model-based expectation-maximization source separation and localization. " *IEEE Transactions ,Audio ,Speech, Language Process*(Vov.18 NO.2), p382–394
- [6] R. J. Weiss, M. I. Mandel, and D. P. W. Ellis, "Source Separation Based on Binaural Cues and Source Model Constraints.", in *Proc. Interspeech* ,Sep. 2008, pp. 419–422 (實驗室論文)
- [7] Ron J. Weiss , Michael I. Mandel, Daniel P.W. Ellis , "Combining Localization Cues And Source Model Constraints For Binaural Source Separation. " , *Lab ROSA, Dept. of Electrical Engineering, Columbia University, New York, NY 10027, USA Available online 12 January 2011* (實驗室論文)
- [8] N. Roman, D. L. Wang, and G. J. Brown, "A Classification-Based Cock-Tail Party Processor.", *Adv. Neural Info. Process. Syst. , S. Thrun, L.Saul, and B. Schölkopf, Eds. Cambridge, MA: M IT Press, 2004*, pp.1425–1432
- [9] S. H. Colburn and A. Kulkarni, "Models of sound localization," in Sound Source Localization .", *Adv. Neural Popper and R. R. Fay, Eds. New York: Springer, 2005, vol. 25, ch. 8*, pp. 272–316.
- [10] Wenyi Zhang and Rao, B.D. , "A Two Microphone-Based Approach for Source Localization of Multiple Speech Sources. " , *IEEE Transactions , Audio, Speech, Language Processing*, (vol. 18, NO. 8), november 2010 , pp 1913 – 1928
- [11] A. Hyvärinen, "Survey on independent component analysis.", *Neural Comput. Surv.*, vol. 2, no. 94–128, pp. 3–1, 1999.
- [12] E. Vincent, R. Gribonval, and C. Fevotte, "Performance Measurement in Blind Audio Source Separation," *IEEE Transactions ,Audio, Speech, Language Process* , (vol. 14, no. 4), pp. 1462–1469, Jul. 2006

