

國立交通大學

電信工程研究所

碩士論文

有關響度與分頻解析度之聽損耳蝸模型的建立

A Construction of Hearing Impaired Cochlea
Model about Loudness and Frequency Selectivity

研究生：李文中

指導教授：冀泰石

中華民國一百年十月七日

有關響度與分頻解析度之聽損耳蝸模型的建立

A Construction of Hearing Impaired Cochlea Model about
Loudness and Frequency Selectivity

研究生：李文中

Student: Wen-Chung Lee

指導教授：冀泰石 博士

Advisor: Dr. Tai-Shih Chi

國立交通大學

電信工程研究所

碩士論文

A Thesis

Submitted to Institute of Communication Engineering
College of Electrical and Computer Engineering

National Chiao-Tung University

In Partial Fulfillment of the Requirements

for the Degree of

Master of Science in

Communication Engineering

October 2011

Hsin-Chu, Taiwan, Republic of China

中華民國一百年十月

有關響度與分頻解析度之聽損耳蝸模型的建立

學生：李文中

指導教授：冀泰石 博士

國立交通大學電信工程研究所

感知訊號處理實驗室

摘要

為了能夠即時驗證助聽器演算法的效果，我們必須有一套能夠模擬聽損病患耳蝸的系統才行；因此我們參考 Moore 團隊所提出的，模擬最小可聽水平提升、響度聚集、以及分頻解析度降低的耳蝸模型，其中有關響度的變化是以濾波器組的方式處理，將訊號分頻之後把能量 N 次方，放大相鄰時間點能量上的差距；而分頻解析度降低則是以頻譜模糊化來取代，將基底膜接收聲音的機制用激發模式表示，並將其矩陣化後以反矩陣的方式來完成，最後在短時傅利葉轉換的架構下進行。然而這樣子的做法有幾個缺點：1. 將能量 N 次方只能決定響度聚集的程度，無法決定最小可聽水平的值。2. 短時傅利葉轉換中保留原相位頻譜會造成模擬程度不匹配，此外重疊的濾波器組解析度被傅利葉轉換所限制。3. 需經過兩次短時傅利葉轉換與一次濾波器組的架構，計算複雜度高。因此，我們首先將響度變化的機制以分段的線性公式一點一點的計算；接著我們將變寬的聽覺濾波器用正常濾波器的線性組合組合而成，計算出原始的濾波器對所有變寬濾波器的貢獻，並將白高斯雜訊濾波過後取出載波，和經過模糊化的能量相乘後合成為聲音訊號，如此整個系統便可在濾波器組的架構下完成。最後我們設計聽力測驗，證明我們的模型與 Moore 團隊的模型會有相似的結果，也符合聽損患者回報出的一些聽損現象。

A Construction of Hearing Impaired Cochlea Model about Loudness and Frequency Selectivity

Student: Wen-Chung Lee

Advisor: Dr. Tai-Shih Chi

Institute of Communication Engineering

National Chiao-Tung University

Perception Signal Processing Laboratory

Abstract

In order to verify the effectiveness of algorithms developed for hearing aids, developing a system that simulates the cochlear of the hearing impaired is in great demand. Thus, we investigate the model developed by Moore's group in addressing threshold elevation, loudness recruitment, and reduced frequency selectivity of hearing impaired. The simulation about loudness was done by extracting the sub-band signal and powering the instantaneous magnitude according to the degree of recruitment. To simulate the frequency smearing, matrix multiplications of filter-bank coefficients of the normal/impaired cochlear were carried out. However, there are some drawbacks in Moore's model. First, the minimum thresholds can not be set alone. Second, preserving the phase spectrum would mitigate the desired degree of magnitude smearing. Also, the spacing of the critical bands is restricted by FFT. Third, the computational cost is very high. To overcome these drawbacks, we (1) use a piecewise formula to decide the loudness sample by sample; (2) model the widened filter by a linear combination of normal filters and compute contributions from normal filters to each widened filter. Besides, we use white Gaussian noise to generate the band-passed carrier and synthesize the signal. Our approaches can be done in the same filter-bank architecture to reduce computational load. Last, listening tests are carried out to verify our model. Results show that our model not only acts like Moore's model, but also exhibits characteristics reported by the hearing impaired.

誌 謝

十八年，該是一段很長的時間。腦中畫面還是剛踏進一年義班、看著陌生人只會愣在原地的的小學生；轉眼間我已穿上碩士長袍，回首向交大道聲珍重再見。

回首「珍重再見」——一句簡短的話語，一個瀟灑的身影，背後卻是由許多人共同成就而得。還記得初上碩班時，秉持著初生之犢不怕死的精神，兀自選了「補償分頻解析度降低演算法」這個題目；然而，在茫茫論文海中要找尋一篇相關的論文，竟是如此的困難！為此，當我找到一篇這樣的論文時，便像是落海客在無邊大海中摸到一根漂流木，緊抓不放；雖然暫時得以安身，卻放棄繼續尋找能夠航向終點的船隻，侷限了自己的眼界。感恩冀泰石老師和 711 實驗室的學長們，總是在我每一次進度報告之後，還花時間討論我研究上的優缺點及可行性，適時的修正我的方向，讓我不致於一頭栽進去，愈走愈偏。感恩學長們總是無私的分享自己研究上的經驗和成果，讓我不用為了相同的原因，相同的害蟲，苦思數日而不得其解；因為站在學長們的肩膀上，讓我能夠看得更高，更遠，也得以在短短的兩年內能夠有初步的研究成果。

一塊璞玉畢生的心願，就是與和氏的邂逅；如果我真能是一塊璞玉，我想冀老師就是那為了玉而四處奔走的和氏。據傳和氏為了與他毫不相干的一塊玉，雙腿俱廢，哭盡眼淚；相信冀老師為了我這個不成材的學生，付出的辛勞也絕不在話下——不知道老師的白頭髮最近又多了幾根？然而若不是老師的提攜，我想我不會有愈磨愈光的一天。一直以來我總認為自己以追根究柢的精神做學問，比別人更認真，比別人更有求知的欲望；然而在漫長的研究過程中，老師一句句的金玉良言，讓我看清了自己在面對真正困難的挑戰時，總是找藉口來說服自己這樣就夠了——才剛走了幾步，卻開始打起退堂鼓。感恩老師每次與我討論研究時，總是和言悅色的面對我理不直而氣壯的說著：這裡好像不行、那裡不知道為什麼；感恩老師總是花時間對我們曉以大義，讓我們不致於因為苦悶的研究過程，而忘了自己研究的初衷。如果我真能是一塊璞玉，我期許自己能夠不辜負這塊玉的價值——「冀氏出品，必屬名品！」

龜山島

每當蘭陽的孩子搭火車外出

當他從車窗望著你時

總是分不清空氣中的哀愁

到底是你的，或是他的 —黃春明《龜山島》

六年前第一次離鄉背井，為了追求夢想的人生，獨自踏上了竹塹風城；而異鄉遊子的惆悵、理工課程的陌生、獨自生活的不便，加上巨蟹特有的多愁善感，卻每每在夜深人靜時，發酵成一股濃濃的鄉愁—好像逃離這個地方。幸運的是，我在這裡遇到了許多生命中的貴人：感恩以十二舍 319 為首的高雄人，你們是我的交大生活中，不可或缺的一部份；感恩慈濟的每一位師姑師伯、學長學姊、還有慈青伙伴們，和你們相處的過程讓我成長，也因為你們的關心和陪伴，讓我對新竹多了一份認同感；感恩系學會和電通營的所有成員，讓我可以擁有多采多姿的大學生活；感恩蘭友會的每一個人，蘭陽週的點點滴滴和精彩的小梅竹，都是我們曾經走過的足跡；感恩國中的同學們，因為每年寒暑的固定聚會，看著大家各奔前程，讓我知道我還可以更努力；感恩實驗室的同屆伙伴—華山、名媛、阿雞，讓我在研究煩悶時，還可以聊天嘴炮瞎扯淡；感恩所有曾經幫助過我的人，因為有你們，才有今天的我。

最後，我要感恩我的父母，因為有你們辛苦賺錢，讓我可以不用擔心學費，專心享受大學生活，揮灑青春；感恩你們在我每一次回家時，總是準備了好多好吃的食物，讓我可以好好休息；感恩你們從小到大的栽培，讓我踏進了社會，才發現原來自己的條件比別人好很多；感恩你們一直牽著我的手，帶我走過 24 年的人生旅程。

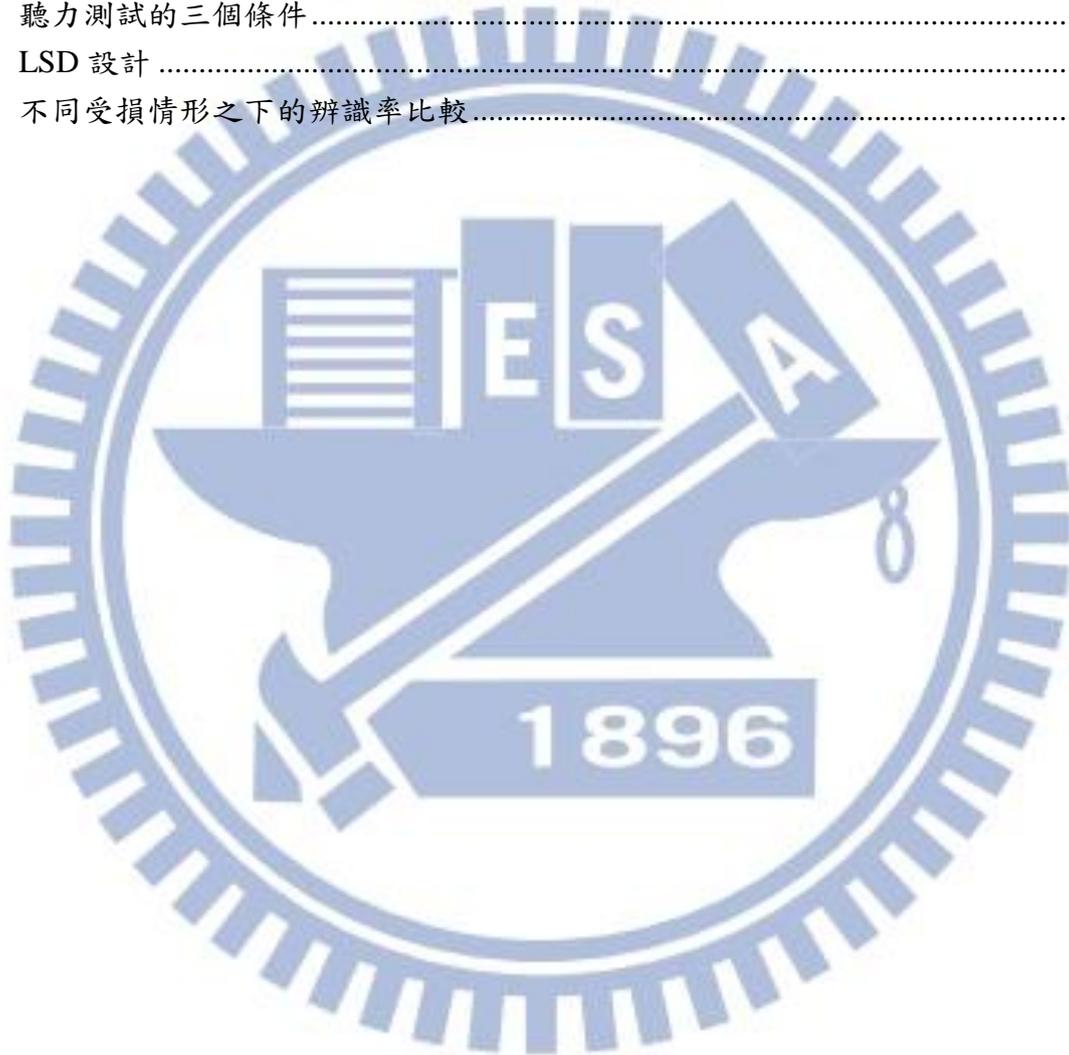
目 錄

中文摘要.....	i
英文摘要.....	ii
誌 謝.....	iii
目 錄.....	v
表 目 錄.....	vii
圖 目 錄.....	viii
第一章 緒論.....	1
1.1 研究背景.....	1
1.2 聽損現象簡介.....	2
1.2.1 最小可聽水平提升與響度聚集.....	2
1.2.2 分頻解析度降低.....	3
1.3 研究方法.....	4
1.4 章節大綱.....	4
第二章 感知訊號處理基礎.....	5
2.1 生理聽覺現象與特性.....	5
2.1.1 聽覺的產生.....	5
2.1.2 行進波在基底膜上的特性.....	6
2.1.3 濾波器組與激發模式 (excitation pattern)	8
2.1.4 響度 (loudness)	10
2.2 短時傅利葉轉換 (STFT)	13
2.2.1 訊框長度、訊框位移、分析視窗.....	15
2.2.2 快速傅利葉轉換點數 (FFT size)	17
2.2.3 OLA 與合成視窗.....	18
2.3 濾波器組的選擇.....	19
2.3.1 小波轉換.....	19
2.3.2 聽覺濾波器組 (auditory filter bank)	21
第三章 聽損耳蝸模型.....	25
3.1 頻譜模糊化模型.....	25
3.1.1 演算法架構與背景.....	25
3.1.2 聽覺濾波器組.....	26
3.1.3 效能分析.....	27
3.1.4 相位頻譜補償.....	29
3.1.5 相位頻譜補償延伸.....	31
3.2 響度模型.....	33

3.2.1 演算法架構與背景.....	33
3.2.2 模型實作概述.....	35
3.2.3 模擬結果與分析.....	35
3.3 混合模型.....	36
第四章 分頻解析度補償.....	39
4.1 多頻帶頻率壓縮演算法.....	39
4.1.1 演算法核心架構.....	39
4.1.2 壓縮方式比較.....	40
4.1.3 演算法分析與討論.....	42
4.2 多頻帶頻率壓縮延伸.....	43
4.2.1 演算法流程與架構.....	44
4.2.2 實作結果分析.....	45
4.2.3 未來方向.....	47
4.3 矩陣化模型的應用.....	48
第五章 基於濾波器組的聽損模型.....	49
5.1 響度模型.....	49
5.1.1 主要架構.....	49
5.1.2 實作結果.....	50
5.1.3 迭代封包補償法.....	51
5.2 頻譜模糊化模型.....	53
5.2.1 演算法概念.....	53
5.2.2 演算法實作詳述.....	54
5.2.3 實作結果與分析.....	57
5.2.4 混合模型.....	58
5.3 取樣頻率與載波比較.....	58
5.3.1 取樣頻率比較.....	59
5.3.2 載波比較.....	60
5.4 實驗設計與結果分析.....	61
5.4.1 實驗一.....	61
5.4.2 實驗二.....	64
5.5 結論.....	66
第六章 未來展望.....	67
參考文獻.....	69

表目錄

表 1	迭代方法對語音品質的影響.....	52
表 2	聽力測試的六個條件.....	62
表 3	LSD 設計.....	63
表 4	在 B6R3 的條件下，兩種演算法的辨識率.....	63
表 5	聽力測試的三個條件.....	64
表 6	LSD 設計.....	65
表 7	不同受損情形之下的辨識率比較.....	65



圖目錄

圖 1	不同的聽覺響度—物理強度的對應曲線.....	2
圖 2	受損患者與正常人的聽覺濾波器比較.....	3
圖 3	耳蝸構造圖.....	5
圖 4	柯替式器.....	6
圖 5	行進波在基底膜的振動模式.....	6
圖 6	相差倍頻單音在基底膜的共振位置示意圖.....	7
圖 7	基底膜上的響應示意圖.....	8
圖 8	臨界頻帶與頻率的關係，縱軸為〔27〕中提出的分頻方式.....	9
圖 9	由濾波器形狀計算激發模式.....	10
圖 10	實線為測量 1KHz 單音的強度-響度對應關係，虛線為〔30〕的公式.....	11
圖 11	人耳可聽到的聲音頻率及能量範圍.....	12
圖 12	以 1KHz 為基準的等響度曲線.....	13
圖 13	非時變訊號與時變訊號之時-頻比較圖.....	14
圖 14	左圖為漢明視窗的時間訊號；右圖為漢明視窗與矩形視窗的頻譜響應比較.....	16
圖 15	矩形視窗與漢明視窗的結果比較圖.....	16
圖 16	OLA 流程示意圖.....	18
圖 17	傅利葉轉換與離散小波轉換的時-頻比較圖.....	20
圖 18	以遞迴濾波器組分割頻譜.....	21
圖 19	129 個不包含遮蔽效應的濾波器的平方頻率響應及總和.....	22
圖 20	128 個包含遮蔽效應的濾波器的平方頻率響應及總和.....	23
圖 21	頻譜模糊化演算法流程圖解.....	25
圖 22	母音/a/的連續框架訊號模糊化前後比較圖.....	28
圖 23	LSEE-MSTFTM 演算法.....	29
圖 24	LSEE-MSTFTM 演算法中連續迭代的變化.....	30
圖 25	結合 LSEE-MSTFTM 與 Baer 演算法的比較.....	31
圖 26	最小平方差變化圖.....	31
圖 27	由上而下分別是框架位移 2、1、1/4 倍的 LS 聲譜圖與理想聲譜圖的差別.....	33
圖 28	模型實作流程圖解.....	35
圖 29	原訊號聲譜圖.....	36
圖 30	(a)、(b)分別為經過 2 次方、3 次方處理的聲譜圖.....	36
圖 31	混合模型實作流程圖解.....	37
圖 32	不同模擬順序的比較圖.....	38
圖 33	多頻帶頻率壓縮演算法的主要想法.....	39
圖 34	點對點對應示意圖.....	40

圖 35	頻率線性相加示意圖	41
圖 36	頻率區段對應示意圖	42
圖 37	取樣頻率不同的訊號頻譜圖	44
圖 38	延伸演算法流程圖	44
圖 39	簡單內插法分解說明	46
圖 40	Power of N 與分段線性公式實現 N=2 的聲譜比較圖	50
圖 41	Power of N 與分段線性公式實現 N=3 的聲譜比較圖	51
圖 42	左右分別為第 28、117 個 channel 的封包比較圖	51
圖 43	迭代封包補償法示意圖	52
圖 44	頻譜模糊化演算法流程圖	54
圖 45	由 a~d 分別為第 1、45、90、128 個變寬 6 倍的 channel 最佳化結果	55
圖 46	加上原訊號載波得到的聲譜圖	56
圖 47	a~d 分別為原訊號、模糊 1.5、3、6 倍的聲譜圖	57
圖 48	混合模型流程圖	58
圖 49	16KHz 取樣環境下，聽損模型的失真現象	59
圖 50	8KHz 取樣頻率下，聽損模型的失真現象	60
圖 51	聽力測試語料表	62



第一章 緒論

1.1 研究背景

現今的助聽器大多只考慮如何彌補因為耳朵受損或老化，而造成的聽力損失，例如 NAL prescription [1]。然而，除了不正常的最小可聽水平（minimum audible level）之外，聽力還受損造成了響度動態範圍縮小（reduced dynamic range/ loudness recruitment）[2]、窄頻時域封包解析度降低（reduced temporal resolution for narrow-band fluctuating stimuli）[3-6]、分頻解析度降低（reduced frequency selectivity）[7-12] 等等結果。一般來說，實作助聽器的演算法較少針對頻率上的受損現象做考量有二大原因：

- 根據測不準原理，我們只能根據應用，在時域及頻域上得到一個較佳的平衡，然而助聽器在設計上有著先天的限制—即時性，在此一情形下無法精確得到訊號的頻率成份，更何況要對頻率成份做複雜的處理；因此，一般看到的助聽器只能處理簡單的、在時域上可觀察到的問題：使用濾波器組（filter-bank）將訊號分成一個個寬窄不一的頻帶，並根據該頻帶聽力受損的程度將此分頻訊號乘上一個增益量[13]，來彌補聽力受損。
- 即使是使用濾波器組的觀念，因為訊號被耳朵接收的最後一級是基底膜，因此經過不同濾波器分別處理過後的訊號，最終還是要先合成為聲音訊號，再通過人耳接收聲音的管道來產生聽覺，因此相鄰濾波器的重疊，會造成相同的頻率成份經過不同的非線性處理，致使最後該聲音在基底膜上產生的波形與我們預期的有落差。

不論發展針對那一種聽損現象的演算法，要如何評估這個演算法的效能也是一個麻煩的問題。為了知道這個演算法實不實用，最直接的方法就是讓病人實際去聽，然而每個病人的聽力受損程度都不一樣，因此在評估之前，必須先對病人的受損情形做詳細的測量，並針對該受損情形調整演算法中相對應的參數；此外，通常聽力受損的症狀會伴隨著出現，因此若是演算法只針對其中某個原因而設計的話，即使對該症狀真的有效果，也可能因為其它的症狀而使得演算法的效果被忽略；最重要的是，在對演算法做細部調整時，沒有辦

法馬上請病患聽，來得知這樣的調整是否有效。

因此，很多學者開始投入研究，如何建立一個模擬聽障患者耳蝸的模型〔14-17〕，其目的是「將處理過後的聲音通過正常人耳，相當於未處理的聲音通過受損的耳蝸」，如此一來，若要評估演算法的效能，只要將處理過後的聲音經過這套系統後，讓正常人聽就可以了。

1.2 聽損現象簡介

1.2.1 最小可聽水平提升與響度聚集

一般聽損問題可直接反應在聲音的最小可聽水平上，最小可聽水平即「人耳可聽到該頻率的聲音所需的最小能量」，意即若聲音能量低於這個值，則人聽不到這個聲音。聽障患者最常產生的問題即是最小可聽水平的提升，然而此現象也常常伴隨其它種類的聽障問題，其一是響度聚集。

根據研究顯示，聽障患者的聽力區間是較小的，意即最小可聽水平提升，而最大可聽水平降低，然而聽障患者與正常人的最小可聽水平會對應到相同響度，最大可聽水平亦然，其結果是聽障患者聽覺響度—物理強度的斜率會比正常人的斜率來得大，如圖 1：

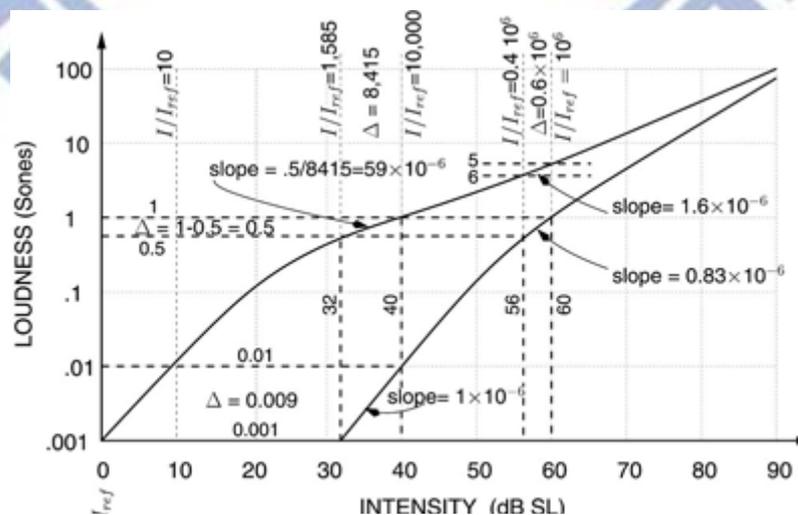


圖 1 不同的聽覺響度—物理強度的對應曲線

資料來源：MITCogNet

圖 1 中，下面的曲線表示聽損患者的響度-強度對應關係，當強度為 30dB 的時候響度相當於 0.001Sones (最小可聽水平提升)，然而當聲音的強度愈大時，下面的曲線與上面的曲線愈來愈重合，代表對於聽損患者而言，每單位強度增加的響度比正常情形還多，此一現象即是響度聚集；可能的影響是聲音響度變化放大，造成聽障患者的不適，無法對該句話做正確的解讀。除此之外，不一定在所有的聽力區間之內，都有響度聚集的現象：研究顯示，當聲音的響度很高的時候，正常人與聽障患者對應的響度是一模一樣的 [18-19]。

1.2.2 分頻解析度降低

除了以上的問題之外，聽損患者常會發生的問題是分頻解析度降低，發生的原因可能是因為耳蝸的老化、或是耳朵受損，使得內毛細胞 (inner hair cell) 對於頻率的敏感度下降。舉例來說，原本只對 1800Hz 到 2100Hz 有反應的內毛細胞，因為敏感度降低的關係，使其對 1400Hz 到 2400Hz 頻率的能量都有反應，這種現象以工程的角度來看，就是人耳相對應的聽覺濾波器變寬了，如圖 2 所示；如此一來造成的影響是，原本不屬於此頻帶的聲音也會被包含進來，使得聲音變得較為吵雜；此外，由於泛頻成份變多的關係，也會使得接收到的聲音音色改變。

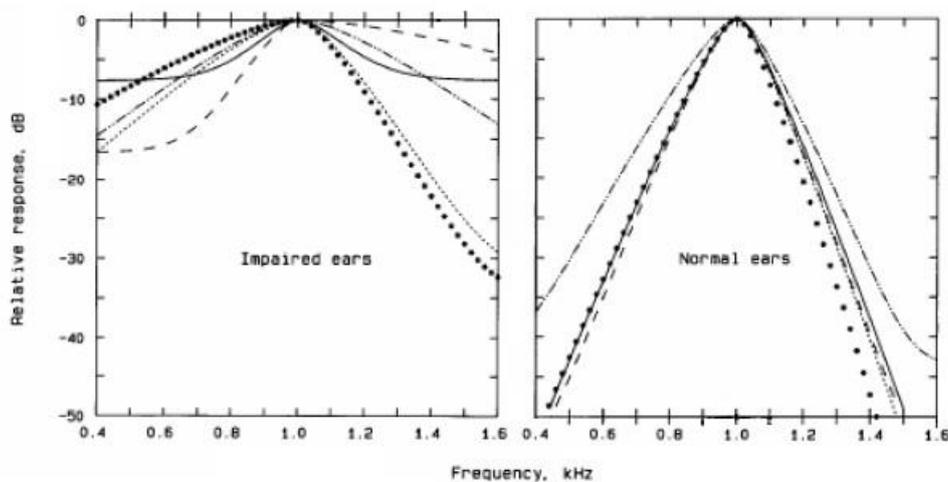


圖 2 受損患者與正常人的聽覺濾波器比較

資料來源：[20]

1.3 研究方法

即使有學者發展出受損耳蝸模型，然而這些研究並不是特別針對助聽器而發展的；換句話說，這些系統只有模擬質的變化，並沒有對量的精確性做較為深入的研究與討論。舉例來說，一般短時傅利葉轉換（Short Time Fourier Transform; STFT）只有強調強度（magnitude）的變化，並在處理訊號時保留原訊號的相位，加上合成時使用 OLA [21] 的影響，使得最後得到的訊號與我們當初模擬的程度有差別 [14][22]；若是用小波轉換（wavelet transform）的方式，則因為濾波器組的相互重疊而造成干擾，因此，若是要模擬出完全符合某位患者的耳蝸受損情形，則必須做更為全面的分析。

本論文將以 Moore 的研究團隊所提出的幾個聽損耳蝸模型為基礎，深入研究比較其優缺點後，提出改良後的響度聚集與分頻解析度降低的模擬方法，建構出一套聽損耳蝸模型；目的不只是为了證明這些聽損現象確實會影響患者的辨識率，更希望藉此準確模擬聽損患者的受損程度，以用來調整演算法的參數，並評估其效能。

除此之外，由於本論文特別注重分頻解析度降低的聽損現象，因此也對一些學者提出的改善分頻解析度降低的演算法 [23-24] 做深入的研究與討論，並在這些基礎之上加以延伸，希望可以得到更好的結果。

1.4 章節大綱

本論文的各章內容如下：第二章為生理聽覺現象與特性簡介，並介紹兩種主要的聲音訊號處理方法—濾波器組與短時傅利葉轉換；第三章介紹 Moore 的研究團隊所提出的幾個模擬聽損耳蝸的模型，並分析其優缺點，以及加上 LSEE-MSTFTM（Least Square Error Estimate from the Modified STFT Magnitude）[22] 修正相位的結果；第四章介紹常見的補償分頻解析度降低的演算法，除了分析優缺點並加以改良之外，也延伸了 Moore 團隊的想法、嘗試提出新的演算法；第五章則是延伸 Moore 團隊的聽損耳蝸模型，加入更多的考量因素之後提出新的模擬方法，並進行辨識率的實驗，來評估這個模型的效能。

第二章 感知訊號處理基礎

2.1 生理聽覺現象與特性

2.1.1 聽覺的產生

圖 3 為耳朵的構造：

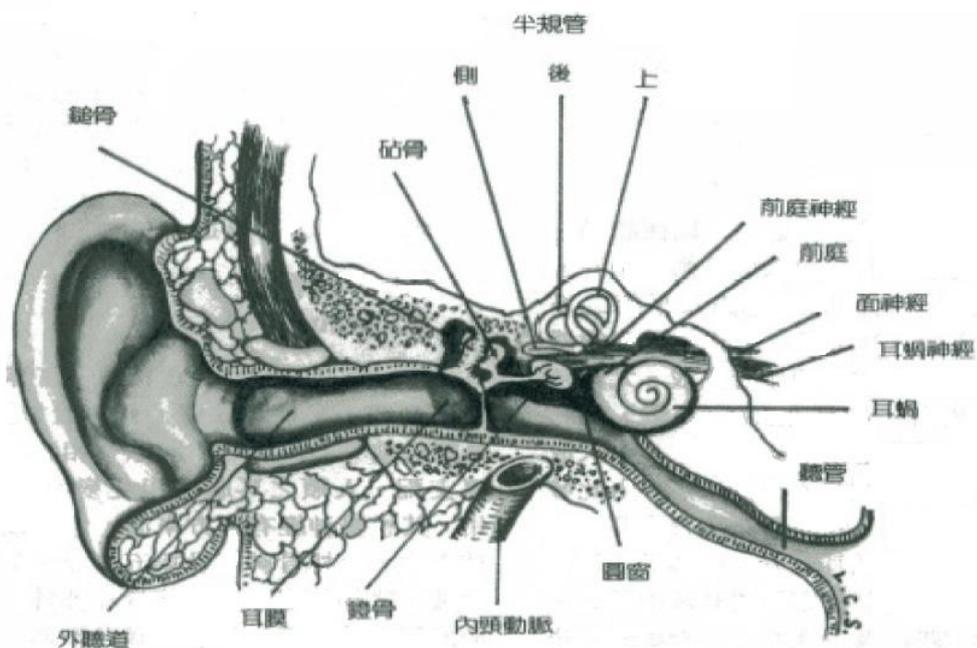


圖 3 耳蝸構造圖

資料來源：〔20〕

空氣中的聲波，經由外耳的耳廓收集之後，通過外聽道振動鼓膜（耳膜）。中耳則是由三小聽骨所構成，前後兩端分別連接與外耳相連的鼓膜，以及與內耳連接的圓窗；聲波在鼓膜產生了振動之後，藉由三小聽骨將聲音轉換為機械能，並傳送到內耳的圓窗。藉由三小聽骨擠壓圓窗內部的液體，使得耳蝸中的基底膜產生了行進波，而基底膜上的柯替式器（organ of Corti; 圖 4）即為聲音的主要接收器；其上分佈的內毛細胞（inner hair cell）與外毛細胞（outer hair cell）的放電程度，與覆膜的拉扯程度、以及組織液的流速快慢成正向關係：當強度大於某個臨界值後，開始有電位差的產生，藉此將行進波的資訊轉換為電

能，並經由其後所連接的聽神經，將聲音傳到大腦，以進行更高階的處理。

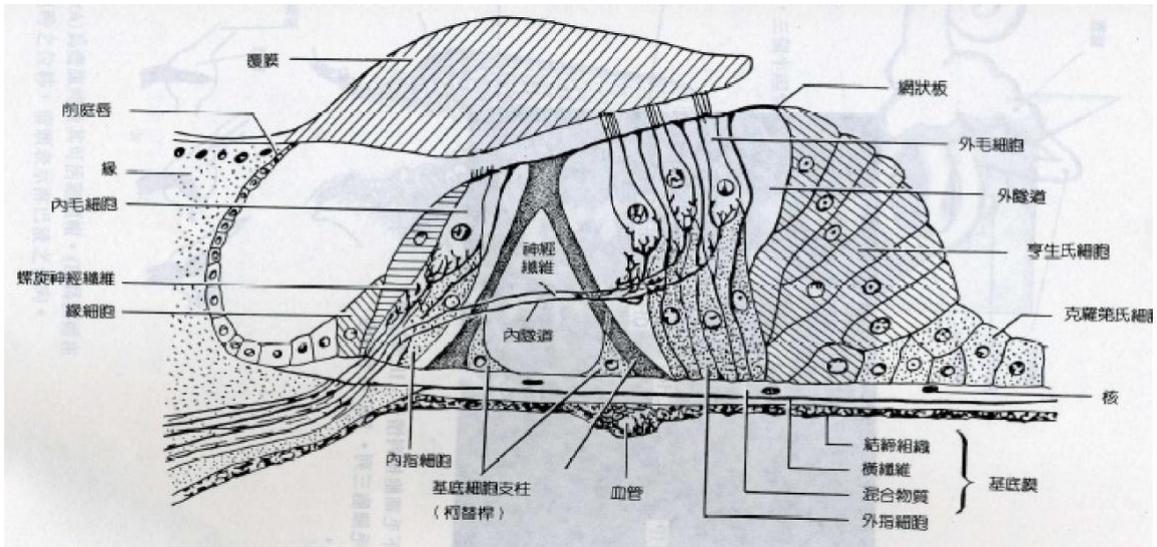


圖 4 柯替式器
資料來源：〔20〕

2.1.2 行進波在基底膜上的特性

基底膜上的行進波如圖 5 所示，其主要性質說明如下：

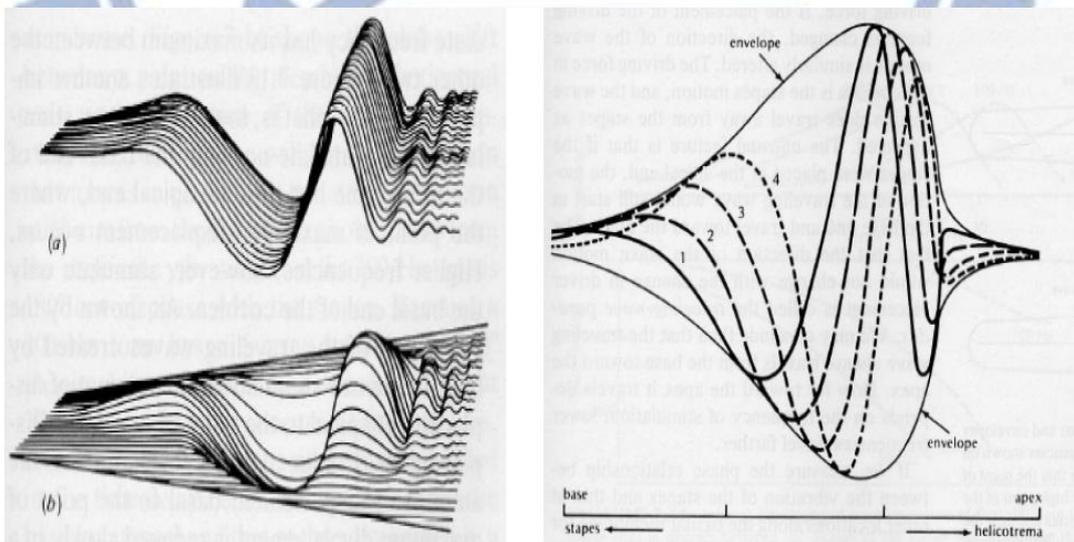


圖 5 行進波在基底膜的振動模式
資料來源：〔20〕

- 不同頻率的聲音會在基底膜不同的地方發生共振，愈高頻的聲音在愈接近圓窗，也就是 base，產生共振；而愈低頻的聲音則在愈深處（apex）產生共振。其現象大概可以用基底膜的質地來解釋：在愈接近 base 的地方，基底膜的質地較為堅硬，其共振頻率較高；而愈接近 apex 的地方，其共振頻率較低。
- 不同頻率的聲音在基底膜上的共振位置是呈現對數分佈的，如圖 6 所示；在圖中的右上角可以發現，相差二倍頻率的聲音，其在基底膜上的共振位置是等距的，而這個現象也會在接下來的聽覺模型中扮演重要的角色。

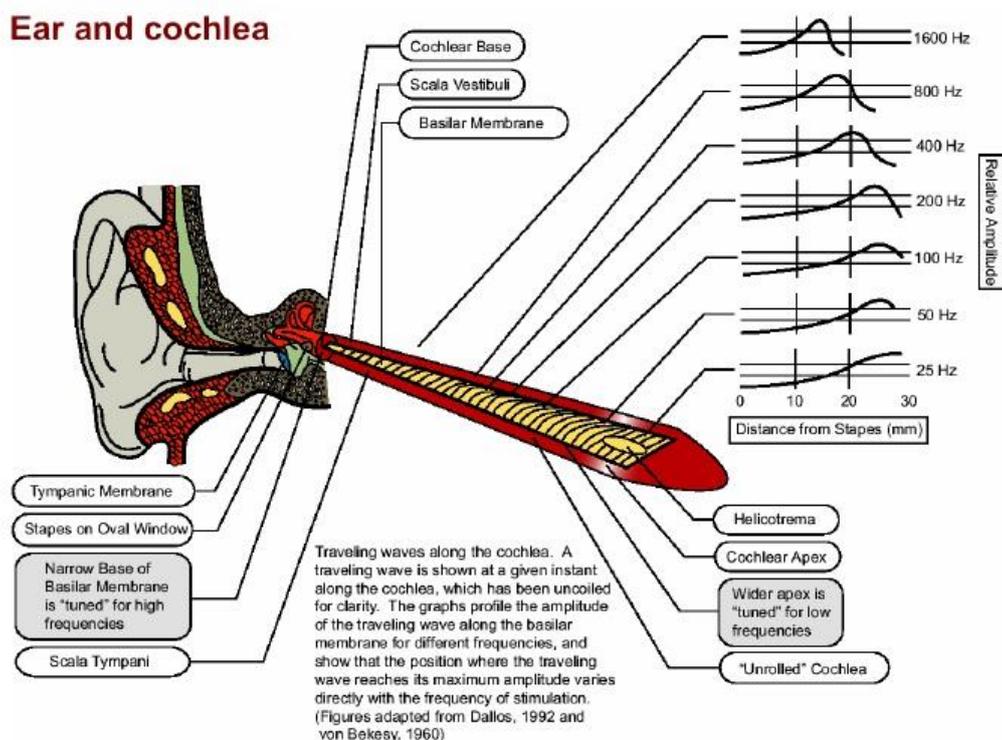


圖 6 相差倍頻單音在基底膜的共振位置示意圖

資料來源：[20]

- 由圖 5 行進波的圖中可以發現，以基底膜產生最強響應的地方為原點，往左右兩邊均可以視為 cosine exponential 的衰減函數，並且往高頻的衰減速度較慢，往低頻的衰減速度較快。當一個聲音為兩個相近頻率的單音所組成時，若是其中一個單音所產生的振動模式幾乎被另一個單音的振動模式所包含時，則會產生遮蔽效應(masking effect)，如圖七所示：(a) 例，當 A 與 B 的距離差很遠時，遮蔽效應不會發生；(b) 例，雖然 A 與 B 的頻率很接近，然而因為 A 的能量與 B 相近，因此 A 也不會被遮蔽掉；(c)

例，A 的頻率與 B 很接近，且能量較弱，因此會被 B 遮蔽掉；(d) 例，雖然 B 頻率與 A 相近，且能量較弱，但因為 B 的頻率較低，而基底膜的振動模式在過了最大值之後就會迅速衰減，因此 B 也不會被 A 遮蔽。

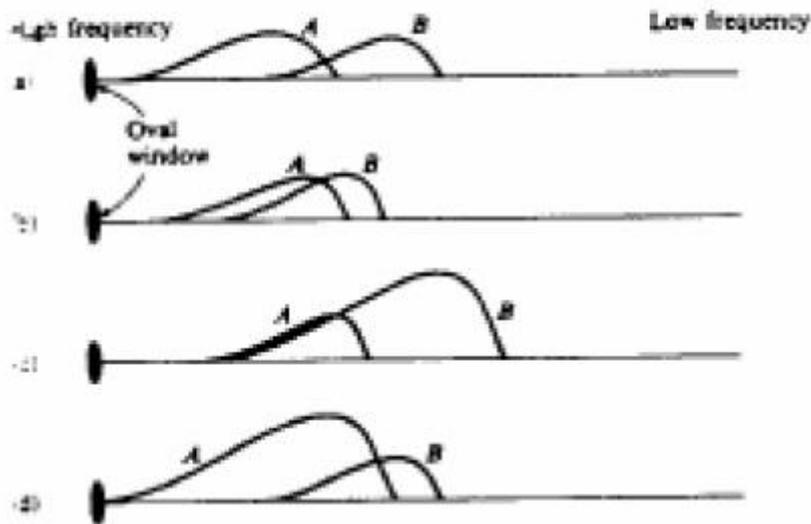


圖 7 基底膜上的響應示意圖

資料來源：〔25〕

由以上的例子我們可以知道，兩相近頻率的單音，若高頻的單音能量較弱的话，很有可能會被低頻的單音遮蔽。根據以上三點，聲音在基底膜上產生的波形，幾乎可以視為是聲音的頻譜，而這種以基底膜來分析聲音頻率成份的方法稱為位置編碼 (place coding)。

2.1.3 濾波器組與激發模式 (excitation pattern)

1. 濾波器組

Fletcher〔26〕發表的聽覺接收相關論文中，提到了有關濾波器組的概念，奠定了工程分析上，以濾波器組來模擬人耳接收聲音的方式。

從聲音在基底膜上的接收來看，當一個單音進來的時候，在與其頻率相對應的位置會有較大的響應，可以視為這附近的內毛細胞，對於該頻率的聲音有較強的接收能力；而距離較遠的內毛細胞，還是會對這個頻率的聲音有反應，並不會完全沒有反應；換言之，某一個位置的內毛細胞，不只會對其共振頻率的聲音有反應，對於其臨近頻率的聲音也會有反應，並且是呈連續變化的。

根據以上的現象，我們可以把這群內毛細胞當作是一個濾波器，或是一個臨界頻帶 (critical band)，此濾波器的頻率響應即為這群內毛細胞對不同頻率的反應程度。然而基底膜上的內毛細胞是連續的，而其對於不同頻率聲音的接收能力也是呈現連續變化，如圖 8 所示：

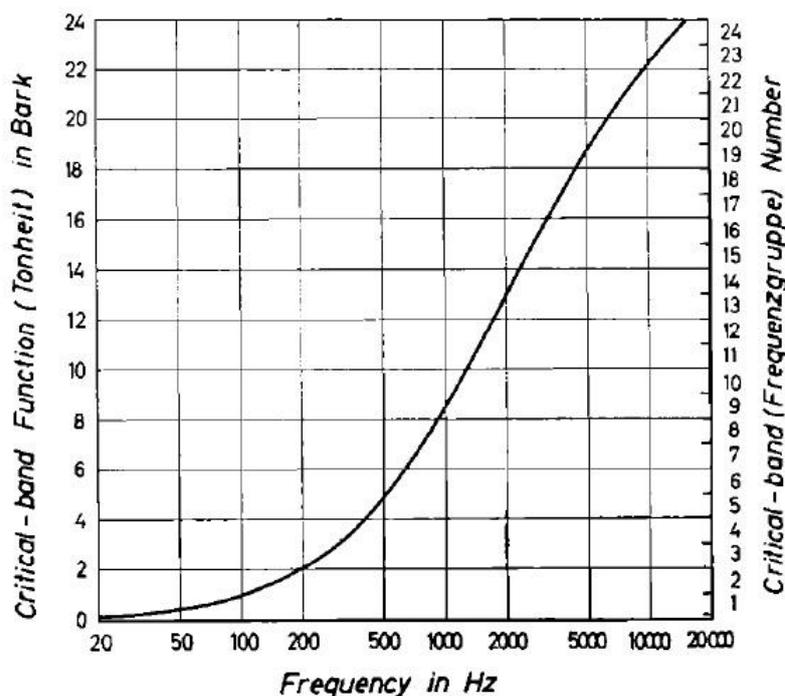


圖 8 臨界頻帶與頻率的關係，縱軸為 [27] 中提出的分頻方式
資料來源：[27]

圖 8 中代表的是某一個臨界頻帶應該包含的頻率範圍；然而從工程的觀點來看，我們在處理聲音訊號的必須面臨的問題是：要取幾個臨界頻帶？中心頻率要如何選取？

一般而言的考量是，不管選取幾個臨界頻帶，我們希望所有臨界頻帶的頻率響應總和為 1，也就是說我們希望每個頻率成份被接收到的能量與原訊號的頻譜是相同的；至於濾波器個數，根據不同的目的，我們會選取不同個數的頻帶，基於聲音訊號大多都以 8Kz 或是 16Kz 來處理，因此一般認為 18 個臨界頻帶是適當的。最後我們只要根據實驗的目的或限制，選取適當的濾波器個數，再根據第一點中提到的原則，就能得到濾波器組中所有濾波器的中心頻率了。

2. 激發模式

一般而言工程學上討論聲音在基底膜上的響應時，不會直接用物理上的行進波來解釋，而是用激發模式來解釋，其原因是因為行進波的產生包含了高度的非線性現象，必須透過較為複雜與精細的實驗才能確切得知其成因，因此目前只能就量測到的結果，將行進波以較為簡單的型式表現出來，並進行進一步的處理。

通常我們透過濾波器組來得到聲音的激發模式，而濾波器組的形狀則可由 notch-noise test [26][28] 來得知；假設我們得到的濾波器組如圖 9 左所示：

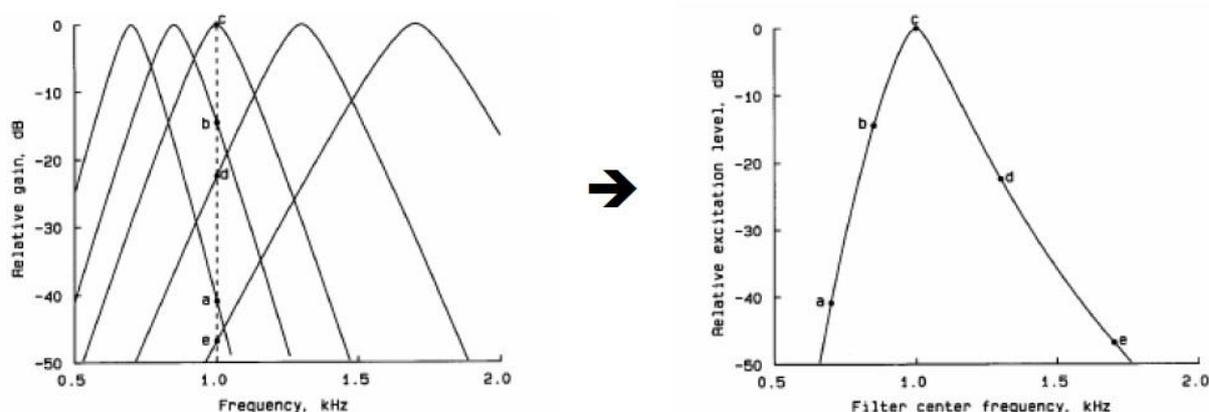


圖 9 由濾波器形狀計算激發模式

資料來源：[20]

當我們想要知道 1KHz 單音的激發模式時，我們觀察 1KHz 單音在不同的濾波器中的能量大小，並將此值以濾波器組的中心頻率做圖，如此一來我們可以得到右邊的圖，解釋為「中心頻率不同的聽覺細胞對於 1KHz 單音的反應程度」，也就是 1KHz 單音的激發模式 (excitation pattern)。

2.1.4 響度 (loudness)

1. 強度 (intensity) 與響度

人耳感覺到聲音的強度稱之為響度，響度與聲音物理強度的關係如圖 10：

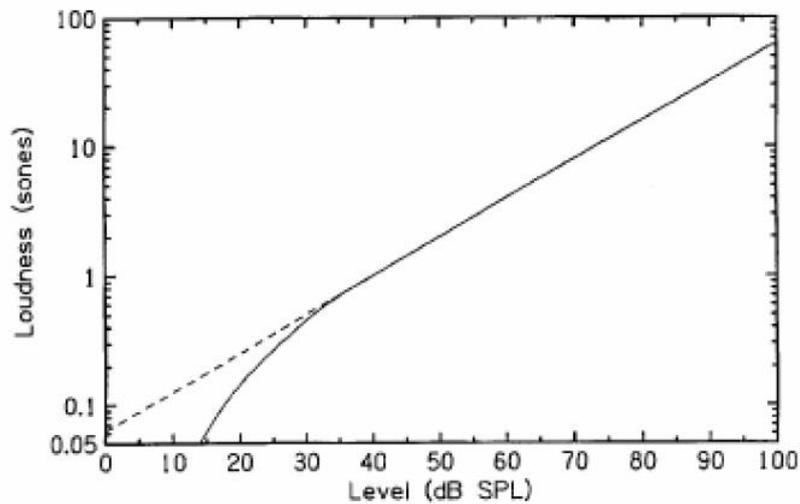


圖 10 實線為測量 1KHz 單音的強度-響度對應關係，虛線為 [30] 的公式
資料來源：[20]

特別注意圖 10 中的縱軸是 log scale。因此我們可以說，當聲音的強度（intensity）愈大的時候，相同強度的差距對應到響度的變化愈來愈小，這可以解釋成為人的自我保護機制，若是人耳沒有這樣的保護機制的話，聽覺細胞可能因為過度激發而提早老化或死亡。

Fletcher [29] 將強度與響度的關係公式化如下：

$$\text{loudness} = \text{level}^{1/3} \quad (1)$$

而 Steven 也以實驗數據推算出該指數關係中的指數約為 0.3 [30]。

2. 最小可聽水平與最大可聽水平

人耳對於聲音的敏感度隨著頻率變化而不同；例如：人耳聽不到 20Hz 以下的聲音，也聽不到 20KHz 以上的聲音，這是頻率與響度交互作用而得到的結果，說明如下：對於某一個頻率的單音，我們漸漸增強它的能量，當我們突然可以聽到這個單音的瞬間，我們稱這個時候的能量稱為最小可聽水平（minimum audible level），也就是人耳聽到這個頻率的聲音所需的最小能量；接著，我們繼續增強它的能量，一直大到我們無法忍受的程度，此時的聲音已經會對我們的耳朵造成破壞性的影響，其能量大小稱為最大可聽水平

(maximum audible level)。

圖 11 為人耳不同頻率所對應的最小可聽水平與最大可聽水平的值。圖 11 中，最高可聽水平幾乎是定值，然而最低可聽水平在小於 100Hz 或是大於 10KHz 時，卻會大幅的升高，並且在約 20Hz 與 20KHz 的時候與最高可聽水平重合，因此我們幾乎聽不到超出這個範圍的聲音，這就是我們一般所熟知的結果。

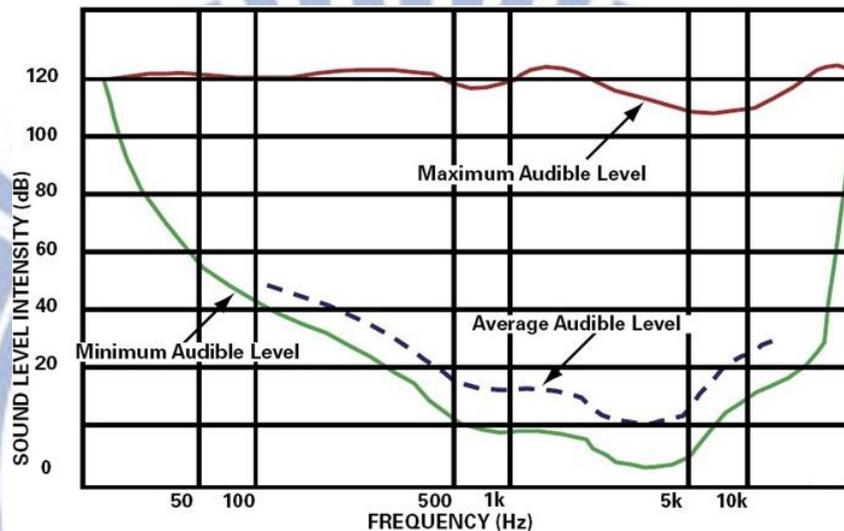


圖 11 人耳可聽到的聲音頻率及能量範圍

資料來源：Mechanical Engineering

3. 等響度曲線 (equal loudness curve)

等響度曲線是以 1KHz 的聲音在一固定強度下的響度為基準，調整其它頻率聲音的能量，使其響度與 1KHz 聲音響度相同，並紀錄其強度值的曲線，如圖 12；在同一條曲線上的值，對於人耳來說響度都是一樣的，而最上面跟最下面的等響度曲線所包含的範圍，則是人耳可聽範圍，相當於圖 11。

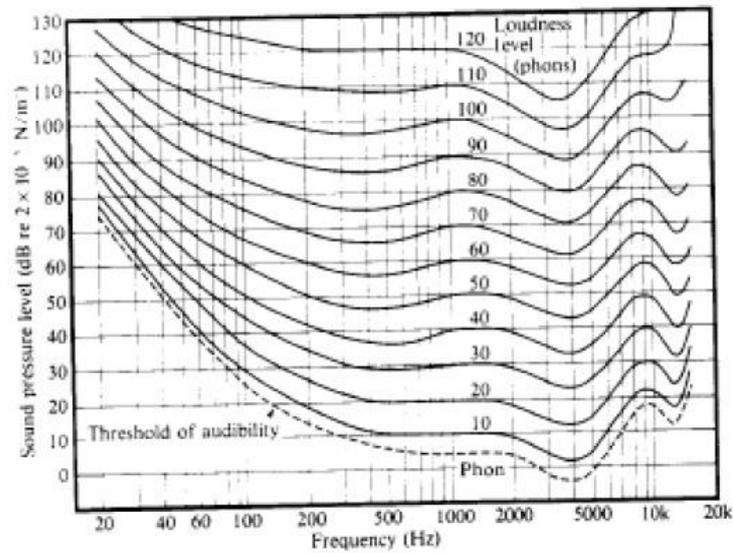


圖 12 以 1KHz 為基準的等響度曲線

資料來源：[20]

2.2 短時傅利葉轉換 (STFT)

雖然傅利葉轉換 (Fourier Transform) 可以讓我們清楚的得知一段訊號的頻譜成份究竟為何，然而在生活中的訊號與系統，卻是無法直接用傅利葉轉換來分析；其主要原因是：傅利葉轉換假定該段訊號為非時變的，換句話說，該訊號的頻譜組成為固定不變的；然而這樣的假設並不符合實際情況，因為現實生活中的訊號隨時都在改變。

若是直接以傅利葉轉換來分析時變訊號，會無法得知某個頻率真正存在的時間點。以圖 13 為例：

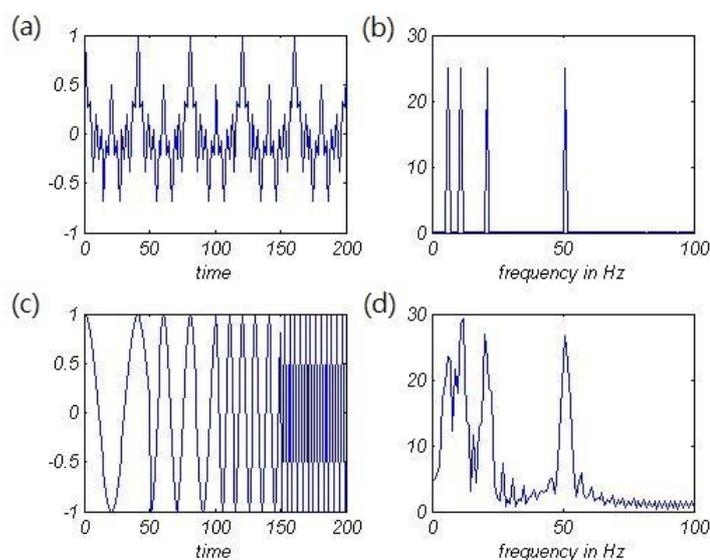


圖 13 非時變訊號與時變訊號之時-頻比較圖

圖 13 中，(a) 與 (b) 是時變訊號的時間訊號及頻譜圖，是由 5Hz、10Hz、20Hz 和 50Hz 的弦波相加混合而成，其頻譜圖也顯示出這樣的特性。而 (c) 小圖中、前 50 點為 5Hz 的弦波，50~100 點為 10Hz 的弦波、100~150 點為 20Hz 的弦波，而最後 50 點為 50Hz 的弦波；觀察 (d) 小圖，似乎可以解釋為：在 5Hz、10Hz、20Hz 和 50Hz 處有最強的能量，而最強的能量稍微散開、且頻譜有高频成份產生是因為在第 50、100、150 點的地方有不平滑而造成的現象。

然而以上的解釋並不合理，因為訊號是時變的，我們無法從頻譜圖中還原時域的圖形；換句話說，我們無法準確說出那一個頻率成份存在那一段時間之內；因此，用傅利葉轉換來分析時變訊號是不合理的，因為頻譜不具代表性。

因此，為了能夠使用 FFT (Fast Fourier Transform) 直接對時變訊號做頻譜上的處理，我們把連續訊號分為一個區塊一個區塊，並假設在該區段之內的訊號是非時變的，再進行傅利葉分析，也就是所謂的短時訊號傅利葉轉換 (STFT)；然而，以 STFT 來分析訊號與合成經過頻譜處理的訊號時，有非常多需要細部注意的事項，以下一一分析說明之。

2.2.1 訊框長度、訊框位移、分析視窗

因為我們將一段段分解後的訊號設為非時變訊號，因此我們必須考慮的是：訊號在什麼狀況之下會近似於非時變訊號？就語音訊號而言，大致可以分為母音（有聲）與子音（無聲）部份；一般我們將子音的產生模擬為雜訊與口腔形狀頻率響應的旋積（convolution）；而母音則為脈波串列（pulse train）與口腔形狀頻率響應的旋積；以此觀念分析，則我們儘需考慮母音，一般而言我們認為 20~30ms 之內的範圍，母音的頻譜是穩定不變動的。

接下來我們可以把原來的連續訊號，視為一個個乘上矩形視窗（rectangular window）後的短時訊號的加總，而此短時訊號是非時變訊號。下一步則考慮此一方形視窗是否會使短時訊號的頻譜與原訊號的頻譜不相同？

一般來說，當我們將一個訊號乘上矩形視窗，相當於原訊號頻譜與該視窗的頻譜響應做旋積；因為矩形視窗的頻譜是 sinc 函數，因此原訊號頻譜會被模糊化，這樣子的現象也可以測不準原理說明：因為時域訊號的成份少了，因此解析出來的頻譜成份自然就較不準確。

根據以上論點，則乘上矩形視窗是一個很不好的選擇，因為其頻譜響應的影響範圍很廣。然而我們必須要注意的是，以上論點的基本假設是原訊號的頻譜組成是固定的，而這個假設並不成立，換句話說，在訊號為非時變的最大時間長度內，本身的頻譜就是被模糊過後的頻譜，也就是說，訊號的最小頻率解析度是固定的，因此，該矩形視窗則不對該訊號有任何影響。

然而以上的論點卻又過於果斷，因為每一個框架訊號並不是毫不相關的，前一個訊框內的後面幾個點，與後一個訊框內的前面幾個點，是連續且可以視為非時變訊號的，因此若我們只考慮該框架裡的訊號頻譜，又沒辦法非常準確呈現頻譜的組成；直覺思考上，若我們將訊號取一個方形視窗，則在邊界點的值會不平滑，而這個不平滑的現象會造成訊號的頻譜擴散，以及高頻雜訊的產生。

因此，為了使訊框在邊界上也可以連續，選擇視窗則變成一個很重要的議題；一般來說，我們希望視窗的頻譜愈像脈衝愈好，也就是指主頻帶的寬度愈窄、副頻帶的高度愈小愈好，而漢明視窗（hamming window）是很常見的一個選擇（圖 14）。

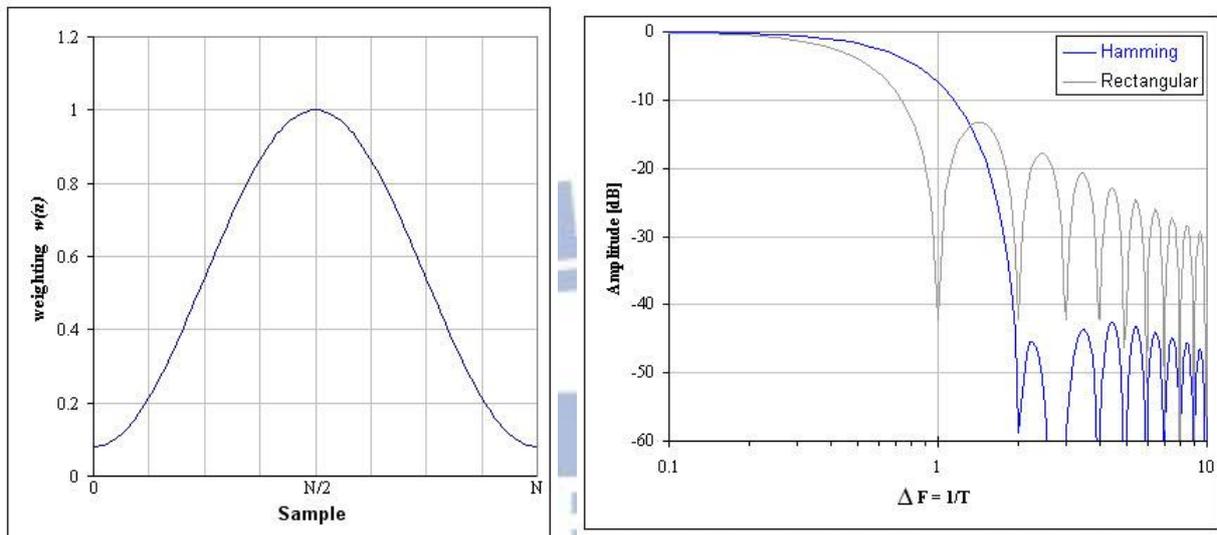


圖 14 左圖為漢明視窗的時間訊號；右圖為漢明視窗與矩形視窗的頻譜響應比較
資料來源：diracdelta.co.uk

舉例來說，圖 15 (a) 是一個從無窮長的弦波取下來的一段訊號，可視為是弦波與矩形視窗相乘而得，(b) 則為其頻譜；從 (b) 圖中得知，因為在邊界的不平滑現象，使得頻譜散開了。而 (c)、(d) 圖則是弦波與漢明視窗相乘而得的訊號及其頻譜；因為漢明視窗的關係，使得訊號在邊界處連續（均為零），因此頻譜也較為集中。

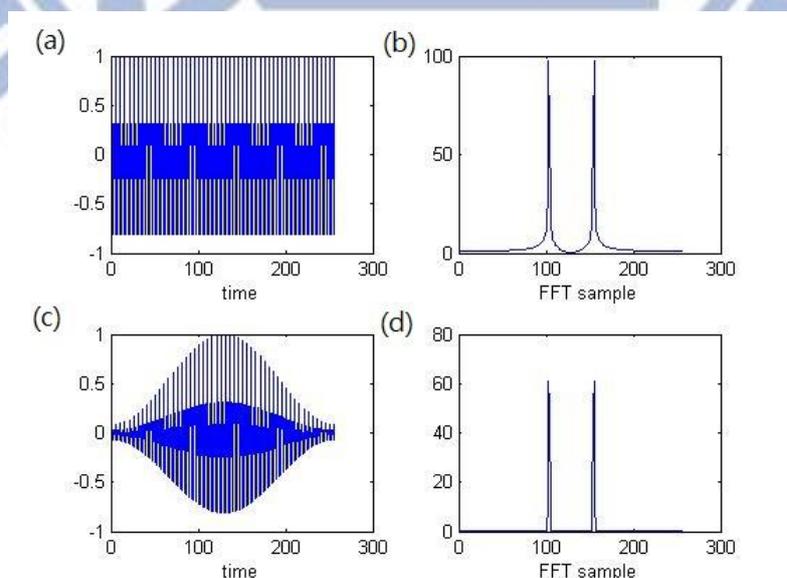


圖 15 矩形視窗與漢明視窗的結果比較圖

除了選取一個好的視窗之外，在分析頻譜的時候，我們也會讓每一段框架訊號有重疊；我們認為這樣子得到的 STFT 是較為連續變化的，而相關參數則是框架位移。

2.2.2 快速傅利葉轉換點數 (FFT size)

我們將連續訊號分解成一個個的框架訊號，並乘上適當的視窗之後，接下來要開始進行傅利葉的分析了，這時候我們要考慮的是，到底要做幾點的 FFT？

從分析的角度來看，如果想要看到比較細部的頻譜變化，則點數愈多愈好，其原因是 DFT (Discrete FT) 的分析是從 DFS (Discrete Fourier Series) 而來，而 DFS 是針對週期性訊號所做的分析，是我們想要分析的訊號的連續頻譜取樣而來的；當點數愈多的時候，頻率軸的取樣率高，因此我們可以看到較為細緻的頻譜。

若是以訊號合成的觀點來看，則根據 [31] 第八章所提到的，連續訊號與一個系統響應做旋積時，我們可以把訊號框架化成有限點訊號，分別與此系統響應做旋積之後，經過 OLA (overlap and add) 或 OLS (overlap and save) 合成訊號，得到與連續訊號直接做旋積相同的結果。而框架訊號與系統響應做旋積的這個步驟，我們可以利用 FFT 將其轉換到頻域相乘之後再轉回來，如此一來可以降低運算量。圖 16 即為 OLA 的分解步驟，左圖是框架訊號補零之後的結果；右圖是分段與系統響應做旋積的結果；而右下圖則是重疊相加後的結果。

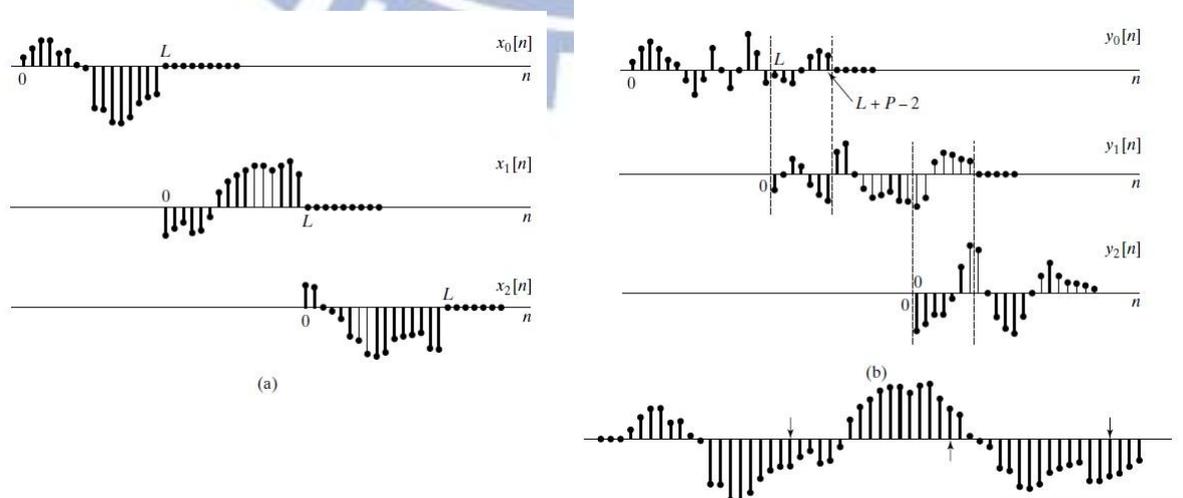


圖 16 OLA 流程示意圖

資料來源：〔31〕

特別注意的是 DFT 架構中，頻域軸的相乘相當於時域軸做迴旋積（circular convolution），而迴旋積與旋積只有在兩個訊號的週期均大或等於兩個訊號旋積結果的長度時才相等，也就是說，我們必須把框架訊號補上足夠數目的零點才行。

連線性非時變系統的旋積，我們都必須補零點，更何況是直接要在頻域上做一些複雜的、非線性的處理—Kulkarni 與 Pandey 在〔23〕提到，雖然框架訊號只有 200 點，然而在進行頻譜上的處理時，大約需要 1024 點的 FFT 才会有比較好的結果—我們認為這些非線性的運算，相當於時間軸為很長的脈衝響應，所以 FFT 的 size 會比訊號框架長度還要多得多。

2.2.3 OLA 與合成視窗

將訊號頻譜做了處理之後，我們以 IFFT 將頻譜轉回時間訊號，並將這些框架訊號加總成為完整的訊號，這時候我們必須注意的是，雖然原框架訊號乘上分析視窗後的頻譜是較為正確的，然而在時間軸上來看訊號強度卻是忽大忽小；除此之外，每一個框架訊號是彼此重疊的，那麼應該要怎麼樣合成訊號才是正確的呢？

直覺上來考慮，每一個時間點被計算的次數要一樣多，因此有了傳統的 OLA 演算法〔21〕：

$$x(n) = \frac{\sum_{m=-\infty}^{\infty} y_w(mS, n)}{\sum_{m=-\infty}^{\infty} w(mS - n)} \quad (2)$$

其中分母項為分析視窗的加總，也就是對每一點訊號的權重做正規化，而分子則是經過處理後的每一個框架訊號的加總；另外也有權重 OLA 演算法〔32〕：

$$x(n) = \sum_{m=-\infty}^{\infty} f(mS - n) y_w(mS, n) \quad (3)$$

其中 f 為一個用於合成的函數；更有以數學推導出來的，使修改能量頻譜的框架訊號最接近時域上存在的訊號的 LSEE-MSTFT (Least Squares Error Estimation-Modified STFT)

[22]:

$$x(n) = \frac{\sum_{m=-\infty}^{\infty} w(mS - n) y_w(mS, n)}{\sum_{m=-\infty}^{\infty} w^2(mS - n)} \quad (4)$$

其中 w 為分析視窗。以上的幾種方法都是由正規化的概念推導出來的，然而一般實作上為了讓計算方便，我們會適當設計框架長度與框架位移，使正規化的參數變成一個常數，如此一來即可簡化我們的系統運算量；例如當框架長度:框架位移=2:1 時，漢明視窗代入[21]的演算法中分母項為一個常數，則演算法減化為一個簡單相加的式子：

$$x(n) = \sum_{m=-\infty}^{\infty} y_w(mS, n) \quad (5)$$

2.3 濾波器組的選擇

在前面的章節中有提到，基底膜上所對應的共振頻率位置是呈現對數分佈的，因此當我們要選擇模擬人耳的濾波器組時，第一個選擇就是頻率響應在對數軸上呈等距分佈的濾波器組，而此想法也剛好與小波轉換 (wavelet transform) 不謀而合，因此以下將簡單介紹小波轉換的特性及實用性。

2.3.1 小波轉換

雖然傅利葉轉換的理論可以幫助我們了解時域和頻域相互的關係，然而在工程觀點上傅利葉轉換是不好的，其原因是因為傅利葉轉換的基底是無窮長的正弦波或是餘弦波，這在工程分析上是不存在的，因此傅利葉轉換無法有效應用在工程問題上。

除此之外，在現實生活中的很多場合，我們不需要看到非常細緻的頻譜；一般來說，低頻的成份比較豐富，在頻譜上的變化多，然而其在時間軸上的變化較慢；而高頻的成份比較少，在頻譜上的變化少，然而其在時間軸上的變化則較快；因此，我們希望在分析訊

號的時候也能以這樣子的解析度來分析，來達到時域與頻域的最佳分配。如圖 17 所示，左圖是傅利葉轉換的時-頻關係圖，其在時間軸上只有固定的、較不理想的解析度；而右圖則是我們希望達到的解析度。

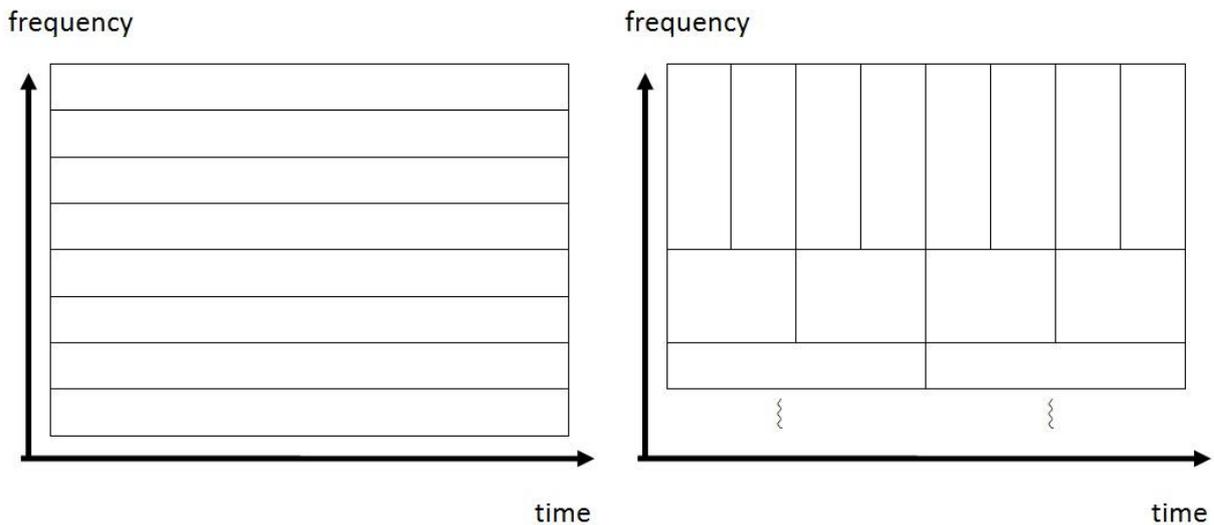


圖 17 傅利葉轉換與離散小波轉換的時-頻比較圖

基於以上的幾個原因，我們開始尋找有限點的基底群，並且其時-頻解析度是我們想要的，於是便有了小波轉換。小波轉換的基底群是由一個母基底（mother basis）而來，其離散形式如下：

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - k2^j}{2^j}\right) \quad (6)$$

其中 $\psi(t)$ 為母基底， j 為 scaling factor，用來決定該小波所涵蓋的頻率範圍； k 為 translation factor，就是時間軸上的位移；特別注意以上的例子中我們提到的離散是指 j 與 k 的值是離散的，並不是指離散的訊號。進一步用傅利葉轉換公式來看母基底與子基底的關係，如下公式：

$$\text{Fourier}\{f(at)\} = \frac{1}{|a|} F\left(\frac{\omega}{a}\right) \quad (7)$$

F 是 f 函數經過傅利葉轉換的頻譜。由此我們得知，當有一個子基底的 j 為 1 時，其頻率範圍約為其母基底的二分之一倍，而脈衝響應則為母基底的兩倍，這樣子的特性也就符合了圖 17 右的時-頻解析度。

真正離散訊號的小波轉換有一個很好的特性：我們只要把原聲音訊號通過以這些小波為轉換函數（transfer function）的濾波器，得到的輸出訊號就是小波轉換的係數。此外，因為小波都是帶通濾波器（band-pass filter），因此定義相對應的縮放函數（scaling function） $\Phi(t)$ 來包含剩餘的頻帶，並進一步與降頻（down-sampling）結合，來減少運算量 [33-34]。

一般我們用小波轉換來分析訊號，會將訊號通過一小波函數與縮放函數，再將通過縮放函數的訊號通過下一級的小波函數與縮放函數（圖 18），一直分析到我們需要的低頻解析度為止。

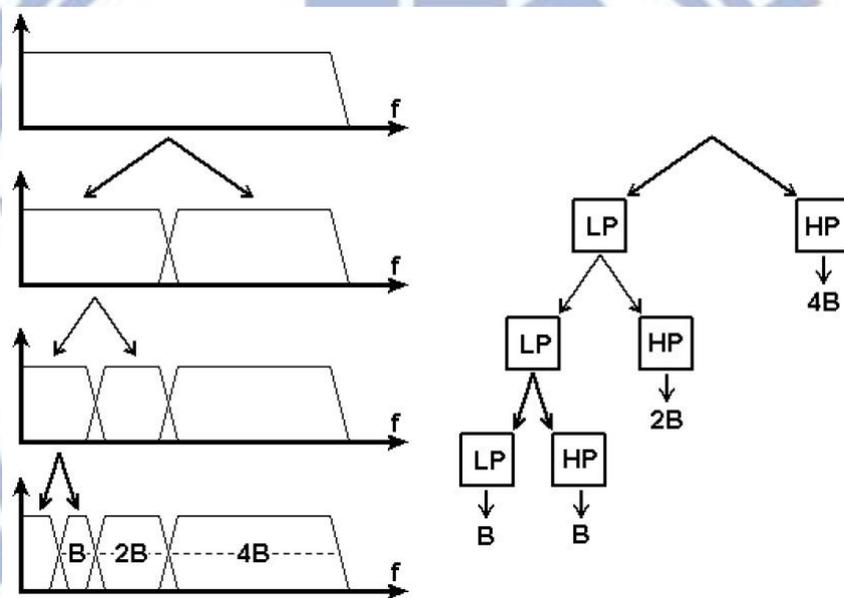


圖 18 以遞迴濾波器組分割頻譜
資料來源：[32]

2.3.2 聽覺濾波器組（auditory filter bank）

最常被用於模擬人耳的濾波器組是 Gammatone filter [35-36]，此濾波器組的小波函數為 cosine exponential 函數；然而因為 Gammatone filter 是左右對稱的，而人耳的濾波器組在某些情形之下會不對稱，因此延伸出 Gammachirp filter [37]；該濾波器組的特性是當訊號能量不同的時候，濾波器的對稱性會隨之改變，而這樣子的特性也較為符合生理實驗的

結果。

在我們的實驗中，使用的濾波器組是〔38〕中人耳聽覺模型中使用的濾波器組，一共有 128 個濾波器（原本有 129 個，但因為模擬了遮蔽效應的結果，因此前後交互作用的結果只有 128 個輸出）。以 128 個濾波器取代 129 個濾波器的原因，是因為從生物實驗的結果來說，大約只需以 18 個濾波器包含所有頻譜，就可以解釋這些現象；然而若是要針對這些經過濾波的副頻帶訊號（sub-band signal）做更進一步的處理及分析，從工程學的角度，則需要較高的解析度才行。

〔38〕中的濾波器組，將聲音從耳廓收集開始，一直到內耳聽覺細胞的接收，都做了非常詳細的模擬；在本論文中，我們僅應用了其中與遮蔽效應相關的觀念以及實作上的延遲補償，以下將分別說明之。

1. 遮蔽效應（masking effect）

我們假設原始的 129 個濾波器是沒有遮蔽效應的（圖 19）。為了得到有遮蔽效應的濾波器組，基於前面章節所探討到的現象——低頻訊號較容易遮蔽高頻的訊號——我們僅考慮單邊的遮蔽效應；在求出 129 個濾波器組的頻率響應之後，將前一個頻帶與後一個頻帶的頻率響應兩兩相減，得到包含遮蔽效應的 128 個濾波器（圖 20）。

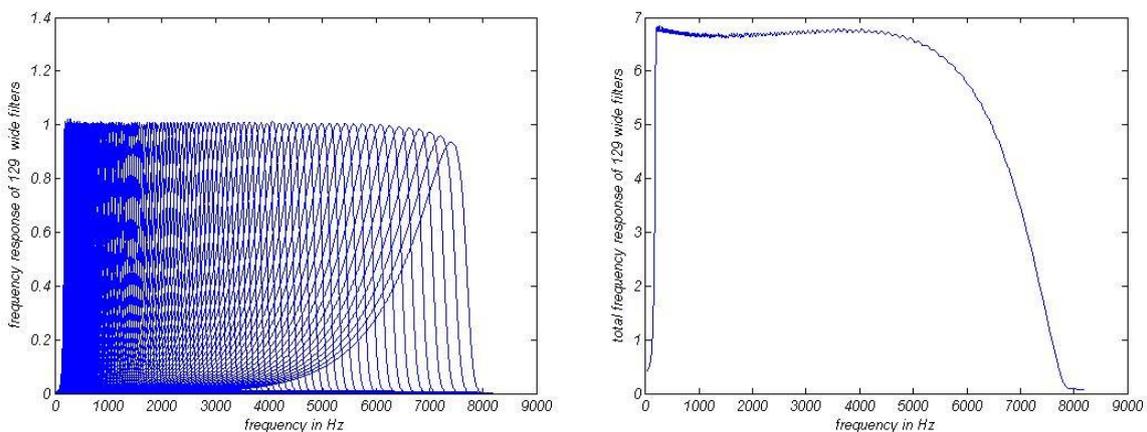


圖 19 129 個不包含遮蔽效應的濾波器的平方頻率響應及總和

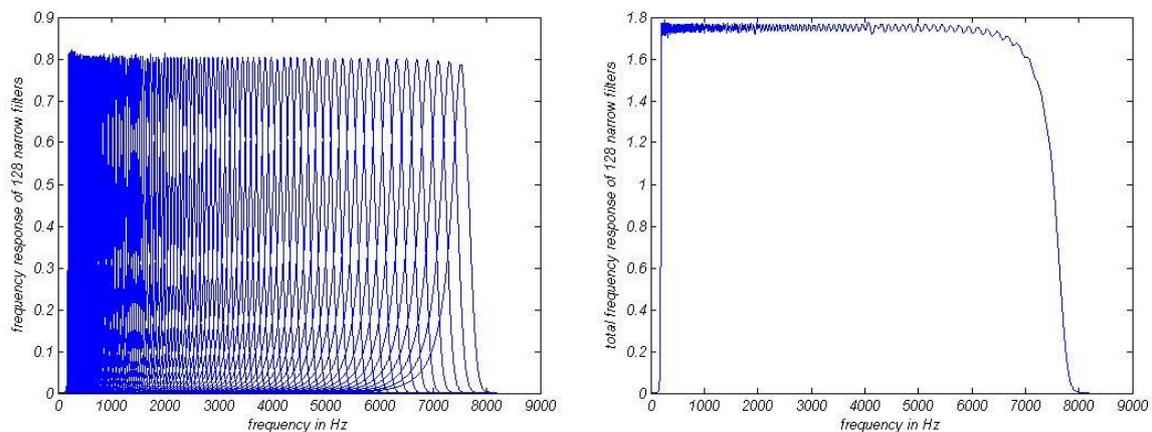


圖 20 128 個包含遮蔽效應的濾波器的平方頻率響應及總和

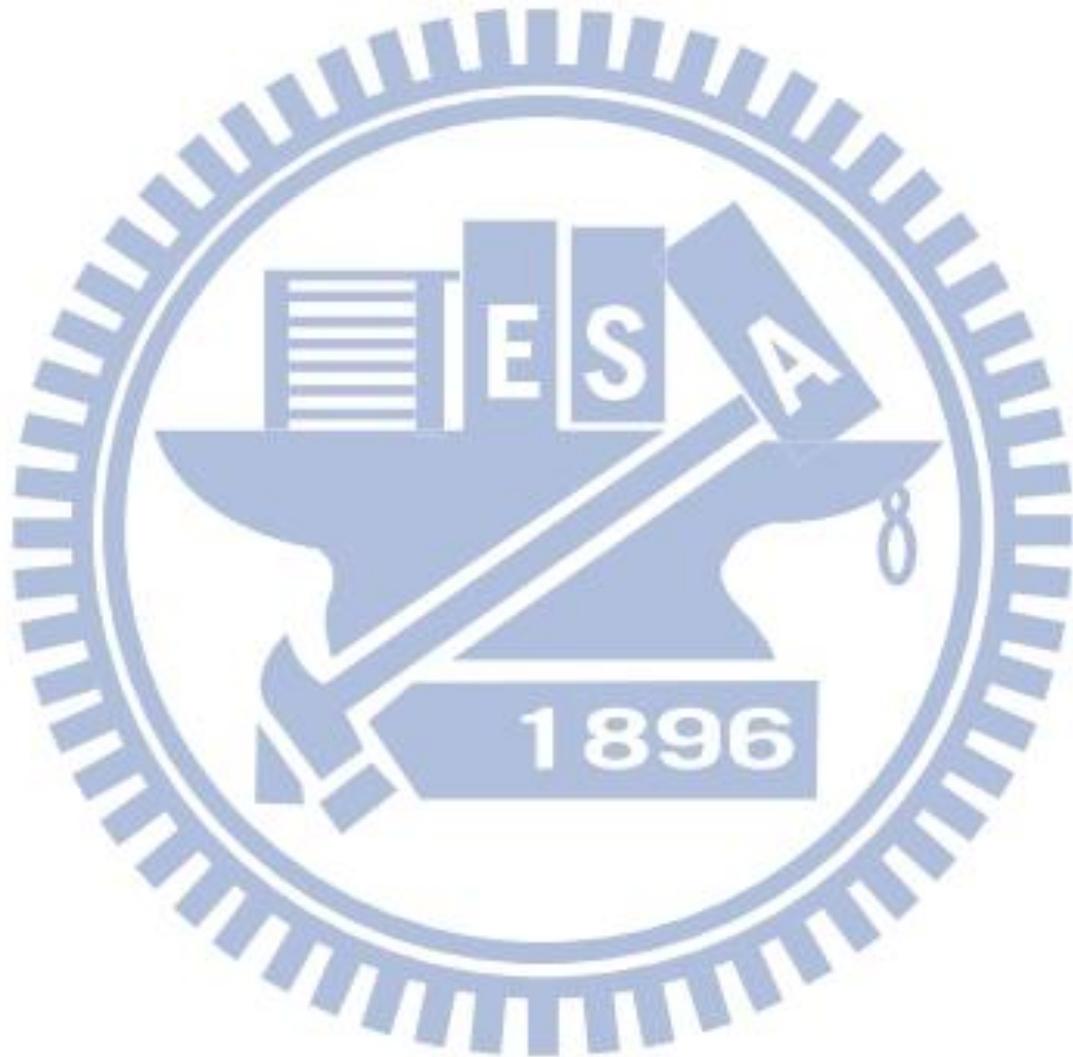
雖然我們可以直接將前後兩個較寬的濾波器的輸出相減得到想要的訊號，然而因為之後必須要做延遲補償的動作，因此在這邊我們還是要求得濾波器的係數才行；在此我們將頻率響應直接做 2048 點的反傅利葉轉換，得到一個 FIR 的濾波器，作為窄濾波器的脈衝響應。

2. 延遲補償 (delay compensation)

因為每一個濾波器的延遲時間都不同，因此當我們要把這些副頻帶訊號加總回來的時候，必須先把延遲時間補回來。最簡單的做法是量測該濾波器的延遲時間為幾個取樣點，再把所有副頻帶延遲到相同的延遲時間後相加；然而因為每一個頻帶包含了很多的頻率，每一個頻率可能都有自己的延遲，並不一定相等，因此這樣子的想法是無法實現的。

Kubin 與 Kleijn 在 1999 年的論文中 [39]，提出「將分析濾波器組的脈衝響應在時間軸上反轉過來，做為合成濾波器組的脈衝響應」；在撰寫程式的時候，我們可以將濾波後的聲音訊號整段反轉過來，再經過同樣的濾波器一次，並將輸出的結果反轉回來，如此一來我們可以得到零延遲的訊號，而其造成的影響僅僅只是被濾波兩次而已，在不需即時性的應用裡，是很好的一個方法，在 matlab 中可以使用 `filtfilt` 執行這個功能。

最後我們必須考慮的是所有濾波器的頻譜響應總和是否為常數？因為延遲補償的關係，在這裡我們要注意的是頻譜響應的平方總和是否為常數；從圖 19、20 中可以發現，在 16KHz 的取樣頻率之下，除了最低頻與 7KHz 以上的頻率成份之外，大部份的頻率響應的波動都很小，可以視為定值，因此也確認了我們使用的濾波器是正確的；接下來就可以在濾波器的架構之下，進行最重要的訊號處理了。



第三章 聽損耳蝸模型

Brian C. J. Moore 是以工程方法研究聽損現象的先驅之一，其團隊的研究包含了各種聽損患者可能產生的現象，例如聽覺濾波器的形狀、響度接收...等等問題 [40-42]，也針對這些可能的聽損現象進行模擬，證明這些現象確實會影響患者對聲音的辨識度 [14-16]；以下將針對這些模型做詳細的介紹，並做進一步的分析與討論。

3.1 頻譜模糊化模型

3.1.1 演算法架構與背景

在 [14] 中，Baer 假設「聲音的激發模式是由不同單音的激發模式線性相加而成」，在這樣的基礎之上，將聲音與基底膜的關係矩陣化，並利用反矩陣的關係，將正常人耳的濾波效果一併考慮進去，最後結合 STFT 演算法，得到模擬聽障患者分頻解析度降低，也就是頻譜模糊化的模型，系統架構圖如圖 21。

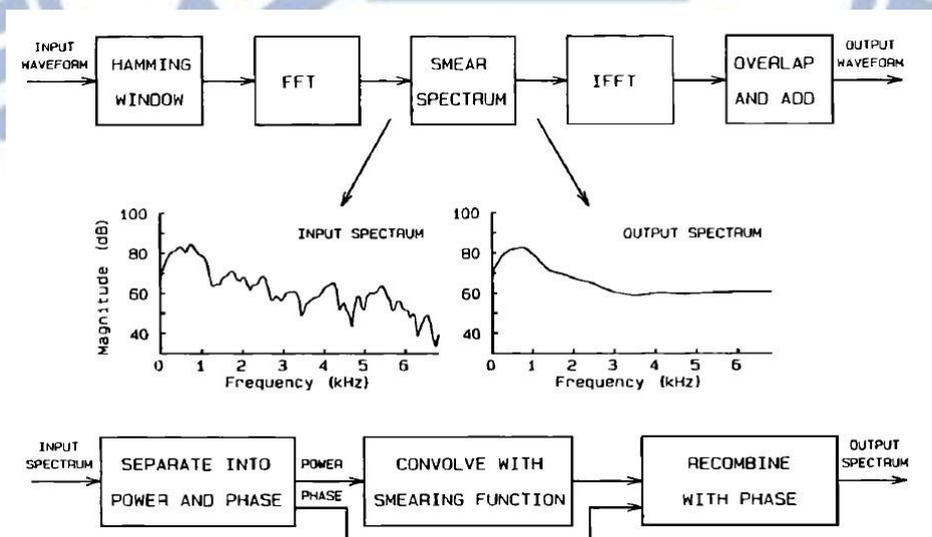


圖 21 頻譜模糊化演算法流程圖解

資料來源：[14]

首先作者將聲音訊號適當的框架化後，做傅利葉轉換得到頻率成份，並以聽覺濾波器組來計算每一個頻率所產生的激發模式，最後將這些激發模式相加起來，得到一個總括激發模式的頻譜，公式如下：

$$\begin{aligned}
 Y &= AX = \sum_i A(:, i)X(i) = \sum_i EP_i \\
 Y(n) &= \sum_i A_N(n, i)X(i)
 \end{aligned} \tag{8}$$

其中 $Y(n)$ 為總括激發模式的頻譜在第 n 個頻率的值、 $A_N(n, i)$ 為中心頻率在第 n 個頻率的正常聽覺濾波器，在第 i 個頻率的增益量、 $X(i)$ 則為框架訊號在第 i 個頻率的值。前式為巨觀形式，較易觀察出激發模式相加的意義；後式則為微觀形式，是實作上的處理方式。

式 (8) 表示了聲音在正常人耳的基底膜上產生的激發模式。為了能夠使正常人聽到相當於聽覺受損者聽到的聲音，我們的目標為：將處理過後的聲音通過正常人耳，相當於未處理聲音通過受損的耳蝸；Baer 根據此目標進行以下的運算：

$$\begin{aligned}
 E &= A_W X = A_N Y_C = A_N (A_S X) \\
 A_S &= A_N^{-1} A_W
 \end{aligned} \tag{9}$$

其中 E 為聲音在受損人耳的基底膜上產生的激發模式、 Y_C 為經過處理的聲音、 A_W 為受損者的聽覺濾波器組、而 X 為原始的聲音；由此可知， A_S 就是我們要對原始訊號做的處理。因此，只要我們有正常與受損的濾波器參數，我們就可以模擬該聽損患者所聽到的聲音。

3.1.2 聽覺濾波器組

為了要能夠模擬不同程度頻譜模糊化的效果，我們要有一個可以參數化模擬正常聽覺濾波器組與受損聽覺濾波器組的模型。許多學者利用 notch-noise test 來測量聽覺濾波器的形狀 [28]，而 Moore 則利用 roex filter (rounded-exponential filter) 來模擬聽覺濾波器組 [43-44]，其數學形式如下：

$$W(g) = (1 + pg)e^{(-pg)}, \quad g = \frac{|f - f_C|}{f_C} \tag{10}$$

其中 f_c 為中心頻率、 g 為目標頻率相對於該頻帶中心頻率正規化後的差值， $W(g)$ 為此濾波器在特定 g 值的增益，而 p 則為控制濾波器寬窄的係數。

一般而言，在某一頻帶下，頻率與增益的關係若以 g -dB 的單位來表示的話，當 g 值介於 0.5 到 3 之間， w 函數為近乎直線；而在 roex filter 的模型中， $-p$ 就是這條直線的斜率，也就大致上決定了這個濾波器的寬窄。正常人的 p 值求法是根據在一般大小聲之下，聽覺濾波器是左右對稱的，因此 roex filter 的等效矩形頻寬（equivalent rectangular bandwidth; ERB）為：

$$ERB = \frac{4f_c}{p} \quad (11)$$

而正常人的等效矩形頻寬與中心頻率的關係為 [28]：

$$ERB = 24.7(0.00437f_c + 1) \quad (12)$$

因此我們可以透過連結式 (11) 與式 (12) 計算出 p 值，因為 p 值與濾波器的頻寬成反比，因此較寬的濾波器頻寬就以 p 值除以一個倍數來模擬。

除此之外，為了讓每一個濾波器對於頻率的增益總量為定值，因此在計算出相對應的 p 值以及增益後，作者將這些值除以濾波器的等效矩形頻寬，作為正規化後的濾波器增益，並將其應用在演算法中。

3.1.3 效能分析

實際將該系統實現分析之後，將此系統的優缺點整理如下。優點方面，因為採用激發模式的想，可以剛好把這種想法以矩陣乘法的形式表現出來，如此一來就可以很輕易的把訊號在受損與正常的聽覺濾波器組之間做轉換，我們只要乘上相對應的方陣或是此方陣的反矩陣就可以了；然而此方法卻會有以下的幾個缺點：

- 為了要做反矩陣的運算，我們必須限制代表濾波器組的矩陣為正方矩陣，也因此實作時會把 FFT 中的每一個頻率當成是一個濾波器的中心頻率，並對其做模糊化；以濾波

器組的想法來看，這種作法的重疊程度太高，而且是頻率間隔固定的，因此平滑化的參數可能與實際聽損患者聽覺濾波器變寬的程度有落差。

- 作者假設每一個頻率成份的激發模式是線性相加的，然而許多生物實驗證明並非如此，相鄰的複音彼此會有非常密切的非線性關係，例如第二章提到的遮蔽效應就是一個最直接的例子；也就是說這樣子的聽覺模型只是非常粗略的版本，若是想要更準確模擬人的聽覺現象，則需要經過更詳細的測量才行。
- 在圖 21 中，因為保留原訊號相位頻譜的關係，使得經過處理的訊號，若再次經過 FFT 的分析，能量頻譜會與我們原先設計的有落差，我們可以解釋為：原訊號的相位頻譜保留了原始頻譜的結構，而我們的演算法只有針對訊號能量頻譜做修改，並沒有將 FFT 所需要的所有資訊—能量、相位—都包含在演算法之中，因此不完整的頻譜經過 IFFT 之後得到的時間訊號，已經與我們設計的不一樣了。

Baer 在 [14] 中針對此現象舉例如圖 22：

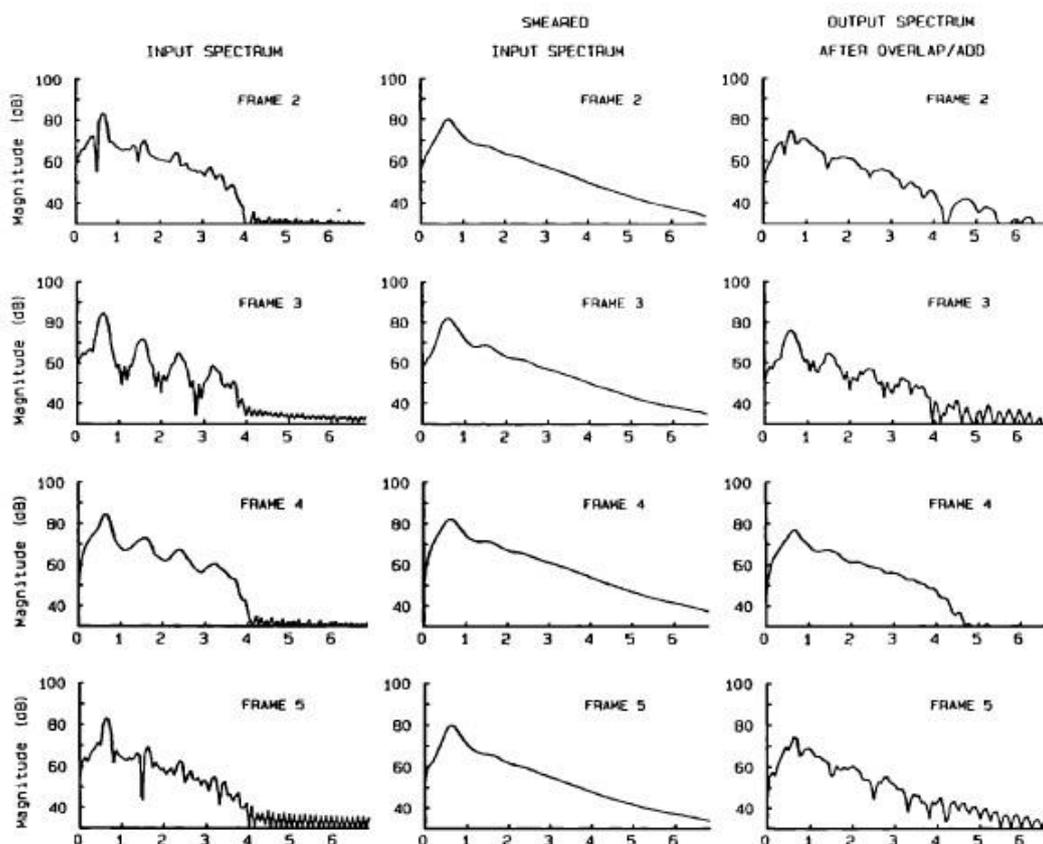


圖 22 母音/æ/的連續框架訊號模糊化前後比較圖

資料來源：[14]

左邊欄代表原訊號能量頻譜、中間欄代表模糊化 3 倍後的能量頻譜、而右邊欄則是經過 OLA 之後再次做 FFT 的能量頻譜；從這三欄中我們可以發現，雖然右邊欄與中間欄較為類似，然而右邊欄還是可以看到左邊欄中母音的頻譜結構；Baer 計算右邊欄的能量頻譜與不同程度的模糊化能量頻譜的最小平方差，發現右邊欄相當於只有模糊化 1.5 倍的結果，與其原先設計的 3 倍有很大的落差。

3.1.4 相位頻譜補償

保留原相位頻譜是一般在聲音訊號經過 STFT 處理時常常用的方式，然而因為以上的缺點，因此 Griffin 與 Lim 在 1984 年提出了最小平方重疊相加，在此基礎上發展了 LSEE-MSTFTM 演算法：不斷的將能量頻譜取代成我們第一次平滑化後的能量頻譜，並搭配原來的相位頻譜，來解決相位頻譜的問題，過程如下：

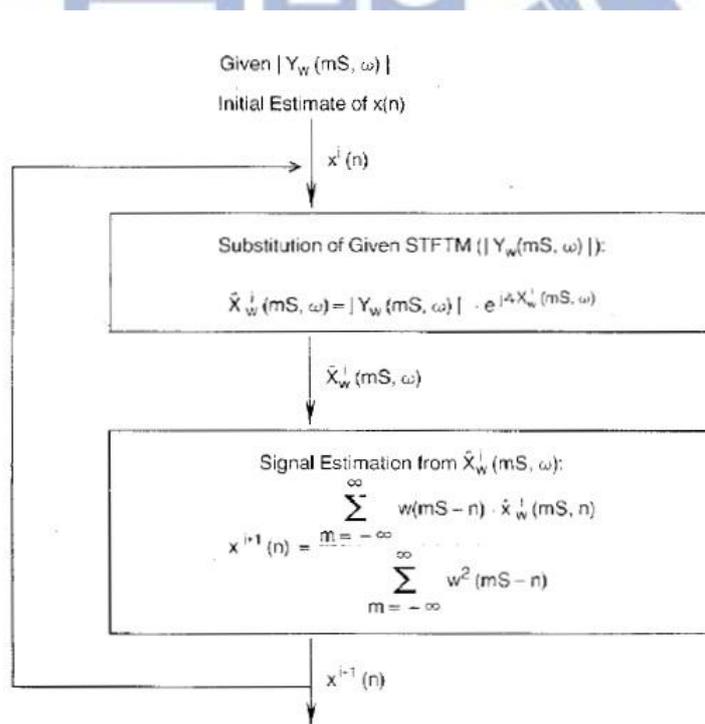


圖 23 LSEE-MSTFTM 演算法
資料來源：[22]

作者並證明此演算法是收斂的，以圖 24 簡略說明之。在每一次的迭代中，我們將能量頻譜修正為我們所期望的值 (a)，然而因為不知道理想頻譜的相位頻譜，因此只能保留原訊

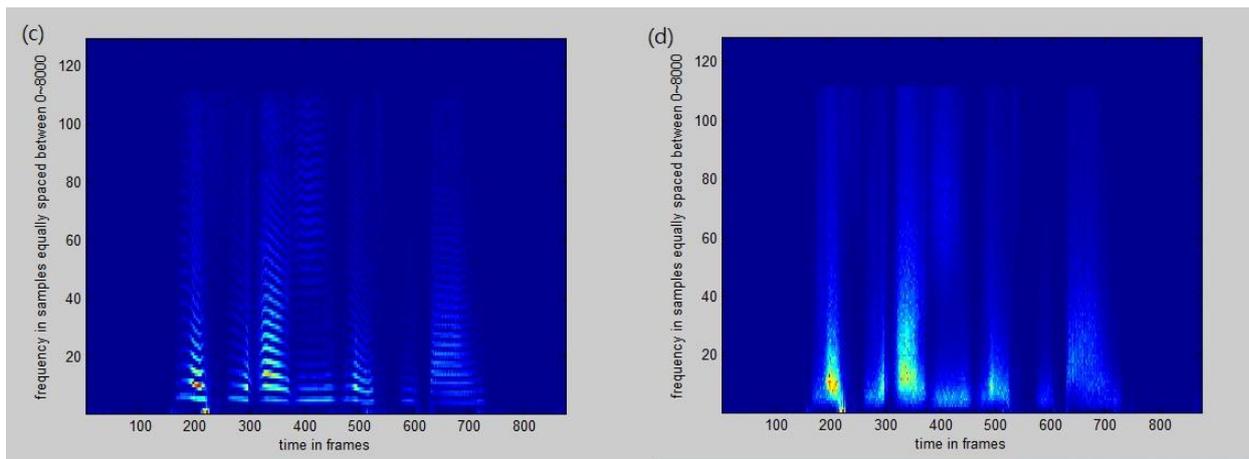


圖 25 結合 LSEE-MSTFTM 與 Baer 演算法的比較

圖 26 則是在迭代的過程中，聲譜圖的最小平方差的變化，由圖可以清楚發現，最小平方差確實會收斂；此外，迭代大約 50 次之後，最小平方差幾乎不再變動了。

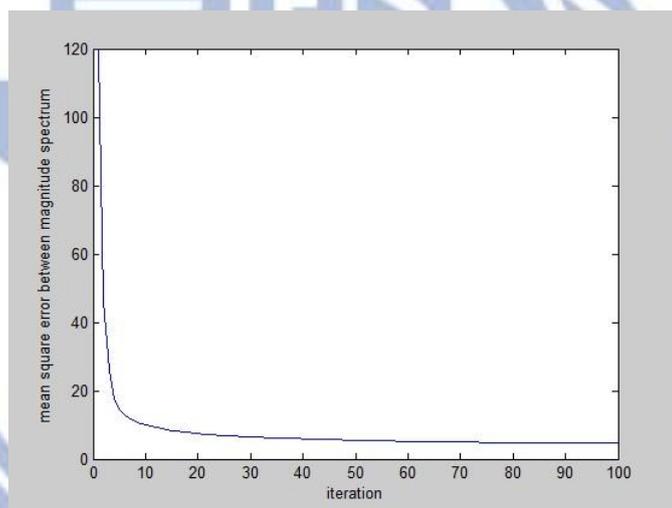


圖 26 最小平方差變化圖

3.1.5 相位頻譜補償延伸

雖然 LSEE-MSTFTM 可以解決相位頻譜的問題，然而若是要用在即時系統上的話，則需要強大的硬體配備，來平行處理每一次的迭代，因此我們試著發展出一步就能得到等同於 LSEE-MSTFTM 效果的演算法。

我們將經過 LSEE-MSTFTM 後的訊號當作是目標訊號，並假設原始訊號經過一個矩陣

乘法後可以得到此目標訊號，如此一來便有希望以最小平方法來解決此問題，不但可以降低運算量，也能達到即時性的要求，公式模型如下：

$$\begin{bmatrix} \text{filter}_1 \\ \text{filter}_2 \\ \vdots \\ \text{filter}_{128} \end{bmatrix} [X_1 \ X_2 \ \dots \ X_N] \cong [Y_1 \ Y_2 \ \dots \ Y_{128}] \quad (13)$$

濾波器矩陣的想法與前面提到的相同， X_i 是代表第 i 個框架的原訊號能量頻譜， Y_i 則是第 i 個框架的處理後訊號能量頻譜。假設以上關係成立，則將矩陣改寫並拆解之後可以得到 128 個小矩陣：

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{128} \end{bmatrix} [\text{filter}_i] \cong \begin{bmatrix} Y_{1,i} \\ Y_{2,i} \\ \vdots \\ Y_{N,i} \end{bmatrix}, Ax \cong b \quad (14)$$

因此只要做 128 次的最小平方法就可以解出原始矩陣的所有參數。

然而，隨著分析的框架數量愈多，最小平方法計算出來的最小平方差會愈來愈大；如圖 27 中，當分析的框架愈多時，左邊的理想聲譜圖，與右邊經過最小平方法求得的聲譜圖差距愈大；換句話說，LSEE-MSTFTM 造成的影響是時變的，我們無法以一個固定的矩陣乘法，使得每一個框架訊號的頻譜都和目標框架訊號的頻譜相同，而此方法無論應用在能量頻譜（實數最小平方法）或是頻譜（複數最小平方法）都得到相同的結果。

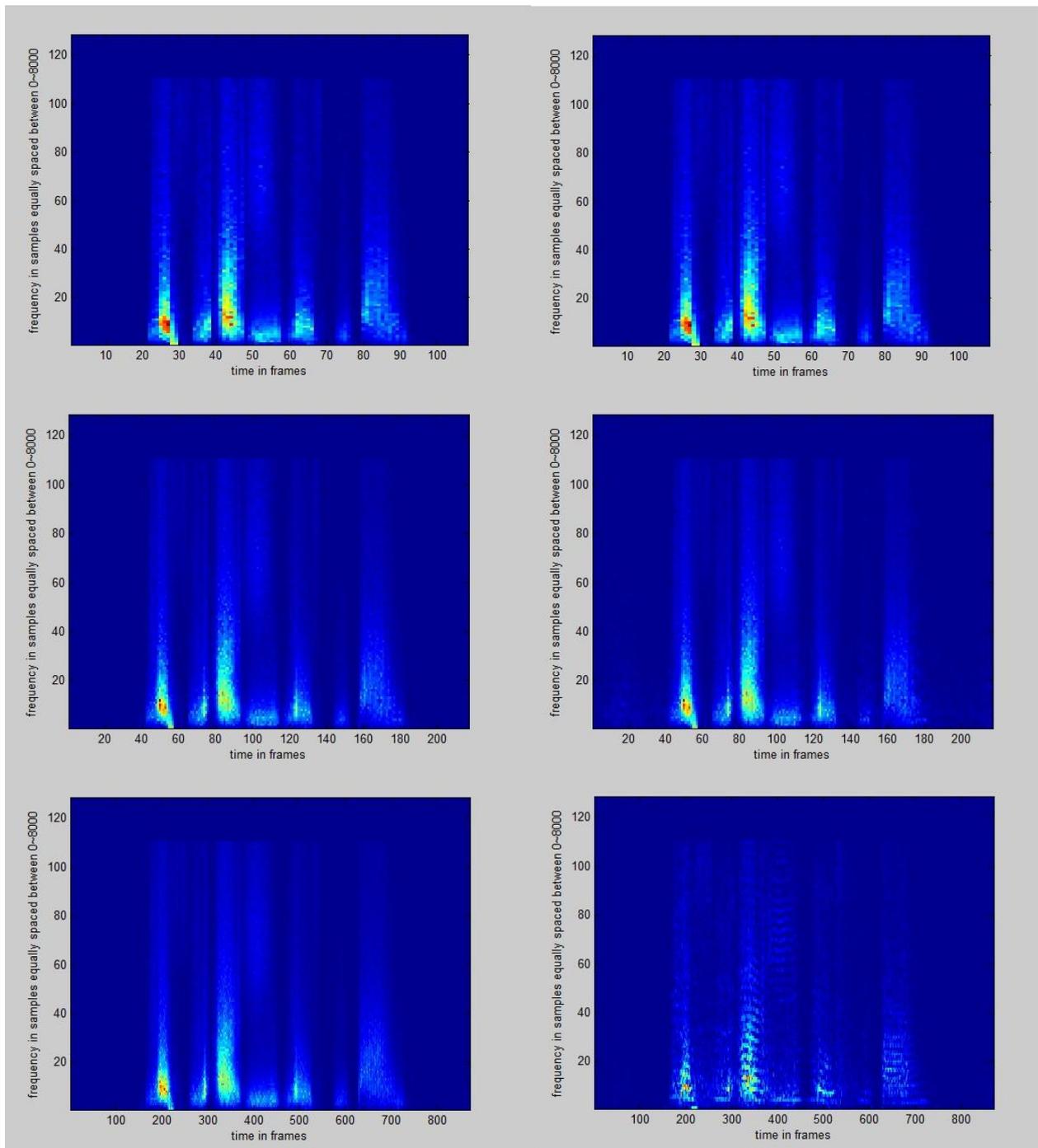


圖 27 由上而下分別是框架位移 2、1、1/4 倍的 LS 聲譜圖與理想聲譜圖的差別

3.2 響度模型

3.2.1 演算法架構與背景

本節將討論 [15] 中提到的模擬最小可聽水平提升與響度聚集的模型。從第二章的介紹中我們知道，不一定在所有的聽力區間之內，都有響度聚集的現象：研究顯示當聲音的響度很高的時候，正常人與聽障患者對應的響度是一模一樣的；此外，當聲音的強度過大的時候，會造成不舒服的感覺，考量到最後評估模型效能時，必須將聲音給受試者聽，因此以下的模擬中，我們設計讓測試語料的瞬時強度小於 100dB，並假設響度聚集的現象只發生在聲音物理強度 100dB 以下。

根據 Moore 在 1985 年的研究結果 [45]，若是在同一聽覺響度之下，將正常人所需的物理強度以聽障患者所需的物理強度來作圖的話，我們可以發現在響度聚集的範圍內，其關係幾乎為一直線，而其斜率則反應聽障患者響度聚集的程度，其值大於 1。此圖不但包含了響度聚集的現象，也包含了最小可聽水平。舉例來說，若是正常人的聽覺響度範圍的聲音強度為 0~100dB，假設聽損患者響度聚集的斜率為 2，則會伴隨著最小可聽水平對應的物理強度提升至 50dB。

假設 L_U 是原聲音訊號的物理強度 dB 值， L_P 是經過處理、模擬聽損患者所聽到的聲音物理強度 dB 值；則根據以上的關係，我們可以列出式 (15)：

$$L_P = N L_U + K \quad (15)$$

K 為某一個常數。若我們要求得在線性振幅之下的關係，則將式 (15) 改寫為式 (16)：

$$\log I_P^2 = \log I_U^{2N} + K \quad (16)$$

其中 I_U 、 I_P 分別原聲音訊號、經處理聲音訊號的線性振幅強度。將式 (16) 化簡成為線性關係後得到：

$$I_P = I_U^N \times 10^{K/2} \quad (17)$$

實際對聲音做處理的時候，我們不需要特別去考慮 $10^{K/2}$ ，因為該項是針對響度聚集的範圍去決定的，因此根據作者的假設，為了讓所有聲音都落在響度聚集的範圍之內 (100dB)，我們把所有聲音的強度乘上 N 次方，再將處理過的訊號，針對其瞬時最大物理強度正規化到 100dB，如此一來即相當於對 K 值做了調整。

3.2.2 模型實作概述

根據不同的頻帶會有不同的最小可聽水平與響度聚集程度，為了準確模擬聽障患者所聽到的聲音，Moore 等人將聲音訊號用較寬的濾波器組（通常聽障患者會有的問題），分成不同的頻帶，並產生該副頻帶訊號的可解析訊號（analytic signal）：

$$f_{\text{analytic}}(t) = f(t) - iF_{\text{HI}}(t) \quad (18)$$

其中 $F_{\text{HI}}(t)$ 為 $f(t)$ 經過喜伯特轉換（Hilbert transform）所得到的訊號。一般而言我們將副頻帶訊號視為一振幅調變訊號；換句話說，將其視為一封包與載波相乘而得的訊號，而可解析訊號的絕對值即為該副頻帶訊號的瞬時強度，也就是該訊號的封包；因此整個訊號處理的過程如圖 21：

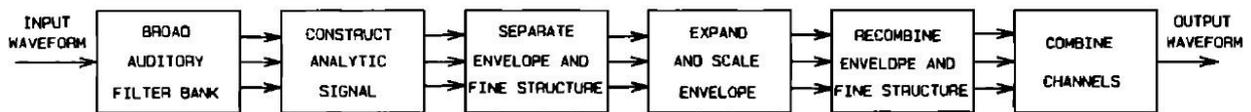


圖 28 模型實作流程圖解

資料來源：[15]

值得注意的是在訊號封包做 N 次方的時候，作者發現在高頻部份會引進一些干擾，因此以式 (19) 取代之：

$$E_p = E \times E_{\text{SM}}^{N-1} \quad (19)$$

其中 E 為原封包， E_p 為經過 Power of N 的封包， E_{SM} 則是以 10-ms 的矩形視窗平滑化後的封包。

3.2.3 模擬結果與分析

由於這個演算法的主要想法是將封包 N 次方，而這樣的做法造成的影響是能量大的會愈來愈大，而能量小的則會愈來愈小，因此我們可以期望經過處理的訊號聲譜圖中，原來能量大的會被保留下來，而能量小的則會消失不見；這樣子的結果可以藉由比較圖 29 的

原訊號聲譜，與圖 30 中不同程度（2、3）次方的聲譜圖中觀察出來。



圖 29 原訊號聲譜圖

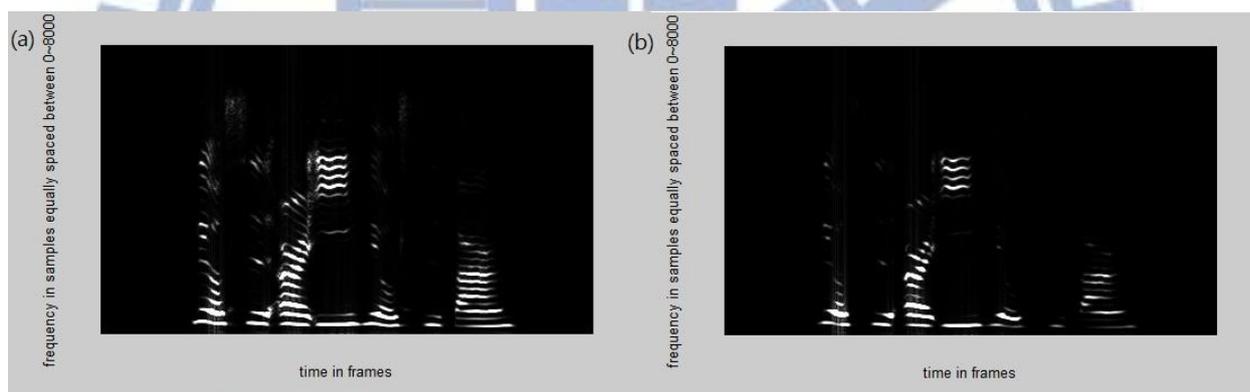


圖 30 (a)、(b)分別為經過 2 次方、3 次方處理的聲譜圖

從圖 30 中看來，Power of N 的方法雖然可以達到很不錯的效果，運算也很容易，然而處理比較簡單所要付出的代價是，基本條件的限制比較多——聲音訊號的能量不能超過 100dB，因為若聲音超過 100dB，則該取樣點的強度是不用被 N 次方的；除此之外，最小可聽水平提升跟響度聚集是被綁在一起同時處理的，只要決定其中一個，另外一項也跟著被決定。

3.3 混合模型

Nejime 在 [16] 中，將 Moore 團隊的兩個個別的聽損模型結合在一起，變成一個模擬三個因素—後小可聽水平提升、響度聚集、與分頻解析度降低—的聽障患者模型；作者並結合了 NAL 的演算法，來驗證一般助聽器能夠彌補多少聽障患者的受損程度，整個系統架構如圖 31。

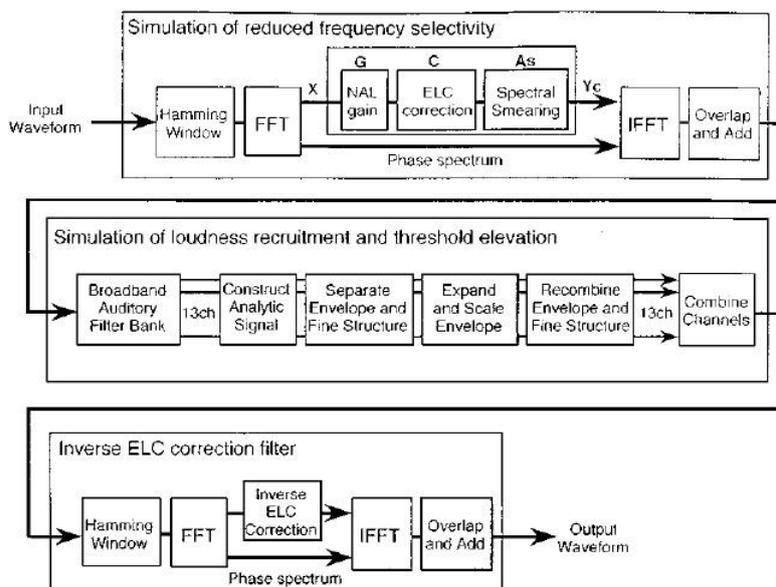


圖 31 混合模型實作流程圖解

資料來源：[16]

在這個混合模型中，加入了一個新的考量—ELC correction，是用來模擬外耳與中耳對不同頻率聲音的增益，使得我們要處理的聲音的頻率成分與人耳真正聽到的聲音頻率成份更為相同。直覺上來考量的話，我們也可以選用最小可聽水平的曲線來當做外耳及中耳的增益量，然而許多研究說明當聲音頻率低於 1KHz 時，內部雜訊會變得很強 [46-47]，使得最小可聽水平曲線在 1KHz 以下會急劇上升 [48-49]；但是當聲音訊號能量較強時，則不會受到內部雜訊的干擾，如圖 12 中強度高的 ELC 曲線幾乎為平行的。

因此 Moore 團隊認為強度大的 ELC 較可以代表外耳及中耳的增益，於是選用 100dB 的 1KHz 單音為基準的 ELC，將其它頻率的強度以此校準之後，再進行原先設計的訊號處理。最後當我們完成所有的處理時，因為聲音最後會經過人耳本身外耳及中耳的增益，因此我們必須將 ELC 的增益值彌補回來，也就是圖 31 中的最後一個階段。

另外，Moore 團隊在此混合模型中先進行了頻譜模糊化的處理，再進行響度聚集的調

整；然而在真正的受損人耳中，這兩件事是同時發生的，因此我們無法從原因來決定到底應該先進行什麼樣的處理。在這裡我們將這兩種先後順序都實作出來，如圖 32 中的左圖是先響度模擬再頻譜模糊化，右圖則是先頻譜模糊化再響度模擬；我們可以發現，右圖的能量較集中，能量差也較大；從此結果來分析，若是先進行響度聚集再頻率模糊化，則原本已經消失不見的成份又會跑出來，相當於響度聚集的效果不見了；因此我們可以斷定，在實作上先進行頻譜模糊化，再進行響度聚集的順序是較為正確的。

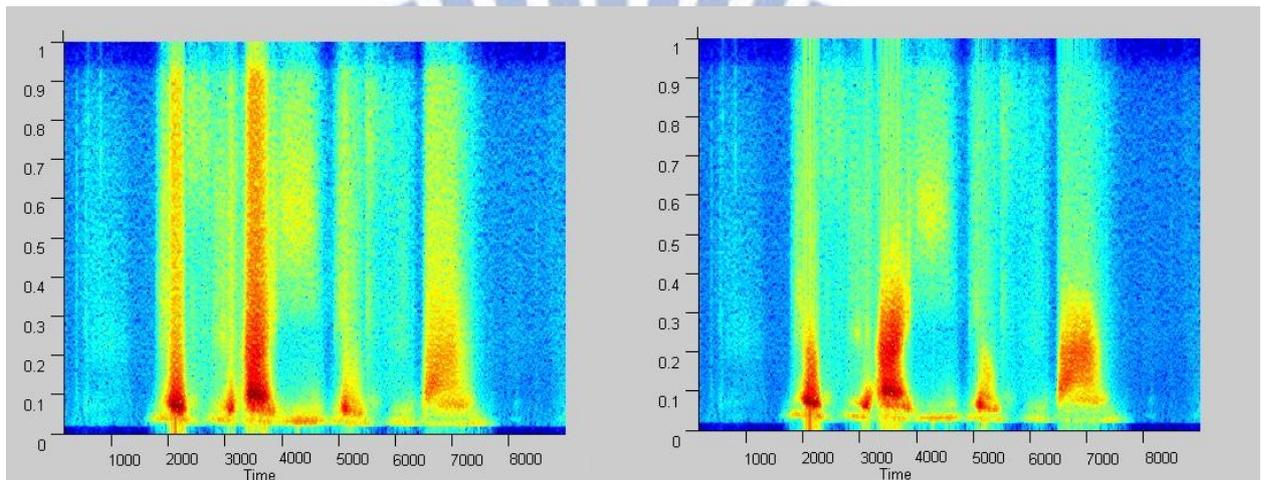


圖 32 不同模擬順序的比較圖

第四章 分頻解析度補償

在本章中我們先介紹「多頻帶頻率壓縮演算法」(multi-band frequency compression algorithm)，是一種經部份學者證實的確可以提升分頻解析度降低患者的辨識率的演算法 [23-24][50]；接著我們分析其優缺點後，將其做法延伸發展成新的演算法；最後我們利用 Moore 的聲音接收的矩陣公式，嘗試發展出不一樣的演算法。

4.1 多頻帶頻率壓縮演算法

4.1.1 演算法核心架構

多頻帶頻率壓縮演算法的主要想法如圖 33：

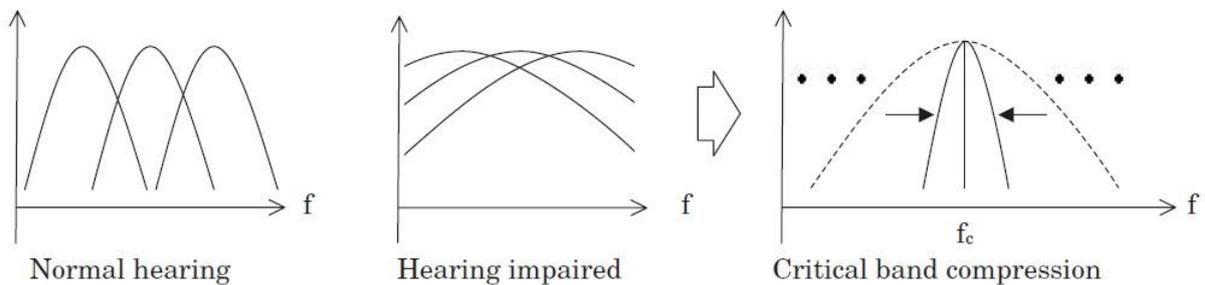


圖 33 多頻帶頻率壓縮演算法的主要想法

資料來源：[50]

從緒論中我們知道，分頻解析度降低的現象，從工程的角度來看，就是人耳相對應的聽覺濾波器變寬了，如圖 33 中所示，如此一來造成的影響是，原本不屬於此頻帶的聲音也會被算進來，使得聲音變得較為吵雜，並且分不清楚語音的音高。

因此，我們將原本屬於該頻帶的能量，以某種演算法，將其往此頻帶的中心頻率集中，如圖 33 右，這麼做的效果是集中能量成份，同時減少干擾成份。當處理過後的聲音經過聽障患者較寬的聽覺濾波器之後，即使因為濾波器較寬，然而因為我們將頻率成份集中，

加上原本屬於其它頻帶的能量成份也往其它頻帶的中心頻率集中，因此在該頻帶內成份一消一長之下，主要成份被遮蔽的機會就變小了。

4.1.2 壓縮方式比較

Kulkarni 與 Pandey 在 2008 年，以分頻帶頻率壓縮的想法，提出了「頻率區段對應」的演算法，並與其它兩種演算法做比較，以下分別介紹之。

1. 點對點對應 (sample to sample mapping)

對應關係如下：

$$\begin{aligned} k' &= k_{ic} + \text{round}(\alpha(k - k_{ic})) \\ k_{ic} &= 0.5(k_{ic} + k_{ie}) \end{aligned} \quad (20)$$

其中 k 為 FFT 的第 k 個頻率成份、 k_{ic} 為第 k 個頻率成份所屬的濾波器的中心頻率相對應的頻率標記、 α 為頻率壓縮的程度（介於 0、1 之間）、 k' 為第 k 個頻率成份經壓縮後得到的新的頻率標記、而 k_{ic} 、 k_{ie} 則為該頻帶最低頻與最高頻的頻率標記，實際例子如圖 34：

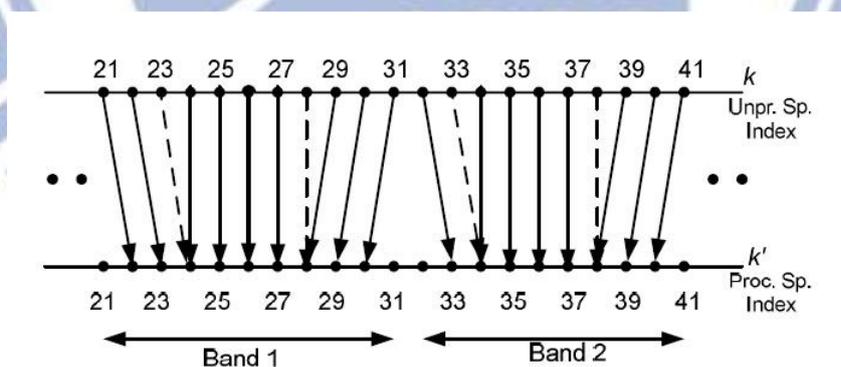


圖 34 點對點對應示意圖

資料來源：[23]

在此方法中，若是有兩個以上的頻率成份對應到同一個頻率標記時，則只會有一個頻率成份被保留下來，因此可能會有部份頻率成份遺失的問題，訊號的總能量也隨著減少。

2. 頻率線性相加 (Superimposition of spectral samples)

為了解決上一個演算法的問題，因此經由點對點對應演算法而得到頻率線性相加演算法，如圖 35：

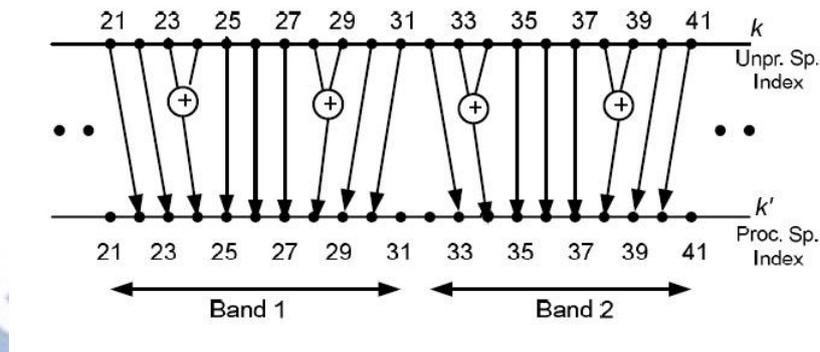


圖 35 頻率線性相加示意圖

資料來源：〔23〕

簡單來說就是把對應後屬於同一個頻率標記的頻率成份加總在一起，如此一來就可以讓總能量維持不變；可是在這些頻率標記與相鄰的頻率標記上會有一些不規則的變化。

3. 頻率區段對應 (Spectral segment mapping)

此一方法是以連續頻譜的觀念，將原訊號頻率與所屬頻帶中心頻率的間距壓縮後，再累加到新頻譜的整數標記。公式與圖解 (圖 36) 如下：

$$\begin{aligned} a &= k_{ic} - [k_{ic} - (k' - 0.5)/\alpha] \\ b &= a + 1/\alpha \end{aligned} \quad (21)$$

$$Y(k') = (m - a)X(m) + \sum_{j=m+1}^{n-1} X(j) + (b - n)X(n) \quad (22)$$

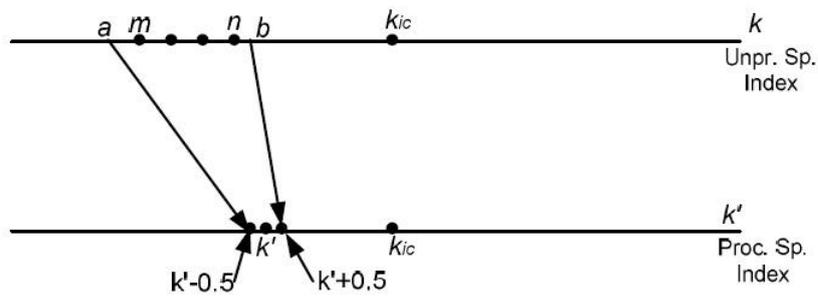


圖 36 頻率區段對應示意圖

資料來源：〔23〕

在計算時我們從 k' 出發，對於每一個頻率標記，我們假設上下各 0.5 單位範圍內的能量都屬於該頻率標記，並計算該範圍的下界對應到原頻率的範圍 (a 值)，而 m 就是在該範圍內的第一個標記，接著依照壓縮比例的不同，計算上界 (b 值)，將在這個範圍內的值累加在新頻譜的同一個標記上。由於此方法引入了平滑化的參數 ($m-a$ 、 $b-n$)，因此讓相鄰頻譜較為圓滑，可以解決線性相加演算法中的缺點。

從以上比較中得知頻率區段對應演算法有較佳的表現〔23〕，因此將之與傳統的 STFT 演算法（保留原相位頻譜，只對能量頻譜做修改）結合，並選取適當的框架長度與 FFT 點數，作為完整的訊號處理流程。

4.1.3 演算法分析與討論

實作此演算法後，與原作者的實驗結果〔24〕相結合，分析得到以下幾個優缺點。優點部份，作者將訊號給真正的聽損患者聽，並將數據做進一步的統計分析，發現對於大部份的聽損患者來說，在壓縮頻率為 0.6 左右時，語音的辨識率都有明顯的提升，而對於少部份的患者來說會有反效果或是沒有效果；這些不明顯的效果可能是因為這些患者有分頻解析度降低之外的其它因素所造成，因此，大體來說，這個演算法的確可以部份改善分頻解析度降低的問題。

然而缺點部份如下：

- 因為保留原相位頻譜的關係，一樣會使得能量頻譜偏移我們一開始所設計的，因此可以視為頻率壓縮的參數被原相位頻譜彌補回來了。
- 雖然作者有提到，這種作法會保留語音頻譜上的泛頻的結構，但是因為在病患實驗中只以辨識率作為演算法好壞的基準，並沒有考慮到語音品質的問題，因此我們合理的猜測，這種做法對於語音品質有一定的破壞程度。
- 在實作演算法、處理分頻問題時，會遇到很多有關邊界上的問題，因此我們必須做一些額外的考量，因此可能再度產生相鄰訊號頻譜連續性上的問題。
- 分頻的選取與病患聽覺濾波器變寬程度的交互關係。當病患的聽覺濾波器含蓋了三個正常的聽覺濾波器時，由於這個演算法不會對位於中心頻率的頻率成份做調整，因此相鄰的中心頻率若被含蓋在變寬的濾波器時，原來屬於其它頻帶的部份能量遠離，部份能量拉近，演算法的效能也有可能變得較不明顯。
- 由於牽涉到 FFT 的處理，因此該演算法還未實際應用在即時的系統中。

4.2 多頻帶頻率壓縮延伸

參考 Kulkarni 與 Pandey 「頻率區段對應」的演算法，在「分頻帶頻率壓縮」的架構之下，我們提出新的演算法，針對上面提到的缺點中的第二點—頻率壓縮演算法對於鄰近頻譜的造成的不連續的現象—做一些分析與改進。

要自然的讓頻率往中心集中的方法，直覺來說就是升頻 (up-sampling)；試比較圖 37，左圖是取樣頻率 8000Hz 的訊號頻譜，而右圖則是升頻兩倍後的頻譜圖。若 x 軸的坐標軸刻度相同的話，則右圖可以視為將左圖把頻率成份往中心壓縮；因為該做法是整體一起考慮，而不是一個點一個點去分析，因此用這種方法得到的頻譜較為完整，語音品質應該會比較好。

除此之外，由於一般助聽器的演算法都是以濾波器組的形式來對訊號做處理，因此我們也希望將這套演算法應用在濾波器組之上。

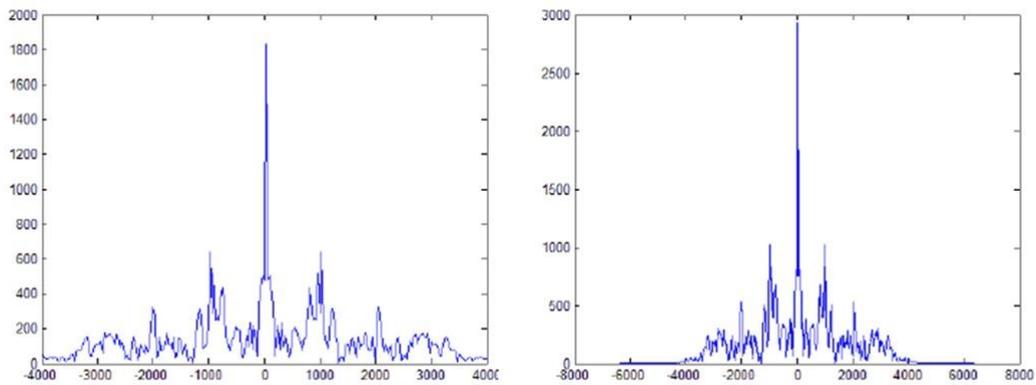


圖 37 取樣頻率不同的訊號頻譜圖

4.2.1 演算法流程與架構

圖 38 為延伸的演算法流程圖，細部說明如下：

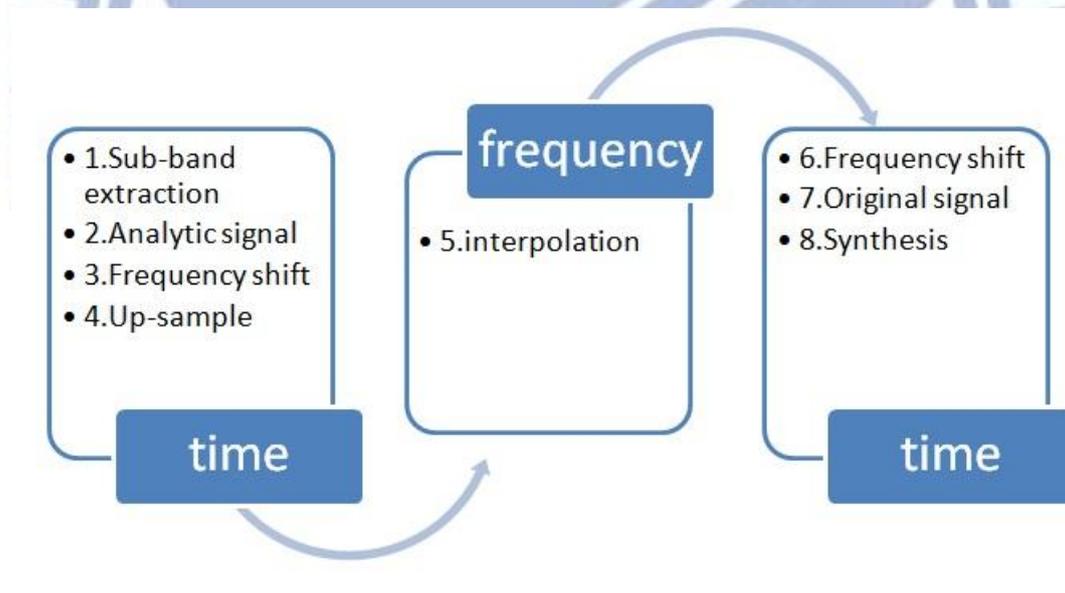


圖 38 延伸演算法流程圖

- 步驟 1：由於我們希望此演算法可以實作在現今助聽器演算法的架構之下，因此採用從 ANSIS1.11 標準中選出的 18 個濾波器組 [51]，將訊號分頻成副頻帶訊號。
- 步驟 2、3、4：由於升頻不只會影響頻譜成份的間距，也會影響中心頻率的頻率，因此我們做 AM 調變，把副頻帶訊號的中心頻率移至原點之後，再做升頻的動作。而在

做調變的同時，考慮到負頻與取樣頻率的關係，很可能會產生 aliasing，因此我們將負頻拿掉，以可解析訊號的形式來進行升頻的動作。

- 步驟 5：升頻之後，我們要考慮的是，如何把 x 軸的坐標換成與左圖一樣的坐標（圖 37）；換句話說，我們要如何把取樣頻率還原，並保留相同的頻譜。假設有一個框架訊號 X 在 8KHz 取樣頻率之下有 200 點，升頻 2 倍變為 Y 訊號，400 點、取樣頻率為 16KHz；另外假設一個 Z 訊號是我們要的訊號，也就是說 Z 訊號為 8KHz、200 點的訊號，而其頻譜與 Y 訊號是一樣的；比較 Y 訊號與 Z 訊號的頻譜，發現 Z 訊號的頻譜相當於把 Y 訊號在頻譜上做降頻（down-sample）的動作，也就是內插，因此在這個步驟中，我們將訊號的頻譜內插成原來的點數。
- 步驟 6、7、8：得到內插訊號之後，透過 IFFT 轉回時域訊號、解 AM 調變後取實數部份，再將副頻帶訊號合成為完整的訊號即可。

4.2.2 實作結果分析

雖然圖 38 的演算法看起來十分完整，然而在實作上卻遇到了不少困難，條列如下：

- 由於將訊號分為框架訊號處理，每一個框架訊號要分為 18 個副頻帶訊號，每一個副頻帶訊號要改變兩次取樣頻率，若以 matlab 內建的 resample 指令，則計算時間過長；因此在此採用簡單的內插方式來改變取樣頻率，方法如下：
- (1) 在原始的軸上，根據壓縮的倍率內插出新的點所對應的標記，接著以左右兩點的值做線性組合得到該點的值。以壓縮 0.6 倍為例，參考圖 39 所示：

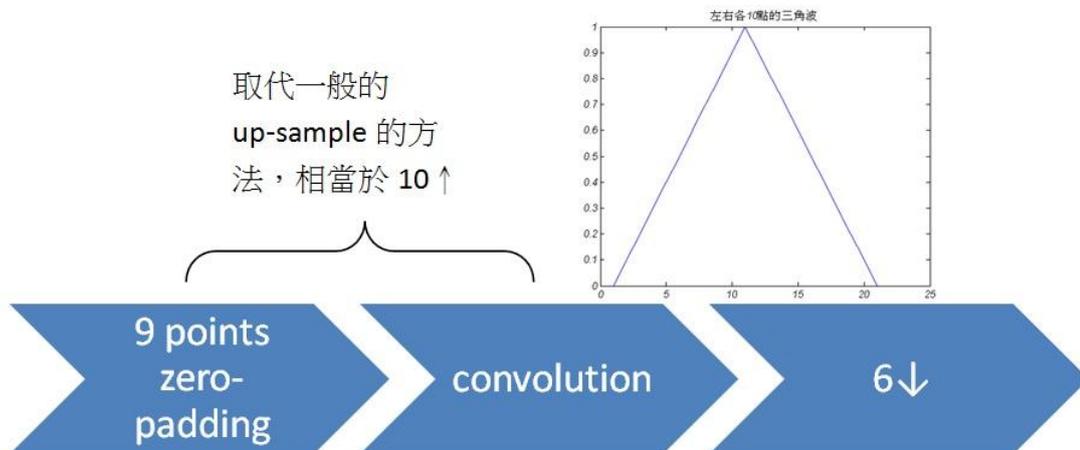


圖 39 簡單內插法分解說明

- (2) 經由取樣理論得知，要得到一個沒有干擾的訊號，我們需要 $\pi/10$ 的低通濾波器，然而上圖中的脈衝響應僅為一個 $\pi/5.25$ 的低通濾波器，因此會濾不乾淨。
 - (3) 假設訊號均為窄頻訊號 (narrow-band signal)，則在經過第一個階段之後，訊號主要能量分佈在「 $\pi n/5$ ， n 為整數」的位置；在基頻的訊號增益可視為定值；而在正負 $\pi/5$ 部份的能量則構成干擾訊號的主要成份，但是此能量在該副頻帶之內並不會造成 aliasing。
 - (4) 若將各副頻帶訊號合成為原聲音訊號，則上述較強能量的干擾訊號會造成 aliasing，因此視情況我們可以另外再通過其它的低通濾波器來確保基頻訊號的增益及干擾訊號的能量。
- 由於在此演算法中同時使用了 FFT 與濾波器組的訊號分析方式，而兩種方式各有各的缺點，因此也特別分析這些方法造成的影響。
- (1) FFT 會因為保留原相位頻譜的關係，使得壓縮的程度減少；然而在此演算法中，我們同時針對能量頻譜與相位頻譜做內插的關係，因此可以認為這樣子得到的頻譜是較為正確的。
 - (2) 因為使用濾波器組的關係，同一個頻率的成份會同時往兩邊移動，因此可以預期當我們把副頻帶訊號合成之後，依然會把壓縮的程度抵消掉。

- 由於這 18 個濾波器組是以 QMF (Quadrature mirror filter bank) [52] 的觀念設計的，而其中包含了濾波器係數降頻，因此若框架的長度太短的話，低頻副頻帶的框架點數會太少，如此一來可能因為測不準原理，造成頻譜的解析度很低，內插的效果也有待進一步的討論。
- 由於 FFT 的關係，我們必須把訊號乘上一個視窗，然而不管所選的視窗型態為何，都會造成頻譜擴散開的效果，這也與我們要把頻譜集中的想法有所抵觸，因此在做分析與合成時都要謹慎分析所使用的視窗的型態與長度。

4.2.3 未來方向

在濾波器組的分析方式中還使用 FFT 並不是常見的做法；除此之外，若要實際應用在助聽器中，則 18 個濾波器的 FFT/IFFT 計算量也太過於龐大，加上使用 FFT 還必須考慮視窗問題、OLA 等等問題，搭配上濾波器組的設計，使得問題過於複雜難以分析；因此我們希望找到某種時域上的處理，來達到「減少點數，但保留原訊號頻譜圖形」的 DSP 演算法。

直覺上的做法有二個：第一個是將時域上多餘的點直接砍掉；此一想法是假定在每一個副頻帶、每一個框架中，聲音訊號都是由很多段非時變的小區間所組成的；換句話說，該框架的長度遠小於一般我們認為語音為非時變的長度—20~30ms；根據傅利葉理論，兩段相同的訊號串連而成的訊號頻譜，與只有一段訊號的頻譜，除了總能量不同之外是一模一樣的，因此，我們可以直接把多餘的點數丟掉。

第二個想法是由第一個想法延伸而來，主要是考慮到時間連續性上的問題。若是把一個副頻帶中的訊號突然丟掉，則可能會有不連續的情形，因此將多餘的訊號直接與之後的框架訊號累加，是一個類似 OLA 的方法。

然而，雖然這些演算法很容易撰寫，卻很難評估其效應，因為需要有真正的病患做測試，也沒有辦法即時得到回饋並修改演算法，因此我們需要一個能模擬聽障患者耳蝸的系統才行。

4.3 矩陣化模型的應用

根據第三章 Baer 提出的模型，其優點是可以將正常人耳聽覺濾波器的影響考慮進去，因此我們試圖利用該矩陣模型，用以解決分頻解析度降低的問題，公式如下：

$$\begin{aligned} E &= A_W Y_C = A_W (A_S X) = A_N X \\ A_S &= A_W^{-1} A_N \end{aligned} \quad (22)$$

式 (22) 與原先 Moore 的公式非常相似，其意義是「將處理過後的聲音通過受損的耳蝸，相當於未處理聲音通過正常人耳」；若是以這樣的想法，則雖然只是很簡略的模型，但是卻可以完美的讓耳蝸受損者聽到正常的聲音。

然而實作後會發現一個非常嚴重的問題：經過處理後的訊號會出現「負」的能量，然而在原先的模型中卻不會有這種問題。直覺上我們可以假設 A_W^{-1} 會抵消 A_W 的效果， A_W 是較寬，能量較大的濾波器組、 A_N^{-1} 會抵消 A_N 的效果， A_N 是較窄，能量較小的濾波器組；因此 $A_N^{-1} A_W$ 會得到一個正的增益，反之 $A_W^{-1} A_N$ 就會有負值的出現。負的能量可以用耳蝸在基底膜上的行進波來解釋，亦即相鄰頻率成份會有遮蔽效應，然而遮蔽效應並沒有被包含在這個模型中，因此也無法反應處理過後的聲音訊號。

根據以上的分析，我們可以得到一個簡單的結論：想要對聲音訊號做預先處理，使聽損患者聽到與正常人一模一樣的聲音是不可能的。因此，在考慮如何補償聽損患者因為分頻解析度降低而造成較差的語音品質時，我們僅能針對語音中的部份特性做考量，並讓分頻解析度降低的影響減到最小，例如本章中探討的演算法一樣。

第五章 基於濾波器組的聽損模型

在第三章中我們探討了 Moore 團隊的幾個模擬聽損現象的模型，因此在本章中我們將以這些模型為基礎，針對其缺點加以改良，並發展出新的聽損模型。我們針對幾個主要問題探討如下：

- 3.1 節中的頻譜模糊化模型，我們希望濾波器組的頻譜間隔可以符合人耳的接收情形，也就是呈 log 分佈的濾波器組。除此之外，由於保留原相位頻譜會造成能量頻譜偏移的問題，使得結果不符合我們原先的設計；我們希望透過其它的處理，來避免這個現象。
- 3.2 節中的響度聚集模型，最小可聽水平是與響度聚集一起變化的，我們無法改變其值，此外聲音大小也被限制在 100dB 以下；因此我們將模型調整得更為彈性，能夠自訂最小可聽水平，也適用於各種大小的聲音。
- 3.3 節中的混合模型，必須先使用 STFT 調整 ELC 增益並進行頻譜模糊化、以濾波器組進行響度聚集，最後再以 STFT 來做反 ELC 增益；如此一來計算複雜度較大，而且 STFT 本身保留原相位頻譜造成的問題也有可能會放大，因此我們期望透過濾波器組的形式來完成混合模型。

5.1 響度模型

5.1.1 主要架構

在 Moore 的模型中，使用了 Power of N 的方式來模擬響度聚集，並透過調整瞬間強度不超過 100dB，讓所有的語料都落在同一個響度聚集的範圍中，而其相對應的計算公式為式 (15)，重寫如下，其中 N 為響度聚集的程度，K 為常數。

$$L_p = NL_U + K \quad (23)$$

然而式 (23) 中，當我們假定當 $L_p=100=L_U$ 時，則 K 值也就固定了；因此為了讓響度

模型更有彈性，我們加入了最小可聽水平的參數，並假設響度聚集的範圍對應到正常人聽到的 0~100dB，若是超過這個範圍，則沒有響度聚集的現象（意即 $N=1$ ）。

根據以上幾個判斷準則，我們將公式重寫如下：

$$\begin{cases} L_P = 0 + (L_U - th), & L_U < th \\ L_P = 0 + N(L_U - th), & th \leq L_U \leq th + 100/N \\ L_P = 100 + [L_U - (th + 100/N)], & th + 100/N < L_U \end{cases} \quad (24)$$

其中 N 為響度聚集的程度、 th 為最小可聽水平、 $100/N$ 則是響度聚集的範圍大小。根據以上的公式，在實作演算法時，我們會一個點一個點考慮訊號的強度，並針對該強度進行相對應的處理；這樣子雖然處理費時，但是也較為精準。

5.1.2 實作結果

在此小節中，我們將驗證以公式求得的響度模型確實會與 Moore 團隊使用 Power of N 的模型相同。首先將新的公式代入原先 Moore 團隊的演算法，並將式 (24) 調整成跟式 (23) 一樣，接著模擬 $N=2$ 與 $N=3$ 時，兩種方法的實作結果如圖 40、圖 41（左邊的圖為 Power of N 的方法，右邊則是根據能量值以分段線性公式而得）。我們發現左右兩邊的聲譜圖幾乎是一致的，因此驗證了 Power of N 的方法與式 (23) 確實是相對應的。

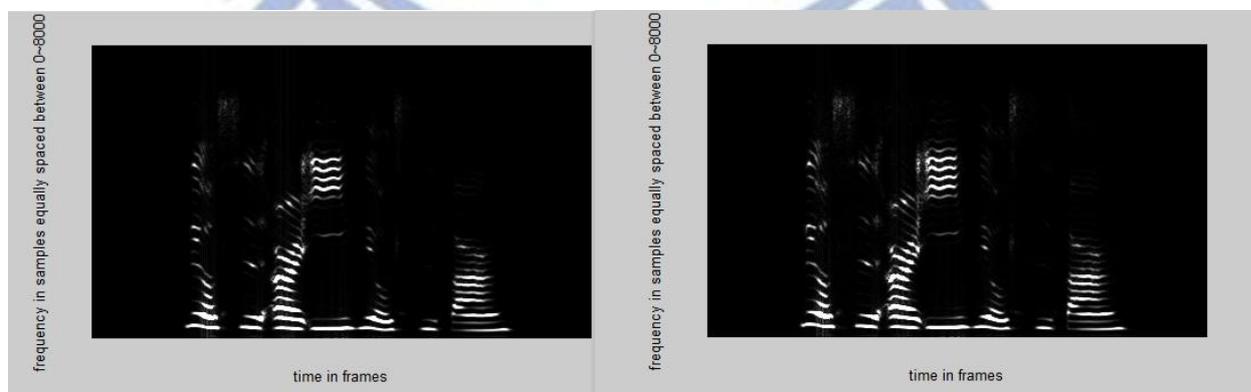


圖 40 Power of N 與分段線性公式實現 $N=2$ 的聲譜比較圖

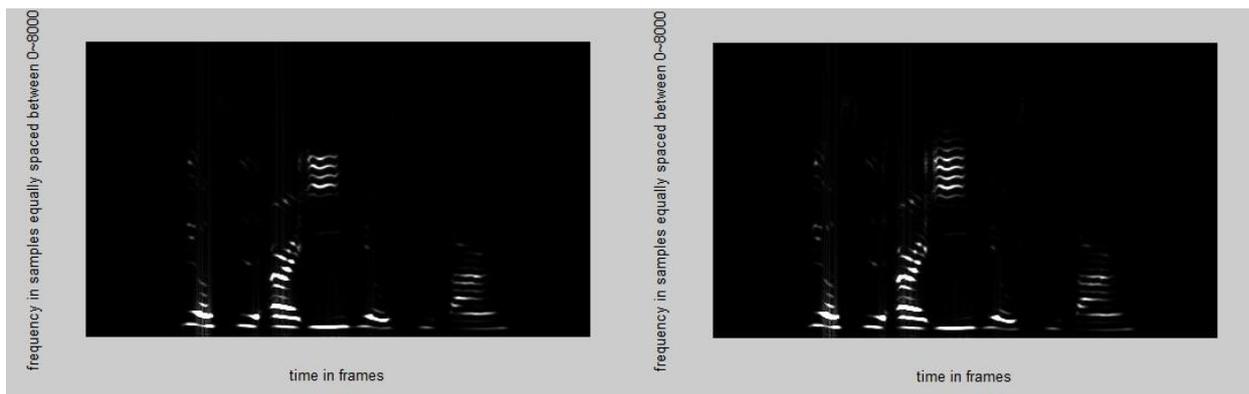


圖 41 Power of N 與分段線性公式實現 N=3 的聲譜比較圖

5.1.3 迭代封包補償法

使用濾波器組的架構來調整封包，會因為濾波器組互相重疊的關係，使得同一個頻率的聲音會對應到不同的非線性處理；因此若再次將處理過的聲音通過濾波器組，該副頻帶訊號會與我們所設計的不一樣，如圖 42。

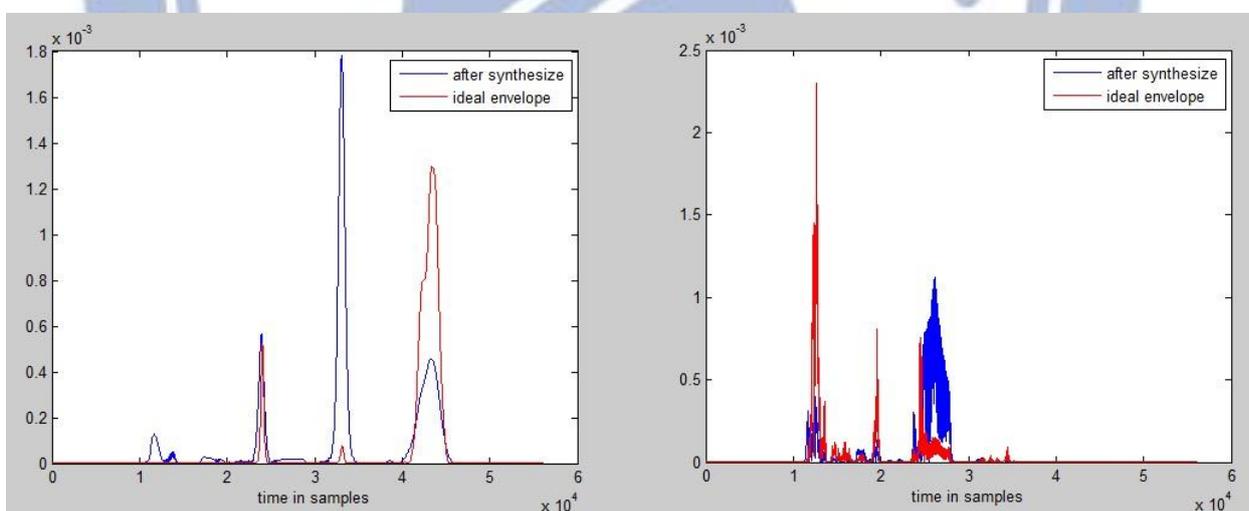


圖 42 左右分別為第 28、117 個 channel 的封包比較圖

為了解決這個狀況，我們參考 LSEE-MSTFTM 演算法，提出迭代波補償演算法(iterative envelope compensation algorithm; IEC)，希望可以透過迭代的方式讓封包愈來愈像我們設計的理想封包，如圖 43：

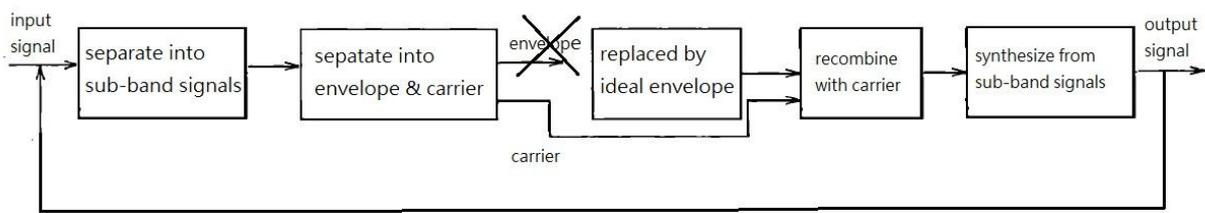


圖 43 迭代封包補償法示意圖

然而結果並不如我們原先所設想的會愈變愈好，反而愈來愈糟；分析原因如下。

LSEE-MSTFTM 透過迭代的方法，讓原相位頻譜造成的誤差愈來愈小；這是一種 convex projection 的方法，其條件是兩個 domain 必須為一對一對應關係。在這樣子的條件之下，我們在兩個 domain 分別加上一些限制，使得在另一個 domain 的限制之下對應回來的訊號，與理想訊號的誤差愈來愈小。LSEE-MSTFTM 就是透過傅利葉轉換中時域與頻域的一對一對應而形成的演算法，我們可以將原相位頻譜造成的誤差視為初始條件的響應，透過迭代的方法，我們讓初始條件逐漸衰減，達到較理想的頻譜。

然而，濾波器組的響應總和只是很趨近於一個常數，並非真的為常數（圖 20）——在極低頻與 7000Hz 以上的頻率響應是變化非常大的；除此之外，在其它的可通頻帶（pass-band）範圍，其響應則是小幅度的上下來回振動。這種頻譜上的缺陷在一次性濾波中通常被忽略不計，然而若我們要不斷的進行分析-合成的步驟，這些小的頻譜上的漣漪則會被放大，造成訊號品質漸漸下降。我們利用 PESQ（Perceptual Evaluation of Speech Quality）計算只經過分析-合成的訊號品質，得到如表 1 的結果。

表 1 迭代方法對語音品質的影響

	Score of PESQ
Original speech	4.5
1-shot speech	4.331
50-shot speech	2.938
100-shot speech	2.461

另外，在 IEC 中，相鄰的濾波器是互相重疊的，即使我們將訊號分頻處理，封包訊號

在每一次迭代中都會被相鄰頻帶影響；這些原因都可以發現這個演算法沒有滿足 convex projection 的基本條件，因此我們無法透過迭代的方式讓誤差愈來愈小。

綜合以上的原因，造成迭代後的封包愈來愈不像理想封包。若是我們能夠將濾波器組的頻譜響應完全補償回來，並在副頻帶訊號的相位加上一些限制，也許可以使 IEC 達到我們想要的效果，然而這些方法必須再往相關的領域深入研究才行。

5.2 頻譜模糊化模型

基於 STFT 會有保留原相位頻譜的問題，此外考慮到必須與響度模型整合成為單一 STFT 架構或單一濾波器組架構，因此我們選用了單一濾波器組架構來建立我們的模型；在這樣的架構之下，我們要思考的問題是：如何從副頻帶訊號得到頻譜模糊化的訊號？

實際情況之下的頻譜模糊化，是因為耳蝸內毛細胞的分頻解析度降低，造成接收聲音的頻率範圍變大，使得雜訊變多，以及分不清楚音高等等；然而在模擬的情形中，我們不可能改變正常人耳蝸聽覺細胞的接收程度，也就是說，我們必須把聽損患者聽到的，多出來的頻率的成份，以某種方式「載」到正常人能夠聽到頻率範圍，來達到頻譜模糊的效果，這是模擬系統與真實的頻譜模糊化一定會有的差異。

5.2.1 演算法概念

根據以上討論的模擬系統的限制，我們採用與 Moore 團隊相同的頻譜模糊化的概念，並將單一的頻率用濾波器組來取代，完整的說明為：根據一特定濾波器變寬的程度，計算其它的濾波器貢獻給該變寬濾波器的增益，並將其它濾波器的訊號能量，依據計算出來的增益，「載」到該濾波器的原始載波上，成為一個被模糊化後的訊號。

以生理現象來解釋此模糊化的概念：一特定變寬的濾波器代表某一群內毛細胞對頻率的敏感度降低，而這群內毛細胞的頻率響應，可以看成是位在基底膜上其它位置的正常內毛細胞的頻率響應，以不同的線性組合而成；因此在計算該濾波器內的能量時，我們

必須將其它內毛細胞產生的響應也一併考慮進去，才能達到模糊化的效果。

最後以小波轉換的理論來詮釋，並將此概念轉換成工程角度上相對應的處理。一聲音訊號被濾波器組分別濾波之後，得到的副頻帶訊號的能量，可以視為該聲音訊號在某一基底（小波）上的投影量，代表該聲音訊號與該小波頻率響應的相似程度；而我們根據該小波頻率響應變寬的程度，計算其它小波對於此變寬小波的貢獻，並計算出模糊化後的能量值；最後我們將係數乘上基底，將副頻帶訊號合成為一模糊化後的聲音訊號。

5.2.2 演算法實作詳述

模糊化演算法的流程圖如 44，以下將詳述其中的每一個部份：

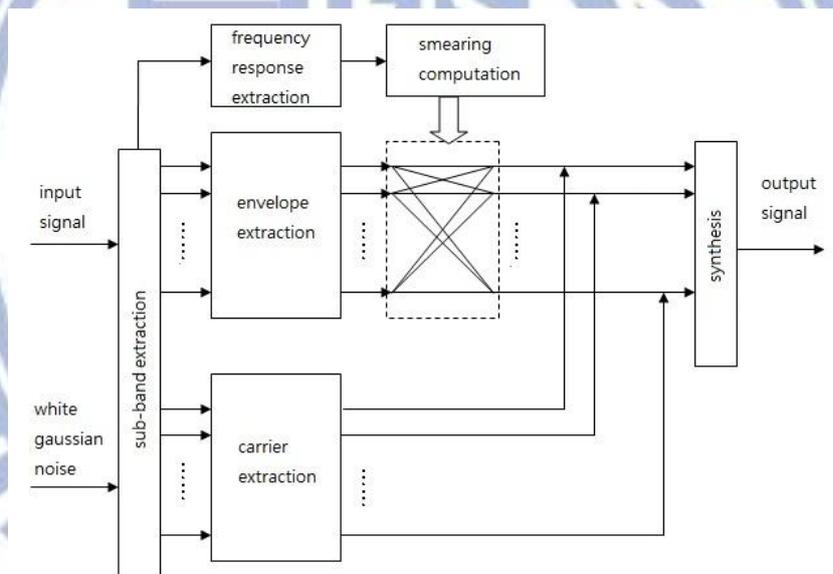


圖 44 頻譜模糊化演算法流程圖

1. 模糊化增益計算

求出濾波器組的頻率響應之後，針對其低頻帶與高頻帶的響應，分別以式 (10) 中的不同 p 值的 roex filter 最其做最佳化，找到相對應的 p 值；並根據該 p 值與頻譜模糊化的參數，計算出變寬的濾波器的頻率響應。

接著我們要計算這個變寬的濾波器頻率響應，是由原始的濾波器組頻率響應，以什麼樣的線性組合組合而成；在此我們將問題公式化，以利於使用最佳化的方式求解，公式如下：

$$\begin{aligned} & \min_x |Ax - b|^2, \\ & \text{subject to } 0 \leq x \leq 1 \\ & A \in \mathbb{R}^{8000 \times 128}, x \in \mathbb{R}^{128 \times 1}, b \in \mathbb{R}^{8000 \times 1} \end{aligned} \quad (25)$$

其中 A 的每一欄代表原始濾波器的頻率響應 (0~7999Hz)、x 為每一個原始濾波器組的增益、而 b 則為變寬的濾波器的頻率響應。透過式 (25)，並以 matlab 中內建的 quadratic programming 求解後，代表性的結果如圖 45。我們可以發現，除了在邊界會因為本身濾波器總響應的不完美，而稍微有一點誤差之外，在大部份的情形之下，我們都可以將變寬的濾波器，以原先濾波器的線性組合取代。將 128 個濾波器組分別經過以上的最佳化運算之後，我們得到一個 128*128 的矩陣，代表濾波器組互相影響的程度。

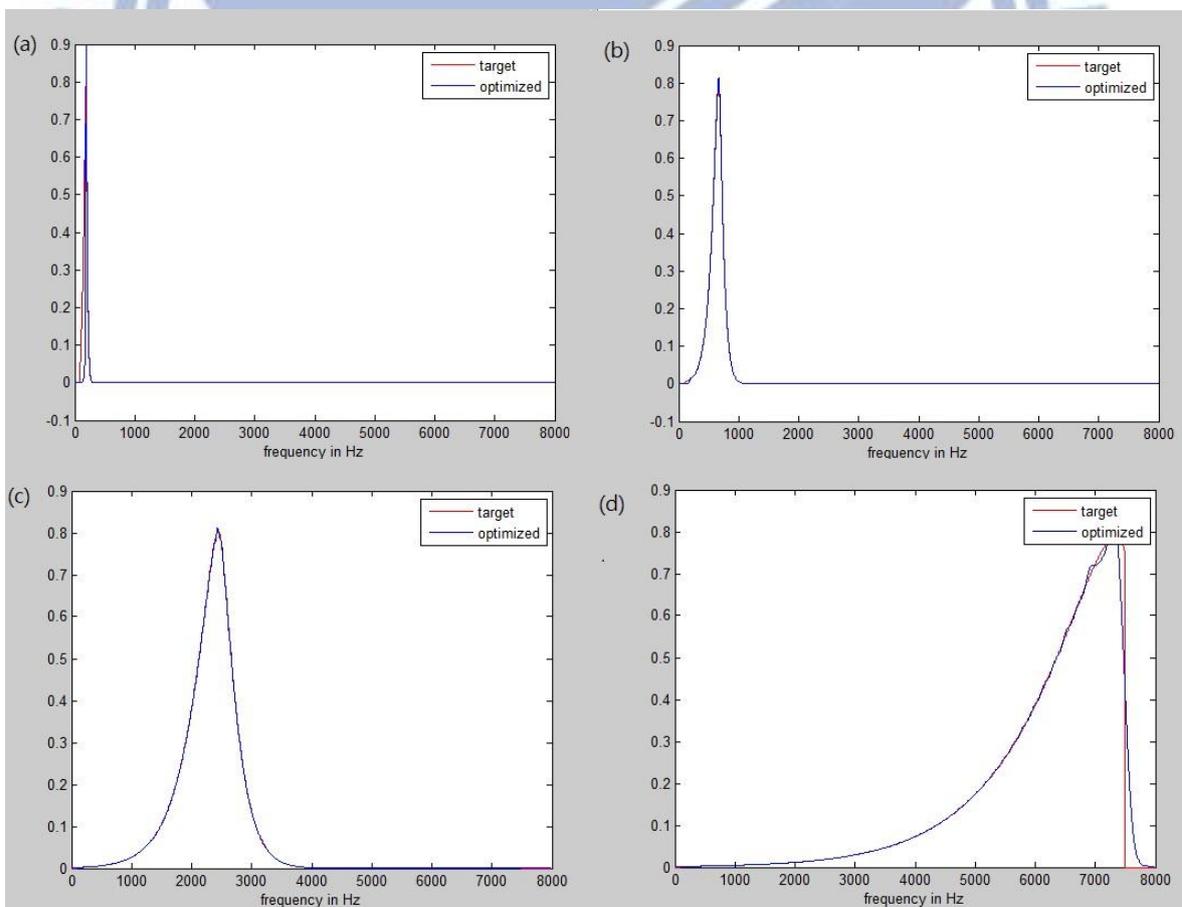


圖 45 由 a~d 分別為第 1、45、90、128 個變寬 6 倍的 channel 最佳化結果

2. 載波計算

有了副頻帶訊號的能量之後，我們要將能量載到對應的載波之上，才能真正達到模糊化的效果，因此我們要有相對應的載波才行。首先考慮的是使用原訊號的載波，然而此方法不論是用於單頻訊號或是寬頻訊號都不適合，說明如下：

- 對於單頻訊號來說，每一個副頻帶訊號的載波都與原來的單頻訊號相同，因此無論再怎麼使用不同頻帶的能量做組合，合成回來的訊號終究還是原來的單頻訊號，並不會有模糊的效果發生，如圖 46 (a)、(b)。
- 對於寬頻訊號而言，因為每一個頻率都有成份，因此每一個副頻帶訊號的載波，其頻率響應會非常接近該濾波器的頻率響應；然而保留原訊號的載波會產生的問題與第三章中 Moore 團隊保留原相位頻譜類似，也就是原來訊號的聲譜結構會被保留下來；因此造成了雖然每一個頻帶有雜訊加入，但是音調的泛頻並沒有變寬，如圖 46(c)、(d)。

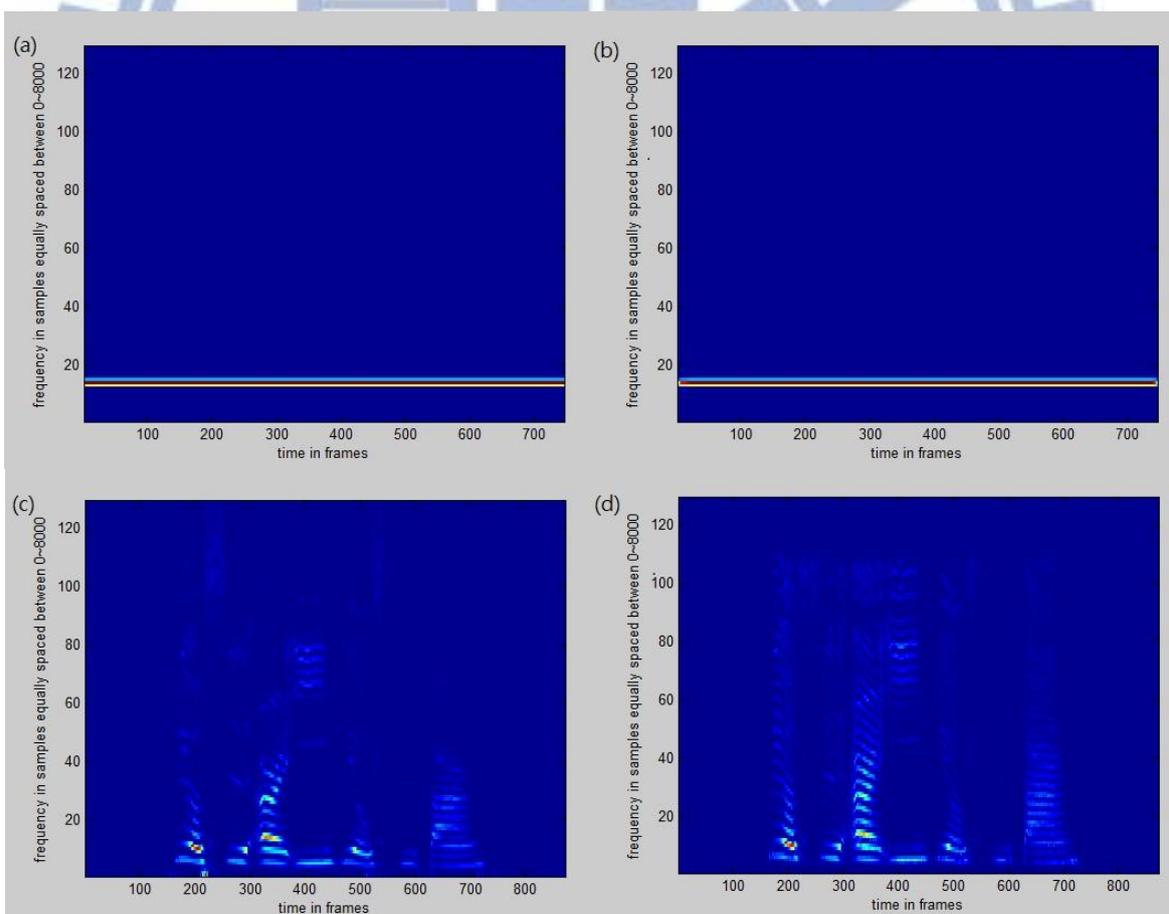


圖 46 加上原訊號載波得到的聲譜圖

為了解決以上的問題，我們將白高斯雜訊通過該濾波器組後產生副頻帶訊號，取出其載波，用於合成訊號；經過模擬發現，這樣子產生的載波的平均頻譜與濾波器組的頻率響應是非常一致的。

5.2.3 實作結果與分析

圖 47 分別為原訊號聲譜與模糊化 1.5、3、6 倍的聲譜圖，觀察之後分析得到結論如下：

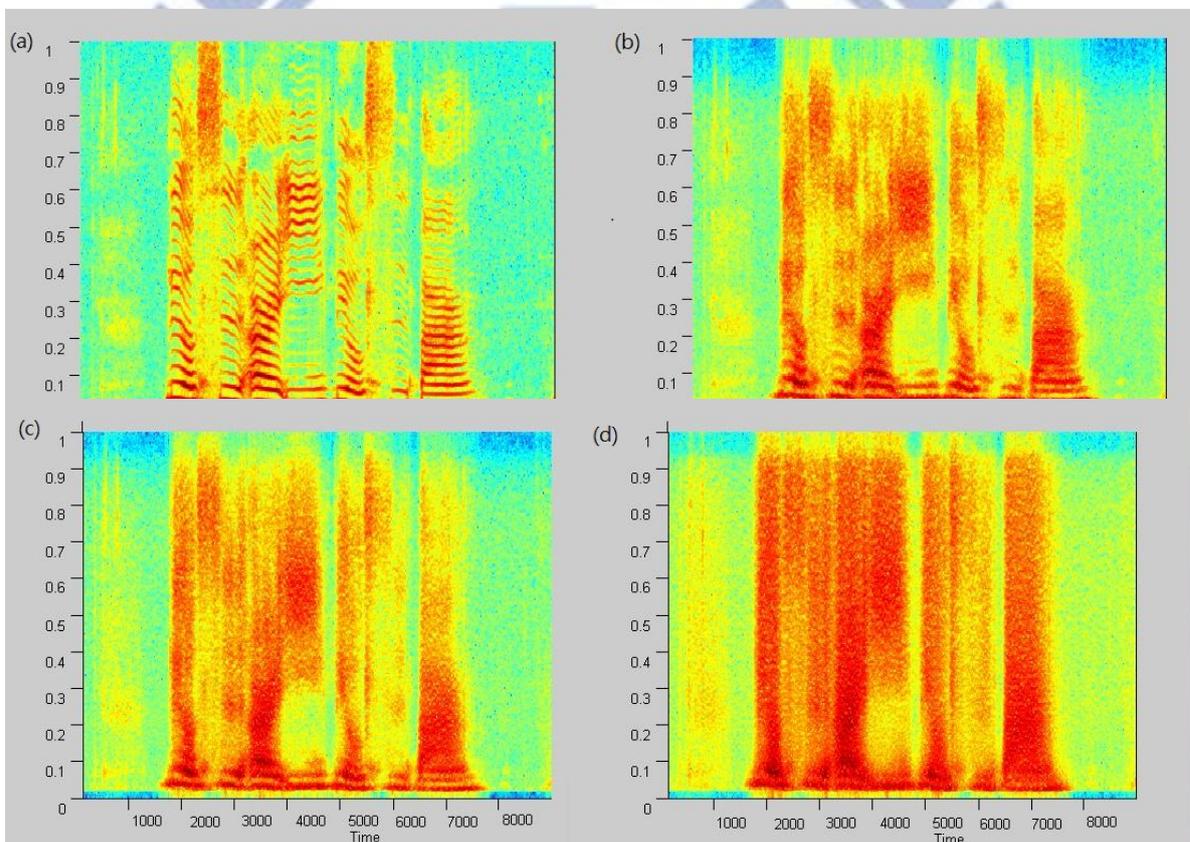


圖 47 a~d 分別為原訊號、模糊 1.5、3、6 倍的聲譜圖

- 當頻譜模糊化只有 1.5 或 3 倍時，我們大概都還可以看到聲音的音高及泛頻，然而當頻譜模糊化 6 倍時，幾乎所有的頻譜都混在一起，音色非常混雜，而這個結果也藉由聽覺測試得到證實。
- 泛頻只有出現在低頻的部份，這是因為我們選定的濾波器組是依照人耳頻率分佈而設

計的，低頻的濾波器較窄，因此在一個濾波器之內可能只有一個泛頻，因此即使經過模糊化的處理，還是可以看出此泛頻；而高頻的濾波器較寬，往往同一個濾波器就包含了好幾個泛頻，而這些泛頻的總能量已經被當作是這個濾波器頻率響應（小波）組成聲音訊號的係數；因此在合成訊號之後，該濾波器內已經不包含泛頻的資訊了。

- 從圖 47 中可以看到，隨著頻譜模糊化程度的加劇，音高的泛頻變得愈來愈寬，雜訊也愈來愈多，因此我們推斷這個演算法是有效果的。

5.2.4 混合模型

由於我們提出的響度模型與頻譜模糊化模型，都只有改變副頻帶訊號的能量而已，因此我們可以很容易的將兩個演算法結合在一起，如圖 48：

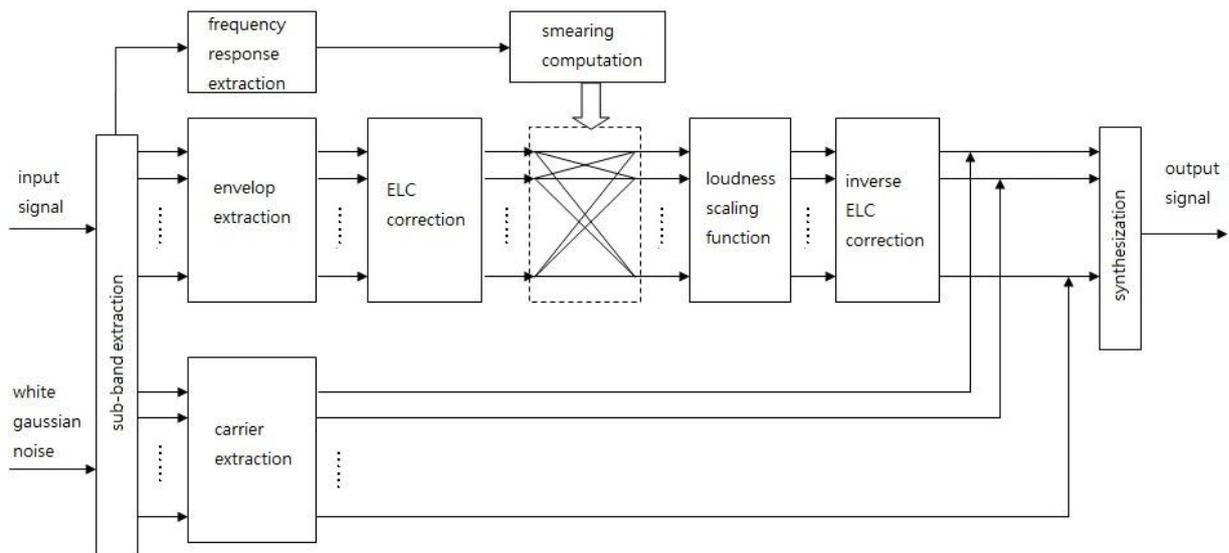


圖 48 混合模型流程圖

在這個演算法中我們依然保留了 ELC 的增益，不過為了符合濾波器組的架構，對於每一個濾波器我們以其中心頻率對於 1KHz 的等響度曲線的強度差值，來代表該副頻帶訊號的增益值，與原先 Moore 團隊以 FFT 方式改變增益的方式不同。

5.3 取樣頻率與載波比較

在我們的模型發展過程中，因為是由 Moore 的模型加以延伸得來的，因此一開始設計的取樣頻率是 16KHz，然而我們使用的濾波器組會依據取樣頻率的不同，而有不同的中心頻率與頻帶，因此我們在本節中會針對取樣頻率的不同加以比較；除此之外，雖然利用白高斯雜訊來產生載波，可以避免因為保留原訊號載波，而造成聽損程度減小的現象，然而可能造成的問題是計算複雜度變高，也可能使得合成的訊號品質較不好，因此我們也將幾種方法實作比較之。

5.3.1 取樣頻率比較

當取樣頻率為 16KHz 時，我們使用的聽覺濾波器組的第一個濾波器的中心頻率約為 190Hz，也就是說太低的頻率成份會被濾掉；這樣子的濾波器組在處理男性的聲音訊號的時候，可能會有基頻消失的現象，因為我們重建聲音訊號時是以 128 個濾波器組的能量乘上其載波之後合成；基頻被濾掉的現象會影響聲音訊號語調的高低起伏，使得聲音聽起來比較像機械音。

如圖 49 所示，左圖是原來的聲音訊號聲譜圖，原始音檔的取樣頻率為 8KHz，在此我們將之升頻為 16KHz，在畫圖的時候為了看得較為仔細，因此只畫出 0~4KHz 範圍的圖；而右圖則是將頻譜模糊化 3 倍後的結果。從圖上我們可以清楚的發現，左圖中語音的基頻，在右圖中完全看不到了，實際聽過後的確會有機械音的現象。

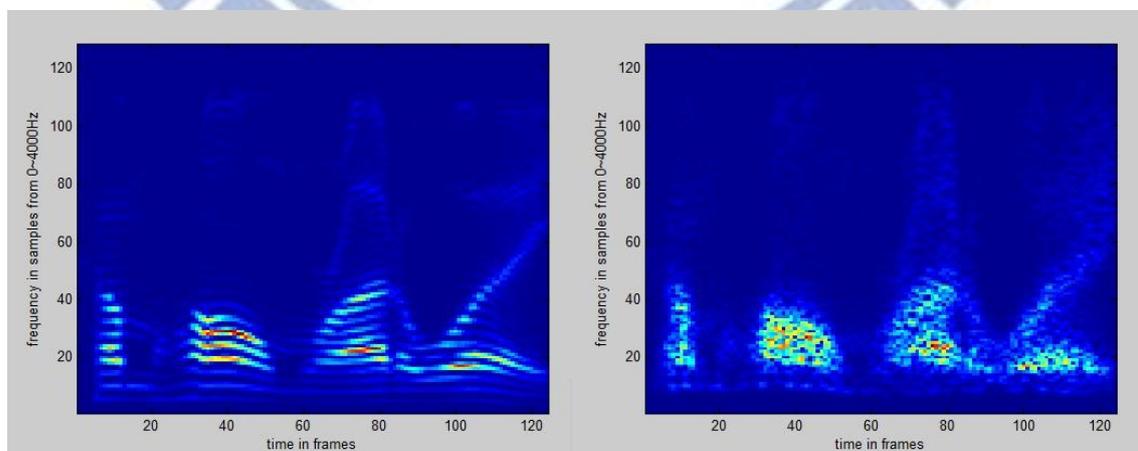


圖 49 16KHz 取樣環境下，聽損模型的失真現象

因此我們將取樣頻率修改為 8KHz，如此一來第一個濾波器的中心頻率大約為 85Hz，落在一般男性的基頻範圍中，因此我們期望在 8KHz 取樣頻率的發展環境下，可以得到較佳的語音品質，如圖 50 所示：圖 50 與圖 49 除了取樣頻率不同之外，其餘的條件均相同，我們可以發現語音的基頻並沒有消失，而且實際比較圖 49（右）與圖 50（右）的聲音之後，可以明顯發現圖 50（右）的聲音的抑揚頓挫較為明顯，因此我們認為使用 8KHz 的取樣在我們的濾波器組下，來進行聽損現象的模型是較為合適的。

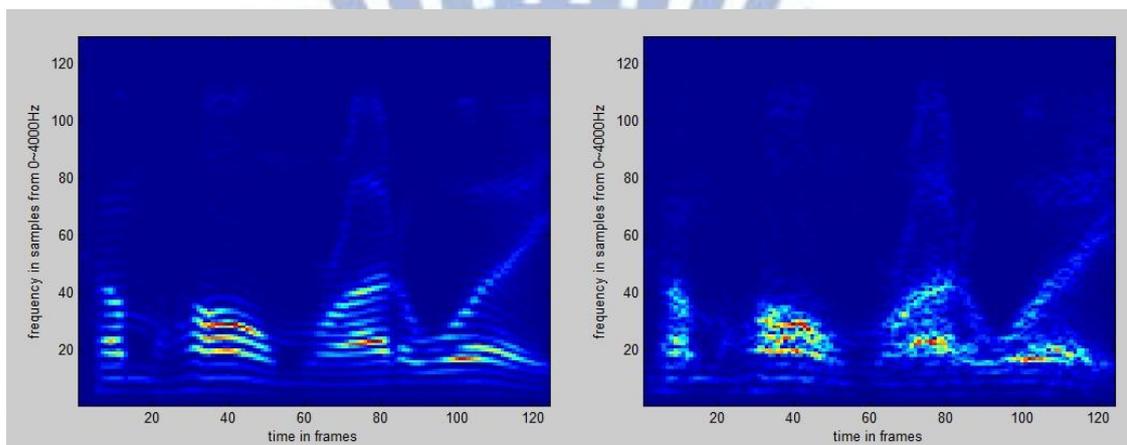


圖 50 8KHz 取樣頻率下，聽損模型的失真現象

5.3.2 載波比較

將白高斯雜訊濾波過後取出載波的做法，可能會造成的影響是計算複雜度變高，以及使得合成的語音品質較差，因此我們分別實驗以下兩個產生載波的方法，並比較其與原來的方法的語音品質。

- 為了避免濾波的過程，我們用以濾波器中心頻率為頻率的單頻弦波，搭配一個隨機相位，來做為這個濾波器的載波；這樣子的做法產生的聲譜圖和原先做法的聲譜圖幾乎相同，然而仔細聽聲音後會發現這樣子合成出的聲音會有單音的成份在，因此語音品質是較為不佳的。
- 在過去發展聽覺模型的研究中，以迭代方代尋找相位的演算法曾被提出 [53]，此演算法的優點是我們只要有每一個濾波器的能量，就可以用迭代的方式來找到對應的相

位；我們比較用此演算法實作出來的結果、與 5.2 節提出來的的方法之後，發現以原來提出的方法合成出來的聲音，聽起來有些破破的，好像有磨喉嚨的感覺，相較之下以迭代方式得到的結果較佳，不過從辨識度的角度來看，兩者幾乎是相同的。

5.4 實驗設計與結果分析

5.4.1 實驗一

1. 實驗目的與設計

若是我們想要驗證聽損模型的效果，最理想的方法就是測試正常人聽經過處理的聲音的辨識率，以及測試聽損患者聽正常聲音的辨識率，並比較兩組數據的相似程度 [17]；然而由於在現階段中並沒有聽損患者可以協助測試，因此我們只能透過與其它模型比較數據的相似程度，簡單的驗證聽損模型的成效。

在這個實驗中我們將以 [54] 中提到的 6 個中文語料列表 (圖 49)，進行中文單字音素 (phones) 的聽力測驗，並與 Moore 團隊的混合模型做比較；語料的條件與 LSD (Latin Square Design) 設計如表 2、3，其中 B6R3 代表聽覺濾波器變寬 (Broaden) 6 倍，響度聚集 (Recruitment) 3 倍的聽損現象，是模擬重度聽損的情形；在此條件之下將我們將聲音加上不同大小的雜訊 (clean、0dB S/N、-3dB S/B) 來進行辨識率的測試。

除此之外，由於演算法與真正聽到的聲音音量有關係，因此在實驗中我們產生一個用來校準的白高斯雜訊，並讓經過 VAD (voice activity detection) 處理後的語料的平均能量大小與該校準訊號相同，接著再以此產生符合表 2 條件的白高斯雜訊，與聲音訊號相加。在實際測試的時候，我們將校準訊號的聲音大小調整至大約 65 分貝，讓人稍微感覺到吵雜的程度，並在相同的條件之下讓所有受試者完成測試。

No.	List A1		List A2		List A3		List B1		List B2		List B3	
	Item	Word										
1	zhi1	知	di4	地	zhi1	知	jing1	經	shi4	是	yi3	以
2	shang4	上	yu2	魚	he2	河	yuan2	元	cheng2	成	cheng2	成
3	hou4	後	jian4	建	yu2	魚	yu3	雨	ji4	計	ru2	如
4	neng2	能	zi4	字	ji3	擠	qi2	其	yu3	雨	si4	四
5	jin4	進	shuo1	說	ying4	硬	xia4	下	bian4	便	du4	度
6	qu4	去	xiao3	小	wu2	吳	jiang4	降	du4	度	cong2	從
7	mu4	木	kan4	看	jie2	結	bian4	便	fei1	飛	ling4	另
8	xian1	先	mu4	木	ban4	半	si4	四	ying3	影	lu4	綠
9	jie3	姐	tou2	頭	chu4	觸	fei1	飛	xi2	習	jian3	檢
10	diao4	掉	jie3	姐	zhan3	展	kao3	考	ru4	入	gei3	給
11	ying2	營	feng1	風	qi1	七	guang1	光	jiao1	交	jun1	軍
12	chu2	除	zhan3	展	diao4	掉	liao4	料	qiang2	強	zhuān1	專
13	li2	梨	gai1	該	xin4	信	tan2	談	liu4	六	yang2	陽
14	wei3	偉	chu2	除	pian4	片	huan1	歡	kuang4	礦	tui1	推
15	ge1	歌	a1	阿	zeng1	增	you1	優	mian3	免	xi3	洗
16	ban3	板	yin3	引	tuo1	脫	cu4	促	tan1	貪	kuang4	礦
17	fan2	凡	zhen4	鎮	long2	龍	re4	熱	zhui1	追	mao4	貿
18	pi2	皮	sui2	隨	gang3	港	xian2	賢	guan4	冠	bang4	棒
19	ao4	傲	ni2	尼	wei1	威	chu3	楚	lun2	輪	pian1	偏
20	kong3	孔	bang1	幫	fa2	罰	zhui1	追	pao3	跑	shun4	順
21	za2	雜	pi2	皮	kao4	靠	shun4	順	song1	松	fan1	翻
22	zhuo1	桌	long2	籠	shou2	熟	dong3	懂	cang2	藏	yao2	搖
23	ai1	哀	ao4	傲	nai4	耐	meng2	蒙	xuan2	玄	qiu1	秋
24	ti4	替	he1	喝	mi3	米	lang2	郎	o1	喔	xuan2	玄
25	sen1	森	qing4	慶	sen1	森	pi4	僻	he4	賀	he4	賀

Note: The items of the word lists are represented by traditional Chinese characters, and their Romanization is in the Hanyu Pinyin system.

圖 51 聽力測試語料表

資料來源：〔54〕

表 2 聽力測試的六個條件

	B6R3, clean	B6R3, 0dB S/N	B6R3, -3dB S/N
Moore's method	甲	丙	戊
Our method	乙	丁	己

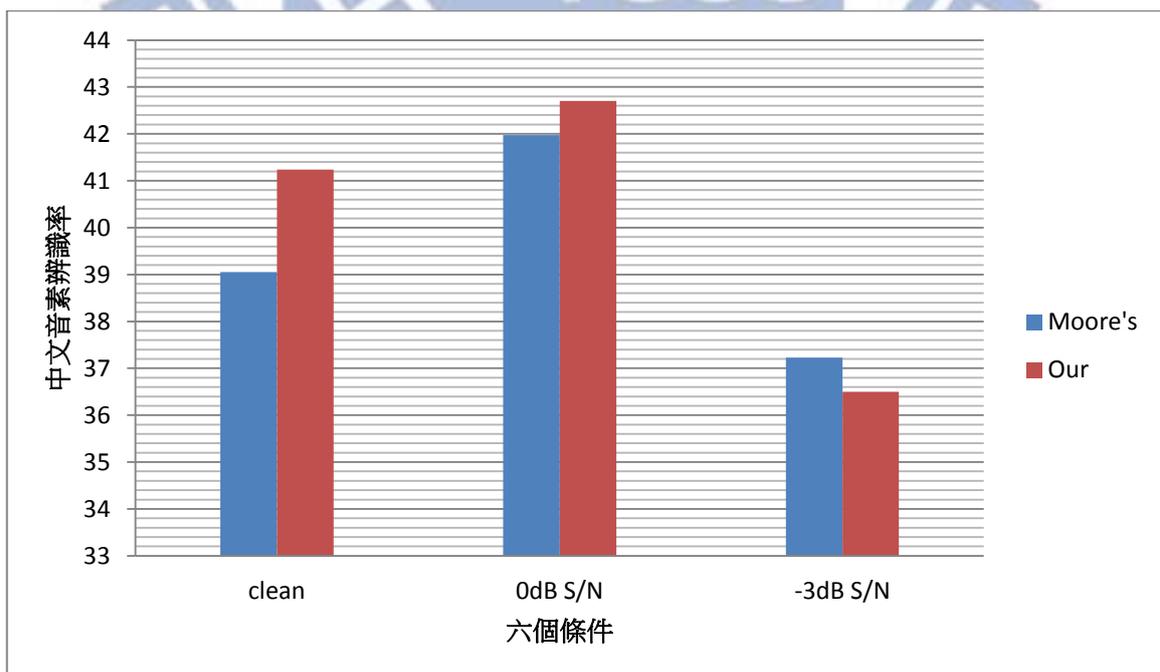
表 3 LSD 設計

Data list	A1	A2	A3	B1	B2	B3
subjects						
S_1	甲	乙	己	丙	戌	丁
S_2	乙	丙	甲	丁	己	戌
S_3	丙	丁	乙	戌	甲	己
S_4	丁	戌	丙	己	乙	甲
S_5	戌	己	丁	甲	丙	乙
S_6	己	甲	戌	乙	丁	丙

2. 實驗結果分析：

辨識率的計算方式是將中文字拆成聲母及韻母（音素），並以聲母或韻母的正確與否為辨識的最小單位，將所有受試者每個條件下的答對數目相加後，除以在該條件之下所有語料的聲母韻母總和，其結果如表 4：

表 4 在 B6R3 的條件下，兩種演算法的辨識率



上表中，我們發現若是不考慮乾淨語音的話，辨識率確實會隨著 SNR 下降而下降；對於乾淨的語音而言，我們原先期望辨識率會比較好，然而在這裡不論是那一種模型，辨識率都會比 0dB S/N 還要低一些，推測可能原因為：響度聚集的主要影響是讓聲音大小聲的差距變大，而對於乾淨的語音來說，在聲音開始與結束的時候，相較於加入雜訊的語音，聲音的大小聲會有較明顯的變化，而頻譜模糊化則是會讓原本乾淨的語音產生雜訊；此外，由於中文字只由包含開頭的聲母和結尾的韻母構成，若是在中文字開頭與結尾的地方被嚴重干擾的話，可能就會造成辨識率比其它兩個條件還要糟的狀況發生。

從以上三種情況的實驗結果來看，我們的模型模擬成果與 Moore 的成果幾乎有相同的結果，因此我們的聽損模型與 Moore 團隊的模型可謂一致，而其最終驗證將待未來正式進行聽損患者的聽力測試來呈現。

5.4.2 實驗二

1. 實驗目的與設計

在實驗二中，我們想要驗證在我們提出的模型之下，不同的受損程度確實會造成辨識率的差異；因此，我們比較在模擬重度受損（B6R3）與中度受損（B3R2）的情況之下，中文單字的音素辨識率的差異。

實驗設計部份，我們沿用實驗一中，在我們提出的模型之下測得的數據，並針對 B3R2 的情形另外設計聽力測試，如表 5、6：

表 5 聽力測試的三個條件

	B3R2, clean	B3R2, 0dB S/N	B3R2, -3dB S/N
Our method	庚	辛	壬

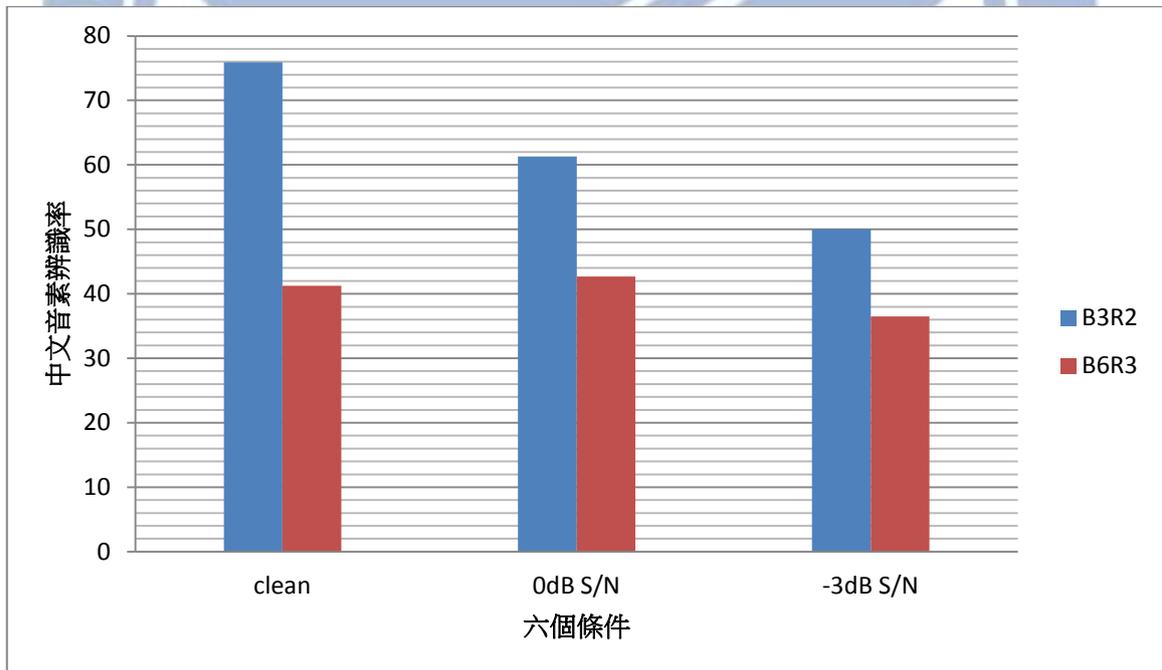
表 6 LSD 設計

Data list	A1	A2	A3	B1	B2	B3
subjects						
S_7	庚	辛	壬	庚	辛	壬
S_8	壬	庚	辛	壬	庚	辛
S_9	辛	壬	庚	辛	壬	庚

比較表 3、4 與表 5、6 後可以發現，針對 B6R3 的條件，我們讓 6 個受試者進行聽力測驗，每個受試者聽一組語料；而針對 B3R2 的條件，我們則是讓 3 個受試者進行測驗，每個受試者聽二組語料，因此每一個條件之下的總音素數目是相同的。

實驗結果如下表 7 所示：

表 7 不同受損情形之下的辨識率比較



從表 7 中可以明顯發現，B3R2 情形下的辨識率確實比 B6R3 還要好；除此之外，乾淨語音條件下的辨識率也比有加雜訊條件之下的辨識率還好，推測原因可能是因為在中度受損的情形之下，語音產生的雜訊與失真較少的緣故，因此在 B3R2 的情形之下，辨識率會隨著雜訊的增加而降低。

另外，我們可以根據表 7 中每個雜訊條件下，B3R2 與 B6R3 辨識率的差值推測：受損程度影響辨識率的程度，隨著雜訊愈大而愈不明顯；這樣的推測也可能與實驗一中討論的原因相同——頻譜模糊化與響度聚集對於乾淨語音有比較明顯的影響。

5.5 結論

Moore 團隊原先的聽損耳蝸模型，只是為了確認響度聚集、分解解析度降低…等等聽損現象，確實會造成辨識率的降低；然而我們希望能夠建立一個模擬聽損患者的耳蝸模型，因此對 Moore 的模型進行深入的分析之後，針對一些工程上的問題提出較佳的做法，不但解決了保留原訊號相位產生的聽損程度不匹配的問題，也將整個演算法整合到濾波器組的架構之下，減少計算複雜度，並在訊號處理上多考慮了人耳對頻率的解析度，讓整個模型更趨於完整。

此外，透過本章各節的分析，我們也驗證了我們提出的模型確實符合聽損患者回報出來的一些特性，如下所述：

- 聲音的大小聲變化較為劇烈。
- 語音的音色改變。
- 雜訊較多。
- 高度受損者比中度受損者更聽不清楚聲音。

基於以上幾點，我們認為我們提出的模型確實可以用來模擬聽損患者的聽損狀況。

第六章 未來展望

在實作出一個初步模擬聽損耳蝸的模型之後，未來的研究方向可以分為四個，以下分別說明之。

1. 加入更多聽損現象，使模型更趨完整

由於聽損現象都是彼此相關的，因此如果沒有考慮周到的話，在實際比較病患的聽損程度以及模擬的聽損程度時，量測結果可能被這些沒有考慮到的因素左右，因此造成我們在調整參數時有程度過少或程度過多的現象；因此我們可以參考〔55〕中提到的有關時域封包模糊化的現象，或是學者針對其它聽損現象所做的研究，再加入一些參數到我們的模型中。

2. 建立符合各種不同聽損患者的個人化模型

為了達到以上的目的，因此我們必須要有聽損患者加入研究才行；如此一來，不但可以讓我們驗證演算法的效果，也有助於參數的選定，舉例如下：

- 測量不同聽損患者的聽損程度，並根據測量結果調整模型參數後，讓聽損患者以及正常人分別以原始語料、經處理語料做聽力測試，並比較兩組數據間的相似程度。
- 不僅僅只是測量語句可辨別度之間的相似性，還可以針對語音品質做調整；例如在 5.3 節當中討論到的，用迭代方式產生載波的合成方法，會比用白高斯雜訊濾波得到載波的合成方法，有較好的語音品質；然而這些語音品質都是正常人的感覺，並不能代表聽損患者聽到的聲音，因此透過聽損患者，我們可以進一步加強語音品質，使得經過此系統的聲音能更接近聽損患者聽到的聲音。

3. 幫助助聽器的演算法的發展

Moore 團隊在 [15-16] 中，不只驗證了響度聚集、最小可聽水平提升、以及分頻解析度降低對於辨識率會有影響，也加入了 NAL prescription 來驗證該模型是否能反應出 NAL 的特性—根據不同頻帶的響度衰減程度，加上適當的增益，來補償因為最小可聽水平提升造成的聽力損失。

我們可以先參考 Moore 團隊的做法，將 NAL 加入我們的模型之中，觀察是否會得到相似的結果；接著再進一步以經過實際測試，確實有助於聽覺濾波器變寬的患者聽力的助聽器演算法 [24]，或者是其它公認有幫助的助聽器演算法，來驗證我們的模型是否可以反應出這些演算法的特性。在經過各種聽損現象的助聽器演算法的測試，並証實我們的演算法確實有效果之後，我們就可以用這套系統來評估發展助聽器演算法了。

4. 發展補償分頻解析度降低的助聽器演算法

如同在緒論中提到的，我們特別注重分頻解析度降低的聽損現象，因此也期望發展出能夠補償此現象造成的聽力損失的演算法；除了參考第四章中所提到的未來方向之外，我們也可以參考 [56] 的有關頻譜銳化 (spectral sharpening) 的想法，並搭配我們發展出的聽損耳蝸模型，加以延伸應用來解決頻譜模糊化的問題。

参考文献

- [1] D. Byrne, and H. Dillon, "The National Acoustic Laboratories'(NAL) new procedure for selecting the gain and frequency response of a hearing aid," *Ear Hear.*, 7, pp. 257-265., 1986.
- [2] E. Villchur, "Simulation of the effect of recruitment on loudness relationships in speech," *J. Acoust. Soc. Am.*, 56, pp. 1601-1611, 1974.
- [3] P. J. Fitzgibbons, and F. L. Wightman, "Gap detection in normal and hearing-impaired listeners," *J. Acoust. Soc. Am.*, 72, pp. 761-765, 1982.
- [4] R. S. Tyler, A. Q. Summerfield, E. J. Wood, and M. A. Fernandes, "Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners," *J. Acoust. Soc. Am.*, 72, pp. 740-752, 1982.
- [5] B. R. Glasberg, B. C. J. Moore, and S. P. Bacon, "Gap detection and masking in hearing-impaired and normal-hearing subjects," *J. Acoust. Soc. Am.*, 81, pp. 1546-1556., 1987.
- [6] B. R. Glasberg, and B. C. J. Moore, "Effects of envelope fluctuations on gap detection," *Hear. Res.*, 64, pp. 81-92., 1992.
- [7] G. Pick, E. F. Evans, and J. P. Wilson, "Frequency resolution in patients with hearing loss of cochlear origin," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), 1977.
- [8] M. Florentine, S. Buus, B. Scharf, and E. Zwicker, "Frequency selectivity in normally-hearing and hearing-impaired observers," *J. Speech Hear. Res.*, 23, pp. 643-669,

1980.

- [9] R. S. Tyler, J. W. Hall, B. R. Glasberg, B. C. J. Moore, and R. D. Patterson, "Auditory filter asymmetry in the hearing impaired," *J. Acoust. Soc. Am.*, 76, pp. 1363-1368, 1984.
- [10] B. R. Glasberg, and B. C. J. Moore, "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," *J. Acoust. Soc. Am.*, 79, pp. 1020-1033, 1986.
- [11] B. C. J. Moore, and B. R. Glasberg, "Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments," *J. Acoust. Soc. Am.*, 80, pp. 93-107, 1986.
- [12] R. S. Tyler, "Frequency resolution in hearing-impaired listeners," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore, (Academic, London), 1986.
- [13] M. Li, H. G. McAllister and N. D. Black, "Multirate Modeling of Human Ear Frequency Resolution for Hearing Aids," *Digital Signal Processing Workshop Proceedings (IEEE)*, pp. 157-160, 1996.
- [14] T. Baer, and B. C. J. Moore, "Effects of spectral smearing on the intelligibility of sentences in the presence of noise," *J. Acoust. Soc. Am.*, 94, pp. 1229-1241, 1993.
- [15] B. C. J. Moore, and B. R. Glasberg, "Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech," *J. Acoust. Soc. Am.*, 94, pp. 2050-2062, 1993.
- [16] Y. Nejime and B. C. J. Moore, "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise," *J. Acoust. Soc. Am.*, 102, pp. 603-615, 1997.
- [17] H. M. Hu, J. Q. Sang, M. Lutman, S. Bleeck, "Simulation of hearing loss using compressive gammachirp auditory filters," *ICASSP*, pp. 22-27, 2011.

- [18] P.M. Sellick, R. Patuzzi, and B. M. Johnstone, "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *J. Acoust. Soc. Am.*, 72, pp. 131-141, 1982.
- [19] M. A. Ruggero, and N. C. Rich, "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," *J. Neurosci*, 11, pp. 1057-1067, 1991.
- [20] T. S. Chi, class notes of Auditory and Acoustical Information Processing, Department of Communication Engineering, National Chiao-Tung University, Taiwan, 2011.
- [21] J. B. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE-ASSP*, 25, pp. 235-238, June 1977.
- [22] D. W. Griffin, and J. S. Lim, "Signal Estimation from Modified Short-Time Fourier Transform," *IEEE-ASSP*, 32, pp.236-243, 1984.
- [23] P. N. Kulkarni, and P. C. Pandey, "Frequency mapping for multi-band frequency compression for improving speech intelligibility," *Proc. 14th National Conference on Communications*, pp. 437-441, 2008
- [24] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, "Multi-band frequency compression for sensorineural hearing impairment," *Proc. 16th Int. Conf. Digital Signal Processing*, Paper S4P.1., 2009.
- [25] T. D. Rossing, *The Science of Sound. Reading*, MA: Addison-Wesley, 1982.
- [26] H. Fletcher, "Auditory patterns," *Rev. Mod. Phys*, 12, pp. 47-61, 1940.
- [27] E. W. Zwicker, "Subdivision of audible frequency range into critical bands (Frequenzgruppen)," *J. Acoust. Soc. Am.*, 33(2), pp. 248, 1961.

- [28] B. R. Glasberg, and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noised data," *Hear. Res.*, 47, pp. 103-138, 1990.
- [29] H. Fletcher, *Speech and Hearing in Communication*, 2nd ed., Bell Telephone Laboratories Series, Van Nostrand, Princeton, NJ , 1953.
- [30] S. S. Stevens, "The Measurement of Loudness," *J. Acoust. Soc. Am*, 27, pp. 815-829, 1955.
- [31] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-time Signal Processing*, 2nd ed., Prentice Hall Signal Processing Series, January 10 1999.
- [32] R. E. Cyochicre, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE-ASSP*, 28, pp. 99-102, February 1980.
- [33] C. Valens, *A Really Friendly Guide to Wavelets*.
<http://polyvalens.pagesperso-orange.fr/clemens/wavelets/wavelets.html#note1>
- [34] R. Polikar, *The Wavelet Tutorial*.
http://person.hst.aau.dk/enk/ST8/wavelet_tutorial.pdf
- [35] R. D. Patterson, and A. Cutler, "Auditory preprocessing and recognition of speech," in *Research Directions in Cognitive Science, Vol.I Cognitive Psychology*, edited by A.D. Baddeley and N. O. Bernsen, (Erlbaum, London), 1989.
- [36] E. de Boer, and C. Kruidenier, "On ringing limits of the auditory periphery," *Biol. Cybern.*, 63, pp. 433-442, 1990.
- [37] T. Irino, and R. D. Patterson, "A Time-Domain, Level-Dependent Auditory Filter: The Gammachirp," *JASA*, 101, pp. 412-419, 1997.
- [38] T. Chi, P. Ru and S. Shamma, "Multiresolution spectro-temporal analysis of complex

sounds,” JASA, 118, pp. 887-906, August 2005.

- [39] G. Kubin, and W. B. Kleijn, “On speech coding in a perceptual domain,” Proc. ICASSP, pp, 205-208, 1999.
- [40] B. R. Glasberg, and B. C. Moore, “Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,” J. Acoust. Soc. Am, 79, pp. 1020–1033, 1986.
- [41] B. C. J. Moore, and B. R. Glasberg, “Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments,” J. Acoust. Soc. Am, 80, pp. 93–107, 1986.
- [42] B. C. J. Moore, B. R. Glasberg, and D. A. Vickers, “Factors influencing loudness perception in people with cochlear hearing loss,” in Psychoacoustics, Speech and Hearing Aids, edited by B. Kollmeier, (World Scientific, Singapore), 1996.
- [43] R. D. Patterson, and I. Nimmo-Smith, “Off-frequency listening and auditory-filter asymmetry,” J. Acoust. Soc. Am, 67, pp. 229-245, 1980.
- [44] R. D. Patterson, I. Nimmo-Smith, D. L. Weber, and R. Milroy, “The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold,” J. Acoust. Soc. Am, 72, pp. 1788-1803, 1982.
- [45] B.C. J. Moore, B. R. Glasberg, R. F. Hess, and J .P. Birchall, “Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears,” J. Acoust. Soc. Am, 77, pp. 1505-1515, 1985.
- [46] D.R. Soderquist, and J.W. Lindsey, “Physiological noise as a masker of low frequencies: the cardiac cycle,” J. Acoust. Soc. Am, 52, pp. 1216-1220, 1972.
- [47] V. Nedzelnitsky, “Sound pressures in the basal turn of the cat cochlea,” J. Acoust. Soc. Am, 68, pp. 1676-1689, 1980.

- [48] T.J. Lynch, V. Nedzelnitsky, and W.T. Peake, "Input impedance of the cochlea in cat," J. Acoust. Soc. Am, 72, pp. 108-130, 1982.
- [49] J.J. Zwislocki, "The role of the external and middle ear in sound transmission," In: D.B. Tower (Ed.), The Nervous System, Vol. 3: Human Communication and its Disorders, Raven Press. New York, 1975.
- [50] K. Yasu, K. Kobayashi, K. Shinohara, M. Hishitani, T. Arai, and Y. Murahara, "Frequency compression of critical band for digital hearing aids," in China-Japan Joint Conf. on Acoustics, pp. 159-162, 2002.
- [51] Y. T. Kuo, T. J. Lin, Y. T. Lee, W. H. Chang, C. W. Liu, and S. T. Young, "Design of ANSI S1.11 filterbank for digital hearingaids," in Proc. ICECS, Dec. 2007
- [52] D. L. Jones, *Quadrature Mirror Filterbanks*.
<http://cnx.org/content/m12770/1.3/>
- [53] T. Chi, and S. Shamma, "Spectrum restoration from multiscale auditory phase singularities by generalized projections," IEEE Transactions on Audio, Speech and Language Processing, vol. 14, no. 4, pp. 1179-1192, July 2006
- [54] K.S. Tsai, L. H. Tseng, C. J. Wu, and S. T. Young, "Development of a Mandarin Monosyllable Recognition Test," Ear and Hearing, 30(1), pp. 90-99, 2009.
- [55] R. Drullman, J.M. Festen, and R. Plomp, "Effect of Temporal Envelope Smearing on Speech Reception," JASA, 95, no. 2, pp. 1053-1064, 1994.
- [56] A. Schaub, and P. Straub, "Spectral sharpening for speech enhancement/noise reduction," Proc. of the International Conference on Acoustics, Speech and Signal Processing, pp. 993-996, 1991.