

國立成功大學
電機工程學系
碩士論文

MPEG-1 Layer III 音訊編碼器於低位元率之改良
Low Data Rate MPEG-1 Layer III Audio Coder
Enhancement

研究生：黃國祥 Student: Kuo-Hsiang Huang
指導教授：楊家輝 Advisor: Jar-Ferr Yang

Department of Electrical Engineering
National Cheng Kung University
Tainan, Taiwan, R.O.C.
Thesis for Master of Science
June 2002

中華民國九十一年六月

國立成功大學

碩士論文

MPEG-1 Layer III 音訊編碼器於
低位元率之改良

研究生：黃國祥

本論文業經審查及口試合格特此證明
論文考試委員

楊家輝
吳宏食

劉修麟
郭鍾榮

指導教授：楊家輝

系(所)主管：劉修麟

中華民國九十一年六月二十六日

Low Data Rate MPEG-1 Layer III
Enhancement

by

Kuo-Hsiang Huang

A thesis submitted to the graduate division in partial
fulfillment of the requirement for the degree of
Master of Science

at

National Cheng Kung University
Tainan, Taiwan, Republic of China

June, 2002

Approved by :

Jaffer Yang

Jueh-Lin Lian

Chy-Hsien Wu

Chung Jung Kuo

Advisor :

Jaffer Yang

Chairman :

[Signature]

MPEG-1 Layer III 音訊編碼器於低位元率之改良

黃國祥* 楊家輝**

國立成功大學 電機工程研究所 通訊組

摘要

本論文主要是以 MPEG-1 Layer III (MP3) 音訊壓縮編碼器之實現與研究為主，並針對 MP3 在低位元率時的缺失，提出編碼效能分析及音質的改善。首先，我們利用低通濾波器與適應性頻譜截斷法將 MP3 低位元率編碼的缺點加以修正；然後，再針對上述方法之缺點加以改良，提出高頻補償法以補償高頻缺失的部分，使得 MP3 音訊壓縮編碼器於低位元率的編碼效能及音質更加完善。

*作者

**指導教授

Low Data Rate MPEG-1 Layer III Enhancement

Kuo-Hsiang Huang^{*} and *Jar-Ferr Yang*^{**}

Department of Electrical Engineering,
National Chen Kung University
1 University Road, Tainan, Taiwan, R.O.C.

ABSTRACT

The research of this thesis is mainly focused on study and implementation of MPEG-1 Layer III (MP3) for low bit-rate applications. First, we propose a lowpass filter approach and its adaptive spectrum reduction method to reduce the defeat of the MP3 codec in low bit-rate applications. Both lowpass and spectrum reduction algorithms will sacrifice the high frequency components of audio signals. To improve the high frequency performance, we further suggest a high frequency compensation method to make up the quasi-harmonics of high frequency spectrum. By using spectrum reduction and quasi-harmonic compensation method, the MPEG-1 Layer III audio coder can achieve a high perceptual quality for low bit-rate applications.

*The Author

**The Advisor

誌 謝

首先感謝指導教授楊家輝老師細心的指導及教誨，以及提供良好的環境與充足的經費，讓我可以專心地完成本論文，在此致上我二十萬分的感謝。另外感謝聲音組的學長陳福坤、陳奇宏及同學林宏偉和陳桂雪學妹所給予之指導與鼓勵，還有林榮三、蔡淑敏、王木良、顏毓良、張榮憲、連桂宏等語音組學長姐和同學林裕斌、黃一展在課業上的討論與幫忙；以及郭晉魁、黃慶祥老師、曾昭雄學長在 C6X 及計畫方面所給予之協助；同時也要感謝實驗室中的學長郝樹聲、李豫輝、張世承、魏永強、陳慶勻、涂育光、康育嘉，學姊鄭玲兒，同學張富貴、林育加，學弟江翼廷、李季峰、沈奕農，在這兩年來不論在生活上或課業上給我的協助或幫忙，使的我可以順順利利、快快樂樂地完成碩士班的學業。

還要感謝我的家人，他們從小到大無微不至的照顧以及關懷，他們永遠是我最深愛的人。最後要感謝我的女朋友在這兩年來所給予我的支持與鼓勵，以及平時的照顧。還有所有關心我的人，真的謝謝你們。

目 錄

摘要.....	i
ABSTRACT.....	ii
誌 謝.....	iii
目 錄.....	iv
表目錄.....	vii
圖目錄.....	viii
第一章 數位音訊編碼簡介.....	1
1.1 簡介.....	1
1.2 感知式音訊編碼.....	2
1.3 聽覺感知.....	3
A. 評鑑頻帶.....	4
B. 絕對臨界與聲音壓力度.....	6
C. 遮罩效應.....	7
C-1 頻域遮罩.....	8
C-2 時域遮罩.....	9
1.4 時頻轉換.....	10
1.5 量化.....	11

1.6 非失真編碼	12
第二章 MPEG-1 Layer III 編解碼器	13
2.1 簡介.....	13
2.2 MPEG-1 Layer III 編碼器.....	16
A. 心理音響模型分析	17
B. 混合式分析濾波器組	21
B-1 多項式分析濾波器	22
B-2 修正式離散餘弦轉換	23
C. 疊代迴圈	24
2.3 MPEG-1 Layer III 解碼器.....	27
A. MPEG-1 Layer III 的位元流格式	29
B. 解量化係數	29
C. 混合式合成濾波器組	31
C-1 逆修正式離散餘弦轉換	32
C-2 多項式合成濾波器	32
第三章 MPEG-1 Layer III 在低位元率的效能.....	34
3.1 簡介.....	34
3.2 MPEG-1 Layer III 在低位元率的效能	36

3.3 對 MPEG-1 Layer III 在低位元率的效能改進	38
3.4 頻譜截斷與低通濾波器	43
3.5 適應性頻譜截斷	44
第四章 MPEG-1 Layer III 之高頻補償	48
4.1 簡介.....	48
4.2 高頻補償法	48
A. MPEG-1 Layer III 的位元流格式修改	49
B. 高頻補償法	49
B-1 編碼端的高頻補償	51
B-2 解碼端的高頻補償	55
B-3 高頻補償之修正	56
4.4 實驗結果	57
第五章 結 論.....	60
參考文獻.....	61
作者簡歷.....	63
Vita	64

表目錄

表(1.1) 關鍵頻帶(Bark)和邊界頻率(Hz)	5
表(2.1) 決定量尺因子的位元數	27
表(3.1) 各音訊壓縮編碼比較 (雙聲道)	35
表(3.2) ITU-R 的試聽評估準則	35
表(3.3) 不同位元率量化後的影響	37
表(3.4) 頻率與頻譜及量尺因子頻帶的關係	44

圖目錄

圖(1.1) 感知式音訊編碼器之方塊圖	3
圖(1.2) 頻率 (Hz) 與評鑑頻帶頻寬的對應	6
圖(1.3) 絕對臨界 (dB) 與頻率 (Hz) 關係.....	7
圖(1.4) 頻域遮罩效應	9
圖(1.5) 時域遮罩效應	10
圖(1.6) 一般音訊壓縮 M 通道的濾波器組.....	11
圖(2.1) 國際標準組織架構	13
圖(2.2) MPEG 音訊編碼器基本架構.....	14
圖(2.3) MPEG 音訊解碼器基本架構.....	15
圖(2.4) MPEG-1 Layer III 編碼器架構	16
圖(2.5) 計算區頻帶的能量頻譜與加權不可預測性的能量頻譜.....	18
圖(2.6) 計算區頻帶上的展開函數	19
圖(2.7) 計算區頻帶中的 $\text{minval}(b)$ 值	20
圖(2.8) 量尺因子頻帶的訊號遮罩比	21
圖(2.9) 混合式分析濾波器組	21
圖(2.10) 多項式分析濾波器	22
圖(2.11) 分析視窗的係數	23
圖(2.12) MPEG-1 Layer III 的解碼架構	27

圖(2.13) MPEG-1 Layer III 的位元流格式	28
圖(2.14) MPEG-1 Layer III 的位元流格式	28
圖(2.15) Layer III 的位元流組成及 main_data_begin	30
圖(2.16) 量化後的係數分佈	31
圖(2.17) 多項式合成濾波器組	32
圖(2.18) 多項式合成濾波器	33
圖(3.1) 不同位元率的量化誤差與可允許的誤差比較	39
圖(3.2) 語音與音樂的頻率範圍	40
圖(3.3) 修正式 MPEG 音訊編碼器架構	40
圖(3.4) 位元率 96 kbps 經過低通濾波器後的 SEGPSNR	41
圖(3.5) 位元率 80 kbps 經過低通濾波器後的 SEGPSNR	42
圖(3.6) 位元率 64 kbps 經過低通濾波器後的 SEGPSNR	42
圖(3.8) 位元率 96 kbps 不同音樂之效能比較	45
圖(3.9) 位元率 80 kbps 不同音樂之效能比較	46
圖(3.10) 位元率 64 kbps 不同音樂之效能比較	46
圖(4.1) 修正式 Layer III 的位元流組成	50
圖(4.2) 利用頻譜相似度來做高頻補償	51
圖(4.3) 頻譜相似度的求法	53
圖(4.4) 高頻補的償解碼端流程	55

圖(4.5) 位元率 96 kbps , MPEG-1 Layer III 音訊編碼器及加強式 MPEG-1

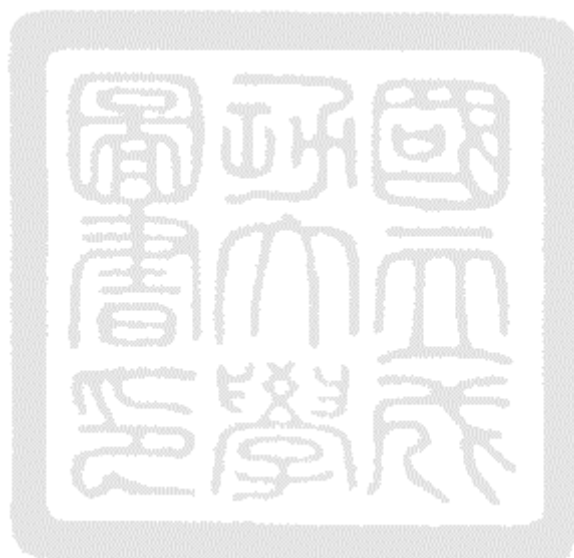
Layer III 音訊編碼器的比較..... 58

圖(4.6) 位元率 80 kbps , MPEG-1 Layer III 音訊編碼器及加強式 MPEG-1

Layer III 音訊編碼器的比較..... 59

圖(4.7) 位元率 64 kbps , MPEG-1 Layer III 音訊編碼器及加強式 MPEG-1

Layer III 音訊編碼器的比較..... 59



第一章 數位音訊編碼簡介

1.1 簡介

近年來，因為電腦與網路的普及率大為提升，造就了數位資訊的時代，相對地亦提昇了人們對資訊的傳達性與擷取性的需求。尤其是在多媒體影音運用的方面更是受到人們的青睞，例如：CD、VCD、DVD、網路會議、網路聊天室等，無一不是耳熟能詳的生活名詞，逐一地，於電子世界加以實現。在聲音（音樂）方面於普及率及使用性而言，最熱門的項目，無非以 MP3（MPEG-1 Layer III）的編碼與解碼器莫屬了，而 MP3 的主要的壓縮原論則屬於數位音訊編碼中的一環，他利用時頻轉換與心理音響之感知量化，而達到音訊壓縮的目的[13]。

數位音訊壓縮是為了要使寬頻聲音訊號（高取樣頻率）經由壓縮後，能夠保持高品質的聲音（近 CD 音質）以及較小容量的儲存空間，以方便人們使用、保存和傳輸。其主要目的在於將聲音訊號處理後以最少的位元（Bit）數來儲存，而訊號重建時希望能達到最小聽覺差異。簡言之，音樂壓縮希望達到即使受過專業訓練的試聽員也聽不出原先聲音以及壓縮處理過後的聲音差別[7]。

目前，高品質的聲音以 CD 音質為主。一般 CD 音質的取樣頻率為 48 kHz 或 44.1 kHz，每一個取樣點經由 16 位元的脈衝碼調變（PCM, Pulse

Code Modulation) 而成。其所需的數據量相當的龐大，每一單聲道聲音在取樣頻率為 48 kHz 每秒可達到 768 kbps (Kilobits per second) 和取樣頻率為 44.1 kHz 傳輸率為 705.6 kbps；若是雙聲道則分別為 1.54 Mbps (Megabits per second) 及 1.41 Mbps[6]。

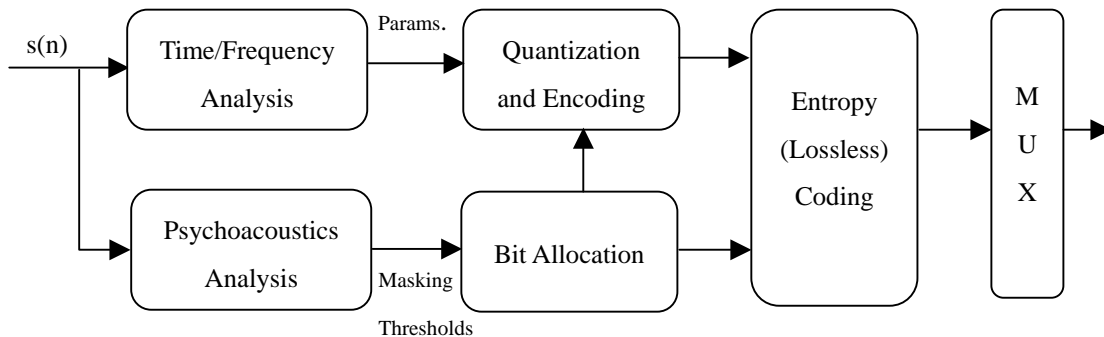
雖然 CD 音質為一般人所認可的高品質聲音，但在網路風潮的數位資訊時代中，卻因為其高容量或高傳輸率使其實用性大打折扣，無論在網路傳輸或資料儲存等方面，均造成使用者的不方便性。而一般使用者都希望能夠享有一個容量小卻又能夠接近 CD 音質的聲音資料，使其無論是在儲存上或傳輸上的便利或是聽音樂的娛樂等方面都能達到質與量的享受。因此，在近十年間，基於上述的多項原因，用戶的需求刺激了在數位音訊壓縮領域上的研究進步，研究者莫不希望能夠提昇以高壓縮率及高品質的聲音編碼為研究目的。

1.2 感知式音訊編碼

音訊壓縮的編碼目的都是為了達到以最少的位元數來代表原始訊號。若要以最少的位元數來做編碼，可經由量化器移除一些不必要或不相關的訊號資料，以達到最少的位元編碼目的。藉著人類聽覺的感官模型 (Human Auditory Perception Model)，可使感知式音訊編碼器更容易達到這個目的[7]。

一個典型的感知式音訊編碼器，乃將時域 (Time Domain) 的聲音訊號轉換成頻域 (Frequency Domain) 的訊號，再藉由聽覺的感官模型在頻域上，計算出人類聽覺可允許的量化誤差，也就是說這個量化誤差在人類聽來是感覺不出或是可容許的，這個聽覺的感官模型又稱之為心理音響模型 (Psychoacoustics Model) [1]。

一般的感知式音訊編碼器的主要架構如圖(1.1)所示，包含了心理音響模型的分析 (Psychoacoustics Analysis)、訊號的時頻轉換分析 (Time/Frequency Analysis)、量化 (Quantization) 及位元配置 (Bit Allocation) 和非失真編碼 (Lossless Coding) 等基本架構[6]。



圖(1.1) 感知式音訊編碼器之方塊圖

1.3 聽覺感知

人類的聽覺可以聽到的聲音範圍在 20 Hz 至 20 kHz 頻率之間，其中尤以 2 kHz 至 4 kHz 這範圍內的聲音對人類聽覺更為敏感。聲音振幅之動

態區間於可聽到的聲音大小範圍（由無聲至最大聲）約 96 dB[1]。

雖然人類的聽覺範圍在 20 Hz 至 20 kHz 之間，但並不是對每一種頻率的聲音其敏感度都相同。一般而言，人類的聽覺對低頻的聲音較為敏感，對高頻的聲音較不敏感。除此之外，聲音的大小、出現的時間亦會影響到人的聽覺系統。首先，我們將介紹幾個應用在聽覺感知的基本技術名詞，其包含：評鑑頻帶(Critical Band) 絕對臨界(Absolute Threshold) 與聲音壓力度(SPL, Sound Pressure Level)以及遮罩效應(Masking Effect) [1], [7]。

A. 評鑑頻帶

聽覺研究者發現，人類可聽到的頻率範圍在 20 Hz 至 20 kHz 之間，但是對不同的頻率的解析度感受有所不同，因此將這區間的頻率分成 25 個評鑑頻帶，使每一個頻帶對應線性頻率域都是非線性。一般而言，其頻帶隨著頻率變高而變寬，如表(1.1) 所示。

評鑑頻帶主要是為了描述人類的聽覺行為系統與工程應用上的頻率單位（赫茲，Hz）之間的關係，其頻寬的定義範圍是來自於人類對同一個評鑑頻帶內的聲音感受是相同的，也就是說訊號在同一個評鑑頻帶內，人對於頻率的變化是很難區分出來的。為了系統的簡單化，我們常將評鑑頻帶轉換至對頻率的測量較為一致性的稱之為評鑑頻帶率（CBR,

Critical Band Rate), 其單位為巴克 (Bark)

表(1.1) 關鍵頻帶(Bark)和邊界頻率(Hz)

Band (Bark)	Lower (Hz)	Center (Hz)	Upper (Hz)
1	0	50	100
2	100	150	200
3	200	250	300
4	300	350	400
5	400	450	510
6	510	570	630
7	630	700	770
8	770	840	920
9	920	1000	1080
10	1080	1170	1270
11	1270	1370	1480
12	1480	1600	1720
13	1720	1850	2000
14	2000	2150	2320
15	2320	2500	2700
16	2700	2900	3150
17	3150	3400	3700
18	3700	4000	4400
19	4400	4800	5300
20	5300	5800	6400
21	6400	7000	7700
22	7700	8500	9500
23	9500	10500	12000
24	12000	13500	15500
25	15500	19500	-

Bark 與頻率的對應以式子 (1-1) 近似表示。

$$b = 13 \cdot \tan^{-1}(0.76f) + 3.5 \cdot \tan^{-1}\left(\left(\frac{f}{7.5}\right)^2\right) \quad (\text{bark}) \quad (1-1)$$

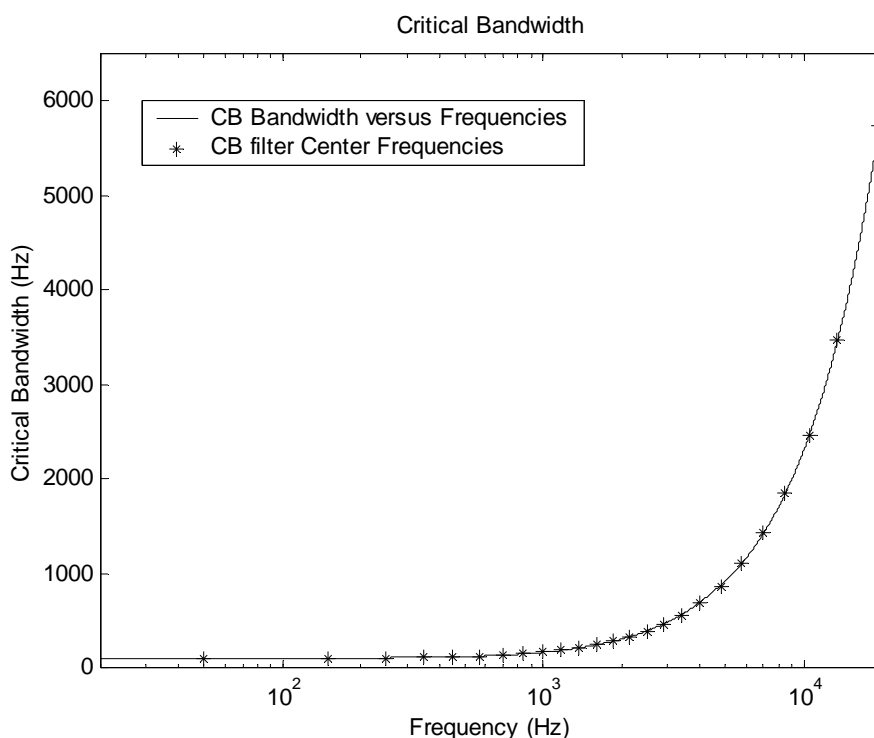
其中 b 為 Bark 表示的評鑑頻帶率, f 為原來 kHz 表示的頻率 由式子(1-1)

可知, 評鑑頻帶率是由連續的頻率函數所表示, 所以我們可由此得每一

個評鑑頻帶頻寬與頻率的關係式, 以式子 (1-2) 表示。

$$\Delta f_G = 25 + 75(1 + 1.4f^2)^{0.69} \quad (1-2)$$

其中 Δf_G 所表示的是個別評鑑頻帶的頻寬大小， f 的單位一樣為 kHz。頻率與評鑑頻帶頻寬的關係圖，如圖(1.2)所示。



圖(1.2) 頻率 (Hz) 與評鑑頻帶頻寬的對應

B. 絕對臨界與聲音壓力度

雖然，人類可聽到的頻率範圍在 20 Hz 至 20 kHz 之間，一旦聲音能量太小，就不會被人類所耳朵察覺，這一個臨界值就稱之為絕對臨界。

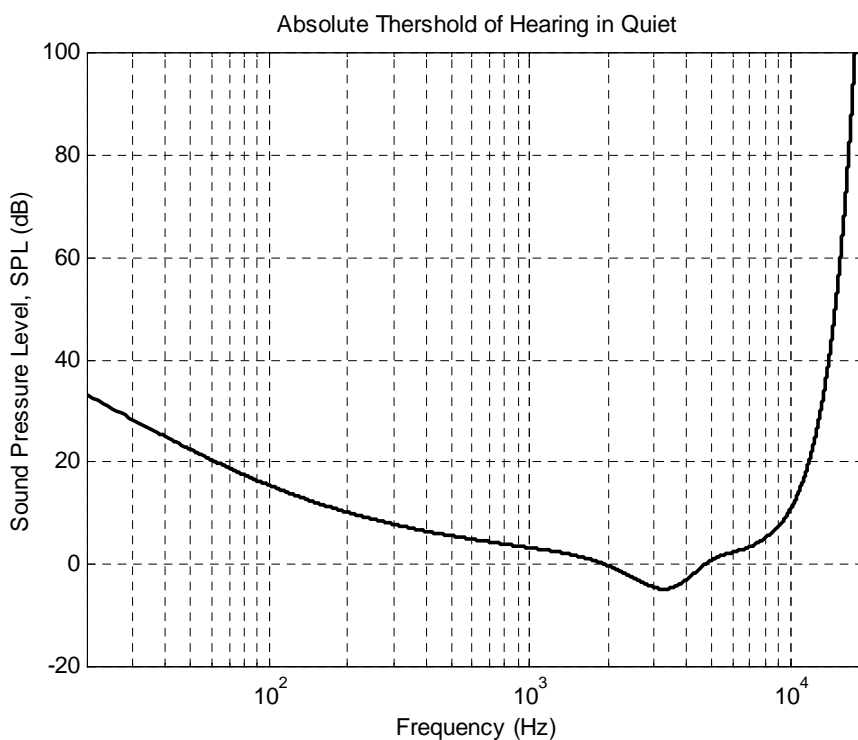
絕對臨界又稱之為靜音臨界 (Threshold in Quiet)，而絕對臨界通常以聲音壓力度 (SPL dB) 來表示。當一個音調 (Tone) 的聲音壓力度低於絕對臨界時，這一個音調就不被人所察覺；相反地，當聲音壓力度高

於絕對臨界時，就會被人耳所聽到。

絕對臨界與頻率的關係式以式子 (1-3) 所表示。

$$TH(f) = 3.64 \cdot \left(\frac{f}{1000}\right)^{-0.8} - 6.5 \cdot \exp\left(-0.6\left(\frac{f}{1000} - 3.3\right)^2\right) + 10^{-3} \cdot \left(\frac{f}{1000}\right)^4 \quad (1-3)$$

f 為頻率，其單位為 Hz， TH 為絕對臨界，單位是 dB，以聲音壓力度表示。兩者關係可以圖(1.3)表示。



圖(1.3) 絕對臨界 (dB) 與頻率 (Hz) 關係

C. 遮罩效應

遮罩效應 (Masking Effect) 在人類的聽覺系統中佔了一個很大的重要性，尤其是在心理音響模型中更是用來當作減少音訊壓縮編碼位元數

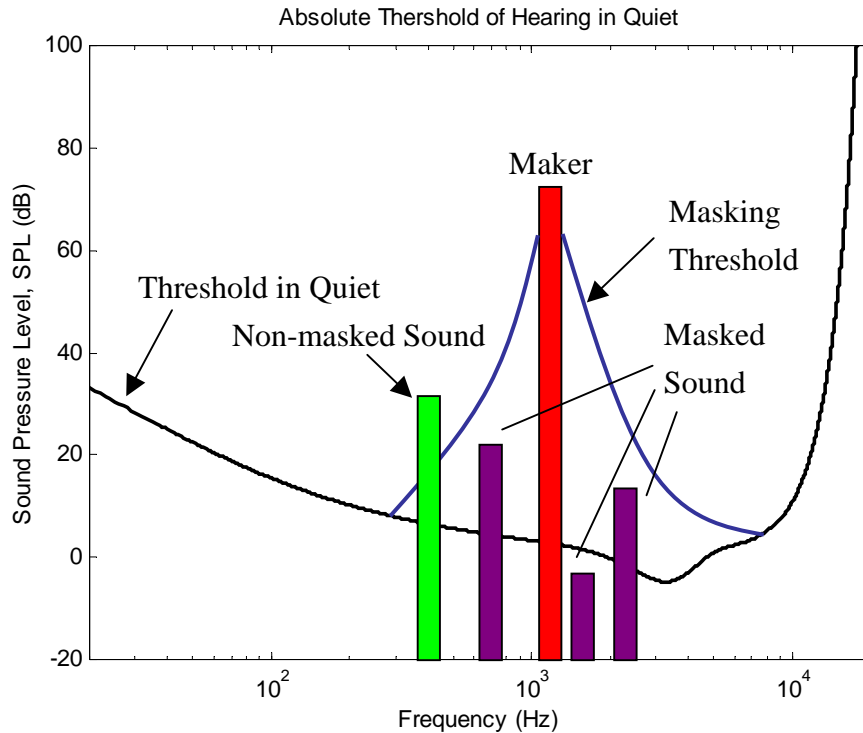
的重要依據。遮罩 (Masking) 可以說是由一個或多個聲音在時間上或是頻率域中所產生的遮蔽效果。假設有一個訊號被遮蔽 (Masked) 了, 就表示這個訊號不被人的聽覺系統所察覺到; 相反地, 人耳只能聽到遮蔽者 (Masker) 這一個訊號的聲音, 被遮蔽者 (Maskee) 在人的聽覺系統中相當於不存在了。

一般遮罩效應簡單來講分為: 頻域遮罩 (Spectral Masking) 與時域遮罩 (Temporal Masking), 以下我們就這兩者的特性再稍加解釋。

➤ C-1 頻域遮罩

遮罩效應在頻率域中通常有以下幾種情形: 訊號遮蔽雜訊、訊號遮蔽訊號、雜訊遮蔽雜訊和雜訊遮蔽訊號等情形發生。不論是哪一種遮蔽, 都取決於聲音壓力度 (SPL) 的大小以及頻率域的位置關係。

由圖(1.4)可以看到在 1 kHz 出現一個聲音壓力度為 60 dB 的音調, 在其頻率附近亦有三個小音調出現, 但其聲音壓力度皆小於由 1 kHz 的音調所產生的遮蔽臨界曲線, 以致於這三個小音調都不被人耳所聽到; 人所能聽到的聲音只有位在 1 kHz 那個音調而已。相反地, 另一個音調的聲音壓力度若高於遮蔽者所產生的遮蔽臨界曲線, 則這個音調將會被人耳所聽到。



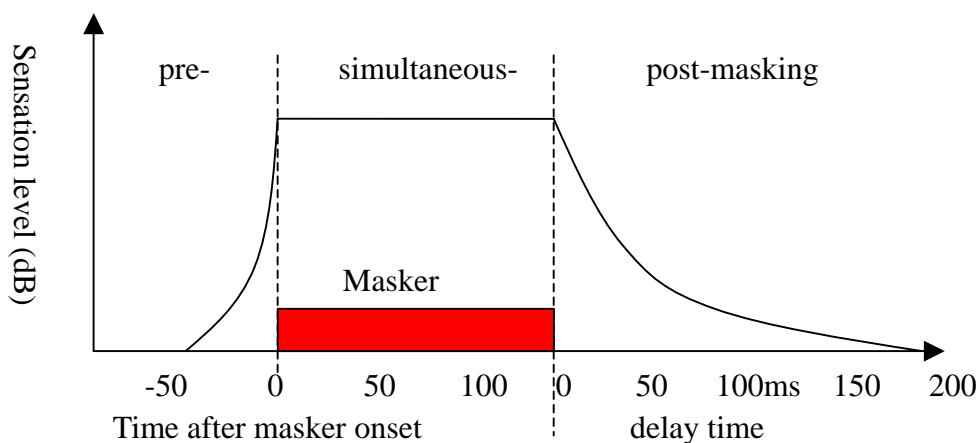
圖(1.4) 頻域遮罩效應

➤ C-2 時域遮罩

除了在頻率域會產生遮罩效應，在時域亦會產生遮罩效應。在時域所產生的遮蔽效應又可分為前遮罩效應（Pre-masking）、同步遮罩效應（Simultaneous masking）以及後遮罩效應（Post-masking），如圖(1.5)所示。

前遮罩效應的發生都是在遮蔽者（Masker）即將產生之前，通常影響的持續時間約 20 ms，再前的影響幾乎可以忽略。前遮蔽效應在時域遮罩所佔的重要性較低。

同步遮罩效應的時效是遮蔽者這個聲音所持續的時間，當遮蔽者消失



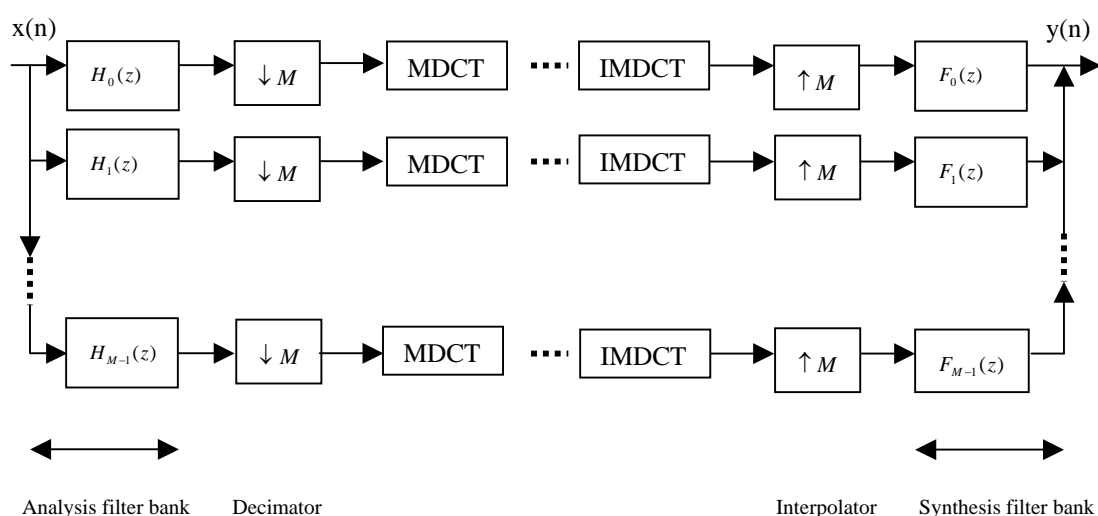
圖(1.5) 時域遮罩效應

時，同步遮罩效應亦消失不見。當遮蔽者的聲音結束時，後遮罩效應接著出現。一般會持續的時間約 200 ms，其遮蔽效應於此對其他聲音之影響大小達到此為止。

1.4 時頻轉換

在音訊編碼中，除了要分析聽覺感知系統之外，另一個重點就是將時域的聲音訊號轉換至頻率域。每隔一段時間，音訊 N 個數據經由濾波器組 (Filter Bank) 和離散轉換 (Discrete Transform) 可將訊號分解頻率域數據，每隔 N 時間之頻域數據稱為時頻成分 (Time-Frequency component)。時頻轉換可以將該段的訊號的相關性去除，以提供編碼端一組低相關性的輸入訊號[11]；編碼器再參考聽覺感知系統模式以量測可允許失真，來作量化編碼的基準，如此即可得到高品質的音訊壓縮結果。

圖(1.6)是一個 M 通道的濾波器組，幾乎所有音訊壓縮編碼系統均可看到。濾波器以鏡像濾波器 (Quadrature Mirror Filter) 為基礎設計，率波器將訊號粗分 M 個頻帶，每個頻帶可再由修正式離散餘弦轉換 (MDCT, Modified Discrete Cosine Transform) 在細分為數個頻率項目；時頻轉換可以達到完全重建 (PR, Perfect Reconstruction) 的結果[註一]。



圖(1.6) 一般音訊壓縮 M 通道的濾波器組

1.5 量化

量化主要是在量尺因子頻帶 (Scalefactor band) 中調整量化步階大小 (Quantizer step size), 以符合由聽覺感知系統所分析出的可允許誤差。

[註一]：完全重建對一個壓縮編碼器而言並不可能，此處是完全重建乃指允許量化失真的結果[11]。

只要量化誤差的結果小於人耳所能容許的範圍內，這個量化步階大小就是最適當的；若是無法用量化步階的大小就使每一個量尺因子頻帶的誤差小於可容許的誤差，這時候就要在誤差較大的量尺因子頻帶中調整量尺因子 (Scalefactor)，以符合誤差需求。

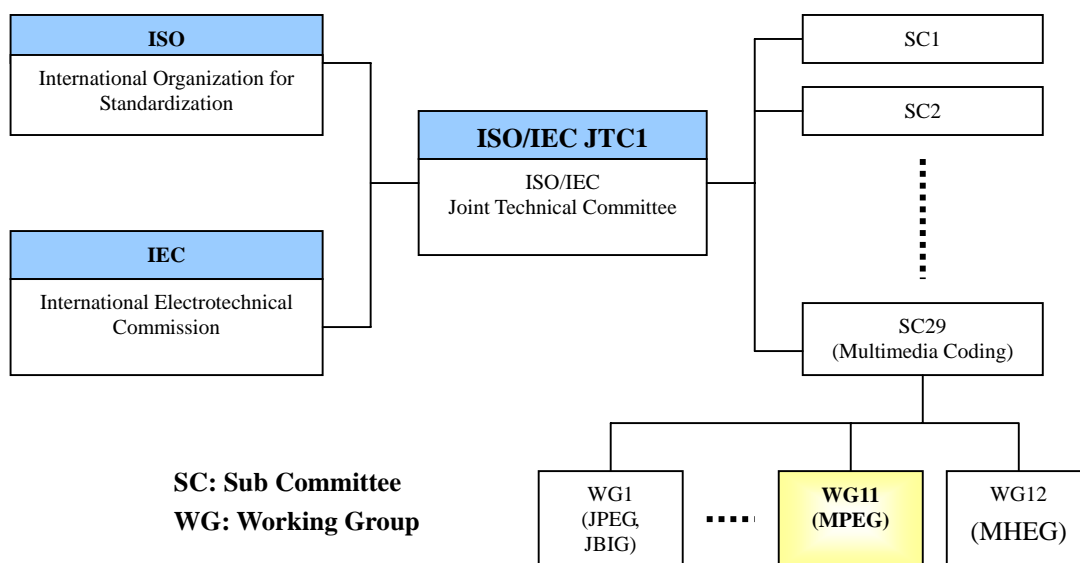
1.6 非失真編碼

利用霍夫曼編碼 (Huffman Coding) 將量化後的係數加以做最有效的非失真編碼。將量化後的頻率係數兩兩為一組，或四個為一組，選擇最佳的霍夫曼碼簿 (Huffman codebook)，利用查表的方式來編碼，不僅可以用最少的位元數編碼，也可以達到完全重建的效果。

第二章 MPEG-1 Layer III 編解碼器

2.1 簡介

MPEG-1 Layer III 又稱之為 MP3，這一套音訊壓縮編碼系統，是由國際標準組織（ISO/IEC, International Organization for Standardization and International Electrotechnical Commission）所制訂出來的 MPEG（Moving Picture Experts Group）音訊壓縮標準。國際標準組織之架構如圖(2.1)所示。

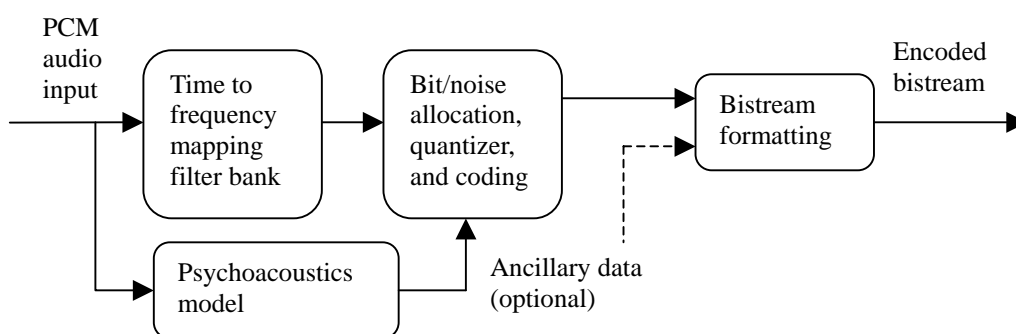


圖(2.1) 國際標準組織架構

在 MPEG-1 多媒體標準中，主要分為三個部分：第一部份（ISO/IEC 11172-1），系統（Systems）標準，以提供音/視訊系統傳輸格式及封包規

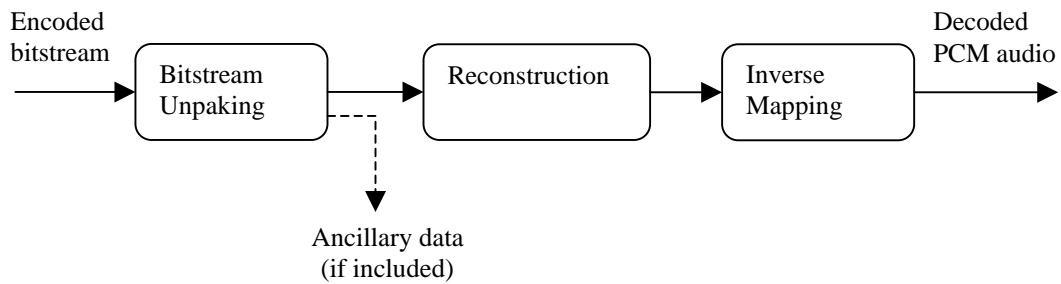
格。第二部分 (ISO/IEC 11172-2), 視訊 (Video) 標準, 訂定視訊解碼流程方法, 以及視訊基本數據流 (Element Stream) 之格式。第三部分 (ISO/IEC 11172-3), 音訊 (Audio) 標準, 制訂音訊編碼流程與方法, 以及音訊基本數據流之格式。

MPEG-1 中的音訊編碼系統屬於第三部分 (Part 3), 其中又提供了三層 (Layer) 互相獨立的壓縮編碼。根據不同的應用, 不同層的編碼系統隨著層數的增加, 其複雜度 (Complexity) 與效能 (Performance) 也隨著增加。無論是哪一層, 其編碼及解碼基本架構分別如圖(2.2)及(2.3)所示[2], [3], [4], [5], [10]。



圖(2.2) MPEG 音訊編碼器基本架構

一般而言, MPEG 僅訂定解碼流程, 在編碼端的演算法並沒有規定, 只要能夠符合編碼完之後的位元流格式 (Bitstream Format) 就可以了; 例如關於聽覺的遮蔽臨界曲線 (Auditory Masking Threshold) 的估測和量



圖(2.3) MPEG 音訊解碼器基本架構

化等過程，均可以依設計者之考量而改寫。除此之外，使用者亦可自己定義的輔助資料 (Ancillary Data)，以增加其編解碼器的彈性。

MPEG-1 音訊壓縮編碼三層的比較如下：第一層：複雜度簡單，每一個通道 (Channel) 較適合的位元率 (Bitrate) 高於 128 kbps，例如：菲力普 (Philips) 的 DCC (Digital Compact Cassette) 所採用的就是 MPEG-1 第一層的壓縮，每個通道的位元率為 192 kbps。

第二層：複雜度普通，每一個通道較適合的位元率大約在 128 kbps 左右，適合的應用包括了數位廣播 (DAB, Digital Audio Broadcasting)、VCD 音訊或視訊及音訊同步系統的儲存等。

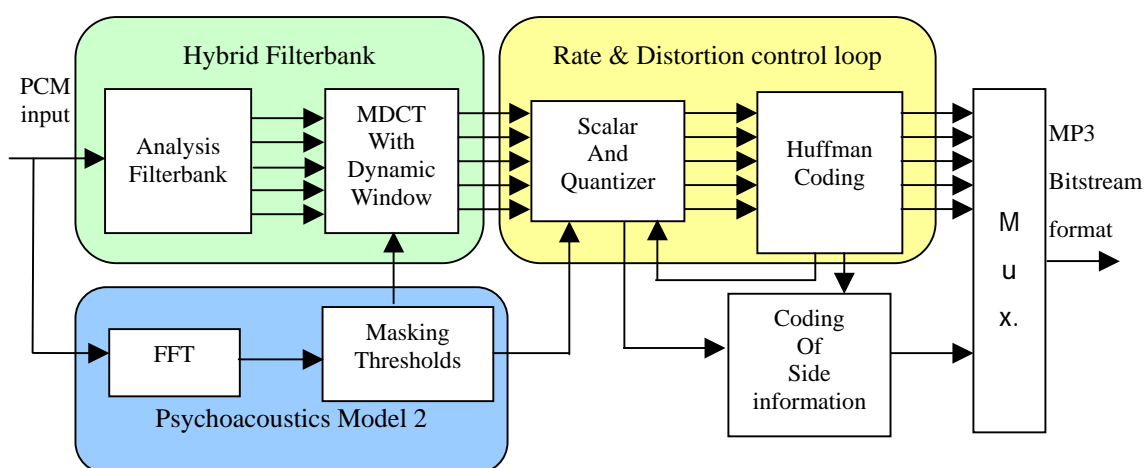
第三層：複雜度最高，但提供了較佳的音訊品質，每一個通道適合的位元率在 64 kbps 左右，由於資料量少，較適合用於網路傳輸；一般稱之為 MP3，主要是因為在 MPEG 音訊編碼系統屬於第三層的架構。

這三層的解碼器都相當簡單，如圖(2.3)，均適合在單一晶片上做即時 (Real-time) 解碼系統。

2.2 MPEG-1 Layer III 編碼器

MPEG-1 Layer III 的編碼架構如圖 (2.4) 所示，與一般的知覺式音訊編碼器 (Perceptual Audio Coder) 近似[參考 1.2 節]，主要架構為心理音響模型 (Psychoacoustics Model)、混合式濾波器組 (Hybrid Filterbank) 與及量化和無失真編碼 (Lossless Coding / Huffman Coding)，其中量化及無失真編碼在 MP3 的架構中主要用來做位元率及誤差的控制 (Rate and Distortion Control Loop)，由於是以迴圈的方式進行微調控制，所以又稱之為疊代迴圈 (Iteration Loop)。

以下將介紹 MPEG-1 Layer III 各個工作原理及方法。



圖(2.4) MPEG-1 Layer III 編碼器架構

A. 心理音響模型分析

MPEG-1 有兩個心理音響模型，第一層及第二層的編碼採用的是心理音響模型 I，第三層的編碼大多採用的是適應性（Adaptation）心理音響模型 II [註二] [2], [3], [4], [5]。

首先，將時域的聲音訊號經由 1024 點的快速傅利業轉換（FFT, Fast Fourier Transform）至頻率域，由傅利業頻譜大小直接評估是否為音調（Tonality）或非音調（Non-tonality）。音調的判別是利用頻譜的不可預測性（Unpredictability），其方法是由前兩個（時間上）頻譜的相位（Phase）及振幅（Amplitude）來做估測，如式子 (2-1)、(2-2) 以及 (2-3) 所示：

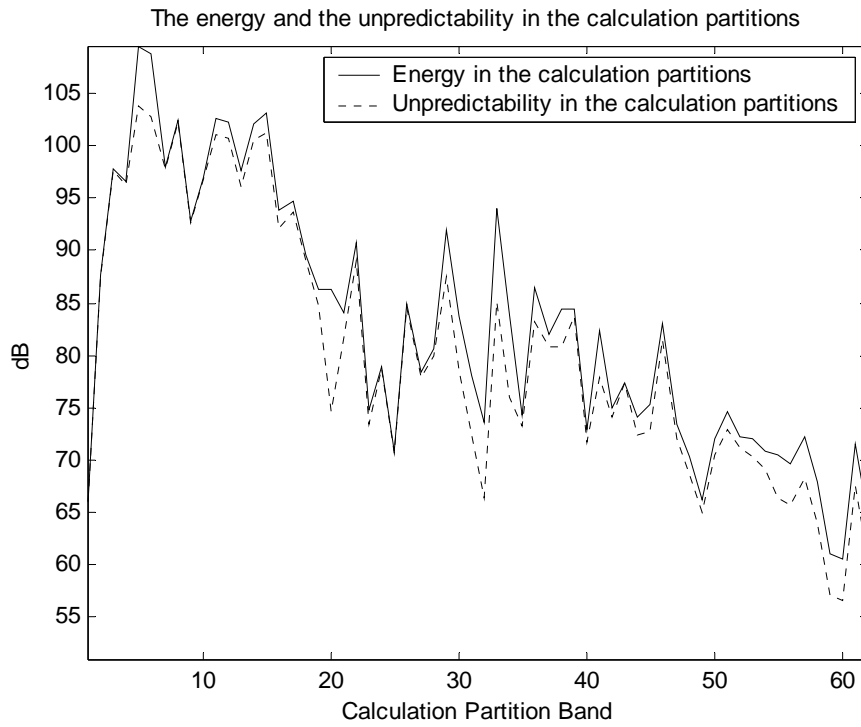
$$\hat{r}_\omega = 2.0r_\omega(t-1) - r_\omega(t-2) \quad (2-1)$$

$$\hat{f}_\omega = 2.0f_\omega(t-1) - f_\omega(t-2) \quad (2-2)$$

$$c_\omega = \frac{((r_\omega \cos f_\omega - \hat{r}_\omega \cos \hat{f}_\omega)^2 + (r_\omega \sin f_\omega - \hat{r}_\omega \sin \hat{f}_\omega)^2)^{0.5}}{r_\omega + \text{abs}(\hat{r}_\omega)} \quad (2-3)$$

其中 r_ω 為振幅， f_ω 為相位， t 是時間（音框），頻譜的不可預測性由 c_ω 來表示。對一個純粹的音調，其 $c_\omega = 0$ ，若是非音調則 $0 < c_\omega \leq 1$ 。將聲音頻譜分為 61 個非線性的頻帶，稱之為計算區頻帶（Calculation Partition Band），並求出每一個計算區頻帶的能量頻譜 e_b 與加權不可預測性的能量頻譜 c_b 分別以式子 (2-4) 及 (2-5) 表示，其結果如圖(2.5)所示。

[註二]：在 MPEG-1 中對心理音響模型 2 所做的部分簡化與修正的版本。



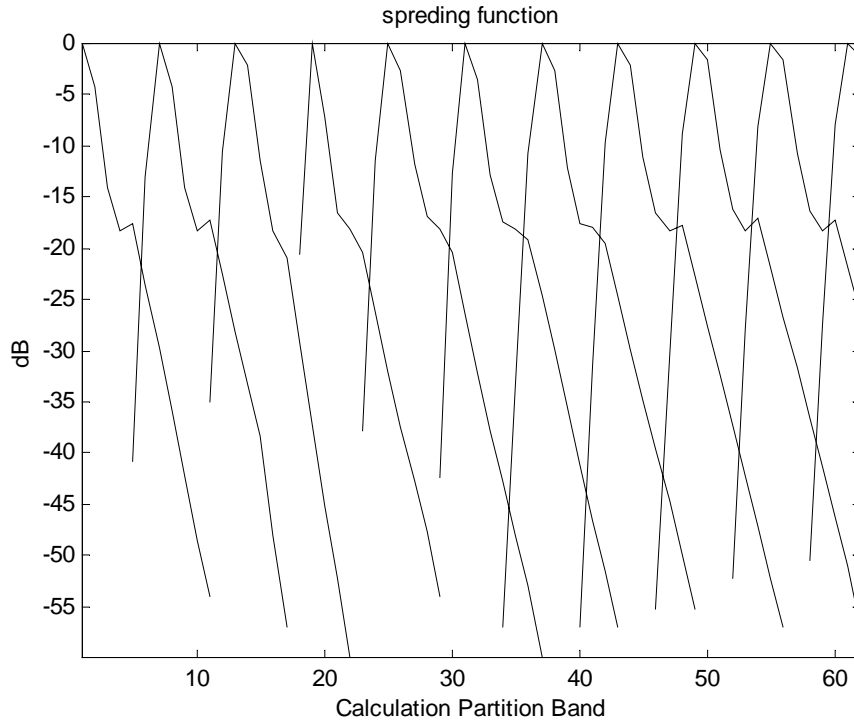
圖(2.5) 計算區頻帶的能量頻譜與加權不可預測性的能量頻譜

$$e_b = \sum_{\omega=\omega_{l,b}}^{\omega_{u,b}} r_{\omega}^2 \quad (2-4)$$

$$c_b = \sum_{\omega=\omega_{l,b}}^{\omega_{u,b}} r_{\omega}^2 c_{\omega} \quad (2-5)$$

其中 $\omega_{u,b}$ 與 $\omega_{l,b}$ 分別為計算區頻帶的上、下界。

將加權 (c_b) 及未加權 (e_b) 不可預測性的能量頻譜與展開函數 (Spreading Function) 做旋積, 分別以 ecb_b 及 ct_b 表示, 其中展開函數如圖(2.6)所示。其加權過後的不可預測性能量頻譜正規化(cb_b), 如式子 (2-6) 所示,



圖(2.6) 計算區頻帶上的展開函數

$$cb_b = \frac{ct_b}{ecb_b} \quad (2-6)$$

繼而求得在頻譜中的音調成分 (範圍 0 1), 如式子 (2-7) 所示。

$$tb_b = -0.299 - 0.43 \log_e (cb_b) \quad (2-7)$$

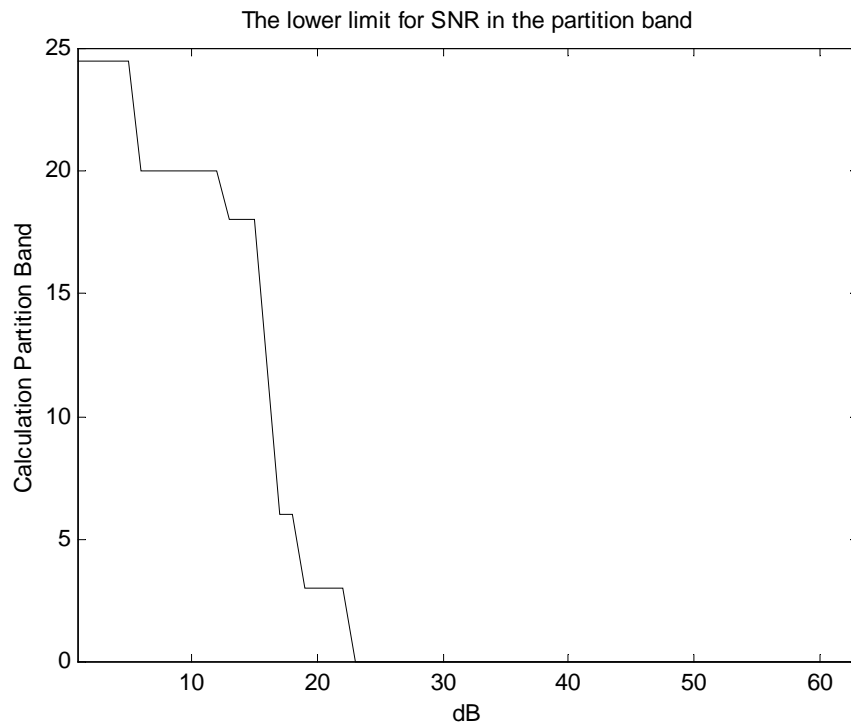
在每一個計算區頻帶中, 遮罩臨界曲線是由能量頻譜乘上衰減因子 (Attenuation Factor)。在 MPEG-1 Layer III 中, 衰減因子的定義如式子 (2-8)、(2-9) 所示。

$$attenuation = 10^{-SNR(b)} \quad (2-8)$$

$$SNR_b = \text{maximum}(\text{minval}(b), tb_b * TMN_b + (1 - tb_b) * NMT_b) \quad (2-9)$$

其中 $\text{minval}(b)$ 是由 MPEG-1 中定義在每一個計算區頻帶的 SNR 最小值如圖(2.7)所示, TMN_b (Tone Masking Noise) 訂為 29 dB, NMT_b (Noise

Masking Tone) 訂為 6 dB。



圖(2.7) 計算區頻帶中的 minval(b)值

將可得到遮罩臨界曲線，如式子 (2-10) 所示，

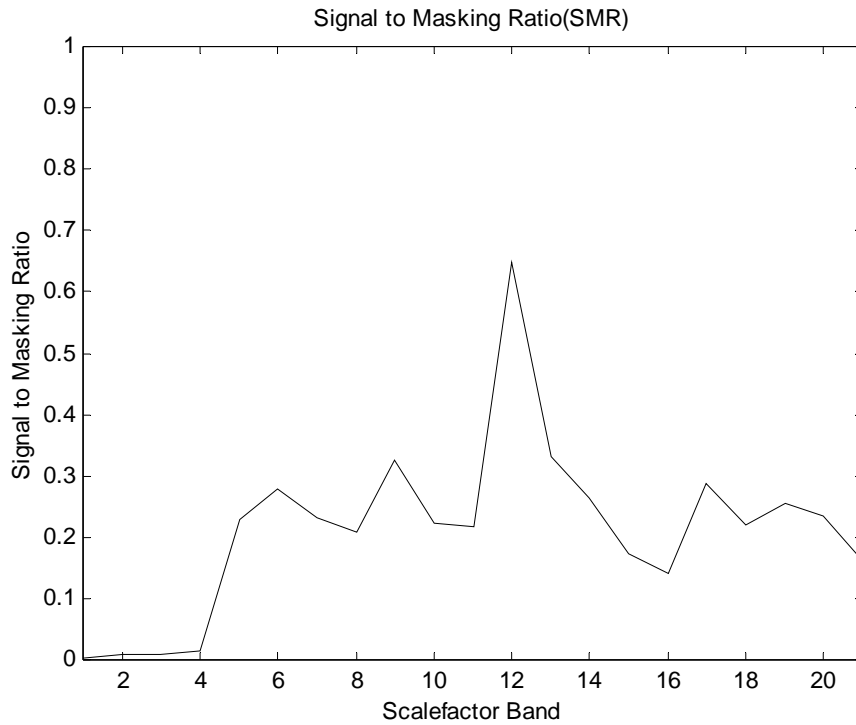
$$masking_threshold = convolved_energy_spectrum * attenuation \quad (2-10)$$

再與靜音臨界比較，將可得到真正的遮罩臨界曲線。最後利用式子 (2-11)

可以計算出每一個量尺因子頻帶的訊號遮罩比 (SMR, Signal to Mask Ratio)，如圖(2.8)所示。

$$SMR_n = 10 \log_{10} \frac{signal_energy_n}{masking_threshold_n} \quad (2-11)$$

其中 n 為量尺因子頻帶。

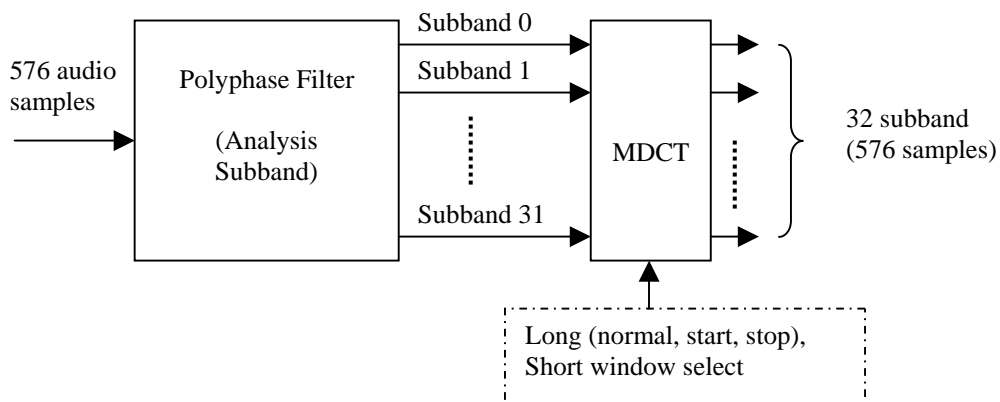


圖(2.8) 量尺因子頻帶的訊號遮罩比

B. 混合式分析濾波器組

混合式分析濾波器組主要分為兩個部分，一個是多項式分析濾波器 (Polyphase Analysis Filter)，另一個是修正式離散餘弦轉換 (MDCT)。

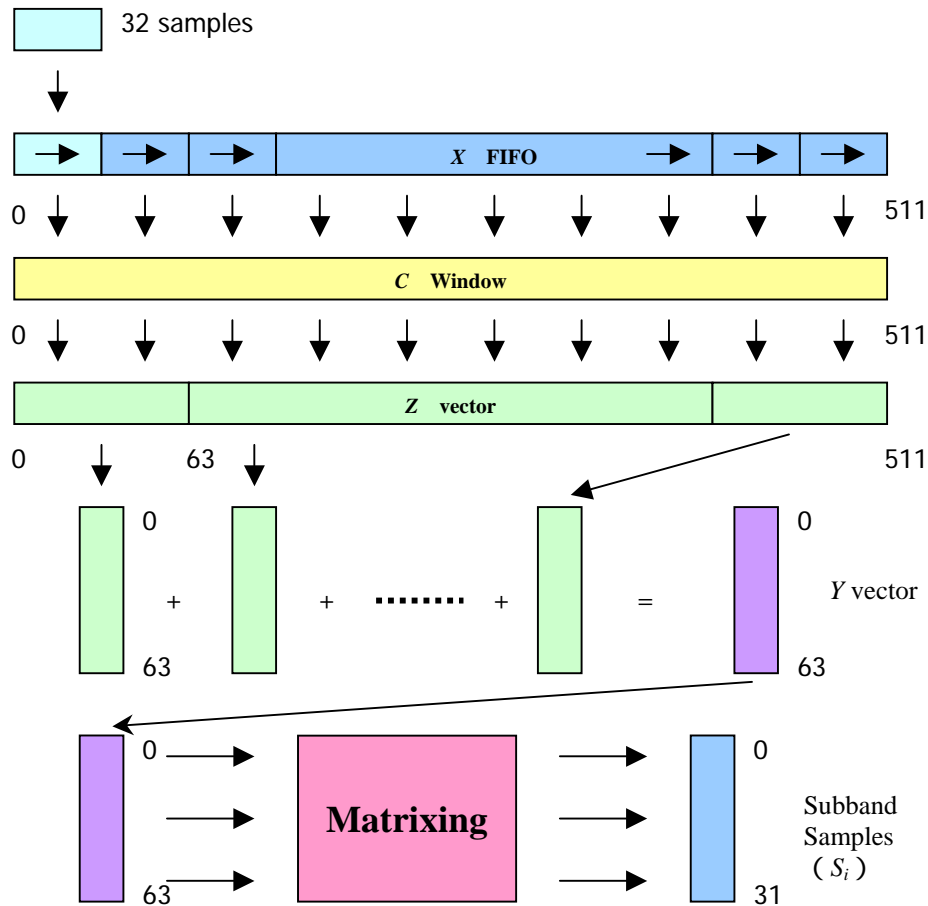
圖(2.9)為混合式分析濾波器組的實現說明圖[2], [11]。



圖(2.9) 混合式分析濾波器組

➤ B-1 多項式分析濾波器

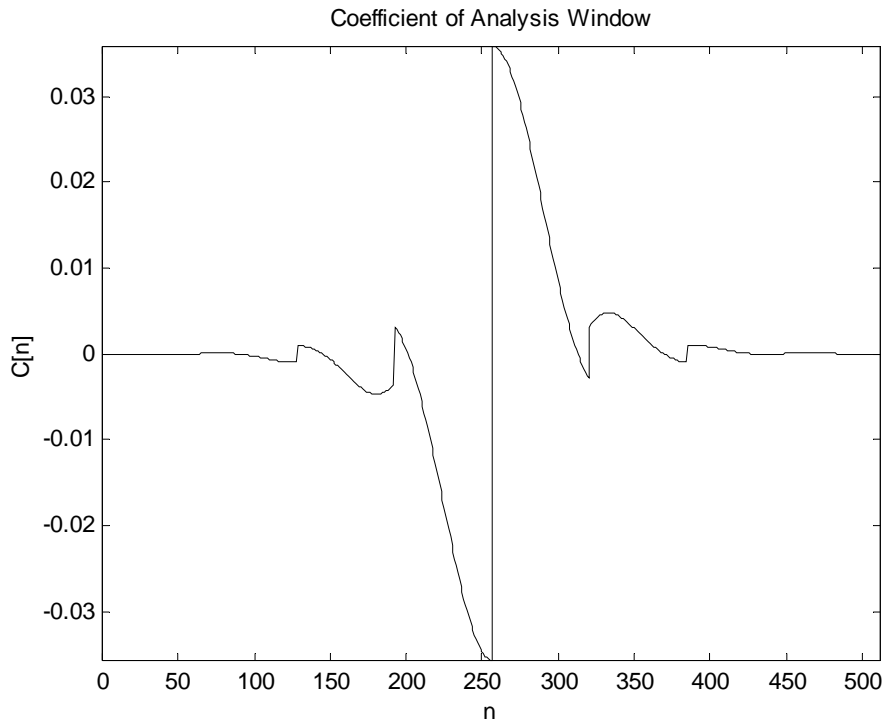
MPEG-1 音訊壓縮利用多項式分析濾波器將訊號分成相同頻寬的 32 個子頻帶，如圖(2.10)所示。



圖(2.10) 多項式分析濾波器

原始數據每 32 點輸入一次，依序置於 X 暫存器中，其中 X 為 FIFO (First In First Out) 暫存器，再與 $C[n]$ 加權窗 (C Window) 相乘產生 Z 向量在利用式子 (2-12) 產生 Y 向量，其中 C 加權窗如圖(2.11)所示。

$$Y_j = \sum_{i=0}^7 Z_{i+64j} \quad \text{for } j = 0 \text{ to } 63 \quad (2-12)$$



圖(2.11) 分析視窗的係數

再經由矩陣運作 (Matrixing), 如式子 (2-13) 所示, 使輸入的 32 聲音訊號產生 32 個子頻帶訊號 (s_i), 此為臨界取樣 (請參考 1.4 節)。

$$S_i = \sum_{k=0}^{63} M_{ik} * Y_k \quad \text{for } i=0 \text{ to } 31$$

$$M_{ik} = \cos[(2i+1)(k-16)\pi / 64] \quad (2-13)$$

➤ B-2 修正式離散餘弦轉換

修正式離散餘弦轉換又稱之為時域重疊負像效應消除法 (TDAC, Time Domain Alias Cancellation), 主要利用重疊 (Over-lap) 轉換原理, 將時域數據轉換為頻域, 並於逆修正式離散餘弦轉換之重疊以消除非理想負像 (Aliasing) 之時域重建誤差[8], [9]。

在 MPEG-1 Layer III 中，修正式離散餘弦轉換，如式子 (2-14) 所示，

$$xr_i = \sum_{k=0}^{n-1} z_k \cos\left(\frac{\pi}{2n}(2k+1+\frac{n}{2})(2i+1)\right) \quad \text{for } i = 0 \text{ to } \frac{n}{2}-1 \quad (2-14)$$

其可視訊號特性，簡單來分有長區塊(Long Block) [註三]及短區塊(Short Block) 兩種，當選擇長區塊時 $n = 36$ ，若為短區塊 $n = 12$ ，兩者之重疊均有 50% 的重疊區域 (Over-lap)。

C. 疊代迴圈

在做音訊壓縮編碼時，必須考慮到人的聽覺，尤其是對於聲音的敏感度；當我們在聽音樂時只要一點點的雜訊就會影響聲音的品質。這個雜訊通常是由量化所產生的量化誤差。疊代迴圈的主要目的為控制這些誤差以及調整位元率。

疊代迴圈主要包含兩個迴圈：

1. 外疊代迴圈 (Outer Iteration Loop)
2. 內疊代迴圈 (Inner Iteration Loop)

其中外疊代迴圈又稱之為誤差控制迴圈 (Distortion Control Loop)，控制著量化誤差 (Quantization Noise)；而內疊代迴圈又稱之為位元率控制迴圈 (Rate Control Loop) 決定量化器階的大小(Quantization Step Size)。

經由心理音響模型可以得到每一個量尺因子頻帶的訊號遮罩比，由訊號遮罩比可評估出所能容許的誤差 (Allowed Distortion)。所容許誤差的計算公式如 (2-15) 所示：

$$x_{\min}(\text{sb}) = \frac{\text{ratio}(\text{sb}) * \text{en}(\text{sb})}{\text{bw}(\text{sb})} \quad (2-15)$$

[註三]：長區塊分三種：normal, start, stop，區塊的決定是由心理音響模型來決定編碼要採用那一種。

其中 sb 表量尺因子頻帶， $xmin(sb)$ 是每一個量尺因子頻帶所能容許的誤差， $ratio(sb)$ 是由心理音響模型所分析出來在量尺因子頻帶中的訊號遮罩比， $en(sb)$ 為每一個量尺因子頻帶能量和， $bw(sb)$ 為量尺因子頻帶的頻寬大小。

在內疊代迴圈中可以找出一個適當的量化步階大小，其所量化出來的大小符合了霍夫曼碼表 (Huffman Code Table) 中編碼的最大範圍內，若超出其範圍，則 $qquant$ 則一步一步往上累加；也就是量化步階大小逐次加 1，直至滿足範圍為止。其量化的公式如 (2-16) 所示：

$$ix(i) = nint\left(\left(\frac{|xr(i)|}{\sqrt[4]{2^{qquant+quantanf}}}\right)^{0.75} - 0.0946\right) \quad (2-16)$$

其中 $xr(i)$ 為經過修正式離散餘弦轉換的頻譜值， $ix(i)$ 為量化後的結果，均為整數， $nint$ (Near Integer)表四捨五入， $qquant$ 則為量化步階的計數器，其初值為 0。 $quantanf$ 為量化步階大小的初值，其定義如下：

$$quantanf = system_const * \log_e (sfm) \quad (2-17)$$

其中

$$sfm = \frac{e^{\frac{1}{n} \left(\sum_{i=0}^{n-1} \log xr(i)^2 \right)}}{\sum_{i=0}^{n-1} \log xr(i)^2} \quad (2-18)$$

表示頻譜平坦度之量測值 (Spectral Flatness Measure)， $system_const$ 訂為 8.0。

上述量化編碼 $ix(i)$ ，經霍夫曼編碼後的位元數，將經過多次疊代後使其總數小於或等於目前所擁有的位元空間 (Bit Reservoir)，為達到目標位元數，量化步階的大小在內疊代迴圈決定出來後，在外疊代迴圈中就要來決定這個量化器所產生的誤差是否在可容許的範圍 ($xmin(sb)$) 內。在每個量尺因子頻帶所產生的誤差則由公式 (2-19) 來得出：

$$xfsf(sb) = \frac{\sum_{i=lbl(sb)}^{i=lbl(sb)+bw(sb)-1} (|xr(i)| - ix(i)^{4/3} * \sqrt[4]{2}^{qquant+quantanf})^2}{bw(sb)} \quad (2-19)$$

其中 $xfsf(sb)$ 為量尺因子頻帶中的量化誤差。

假使在這個量尺因子頻帶中， $xfsf(sb) > xmin(sb)$ ，那我們將會加強這個量尺因子頻帶中的每一個頻譜值 (xr)，也就是將每一個頻譜值乘上一個 $ifqstep$ 。其中：

$$ifqstep = 2^{(0.5(1+scalefac_scale))} \quad (2-20)$$

而 $scalefac_scale$ 的初始值為 0，可調整為 1。所以乘上的 $ifqstep$ 為 2 或是 $\sqrt{2}$ 。除此之外，所能容許的誤差 ($xmin$) 也要相對應的放大，但卻是乘上 $ifqstep^2$ 。其中：

$$ifqstep2 = ifqstep * ifqstep \quad (2-21)$$

因此我們會將這個量尺因子頻帶中的量尺因子 (Scalefactor) 加 1。所以放大幾次就加幾次，最後要送給解碼端的則是所放大的次數，也就是每一個量尺因子頻帶中的量尺因子。

由以上我們可以知道：在外疊代迴圈所調整的量尺因子，也就是放大某個量尺因子頻帶中的頻譜值；之後在內疊代迴圈再量化。最後，迴圈終止的條件為：

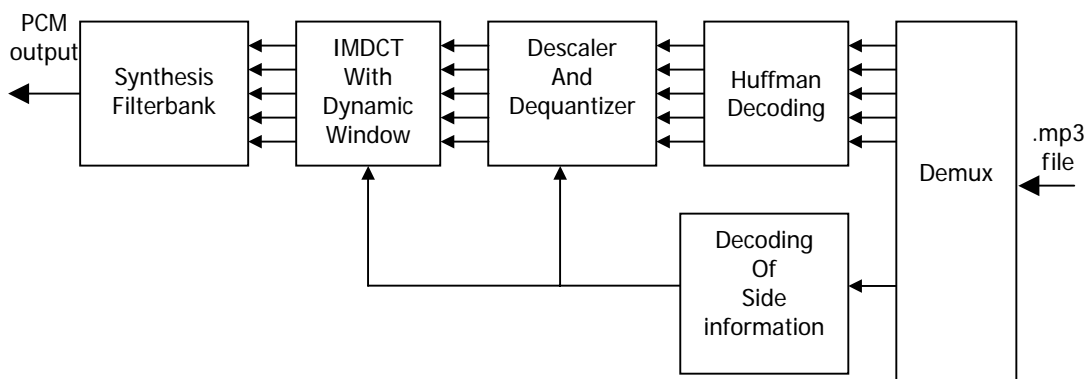
- a) 所有的量尺因子頻帶的頻譜都已經放大了，或
- b) 在量尺因子頻帶所增強的因子量尺因子都達到上限（參考 $scalefac_compress$ 表(2.1)中 $slen1$ 和 $slen2$ 的位元數）內。

表(2.1) 決定量尺因子的位元數

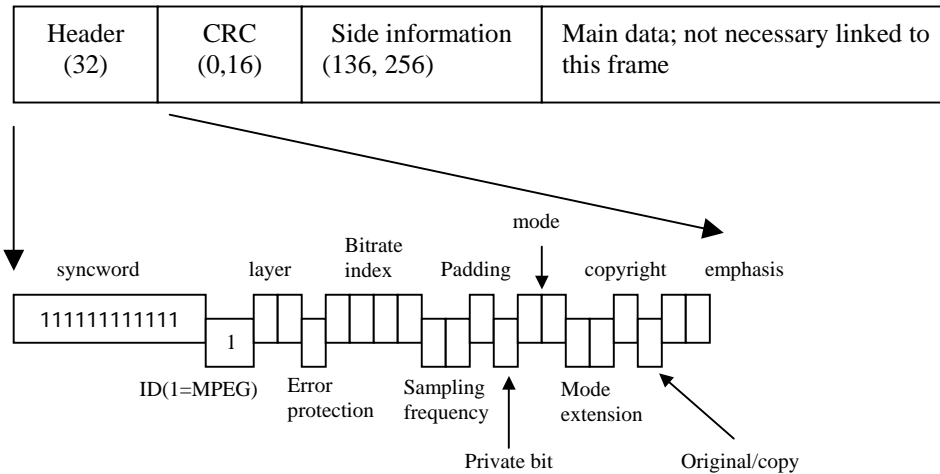
scalefac_compress	slen1	slen2
0	0	0
1	0	1
2	0	2
3	0	3
4	3	0
5	1	1
6	1	2
7	1	3
8	2	1
9	2	2
10	2	3
11	3	1
12	3	2
13	3	3
14	4	2
15	4	3

2.3 MPEG-1 Layer III 解碼器

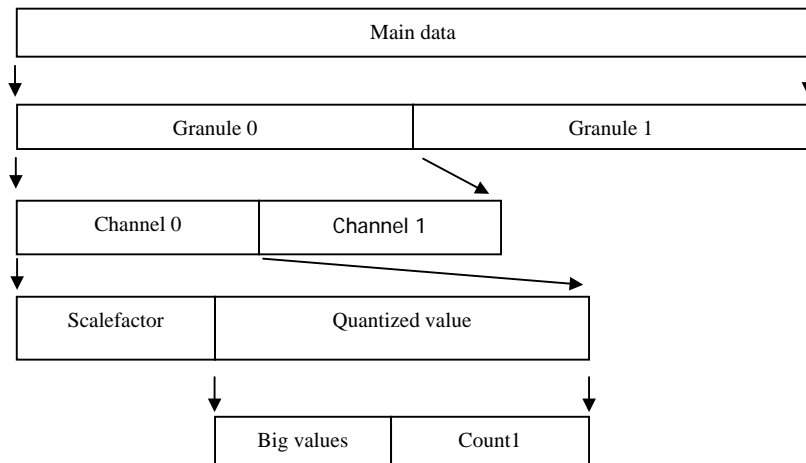
MPEG-1 Layer III 的解碼架構如圖(2.12)所示,經由編碼後的位元流格式如圖(2.13)、(2.14)所示。



圖(2.12) MPEG-1 Layer III 的解碼架構



圖(2.13) MPEG-1 Layer III 的位元流格式



圖(2.14) MPEG-1 Layer III 的位元流格式

MPEG-1 Layer III 的解碼端相較於編碼端簡單多了，少了心理音響模型的分析，也不用作量化疊代迴圈，適合於在不同的平台上做及時播放系統。整體而言，MPEG-1 Layer III 屬於非對稱式音訊壓縮編碼系統。

A. MPEG-1 Layer III 的位元流格式

由圖(2.13)得知，每一個音框的最前面都有一個標頭 (Header)，固定為 4 個位元組 (Byte)；接著是副資訊 (Side Information)，若是單聲道則有 17 個位元組，若是雙聲道則有 32 個位元組；最後才是主要資料 (main_data) 段。而其中的循環冗贅核對碼 (Cyclic Redundancy Code) 為選擇性存在，由使用者決定。

由圖(2.14)可以知道，一個主要資料段是由兩個細群 (Granule) 所組成，每一個細群又包含了兩個聲道 (若聲音為單聲道就只有一個)，每個聲道均有量尺因子及量化編碼係數。

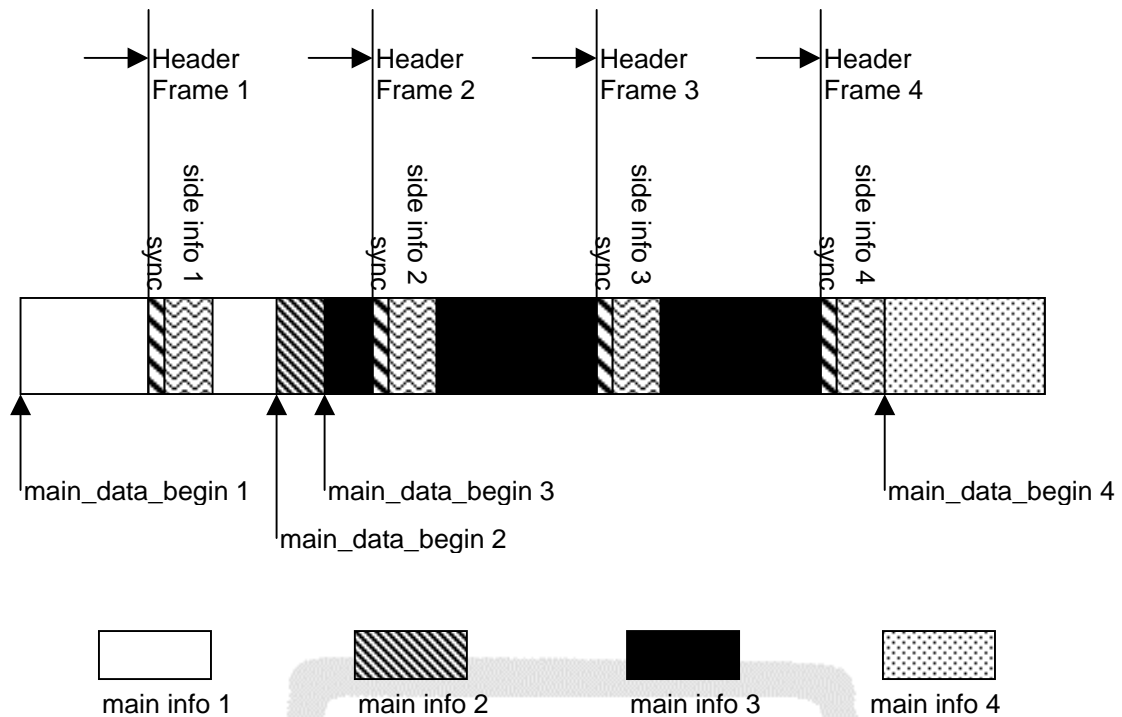
在副資訊中有一個指標，稱為 main_data_begin，共有 9 個位元，其功用為決定主要資料段的起始位置 (第一個位元組)，如圖(2.15)所示。

B. 解量化係數

由圖(2.16)可以知道量化後的係數分配，rzero 區域均為 0，所以真正解量化的過程只有 big_vales 區域及 count1 區域。

解量化係數的分為短區塊及長區塊的解碼，如果是短區塊時，其公式為 (2-22)，若為長區塊，則為式子 (2-23)。

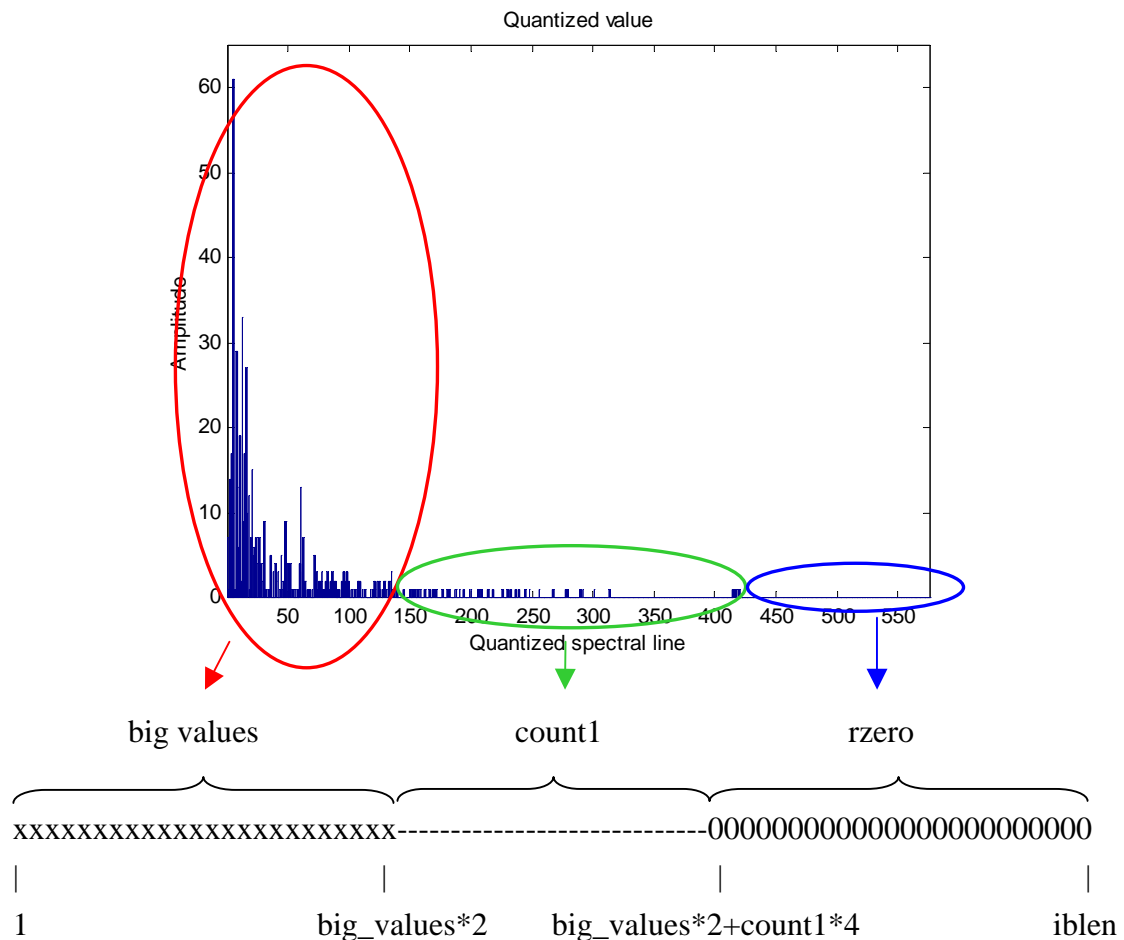
$$x_{r_i} = \text{sign}(is_i) * |is_i|^{\frac{4}{3}} * 2^{\frac{1}{4}(\text{global_gain}[gr]-210-8*\text{subblock_gain}[\text{window}][gr])} * 2^{-(\text{scalefac_multiplier}*\text{scalefac_s}[gr][ch][\text{fb}][\text{window}])} \quad (2-22)$$



圖(2.15) Layer III 的位元流組成及 main_data_begin

$$\begin{aligned}
 xr_i = & \text{sign}(is_i) * |is_i|^{\frac{4}{3}} * 2^{\frac{1}{4}(global_gain[gr]-210)} \\
 & * 2^{-(scalefac_multiplier*(scalefac_l[sfb][ch][gr]+preflag[gr]*pretab[sfb]))}
 \end{aligned}
 \tag{2-23}$$

其中 is_i 為經過霍夫曼解碼後的值， $global_gain[gr] - 210$ 就等於量化步階的大小， $subblock_gain$ 只有在短區塊中出現，代表一個增益的修正， $scalefac_multiplier$ 與 $scalefac_l$ 及 $scalefac_s$ 相乘所代表的就是量尺因子的大小。 $preflag * pretab$ 代表高頻的增益。以上各係數的大小及增益的改變都由編碼端所調整。

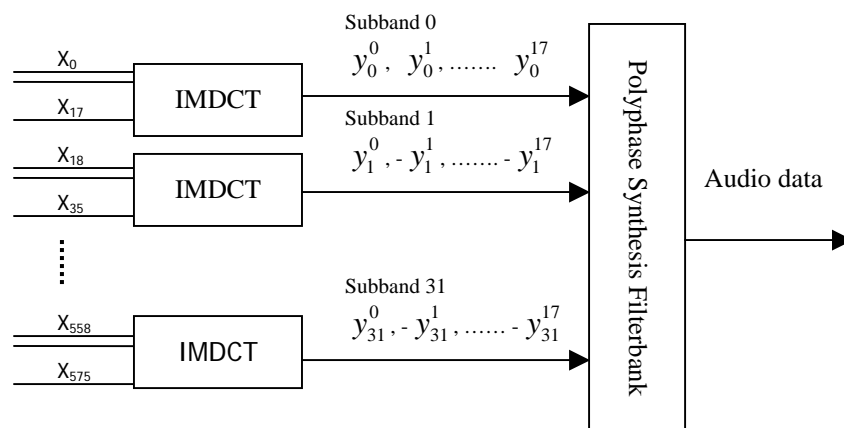


其中 000 的區域值都為零，稱為 rzero 區域，其數目為 2 的倍數。
 而 --- 的區域值為 -1,0,或 +1，稱為 count1 區域，其數目為 4 的倍數。區域 xxx 的值則沒有界線，稱為 big_value 區域，其數目為 2 的倍數。
 iblen 為 576。
 因此， $big_values * 2 + count1 * 4 + rzero * 2 = 576$ 。

圖(2.16) 量化後的係數分佈

C. 混合式合成濾波器組

混合式合成濾波器組主要分為兩個部分，一個是反修正式離散餘弦轉換 (IMDCT)，另一個是多項式合成濾波器 (Polyphase Synthesis Filter)。與編碼器的混合式分析濾波器組相對應，如圖(2.17)所示。



圖(2.17) 多項式合成濾波器組

➤ C-1 逆修正式離散餘弦轉換

其與修正式離散餘弦轉換互為反轉換，其反轉換公式為式子 (2-24) 所示。

$$x_i = \sum_{k=0}^{\frac{n-1}{2}} X_k \cos\left(\frac{n}{2\pi} \left(2i+1 + \frac{n}{2}\right)(2k+1)\right) \quad \text{for } i=0 \text{ to } n-1 \quad (2-24)$$

其中 n 若是短區塊，則 $n = 12$ ； n 若為長區塊，則 $n = 36$ 。逆修正式離散餘弦轉換目的除了將訊號恢復成時域，並利用重疊相加以消除時域上的重疊效應。

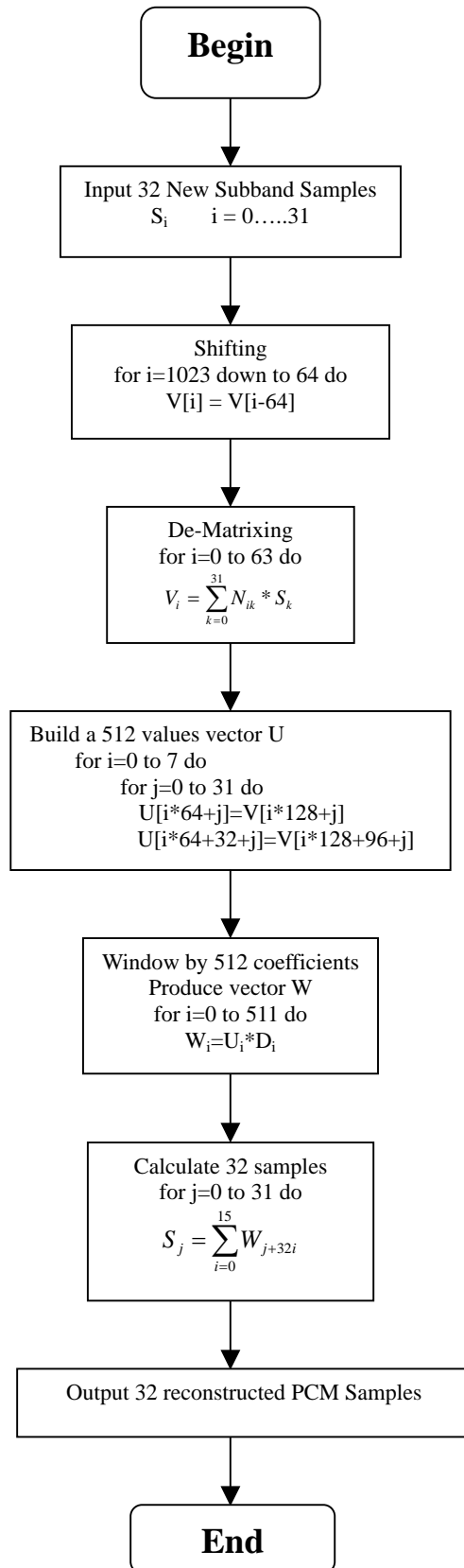
➤ C-2 多項式合成濾波器

將經過多項式分析濾波器的訊號還原，如圖(2.18)所示。將 32 個子頻帶訊號，經由反矩陣運算 (De-Matrixing) 後，如式子 (2-25) 所示，

$$V_i = \sum_{k=0}^{31} N_{ik} * S_k \quad \text{for } i=0 \text{ to } 63$$

$$N_{ik} = \cos[(16+i)(2k+1)\pi/64] \quad (2-25)$$

產生 Y 向量，再經過重排可以得到 U 向量，再與 D 加權窗相乘，最後可以得到 32 個重建訊號 (S_j)。



圖(2.18) 多項式合成濾波器

第三章 MPEG-1 Layer III 在低位元率的效能

3.1 簡介

在音訊壓縮編碼的目的無非就是為了要得到明顯的聲音品質，使試聽者難以區分原音與經過壓縮編碼處理後的聲音。不可否認地，提高位元率可以加強聲音的品質；相反地，一旦位元率降低到某一個程度時，經過編碼處理的聲音聽起來會較難讓人們所接受。因此在這一章節將會討論 MPEG-1 Layer III 位元率與其效能的關係。

目前較為熱門的音訊壓縮編碼有 MPEG-1 Layer II、MPEG-1 Layer III、MPEG-2 AAC(Advanced Audio Coding) 及 AT&T PAC 和 Dolby AC-3 這幾種編碼標準。表(3.1)為各音訊編碼器的品質比較[6]。

由表(3.1)、(3.2)可以知道各音訊編碼較佳的位元率及其聲音品質。MPEG-1 Layer III 以雙聲道的聲音編碼，位元率在 128 kbps 以上可得到較好的聲音品質，相反地，若是在 128 kbps 以下則會讓人越來越無法忍受品質的降低，以致於一般的 MPEG-1 Layer III 壓縮過的位元率最低都保持在 128 kbps 這個臨界點[註四]。

[註四]：此處的比較標準都在高取樣頻率的情況下，如：44.1 kHz。

表(3.1) 各音訊壓縮編碼比較 (雙聲道)

Group	Algorithm	Rate (kbps)	Mean Diff. Grade
1	AAC	128	-0.47
	AC-3	192	-0.52
2	PAC	160	-0.82
3	PAC	128	-1.03
	AC-3	160	-1.04
	AAC	96	-1.15
	MP-1 L2	192	-1.18
4	MP-1 L3	128	-1.73
	MP-1 L2	160	-1.75
	PAC	96	-1.83
5	AC-3	128	-2.11
	MP-1 L2	128	-2.14
6	PAC	64	-3.09

表(3.2) ITU-R 的試聽評估準則

Absolute Grade	5.0	Imperceptible	0.0	Difference Grade
	4.9	Perceptible but NOT Annoying	-0.1	
	4.8		-0.2	
	4.7		-0.3	
	4.6		-0.4	
	4.5		-0.5	
	4.4		-0.6	
	4.3		-0.7	
	4.2		-0.8	
	4.1		-0.9	
	4.0		-1.0	
	3.9	Slightly Annoying	-1.1	
	3.8		-1.2	
	3.7		-1.3	
	3.6		-1.4	
	3.5		-1.5	
	3.4		-1.6	
	3.3		-1.7	
	3.2		-1.8	
	3.1		-1.9	
3.0	-2.0			
2.9	Annoying	-2.1		
2.8		-2.2		
2.7		-2.3		
2.6		-2.4		
2.5		-2.5		
2.4		-2.6		
2.3		-2.7		
2.2		-2.8		
2.1		-2.9		
2.0		-3.0		
1.9	Very Annoying	-3.1		
1.8		-3.2		
1.7		-3.3		
1.6		-3.4		
1.5		-3.5		
1.4		-3.6		
1.3		-3.7		
1.2		-3.8		
1.1		-3.9		
1.0		-4.0		

3.2 MPEG-1 Layer III 在低位元率的效能

MPEG-1 Layer III 經過量化編碼後會區分為三個區域，如 2.3 節中圖 (2.16) 所示。為了要能夠達到所需的位元率，在內疊代迴圈內會調整量化階的大小，但也因此影響了量化的結果：big_values 區域、count1 區域與 rzero 區域的大小變動，當然，聲音的品質亦會受到影響。由表(3.3)可以明顯看出來影響的程度，當位元率降的越低，其 SEGPSNR (Segmental Peak Signal to Noise Ration) 與 SEGSNR (Segmental Signal to Noise Ration) 也隨著降低，其中 SEGPSNR 與 SEGSNR 的定義如式子 (3-1) 及 (3-2) 所示。

$$PSNR = \frac{V_{peak_value}^2}{\frac{1}{N} \sum (V_s - \hat{V}_s)^2}$$
$$SEGPSNR = \frac{1}{N} \sum PSNR_i \quad (3-1)$$

$$SNR = \frac{\sum V_s^2}{\sum (V_s - \hat{V}_s)^2}$$
$$SEGSNR = \frac{1}{N} \sum SNR_i \quad (3-2)$$

其中 V_{peak_value} 是訊號的最大值， V_s 為訊號， \hat{V}_s 為重建訊號，所以 $(V_s - \hat{V}_s)$ 為雜訊。雖然一般採用的評估標準大多是 SEGSNR，本論文特別加入 SEGPSNR 來當作評估準則，主要是因為當聲音訊號越大，所容忍的誤差也越大，還有加上聲音有遮罩效應的考量，所以加入 SEGPSNR 來評估會有較準確之評量。例如：若小訊號的誤差過大 (超過其能量一半)，若

是在頻域或時域上，其周遭有一個大訊號，這大訊號將會遮蔽小訊號的誤差，而不讓人所察覺。所以用 SEGPSNR 客觀評量較一般訊雜比 SEGSNR 接近真正人耳聽覺許多。

表(3.3) 不同位元率量化後的影響

位元率 (bps)	Big_values 區 (spectra lines)	Count1 區 (spectra lines)	SEGPSNR (dB)	SEGSNR (dB)
256 k	410	84	85.0	38.8
128 k	166	92	67.6	21.4
96 k	190	176	63.0	16.8
80 k	148	240	61.7	15.5
64 k	72	264	60.3	14.1

(a) 爵士樂 (sax)

位元率 (bps)	Big_values 區 (spectra lines)	Count1 區 (spectra lines)	SEGPSNR (dB)	SEGSNR (dB)
256 k	370	92	73.4	32.2
128 k	278	116	56.4	15.2
96 k	148	216	51.3	10.1
80 k	88	244	49.5	8.3
64 k	44	228	48.4	7.2

(b) 流行樂 (歌曲)

位元率 (bps)	Big_values 區 (spectra lines)	Count1 區 (spectra lines)	SEGPSNR (dB)	SEGSNR (dB)
256 k	418	84	87.6	36.7
128 k	330	84	72.4	21.4
96 k	226	172	66.4	15.4
80 k	152	220	64.0	13.1
64 k	84	220	61.7	10.8

(c) 古典樂 (弦樂)

位元率 (bps)	Big_values 區 (spectra lines)	Count1 區 (spectra lines)	SEGPSNR (dB)	SEGSNR (dB)
256 k	414	132	86.5	47.9
128 k	162	244	70.3	31.7
96 k	124	284	62.8	24.3
80 k	90	312	58.6	20.0
64 k	58	292	53.8	15.3

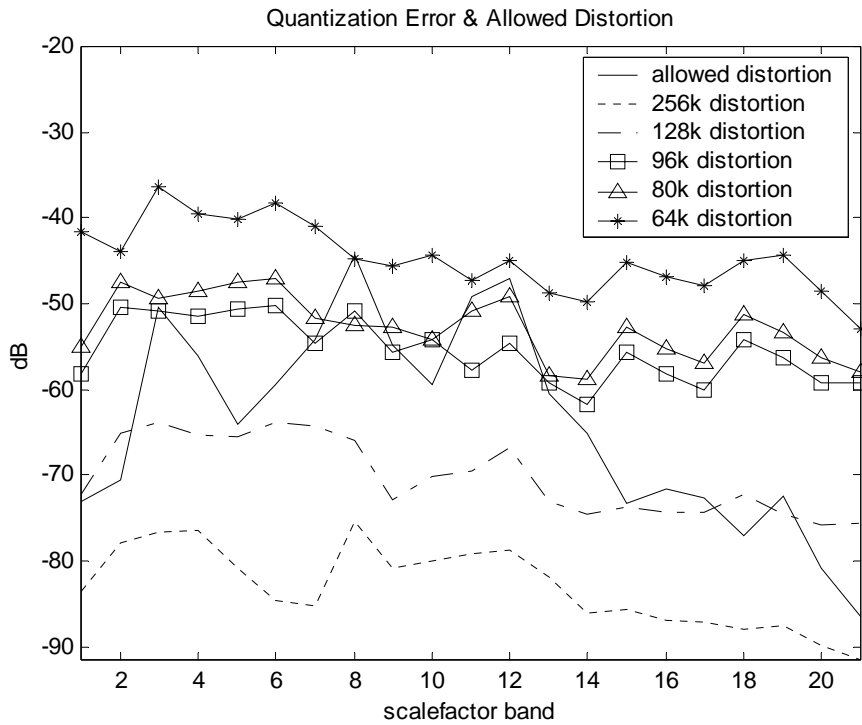
(d) 輕音樂 (鋼琴)

由表(3.3)可以看出，big_values 區域隨著位元率降低而降低，count1 區域隨著位元率降低而升高；主要是因為 count1 區域所需編碼的位元數較少，反之，要將 big_values 區域編碼，對位元數有相當大的需求。雖然這樣編碼可以達到低位元率的目的，但也因為如此，量化誤差也隨之提升，以致於編碼出來的聲音聽起來較不悅耳。

3.3 對 MPEG-1 Layer III 在低位元率的效能改進

由 3.2 節可以知道，位元率的降低對聲音品質的影響；又從 2.2 節 C 疊代迴圈的部分得知，要達到位元率的需求是由內疊代迴圈來決定，但這也影響了量化的誤差的大小；除此之外，如果量化誤差無法利用外疊代迴圈來控制壓低，就已經達到了迴圈終止的條件，那麼量化誤差將會一直存在。如圖(3.1)可以明顯看出位元率越低，量化誤差在量尺因子頻帶中超過所允許的誤差也越多，同樣地，表(3.3)也說明了編碼後的聲音品質也隨之降低。

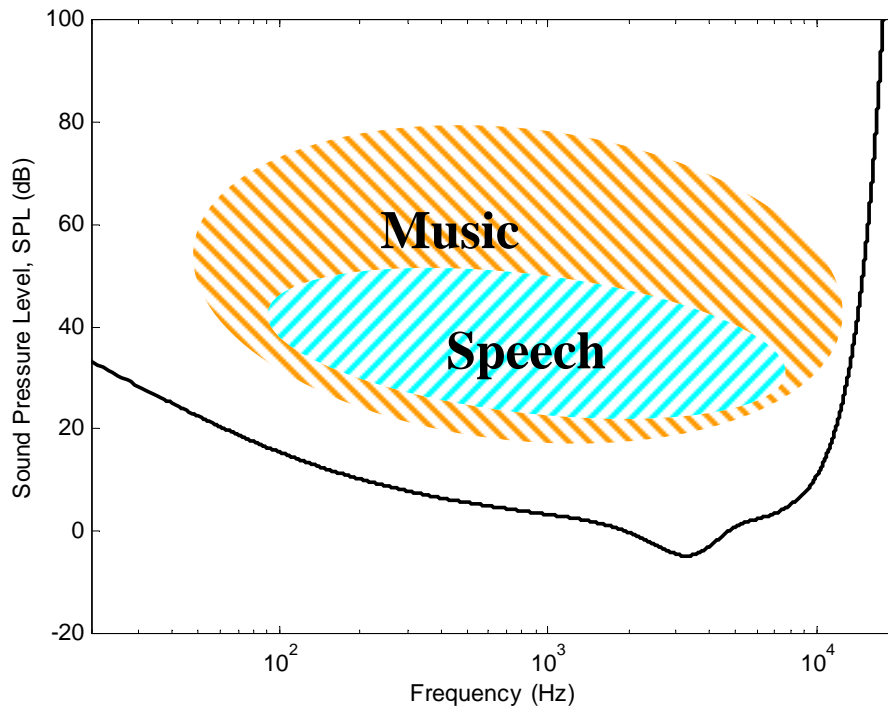
在 1.3 節中有提到，人所能聽到的頻率範圍在 20 Hz 至 20 kHz 之間，但人耳對 2 kHz 至 4 kHz 這一段的頻率較為敏感，在這一段頻率只要有一點點的誤差很容易讓人聽起來不悅耳；相對應至量尺因子頻帶約第 10 個頻帶至第 13 個頻帶之間（取樣頻率為 44.1 kHz）。為了要達到低位元率的目的，而且希望能夠編碼後的結果讓人聽起來滿意。目標為不因降低



圖(3.1) 不同位元率的量化誤差與可允許的誤差比較

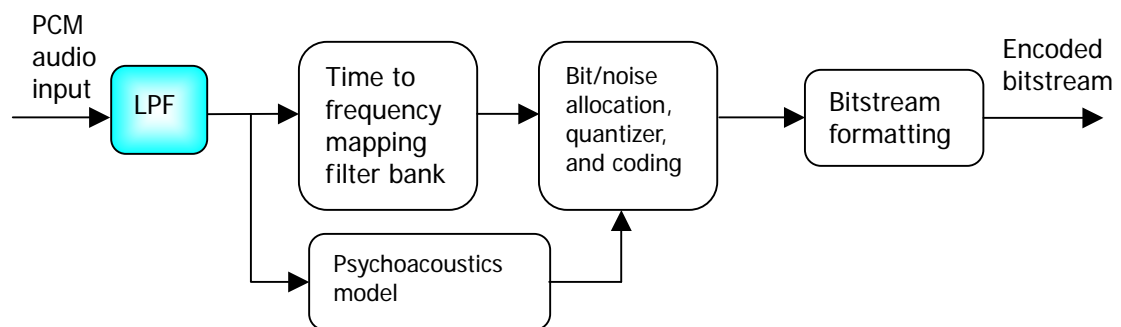
位元率而造成量尺因子頻帶的量化誤差過大，最好能夠低於所允許的誤差之內；當然，如果能夠讓 40 Hz 至 10 kHz 之間的誤差越低越好，因為在這頻帶以外的聲音，人聽起來更不敏感；而且一般的語音頻寬大約在 100 Hz 至 7 kHz 之間。而樂器所演奏的音樂頻寬範圍較廣，分佈範圍大約在 40 Hz 至 10 kHz 之間，如圖(3.2)所示，而高於 10 kHz 的聲音也大多是諧波(Harmonic)所產生的，當然也有少數基頻(Fundamental frequency) 是在 10 kHz 以上的[1]。

40 Hz 至 10 kHz 相對應至量尺因子頻帶約第 1 個頻帶至第 18 個頻帶之間（取樣頻率為 44.1 kHz）。因此，只要能夠讓可以編碼的位元數集中



圖(3.2) 語音與音樂的頻率範圍

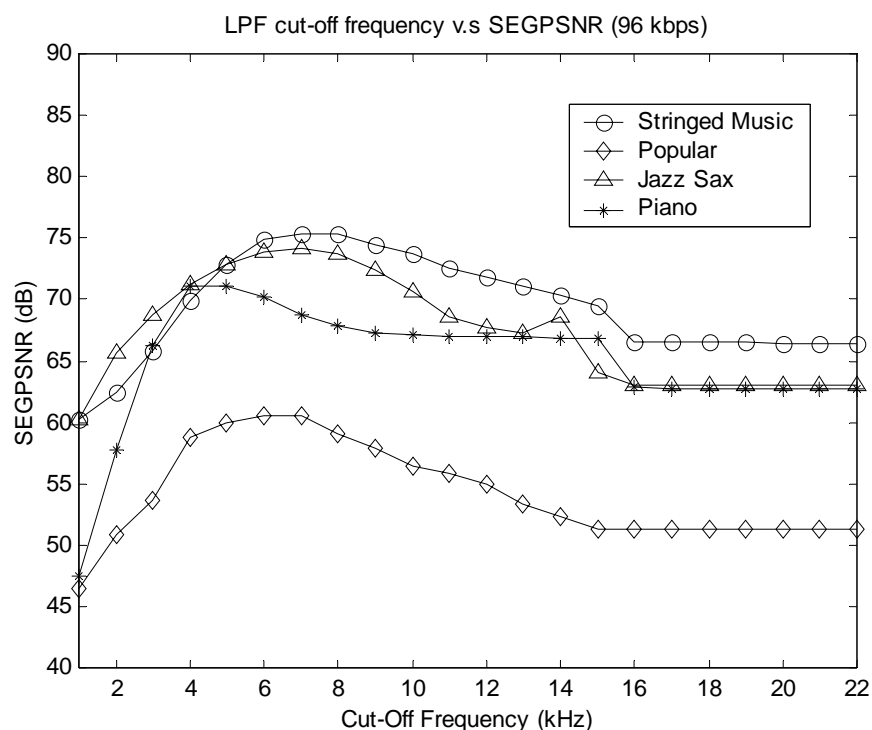
在 10 kHz 或某一個頻率以下的聲音，就算是採用低位元率的編碼，其結果也可以讓人所接受；最簡單的方法就是在聲音輸入 MPEG-1 Layer III 編碼器之前先通過一個低通濾波器，如圖(3.3)所示，



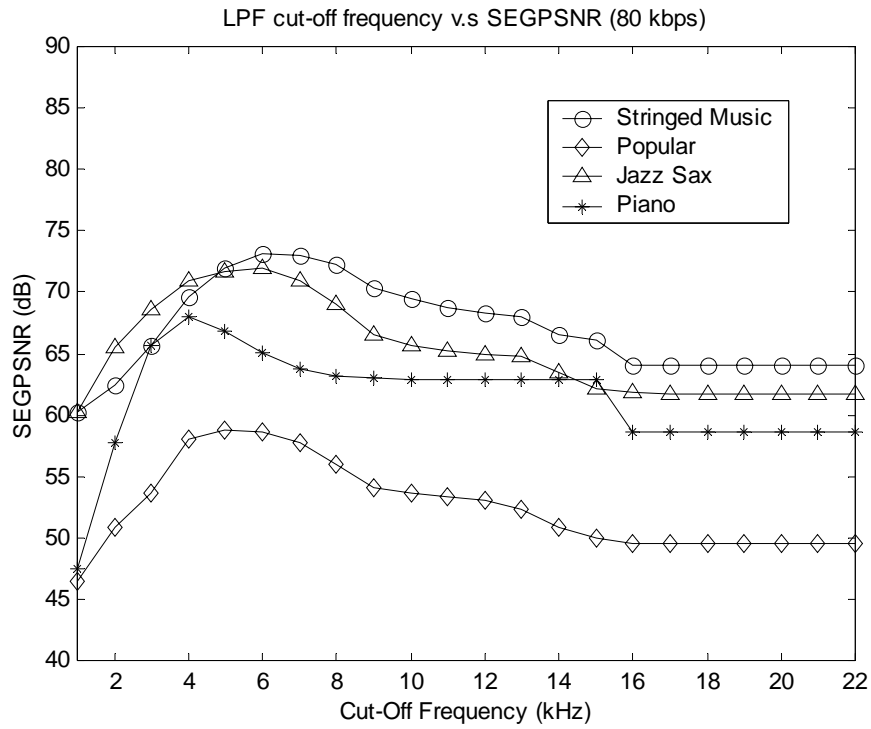
圖(3.3) 修正式 MPEG 音訊編碼器架構

低通濾波器的截止頻率 (Cut-off Frequency) 與 SEGPSNR 的關係，如圖 (3.4)、(3.5)、以及(3.6)所示。其中頻率為 22.05 kHz 時，代表沒有通過低通濾波器，即原始 ISO MPEG-1 Layer III 所編碼的效果。

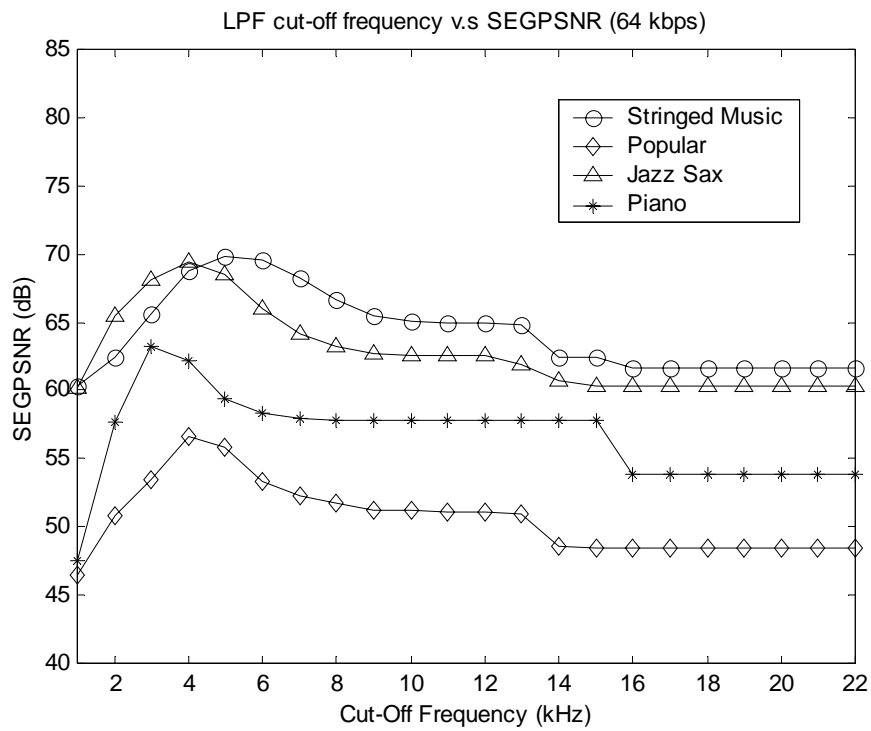
由圖(3.4)、(3.5)、及(3.6)可知，在輸入 MPEG-1 Layer III 編碼器之前加上一個低通濾波器，當截止頻率設在 4 kHz 至 8 kHz 其 SEGPSNR 都比 22.05 kHz 的 SEGPSNR 要高，可知其聲音品質的提升。這是因為把原先編碼器將所有的位元數平均分散在各個頻率中，以致於位元數不足以表示所有頻率的聲音，而導致量化誤差過大，造成不悅耳的聲音；當加入一個低通濾波器時，可以將位元數全部集中在低頻的聲音，盡量避免低頻的量化誤差出現。



圖(3.4) 位元率 96 kbps 經過低通濾波器後的 SEGPSNR



圖(3.5) 位元率 80 kbps 經過低通濾波器後的 SEGPSNR



圖(3.6) 位元率 64 kbps 經過低通濾波器後的 SEGPSNR

而低通濾波器的截止頻率應該設為多少也可由以上各圖得知，最佳的截止頻率應該設於 SEGPSNR 最高點時的截止頻率。因為每種音樂的特性不同，所以當位元率為 96 kbps 時，其截止頻率可設為在 5 kHz 至 10 kHz 之間；位元率為 80 kbps 時，其截止頻率可設為 4 kHz 至 8 kHz 之間；而位元率為 64 kbps 時，其截止頻率可設為 3 kHz 至 7 kHz 之間。

不過這個方法的缺點，就是當聲音訊號經過低通濾波器後，高頻的聲音不被人所聽到，只能確保低頻的聲音保持在不錯的品質，解決低頻聲音的量化誤差，不能改善高頻的編碼品質。

3.4 頻譜截斷與低通濾波器

當聲音通過低通濾波器，相當於在頻率域將截止頻率以上的頻譜截斷，留下截止頻率以下的頻譜。

如此可簡化通過低通濾波器的 MPEG-1 Layer III 編碼器，只要訊號通過多項式分析濾波器時，在修正式離散餘弦轉換域中將頻譜截斷在與低通濾波器相同的截止頻率，如此就等效於訊號通過低通濾波器的音效編碼。

在修正式離散餘弦轉換域中的主要資料共有 576 (0 ~ 575) 的頻譜，代表著 0 kHz 至 $f_s/2$ kHz (f_s 為取樣頻率)，表(3.4)為取樣頻率 44.1 kHz 的頻譜與頻率及量尺因子頻帶的對照表。

表(3.4) 頻率與頻譜及量尺因子頻帶的關係

Frequency (kHz)	Spectra Line	Scalefactor band
1	26	6
2	52	10
3	78	12
4	104	13
5	130	14
6	156	15
7	182	16
8	208	17
9	236	17
10	262	18
11	288	19
12	314	19
13	340	19
14	366	20
15	392	20
16	418	21
17	444	21
18	470	21
19	496	21
20	522	21
22.05	575	21

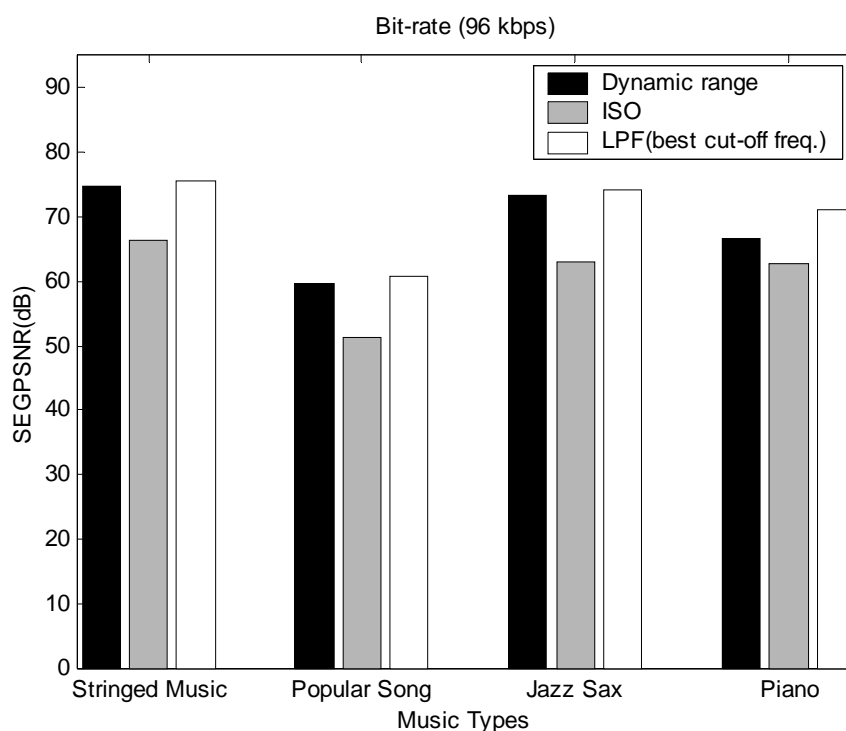
3.5 適應性頻譜截斷

由圖(3.4)、(3.5)、(3.6)所知，不同的音樂（不同的樂器或聲音），在不同的位元率下，其低通率波器的最佳截止頻率（圖中曲線 SEGPSNR 的最高點）也有所不同。為了要適合不同種類的音樂，將採用適應性頻譜截斷，可盡量避免因為不同頻率的聲音，在量化編碼時，對截止頻率的需求有所不同。

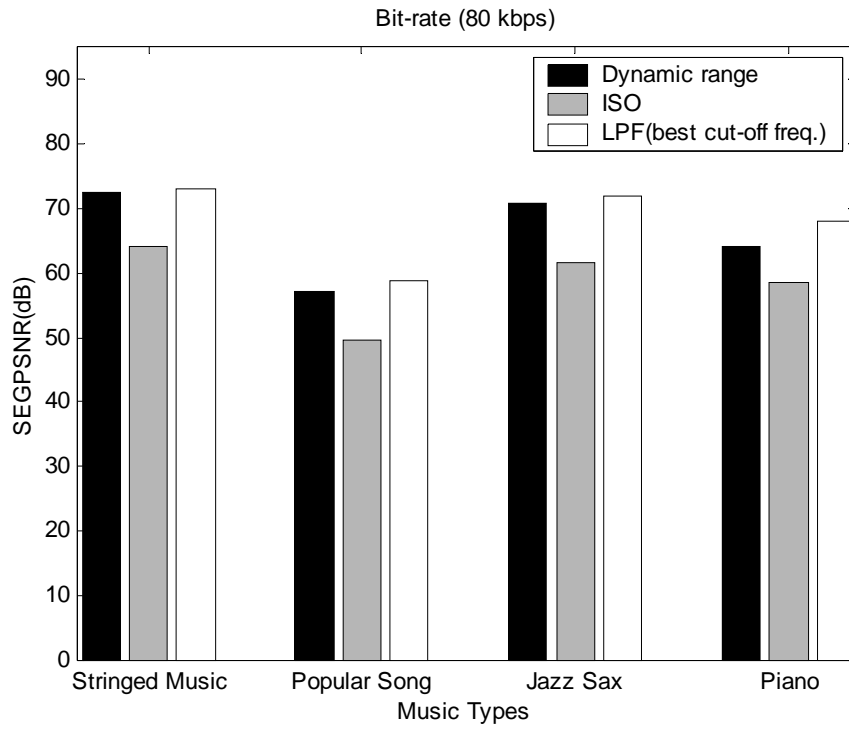
由 3.2 節可知，當位元率降低，其 big_values 區域與 count1 區域為了符合位元率之需求也隨之降低。這也表示著經過內疊代迴圈後，量化後的值經由霍夫曼編碼後的位元在目前所擁有的位元空間範圍內。也就是

指在這一個音框中，量化後所能表示的最大頻譜區域範圍由 0 至 $2 * \text{big_values} + 4 * \text{count1}$ 之間，則低通率波器的截止頻率若設在此處不會影響編碼後的結果。

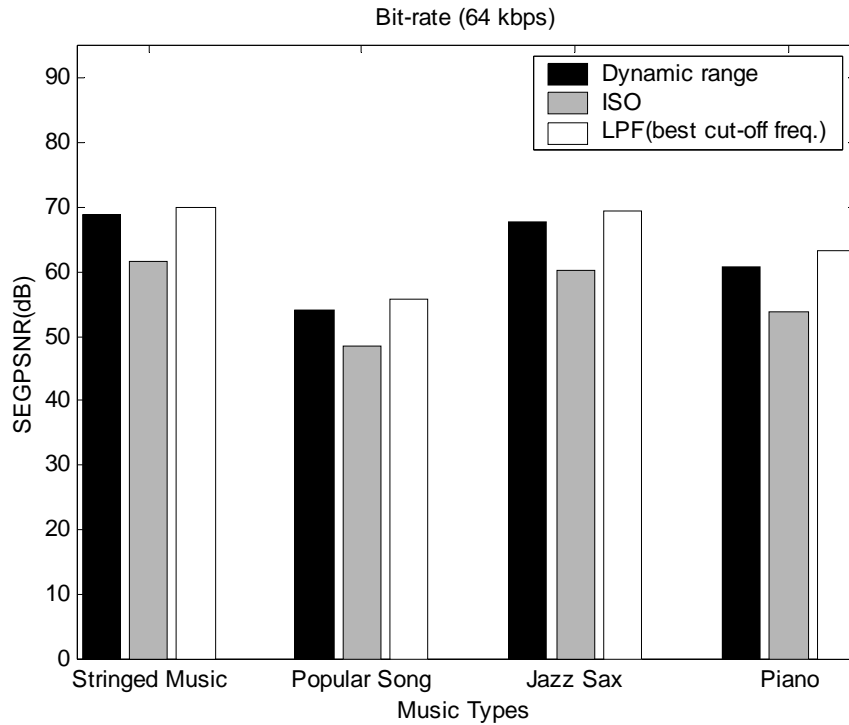
為了要讓大多數的位元率集中在低頻，先經過一次內疊代迴圈後，再將截止頻率（頻譜截斷點）設在 $2 * \text{big_values} + 2 * \text{count1}$ ；因為在低位元率編碼時，通常 count1 區域較大，取其一半加上 big_values 區域則為一個參考的截止頻率（頻譜截斷點）。其結果與原始（ISO）編碼器及聲音訊號通過較佳的截止頻率之低通率波器編碼器比較，如圖(3.8)、(3.9)、以及(3.10)所示。



圖(3.8) 位元率 96 kbps 不同音樂之效能比較



圖(3.9) 位元率 80 kbps 不同音樂之效能比較



圖(3.10) 位元率 64 kbps 不同音樂之效能比較

由圖(3.8)、(3.9)、及(3.10)可以得知，利用適應性頻譜截斷法可以改善 MPEG-1 Layer III 音訊編碼器在低位元率時的編碼效果，並且其 SEGPSNR 值與通過低通率波器的最大 SEGPSNR 值相當接近。

第四章 MPEG-1 Layer III 之高頻補償

4.1 簡介

在高取樣頻率的情況下，如：48 kHz 或 44.1 kHz，如果希望要達到低位元率的音訊壓縮編碼，常常會因為位元數的不足，而導致量化誤差過大，聲音品質也隨著下降。

若想改善這樣子的個情況，可在聲音訊號輸入 MPEG-1 Layer III 編碼器之前，加上一個低通濾波器；當聲音訊號通過低通濾波器時，高頻的聲音將被消弭，所造成的結果是將位元數的分配主要配置在低頻，避免低頻的量化誤差過大；同樣地，頻譜截斷法也是相同的效果。

固然利用低通濾波器或頻譜截斷法可以提升聲音的低頻品質，但缺點就是高頻的聲音消失了，音樂聽起來也比較沒有豐富性。為了要彌補這個缺點，在這一個章節將會提到高頻補償的方法，其方法可使經過低通濾波器後編解碼出來的聲音不再只有低頻的成分。

4.2 高頻補償法

為了要彌補含有低通濾波器或經過頻譜截斷法的 MPEG-1 Layer III 音訊編碼器的缺點，必須要將高頻的聲音成分加以補償，以達到高品質音訊壓縮編碼的目的。當然，要補償的高頻資訊 (Information) 就要加入

在 MPEG-1 Layer III 的位元流中；又要防止加入的資訊會影響編碼位元率，所以這一部份的資訊位元量必須加以考慮，以維持位元率的不變，而音質提升。

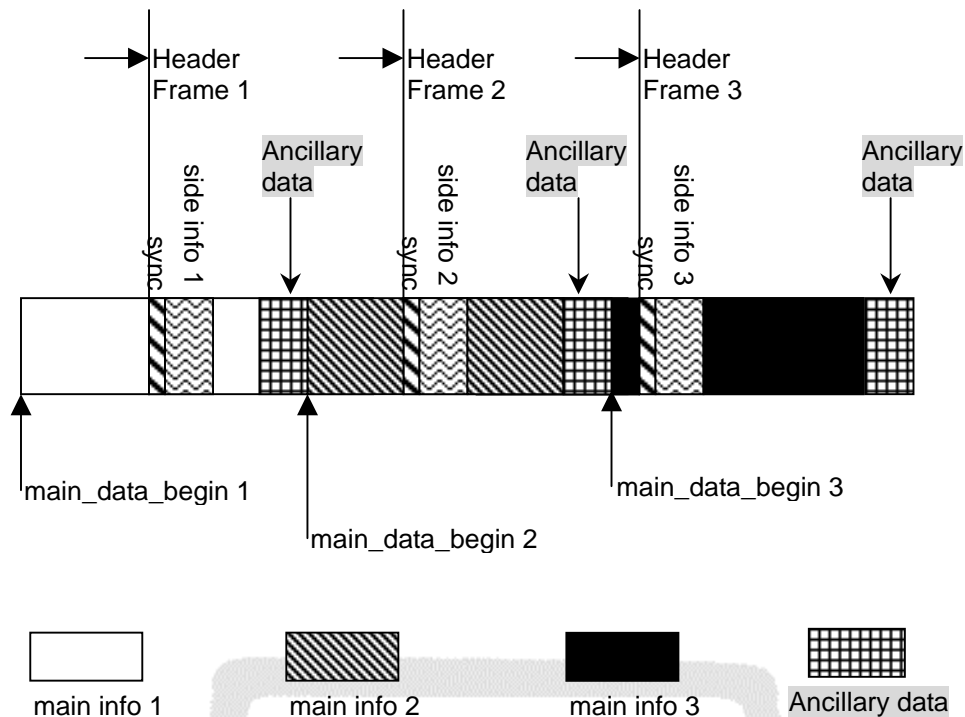
A. MPEG-1 Layer III 的位元流格式修改

在 MPEG-1 Layer III 的位元流格式中並沒有輔助的資料段 (ancillary data) [註五]，如圖(2.13)所示。如今為了要將高頻補償的資訊加入位元流中，又希望能夠讓一般的解碼器順利解碼，例如：winamp 播放器，唯一的方法就是利用 main_data_begin 的特性，如圖(2.15)所示，將高頻資訊加入輔助的資料段中，而不影響整個 MPEG-1 Layer III 的位元流解碼，如圖(4.1)所示。除此之外，要將輔助資料段所佔的位元數亦要考慮進編碼後的位元率中，相對地，主要資料段的位元數就會有些微的減少。

B. 高頻補償法

由量化的結果可以顯示出，MPEG-1 Layer III 的量化編碼較注重聲音的低頻成分 (big_values 區域)；中間頻帶則用少量的位元數來做編碼 (count1 區域)；最後捨去聲音高頻的成分 (rzero 區域) 當 MPEG-1 Layer III 的位元率降的越低，為了符合位元率的需求，量化

[註五]：在 MPEG-1 標準中，只有 Layer I 及 Layer II 才有定義輔助資料段。



圖(4.1) 修正式 Layer III 的位元流組成

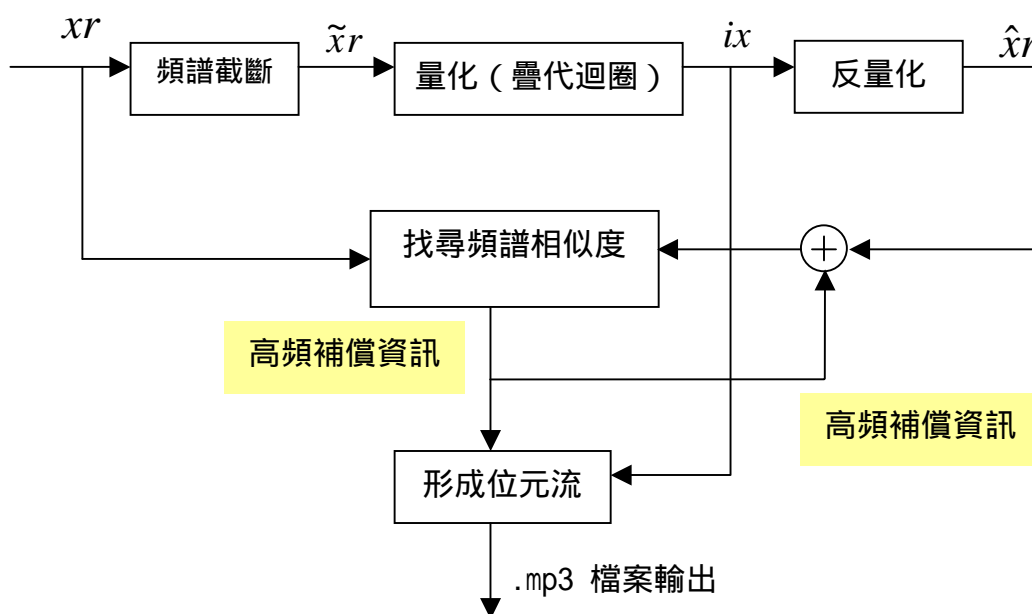
步階的大小改變使得 big_values 的區域會減少，但也造成了不少的量化誤差；一旦這些量化誤差大於人耳朵所能容忍的範圍內，也就使的耳朵更容易聽出不悅耳的雜音。

由於人的耳朵對低頻的敏感度較大，如果低頻的量化編碼效果很好，高頻的誤差較大，則音樂聽起來較可以接受，影響不大；相反地，如果低頻的部分量化編碼表現較差，而高頻的部分表現較好，由於人類的聽覺對高、低頻的聲音敏感度不同，則聲音聽起來則較難令人所接受。經過低通濾波器後或在頻率域中截斷頻譜來做音訊編碼，可以確保低頻的品質，也就是降低在低頻的量化誤差。當低頻的品質良好，要對高頻補償就可以利用頻譜之間的相似程度來用低頻來做高頻補償的動作。

➤ B-1 編碼端的高頻補償

在疊代迴圈(量化)的輸入端為經過修正式離散餘弦轉換的 576 條頻譜，稱為 x_r ；經過頻譜截斷後(等效低通濾波器)的頻譜稱之為 \tilde{x}_r ；將 \tilde{x}_r 經過疊代迴圈量化過後的頻譜稱之為 i_x 。其中 \tilde{x}_r 與 i_x 兩者因為經過頻譜截斷，所以高頻的聲音成分已經消除了，所剩下的只有在截止頻率以下的低頻成分；而 x_r 為原始的頻譜。

為了要與解碼端的高頻補償一致，將會利用 i_x 經過反量化後的結果，稱為 \hat{x}_r ，來補償 x_r 在截止頻率以上的頻譜，如此就可以達到高頻補償的目的了，如圖(4.2)所示。



圖(4.2) 利用頻譜相似度來做高頻補償

利用式子 (4-1) 可以計算出 $\hat{x}r$ 與 xr 之間在頻率域上某一段較為相似的頻譜，

$$cor_j(\kappa) = \frac{\sum_{i=0}^{bd_j} xr(idx+i) * \hat{x}r(idx-k+i)}{\sqrt{\sum_{i=0}^{bd_j} xr(idx+i) * xr(idx+i)}} \quad \text{for } k = 0 \text{ to } idx \quad (4-1)$$

$$idx = 2 * big_values + 4 * count1 + d_j$$

$$if \begin{cases} j = 0 \Rightarrow d_j = 0 \\ j > 0 \Rightarrow d_j = d_{j-1} + p(j-1) \end{cases}$$

$$if \begin{cases} k \geq p(j) \Rightarrow bd_j = p(j) \\ k < p(j) \Rightarrow bd_j = k \end{cases}$$

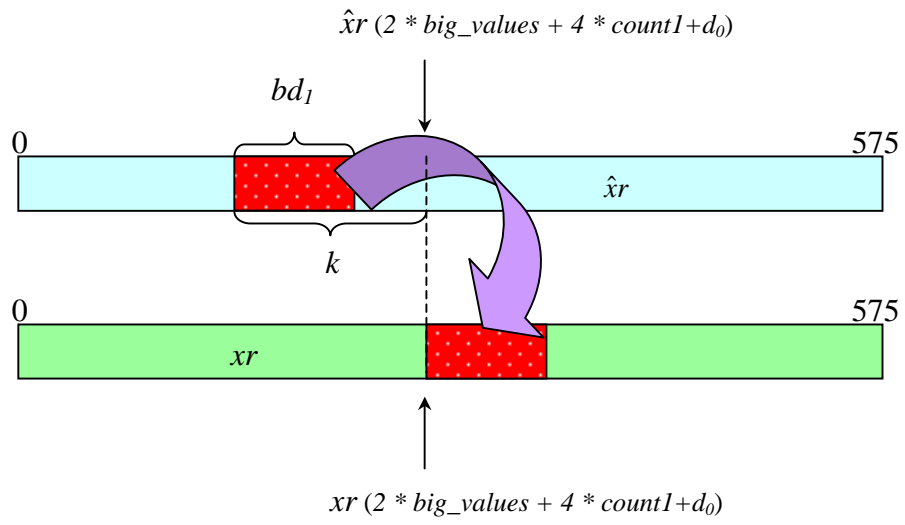
$$p(j) = \{24, 28, 34, 42, 50, 54, 76\}, \quad \text{for } j = 0 \text{ to } 6$$

其中 bd_j 為區塊範圍 (Block Boundary)， $p(j)$ 為找尋的段落大小， j 為找尋的次數， k 為與 $2 * big_values + 4 * count1 + d_j$ 之間的距離。以圖(4.3)說明。

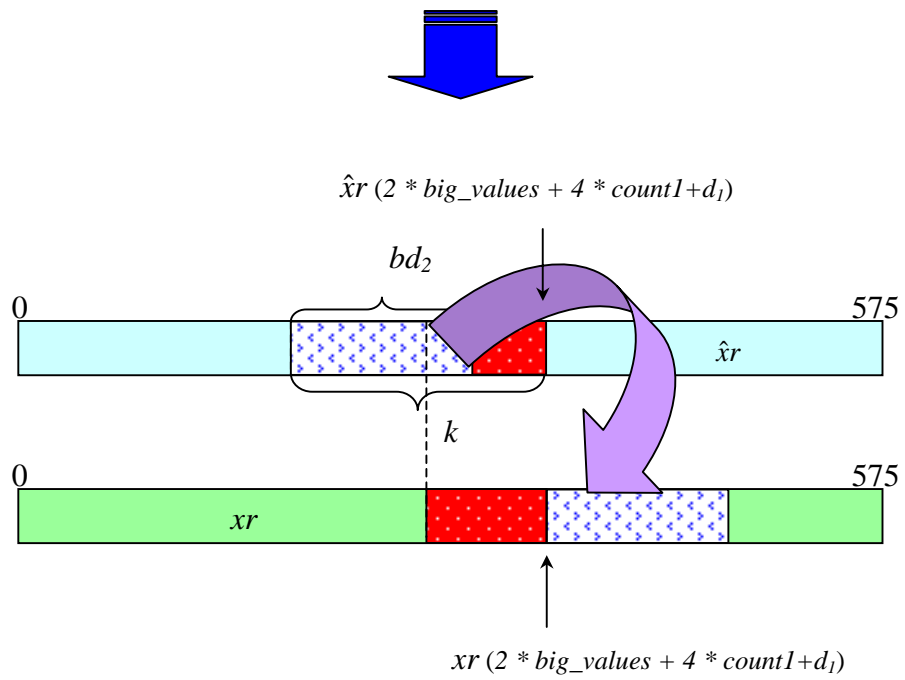
若要找出最相似的頻譜區段只要找出 $max(cor_j(k))$ 即可。如此可知道在 $\hat{x}r$ 由第幾 (k) 點開始的區塊範圍 (bd_j) 與 xr 的欲補償的頻譜區塊範圍最為接近。重複尋找相似頻譜區塊，就可以將高頻被濾掉的訊號逐一補齊。

一旦找到了適當的高頻補償區塊後，接下來就要找補償區塊的增益 (gain) 了，其增益 (G_j) 的大小由式子 (4-3) 可以得知， s 為所找到的相似頻譜的起始點，如式子 (4-2) 所示，

High frequency compensation (1st step)



High frequency compensation (2nd step)



圖(4.3) 頻譜相似度的求法

$$\text{if } \max(\text{cor}_j(k)) \text{ then } s = k \quad (4-2)$$

$$G_j = \frac{\text{cor}_j(s)}{\sqrt{\sum_{i=0}^{bd_j} xr(idx+i) * xr(idx+i)}} \quad (4-3)$$

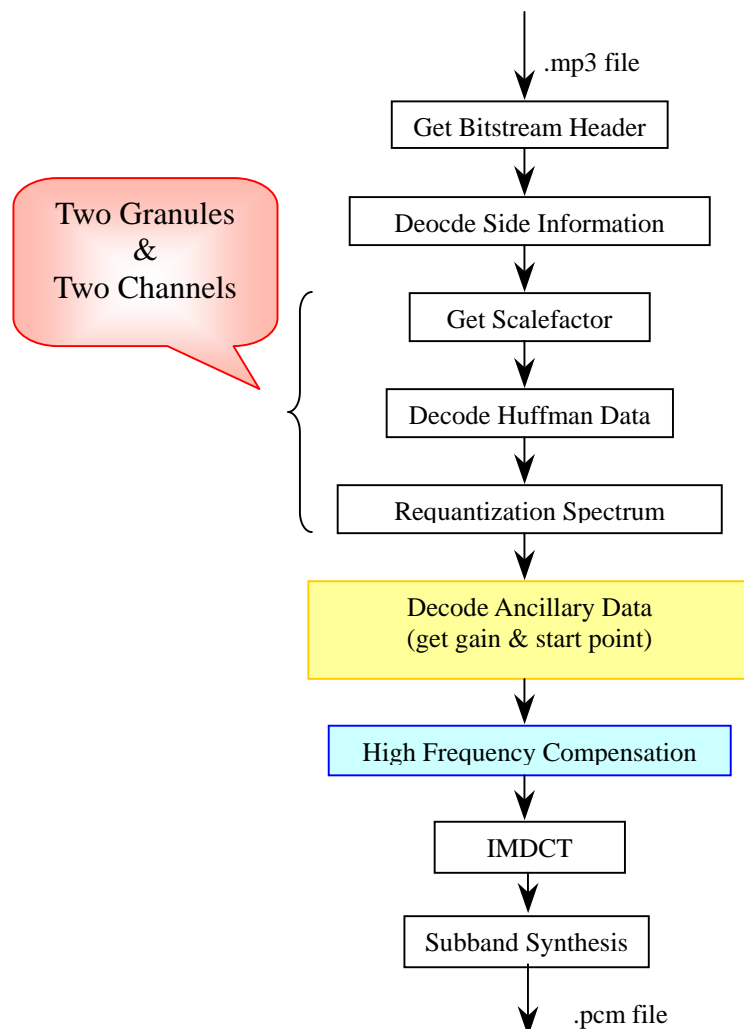
其中 bd_j 為相似的頻譜區段大小， idx 與 (4-1) 式的定義相同，為欲補償區塊的起始位置。經由以上方法，將可以找到較接近的頻譜區段來做高頻頻譜補償。

由 $p(j)$ 可知，每次的補償區塊是由小漸大，主要是因為人的聽覺對高頻的聲音敏感度較低的緣故，而此一部份則是參考量尺因子頻帶的大小而決定。當第一次的補償完畢，需將補償的區塊乘上增益後複製於 \hat{x}_r 的高頻部分，接著再由補償過後的 \hat{x}_r 與 x_r 做頻譜相似度的測量與及求得相似區塊的起始點及增益，如圖(4.3)的第二步驟所示，經由一連串的補償工作後，也就完成了編碼端的高頻補償部分。

最後將解碼端所需的資訊，增益 (G_j) 及補償區塊起始點 (s)，依照細群及聲道的順序，如圖(2.14)所示，依序加入修正式 MPEG-1 Layer III 位元流中的輔助資料段中，這樣一來解碼端將可以利用由輔助資料段所得到的資訊來實行高頻補償。

➤ B-2 解碼端的高頻補償

由圖(4.1)可以得知，輔助資料段位於主要資料段的後面，所以解碼端必須一次將所有的解碼資訊讀出（標頭、副資訊、量尺因子及主要資料段先解碼出來），最後才由輔助資料段將高頻補償的資訊依照細群及聲道的順序依次解碼出來（增益及補償區塊的起始位址），其流程如圖(4.4)所示。



圖(4.4) 高頻補的償解碼端流程

而高頻補償的方法如式子 (4-4) 所示，

$$\hat{x}r(idx+i) = G_j * \hat{x}r(idx-s+i) \quad , \quad \text{for } i=0 \text{ to } bd_j \quad (4-4)$$

其中， $\hat{x}r$ 為解碼端經過反量化後的 MDCT 頻譜值， s 為補償區塊的起始位址， G_j 為補償區塊之增益， j 為補償的次數，依不同位元率而有所不同， idx 及 bd_j 與式子 (4-1) 定義相同；經由一次又一次的高頻補償動作，如圖(4.3)的第二步驟，即可將因為量化及頻譜截斷後的高頻部分加以回復。

➤ B-3 高頻補償之修正

雖然照上述之方法可以達到高頻補償的效果，卻會造成輔助資料段的位元數過多（其中補償區塊的起始位址需 8 個位元來表示，補償區塊之增益需 16 個位元表示），每補償一次需要 24 個位元數，反而使得可以表示霍夫曼編碼的位元數減少（因為要符合位元率的需求，將輔助資料段的位元數考慮進去），最後需以較大的量化步階來量化頻譜，導致量化誤差增加，聲音品質降低。

為了改善上述之情況，希望將輔助資料段的位元數減少，留下較多的位元數在主要資料段的編碼部分，以減少量化誤差，依然可以達到高頻補償的效果。因此將當初所設的補償區塊的起始位址及其增益大小製成碼簿，分別以式子 (4-5) 及 (4-6) 所示。

$$s_idx[i] = 4 + i * 4 \quad , \quad \text{for } i=0 \text{ to } 63 \quad (4-5)$$

$$G_{j_idx}[i] = \frac{i}{16}, \quad \text{for } i = 0 \text{ to } 15 \quad (4-6)$$

經由此簡化結果，不但編碼端在搜尋補償區塊的起始位址可以減少，而且每補償一次的位元數由原先的 24 個位元數減為 10 個位元數（其中補償區塊的起始位址 (s_idx) 需 6 個位元來表示，補償區塊之增益 (G_{j_idx}) 需 4 個位元表示）。

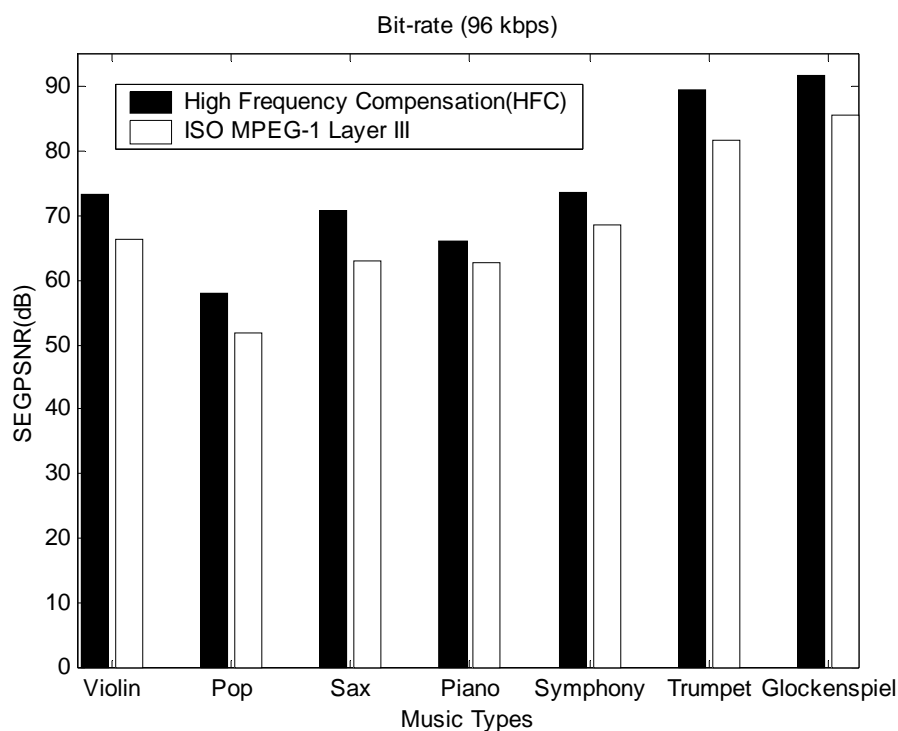
除此之外，在 3.6 節中所提到的適應性頻譜截斷法在此並不適用，主要原因是適應性頻譜截斷法的頻譜截斷點範圍過大，而補償的區塊大小亦需隨其調整，否則易造成補償上的誤差；因此適應性頻譜截斷法較適合改善 MPEG-1 Layer III 在低位元率的編碼效能，避免產生過多在聽覺上難以忍受的量化誤差。在此處的高頻補償法依然採用固定點的頻譜截斷法，而非適應性頻譜截斷法。

4.4 實驗結果

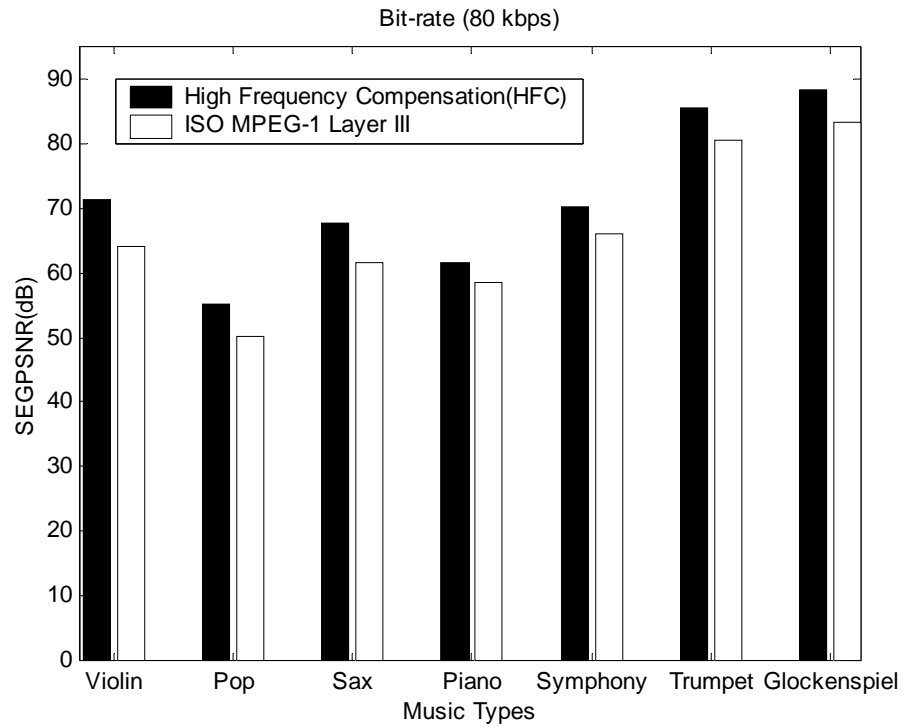
實驗結果如圖(4.5)、(4.6)以及(4.7)所示，這裡採用幾種不同種類的音樂型態來做比較（均為雙聲道聲音），有變化複雜的交響樂、輕柔的古典樂（小提琴）和輕音樂（鋼琴）、即興的爵士樂（薩克斯風）、銅管樂（小號）、流行音樂以及打擊樂（鐘琴）等幾個一般生活中常聽到的音樂或樂器來做評估；由國際標準組織所訂定的 MPEG-1 Layer III 音訊編碼器及加強式 MPEG-1 Layer III 音訊編碼器（包含了頻譜截斷法以及高頻補償

法) 來互相做比較, 採取的低位元率有以下三種: 96 kbps、80 kbps 以及 64 kbps。

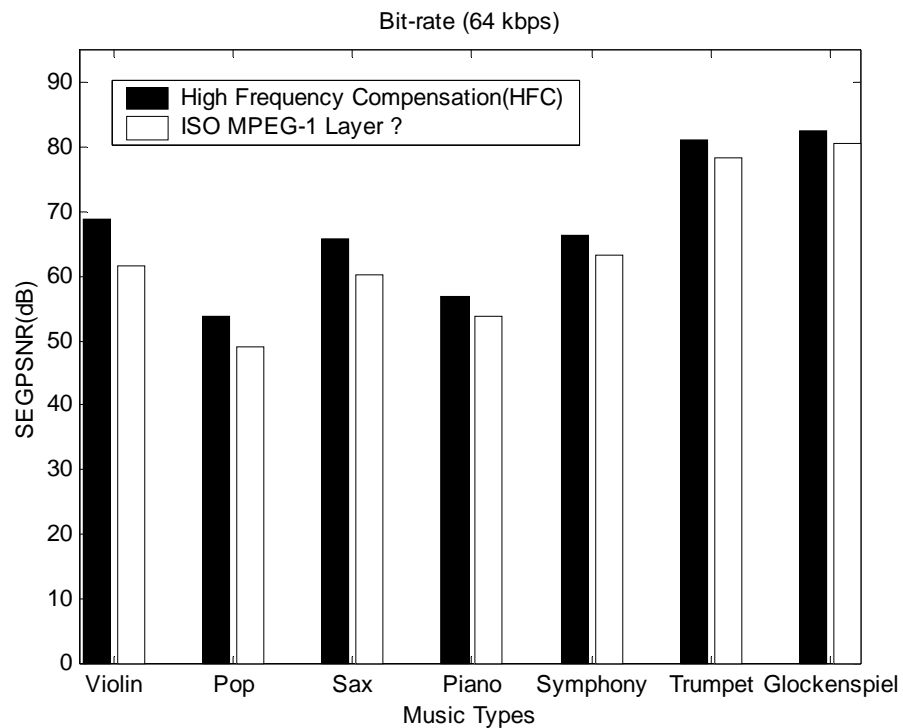
由以下各圖可以得知, 加強式 MPEG-1 Layer III 音訊編碼器在低位元率的編碼效果都比原先的 MPEG-1 Layer III 音訊編碼器要好; 除了利用頻譜截斷法外, 還有為了要彌補頻譜截斷法的缺點-高頻缺失, 利用頻譜相似度來做高頻補償, 使得加強式 MPEG-1 Layer III 音訊編碼器在低位元率時的表現相當不錯, 充分改善原先的效能。



圖(4.5) 位元率 96 kbps, MPEG-1 Layer III 音訊編碼器及加強式 MPEG-1 Layer III 音訊編碼器的比較



圖(4.6) 位元率 80 kbps , MPEG-1 Layer III 音訊編碼器及加強式 MPEG-1 Layer III 音訊編碼器的比較



圖(4.7) 位元率 64 kbps , MPEG-1 Layer III 音訊編碼器及加強式 MPEG-1 Layer III 音訊編碼器的比較

第五章 結 論

MPEG-1 Layer III 音訊編碼器是目前網路上或市面上最普遍熱門的音訊壓縮編碼器，為了要維持高品質的音訊壓縮，其位元率必須高於 128 kbps，所編碼出來的聲音也較讓人們（使用者）所能接受；一旦將位元率降低，將導致過多在聽覺上的雜訊誤差，反而難以忍受；相對地，其應用價值也隨之下降。

為了改善 MPEG-1 Layer III 音訊編碼器在低位元率時的缺點，在本論文中提出了兩個解決方法，一個是先將訊號通過低通濾波器或將訊號的頻譜截斷來加以改善其量化誤差，但是唯一的缺點就是高頻的聲音被消弭了；為了要加強頻譜截斷法，在本論文中另提出高頻補償法以改善高頻聲音缺失的缺點；經由以上兩個方法，可以充分改善 MPEG-1 Layer III 音訊編碼器在低位元率時的效能，不但降低了聽覺上的雜訊誤差，也將失去高頻的聲音補償回來，使得聲音較接近原始的聲音，也較讓人們所接受。除此之外，經過改善後的加強式 MPEG-1 Layer III 音訊編碼器以低位元率編碼也較適合在網路上傳輸，同時儲存的聲音資料量也隨之增加，這些優點都提升 MP3 音訊壓縮編碼器的目的。

參考文獻

- [1] E. Zwicker and H. Fastl, “Psychoacoustics - Facts and Models,” Springer Press, 1999.
- [2] ISO/IEC JTC1/SC29/WG11 MPEG, “Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, Part 3: Audio,” IS11172-3, 1993.
- [3] S. Shlien, “Guide to MPEG-1 Audio Standard,” IEEE Transactions on Broadcasting, Vol. 40, No. 4, December 1994.
- [4] D. Pan, “A Tutorial on MPEG/Audio Compression,” IEEE Trans. on Multimedia, Vol.2, No. 2, pp. 60-74, 1995.
- [5] P. Noll, “MPEG Digital Audio Coding,” IEEE Signal Processing Magazine, September 1997.
- [6] T. Painter and A. Spanias, “Perceptual Coding of Digital Audio,” Processing of the IEEE, Vol. 88, No. 4, April 2000.
- [7] C.A. Lanciani, “Audio Perception and the MPEG Audio Standard – A Qualifying Examination Report,” Georgia Institute of Technology School of Electrical and Computer Engineering, August 11, 1995.
- [8] J. Princen and A. Bradley, “Analysis/Synthesis Filterbank Design Based on Time Domain Aliasing Cancellation,” IEEE Trans. on Acoust. Speech, and Signal Process. Vol. ASSP-34, pp. 1153-1161, 1986.
- [9] J. Princen, A. Johnson, and A. Bradley, “Subband/Transform Coding Technique Based on Time Domain Aliasing Cancellation,” Proc. Int’l Conf. IEEE ASSP, IEEE Press, Piscatawan, N.J., pp. 2161-2164, 1987.
- [10] J. Watkinson, “The MPEG Handbook,” Focal Press, 2001.

- [11] Y.H Lam and R.W. Stewart, "Digital Filtering for Audio Coding," An Enabling Technology (Ref. No. 1998/252), IEE Colloquium on, 1998.
- [12] D.Y Pan, "Digital Audio Compression," Digital Technical Journal Vol. 5 No. 2, Spring, 1993.
- [13] F. Wylie, "Digital audio data compression," Electronics & Communication Engineering Journal, February 1995.

作者簡歷

作者：黃國祥

生日：民國 66 年 10 月 30 日

籍貫：江西省贛縣

學歷：省立台中一中 (民國 85 年 6 月畢業)

國立成功大學工程科學系 (民國 89 年 6 月畢業)

國立成功大學電機工程研究所碩士班 (民國 91 年 6 月畢業)

Vita

Kuo-Hsiang Huang was born in Taichung, Taiwan, R.O.C., on October 30, 1977. He received the B.S. in engineering science and M.S. degrees in electrical engineering from the National Cheng Kung University, Taiwan, in 2000 and 2002, respectively. His research interests include digital signal processing and audio coding algorithms.