國立交通大學

資訊科學與工程研究所

碩 士 論 文

MPEG-4 HE-AAC 中耦合編碼之設計

Design of Coupling Coding in MPEG-4 HE-AAC

研 究 生：張家銘

指導教授：劉啟民　教授

李文傑　博士

中 華 民 國　九 十 五　年 六 月

MPEG-4 HE-AAC 中耦合編碼之設計

Design of Coupling Coding in MPEG-4 HE-AAC

研 究 生：張家銘　　Student：Chia-Ming Chang

指導教授：劉啟民　　Advisor：Dr. Chi-Min Liu

李文傑　　Dr. Wen-Chieh Lee

國 立 交 通 大 學

資 訊 科 學 與 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

June 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年六月

MPEG-4 HE-AAC 中耦合編碼之設計

學生：張家銘　　指導教授：劉啓民 博士
李文傑 博士

國 立 交 通 大 學

資 訊 科 學 與 工 程 研 究 所

中文論文摘要

SBR 的耦合編碼機制乃藉由轉換資料的紀錄方式，以去除資料的相關性，達到節省位元的目的。然而，礙於耦合編碼的先天限制，若干原先獨立的資訊必須被雙聲道共用。由於資訊的共用，產生兩項關係到品質的設計議題，其中包含時頻格子與啾聲參數。再者，量化處理所造成的品質衰退的風險，也必須進一步的審視。本論文考慮可能產生的人工缺陷，來考量共用參數的決定，並提出一個兼顧高頻品質及需求位元數的耦合轉換方法。在品質的量測上，除了主觀的聆聽評量之外，並採用 ITU 所發展的 PEAQ 測試系統來評估音訊壓縮後的改良程度。

Design of Coupling Coding in MPEG-4 HE-AAC

Student: Chia-Ming Chang Advisor: Dr. Chi-Min Liu

Dr. Wen-Chieh Lee

Computer Science

National Chiao Tung University

ABSTRACT

The coupling coding in SBR is adapted to transform the data domain to de-correlation and save more bits. However, because of the inherent constraint of the coupling coding, some side information need to be shared by the stereo channels. There are two considerable issues related to quality due to the sharing, including the determining of the shared T/F grid and the shared chirp factor. On the other hand, the quantization process causes of the risk of quality degradation also need to be inspected. This thesis considers the possible artifacts to examine the decision of the shared parameter, and proposes a coupling decision method for the tradeoff between high band quality and demand bits. Both subjective and objective tests are conducted to check the quality improvement. The objective test measures used is the recommendation system by ITU-R Task Group 10/4.

致謝

感謝劉啓民老師一直以來的栽培及李文傑博士給予的指導，實驗室的楊宗翰學長、許瀚文學長，同學唐守宏、李侃峻和楊詠成，以及學弟曾信耀和胡正倫的協助，在研究上提供我寶貴的意見，讓我在專業知識及研究方法獲得非常多的啓發。

最後，感謝我的父母與家人及系上同學，在我研究所兩年的生活中，給予我無論在精神上以及物質上的種種協助，使我能全心全意地在這個專業的領域中研究探索在此一併表達個人的感謝。

Contents

Figure List

Table List

Chapter 1 Introduction

MPEG-4 HE-AAC (High Efficiency Advanced Audio Coding) is the extension of the conventional AAC [1] by supporting the SBR (Spectral Band Replication) [2][3][4][5]. The basic principle of SBR is to reconstruct the high frequency spectral bands by replicating the low frequency spectral bands. The resulting codec is referred to as the MPEG-4 HE-AAC or AACplus. Besides taking the SBR as the bandwidth extension tool, the PS (parametric stereo) [6][7][8] coding is further incorporated as the channel reduction tool. The integrated codec is referred to MPEG-4 HE-AAC version 2. Figure 1 illustrates the scheme of HE-AAC and HE-AAC version 2.

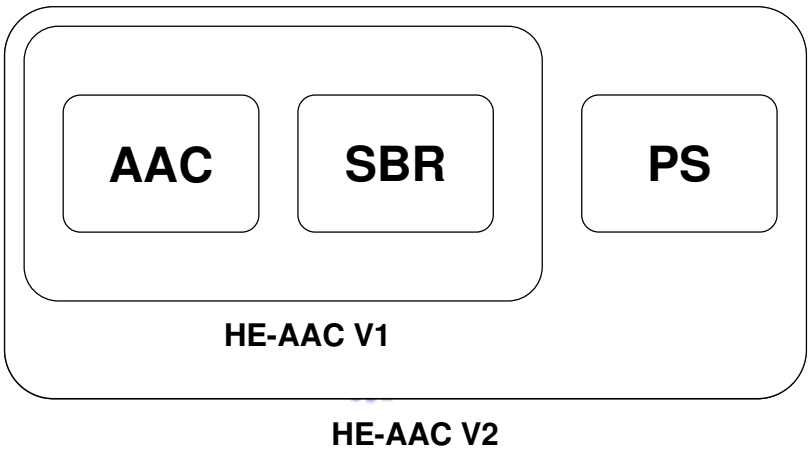


Figure 1: Diagram of the HE-AAC and HE-AAC v.2 [9]

To fit a variety of situations of demand, the three coding schemes are applied to the different bit rates. In order to maintain the audio quality at low bit rate, HE-AAC is adapted among the 48 ~ 96 kbps. Furthermore, to satisfy the requirement of very low bit rate lower than 48 kbps, HE-AAC version 2 is proposed to overcome the challenge. Figure 2 illustrates the relationship between the bit rate and the perceptual quality. However, the efficiency of the complete system is determined largely by the cooperation of the three modules. Any unsuitable design of anyone of the three modules will affect the effect of the remainders, and hence destroy seriously the quality of the HE-AAC version 2.

This thesis will focus on the coupling coding in SBR. The principle of the coupling coding is to transform the left/right (L/R) energy signals into average/ratio (A/R) mode to eliminate signal correlation and take advantage of parameter sharing to

save bits. Especially, at very low bit rate, the bit shortage will result in many annoying artifacts at the low frequency in AAC. Hence, the purpose of coupling coding is to save the consuming bits of the high frequency to promote the low frequency quality. Figure 3 shows an example of a common artifact, known as “birdies effect” [10], at the low frequency part due to lack of bits in AAC. The spectral valley is visible in the low frequency spectrum encoded by AAC. Figure 4 shows the artifact is enhanced largely once AAC is supplied with enough bits.

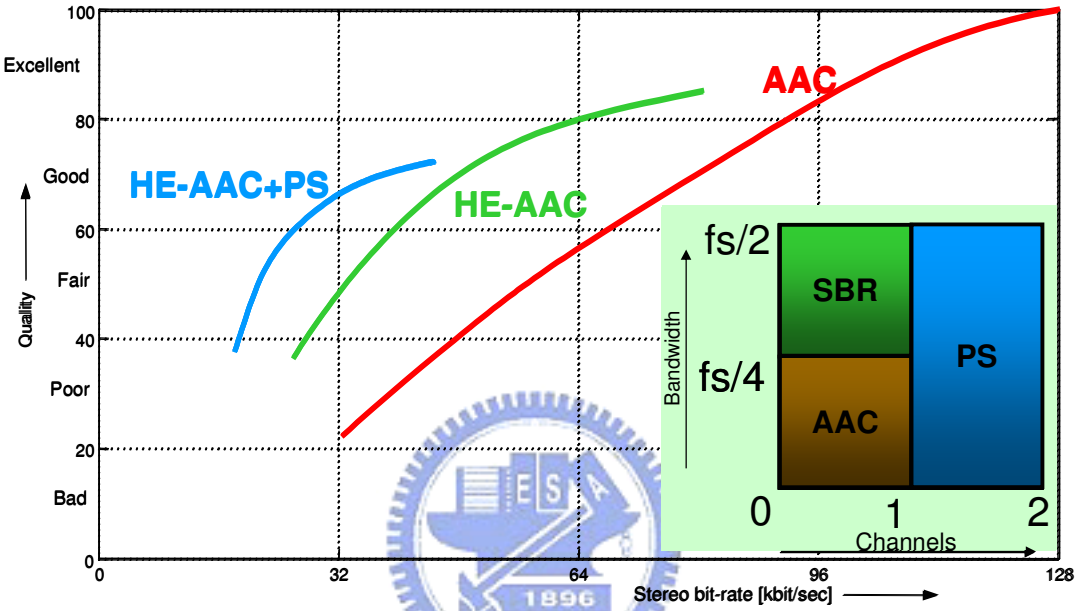


Figure 2: Quality comparison at different bit-rate among AAC, HE-AAC, and HE-AAC v.2 [6]

Figure 5 illustrates the block diagram of HE-AAC encoder. This thesis considers the coupling coding design through four design issues. The first and second issues are decision of shared T/F grid [9] parameter and shared inverse filtering intensity [11][12]. Furthermore, according to the constraint in the coupling mode, the difference values of quantized value of the specific parameters, named as noise floor scale factors [11][12], between two channels should be limited. Therefore, selecting suitable quantized noise floor scale factors for the two channels is the third issue. Finally, the last issue is coupling switch method which need to consider the tradeoff between high band quality and demand bits, and some possible risk of quality degradation.

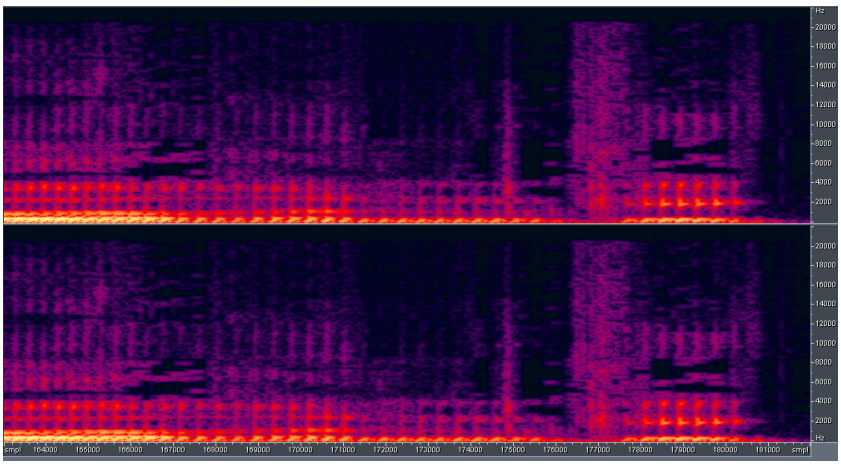


Figure 3: Birdies effect occurring in LF of HE-AAC due to the insufficient bits

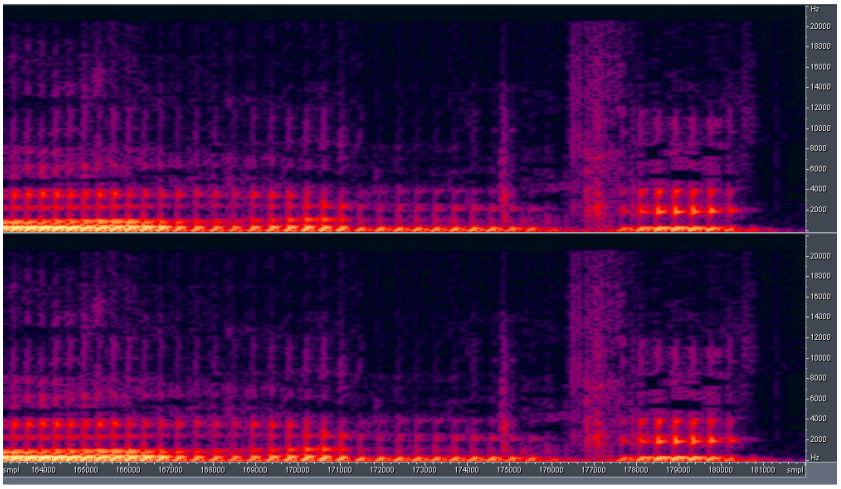


Figure 4: Enhanced Birdies effect by taking advantage of coupling coding

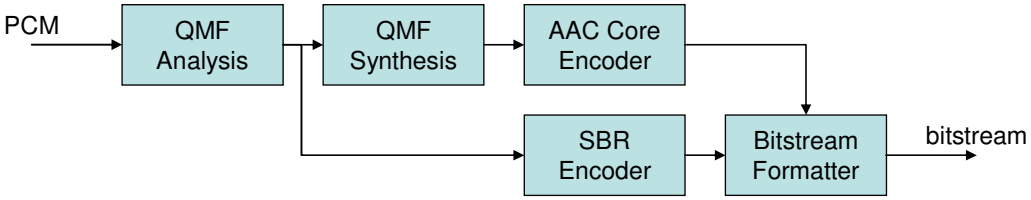


Figure 5: Basic architecture of HE-AAC encoder

This thesis is organized as follow. Chapter 2 introduces the backgrounds on the fundamental knowledge of MPEG-4 HE-AAC. In Chapter 3, the five design issues of the coupling coding in MPEG-4 HE-AAC will be investigated, which include coupling switch method, decision of shared T/F grid parameter, decision of shared inverse filtering intensity, decision of the noise floor scalefactors. Chapter 4 conducts

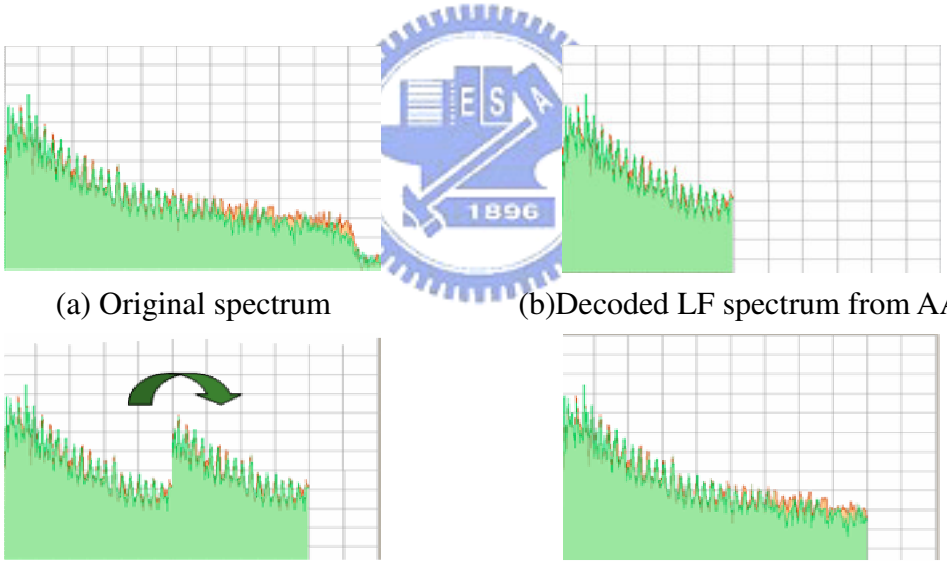
experiment to verify performance of the proposed coupling coding method. Chapter 5 gives a conclusion on this thesis.

Chapter 2 Backgrounds

This chapter introduces some fundamental background of MPEG-4 HE-AAC. Especially, the three modules related to the coupling coding in SBR are also described.

2.1 MPEG-4 High Efficiency AAC

By the cooperation of both SBR and the conventional AAC, the HE-AAC can maintain the high audio quality at very low bit rate. The SBR takes care of the high frequency contents, and relatively the AAC encoder compresses the low frequency contents. Because of the few bits consuming of SBR, the most of available bits are supplied for AAC to maintain the quality of low frequency. Figure 6 illustrates the reconstruction process of SBR.



(a) Original spectrum (b)Decoded LF spectrum from AAC

(c)HF regeneration by replicating LF bands (d)Reconstructed HF after envelope adjustment

Figure 6: An example of reconstruction process of SBR

The HE-AAC decoder reconstructs the high frequency by replicating the low frequency decoded from AAC, and then it will adjust the tonality of the replicated low band to be close to the tonality of the original high band by comparing the difference between the content of the high and low bands. Furthermore, there are two main modules in SBR encoder. One is the time/frequency grid module, and another is the high frequency adjustment module. In the time/frequency grid module, it splits the high bands into several T/F grids. Each T/F grid records the average energy to the

HE-AAC decoder. On the other hand, the HF adjustment module records the difference between the original and replicated contents of high frequency part. Besides the basic reconstruction operation, the data-rate reduction tool, Coupling coding, is adopted to further eliminate the signal correlation and save consuming bits of high frequency by domain transformation of envelope data and parameter sharing. The saved bits by the coupling coding mechanism can supply to AAC to promote low frequency quality effectively, and achieve the optimal overall quality.

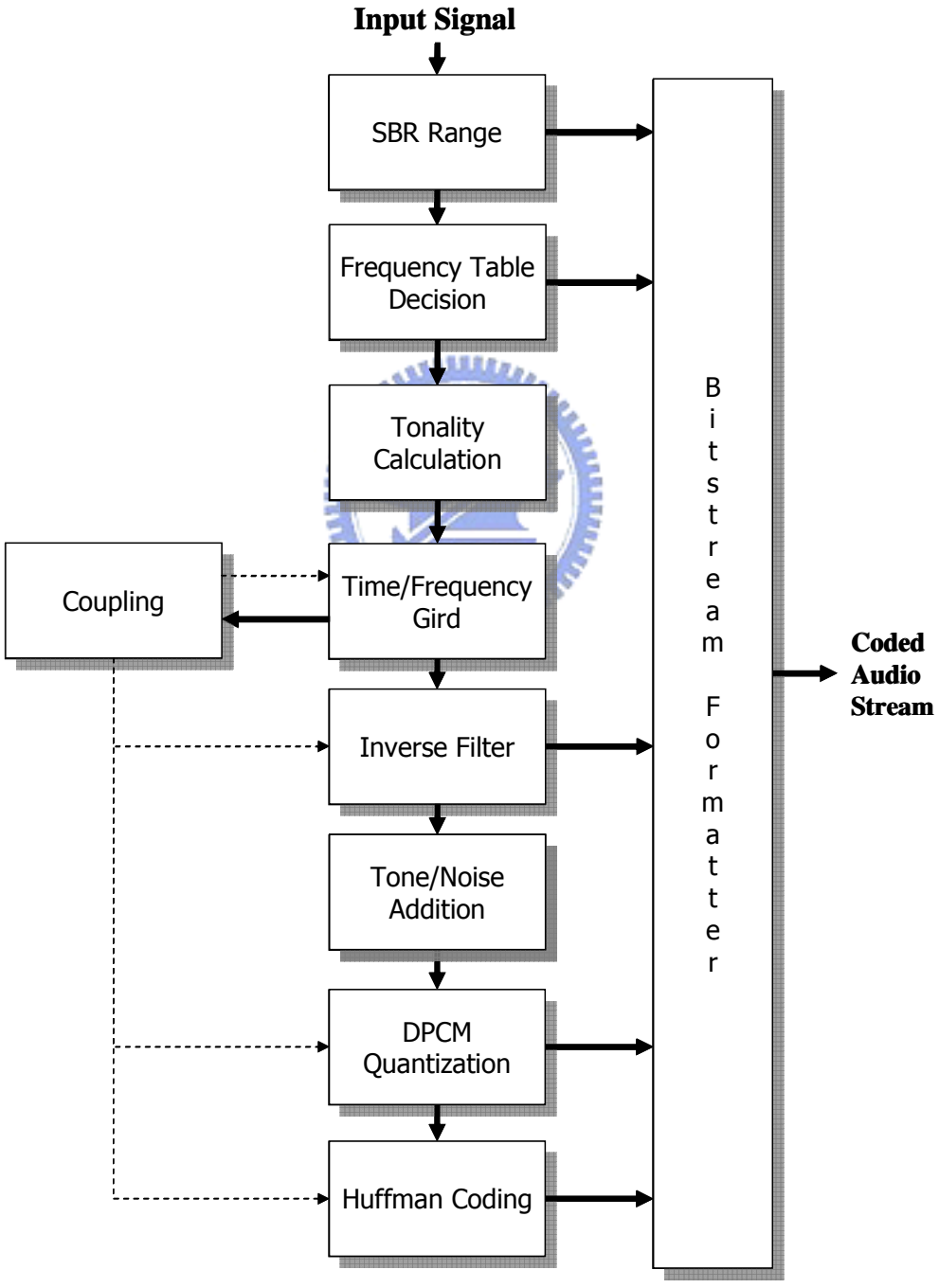


Figure 7: Block diagram of HE-AAC encoder

2.2 Related Modules in SBR to Coupling Coding

In this section, three important modules related to coupling coding will be introduced. They include the time/frequency grid, the chirp factor used in the inverse filtering operation, and the tonality controlling factor Q. The understanding of the three issues will affect the design of the coupling coding largely.

2.2.1 Time/Frequency Grid in HE-AAC

This subsection introduces the protocol about the time/frequency grid in HE-AAC. Adaptive time and frequency resolution are incorporated into SBR for the envelope coding and adjustment. The SBR replicates the low frequency signal to high frequency signal. The QMF subbands in SBR range will be segmented into several grids by divisions from the T/F dimension. The successive samples in QMF subbands are integrated into a “time envelope”. The successive subbands are segmented into several uniform or non-uniform bands with different bandwidth by choosing one of the frequency band tables. T/F grid which describes time segments and associated frequency tables is the basic reconstructed unit in the subsequent SBR coding process. The average energy in a grid will be used to get the rescale ratio, which is over the energy of the duplicated low bands. It implies that the locations of time borders and the resolution of the envelopes determine the accuracy of the replication and the audio quality. There are three components in the time/frequency grid module as follows

- Frequency table
- Time segment
- Frame class

The frequency domain resolution is determining by choosing from the different frequency tables in the SBR. Frequency table affect the precision of tone addition and the low frequency which will be replicated for reconstruction high frequency in the decoder. There are five frequency tables in SBR. High resolution frequency band table and low resolution frequency band table are two available resolution tables that can be selected for every envelope of SBR frame. Noise floor frequency band table and limiter frequency band table correspond respectively to the noise floor and the limiter. All frequency band tables are derived from the master frequency band table. The master frequency tables are defined by functions and the arguments are transmitted in the SBR header. The time borders affect the resolution in the time domain. The time

borders are more flexible than frequency band table. It contains the number of envelopes in SBR frame and locations of time borders. There are 32 samples in the time domain of one SBR frame. But there are only 16 locations in the SBR frame for the time borders. Because of the constraint in SBR, there are 5 time borders if frame class is VARVAR. The other frame class only has 4 time borders at most.

There are four different SBR frame classes, FIXFIX, FIXVAR, VARFIX, and VARVAR in the SBR time/frequency grid. The four classes refer to whether locations of leading and trailing SBR frame boundaries are variable. HE-AAC allows the boundary between the two frames to shift few sample. It make time domain segment more flexible. Figure 9 is an example of the boundary shift few sample to match the signal.

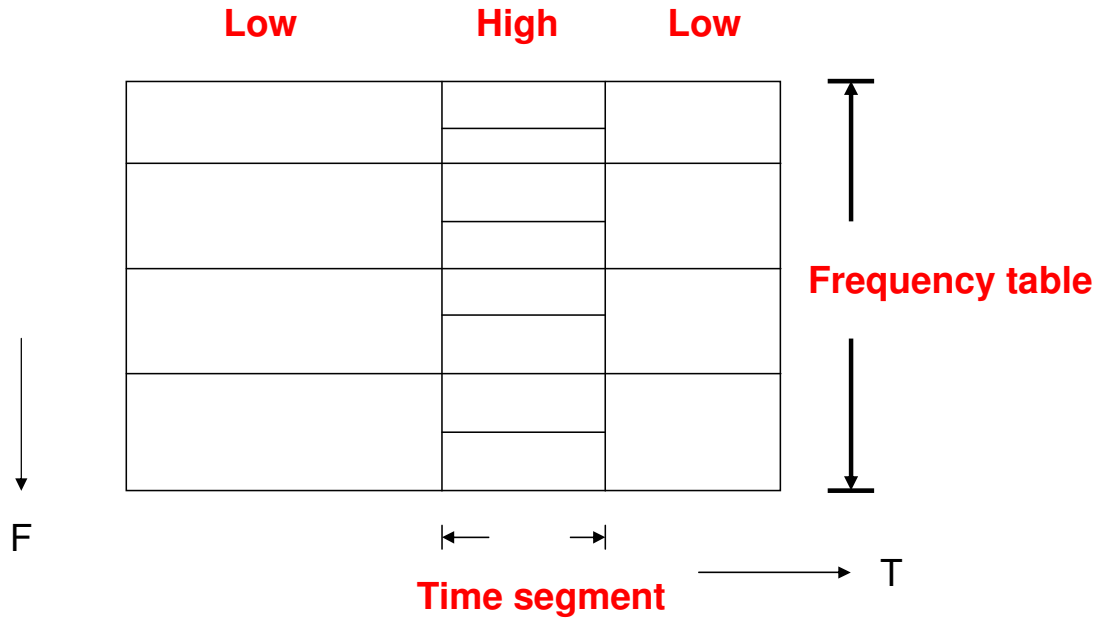


Figure 8: An example of the frequency table and time segment

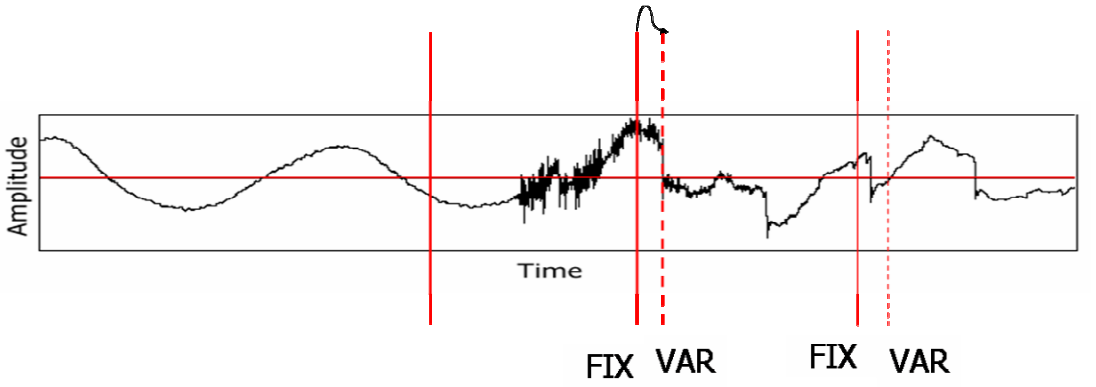


Figure 9: An example of VAR frame border to shift few sample points

2.2.2 Chirp Factor of Inverse Filtering

The high bands in SBR decoder are reconstructed from the low bands. Hence, if the tonality of the replicated low bands will not match the features of the original high bands, the inverse filtering is applied to eliminate the excess tone component of the replicated low bands. The inverse filtering process is performed in the two steps. First, the linear prediction is applied on the replicated low bands. Then the actual inverse filtering is performed respectively for each of the replicated low bands patched to the high bands. The resultant high band generated in SBR decoder is obtained as

$$X_{High}(k,l) = X_{Low}(p,l) - \{\alpha \cdot a_0 \cdot X_{Low}(p,l-1) + \alpha^2 \cdot a_1 \cdot X_{Low}(p,l-2)\}, \quad (1)$$

where α is the chirp factor that can control the inverse filtering level, X_{Low} is the low band signal analyzed from the output of AAC decoder, and a_0, a_1 are the prediction coefficients which are used to filter the subband signal in the inverse filtering .The chirp factor can remove the location of the poles of the inverse filter and affect the degree of the harmonic attenuation in the high frequency generator. According to the standard [2] , the calculation of the chirp factor is defined as

$$\alpha(i) = \begin{cases} 0 & if\ tempBw(i) < 0.015625 \\ tempBw(i) & if\ tempBw(i) \geq 0.015625 \end{cases}, 0 \leq \mathrm{i} < \mathrm{N_Q}, \quad (2)$$

where N_Q is number of noise floor bands, and $tempBw(i)$ is calculated as

$$tempBw(i) = \begin{cases} 0.75 \cdot newBw + 0.25 \cdot \alpha'(i) & if\ \mathrm{newBw} < \alpha'(i) \\ 0.90625 \cdot newBw + 0.09375 \cdot \alpha'(i) & if\ \mathrm{newBw} \geq \alpha'(i) \end{cases}, 0 \leq i < N_Q, \quad (3)$$

where α' is the α values calculated in the previous SBR frame, and newBw is decided by inverse filtering mode of current frame and previous frame according to Table 1.

Table 1: The parameter *newBw* decided by inverse filtering mode

bs_invf_mode(i)´	**bs_invf_mode**(i)			
	Off	Low	Intermediate	Strong
Off	0.0	0.6	0.9	0.98
Low	0.6	0.75	0.9	0.98
Intermediate	0.0	0.75	0.9	0.98
Strong	0.0	0.75	0.9	0.98

2.2.3 Noise Floor Scale Factor *Q*

The noise floor scale factor is the parameter which affects the magnitude

adjustment of the each subband in the high resolution grid and the tone-noise additive level. There are two different types of the envelope scaling for the noise-demand and the tone-demand grid according to their requirement of content. In the noise-demand grid, the magnitudes of the replicated bands are adjusted by a gain control factor defined as [11]

$$G_k^{ND} = \sqrt{\frac{E_k^o}{E_k^r} \cdot \frac{1}{1+Q_k}}, \tag{4}$$

where E_k^o and E_k^r are the average energy of the original high frequency signal and the replicated low frequency signal respectively in the k^{th} high resolution grid, and Q_k is the noise floor scale factor. After the envelope scaling, the random noise is added to the high frequency with level defined as

$$C_k^n = E_k^o \cdot \frac{Q_k}{1+Q_k}. \tag{5}$$

Similarly, the gain control factor in the tone-demand grid is defined as

$$G_k^{TD} = \sqrt{\frac{E_k^o}{E_k^r} \cdot \frac{Q_k}{1+Q_k}}. \tag{6}$$

Also, the energy amount for the compensated tone is defined as

$$C_k^t = E_k^o \cdot \frac{1}{1+Q_k}. \tag{7}$$

Figure 10 and Figure 11 illustrate the two different adjustment results for the noise- and tone- demand grids.

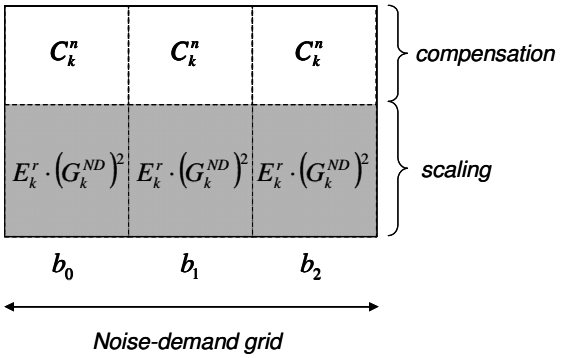


Figure 10: HF adjustment process of a noise-demand high resolution grid with three subbands b_0, b_1, b_2

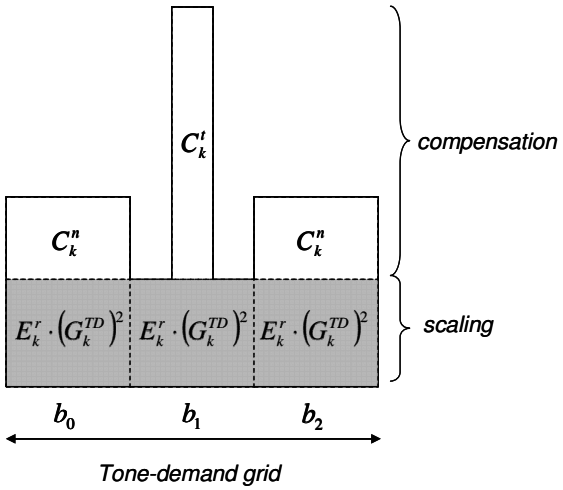


Figure 11: HF adjustment process of a noise-demand high resolution grid with three subbands b_0, b_1, b_2

Chapter 3 Design of Coupling Coding in SBR

As the fundamentals of the coupling coding design, this chapter reviews firstly the coupling coding schemes defined in the HE-AAC standard [2]. Furthermore, based on our related work [9][11] about the designs of the T/F grid and tonality compensation, the thesis extends the works into the coupling coding method. The extension should consider the risk of parameter sharing. For example, the T/F grid sharing will result in the coarse reconstructed envelope or the more bits consuming. Also, the sharing of the chirp factor and noise-floor factor may destroy the tonality accuracy of some channel and lead to the quality degradation. Finally, the huge ratio of the left and right energy channels will result in the data correlation reduction and the large quantization error. A coupling switching method considering the abnormal phenomenon is proposed to compromise the tradeoff between the quality and the bits reduction.

3.1 Overview of Coupling Coding Schemes in HE-AAC

For the energy data of the spectral envelop extracted from time/frequency (T/F) grids, there are several processes, including quantization operation, the DPCM and the Huffman entropy coding, to be applied to reduce the data rate in turn. Furthermore, to reduce the data redundancy of the stereo energy channels, the coupling mode is adapted to transform the Left/Right (L/R) energy channels into Average/Ratio (A/R) mode to eliminate signal correlation. To meet the inherent requirement of the coupling, the parameters for T/F grid and inverse filtering need to be shared by A/R channels in coupling mode. Figure 12 illustrates the syntax of the SBR extension data elements in the two modes. It shows that both the T/F grid and the controlling parameter of inverse filtering level, that is chirp factor, should be shared.

In summary, there are five critical points when the coupling mode is switched on. The five terms are as follows:

- Transform L/R mode into A/R mode
- Share Time/Frequency grid
- Share chirp factor of inverse filtering
- DPCM values of the ratio channel is further quantized with step size 2.

- The difference of quantized noise-floor scalefactors between the two channels is restricting to the range from 0 to 12 by the syntax constraint.

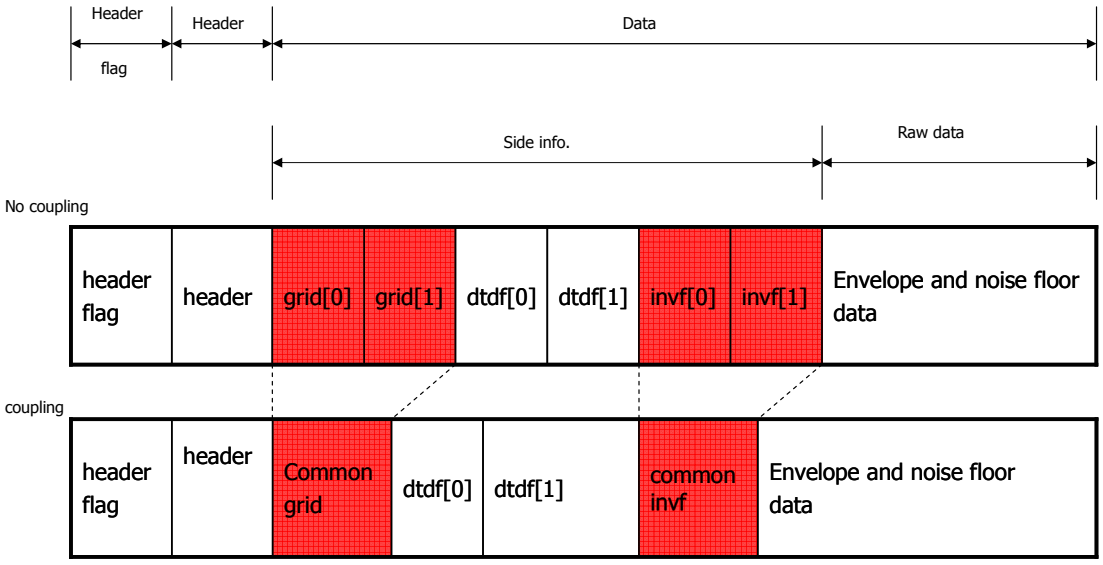


Figure 12: The syntax of the SBR extension data elements in coupling and non-coupling modes

3.2 Decision of Shared T/F Grid

Instead of using the individual set of time segments as the normal L/R mode, there can be only a common segment set in the coupling mode. Although the sharing of side information can save bits, the quality artifact may occur due to inaccurate segment. Hence, the quality degradation should be considered. For the optimal time segments of the signal subbands of the L/R channels in the normal mode, a decision method based on dynamic programming approach has been proposed in our other work [9]. In this section, the modified decision method for coupling mode is proposed to determine the optimal common segment set and measure the affect for quality.

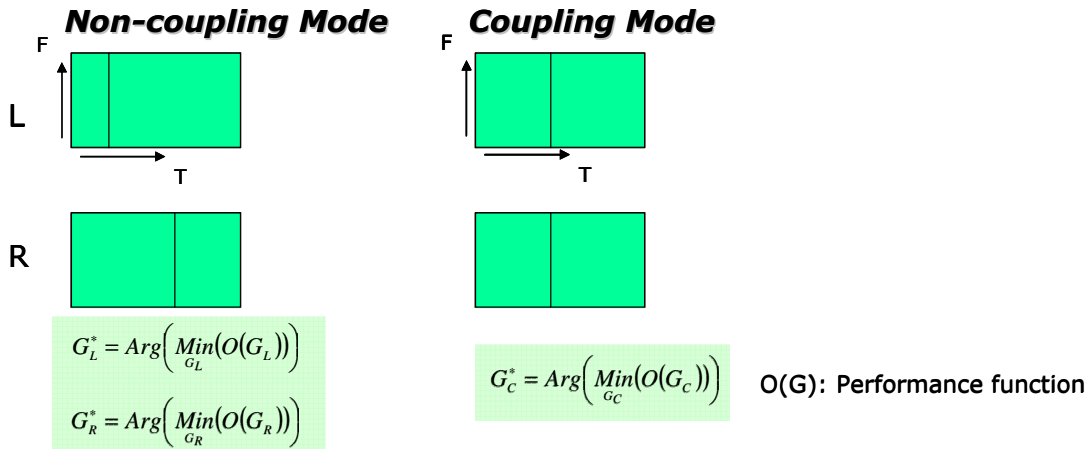


Figure 13: Diagram of shared time segment in coupling and non-coupling modes

3.2.1 Design of T/F Grid by Dynamic Programming in Non-coupling Mode

In [9], a decision method of T/F grid by the dynamic programming (DP) in non-coupling mode has been proposed. The basic concept of the DP method is to search the optimal grid in the all possible grids in individual channel by an efficient recursive procedure. The resultant grid G^* searched by the method will be an optimal solution to make the average of the energy difference (reconstructed energy error) to the original signal energy ratios (DSR) in all quality measurement units, named critical units, minimum. That is,

$$G^* = Arg\left(\underset{G}{Min}\left(\overline{DSR}(G)\right)\right), \tag{8}$$

where $\overline{DSR}(G)$ is defined as

$$\overline{DSR}(G) = \frac{\sum_{c \in G} DSR_c}{\#(c)}, \tag{9}$$

where c is the critical unit, $\#(c)$ means the number of critical units, and DSR_c is the reconstruction error of the critical unit c in the frame. The lengths of the critical unit are defined as four sample points and the critical band bandwidth for time and frequency direction respectively in [9].

The number of the time borders and the associated frequency resolution determine the total number of the girds and also affects the resultant DSR. The dynamic programming for DSR analysis is shown as

$$DSR_{2i,2j}^{k,u} = \underset{i+1\le t\le j-1}{Min}\left\{DSR_{2i,2t}^{0,0} + DSR_{2t,2j}^{k-1,u}, DSR_{2i,2t}^{0,1} + DSR_{2t,2j}^{k-1,u-1}\right\}$$
$$0 \le i < j \le 8;$$
$$0 \le k \le 4, \qquad (10)$$
$$0 \le u \le k+1$$

where i, j are the border of the time slot consisting of two samples, k is the number of the time borders, and u is the number of the high resolution envelopes. The notation $D_{i,j}^{k,u}$ means the optimal DSR from i to j with k time borders and u high resolution envelope. According to [9], there are ten different bit-consuming stages defined in the dynamic programming method. Each stage indicates the different number of time borders and high resolution envelopes in one SBR frame. The scenarios of the ten stages are described in Table 2. Figure 14 illustrates the optimal partition from i to j with 3 time borders and 2 high resolution envelopes.

Table 2: The scenarios of the ten bit-consuming stage

Bit-Overhead Stage	k (number of time borders)	u (number high resolution envelopes)
0	0	0
1	0	1
	1	0
2	2	0
	1	1
3	3	0
	2	1
	1	2
4	4	0
	3	1
	2	2
5	4	1
	3	2
	2	3
6	4	2
	3	3
7	4	3
	3	4
8	4	4
9	4	5

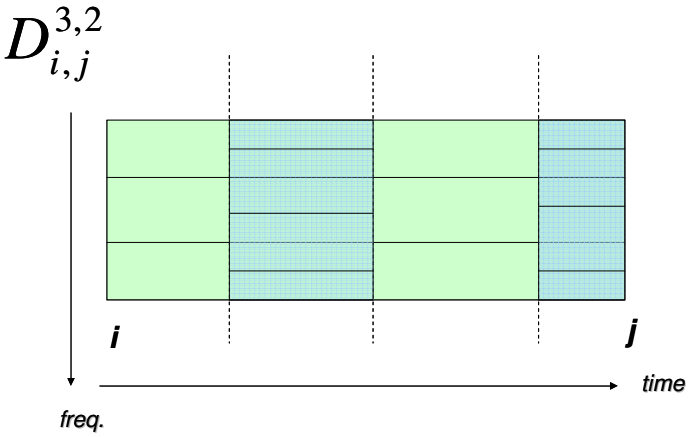


Figure 14: An example of the optimal partition from *i* to *j* time unit with 3 time borders and 2 high resolution envelopes

Figure 15 is the flowchart of the dynamic programming method for searching optimal T/F grid. The loop will consider all passable resolution grids. In the loop, it will have an objective function for determining the optimal T/F grid in the same bit-consuming stage. There is another efficiency checking for switching different stages. The dynamic programming method searches the optimal grid from the lower bit-consuming stages to the higher bit-consuming stages. Because the different bit-consuming stages have different requirement of bits, the grid decision in the different stages must consider the tradeoff of bits and quality.

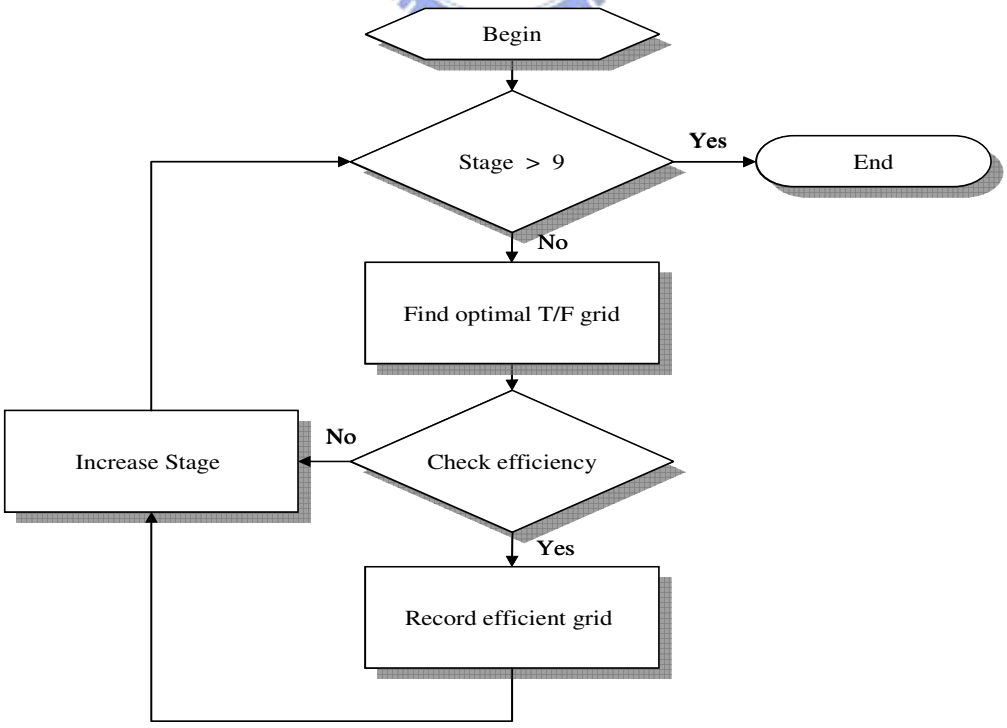


Figure 15: Flowchart of the DP method proposed in [9]

3.2.2 Design of T/F Grid by Dynamic Programming in Coupling Mode

Because of the sharing of the T/F grid, the two criterions used in the above DP method must be modified to simultaneously consider the content of two channels in the coupling mode. There is an objective function which measures the grids in the DP search method. In the normal mode, the objective function is defined as the DSR value described above. To consider both the two DSR values from L/R channels in the coupling mode, the objective function is modified as

$$O(G) = Max\left(DSR_0, DSR_1\right), \tag{11}$$

where DSR_0 and DSR_1 are the DSR values of left and right channel respectively. To ensure the quality of worst channel, the conservative choice of the resultant grid is adopted in the criterion. The optimal grid is the minimum solution of the objective function (11).

On the other hand, the iteration criterion of the DP method in the normal mode involves the improvement of DSR in the current resolution. If the improvement of DSR is over the threshold depending on the bit rate, it will update the higher resolution T/F grid to improve the quality. The improvement is defined as

$$E = DSR' - DSR, \tag{12}$$

where DSR' is the optimal DSR for the preceding bit-consuming stage. Similarly, in the coupling mode, there are two improvements of DSR which are defined as

$$\begin{aligned} E_0 &= DSR_0' - DSR_0 \\ E_1 &= DSR_1' - DSR_1 \end{aligned}. \tag{13}$$

The modified iteration criterion is to satisfied the two conditions as follows

$$E_{\max} = \max\left(E_0, E_1\right) > \Phi_1, \tag{14}$$

$$E_{\min} = \min\left(E_0, E_1\right) > \Phi_2, \tag{15}$$

where Φ_1 and Φ_2 are the threshold of iteration criterion. The modification of the iteration criterion can ensure that the improvement of both the DSR values of the two channels are over a low bound, and at least one of the DSR improvements can exceed the large degree to show the efficiency of the new stage.

Table 3: Comparison of the grid criterion in the normal/coupling mode

	Normal mode	Coupling mode
Performance function	DSR	Max(DSR_0 ,DSR_1)
Optimal candidate	Min(O(G))	Min(O(G))
DSR improvement	E = DSR' – DSR	$E_0 = DSR'_0 - DSR_0$ $E_1 = DSR'_1 - DSR_1$
Iteration criterion	E > Φ	Max (E_0,E_1) > Φ_0 and Max (E_0,E_1) > Φ_1

O(G): Performance function

Φ:Threshold

DSR': Optimal DSR for preceding resolution

3.3 Decision of Shared Inverse Filtering Intensity

In SBR, the inverse filtering is adopted to eliminate excess tones in low bands to fit the tonality of high bands. The different chirp factors, which are the parameter to control the intensity level of the inverse filtering, are assigned to L/R channels in the normal mode. Similarly, according to the regulation of the standard [2], only single chirp factor can be used in the coupling mode. Once the difference of the tonality contents is dominate, the possible artifact suffering from the unsuitable inverse filtering under the constraint is able to be anticipated. Hence, the risk of the inverse filtering with the same intensity level for the stereo channels in the coupling mode should be considered.

3.3.1 Decision of Inverse Filtering Intensity in Non-coupling Mode

In [11][12] of our another work, a compensation method of additional tone and noise to maintain the dB-difference of the tone and noise component has been proposed. In the non-coupling mode, based on the method, the selecting method of inverse filtering level is depended on the tonality between the high bands and the low bands. According to the syntax, each specific "noise band" has an inverse filtering mode individually. Furthermore, a noise band may include several high-resolution grids. Therefore, the tonality of the noise band is defined in [11][12] as

$$\Delta_{NB} = \max\left\{\frac{T_i}{N_i} \mid i \in NB\right\}, \tag{16}$$

where T_i is the energy of tone in the i^{th} high resolution grid of the noise band, N_i is the energy of noise floor in the i^{th} high resolution grid of noise band, and NB is the noise bands in the SBR frame. From (16), the maximum tonality among the high resolution grids stands for the tonality of the noise band. The goal of the inverse filtering mode is to imply that the tonality of the replicated low frequency band can approach the tonality of the original high frequency band, that is

$$\hat{\Delta}^l_{NG} = \Delta^h_{NG}, \tag{17}$$

where $\hat{\Delta}^l_{NG}$ is the tonality of the replicated low frequency band, Δ^h_{NG} is the tonality of the original high frequency band. The optimal chirp factor can be evaluated from

$$\frac{(1-\alpha^2)T_l}{N_l} = \frac{T_h}{N_h}, \tag{18}$$

where α is chirp factor. Then the practice chirp factor can be searched in the Table 1 to approximate the optimal one.

3.3.2 Decision of Inverse Filtering Intensity in Coupling Mode

However, in the coupling mode, there is only one inverse filtering mode for two channels. To measure the influence to the resultant tonalities, the distortion function of the shared chirp factor is defined as

$$f(x) = \left|x\Delta^0_l - \Delta^0_h\right|^2 + \left|x\Delta^1_l - \Delta^1_h\right|^2, \tag{19}$$

where $x = 1-\alpha^2$, Δ^0_l, Δ^1_l are the tonality of low frequency in the left and right channel respectively, and Δ^0_h, Δ^1_h are the tonality of high frequency in the left and right channel respectively. Hence, the optimal chirp factor shared by the two channels in the coupling mode should be the minimum solution of the distortion function (19). The optimal solution can be calculated by solving the equation of the one order differential of the distortion function,

$$2\Delta^0_l\left(x\Delta^0_l - \Delta^0_h\right) + 2\Delta^1_l\left(x\Delta^1_l - \Delta^1_h\right) = 0. \tag{20}$$

Form (20), the optimal chirp factor in the coupling mode can be evaluated from

$$x = 1-\alpha^2 = \frac{\Delta^0_l\Delta^0_h + \Delta^1_l\Delta^1_h}{\left(\Delta^0_l\right)^2 + \left(\Delta^1_l\right)^2}. \tag{21}$$

But there is an exception when the original signal is tone-rich. The inverse filtering

mechanism must be turned off for the tone-rich frame, because the tone-rich signal must not be processed in order to maintain the structure completeness of the tone rich signal. Figure 16 illustrates the inverse filtering mode decision in the coupling mode.

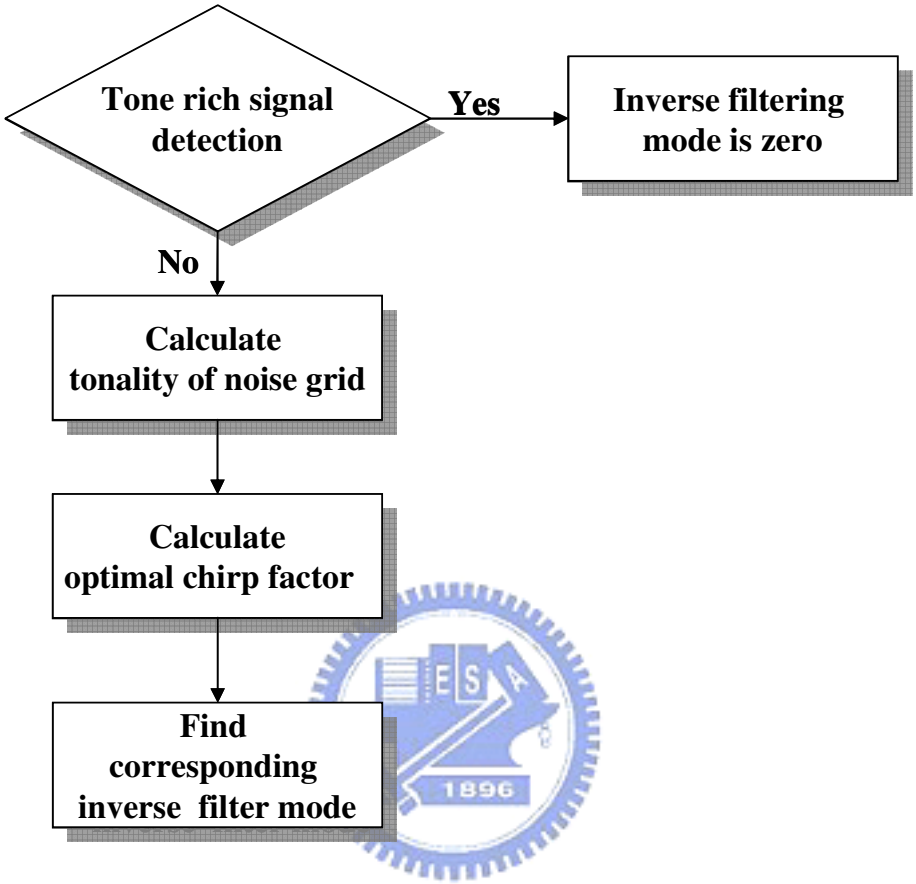


Figure 16: Decision flowchart of inverse filtering mode

3.4 Decision of Noise Floor Scalefactor

As the quantization value of the noise floor factor Q, the quantized noise floor scale factor q is the parameter adjusting the reconstructed tone-noise content. However, according to the syntax constraint, in the coupling mode, the difference of q values between the two channels can not be larger than twelve. Because of the restriction, the optimal q values in the coupling mode may be different from the ones in the normal mode. In this section, the modified decision method for coupling mode is proposed to determine the optimal quantized noise floor scale factor.

In [11][12] of our another work, the proposed method in the normal mode can decide the optimal q to maintain the minimum distortion of the dB-differences among all the high resolution grids contained in the single noise grid. There is only one q

value shared in a single noise grid which includes several high resolution grids. The distortion function is defined as

$$D(Q)=\sum_{k\in ND}\left(\Delta_k^{ND}(Q)-\Delta_k\right)^2+\sum_{k\in TD}\left(\Delta_k^{TD}(Q)-\Delta_k\right)^2, \tag{22}$$

where Δ_k is the ideal tone/noise dB-difference of original signal in k^{th} high resolution grid, $\Delta_k^{ND}, \Delta_k^{TD}$ are the resultant tone/noise dB-difference by q, if k^{th} high resolution grid is noise demand grid and tone demand grid respectively. Hence, the optimal q value must be chosen to minimize the distortion function.

$$Q^*=Arg\left\{\min_Q\left[D(Q)\right]\right\} \tag{23}$$

In the coupling mode, for the quantization process of the noise floor factor Q values, the Q values should be firstly changed into A/R mode from the L/R mode to eliminate signal correlation. As mentioned above, in the coupling mode, the ratio channel of the quantized noise floor needs to be restricted as

$$q_{ratio}\in[0,24]. \tag{24}$$

And, the quantization formula of q_{ratio} is defined as

$$q_{ratio}(k,l)=INT\left(\log_2\left(\frac{Q_{left}(k,l)}{Q_{right}(k,l)}\right)+0.5\right)+12, \tag{25}$$

where Q_{left} and Q_{right} are the noise floor of left and right channel respectively, k is the index of the frequency table band in SBR, l is the index of the envelope. On the other hand, the dequantization formula of the q value in the L/R mod is defined as

$$Q=2^{6-q}. \tag{26}$$

By substituting (26) into (25), it results in

$$q_{ratio}\approx\log_2\left(\frac{2^{6-q_L}}{2^{6-q_R}}\right)+12, \tag{27}$$

where q_L and q_R are the quantized noise scale factor in left and right channel respectively. Hence, from (24) and (27), the constraint of the noise floor scale factor pair of the L/R mode is derived as

$$q_R-q_L\in[-12,12]. \tag{28}$$

Hence, the search range of the candidates of quantized L/R noise scale factor pair in the coupling mode is smaller than the range in the normal mode. As shown in Figure 17, range K is the search scope in the normal mode, the contracted range χ is the search scope in the coupling mode, where

$$\kappa=\left\{(q_L,q_R)\in Z^2\mid 0\le q_L,q_R\le 30\right\}, \tag{29}$$

$$\chi = \{(q_L, q_R) \in \kappa \mid |q_L - q_R| \le 12\}. \quad (30)$$

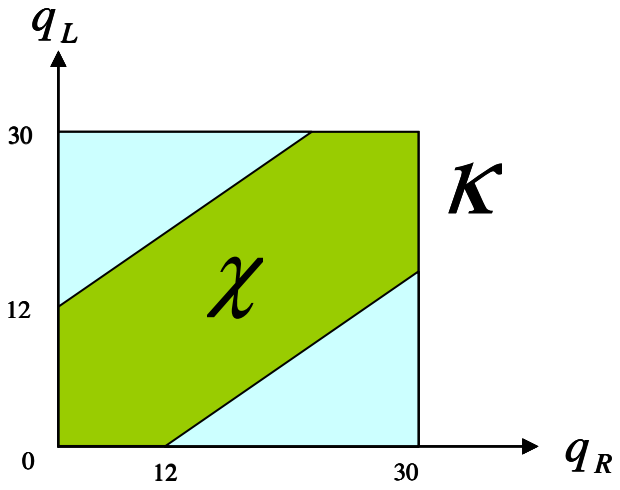


Figure 17: The search range of quantized noise floor for different modes

Hence, the distortion measure function needs to be modified in the coupling mode to consider the two channel distortion under the inherent constraint. As an extension of (22), the modified distortion function for the coupling mode is defined as

$$D_C(q_L, q_R) = D(Q_L(q_L)) + D(Q_R(q_R)), \ (q_L, q_R) \in \chi \quad (31)$$

where $Q(q)$ is the de-quantization function of q. Furthermore, the optimal (Q_L,Q_R) should be decided by choosing the minimum solution pair *(q_L, q_R)* for the distortion function, that is

$$(q_L^*, q_R^*) = Arg\left\{\min_{(q_L, q_R) \in \chi} [D_C(q_L, q_R)]\right\}. \quad (32)$$

Different from the brute-force method to search optimal solution, a modified search method is proposed to reduce the time cost. Figure 18 illustrates the proposed noise floor scale factor decision. We expected that the most pairs of the optimal quantized noise floor scale factor for each individual channel can conform to the constraint of coupling mode. It means the first searching target should be the ones. Once the special pair can fit the constraint, the final optimal solution is also found and the search procedure can be stopped. In summary, we search the optimal scale factors for each channel after the distortion calculation and check of the difference between two channels. If the optimal scale factors for each individual channel can't conform to the constraint, it will search the remained candidate pairs for the optimal solution of the coupling distortion function.

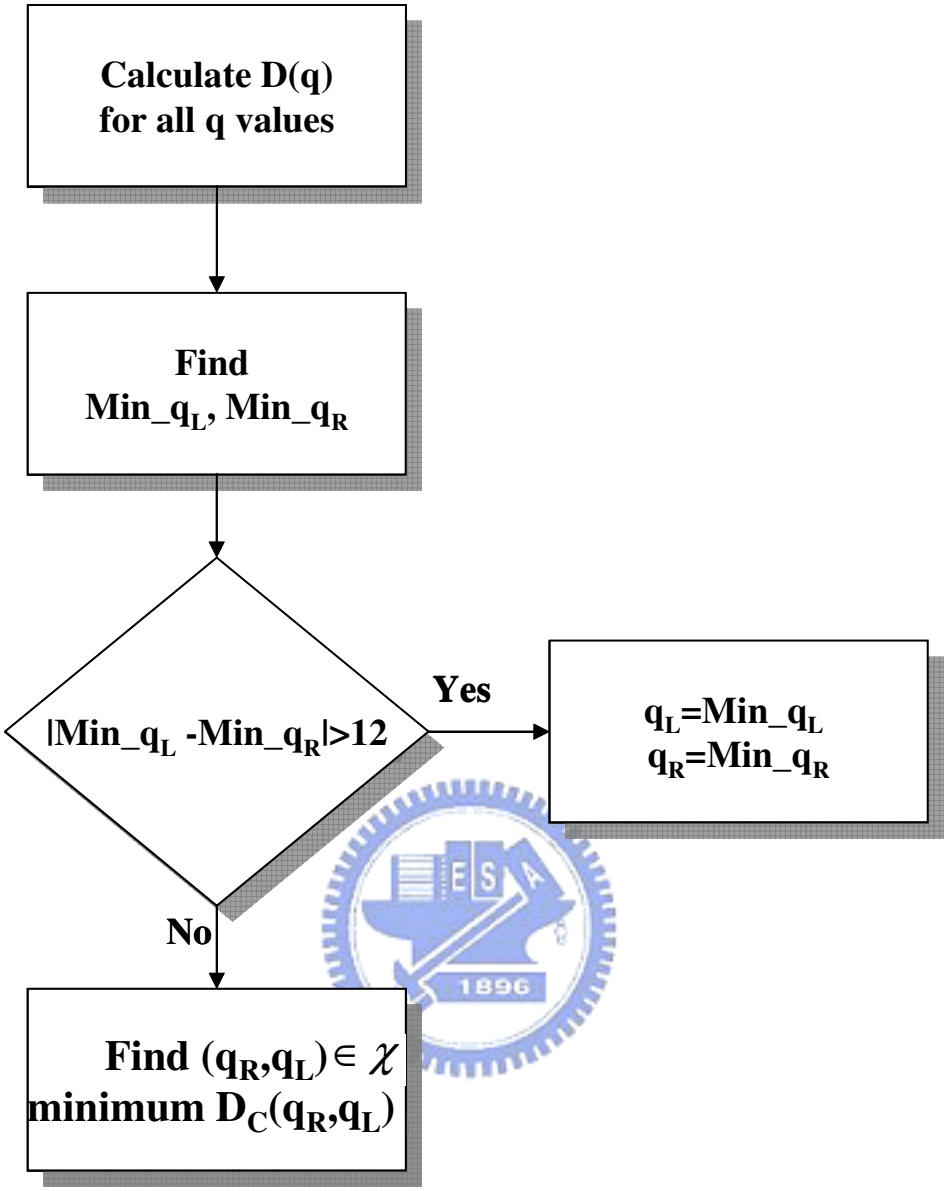


Figure 18: Flowchart of the propose search method of the noise floor scale factor decision

3.5 Coupling Switch Method

The human hearing is relatively more sensitive for the low frequency then the high frequency. To improve the quality of the LF component, the basic objective of the coupling is to save the consuming bits of the HF part as many as possible under the constraint of the least quality requirement. Unlike the aspects of the intensity coding or M/S coding, the main risk of the coupling coding is from the more inaccurate envelope of reconstructed high bands, not fine structure loosing. Because the robustness of SBR, in general, the coupling coding can save a large amount of consuming bits, and promote overall quality under very small risks. However, from

both the two aspects, including the data correlation degree and quantization error variation, the huge difference of the left and right energy channels may result in the degradation of the coding gain and the increase of the reconstruction error. Based on the two points, a coupling switch decision method is proposed for the tradeoff between HF quality and demand bits. Taking into account the above risks with the least bits consuming, the optimal represent domain of the signal, either L/R or A/R mode, will be chosen.

3.5.1 Quantization error analysis

As the following, the variations of the quantization errors at different mode will be analyzed. The spectral energy is quantized into the scale factor by taking time-frequency grid of the current frame as the recorder units. There are different quantization methods in the normal mode and the coupling mode. The quantization formula in the normal mode is defined as

$$E_Q(k,l) = INT\left(a \cdot \max\left(\log_2\left(\frac{E(k,l)}{64} \right), 0 \right) + 0.5 \right), 0 \le l \le L_E, \tag{33}$$

where $a = \begin{cases} 2 & ,if\ bs\_amp\_res = 0 \\ 1 & ,if\ bs\_amp\_res = 1 \end{cases}$, k is the index of the frequency table band in SBR, l is the index of the envelope, L_E is the number of envelope in current SBR frame, and $E(k,l)$ means the energy of input signal in the normal mode, and also means the average energy between the stereo channel in coupling mode. Furthermore, the right channel quantization formula in the coupling mode is defined as

$$E_{QRight}(k,l) = INT\left(a \cdot \log_2\left(E(k,l)\right) + 0.5\right) + panOffset\left(bs\_amp\_res\right), \tag{34}$$

where $E(k,l)$ is the energy ratio between the stereo channel. On the other hand, the de-quantization method for the normal mode in HE-AAC decoder is defined as

$$E_{Orig}(k,l) = 64 \cdot 2^{\frac{E(k,l)}{a}}, \begin{cases} 0 \le l < L_E \\ 0 \le k < n\left(r(l)\right) \end{cases}, \tag{35}$$

where $E(k,l)$ is decoded envelope scale factor. Also, by the incorporation of the de-A/R operation, the de-quantization method in the coupling mode is defined as

$$E_{LeftQrig}(k,l) = 64 \cdot \frac{2^{\frac{E_0(k,l)}{a}+1}}{1+2^{\frac{panOffset(bs\_amp\_res) - E_1(k,l)}{a}}}, \begin{cases} 0 \le l < L_E \\ 0 \le k < n\left(r(l)\right) \end{cases}, \tag{36}$$

$$E_{RightQrig}(k,l) = 64 \cdot \frac{2^{\frac{E_0(k,l)}{a}+1}}{1+2^{\frac{E_1(k,l)-panOffset(bs\_amp\_res)}{a}}}, \begin{cases} 0 \le l < L_E \\ 0 \le k < n(r(l)) \end{cases}, \tag{37}$$

where E_0,E_1 represent the decoded A/R envelope scale factors.

Through the quantization formulas above, the reconstruction error can be estimated. From (33) and (35), the reconstruction value in the normal mode can be calculated as

$$E' = 64 \times 2^{\frac{\text{int}(a \times \max(\log_2(\frac{E}{64}),0)+0.5)}{a}}, \tag{38}$$

where E is the energy of original signal. From (38), If E is smaller than 64, E' is always 64. If E is greater than 64, E' is calculated as

$$E' = 64 \times 2^{\frac{\text{int}(a \times \log_2(\frac{E}{64})+0.5)}{a}}. \tag{39}$$

Form (39) , the relative error in the normal mode can be evaluated from

$$\frac{E - E'}{E} = 1 - 2^{\frac{\varepsilon}{a}}, \tag{40}$$

where some ε locates at the range from -0.5 to 0.5. If ε is zero value, it implies there is no reconstruction error.

From (33), (34), (36), and (37), the reconstruction value of two channels in the coupling mode is derived as

$$E_L' = \left\{ 64 \cdot 2^{\frac{\text{int}(a \times \log_2(\frac{E_L+E_R}{128})+0.5)}{a}} \right\} \cdot \left\{ 2 \cdot \left(1+2^{\frac{-\text{int}(a \times \log_2(\frac{E_L}{E_R})+0.5)}{a}} \right)^{-1} \right\}, \tag{41}$$

$$E_R' = \left\{ 64 \times 2^{\frac{\text{int}(a \times \log_2(\frac{E_L+E_R}{128})+0.5)}{a}} \right\} \cdot \left\{ 2 \cdot \left(1+2^{\frac{\text{int}(a \times \log_2(\frac{E_L}{E_R})+0.5)}{a}} \right)^{-1} \right\}, \tag{42}$$

where E_L and E_R are the energy of original signal. The ratio of the reconstructed and original energies in the coupling mode can be evaluated from (41), (42), that is

$$\frac{E_L'}{E_L} = \frac{\Psi \cdot 2^{\frac{\varepsilon_1+\varepsilon_2}{a}} + 2^{\frac{\varepsilon_1+\varepsilon_2}{a}}}{\Psi \cdot 2^{\frac{\varepsilon_2}{a}} + 1}, \tag{43}$$

$$\frac{E_R'}{E_R} = \frac{\Psi \cdot 2^{\frac{\varepsilon_1}{a}} + 2^{\frac{\varepsilon_1}{a}}}{1 + \Psi \cdot 2^{\frac{\varepsilon_2}{a}}}, \tag{44}$$

where $\Psi = \frac{E_L}{E_R}$, $\varepsilon_1 \in [-0.5, 0.5]$, and $\varepsilon_2 \in [-1.5, 0.5]$. The constant ε_1 is the quantization error of the energy quantization process, and the constant ε_2 is the reconstruction error which results from the quantization process and the DPCM operation in the SBR.

3.5.2 Energy Abnormal Phenomenon

From (43) and (44), if Ψ is huge, the reconstruction error will approximate the limit value as follows

$$\lim_{\Psi \to \infty} \frac{E_L^{'}}{E_L} = 2^{\frac{\varepsilon_1}{a}}, \tag{45}$$

$$\lim_{\Psi \to \infty} \frac{E_R^{'}}{E_R} = 2^{\frac{\varepsilon_1 - \varepsilon_2}{a}}. \tag{46}$$

On the other hand, if Ψ is very small, the reconstruction error will approximate the limit value as follows

$$\lim_{\Psi \to 0} \frac{E_L^{'}}{E_L} = 2^{\frac{\varepsilon_1 + \varepsilon_2}{a}}, \tag{47}$$

$$\lim_{\Psi \to 0} \frac{E_R^{'}}{E_R} = 2^{\frac{\varepsilon_1}{a}}. \tag{48}$$

If the values $\frac{E_R^{'}}{E_R}$, $\frac{E_L^{'}}{E_L}$ are more close to one, it means the relative error is more close to zero. However, from (40), (46), and (47), the two value $\varepsilon_1 - \varepsilon_2$ and $\varepsilon_1 + \varepsilon_2$ occurring in the coupling mode may be much larger than the value ε in the normal mode. This is because that the ε belongs the range from 0.5 to -0.5, and the distribution of $\varepsilon_1 - \varepsilon_2$ and $\varepsilon_1 + \varepsilon_2$ is larger, that is $\varepsilon_1 - \varepsilon_2 \in [-1, 2]$, and $\varepsilon_1 + \varepsilon_2 \in [-2, 1]$. It implies that the reconstructed error of weaker channel may abnormally become very large when the energy difference between the two channels is large. For example, Figure 19 is the spectrum of the test stereo signal to illustrate the error augment phenomenon. The test track has high energy difference in HF between L/R channels. As shown in Figure 20, it is the comparison of reconstruction signal between coupling and normal mode. The energy difference is about 3 dB between the reconstruction signals in the coupling and normal mode. Figure 21 indicates that the stronger channel has the nearer reconstruction signal between coupling and normal mode.

In order to limit the reconstruction error to the endurable range, it needs to find the reasonable relative range of Ψ to switch coupling mode. Figure 22 and Figure 23 illustrate the relationship between Ψ and the mean of relative error in the right channel. The mean of relative error is calculated as follows

$$mean(\Psi) = \frac{1}{2.5}\int_{-0.5}^{0.5}\int_{-1.5}^{1}\left(1-\frac{2^{\frac{\varepsilon_1}{a}}+\Psi\cdot 2^{\frac{\varepsilon_1}{a}}}{1+\Psi\cdot 2^{\frac{\varepsilon_2}{a}}}\right)d\varepsilon_1 d\varepsilon_2 . \tag{49}$$

The mean of relative error is close rapidly to the upper bound when Ψ is over 8, and the coupling mechanism should be turned off to avoid the annoying phenomenon of the error augment. Figure 24 and Figure 25 indicate the relationship between Ψ and the variance of relative error in the right channel. The variance of relative error is calculated as follows

$$\text{var}iance(\Psi) = \frac{1}{2.5}\int_{-0.5}^{0.5}\int_{-1.5}^{1}\left(\left(1-\frac{2^{\frac{\varepsilon_1}{a}}+\Psi\cdot 2^{\frac{\varepsilon_1}{a}}}{1+\Psi\cdot 2^{\frac{\varepsilon_2}{a}}}\right)-mean(\Psi)\right)^2 d\varepsilon_1 d\varepsilon_2 . \tag{50}$$

The variance of relative error is also close rapidly to the upper bound when Ψ is over 8. There is a trade-off between the reconstruction error and the saved bit, and hence a switch threshold is required to avoid extreme reconstruction error. Similarly, Figure 26 and Figure 27 illustrate the relationship between Ψ and the mean of relative error in the left channel. Figure 28 and Figure 29 indicate the relationship between Ψ and the variance of relative error in the left channel. The mean and variance of relative error is calculated as

$$mean(\Psi) = \frac{1}{2.5}\int_{-0.5}^{0.5}\int_{-1.5}^{1}\left(1-\frac{2^{\frac{\varepsilon_1}{a}}+\Psi\cdot 2^{\frac{\varepsilon_1}{a}}}{\Psi+2^{-\frac{\varepsilon_2}{a}}}\right)d\varepsilon_1 d\varepsilon_2 \tag{51}$$

and

$$\text{var}iance(\Psi) = \frac{1}{2.5}\int_{-0.5}^{0.5}\int_{-1.5}^{1}\left(\left(1-\frac{2^{\frac{\varepsilon_1}{a}}+\Psi\cdot 2^{\frac{\varepsilon_1}{a}}}{\Psi+2^{-\frac{\varepsilon_2}{a}}}\right)-mean(\Psi)\right)^2 d\varepsilon_1 d\varepsilon_2 \tag{52}$$

for left channel, respectively. According to the mean and variance of relative error, the switch threshold is set to 8 to avoid the high relative error.

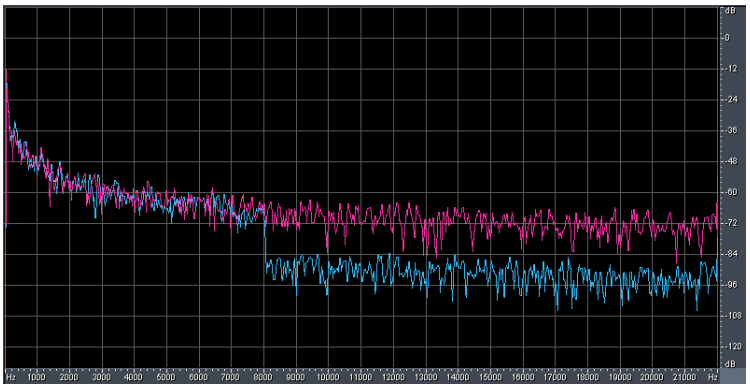


Figure 19: The test track has high energy difference in high frequency

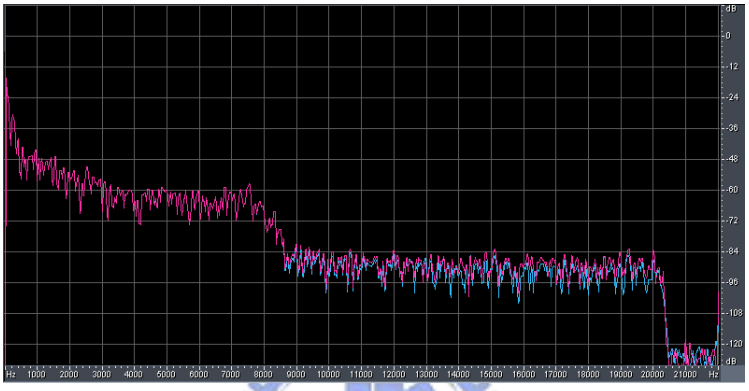


Figure 20: The high reconstruction error in weaker channel

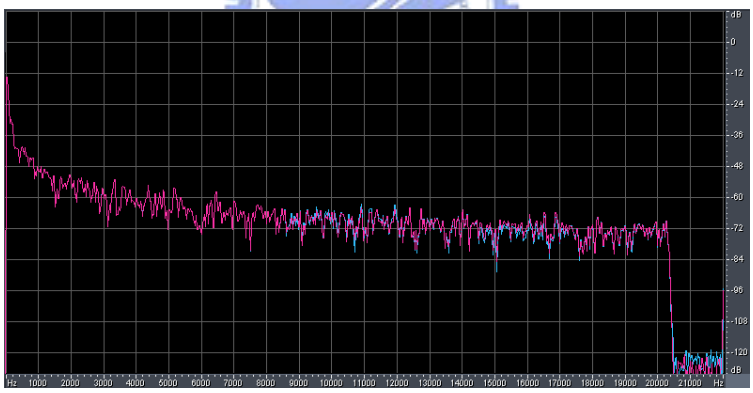


Figure 21: The low reconstruction error in stronger channel

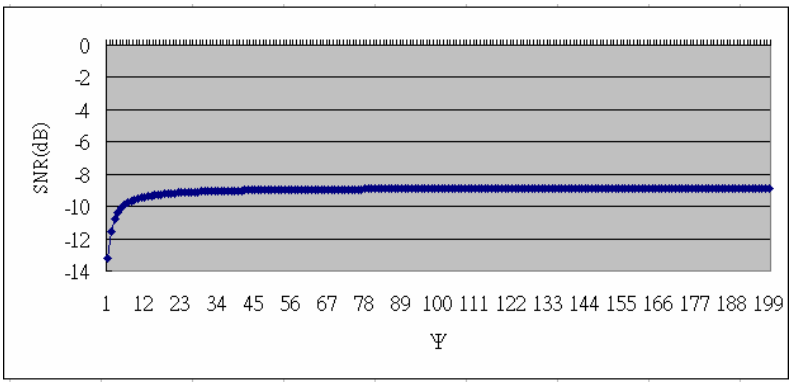


Figure 22: The relationship between mean of relative error and Ψ when Ψ is more

than one in the right channel

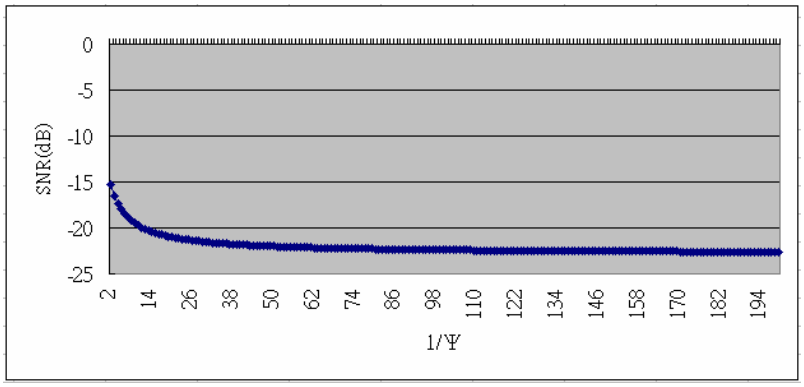


Figure 23: The relationship between mean of relative error and Ψ when Ψ is less than one in the right channel

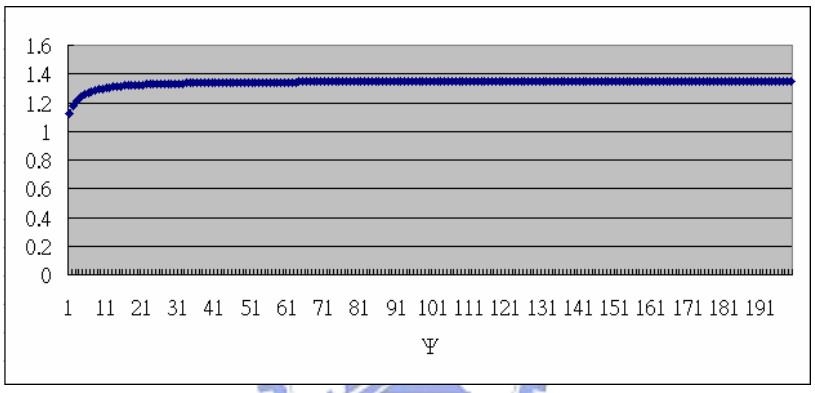


Figure 24: The relationship between variance of relative error and Ψ when Ψ is more than one in the right channel

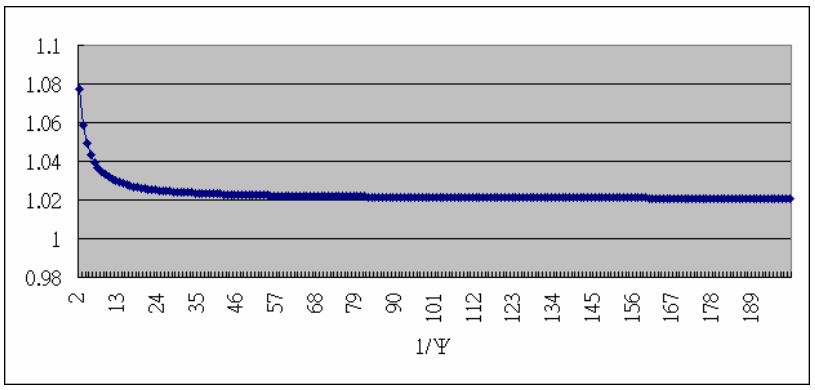


Figure 25: The relationship between variance of relative error and Ψ when Ψ is less than one in the right channel

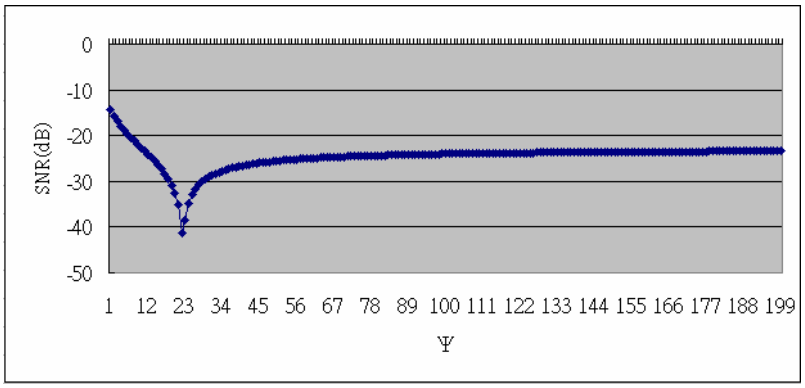


Figure 26: The relationship between mean of relative error and Ψ when Ψ is more than one in the left channel

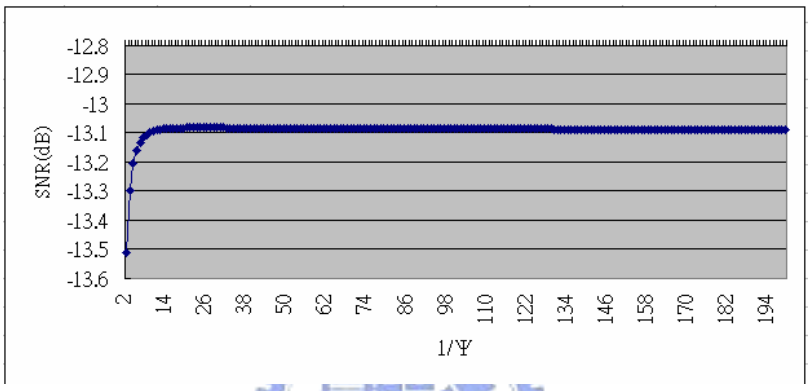


Figure 27: The relationship between mean of relative error and Ψ when Ψ is less than one in the left channel

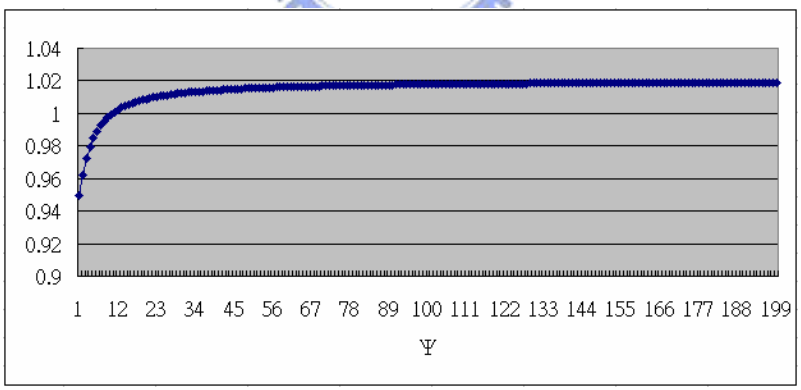


Figure 28: The relationship between variance of relative error and Ψ when Ψ is more than one in the left channel

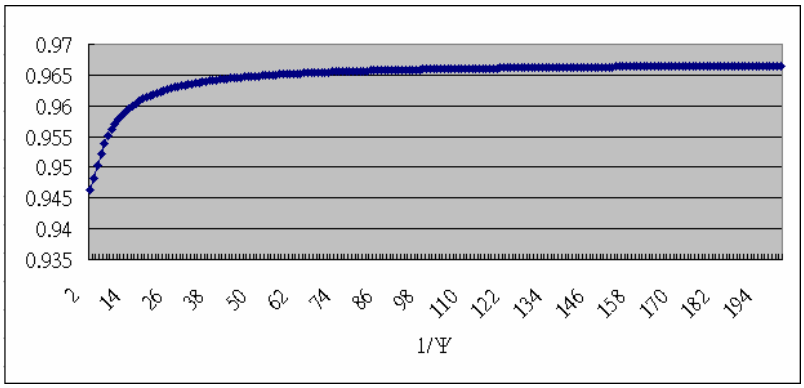


Figure 29: The relationship between variance of relative error and Ψ when Ψ is less than one in the left channel

3.5.3 Summary

Therefore, the criterion of the proposed coupling switch method focuses on the average energy difference in the SBR range to avoid the abnormal phenomenon. The average energy difference is calculated from the energy ratio which is divided by critical unit. The number of samples of critical band in high frequency is different from the number in the low frequency, hence the energy ratio must be normalized as follows,

$$E_{Diff} = \sum_{c \in F} \left(\frac{\sum_{j \in c} \left| \log_2 \left(E_j^{c,l} \right) - \log_2 \left(E_j^{c,r} \right) \right|}{|c|} \right), \tag{53}$$

where c is the critical unit we used in Section 3.2, $E_j^{c,i}$ is the j[th] sample energy of critical unit c of channel i, and $|c|$ is the number of samples of critical unit. Figure 30 illustrates the proposed coupling switch method which detects of the huge average energy difference.

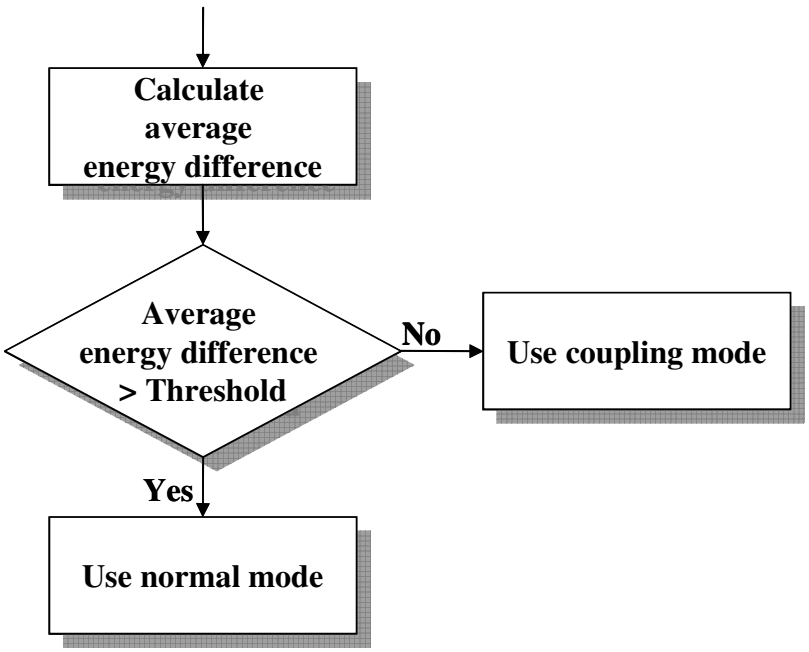


Figure 30: Coupling Switch Flowchart.

Chapter 4 Experiments

In this chapter, the quality measurement is conducted on the NCTU_HEAAC platform. Extensive experiments are performed to prove the enhancement of the proposed methods on the MPEG test tracks and the music database [13] collected in PSPLAB.

4.1 Experiment Environment

Objective Quality Measurement Tool:

The tool called "EAQUAL" [14] is chosen to measure the audio quality in the objective test. EAQUAL stands for "Evaluation of Audio Quality". The purpose of EAQUAL is to supply the audio objective quality measurement for coded/decoded audio signals especially useful for audio codec development. The implementation of EAQUAL is based on the ITU-R recommendation BS.1387 [15].

Subjective Quality Measurement Tool:

In subjective quality test, we use MUSHRA [16] to assist the assessment. Multi stimulus test with hidden anchors and reference has been designed to give a reliable and repeatable measure of the audio quality of intermediate-quality signals. MUSHRA has the advantage that it provides an absolute measure of the audio quality of the codec which can be compared directly with the reference. MUSHRA follows the test method and impairment scale recommended by the ITU-R BS.1116 [17].

4.2 Objective Quality Measurement in MPEG Test Tracks

MPEG twelve tracks include critical music balancing on the percussion, string, wind instruments, and human vocal. The features of these twelve tracks are shown in the Table 4. In this section, it will verify the quality enhancement of proposed methods in different bit rates based on the MPEG test tracks.

Table 4: The twelve tracks recommended by MPEG

Tracks		Signal Description			
		Signals	Mode	Time (sec)	Remark
1	es01	Vocal (Suzan Vega)	stereo	10	(c)
2	es02	German speech	stereo	8	(c)
3	es03	English speech	stereo	7	(c)
4	sc01	Trumpet solo and orchestra	stereo	10	(b) (d)
5	sc02	Orchestral piece	stereo	12	(d)
6	sc03	Contemporary pop music	stereo	11	(d)
7	si01	Harpsichord	stereo	7	(b)
8	si02	Castanets	stereo	7	(a)
9	si03	pitch pipe	stereo	27	(b)
10	sm01	Bagpipes	stereo	11	(b)
11	sm02	Glockenspiel	stereo	10	(a) (b)
12	sm03	Plucked strings	stereo	13	(a) (b)

Remarks:

(a) Transients: pre-echo sensitive, smearing of noise in temporal domain.

(b) Tonal/Harmonic structure: noise sensitive, roughness.

(c) Natural vocal (critical combination of tonal parts and attacks): distortion sensitive, smearing of attacks.

(d) Complex sound: stresses the device under test.

Table 5: Objective measurements through the ODGs for proposed coupling approach at 80 kbps

Codec	NCTU-HEAAC	
Bit Rate	80k kbps	
Tracks	M0	M1
es01	-0.68	-0.67
es02	-0.58	-0.56
es03	-0.68	-0.64
sc01	-0.95	-0.93
sc02	-1.08	-1.05
sc03	-1.1	-1.09
si01	-1.56	-1.55
si02	-1.02	-1.02
si03	-1.62	-1.6
sm01	-1.56	-1.52
sm02	-1.55	-1.51
sm03	-1.29	-1.27
Max	-0.58	-0.56
Min	-1.62	-1.6
Average	-1.1392	-1.1175
M0: coupling mode disabled		
M1: coupling mode enabled		

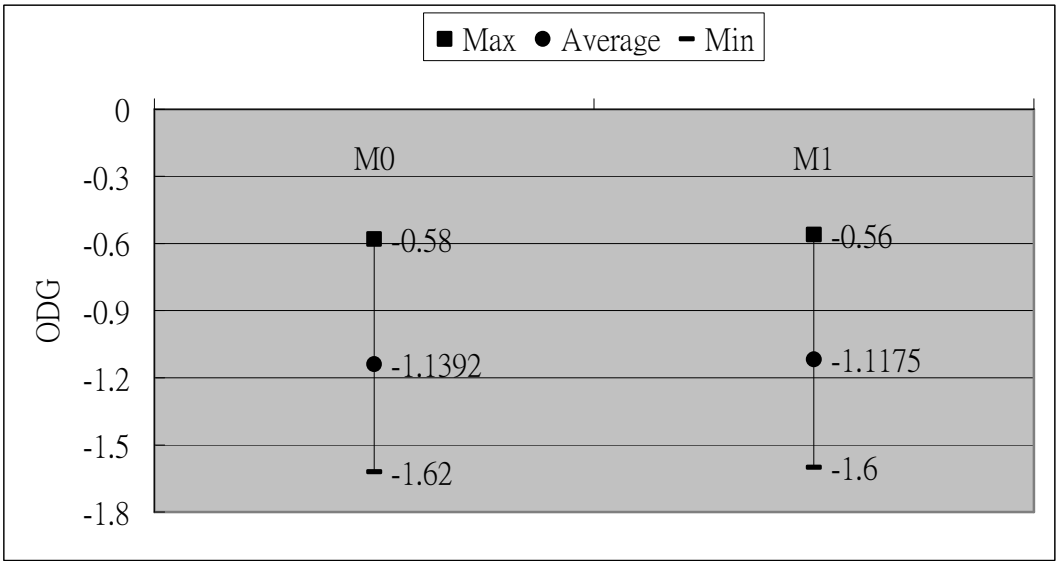


Figure 31: The variance in the ODGs of proposed coupling approaches at 80 kbps

Table 6: Objective measurements through the ODGs for proposed coupling approach at 64 kbps

Codec	NCTU-HEAAC	
Bit Rate	64k kbps	
Tracks	M0	M1
es01	-0.92	-0.92
es02	-0.8	-0.77
es03	-0.95	-0.89
sc01	-1.6	-1.54
sc02	-1.66	-1.63
sc03	-1.59	-1.56
si01	-1.89	-1.87
si02	-1.35	-1.3
si03	-2.07	-2.04
sm01	-2.11	-2.06
sm02	-2.19	-2.19
sm03	-1.64	-1.63
Max	-0.8	-0.77
Min	-2.19	-2.19
Average	-1.5642	-1.5333
M0: coupling mode disabled		
M1: coupling mode enabled		

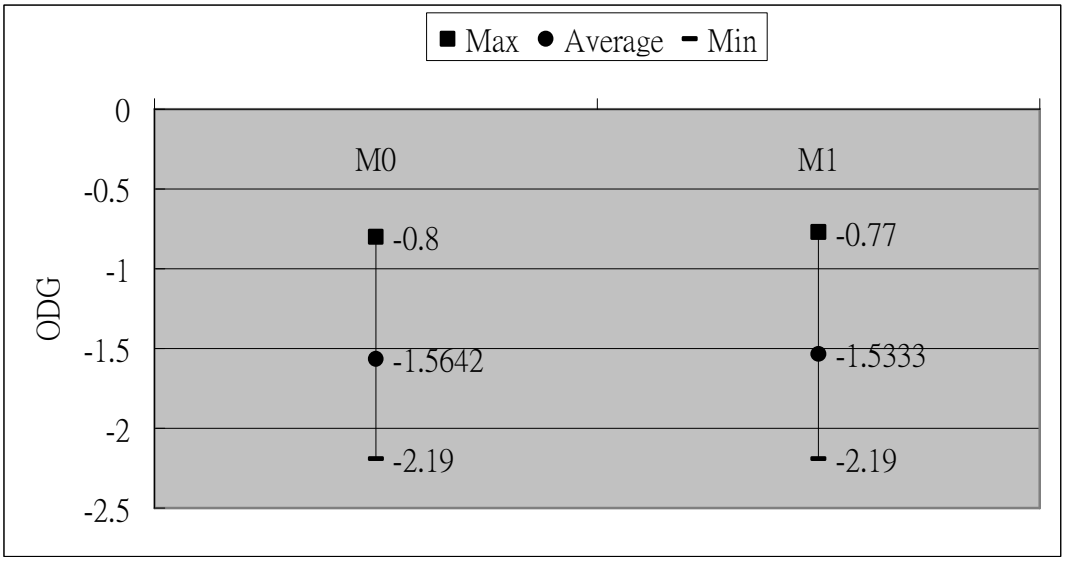


Figure 32: The variance in the ODGs of proposed coupling approaches at 64 kbps

Table 7: Objective measurements through the ODGs for proposed coupling approach at 48 kbps

Codec	NCTU-HEAAC	
Bit Rate	48k kbps	
Tracks	M0	M1
es01	-1.47	-1.37
es02	-1.34	-1.27
es03	-1.64	-1.51
sc01	-2.36	-2.35
sc02	-2.53	-2.49
sc03	-2.28	-2.22
si01	-2.69	-2.69
si02	-2.25	-2.19
si03	-3.16	-3.11
sm01	-3.15	-3.07
sm02	-3.23	-3.21
sm03	-2.22	-2.2
Max	-1.34	-1.27
Min	-3.23	-3.21
Average	-2.36	-2.3067
M0: coupling mode disabled		
M1: coupling mode enabled		

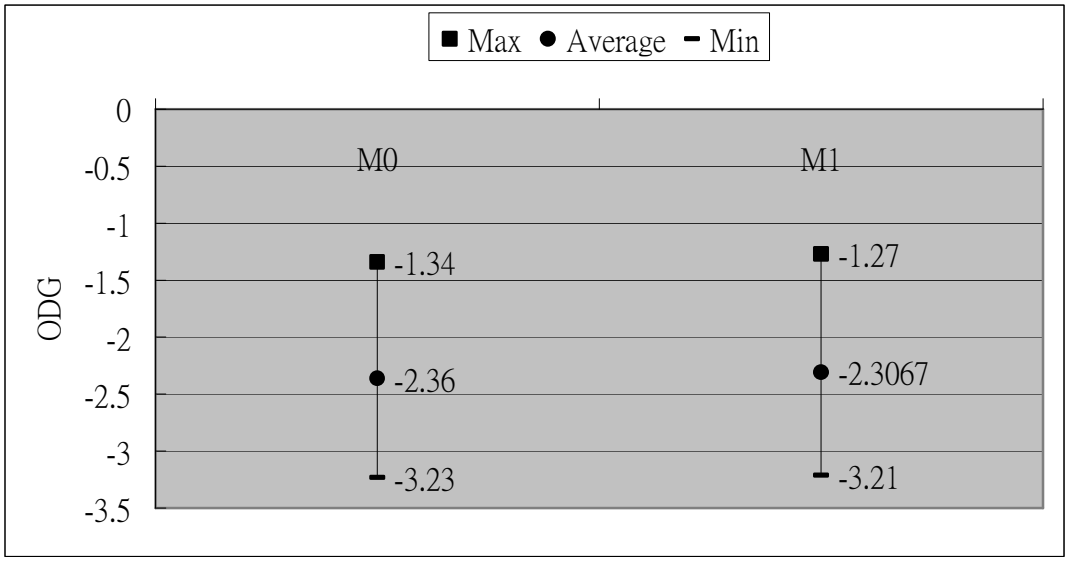


Figure 33: The variance in the ODGs of proposed coupling approaches at 48 kbps

According to the result of the objective test, the encoder which enables coupling mode has a better quality. Especially, because of the similarity of L/R channels, the voice tracks including “es01”, “es02” and “es03” have the conspicuous improvement among all the test tracks. On the other hand, the improvement degree is larger with the bit-rate decreases. This is because that the saved bits by coupling coding are more important at the lower bit-rate.

4.3 Objective Quality Measurement in Music Database

To confirm the possible risk and robustness of proposed methods, extensive tests are adapted to verify the qualities of these methods. In PSPLab audio database [13], here are 15 sets. For each bitstream set, they are described in Table 8.

Table 8: The PSPLab audio database [13]

Bitstream Categories		# of tracks	Remark
1	Ff123	103	Killer bitstream collection from ff123 [18].
2	Gpsycho	24	LAME quality test bitstream [19].
3	HA64KTest	39	64 kbps test bitstream for multi-format in HA forum [20].
4	HA128KTestV2	12	128 kbps test bitstream for multi-format in HA forum [20].
5	Horrible_song	16	Collections of critical songs among all bitstreams in PSPLab.
6	Ingets1	5	Bitstream collection from the test of OGG Vorbis pre 1.0 listening test [21].
7	MPEG	12	MPEG test bitstream set for 48000Hz.
8	MPEG44100	12	MPEG test bitstream set for 44100 Hz.
9	Phong	8	Test bistream collection from Phong [22].
10	PSPLab	37	Collections of bitstream from early age of PSPLab. Some are good as killer.
11	Sjeng	3	Small bitstream collection by sjeng.
12	SQAM	16	Sound quality assessment material recordings for subjective tests [23].
13	TestingSong14	14	Test bitstream collection from rshong, PSPLab.
14	TonalSignals	15	Artificial bitstream that contains sin wave etc.
15	VORBIS_TESTS_ Samples	8	Eight Vobis testing samples from HA [20].

M0: coupling mode disabled

M1: coupling mode enabled

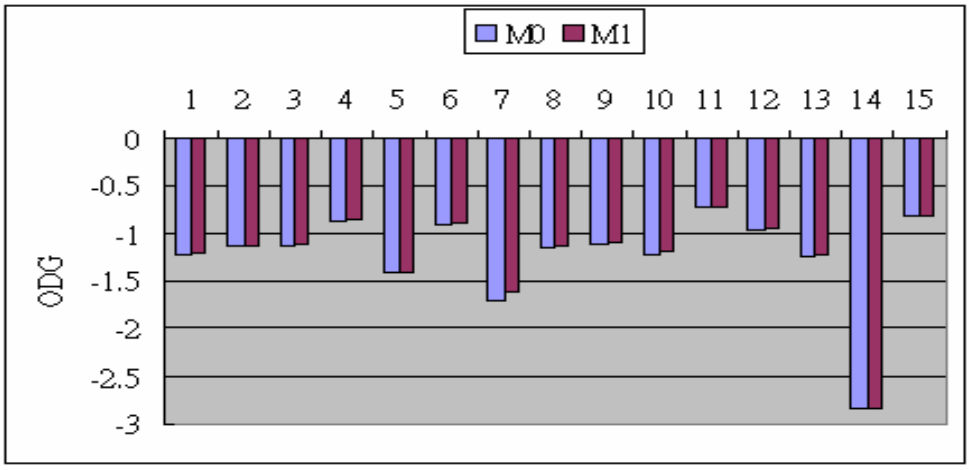


Figure 34: The average ODGs of method M0 and M1 at 80kbps in 16 categories

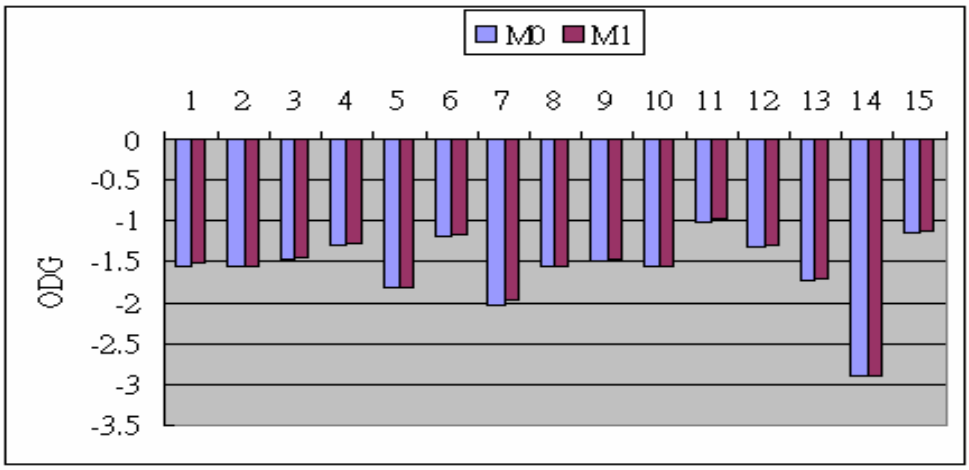


Figure 35: The average ODGs of method M0 and M1 at 64kbps in 16 categories

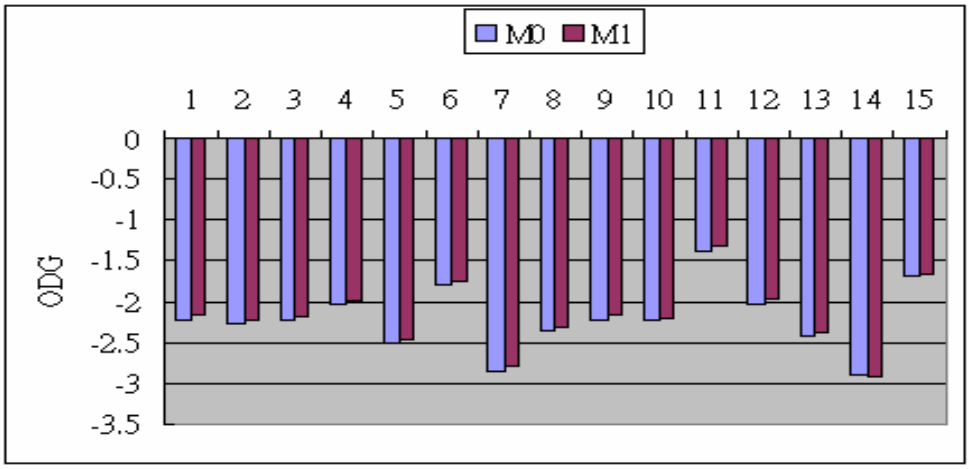


Figure 36: The average ODGs of method M0 and M1 at 48kbps in 16 categories

Figure 34 – 25 illustrate the experiments for the 15 bitstream sets at the different bit-rates, where the bars indicate the average ODG of the bitstream sets. For the most cases, the quality under the coupling mode is better than the one under non-coupling mode. The main characteristic of the improvement tracks is that L/R channels are similar in HF component. On the other hand, there are two tracks which degrade quality in the coupling mode. The two tracks are “impulse_m20_0db” and “triangle1” in the “TonalSignals” set. As shown in Figure 37, the cause of the degradation of the “impulse_m20_0db” track is that the silence suffers the spreading noise in coupling mode. Furthermore, the mechanism of the noise-floor correction [11] will miss detection in coupling mode for “triangle1”. Figure 39 and Figure 40 show the spectrums which are reconstructed under the normal mode and the coupling mode, respectively.

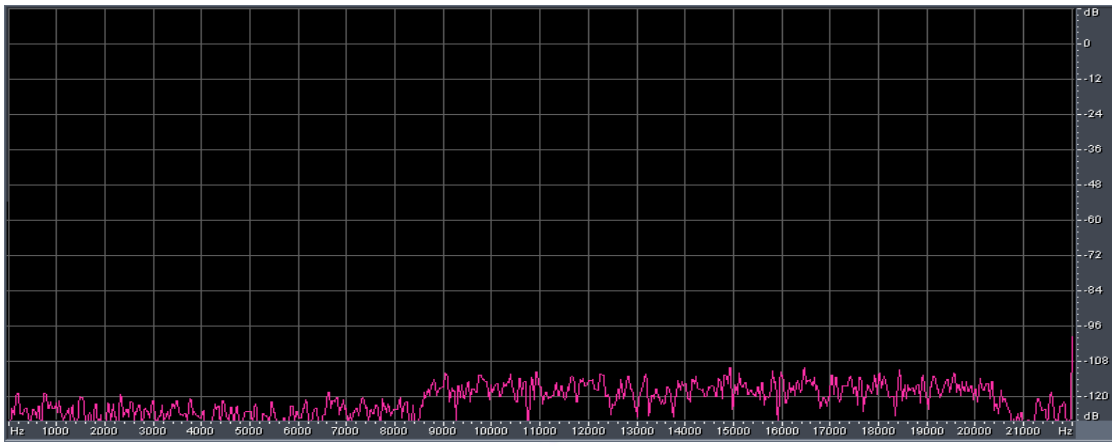


Figure 37: The suffered spectrum of the silence in the “impulse_m20_0db” track

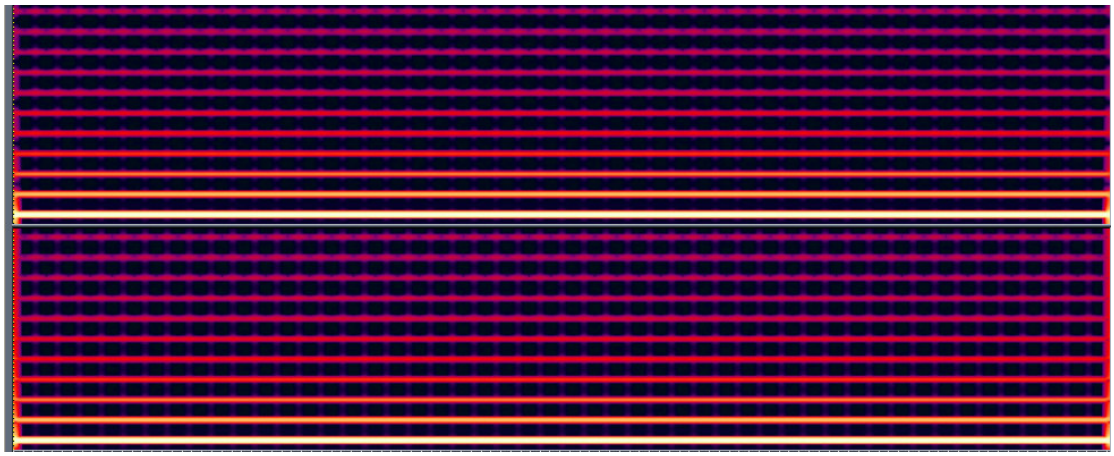


Figure 38: The spectrum of the “triangle1” track in the “TonalSignals” set

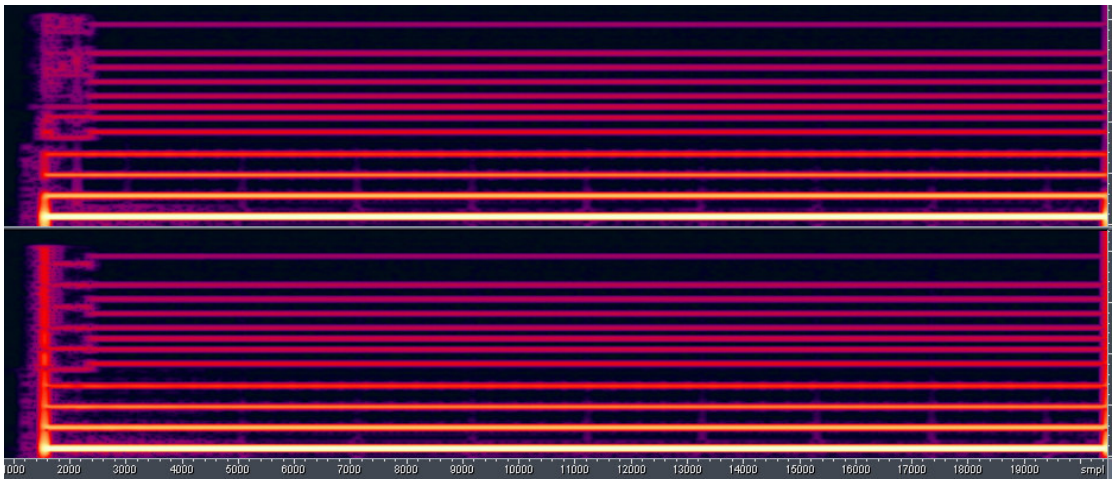


Figure 39: Reconstructed spectrogram of the “triangle1” track in normal mode

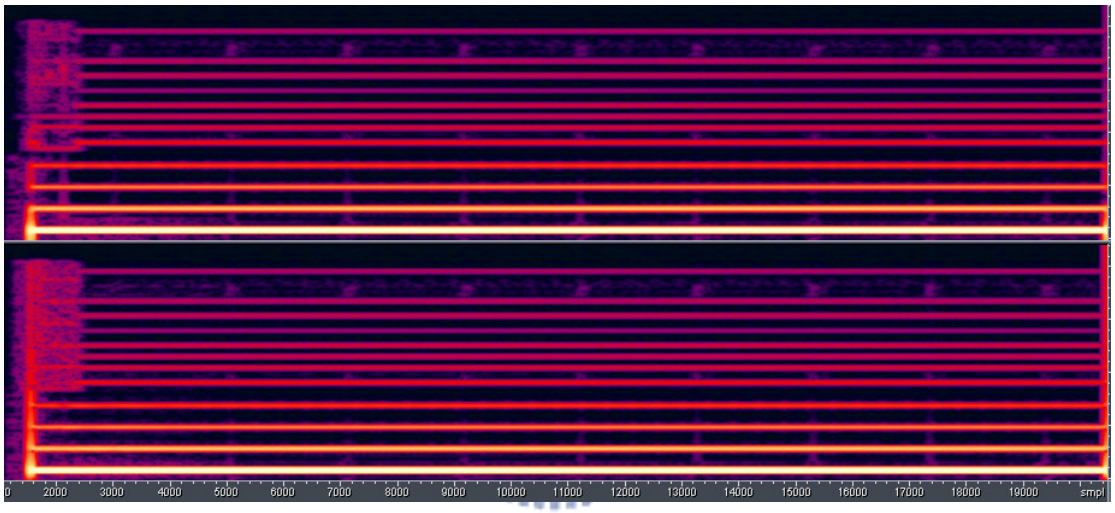


Figure 40: Reconstructed spectrogram of the “triangle1” track in coupling mode

4.4 Subjective Quality Measurement

After the objective quality measurement, the subjective listening test needs to be performed to verify the quality improvement and possible risk of proposed methods in this thesis. The twelve test tracks are selected from the MPEG test tracks. The subjective listening test is performed on the codec “NCTU-HEAAC” and use the tool called “MUSHRA” to assist the assessment. The result of the subjective quality measurement is shown in the following figure.

M0: coupling mode disabled
M1: coupling mode enabled

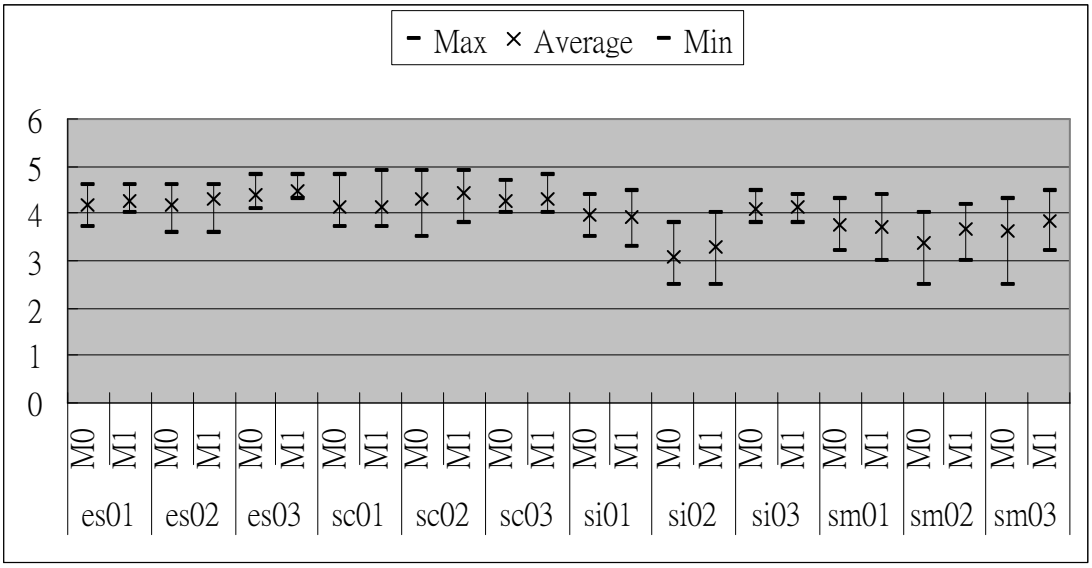


Figure 41: The result of the subjective test for coupling coding at 48kbps

Chapter 5 Conclusion

This thesis focused on the design of coupling coding method in MPEG-4 HEAAC to improve the audio quality at very low bit rate. Based on the previous work [9][11], the thesis extended the several decision method in normal mode into the coupling mode and considered possible risk. First, the decision methods of the shared T/F gird parameter, shared inverse filtering intensity, and the noise-floor factor have been proposed. Finally, a switch method considering the channel energy ratio and the frame content has also been presented. On the other hand, the extensive experiments have been conformed to verify the quality enhancement. Through both subjective listening test and objective quality measurement, the proposed methods are verified to improve the quality of the coded audio signals.

References

[1] ISO/IEC 14496-3:1999, “Information Technology–Coding of Audiovisual objects, Part3: Audio.”

[2] ISO/IEC, “Text of ISO/IEC 14496-3:2001/FDAM1, Bandwidth Extension,” ISO/IEC JTC1/SC29/WG11/N5570, March 2003, Pattaya, Thailand.

[3] M. Dietz, L. Liljeryd, K. Kjörling, O. Kunz, “Spectral Band Replication, a novel approach in audio coding,” at the 112th AES Convention, Munich, May 10–13, 2002.

[4] M. Wolters, K. Kjörling, D. Homm, H. Purnhagen, “Acloser look into MPEG-4 High Efficiency AAC,” at the 115th AES Convention, New York, USA, October 10–13, 2003.

[5] H.W. Hsu, C.M. Liu, and W.C. Lee, “Audio Patch Method in MPEG-4 HE-AAC Decoder,” at the 117th AES Convention, San Francisco, USA, October 28~31, 2004.

[6] Kan-Chun Lee, “Design of T/F Stereo Parameter Extraction and Downmix Procedure in Parametric Stereo Coding”, CSIE Master Thesis of NCTU, 2006

[7] H. Purnhagen: “Low Complexity Parametric Stereo Coding in MPEG-4”, 7th International Conference on Audio Effects (DAFX-04), Naples, Italy, October 2004.

[8] E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegård: “Low complexity parametric stereo coding”, Proc. 116th AES convention, Berlin, Germany, 2004, Preprint 6073.

[9] Shou-Hung Tang, “Efficient Design of Time/Frequency Grid in HE-AAC Encoder”, CSIE Master Thesis of NCTU, 2006

[10] AES Technical Committee of Coding of Audio Signals, “Perceptual Audio Coders: What to listen for”, CD-ROM with tutorial information and audio examples, AES publications

[11] Yung-Cheng Yang, “Design for High Frequency Adjustment Module in MPEG-4 HE-AAC” CSIE Master Thesis of NCTU, 2006

[12] H. W. Hsu, Y. C. Yang, C. M. Liu, and W. C. Lee, “Design for High Frequency Adjustment Module in MPEG-4 HE-AAC Encoder Based on Linear Prediction Method,” at the 120th AES Convention, Paris , France, May 20~23, 2006.

[13] PSPLAB audio database, website

http://psplab.csie.nctu.edu.tw/projects/index.pl/testbitstreams.html .

[14] EAQUAL, mirror site: http://public.planetmirror.com/pub/sourceforge/e/ea/eaqual/, Date: June 03, 2005.

[15] International Telecommunications Union, Radiocommunication Sector Bs.1387, "Method for the ObjectiveMeasurement of Perceived Audio Quality," Geneva 1998.

[16] MUSHRA, website: http://ff123.net/abchr/abchr.html, Date: June 03, 2005.

[17] International Telecommunications Union, Radiocommunication Sector BS.1116 (rev. 1), "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," Geneva, 1997.

[18] Samples for Testing Audio Codecs from ff123, website: http://ff123.net/samples.html, Date: June 03, 2005.

[19] Quality and Listening Test Information for LAME, website: http://lame.sourceforge.net/gpsycho/quality.html, Date: June 03, 2005.

[20] Hydrogen Audio, website: http://www.hydrogenaudio.org, Date: June 03, 2005.

[21] OGG Vorbis Pre 1.0 Listening Test, website: http://hem.passagen.se/ingets1/vorbis.htm, Date: June 03, 2005.

[22] Phong's Audio Samples, website: http://www.phong.org/audio/samples.xhtml, Date: June 03, 2005.

[23] Sound Quality Assessment Material, website: http://sound.media.mit.edu/mpeg4/audio/sqam/, Date: June 03, 2005.