

國立交通大學

資訊科學與工程研究所

碩士論文



HE-AAC 編碼器上之有效率的 Time/Frequency
Grid 設計

Efficient Design of Time/Frequency Grid in HE-AAC Encoder

研究生：唐守宏

指導教授：劉啟民 教授

李文傑 教授

中華民國九十五年七月

HE-AAC 編碼器上之有效率的 Time/Frequency Grid 設計
Efficient Design of Time/Frequency Grid in HE-AAC Encoder

研究生：唐守宏

Student : Shou-Hung Tang

指導教授：劉啟民

Advisor : Chi-Min Liu

李文傑

Wen-Chieh Lee

國立交通大學

資訊科學與工程研究所



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

中華民國九十五年七月

HE-AAC 編碼器上有效率之 Time/Frequency Grid 設計

學生：唐守宏

指導教授：劉啓民 博士
李文傑 博士

國立交通大學資訊工程所碩士班

中文論文摘要

SBR (Spectral Band Replication) 編碼技術主要是一種頻寬擴展工具，並且和 AAC 結合，這種新的壓縮技術稱為 MPEG-4 High Efficiency (HE) AAC。在 HE-AAC 中，經由 SBR 負責高頻的訊號壓縮，AAC 就可以利用大部分的位元數來壓縮低頻的訊號。從訊號的相似性來看，SBR 的基本精神就是根據高低頻訊號之間的相似性，利用複製低頻訊號來重建高頻訊號。時頻格子 (Time/Frequency grid) 主要在決定重建訊號的單位，並且是 HE-AAC 中的核心模組。時頻格子主要包含了頻率表格的決定、時間區塊的決定，每個時間區塊的頻率解析度，以及每一個訊號框的格式。其中，頻率表格的選取以及時間區塊的大小決定了重建訊號的品質以及所花費的位元數；訊號框格式則是會影響訊號框中時間區塊的數目以及分佈型態。因此，時頻格子同時影響了品質以及所花費位元個數。

要成功的設計出一個有效率的時頻格子機制，必須要同時考慮到品質以及位元的影響。因此，在本篇論文中提出了一個衡量品質的標準，然後將時頻格子的決定轉化成格子狀的搜尋問題，最後，提出一個有效率的搜尋法來來找出最佳的解答。除此之外，也在搜尋法中加入所消耗位元數目的考量。在品質的量測上，主觀跟客觀的測試都會拿來驗證本篇論文所提出的設計，客觀的測試是採用 ITU 所發展的 PEAQ 測試系統來評估音訊壓縮後的誤差程度。

Efficient Design of Time/Frequency Grid in HE-AAC Encoder


Student: Shou-Hung Tang

Advisor: Dr. Chi-Min Liu

Dr. Wen-Chieh Lee

Institute of Computer Science and Information Engineering
National ChiaoTung University

ABSTRACT

The logo of National Chiao Tung University is a circular emblem. It features a central shield with a book and a graduation cap. The letters 'ES' are prominently displayed on the shield. Below the shield, the year '1996' is inscribed. The entire emblem is surrounded by a gear-like border.

Spectral Band Replication (SBR) has been combined with MPEG AAC as bandwidth extension tool. The resulting scheme is referred to as the MPEG-4 High Efficient (HE) AAC or AACplus. With SBR module taking care of the high frequency contents, the conventional AAC encoder can compress the low frequency part using most of the available bits. From the similarity between the low and high bands, SBR reconstructs the high bands by replicating the low bands. In the SBR, the time-frequency (T/F) grids deciding the replication unit in high frequency bands are the kernel module in SBR. Frequency table decision, frame class decision, time segments and associated frequency resolution decision are the main design issues in T/F grid. The chosen frequency table and number of time segments determine the quality and consumed bits of reconstructed audio. Frame class restricts the number of time segments and the distribution of time borders. Therefore, the design of T/F grid should be greatly involved with quality and consumed bits.

This thesis proposes an approach that formulates the decision of the T/F grid into a trellis-lattice search problem and presents an efficient search algorithm to find the optimum path. Both subjective and objective tests are conducted to check the quality improvement over existing methods. The objective test measures used is the recommendation system by ITU-R Task Group 10/4.

致謝

感謝劉啓民老師兩年來的栽培以及李文傑博士給予的指導，實驗室的楊宗翰、許瀚文學長，同學李侃峻、張家銘、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，讓我在專業知識以及研究方法獲得非常多的啟發。

最後，感謝我的父母與家人，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

| | |
|---|----|
| Contents | iv |
| Figure List..... | vi |
| Table List | ix |
| Chapter 1 Introduction | 1 |
| Chapter 2 Backgrounds..... | 6 |
| 2.1 SBR Decoder Overview..... | 6 |
| 2.2 HF Generator | 7 |
| 2.3 Envelope Adjuster..... | 7 |
| 2.3.1 Parameter Mapping..... | 8 |
| 2.3.2 Current Envelope Estimation..... | 8 |
| 2.3.3 Additional HF Components and Gain Calculation | 10 |
| Chapter 3 SBR Range Decision..... | 12 |
| 3.1 Adaptive SBR Range Adjustment..... | 13 |
| 3.1.1 SBR Header Overhead..... | 14 |
| 3.1.2 Tone Trembling..... | 15 |
| 3.1.3 Reduced Efficiency of DPCM..... | 15 |
| 3.1.4 Fluctuated Signal Bandwidth..... | 16 |
| 3.2 Error Concealment | 17 |
| Chapter 4 Related Work for Time/Frequency Grid..... | 18 |
| 4.1 Time/Frequency Grid Design..... | 18 |
| 4.1.1 Transient Detector..... | 19 |
| 4.1.2 Frame Splitter..... | 20 |
| 4.1.3 T/F Grid Generator | 20 |
| 4.2 Summary | 21 |
| Chapter 5 Efficient Design of Time/Frequency Grid..... | 22 |
| 5.1 Analysis of Reconstructed Error | 22 |
| 5.2 Frequency Band Table Decision | 25 |
| 5.3 Time Borders and Envelope Resolution | 26 |
| 5.4 Frame Class Decision | 31 |
| Chapter 6 Artifacts in SBR | 32 |
| 6.1 Tone Trembling..... | 32 |
| 6.2 Tone Shift..... | 33 |
| 6.3 Sawtooth | 34 |
| 6.4 Noise Floor Overflow | 35 |

Chapter 7 Experiments.....38

 7.1 Measurement Tools Description38

 7.2 Objective Quality Measurement in MPEG Test Tracks38

 7.3 Objective Quality Measurement in Music database45

 7.4 Subjective Quality Measurement55

 7.5 Objective Quality Measurement with Existing Codecs56

 7.6 Objective Quality Measurement by SBR range with Error Concealment58

Chapter 8 Conclusion and Future Works63

References.....64



Figure List

| | |
|--|----|
| Figure 1: Components of HE-AAC and HE-AAC version 2..... | 1 |
| Figure 2: The basic principle of SBR for reconstructions. | 2 |
| Figure 3: bit-rate quality comparison among AAC, HE-AAC, and HE-AAC version 2..... | 2 |
| Figure 4: Basic architecture of HE-AAC encoder. | 3 |
| Figure 5: Block diagram of HE-AAC encoder. | 4 |
| Figure 6: Block diagram of HE-AAC decoder [3]..... | 7 |
| Figure 7: Extracted parameters from bitstream mapping in HE-AAC decoder. | 8 |
| Figure 8: Envelope estimation by interpolation mode and relating adjustment..... | 9 |
| Figure 9: Envelope estimation by non-interpolation mode and relating adjustment..... | 10 |
| Figure 10: Spectral valley from inappropriate range decision..... | 12 |
| Figure 11: Distortion for SBR range due to bad quality of AAC part. | 13 |
| Figure 12: The spectral valley revising due to adaptive range adjustment.. | 14 |
| Figure 13: The parameters include in SBR header bitstream. | 15 |
| Figure 14: The example for the characteristic of attack signal,..... | 16 |
| Figure 15 Time-direction dpcm disability..... | 16 |
| Figure 16: The range constraint of SBR [3]..... | 16 |
| Figure 17: Block diagram of 3GPP HE-AAC encoder [15]. | 18 |
| Figure 18: The illustration of detection mechanism in transient detector.... | 19 |
| Figure 19: An example for T/F grid generator [15]. | 21 |
| Figure 20: Illustration of DSR notation. | 27 |
| Figure 21: The trellis-lattice deducing path by dynamic programming..... | 27 |
| Figure 22: DP flow chart with quality constraint..... | 29 |
| Figure 23: DP flow chart with both quality and efficiency constraint. | 30 |
| Figure 24: An example for variable frame border..... | 31 |
| Figure 25: Tone trembling effect..... | 32 |
| Figure 26: An example for characteristics of tone-rich signals | 33 |
| Figure 27: Tone shift effect. | 33 |
| Figure 28: The envelope comparison between high bands and low bands. The envelope of high bands is flat, and the envelope of low bands is | |

| | |
|--|----|
| sharp. | 34 |
| Figure 29: Sawtooth effect due to the limiter gain mechanism. | 35 |
| Figure 30: Noise floor overflow due to failure of detecting tones in high bands. | 35 |
| Figure 31: Noise floor overflows due interpolation mode. The target circle indicates the noise floor overflow is from the averaged energy with tone component. | 36 |
| Figure 32: A comparison to Figure 31. It incident the result without tone addition mechanism. | 37 |
| Figure 33: The ODG variance comparison of Table 4. | 41 |
| Figure 34: The ODG variance comparison of Table 5. | 42 |
| Figure 35: The ODG variance comparison of Table 6. | 43 |
| Figure 36: The ODG variance comparison of Table 7. | 44 |
| Figure 37: The ODG variance comparison of Table 8. | 45 |
| Figure 38: ODG-bit rate curve comparison among different T/F grid design. | 45 |
| Figure 39: The average ODG of three coding methods in PSPLAB audio database at bit rate 80kbps and sampling rate 44100 Hz. M1 is uniform 1 cut in T/F grid and M2 is uniform 7 cuts in T/F grid. M3 is our design. | 47 |
| Figure 40: The average ODG of three coding methods in PSPLAB audio database at bit rate 64kbps and sampling rate 44100 Hz. M1 is uniform 1 cut in T/F grid and M2 is uniform 7 cuts in T/F grid. M3 is our design. | 47 |
| Figure 41: The average ODG of three coding methods in PSPLAB audio database at bit rate 48kbps and sampling rate 44100 Hz. M1 is uniform 1 cut in T/F grid and M2 is uniform 7 cuts in T/F grid. M3 is our design. | 48 |
| Figure 42: The example for signal whose high band envelope alters rapidly. | 49 |
| Figure 43: The example for signal which has a sharp high band envelope. | 50 |
| Figure 44: The error of noise floor due to tone addition. | 50 |
| Figure 45: The spectrum of “sweep” which has continuous tone from 10 KHz to 22 KHz. | 51 |
| Figure 46: The reconstructed spectrum of Figure 45 through uniform 7 cuts in T/F grid. | 51 |
| Figure 47: The reconstructed spectrum of Figure 45 through DP T/F grid design. | 52 |

Figure 48: The reconstructed spectrum of Figure 45 through aacPlus.52

Figure 49: The spectrum of “sin_300_625_1k_5k_10k_15K_20k_m20db”
which has interrupted tone from low frequency bands to high
frequency bands53

Figure 50: The reconstructed spectrum of Figure 49 through DP T/F gird
design.53

Figure 51: The reconstructed spectrum of Figure 49 through aacPlus.54

Figure 52: The frequency envelope of “sin_9kind_valious”.54

Figure 53: Noise floor overflow due to interpolation mode.55

Figure 54: The result of subjective test at 80 kbps.55

Figure 55: The ODG-bit rate comparison curve among different codecs....58

Figure 56: The result of objective quality measurement for error
concealment based on MPEG test tracks at bit rate 80 kbps.59

Figure 57: The result of objective quality measurement for error
concealment based on MPEG test tracks at bit rate 64 kbps.59

Figure 58: The result of objective quality measurement for error
concealment based on MPEG test tracks at bit rate 48 kbps.60

Figure 59: The ODG and bit rate comparison curve for NCTU HE-AAC
with and without error concealment.61

Figure 60: The frequency envelope of “sc01” by NCTU HE-AAC without
error concealment.....61

Figure 61: The frequency envelope of “sc01” by NCTU HE-AAC with error
concealment. The blue line represents the original signal, and the red
one is coded signal.62

Table List

| | |
|--|----|
| Table 1: The combination of transient and trailing frame borders..... | 20 |
| Table 2: Combinations of bit-consuming stages..... | 29 |
| Table 3: The twelve tracks recommended by MPEG | 39 |
| Table 4: Objective measurements through ODGs for different T/F grid design in HE-AAC at 112 kbps..... | 40 |
| Table 5: Objective measurements through ODGs for different T/F grid design in HE-AAC at 96 kbps. | 41 |
| Table 6: Objective measurements through ODGs for different T/F grid design in HE-AAC at 80 kbps. | 42 |
| Table 7: Objective measurements through ODGs for different T/F grid design in HE-AAC at 64 kbps. | 43 |
| Table 8: Objective measurements through ODGs for different T/F grid design in HE-AAC at 48 kbps. | 44 |
| Table 9: The PSPLAB audio database. | 46 |
| Table 10: The objective quality measurement among different codecs at bit rate 80 kbps. | 57 |
| Table 11: The objective quality measurement among different codecs at bit rate 64 kbps. | 57 |
| Table 12: The objective quality measurement among different codecs at bit rate 48 kbps. | 58 |

Chapter 1

Introduction

Perceptual audio codecs (MPEG I Layer3 [1] or MPEG-II AAC [2]) which exploit the properties of human psychoacoustic model not only reduce the transmitted audio data through eliminating the unheard frequencies and tones, but also provide “CD-quality” or “transparent” audio quality (indistinguishable source from encoded one) at a bit rate of 128kbps. Below 128kbps, the perceived audio quality of most codecs collapses rapidly. Due to insufficient bits, the codecs usually have two possible solutions that are the audio bandwidth limited or keeping the complete bandwidth but introducing annoying coding artifacts. The first solution makes the audio dull and the other results in unacceptable artifacts.

SBR (Spectral Band Replication) [3][4][5][6][7] which is standardized in ISO/IEC 14496-3:2001/Amd.1:2003 is a new audio coding enhancement tool and a significant breakthrough in the area of audio coding for low bit rates. SBR is a bandwidth extension tool used in combination with conventional audio codecs, e.g. AAC (called aacPlus or HE-AAC), and MP3 (called mp3PRO [8]). **Figure 1** demonstrates the components of HE-AAC and HE-AAC version 2. In addition to AAC and SBR codec, HE-AAC version 2 extra contains PS (Parameter Stereo) coding [9][10].

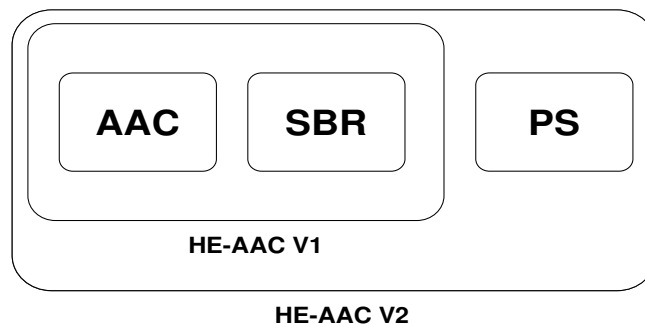


Figure 1: Components of HE-AAC and HE-AAC version 2.

The basic principle of SBR is to reconstruct the high frequency bands by replicating the low frequency bands and adjusting the replicated bands perceptually similar to original ones. The reconstruction procedure of SBR is illustrated in Figure 2. The replication and adjustment is achieved by a small amount of control parameters. With SBR taking care of the high frequency contents, the underlying perceptual audio

encoder can compress the low frequency part with most of the available bits. Hence, SBR not only can increase the audio bandwidth, but improve the quality of underlying codec at low bit rates. The bit rate-quality comparison among AAC, HE-AAC version 1 and HE-AAC version 2 is illustrated in Figure 3. AAC has good performance at the bit rate over 96K, HE-AAC (SBR) targets at the range from 48K to 80K, and HE-AAC version 2 (PS) is responsible for the bit rate lower than 48K.

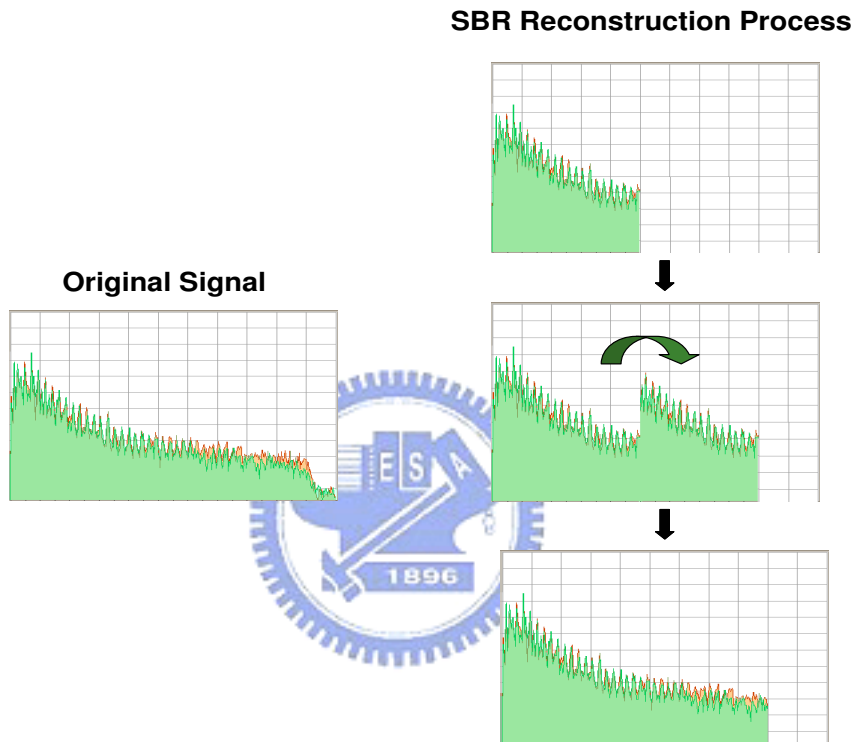


Figure 2: The basic principle of SBR for reconstructions.

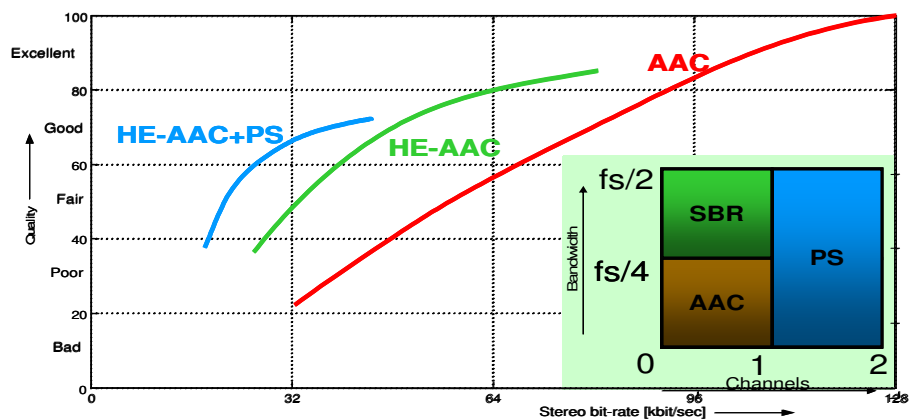


Figure 3: bit-rate quality comparison among AAC, HE-AAC, and HE-AAC version 2.

SBR is an advanced scheme to compress high frequency contents efficiently at very low bit rate, commonly about 1K~3K bps per channel [5]. Under bit rate constraint, the existing audio codecs sacrifice the high frequency component of

signals to obtain good perceptual quality. However, when the audio bandwidth is getting lower, the hearing perception becomes duller. SBR is responsible for retaining the signal bandwidth at low bit rate. Through SBR taking care of the high frequency parts of the audio signals with small amount of bits, the conventional encoder only needs to handle the low frequency parts. Consequently, the signal fed into underlying encoder half of the original sample rate is enough based on Nyquist's theorem. Inherently, HE-AAC is a dual rate system, where the AAC encoder is operated at half the sampling rate of SBR encoder. The basic architecture of HE-AAC encoder is depicted in Figure 4. The audio signal is fed into the 64-bands Analysis Filter. After the Analysis Filter, there are two branches where one is SBR encoder and the other is 32-bands Synthesis Filter. Through 64-bands Analysis Filter and 32-bands Synthesis Filter, the signal fed into AAC encoder is half the sampling rate of the original signal.

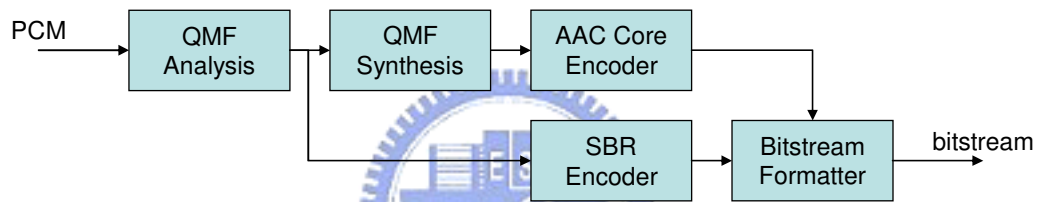


Figure 4: Basic architecture of HE-AAC encoder.

In the SBR encoder, the control parameters is estimated to ensure that the reconstructed high frequency bands are as similar as original ones. These parameters mostly for spectral envelope representation are used to rescale the spectral envelope and control the tonal-to-noise ratio of high frequency bands. Time-frequency (T/F) grid and High frequency (HF) generator are the two main modules in SBR codec. The former decides the reconstruction unit in time and frequency domains for rescaling the replicated envelopes to original ones, and the latter keeps the similarity of noise-to-tone ratio between replicated contents and original ones. The resolutions of reconstruction units dominate the accuracy of the reconstructed contents and required bits. The higher is resolution of these units; the better is the accuracy of reconstruction, but taking more bits. Hence, T/F grid plays a key role to determine the resolutions of reconstruction units and dominates the resulting audio quality of whole HE-AAC.

The block diagram of SBR encoder is illustrated in Figure 5. In SBR encoder, the SBR range is determined in the first. Frequency table decision is responsible for choosing the most suitable frequency resolution from eight different tables. Tonality calculation estimates the tonality of original signal. T/F grid determines the number of time envelopes, time borders, and envelope resolutions. According to the information of grid unit and tonality, Inverse Filter tries to decrease the tonality of the replicated band contents according to demands. In order to maintain the similar tone-to-noise

ratio, additional tone and noise is established by Tone/Noise Addition module. Finally, through quantization and delta coding eliminating redundancy between encoded data, the information from SBR encoder and AAC encoder is combined into bitstream packet. The Coupling module determines that stereo or coupled to mono compressing in SBR encoder. The Bit Reservoir module plays the role of allotting bits among SBR encoder and AAC encoder. The shadowed modules are the main topics in this thesis, including SBR range decision, frequency table decision and Time/Frequency grid.

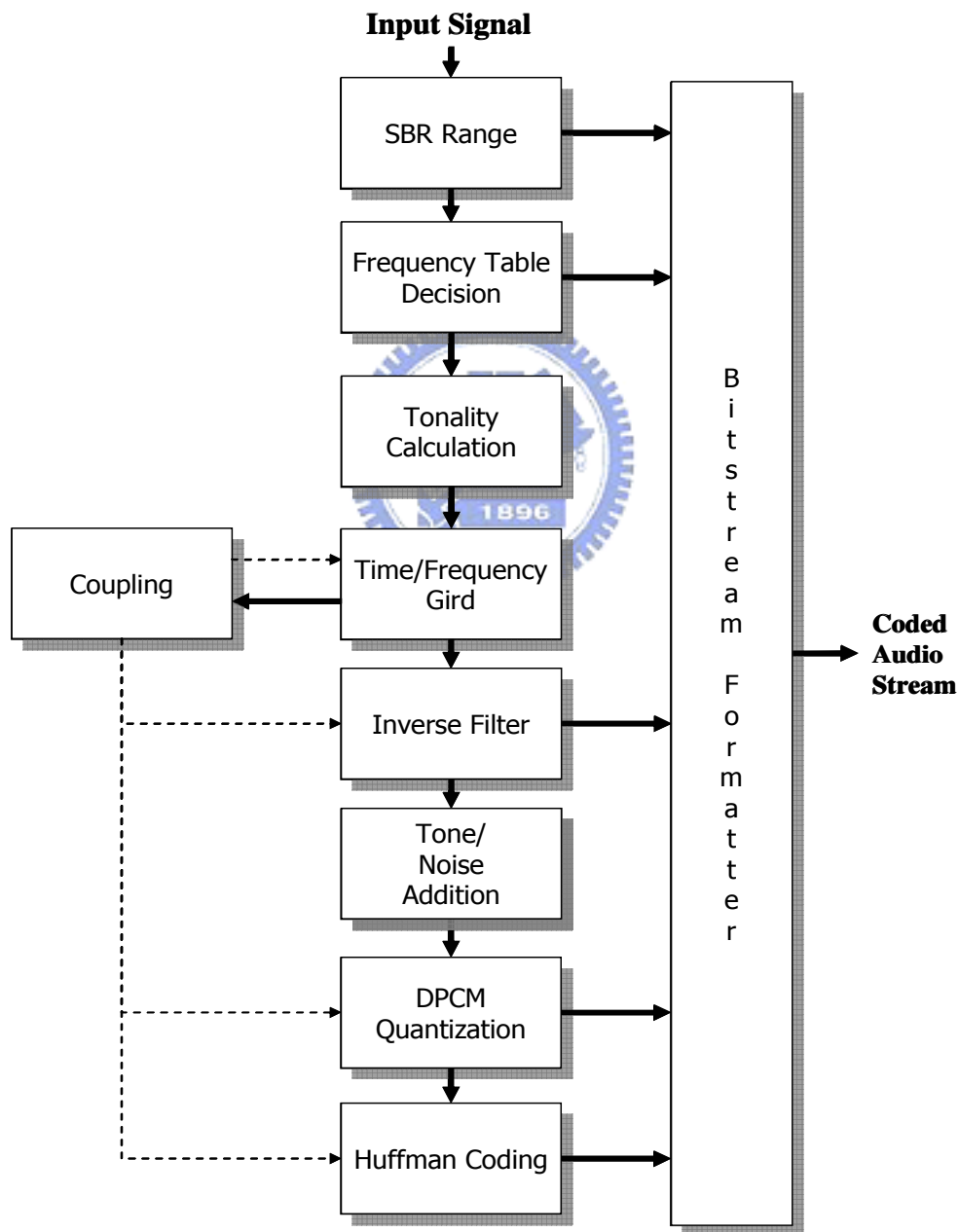


Figure 5: Block diagram of HE-AAC encoder.

Since HE-AAC comprises SBR and AAC, the cooperation among them is very important. The reconstructed high frequency bands by SBR depend on replicated low

frequency bands, and therefore, the quality of AAC encoder has significant effect on HE-AAC. The SBR range and associated AAC range decision is the first important design issue in HE-AAC encoder. With inappropriate allocation for AAC and SBR range, it may bring the artifact “spectral valley” around range boundary or reduce the quality of HE-AAC encoder. The range decision is involved with contents of signal, compressing bit rate and sampling rate. In existing SBR encoder, including 3GPP [11], Coding Technology [12] and Nero [13], the SBR range is determined only by bit rate and sampling rate. This thesis proposes a method to determine the most appropriate SBR range taking account of above factors.

The resolution of reconstruction unit used in SBR process is determined by Time/Frequency grid module. This module can be incised into three parts, which are frequency table decision, time borders distribution and associated envelope frequency resolution decision and frame class decision. Frequency tables describe the approximate resolution of reconstruction unit in frequency domain, and time borders revolve the resolution in time domain. Envelope resolution define the detailed frequency resolution each frame. There are four different SBR frame classes, FIXFIX, FIXVAR, VARFIX, and VARVAR used, each of which has different capabilities to describe the distribution of time borders. Appropriate frame classes selection can increase the coding efficiency. Bit rate, content of signal greatly affect T/F Grid decision. In 3GPP SBR encoder, it introduces a transient detector to detect the start position of transients and labeling time borders. It only considers energy difference of neighbor samples without the correlation between replicated samples and original ones. This thesis formulates the decision of the T/F grid into a trellis-lattice search problem and proposes an efficient search algorithm to find the optimum solution.

This thesis is organized as follows. Chapter 2 introduces the fundamental knowledge of HE-AAC. Chapter 3 introduces the design of SBR range cooperating with AAC encoder. In Chapter 4, the existing T/F grids designs are described. Chapter 5 presents an efficient design of T/F grids through dynamic programming. In Chapter 6, extensive experiments are made to prove the improvement of the proposed T/F grid design. Both subjective and objective measurements are conducted to verify the quality and efficiency of our T/F grids in Chapter 5. Chapter 7 gives a conclusion and future work on this thesis.

Chapter 2

Backgrounds

2.1 SBR Decoder Overview

The block diagram of HE-AAC SBR decoder [3] is illustrated in Figure 6. It shows the relationship between AAC decoder and SBR enhancement parts. First, the bitstream payload deformatter divides the bitstream payload into AAC parts and SBR parts. The AAC bitstream part is fed to the AAC decoder, and the SBR bitstream part is fed to the bitstream parser. After the parser, de-quantization follows and the raw data is Huffman decoded. Through an analysis QMF bank, the low frequency part of signal from AAC decoder separated into 32 subbands is fed into HF Generator which is responsible for deriving the high frequency part according to SBR data and the low frequency part of signal. Envelope Adjuster is guided by the SBR data extracted from the bitstream to adjust the reconstructed components as similar as original ones. Finally, the low frequency parts and the high frequency parts are synthesized by a synthesis QMF bank.

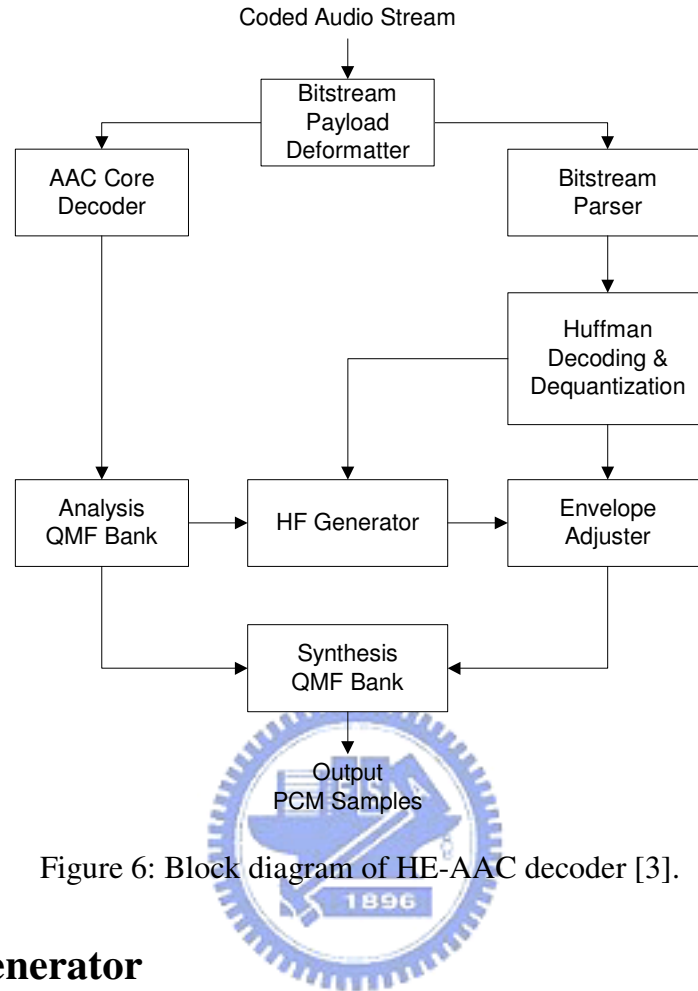


Figure 6: Block diagram of HE-AAC decoder [3].

2.2 HF Generator

In the HF generator, the goal is to copy or patch a number of low frequency subband signals obtained from the analysis filter bank to consecutive high frequency subband signals. The patching determines the corresponding relation between high bands and replicated low bands. In addition, in order to remove the unwanted tone components, the inverse filtering is done in this module. Hence, the output of HF generator is the corresponding subband signal for reconstructing original high frequency subband.

2.3 Envelope Adjuster

The objective of envelope adjuster is to adjust the reconstructed envelop as similar as original one. The envelope adjustment is accomplished according to the parameters extracted from bitstream. With the original high band energies and additional components for each reconstruction unit from bitstream, the corresponding gain values can be derived by estimating the current envelope. The process of calculating gain values and signal reconstruction is described below.

2.3.1 Parameter Mapping

Some of the parameters extracted from the bitstream are vectors or matrices. Out of necessity, this grouped data is mapped to the highest available frequency resolution for the envelope adjustment. This means that the adjacent subbands in the grouped data will have the same value. However, the mapping is only in the frequency domain, and time resolution will be preserved. Figure 7 shows the mapping of envelope scalefactor.

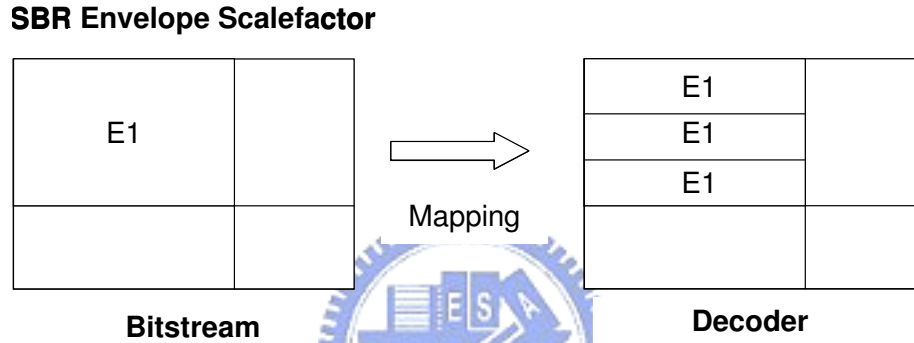


Figure 7: Extracted parameters from bitstream mapping in HE-AAC decoder.

2.3.2 Current Envelope Estimation

In order to adjust the envelope of the present SBR frame, the envelope of the current SBR signal needs to be estimated. There are two different estimation mode used in SBR codec, interpolation and non-interpolation. With interpolation, the estimated current envelop E' is given by

$$E'(m, l) = \frac{\sum_{i=2t_E(l)}^{2t_E(l+1)-1} |X_h(m + k_x, i)|^2}{2 \cdot (t_E(l-1) - t_E(l))}, 0 \leq m \leq M - 1, 0 \leq l \leq L_E - 1 \quad (1)$$

The notations are illustrated below

m : QMF subband index

l : Time envelope index

k_x : the first QMF subband in the SBR range. (SBR start boundary)

t_E : contains the time borders for all SBR envelopes in the current SBR frame.

M : The total amount of QMF subband in SBR range.

L_E : Number of SBR envelopes.

If non-interpolation is used, the energies are averaged over every frequency band. The estimation is

$$E'(m, l) = \text{Mapping}(E'(k_h - k_l, l)),$$

$$E'(k_h - k_l, l) = \frac{\sum_{i=2^{t_E(l)}}^{2^{t_E(l+1)}-1} \sum_{j=k_l}^{k_h} |X_h(j, i)|^2}{2 \cdot (t_E(l+1) - t_E(l)) \cdot (k_h - k_l - 1)}, k_l \leq m \leq k_h, \quad (2)$$

$$\begin{cases} k_l = F(p, r(l)) \\ k_h = F(p+1, r(l)) - 1 \end{cases}, 0 \leq p \leq n(r(l)) - 1$$

$r(l)$: Envelope resolution

$n(r(l))$: Number of frequency band

$F(\cdot)$: Frequency band table

The difference between interpolation mode and non-interpolation mode is illustrated in Figure 8 and Figure 9. In interpolation mode, the energies are averaged over every QMF filter band, and each QMF subband derives respective gain value. All the QMF subbands in one frequency band will be adjusted to the same energy. In non-interpolation mode, the energies are averaged over every frequency band. All the QMF subbands in one frequency band use the same gain value, and the envelope of replicated signal will be maintained.

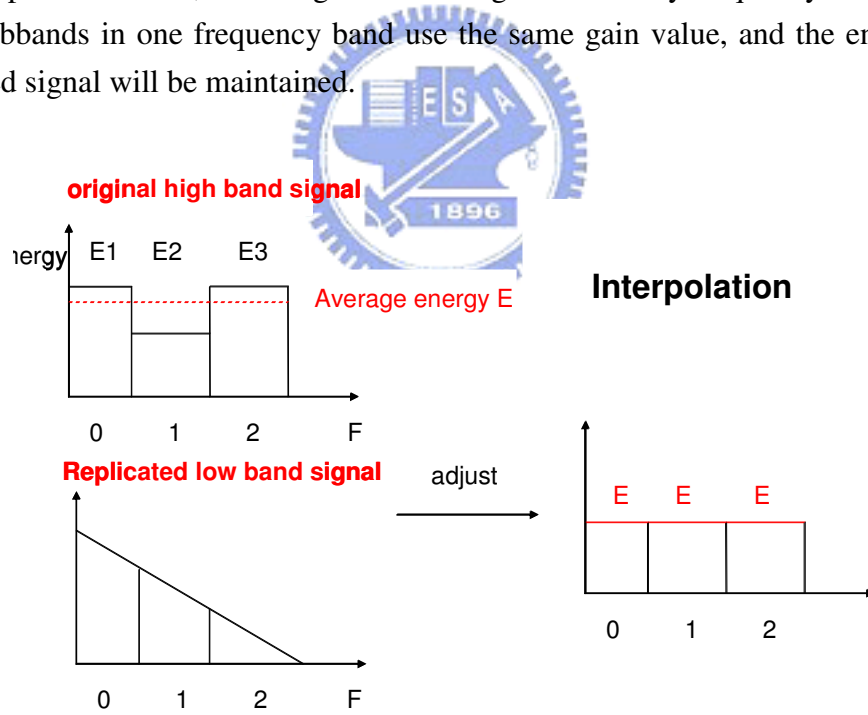


Figure 8: Envelope estimation by interpolation mode and relating adjustment

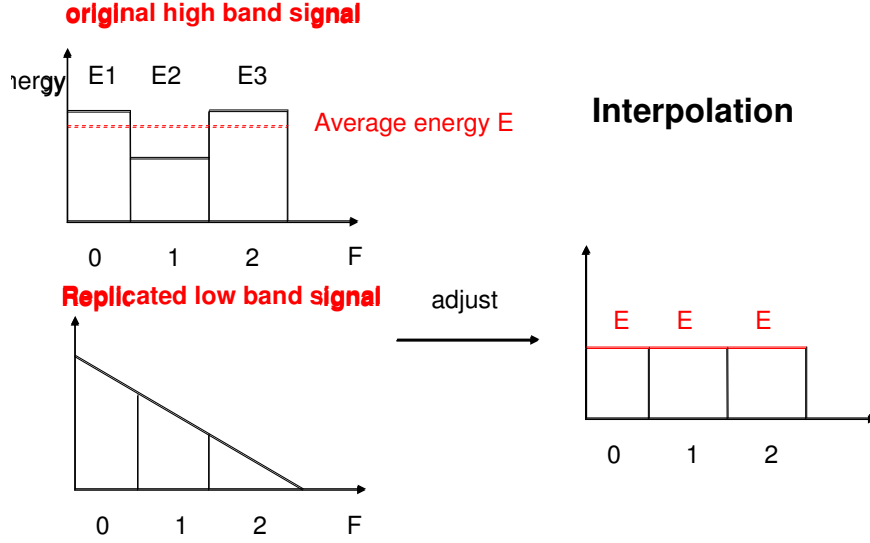


Figure 9: Envelope estimation by non-interpolation mode and relating adjustment

2.3.3 Additional HF Components and Gain Calculation

The noise floor scalefactor is a ratio, and in order to add the correct amount of noise, it needs to be converted to a proper amplitude value, according to the following.

$$Q(m,l) = \sqrt{E(m,l) \frac{Q(m,l)}{1+Q(m,l)}}, 0 \leq m \leq M-1, 0 \leq l \leq L_E-1 \quad (3)$$

The level of sinusoids are derived as

$$S(m,l) = \sqrt{E(m,l) \frac{S(m,l)}{1+Q(m,l)}} \quad (4)$$

And the gain value are derived by

$$G(m,l) = \begin{cases} \sqrt{\frac{E(m,l)}{E'(m,l) \cdot (1+Q(m,l))}}, & \text{if } S(m,l) = 0 \\ \sqrt{\frac{E(m,l)}{E'(m,l)} \cdot \frac{Q(m,l)}{1+Q(m,l)}}, & \text{if } S(m,l) \neq 0 \end{cases} \quad (5)$$

In order to avoid unwanted noise substitution, the gain values are limited according to the limiter gain mechanism:

$$\begin{aligned}
G_{Max}(m,l) &= Mapping(G_{Max}(k,l)) \\
&= \min \left(\sqrt{\frac{\sum_{i=f_{lim}(k)}^{f_{lim}(k+1)-1} E(i,l)}{f_{lim}(k+1)-1}}{\sum_{i=f_{lim}(k)}^{f_{lim}(k+1)-1} E'(i,l)}} \cdot \text{limGain}(bs_limiter_gains), 10^5 \right), \quad (6) \\
0 \leq k \leq N_L - 1, f_{lim}(k) \leq m \leq f_{lim}(k+1)
\end{aligned}$$

where $\text{limGain} = [0.70795, 1.0, 1.41254, 10^{10}]$, and f_{lim} presents the limiter frequency band. Hence, the gain values are limited according to

$$G_{lim}(m,l) = \min(G(m,l), G_{Max}(m,l)) \quad (7)$$

The additional noise component needs to be revised by

$$Q_{lim}(m,l) = Q(m,l) \cdot \frac{G_{lim}(m,l)}{G(m,l)} \quad (8)$$

Due to the limitation, the total energy for a limiter band will have loss, and it is compensated by

$$\begin{aligned}
G_C(m,l) &= Mapping(G_C(k,l)) \\
&= \min \left(\left\{ \begin{array}{l} \sqrt{\frac{\sum_{i=f_{lim}(k)}^{f_{lim}(k+1)-1} E(i,l)}{f_{lim}(k+1)-1}}, \text{if } S(m,l) = 0 \\ \sqrt{\sum_{i=f_{lim}(k)}^{f_{lim}(k+1)-1} (E(i,l) * G_{lim}^2 + Q_{lim}^2(i,l))} \\ \sqrt{\frac{\sum_{i=f_{lim}(k)}^{f_{lim}(k+1)-1} E(i,l)}{f_{lim}(k+1)-1}}, \text{if } S(m,l) \neq 0 \\ \sqrt{\sum_{i=f_{lim}(k)}^{f_{lim}(k+1)-1} (E(i,l) * G_{lim}^2 + S^2(i,l))} \end{array} \right\}, 1.584893192 \right) \quad (9)
\end{aligned}$$

Finally, the resulting gain values and additional components are

$$\begin{aligned}
G_{final}(m,l) &= G_{lim}(m,l) \cdot G_C(m,l), \\
Q_{final}(m,l) &= Q_{lim}(m,l) \cdot G_C(m,l), \\
S_{final}(m,l) &= S(m,l) \cdot G_C(m,l)
\end{aligned} \quad (10)$$

HF Signal Assembling

Before the gain values are applied to the subband samples, there is a filter to smooth the gain values. The smooth filter is applied according to the parameters extracted from bitstream. Finally, the subband samples are adjusted by these gain values, and the additional noise and tone components are added.

Chapter 3

SBR Range Decision

SBR range decision here is to decide a boundary in frequency domain. This boundary separates QMF data into two portions. Frequency bands lower than the boundary is to be coded by AAC encoder, and frequency bands higher than the boundary is to be coded by SBR encoder. SBR replicates the subband signals encoded by AAC to reconstruct high frequency subband signals. Therefore, the quality of SBR encoder greatly relies on AAC encoder. In other words, the objective of SBR range decision is to ensure that the low frequency parts encoded by AAC can have a satisfactory quality. The range decision allots the burdens between SBR encoder and AAC encoder, which is a decided module for HE-AAC quality. With inappropriate range disposing, it may bring perceptible artifacts therefore reduce the audio quality. While selected SBR start frequency is too high, available bits may not be enough for AAC to encode the assigned bandwidth. HE-AAC may produce “spectral valley” around range boundary due to insufficient bits. Figure 10 illustrates an example of resultant spectrum with spectral valley. The quality of low frequency components also influences the high frequency components reconstructed by SBR. Figure 11 shows an example of distortion in high frequency due to bad quality of low frequency component. Oppositely, if bits are sufficient and the bandwidth assigned to AAC is too narrow, since the maximum bandwidth of SBR is restricted, the overall bandwidth of HE-AAC is getting lower and decreases the audio quality. Obviously, bit rates and sampling rates are the main factors to this range decision.

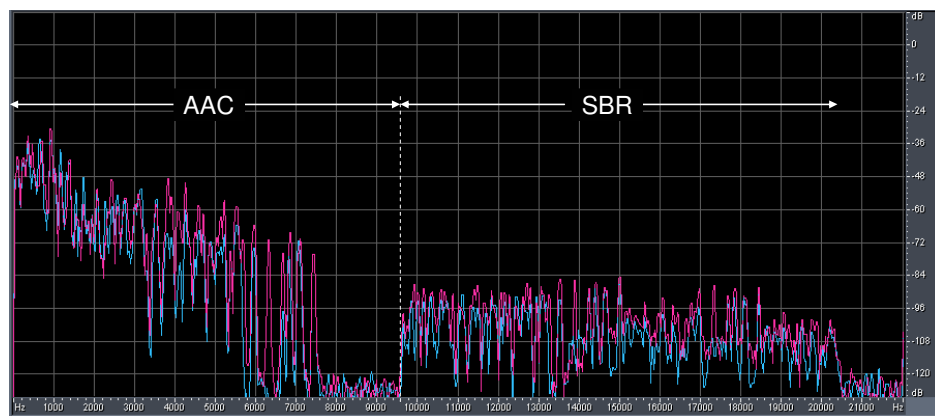


Figure 10: Spectral valley from inappropriate range decision.

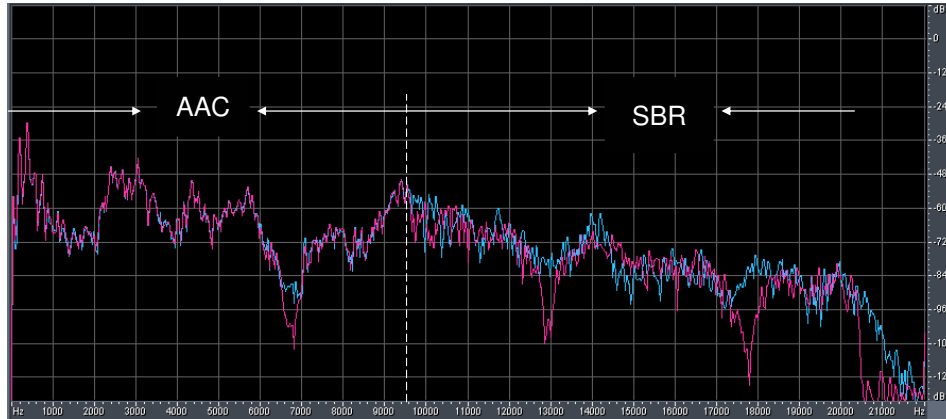


Figure 11: Distortion for SBR range due to bad quality of AAC part.

Another significant factor to SBR range decision is the contents of signal. The required bits of AAC encoder are related to the audio content. Thus, whether or not the available bits are sufficient is related to signal contents. Even at high bit rates, the spectral valley still occurs in the trailing of AAC parts because the relating audio content is hard to be encoded. On the contrary, at the low bit rates, the bandwidth of AAC is not necessary to be cut off.

Therefore, SBR range decision is greatly involved with bit rates and audio content. The consideration of audio content is aggressive and active, and the consideration of bit rates is steady and conservative. According to the two factors, this thesis proposes two possible approaches, which are adaptive SBR range and error concealment.

3.1 Adaptive SBR Range Adjustment

The required bits for each frame in AAC encoder are different due to the contents of signal. Consequently, the most flexible method is adaptively adjusting SBR range according to condition of AAC encoder. Through detecting the zero bands in AAC bit allocation, SBR range can be determined adaptively frame by frame. The adaptive method not only eliminates the spectral valley artifact, but achieves good interconnection between AAC encoder and SBR encoder. Giving Figure 6 for an example, there are zero bands in the trailing of AAC parts. By detecting these zero bands, the SBR range can move ahead to avoid the spectral valley. Figure 12 shows the result.

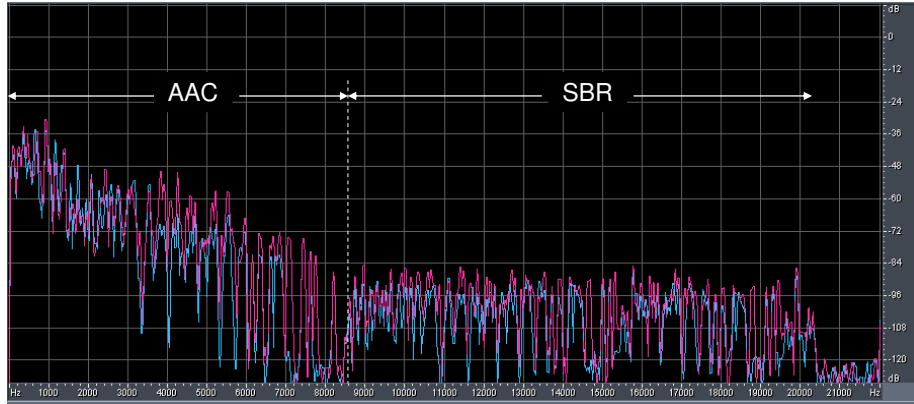


Figure 12: The spectral valley revising due to adaptive range adjustment.

However, this method may face four shortcomings: the SBR header overhead, tone trembling artifact, reduced efficiency for the DPCM and the fluctuated bandwidth.

3.1.1 SBR Header Overhead

SBR encoder uses several different frequency band tables, which are master band frequency table, high resolution frequency band table, low resolution frequency band table, noise floor frequency band table and limiter frequency band table. The parameters in SBR bitstream header are needed to define all frequency band tables, SBR start boundary and stop boundary. If the bitstream parameters used for this frame are the same as the last one, then the bitstream header needs not to be transmitted again. On the contrary, a transmission of the header is only needed when the parameters differ from the last ones. Therefore, adaptively revising SBR range needs to consume bits for transmitting new bitstream header. The syntax of SBR header is illustrated in Figure 13. In SBR header, the two parameters **bs_start_freq** and **bs_stop_freq** define the SBR range. According to Figure 13, the overhead for varying **bs_start_freq** and **bs_stop_freq** is 16bits. Therefore, altering SBR range each time takes 16 bits. This header overhead may be serious when SBR range changes often.

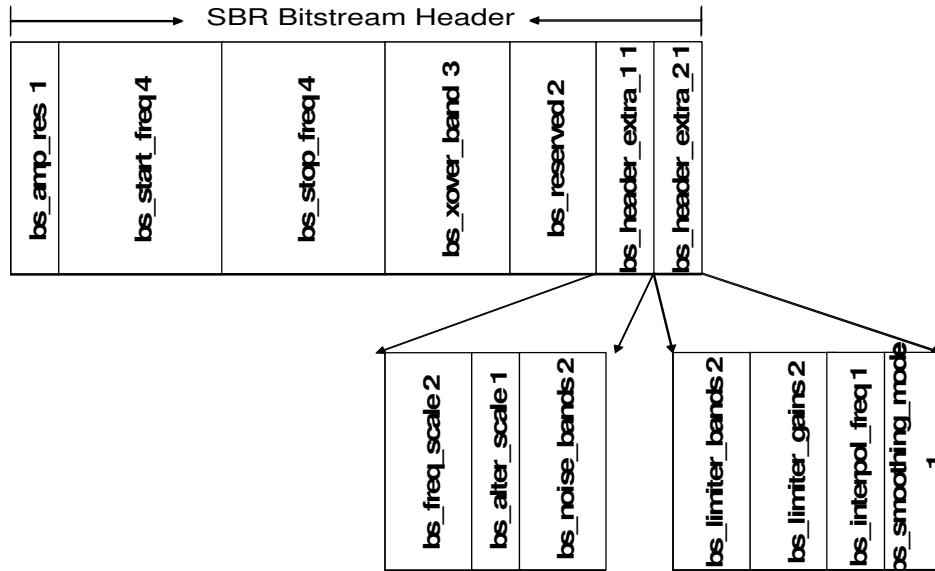


Figure 13: The parameters include in SBR header bitstream.

3.1.2 Tone Trembling

Tone trembling is an artifact which greatly decreases the audio quality in the perceptual hearing. The characterization of this artifact will be described in Chapter 6.

3.1.3 Reduced Efficiency of DPCM

A single sinusoid in the frequency domain transforms to a stable signal in the time domain, and on the contrary, a pulse in the time domain corresponds to a constant in the frequency domain. Figure 14 describes the above property of audio signal. On a word, most signal is usually stable in either time or frequency domain. Therefore, using delta coding in one of these two domains according to the signal characteristics can increase the coding efficiency. In Figure 15, since the SBR range of this frame differs from the last one, the number of subband included in SBR range is different between two consecutive frames. It disables time-direction DPCM for the first envelope of this frame. Consequently, the incomplete DPCM takes more bits and decreases the coding efficiency.

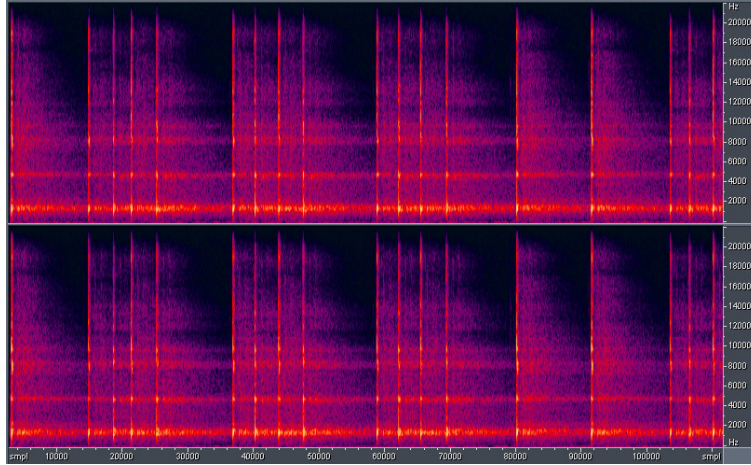


Figure 14: The example for the characteristic of attack signal,

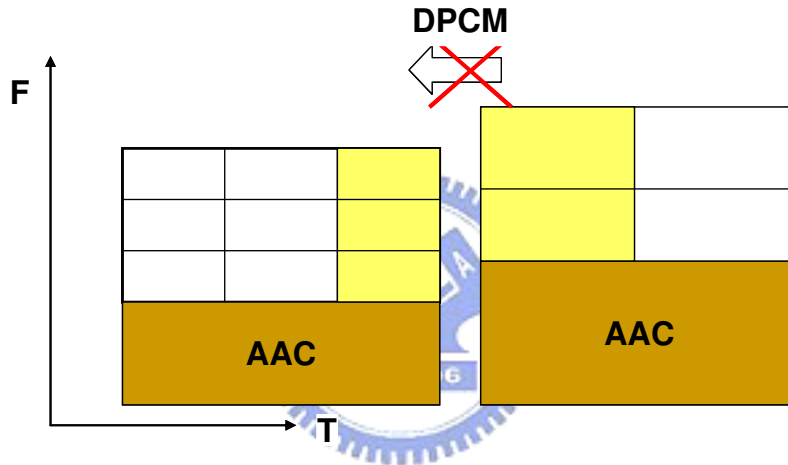


Figure 15 Time-direction dpcm disability

3.1.4 Fluctuated Signal Bandwidth

The subband number of SBR range is restricted by standard in Figure 16. According to different sampling rate, the different maximum ranges are defined. Since the SBR start boundary is adjusted adaptively, in order to observe this restriction, the stop boundary may need to be moved ahead. Consequently, the bandwidth of signal may vary with frames. In addition, the changing of SBR stop boundary may also cause tone trembling artifact. Regardless of tone trembling, the fluctuated bandwidth makes signal intermittent and reduce the audio quality in perceptual hearing.

$$k_2 - k_0 \leq \begin{cases} 48 & , F_{SBR} \leq 32kHz \\ 35 & , F_{SBR} = 44.1kHz \\ 32 & , F_{SBR} \geq 48kHz \end{cases}$$

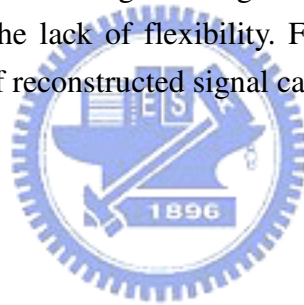
Figure 16: The range constraint of SBR [3].

Summary

Integrating the above discussion, the adaptive SBR range method will face many difficulties. The most serious problem is the bit overhead. In order to record the change of range, more bits are used for transmitting bit stream. Hence, this method has the highest flexibility but loses the coding efficiency. However, SBR is used for low bit rate coding. Instead of high coding precision, SBR turns to economize required bits for acceptable reconstructed signal. Therefore, at low bit rates, the adaptive range approach seems unsuitable.

3.2 Error Concealment

This method determines SBR range depending on bit rates and sampling rates. The SBR range is fixed among frames. According to bit rates, the burden allocation between AAC encoder and SBR encoder is fine in most frames. The occasional spectral distortion is handled by error concealment mechanism [14][15][16][17] in SBR decoder. This approach leads high coding efficiency, and error concealment mechanism can compensate the lack of flexibility. Further, due to the help of error concealment, the bandwidth of reconstructed signal can be aggressively extended.



Chapter 4

Related Work for Time/Frequency Grid

This chapter introduces the design of T/F grid in 3GPP HE-AAC encoder [15]. The block diagram of 3GPP HE-AAC encoder is illustrated in Figure 17. T/F grid decision contains transient detector, frame splitter, and T/F grid generator. Transient detector detects the start position of transient. Frame splitter is only operated in the frame without transient, and it determines this frame separated into two envelopes or not. T/F grid generator receives the information from transient detector and determines the time borders and the related envelope resolution.

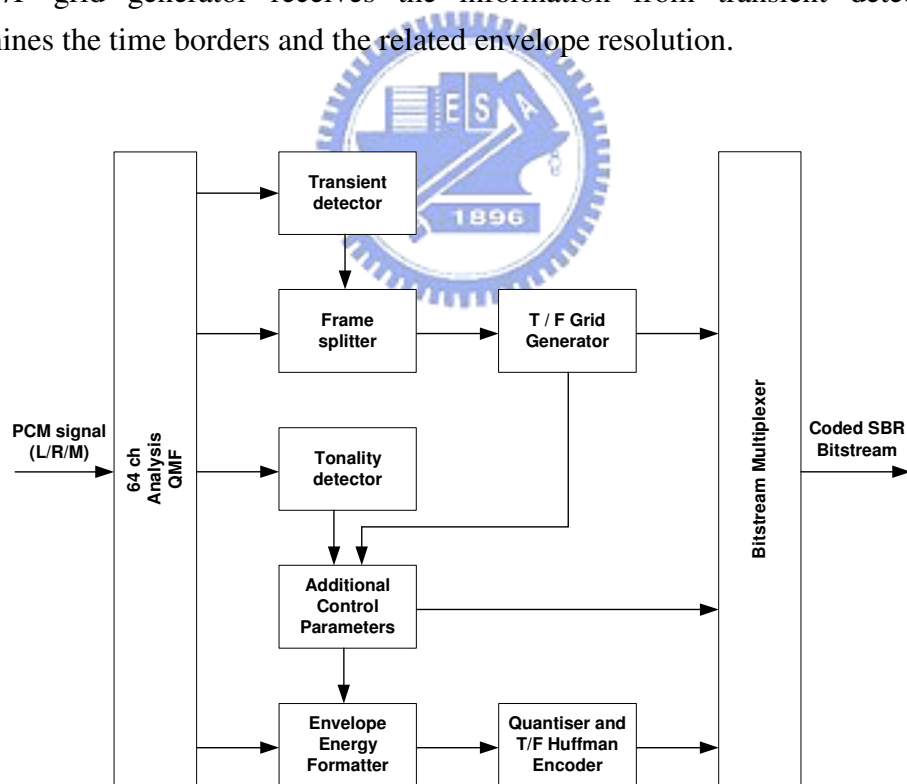


Figure 17: Block diagram of 3GPP HE-AAC encoder [15].

4.1 Time/Frequency Grid Design

In 3GPP SBR encoder, the frequency band table is selected depending on bit rate and sampling rate and it does not alter among frames. Time segments, envelopes

resolution and frame classes are determined according to below mechanism.

4.1.1 Transient Detector

Transient detector is the most important module in 3GPP T/F grid. The following frame splitter and T/F grid generator operates according to the information from it. On a word, the SBR coding quality relies on this module.

The objective of transient detector is to determine whether a transient occurs in the present frame and find the position for the on-set of the transient. The output variables of transient detector, tranFlag and tranPos are used for recording the above information.

Transient detector operates on subband samples of one frame length and starts from sample 8. The basic principle of transient detector is to estimate the energy difference among samples, and determines whether a transient exists depending on information of energy difference. At first, calculate the average energy of each subband in the processed frame and then derive the standard deviations. Next, for each subband, calculate the neighbor energy difference of each sample and compare it to the relating standard deviations. If the energy difference is larger then the relating standard deviations, then take down the value which exceeds the standard deviation. For each time samples, the estimation procedure is executed 64 times, and the values indicating “large energy difference” are added. Finally, check each sample whether one with the energy difference value exceeds the threshold. If it does, then set tranFlag to true, and record the position of this sample. The diagram of transient detector is illustrated in Figure 18.

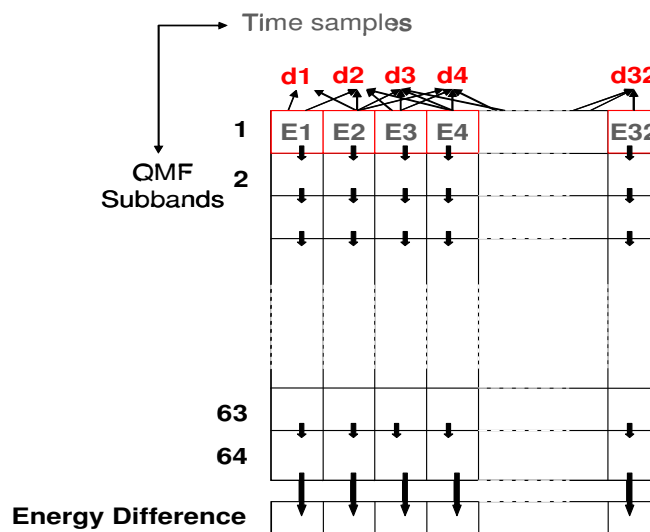


Figure 18: The illustration of detection mechanism in transient detector.

4.1.2 Frame Splitter

Frame Splitter only operates when transient detector has detected the absence of a transient in the current frame. It decides whether the current frame is split into two envelopes of equal size and uses a variable splitFlag to store the result. The concept of frame splitter similar to transient detector is to estimate the energy difference, but not as precise as transient detector. The estimation unit used in this module is half of frame length. Compare the energies of the two half frame, the variable splitFlag can be determined.

4.1.3 T/F Grid Generator

The T/F grid generator creates the time/frequency grid for one SBR frame. Input parameters are provided by transient detector and frame splitter. Frame class is determined at first. It is accomplished depending on the trailing frame border of last frame (FIX or VAR) and the parameter tranFlag of the present frame. On a word, if there is a transient in the current frame, the trailing frame border is VAR; else, the trailing frame border is FIX. The total combination of the leading frame border and transient is described in Table 1. When most transients are sparse, the FIXVAR-VARFIX pair is used. The current frame is encoded with the FIXVAR portion, and the VARFIX grid is stored for the next frame. If no transient occurs in the next frame, the stored VARFIX grid is used; else, the new calculations are needed for the new transient, and merged with the already calculated grid, whereby, a VARVAR class frame is used.

| Leading Frame Border \ TransFlag | 0 | 1 |
|----------------------------------|--------|--------|
| FIX | FIXFIX | FIXVAR |
| VAR | VARFIX | VARVAR |

Table 1: The combination of transient and trailing frame borders.

The positions of time borders in one frame are determined mainly on the position of transient, i.e. the input parameter tranPos provided by transient detector. Each

Chapter 5

Efficient Design of Time/Frequency Grid

Time/Frequency Grid decides the format of reconstruction unit in both time and frequency domain. The resolution of reconstruction unit determines the accuracy degree of reconstructed signal and required bits. It is obvious that for stable signal, the format of T/F grid should be “simple” to reduce the required bits. Oppositely, in order to reduce the distortion of reconstructed signal, the format of T/F grid should be dedicated. In addition, if there are more available bits, the resolution of T/F grid can be higher. Consequently, T/F grid decision is greatly involved with bit rate and audio contents. The efficient design of T/F Grid which this thesis proposes emphasize on these two main factors.

The first issue is how to judge a T/F grid assignment is good or poor. This paper introduces a method to measure the reconstruction error of signal objectively. With the analysis of error, collocating different bit rates, the most suitable form of T/F grid can be selected.

T/F grid comprises frequency table decision, time borders distribution, envelope resolution decision and frame class decision. Frequency tables and envelope resolution codetermine the frequency resolution of T/F grid, and time borders distribution and frame class are responsible for time resolution. The reconstruction error measurement is described first, and the designs of other sub-modules are followed.

5.1 Analysis of Reconstructed Error

The process of SBR decoder has been described in Chapter 2. Through simulating the concept of reconstruction in decoder, the corresponding reconstructed error can be estimated. The reconstructed error used in this thesis is defined as error of power spectrum. The estimation process is accomplished as follows:

First, the rescaling gain values can be hypothesized to the power spectrum of original high frequency contents divided by corresponding low frequency ones. It is given by

$$\alpha_g = \frac{\sum_{i \in g} H_{i,g}}{\sum_{i \in g} L_{i,g}} \quad (11)$$

Where $H_{i,g}$ represents the power spectrum of high frequency band samples in one grid unit, and $L_{i,g}$ stands for the corresponding power spectrum of low frequency band samples, i.e.

$$\begin{aligned} H_i &= |h_i|^2 \\ L_i &= |l_i|^2 \end{aligned} \quad (12)$$

Where the h_i is the high frequency band sample, and l_i is the related low frequency band sample.

The number of grid units and related α_g is determined by the form of T/F grid, using G as notation. Consequently, the reconstructed envelope error of T/F grid is estimated by

$$E(G) = \sum_{g \in G} \sum_{i \in g} |H_{i,g} - \alpha_g L_{i,g}|^2 \quad (13)$$

Next, the goal is to find one kind of T/F grid to minimize $E(G)$, which can be expressed by

$$G = \underset{G}{\text{Arg}} [\min E(G)] \quad (14)$$

And (13) is expressed as follows.

$$\begin{aligned} E(G) &= \sum_{g \in G} \sum_{i \in g} |H_{i,g} - \alpha_g L_{i,g}|^2 = \sum_{g \in G} \sum_{i \in g} |H_{i,g}^2 - 2\alpha_g L_{i,g} H_{i,g} + \alpha_g^2 L_{i,g}^2| \\ &= \left(\sum_{g \in G} \sum_{i \in g} H_{i,g}^2 \right) - 2 \left(\sum_{g \in G} \alpha_g \sum_{i \in g} L_{i,g} H_{i,g} \right) + \left(\sum_{g \in G} \alpha_g^2 \sum_{i \in g} L_{i,g}^2 \right) \\ &= \left(\sum_{g \in G} \sum_{i \in g} H_{i,g}^2 \right) - 2 \left(\sum_{g \in G} \alpha_g \sum_{i \in g} L_{i,g} H_{i,g} \right) \end{aligned} \quad (15)$$

Obviously, the minimum $E(G)$ is involved with the latter part of (15). Therefore, (13) can be reduced to

$$G = \underset{G}{\text{Arg}}[\min E(G)] = \underset{G}{\text{Arg}} \left[\max \left(\sum_{g \in G} \alpha_g \sum_{i \in g} L_{i,g} H_{i,g} \right) \right] \quad (16)$$

Furthermore, it can be expressed as follows

$$\sum_{g \in G} \alpha_g \sum_{i \in g} L_{i,g} H_{i,g} = \sum_{g \in G} \left(\frac{\sum_{i \in g} H_{i,g}}{\sum_{i \in g} L_{i,g}} \cdot \sum_{i \in g} L_{i,g} H_{i,g} \right) = \left\{ \sum_{g \in G} \left(\sum_{i \in g} H_{i,g} \right)^2 \cdot \left(\frac{\sum_{i \in g} L_{i,g} H_{i,g}}{\sum_{i \in g} L_{i,g} \cdot \sum_{i \in g} H_{i,g}} \right) \right\} \quad (17)$$

It is clear to see that the error is related to the energy of original high frequency contents and the correlation between high frequency bands and replicated low frequency bands.

According to (17), the reconstructed error will be affected more by the high energy samples than small energy ones. Therefore, the T/F grid which is picked out through minimum $E(G)$ tends to take care of samples with large energy, and may ignore the others with small energy. However, the distortion of small energy samples is huge, and makes the reconstructed signal sounds noisy. In order to overcome this problem, this paper introduces critical unit to revise the criterion. Critical unit is used for energy normalization and defined as follows, in the time domain, each critical unit contains four samples(two timeslots), and in the frequency domain, the resolution of critical unit is involved with critical band bandwidth, i.e. the critical unit contains fewer subbands in low frequency and more subbands in high frequency. Instead of minimum error, the objective is the minimum distortion rate, and (13) is revised by

$$\sum_{c \in G} \frac{\left| \sum_{i \in c} H_{i,c} - \sum_{i \in c} \alpha_{g(i)} L_{i,c} \right|^2}{\sum_{i \in c} H_{i,c}^2} \quad (18)$$

Comparing (13) with (18), the measurement unit is magnified form sample to critical unit. In order to reduce the calculation complexity, (18) can be changed into a radical expression

$$\sum_{c \in G} \sqrt{\frac{\left| \sum_{i \in c} H_{i,c} - \sum_{i \in c} \alpha_{g(i)} L_{i,c} \right|^2}{\sum_{i \in c} H_{i,c}^2}} = \sum_{c \in G} \frac{\left| \sum_{i \in c} H_{i,c} - \sum_{i \in c} \alpha_{g(i)} L_{i,c} \right|}{\sum_{i \in c} H_{i,c}} \quad (19)$$

The calculation of (19) is defined as DSR (energy difference to original signal ratio). Finally, the reconstructed error estimation is given by

$$G = \underset{G}{\text{Arg}} \left[\min_G \text{DSR}(G) \right] = \underset{G}{\text{Arg}} \left[\sum_{c \in G} \frac{\left| \sum_{i \in c} H_{i,c} - \sum_{i \in c} \alpha_{g(i)} L_{i,c} \right|}{\sum_{i \in c} H_{i,c}} \right] \quad (20)$$

Due to the frame boundaries can be variable; the number of critical unit in each frame is different. Consequently, (20) is revised by

$$G = \underset{G}{\text{Arg}} \left[\min_G \overline{\text{DSR}}(G) \right] = \underset{G}{\text{Arg}} \left[\frac{\sum_{c \in G} \frac{\left| \sum_{i \in c} H_{i,c} - \sum_{i \in c} \alpha_{g(i)} L_{i,c} \right|}{\sum_{i \in c} H_{i,c}}}{\sum_{c \in G} 1} \right] \quad (21)$$

To summarize, the objective of T/F grid decision is to find a format of T/F grid to minimize the averaged DSR.

5.2 Frequency Band Table Decision

Frequency band tables determine the resolution of T/F grid in the frequency domain and the precision of tone addition. Hence, frequency band tables dominate the frequency resolution in SBR. The frequency band tables used in SBR include master frequency band tables, high resolution frequency band tables, low resolution frequency band table, noise floor frequency band tables and limiter frequency band tables. All the frequency tables can be built from master frequency band tables. Consequently, the design issue is the way to select the most suitable master frequency band table.

There are eight different master frequency band tables defined in SBR codec. The delicate tables have fine resolution but take more bits, and inversely, the coarse tables economize the required bits but bring more distortion. In addition, contents of signal

also have great influence on table selection. Therefore, table decision should be considered with bit rate and audio content.

From the aspect of reconstructed error, the most suitable frequency table can be picked out through calculating the relating DSR. However, the method depending on DSR greatly increases calculation complexity and may change frequency table between frames easily. Regardless of complexity, the adaptive method faces shortcomings similar to adaptive SBR range decision mentioned above, which contain SBR header overhead, tone trembling artifact and disable for DPCM in time domain. Therefore, changing master frequency band table between frames too often consumes more bits and may reduce the coding efficiency. Furthermore, the resolution of frequency tables is not as flexible as time borders. In short, it is not allowed to choose arbitrary subband as frequency band boundary. Consequently, selecting frequency band table by DSR is inappropriate. From the other aspect, the resolution of chosen frequency band table greatly affects the precision of adding tones. To summarize, the frequency band table should be coarse to save bits at most time, and when additional tone components is needed, the higher resolution of frequency band table can be selected according to the information from tone-addition mechanism. In this thesis, we choose the coarsest frequency table for saving consumed bits.

5.3 Time Borders and Envelope Resolution

This sub-module is responsible for determining the time resolution of T/F grid and corresponding envelope resolution. The former contains number of envelopes in one frame and locations of time borders. The latter defines the detailed frequency resolution of each envelop in one frame. According to the constraint of SBR standard, there are four time borders and related five envelopes in one frame at most. With calculating the DSR for each form of T/F grid, the one with the minimum DSR can be selected. If the highest resolution is 4 time samples (2 timeslots), in one frame with 32 time samples, the total combinations of time borders and envelope resolution is given by

$$\sum_{i=0}^4 2^{i+1} * C_i^7 = 1878 \quad (22)$$

However, the resulting calculation complexity is very high. In order to simplify the calculation, this thesis proposes an efficient search algorithm through dynamic programming. The notation $DSR_{i,j}^{k,u}$ used in the following presents the minimum DSR value for the range from $2i$ -th timeslot to $2j$ -th timeslot of the current frame, with k

time borders and u high resolution envelopes. For example, $DSR_{2,7}^{3,2}$ is illustrated in Figure 20. Hence, the objective is to find the $DSR_{0,8}^{k,u}$. Furthermore, the DSR with higher number of “k and u” is deducted from the lower ones, i.e. $DSR_{i,j}^{1,1}$ can be derived from two possible combinations, one is $DSR_{i,t}^{0,0} + DSR_{t,j}^{0,1}$, and the other is $DSR_{i,t}^{0,1} + DSR_{t,j}^{0,0}$. The sketch for the deduction of time borders and high resolution envelopes is described in Figure 21.

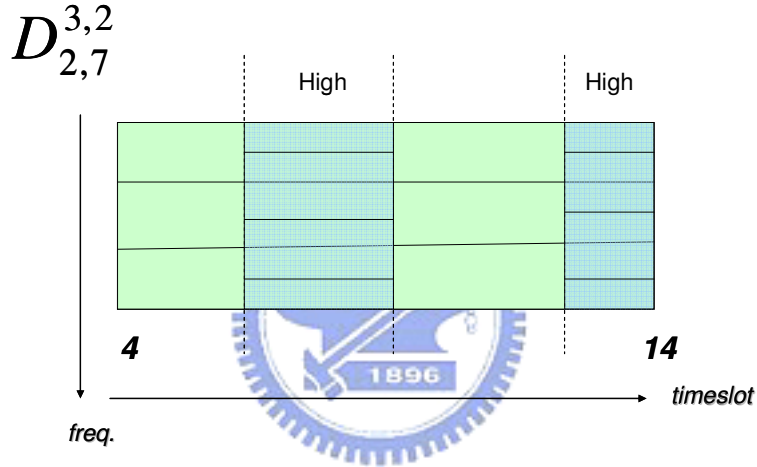


Figure 20: Illustration of DSR notation.

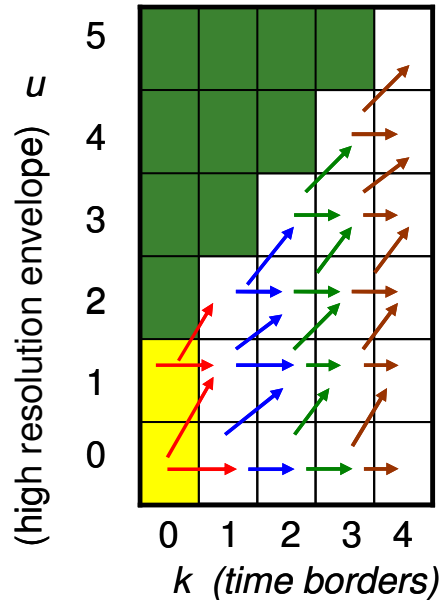


Figure 21: The trellis-lattice deducing path by dynamic programming. Therefore, the structure of dynamic programming is given by

$$\begin{aligned}
DSR_{2i,2j}^{k,u} &= \underset{i+1 \leq t \leq j-1}{Min} \left\{ DSR_{2i,2t}^{0,0} + DSR_{2t,2j}^{k-1,u}, DSR_{2i,2t}^{0,1} + DSR_{2t,2j}^{k-1,u-1} \right\} \\
0 &\leq i < j \leq 8; \\
0 &\leq k \leq 4, \\
0 &\leq u \leq k+1
\end{aligned} \tag{23}$$

And the initial cases are calculated to be the bottom of the structure for dynamic programming.

$$\begin{cases} DSR_{2i,2j}^{0,0} \\ DSR_{2i,2j}^{0,1} \end{cases} \tag{24}$$

$$0 \leq i < j \leq 8$$

Through deriving the minimum DSR of each sub-structure, the probable combinations of target structure are greatly reduced. The total combinations of this search algorithm with dynamic programming are

$$2^2 * \sum_{i=1}^7 C_1^i + 2^2 * \sum_{i=1}^6 C_1^i + 2^2 * \sum_{i=1}^5 C_i^5 + 2^2 * \sum_{i=1}^4 C_i^4 = 296 \tag{25}$$

Compared to (12), it is clear to that the complexity is much lower. However, the most time-consumed portion is to derive the DSR of each possible region. In this proposed dynamic programming algorithm, only the initial cases need to be calculated. With the initial DSR, the other DSR of possible regions can be “pieced” out easily. Consequently, the total calculations needed for DSR are only

$$2 * (8 + 7 + 6 + 5 + 4 + 3 + 2 + 1) = 72 \tag{26}$$

In addition, the factor about consumed bits needs to be taken account of into the T/F grid decision. The first issue is how to estimate the consumed bits of each form of T/F grid. In the SBR bit stream, the energies of grid units are quantized and then transmitted. Therefore, the number of grid units within T/F grid can be assumed to present the consumed bits, i.e. more is the number of grid unit, more bits this T/F grid takes. From this aspect, one time border is regarded equivalent as one high resolution envelope, due to the both creating the same number of grid units. Thus, the total amount of time borders and high resolution envelopes can present the degree of bit-consuming.

In order to take consideration for the bit overhead, there are ten bit-consuming stages set in the dynamic programming. Each stage indicates the different degree of

bit-consuming. The relation between these stages and relating number of time borders and high resolution envelopes is described in Table 2. From the lower stages to the higher ones, the T/F grid with the minimum DSR of each stage is derived. If there is one relating DSR value under the threshold, then the search terminates. The flow chart is illustrated in Figure 22.

| 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 |

Table 2: Combinations of bit-consuming stages.

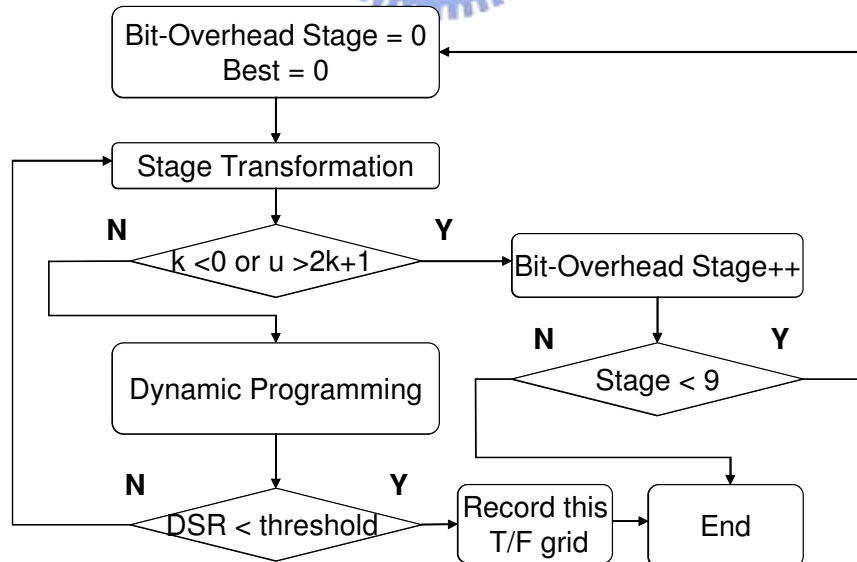


Figure 22: DP flow chart with quality constraint.

It is clear to see that the resulting performance is greatly involved with the threshold. This threshold is named as “quality threshold” because it stands for the satisfactory reconstructed error. Further, the quality threshold should be different on different bit rates. At higher bit rate, this threshold needs to be stricter; on the contrary, it should be looser at the lower bit rate. In a word, the quality constraint should be

adaptive bit rates.

Through the above algorithm, the derived T/F grid presents that the error for this grid format is acceptable. However, the situation that no any T/F grid meets the quality constraint may happen. In such case, the highest resolution T/F grid may not be the best solution. Therefore, another threshold is needed. This constraint is referred to as “efficiency threshold” because it restricts efficiency of consumed bits. The new form of T/F grid is adapted only when it improves some percentage over the best one. The efficiency constraint ensures that each additional time border and high resolution envelop is worth. The modified flow chart with efficiency threshold is illustrated in Figure 23. The proposed T/F grid decision take account of quality, bit overhead, and encoding bit rates at the same. The experiments in Chapter 6 will show the efficiency compared to other codecs.

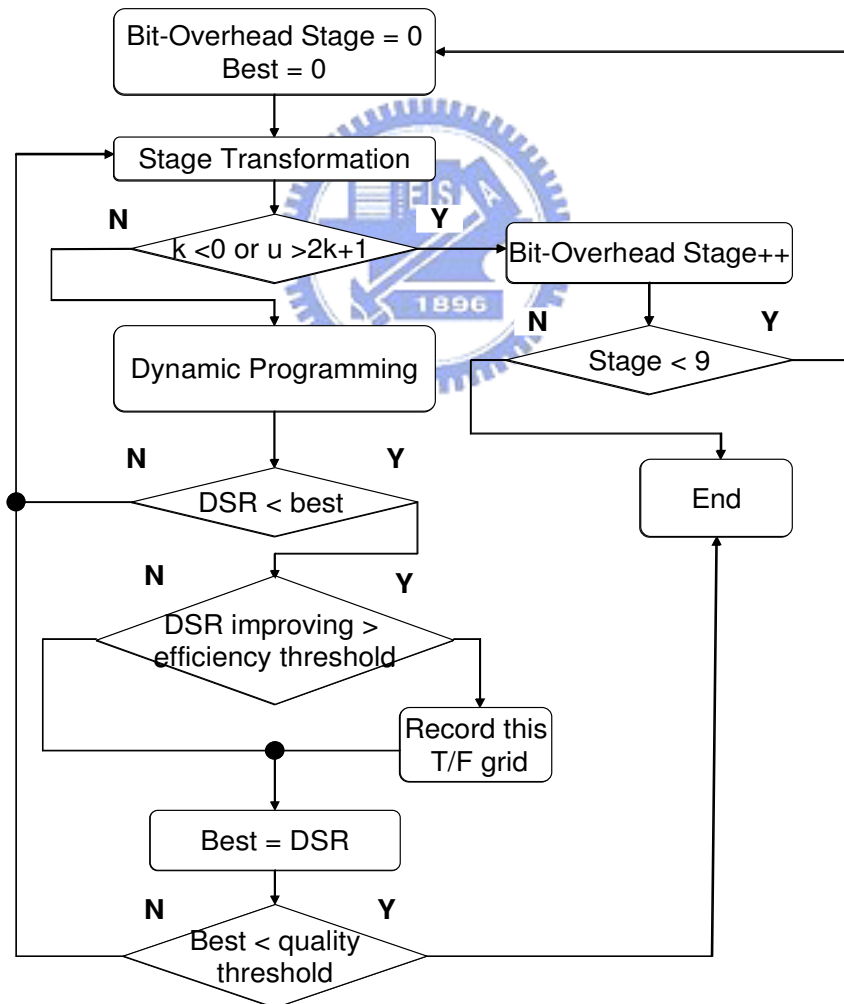


Figure 23: DP flow chart with both quality and efficiency constraint.

5.4 Frame Class Decision

Four frame classes are used in SBR codec. Each frame class has different flexibility to describe the distribution of time borders. According to the position of time borders, leading frame border and trailing frame border, frame class of each frame can be determined. In addition to record the format of time borders, the objective of frame class for variable frame border is to spare bits for time borders. If both frame borders are fixed, it means that there are two “time borders” wasted. In Figure 24(a), there are two consecutive frames and respective time borders. If the frame borders are always constant, the latter frame needs an extra time border. Comparing to Figure 24(b), due to the variable trailing frame border of the first frame, the first time border of the second frame can be removed.

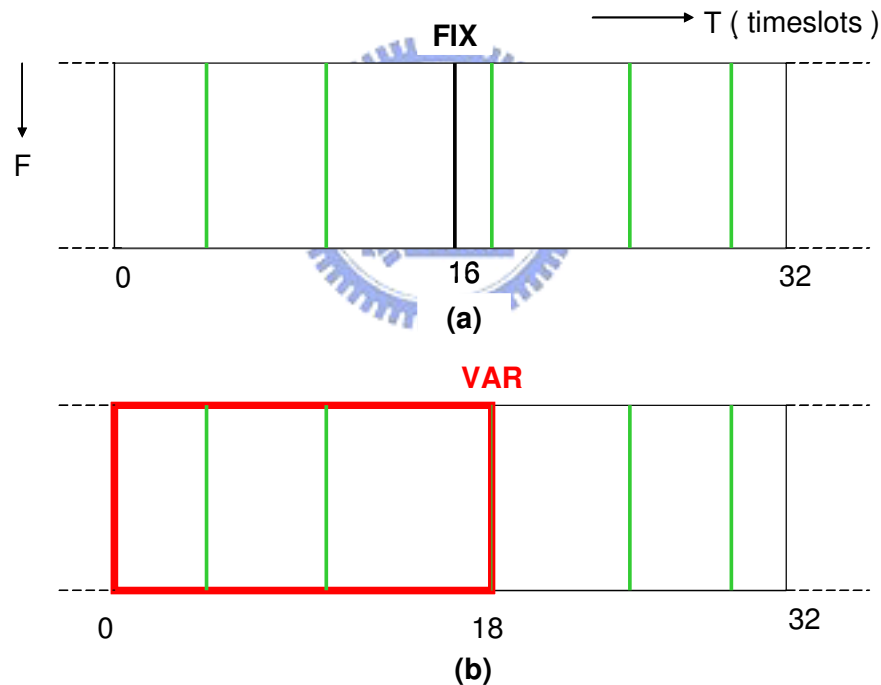


Figure 24: An example for variable frame border.

Consequently, in order to determine the position of the trailing frame border of each frame, the information for time borders of the next frame is necessary. According to looking ahead of the next frame and the distribution of time borders in this frame, the most suitable frame class can be determined.

Chapter 6

Artifacts in SBR

6.1 Tone Trembling

The patching which determines the corresponding relation between replicated low frequency bands and original high frequency bands is different depending on different master frequency band tables, SBR start and stop boundaries. If one of these three factors changes, then patching changes, i.e. assume that the 8th subband is replicated for someone high frequency subband this frame, and the next frame, the replicated low frequency subband change into 10th subband. This phenomenon may cause the spectrum discontinued in time domain. In noise-like signal, the discontinuous spectrum is hard to be discovered in both perceptual hearing and spectral envelope. However, in tone-rich signal, this phenomenon is much more serious. Comparing Figure 25(a) to Figure 25(b), and it is easy to see the discontinuity of spectrum. The major discontinued envelope locates on tones. In perceptual hearing, this artifact causes the signal sounds like trembling, or sparkling. Therefore, this phenomenon is referred to tone trembling artifact.

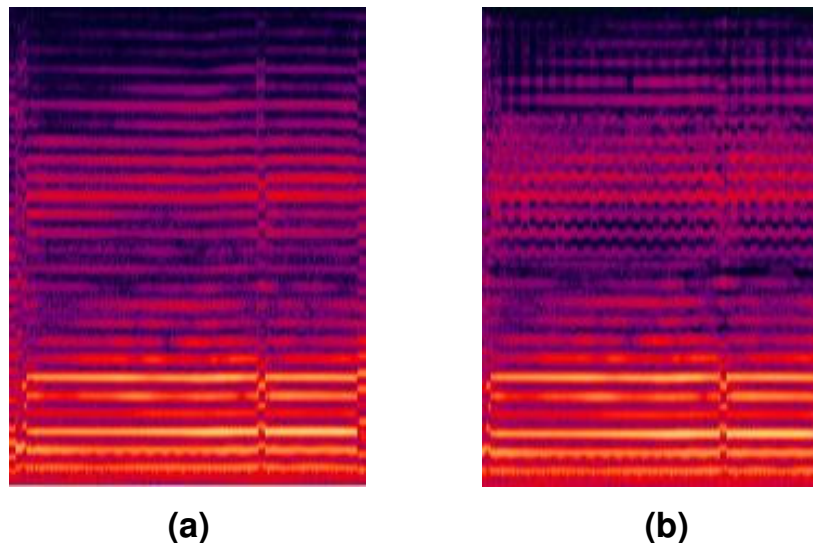


Figure 25: Tone trembling effect.

6.2 Tone Shift

In tone-rich signal, the tone components usually distribute regularly. Figure 26 shows this phenomenon. In this kind of signal, using inverse filter and adding additional tone components is ineffective and bit-consuming. Furthermore, sinusoids in the frequency transform to signals with constant magnitude in the time domain. Therefore, the non-clipping method should be better than the clipping method either in time or frequency domain, i.e. no any time borders or additional components are needed. However, it will cause tone shift artifact in the reconstructed signal. This artifact is referred to “tone shift” because of its spectrum shape. In Figure 27, it shows this phenomenon. The blue line represents the spectrum of original signal, and the red one is reconstructed by HE-AAC. It is clear to see that, in SBR range, the tone components have offsets comparing to original ones, but they still keep regular. However, in the perceptual hearing, this artifact is almost hard to be discovered.

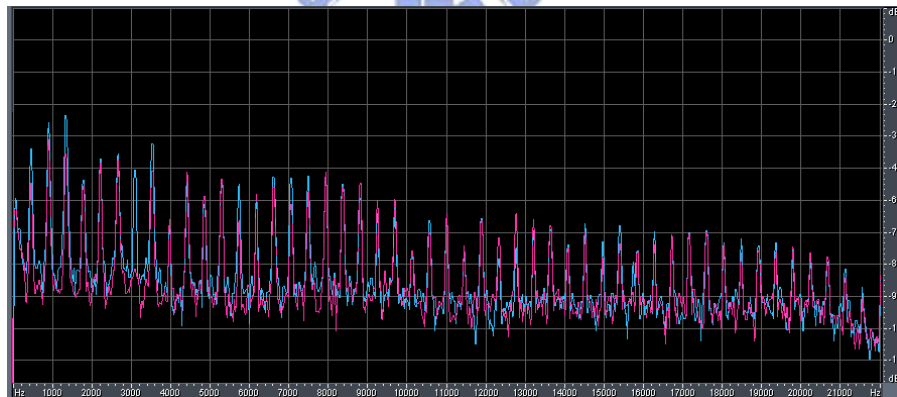


Figure 26: An example for characteristics of tone-rich signals

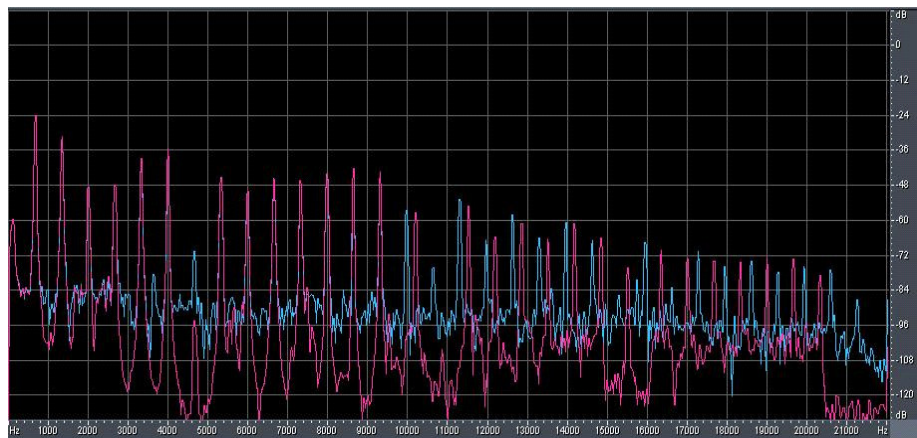


Figure 27: Tone shift effect.

6.3 Sawtooth

The limiter gain mechanism in SBR decoder is to avoid unwanted noise substitution. It restricts the maximum values for gain. As mentioned in Chapter 2, these maximum gain values are calculated according to a limiter frequency band table which has coarser frequency resolution than original gain values use. Hence, the limiter gain mechanism implicates to preserve the envelope of the replicated signal. Limiter gain values can be assumed as the maximum rescaling value for adjusting the replicated contents to original ones. If the rescaling value exceeds limiter gain, it represents that the adjusting of replicated signal is immoderate and may destroy the continuity of spectrum. However, this protection mechanism may bring artifact when the envelope of the replicated signal highly differs from the envelope of original one. In Figure 28, the envelope of high bands is flat, but the envelope of low bands is sharp. In such situation, in order to adjust reconstructed contents as similar as original ones, revising the envelope of replicated signal is necessary. Therefore, if the limiter gain mechanism always turns on, the adjustment is restricted, and the resulting envelope of reconstructed signal may be discontinuous. This phenomenon is named as sawtooth artifact due to the discontinuous envelope, which is illustrated in Figure 29.



Figure 28: The envelope comparison between high bands and low bands. The envelope of high bands is flat, and the envelope of low bands is sharp.

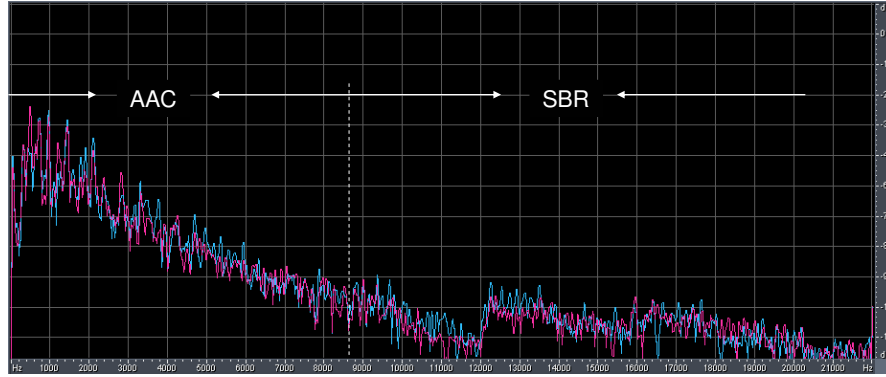


Figure 29: Sawtooth effect due to the limiter gain mechanism.

6.4 Noise Floor Overflow

Noise floor overflow is a common artifact in SBR codec. There are two main reasons causing this phenomenon. The first one is the missing of tone detection in T/F grid. After the envelope adjustment in decoder, the inconsistent content of the noise-like replicated low band and the tonal original high band will cause the huge promotion of the noise floor in the low band to compensate the energy of the lost tones. The “noise-floor overflow” phenomenon, as shown in Figure 30, produces a fizzy sound in perception and results in serious quality degradation.

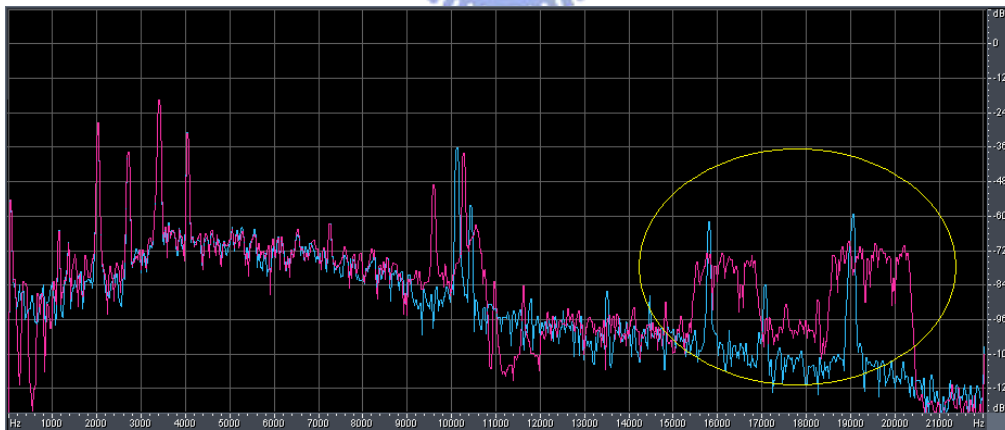


Figure 30: Noise floor overflow due to failure of detecting tones in high bands.

The accuracy of tonality measurement is critical for the artifact, because either the underestimation of the tonal energy or the overestimation of the energy of noise component will lead to the noise-floor overflow. However, the constraint of SBR syntax will restrict the location and the number of the added additional tones, and hence the noise-floor overflow effect is still unavoidable possibly even with accurate tonality measurement. This problem can be overcome by a method of noise-floor correction in encoder part.

On the other hand, the second reason causing noise floor overflow is the interpolation mode. As mention in Chapter 2, the estimation for current SBR frame envelope has two different modes, interpolation and non-interpolation. By comparing the resultant envelops in the two modes, the interpolation mode will generate the flat envelop, and oppositely the non-interpolation mode will maintain the original envelop structure of the replicated low bands. In other word, under the interpolation mode, the inherent characteristic of the envelop flatness does not agree with the sharp envelop of the tonal bands. Hence, due to the noise-floor overflow effect, the envelope estimation mode should be switched between interpolation and non-interpolation. When the envelope of high bands is flat, interpolation mode is selected. Oppositely, as the envelope of high band is sharp, it needs to change into non-interpolation mode. However, the information of tonality can presents the characteristic of relating envelope. Therefore, the estimation mode can be determined according to the tonalities. As shown in Figure 31, there is a serious noise-floor overflow around the first tone which is replicated from low bands and almost overwhelmed by the amplified noise. The last two tones which are compensated additionally have no the artifact. This is because the tonality information is kept by the tone adding mechanism. Once without the mechanism as in Figure 32, the artifact occurs again. This also presents the immunity of the tone adding mechanism against the noise-floor overflow effect under interpolation mode.



Figure 31: Noise floor overflows due interpolation mode. The target circle indicates the noise floor overflow is from the averaged energy with tone component.

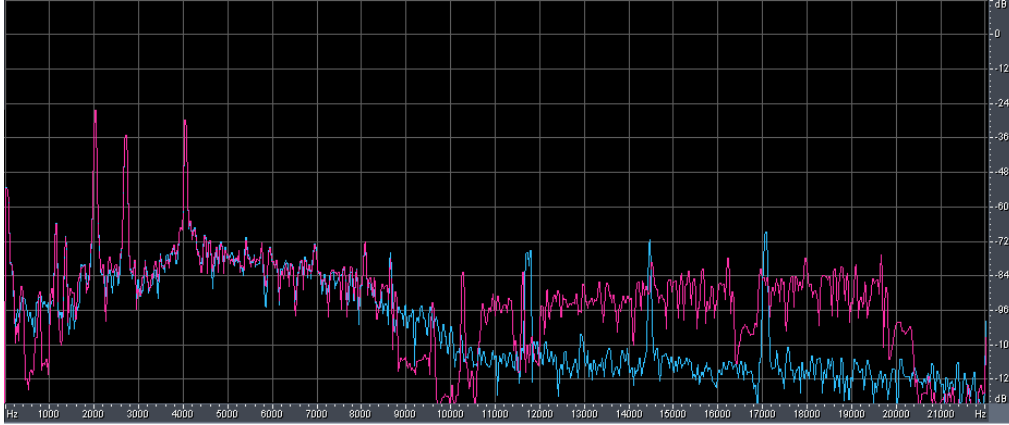


Figure 32: A comparison to Figure 31. It incident the result without tone addition mechanism.



Chapter 7

Experiments

In this chapter, extensive experiments are made to prove the enhancement of proposed methods through objective and subjective measurements. NCTU-AAC [20] is adopted as core encoder in our NCTU HE-AAC [21]. The MPEG test tracks are chosen as our default track database. Next, the experiment on the PSPLAB audio database [22] is executed to prove the robustness of our design. Through both objective and subjective tests, the efficiency and quality of our proposed methods are well examined.

7.1 Measurement Tools Description

Objective Quality Measurement Tool

In the objective test, we choose EAQUAL as the tool to assess the audio quality. EAQUAL stands for Evaluation of Audio Quality. The objective difference grade (ODG) is the output variable from this objective measurement tool. The range of the ODG value is from 0 to -4, where 0 presents an imperceptible impairment and -4 correspond to a very annoying impairment. The improvement up to 0.1 is usually perceptually audible. The implementation of EAQUAL is based on the ITU-R recommendation BS.1387 [23].

Subjective Quality Measurement Tool

We choose the tool called “MUSHRA” to conduct the subjective test. MUSHRA (Multiple Stimulus with Hidden Reference and Anchors) is proposed 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 a codec which can be compared directly with the reference, i.e. the original audio signal as well as the anchors [24]. MUSHRA follows the test method and impairment scale recommended by ITU-R BS.1116 [25].

7.2 Objective Quality Measurement in MPEG Test Tracks

The twelve tracks, which contain critical music balancing on the percussion,

string, wind instruments and human vocal, recommended by MPEG, are included in the assessment of audio quality. Table 3 shows the characteristics and details of these tracks. In the section, the quality improvement of the proposed methods at different bit rates is considered based on these MPEG test tracks.

| Tracks | | Signal Description | | | |
|--|------|----------------------------|--------|------------|---------|
| | | Signal | 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) |
| <p>Remarks:</p> <p>(a) Transients: pre-echo sensitive, smearing of noise in temporal domain.</p> <p>(b) Tonal/Harmonic structure: noise sensitive, roughness.</p> <p>(c) Natural vocal (critical combination of tonal parts and attacks): distortion sensitive, smearing of attacks.</p> <p>(d) Complex sound: stresses the Device Under Test.</p> | | | | | |

Table 3: The twelve tracks recommended by MPEG

Six different methods are compared at different bit rates from 48kbps to 112kbps. The first one is NCTU-AAC; from the second one to the fifth one are NCTU HE-AAC with uniform “cuts” in T/F grid with 0, 1, 3, and 7 cuts respectively. The frequency table is suggested by SBR standard [3]. The sixth one is NCTU HE-AAC with the proposed T/F grid design.

| Bit Rate | 112 kbps | | | | | |
|----------------|----------|--------|---------|---------|---------|---------|
| Coding Methods | 1 | 2 | 3 | 4 | 5 | 6 |
| es01 | -0.35 | -0.88 | -0.67 | -0.6 | -0.56 | -0.56 |
| es02 | -0.11 | -0.86 | -0.61 | -0.57 | -0.55 | -0.49 |
| es03 | -0.19 | -1.63 | -0.87 | -0.6 | -0.57 | -0.58 |
| sc01 | -0.61 | -0.6 | -0.59 | -0.6 | -0.6 | -0.57 |
| sc02 | -1.13 | -0.63 | -0.59 | -0.58 | -0.58 | -0.62 |
| sc03 | -0.92 | -0.89 | -0.76 | -0.73 | -0.72 | -0.74 |
| si01 | -1.03 | -1.28 | -1.05 | -1.03 | -0.95 | -1 |
| si02 | -0.92 | -2.21 | -1.38 | -1.04 | -0.99 | -0.7 |
| si03 | -1.49 | -1.06 | -1.07 | -1.07 | -1.05 | -1 |
| sm01 | -1.18 | -1.19 | -1.18 | -1.18 | -1.11 | -1.07 |
| sm02 | -0.62 | -1.72 | -1.16 | -1.16 | -1.11 | -1.21 |
| sm03 | -1.19 | -1.75 | -0.95 | -0.91 | -0.9 | -0.89 |
| Average | -0.8117 | -1.225 | -0.9067 | -0.8392 | -0.8075 | -0.7858 |

Sample Rate: 44100 Hz
Coding Method:
1: NCTU-AAC; 2: NCTU HE-AAC without any cuts in T/F grid; 3: NCTU HE-AAC with uniform 1 cut; 4: NCTU HE-AAC with uniform 3 cuts; 5: NCTU HE-AAC with uniform 7 cuts; 6: Proposed DP design of T/F grid.

Table 4: Objective measurements through ODGs for different T/F grid design in HE-AAC at 112 kbps.

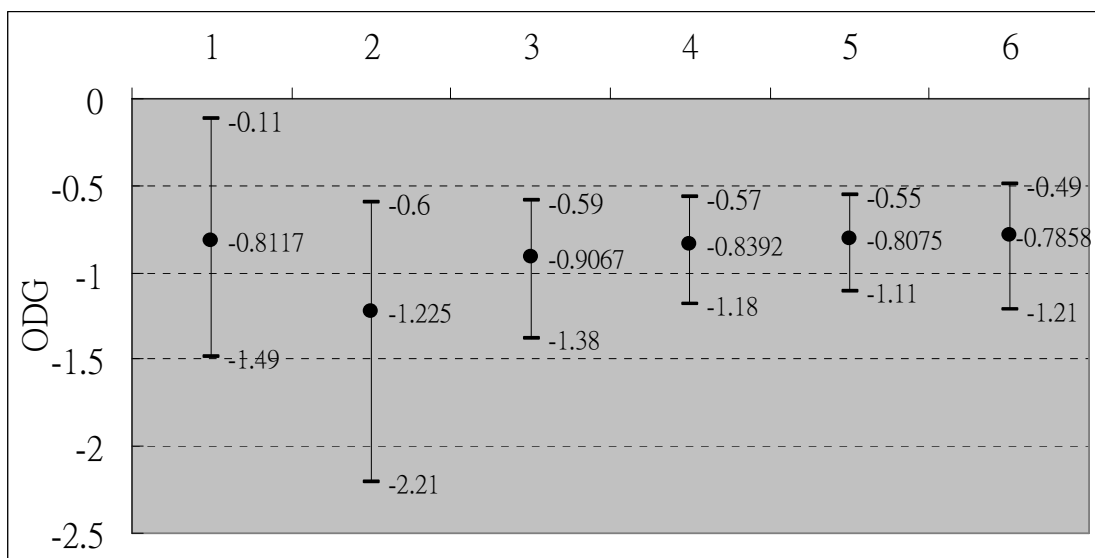


Figure 33: The ODG variance comparison of Table 4.

| Bit Rate | 96 kbps | | | | | |
|----------------|---------|---------|-------|--------|---------|---------|
| Coding Methods | 1 | 2 | 3 | 4 | 5 | 6 |
| es01 | -0.54 | -0.91 | -0.7 | -0.64 | -0.6 | -0.6 |
| es02 | -0.23 | -0.86 | -0.62 | -0.58 | -0.57 | -0.51 |
| es03 | -0.37 | -1.65 | -0.9 | -0.62 | -0.61 | -0.6 |
| sc01 | -1.01 | -0.74 | -0.71 | -0.72 | -0.74 | -0.68 |
| sc02 | -1.7 | -0.75 | -0.72 | -0.71 | -0.72 | -0.74 |
| sc03 | -1.5 | -1.01 | -0.88 | -0.85 | -0.84 | -0.86 |
| si01 | -1.77 | -1.5 | -1.22 | -1.23 | -1.19 | -1.18 |
| si02 | -1.27 | -2.31 | -1.51 | -1.13 | -1.07 | -0.79 |
| si03 | -2.56 | -1.28 | -1.31 | -1.34 | -1.34 | -1.22 |
| sm01 | -2.22 | -1.36 | -1.34 | -1.34 | -1.3 | -1.23 |
| sm02 | -1.05 | -1.84 | -1.3 | -1.29 | -1.26 | -1.3 |
| sm03 | -1.87 | -1.97 | -1.15 | -1.13 | -1.14 | -1.08 |
| Average | -1.3408 | -1.3483 | -1.03 | -0.965 | -0.9483 | -0.8992 |

Sample Rate: 44100 Hz
Coding Method:
1: NCTU-AAC; 2: NCTU HE-AAC without any cuts in T/F grid; 3: NCTU HE-AAC with uniform 1 cut; 4: NCTU HE-AAC with uniform 3 cuts; 5: NCTU HE-AAC with uniform 7 cuts; 6: Proposed DP design of T/F grid.

Table 5: Objective measurements through ODGs for different T/F grid design in HE-AAC at 96 kbps.

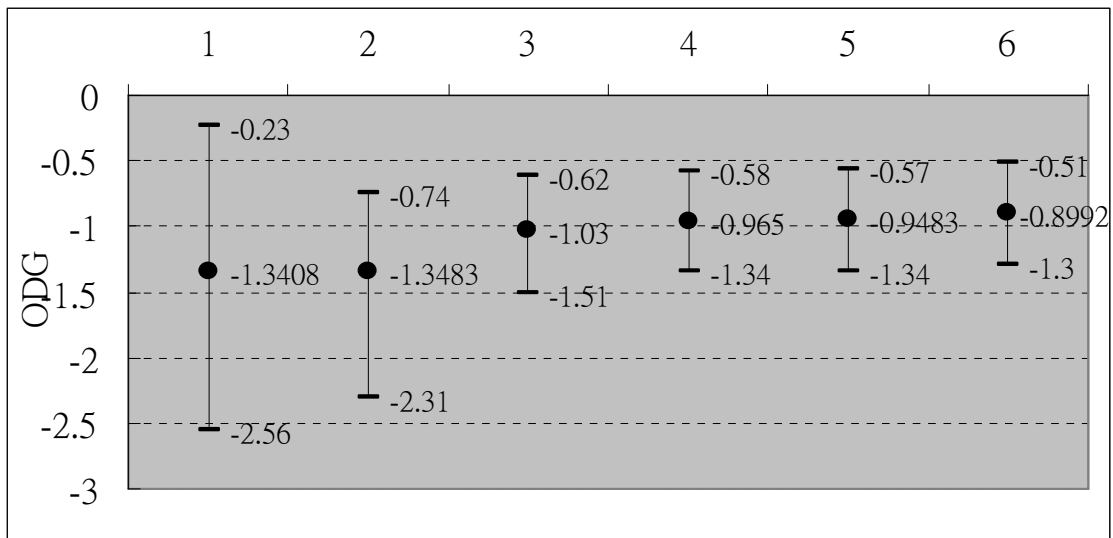


Figure 34: The ODG variance comparison of Table 5.

| Bit Rate | 80 kbps | | | | | |
|----------------|---------|---------|---------|--------|---------|---------|
| Coding Methods | 1 | 2 | 3 | 4 | 5 | 6 |
| es01 | -0.8 | -0.97 | -0.77 | -0.71 | -0.69 | -0.67 |
| es02 | -0.49 | -0.92 | -0.68 | -0.66 | -0.64 | -0.59 |
| es03 | -0.74 | -1.71 | -0.96 | -0.71 | -0.71 | -0.68 |
| sc01 | -1.61 | -1 | -1 | -1.02 | -1.08 | -0.96 |
| sc02 | -2.49 | -1.07 | -1.05 | -1.05 | -1.1 | -1.08 |
| sc03 | -2.47 | -1.27 | -1.13 | -1.1 | -1.11 | -1.11 |
| si01 | -2.78 | -1.97 | -1.63 | -1.61 | -1.57 | -1.59 |
| si02 | -1.94 | -2.43 | -1.65 | -1.31 | -1.27 | -1 |
| si03 | -3.66 | -1.65 | -1.67 | -1.71 | -1.71 | -1.6 |
| sm01 | -3.38 | -1.63 | -1.62 | -1.65 | -1.66 | -1.55 |
| sm02 | -1.82 | -2.02 | -1.58 | -1.56 | -1.55 | -1.52 |
| sm03 | -2.6 | -2.21 | -1.41 | -1.37 | -1.44 | -1.3 |
| Average | -2.065 | -1.5708 | -1.2625 | -1.205 | -1.2108 | -1.1375 |

Sample Rate: 44100 Hz
Coding Method:
1: NCTU-AAC; 2: NCTU HE-AAC without any cuts in T/F grid; 3: NCTU HE-AAC with uniform 1 cut; 4: NCTU HE-AAC with uniform 3 cuts; 5: NCTU HE-AAC with uniform 7 cuts; 6: Proposed DP design of T/F grid.

Table 6: Objective measurements through ODGs for different T/F grid design in HE-AAC at 80 kbps.

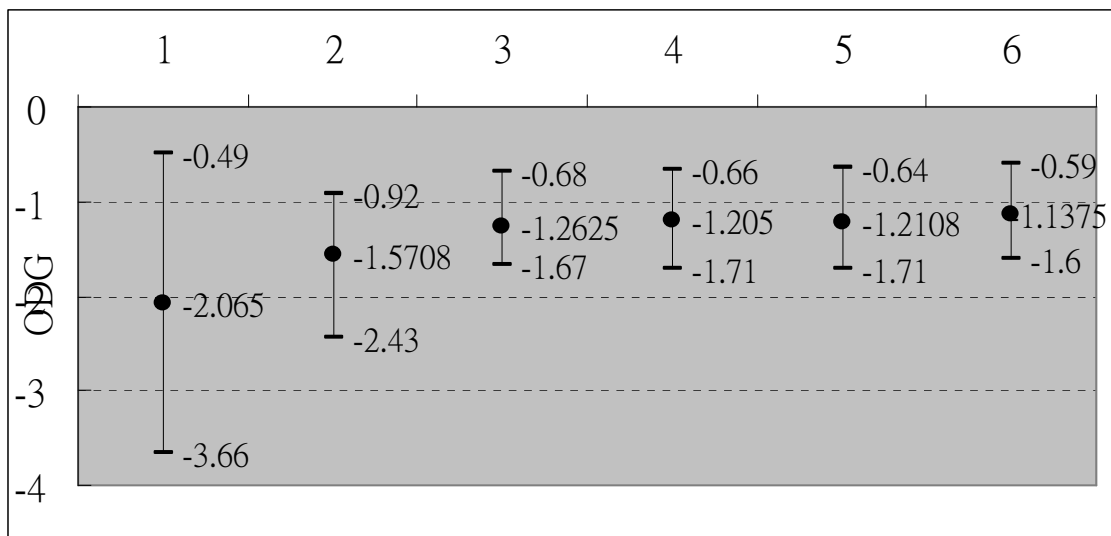


Figure 35: The ODG variance comparison of Table 6.

| Bit Rate | 64 kbps | | | | | |
|----------------|---------|---------|---------|---------|---------|---------|
| Coding Methods | 1 | 2 | 3 | 4 | 5 | 6 |
| es01 | -1.55 | -1.58 | -1.09 | -0.98 | -0.98 | -0.93 |
| es02 | -1.16 | -1.45 | -0.97 | -0.89 | -0.93 | -0.82 |
| es03 | -1.53 | -2.49 | -1.46 | -1 | -1 | -0.96 |
| sc01 | -1.96 | -1.79 | -1.76 | -1.84 | -1.98 | -1.56 |
| sc02 | -3 | -1.68 | -1.67 | -1.69 | -1.82 | -1.65 |
| sc03 | -3.38 | -1.79 | -1.63 | -1.55 | -1.62 | -1.58 |
| si01 | -3.54 | -2.31 | -1.95 | -1.95 | -2.05 | -1.93 |
| si02 | -2.85 | -2.77 | -1.95 | -1.68 | -1.72 | -1.35 |
| si03 | -3.85 | -2.07 | -2.12 | -2.2 | -2.33 | -2.04 |
| sm01 | -3.78 | -2.18 | -2.19 | -2.21 | -2.3 | -2.12 |
| sm02 | -2.77 | -3.03 | -2.17 | -2.09 | -2.36 | -2.18 |
| sm03 | -3.2 | -2.64 | -1.78 | -1.77 | -1.86 | -1.64 |
| Average | -2.7142 | -2.1483 | -1.7283 | -1.6542 | -1.7458 | -1.5633 |

Sample Rate: 44100 Hz
Coding Method:
1: NCTU-AAC; 2: NCTU HE-AAC without any cuts in T/F grid; 3: NCTU HE-AAC with uniform 1 cut; 4: NCTU HE-AAC with uniform 3 cuts; 5: NCTU HE-AAC with uniform 7 cuts; 6: Proposed DP design of T/F grid.

Table 7: Objective measurements through ODGs for different T/F grid design in HE-AAC at 64 kbps.

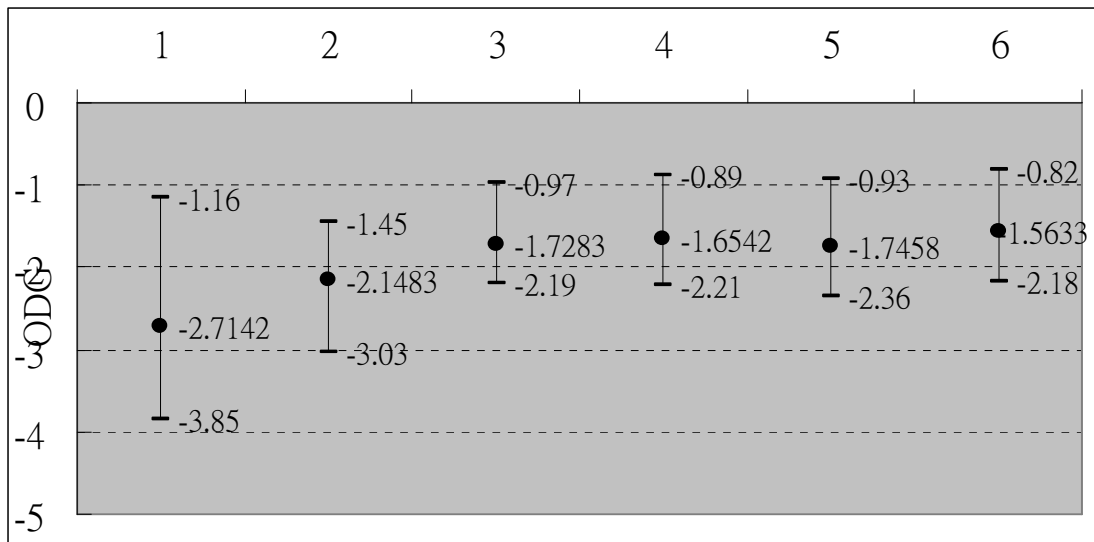


Figure 36: The ODG variance comparison of Table 7.

| Bit Rate | 48 kbps | | | | | |
|----------------|---------|-------|-------|---------|---------|---------|
| Coding Methods | 1 | 2 | 3 | 4 | 5 | 6 |
| es01 | -3.12 | -2.02 | -1.58 | -1.58 | -1.74 | -1.48 |
| es02 | -2.49 | -1.81 | -1.42 | -1.47 | -1.8 | -1.28 |
| es03 | -3.23 | -2.85 | -1.97 | -1.71 | -2.08 | -1.67 |
| sc01 | -2.5 | -2.73 | -2.71 | -2.73 | -2.92 | -2.34 |
| sc02 | -3.14 | -2.55 | -2.53 | -2.51 | -2.54 | -2.47 |
| sc03 | -3.63 | -2.54 | -2.35 | -2.32 | -2.41 | -2.29 |
| si01 | -3.77 | -3.03 | -2.75 | -2.75 | -2.9 | -2.68 |
| si02 | -3.58 | -3.26 | -2.67 | -2.61 | -2.66 | -2.26 |
| si03 | -3.88 | -3.15 | -3.21 | -3.29 | -3.42 | -3.15 |
| sm01 | -3.88 | -3.18 | -3.19 | -3.25 | -3.41 | -3.17 |
| sm02 | -3.33 | -3.63 | -3.21 | -3.26 | -3.29 | -3.25 |
| sm03 | -3.51 | -3.09 | -2.41 | -2.41 | -2.54 | -2.2 |
| Average | -3.3383 | -2.82 | -2.5 | -2.4908 | -2.6425 | -2.3533 |

Sample Rate: 44100 Hz
Coding Method:
1: NCTU-AAC; 2: NCTU HE-AAC without any cuts in T/F grid; 3: NCTU HE-AAC with uniform 1 cut; 4: NCTU HE-AAC with uniform 3 cuts; 5: NCTU HE-AAC with uniform 7 cuts; 6: Proposed DP design of T/F grid.

Table 8: Objective measurements through ODGs for different T/F grid design in HE-AAC at 48 kbps.

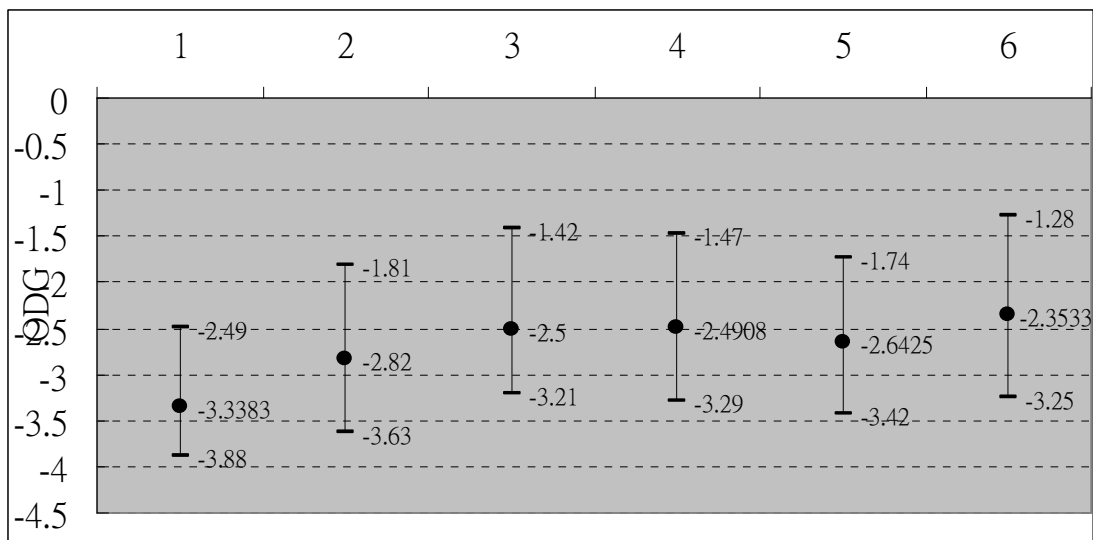


Figure 37: The ODG variance comparison of Table 8.

Summary

In above experiments, method 3 represents the low bit-consuming in T/F grid, and methods 4 and 5 represent medium and high bit-consuming respectively. Therefore, at high bit rates, such as 112 kbps and 96 kbps, method 5 is better than methods 3 and 4. Oppositely, at low bit rates, such as 64kbps and 48kbps, method 5 is getting worse than methods 3 and 4. However, the ODG of our proposed design is the best among these coding methods at every bit rate. The ODG and bit rates comparison curves are illustrated in Figure 38. The curve demonstrates the relation between ODG degeneracy with respect to bit rates. From Figure 38, the quality of AAC reduces with the bit rates rapidly. On the other hand, the curves of SBR codec with bit rates can be controlled to be more smooth than AAC codec. Furthermore, the curve of our design is on top of others, which shows the efficiency of the proposed method.

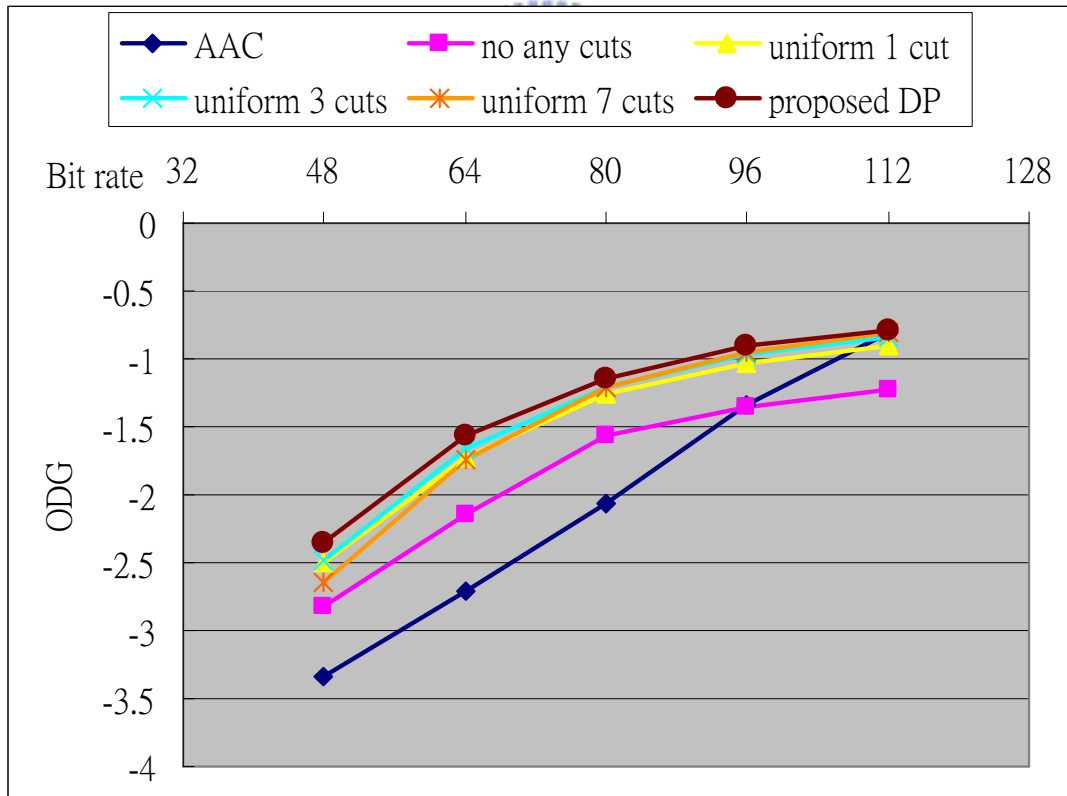


Figure 38: ODG-bit rate curve comparison among different T/F grid design.

7.3 Objective Quality Measurement in Music database

In order to verify the robustness of the proposed design and evaluate the possible

risk for a variety of audio categories, PSPLAB audio database [22] is adopted as testing samples. The database includes 327 tracks which are separated into 16 sets with different signal properties as shown in Table 9.

| Bitstream Categories | | Number of Tracks | Remark |
|----------------------|----------------------|------------------|---|
| 1 | FF123 | 103 | Killer bitstream collection from ff123. |
| 2 | Gpsycho | 24 | LAME quality test bitstream. |
| 3 | HA64KTest | 39 | 64 kbps test bitstream for multi-format in HA forum. |
| 4 | HA128KTestV2 | 12 | 128 kbps test bitstream for multi-format in HA forum. |
| 5 | Horrible_song | 16 | Collections of critical songs among all bitstream in PSPLab. |
| 6 | Ingets1 | 5 | Bitstream collection from the test of OGG Vorbis pre 1.0 listening test. |
| 7 | Mono | 3 | Mono test bitstream. |
| 8 | MPEG | 12 | MPEG test bitstream set for 48KHz. |
| 9 | MPEG44100 | 12 | MPEG test bitstream set for 44100 Hz. |
| 10 | Phong | 8 | Test bistream collection from Phong. |
| 11 | PSPLab | 37 | Collections of bitstream from early age of PSPLab. Some are good as killer. |
| 12 | Sjeng | 3 | Small bitstream collection by sjeng. |
| 13 | SQAM | 16 | Sound quality assessment material recordings for subjective tests. |
| 14 | TestingSong14 | 14 | Test bitstream collection from rshong. |
| 15 | TonalSignals | 15 | Artificial bitstream that contains sin wave etc. |
| 16 | VORBIS_TESTS_Samples | 8 | First 8 Vobis testing sample from HA. |
| Total | | 327 | |

Table 9: The PSPLAB audio database.

In this section, there are two different coding methods used to be compared with our design. The first one is NCTU HE-AAC with uniform 1 cut in T/F grid, and the second one is NCTU HE-AAC with uniform 7 cuts. The two coding methods represent different bit-consuming degree for T/F grid respectively and both adopt the frequency table suggested in SBR standard.

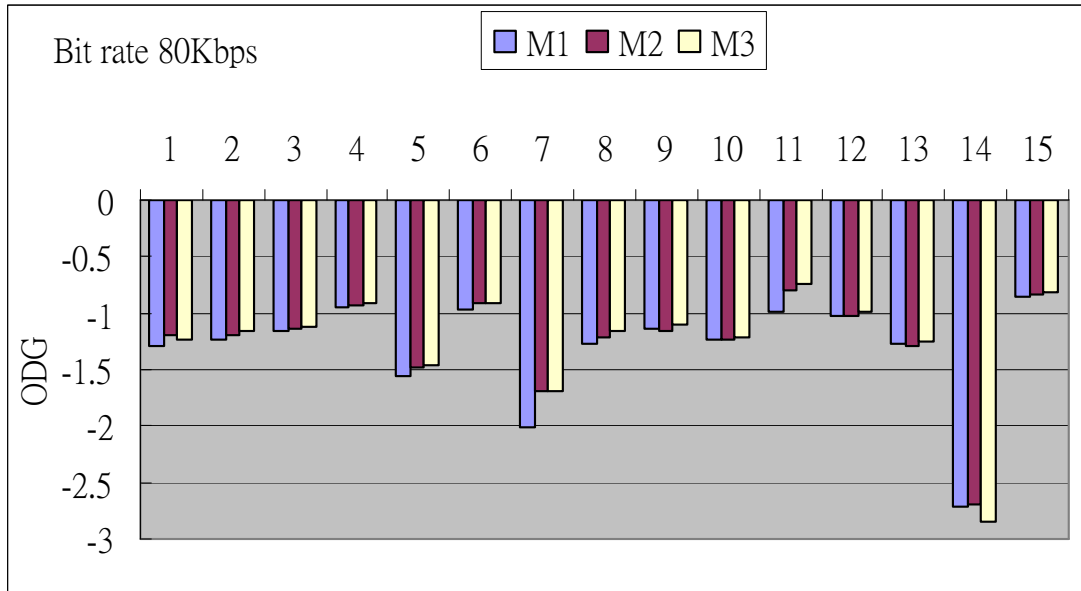


Figure 39: The average ODG of three coding methods in PSPLAB audio database at bit rate 80kbps and sampling rate 44100 Hz. M1 is uniform 1 cut in T/F grid and M2 is uniform 7 cuts in T/F grid. M3 is our design.

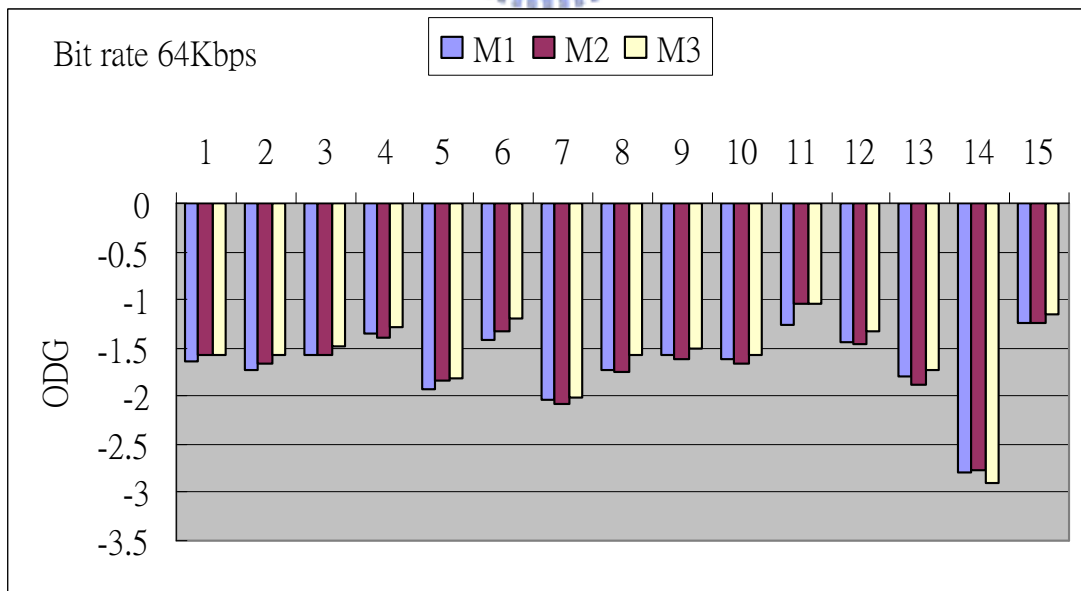
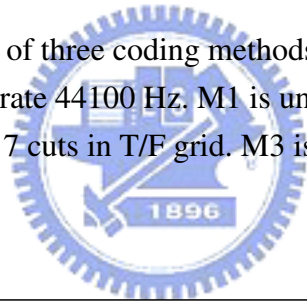


Figure 40: The average ODG of three coding methods in PSPLAB audio database at bit rate 64kbps and sampling rate 44100 Hz. M1 is uniform 1 cut in T/F grid and M2 is uniform 7 cuts in T/F grid. M3 is our design.

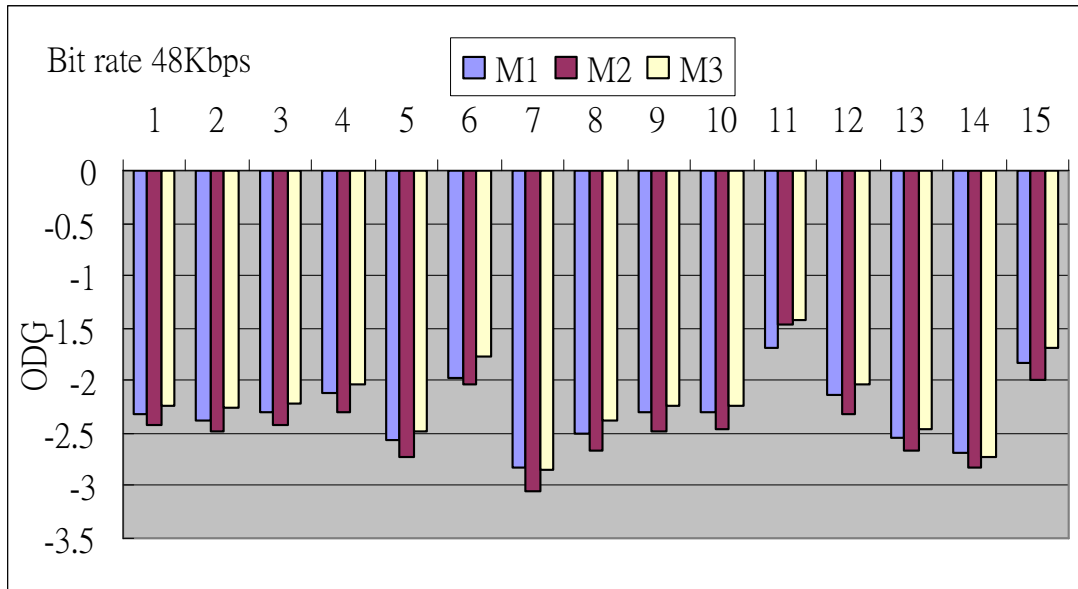


Figure 41: The average ODG of three coding methods in PSPLAB audio database at bit rate 48kbps and sampling rate 44100 Hz. M1 is uniform 1 cut in T/F grid and M2 is uniform 7 cuts in T/F grid. M3 is our design.

Summary

As mention above, at bit rate 80kbps, M2 is better than M1, and on the contrary, M1 is better than M2 at bit rate 48kbps. However, our T/F grid design is better than the other two coding methods for most sets in audio database. It proves the robustness and flexibility of our design.

There are two sets needed to be observed, FF123 and TonalSignals, because the ODG of our design is worse in both sets. Through analyzing the signals in these two sets, the problem tracks can be separated into three categories, frequency table selection, tone-vanishing and noise floor overflow. The higher resolution frequency table is needed due to two situations; one is the envelope of high bands alters rapidly, and the other one is adding tone components. In Figure 42 and Figure 43, there are two examples showing the unstable high band envelopes. Figure 42 shows that the envelope alters hugely by time samples, and in Figure 43, the envelope is very sharp. Figure 44 shows another situation which needs higher resolution frequency table. When additional tone components are adding, detailed frequency table can increase the coding precision. Therefore, the policy of frequency table selection should be revised. In the most time, the frequency table should be coarse to save bits, and when the high band envelope is unstable, the higher resolution table needs to be used.

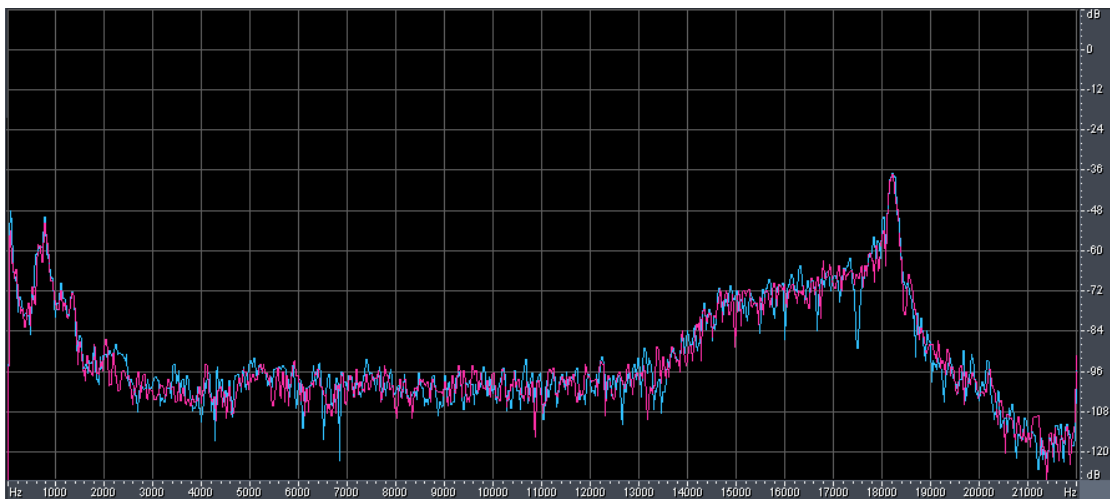
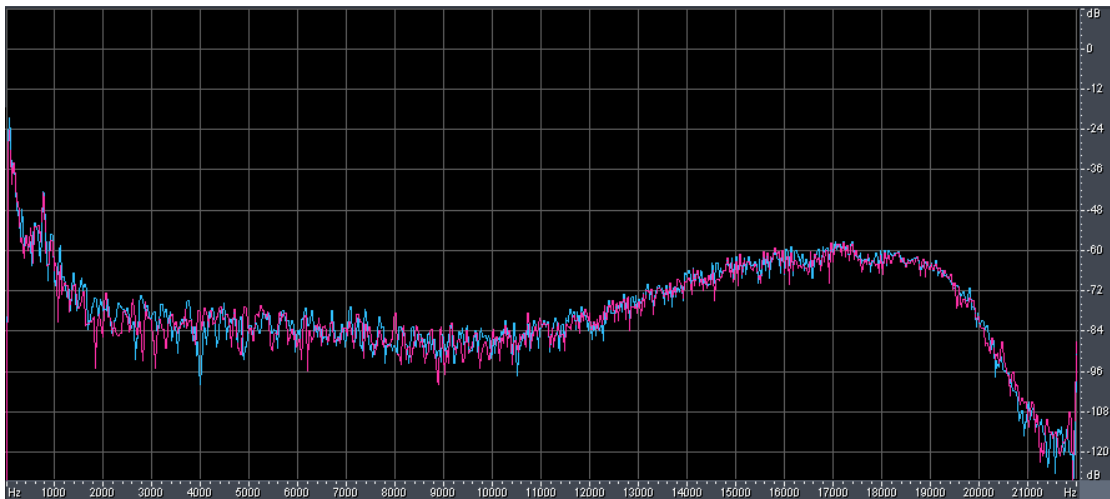
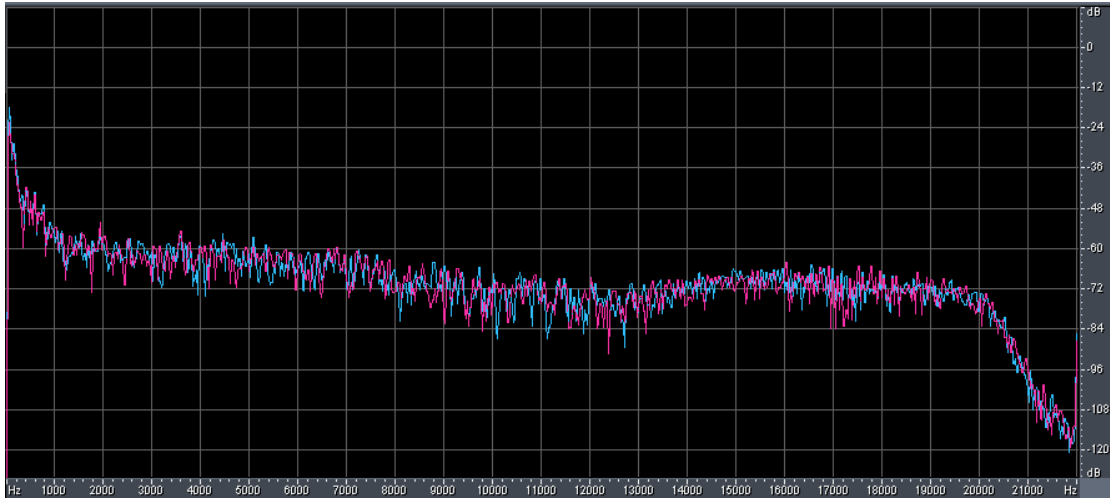


Figure 42: The example for signal whose high band envelope alters rapidly.

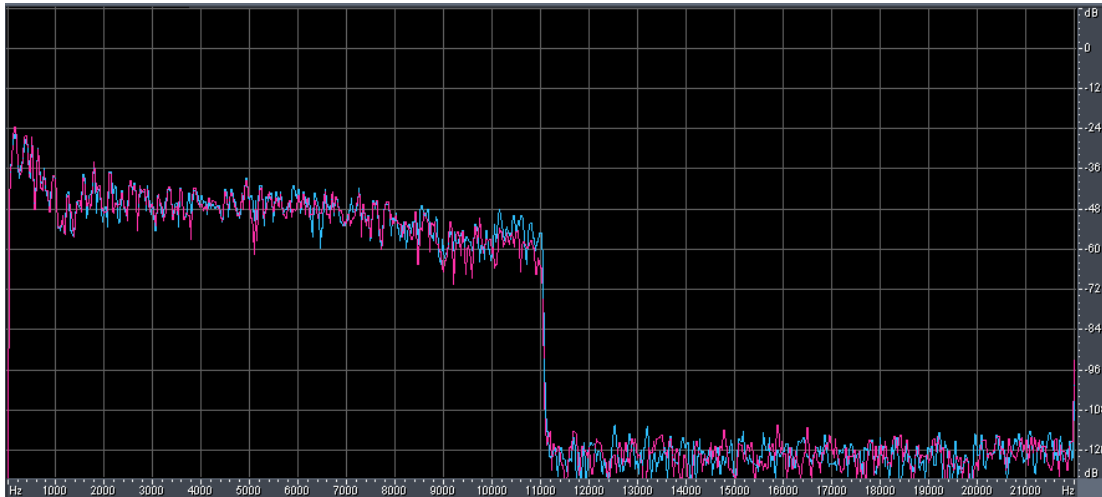


Figure 43: The example for signal which has a sharp high band envelope.

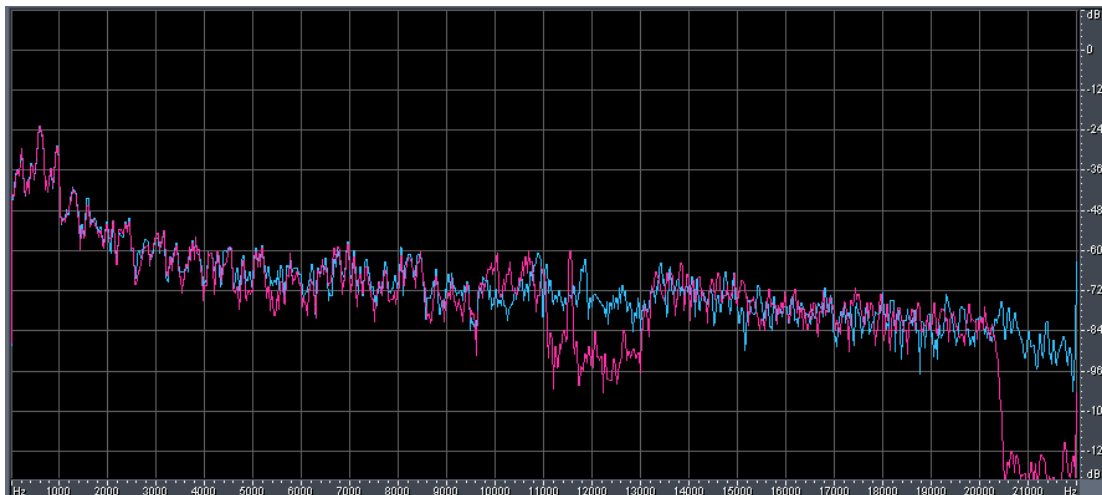


Figure 44: The error of noise floor due to tone addition.

Another problem occurs in the audio tracks showed in **Figure 45**. Through SBR codec, the spectrum of reconstruction signal becomes discontinuous segments. However, comparing to method 2 which uses uniform 7 cuts in T/F grid, our T/F grid design makes tone vanish for some frames in reconstruction signal. From Figure 46 and Figure 47, the tone-vanishing phenomenon is evident. There are four tracks have this property, which are sweep, halvesweep, halvesweepinvert and 20k-20. In addition, we also take aacPlus [12] to test these tracks, and the result is showed in Figure 48. It is clear to see that tone component is missing and replaced by noise. Figure 49 shows another similar spectrum, the difference is the tone component is not continuous in frequency domain. The reconstruction signal by NCTU HE-AAC and aacPlus is illustrated in Figure 50 and Figure 51 respectively. Besides to tone-vanishing phenomenon, the energy of added tone by NCTU HE-AAC is lower than original one.

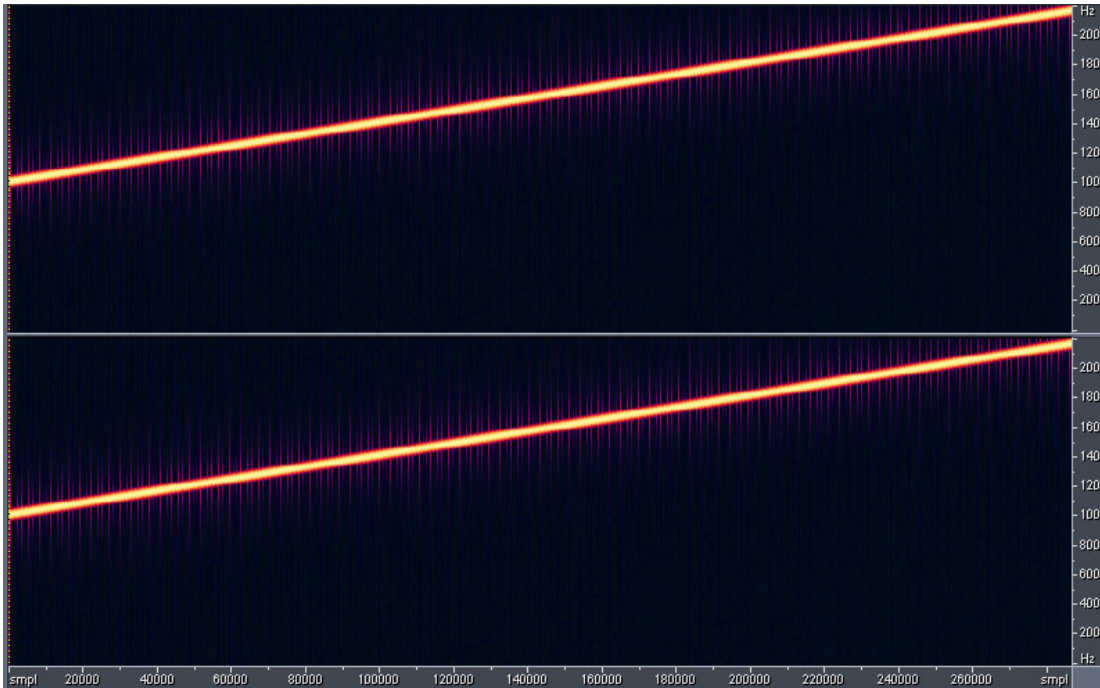


Figure 45: The spectrum of “sweep” which has continuous tone from 10 KHz to 22 KHz.

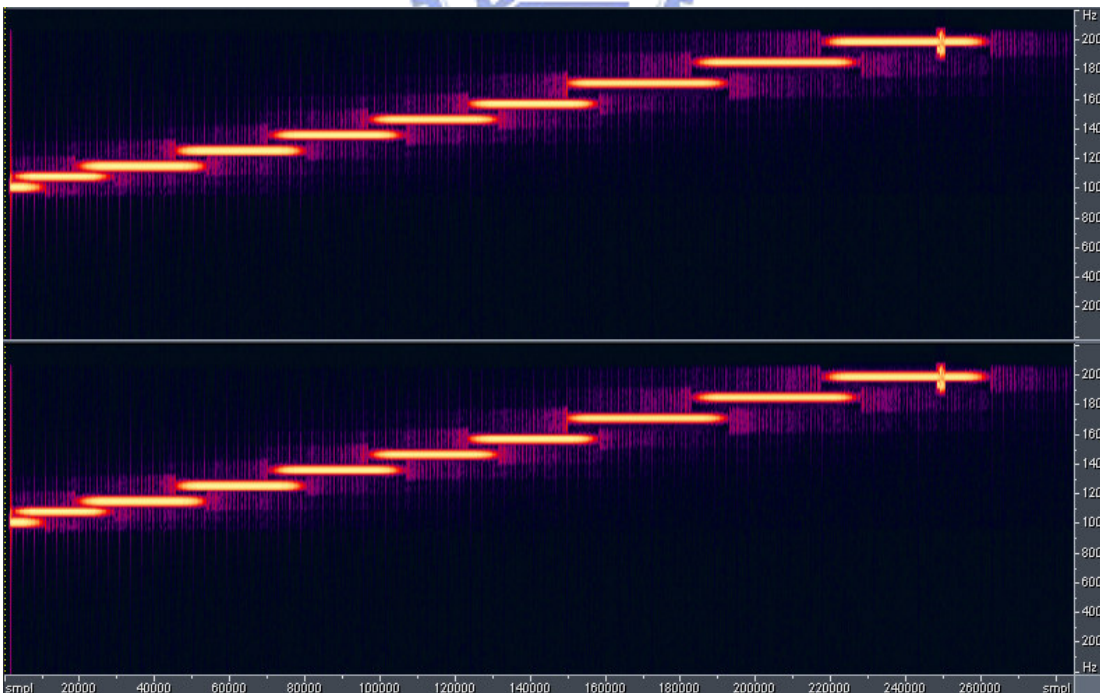


Figure 46: The reconstructed spectrum of Figure 45 through uniform 7 cuts in T/F grid.

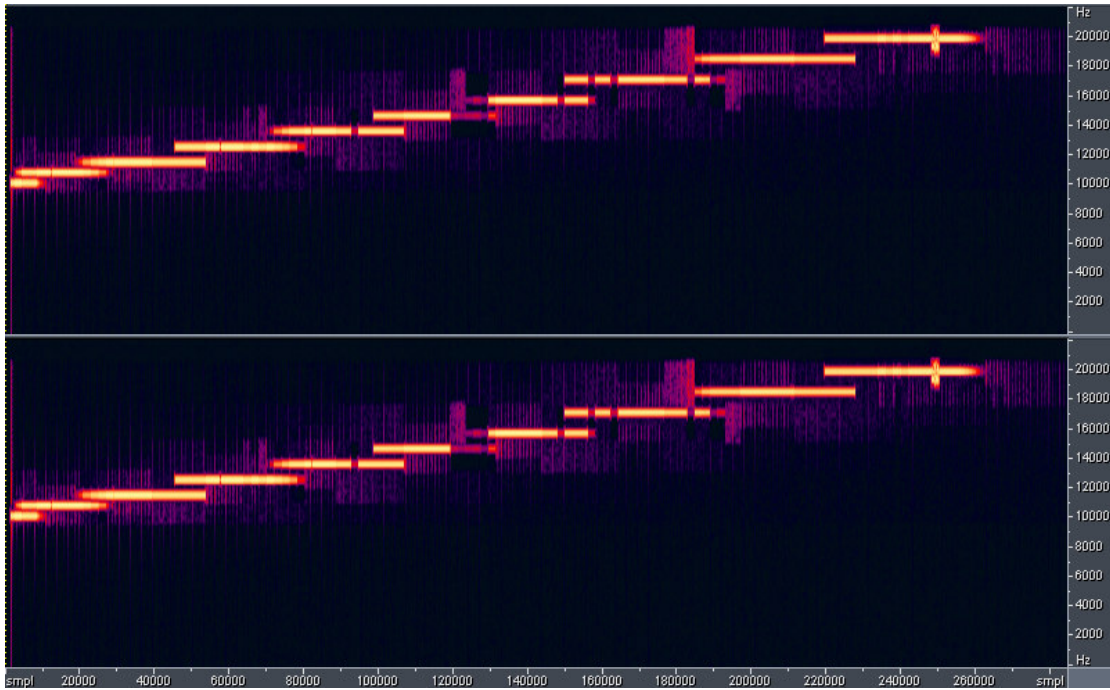


Figure 47: The reconstructed spectrum of Figure 45 through DP T/F grid design.

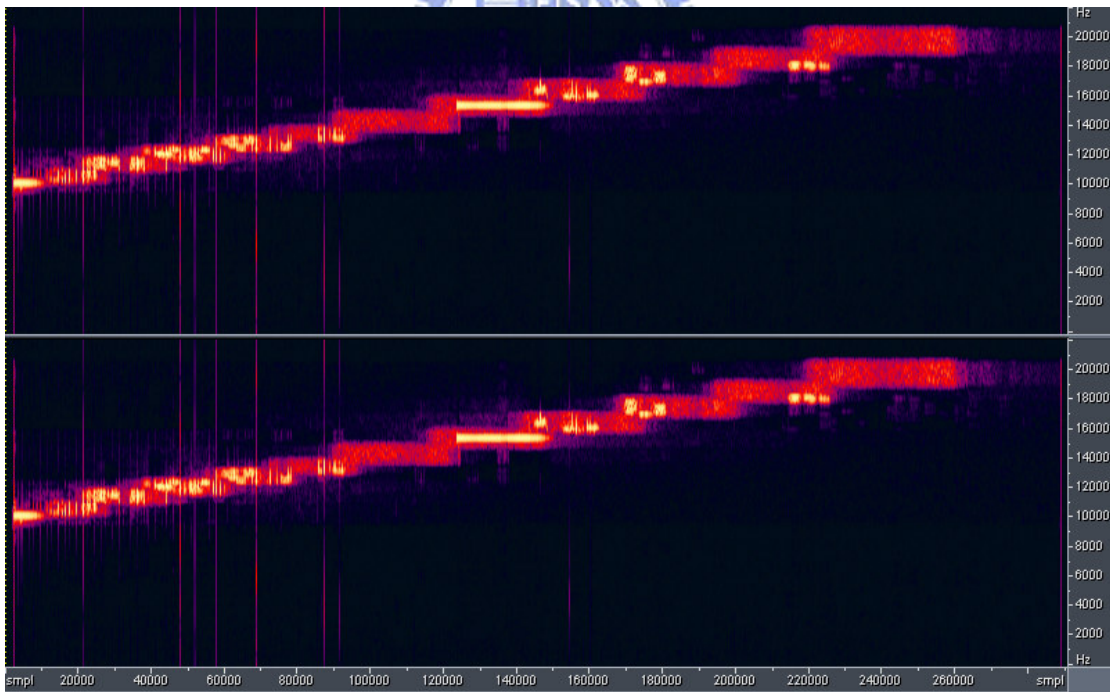


Figure 48: The reconstructed spectrum of Figure 45 through aacPlus.

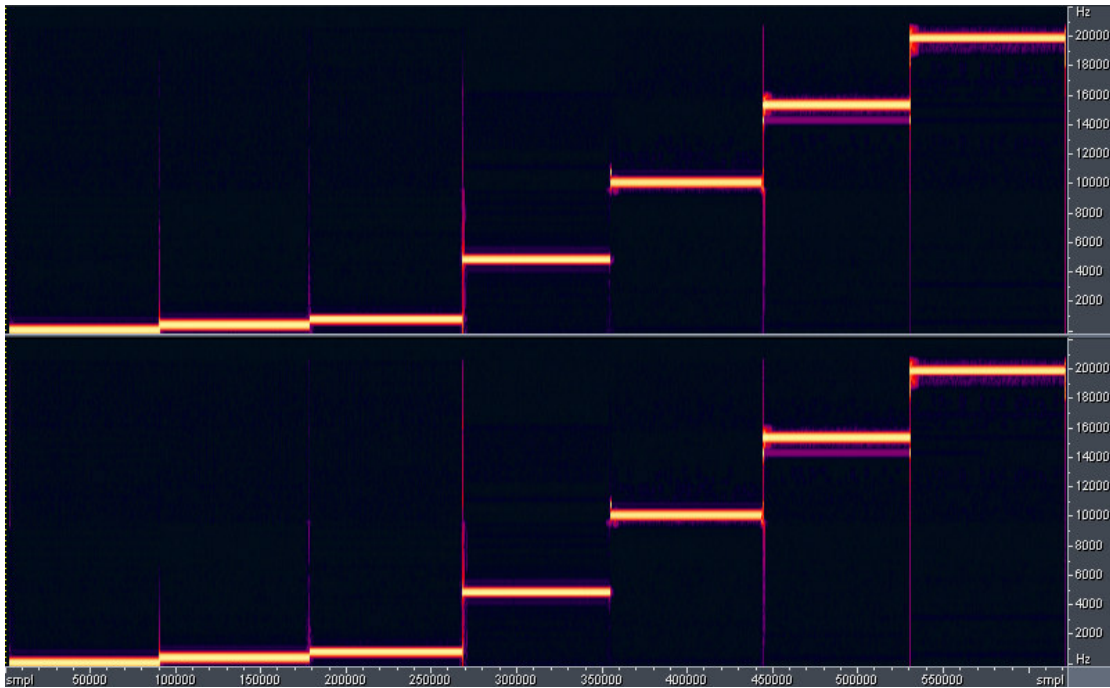


Figure 49: The spectrum of “sin_300_625_1k_5k_10k_15K_20k_m20db” which has interrupted tone from low frequency bands to high frequency bands

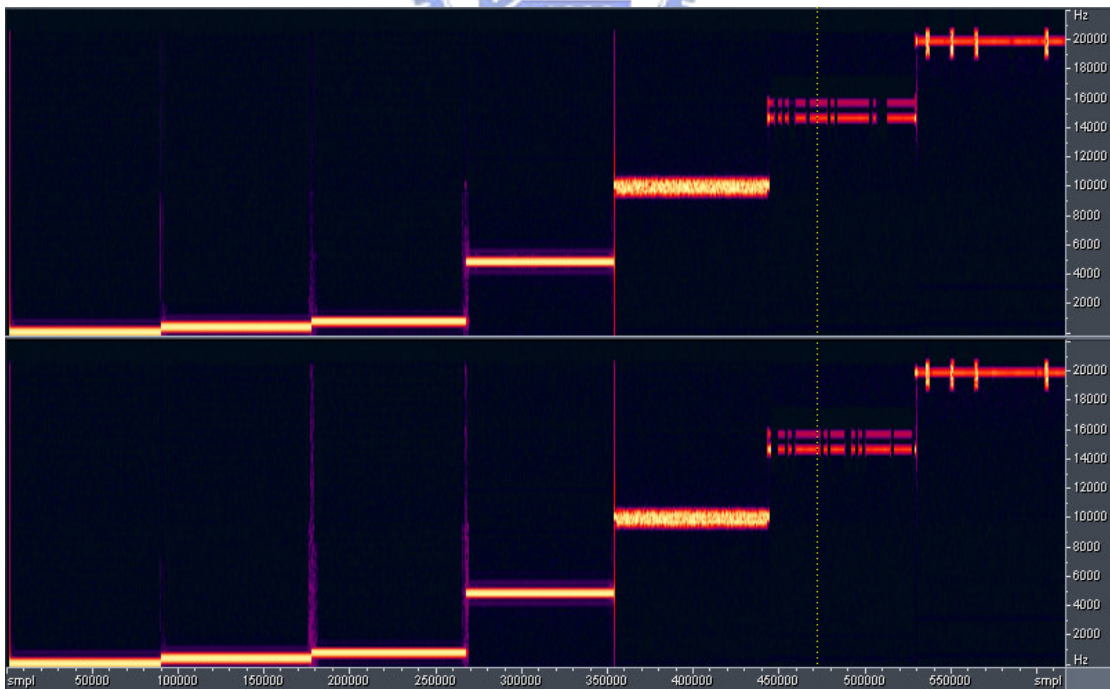


Figure 50: The reconstructed spectrum of Figure 49 through DP T/F grid design.

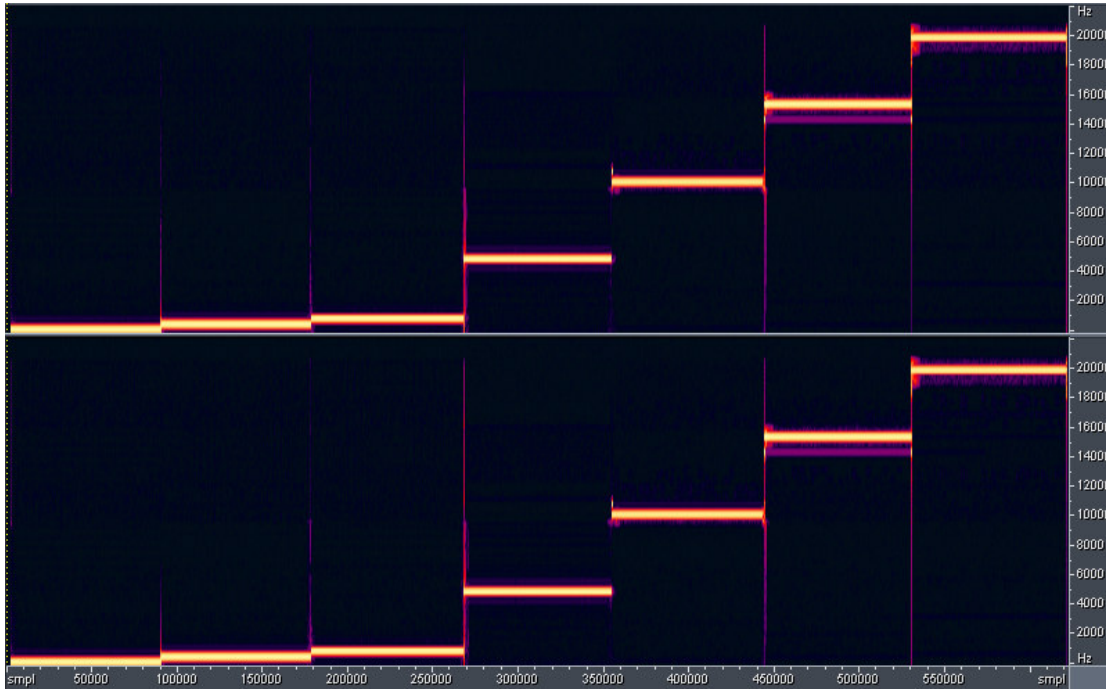


Figure 51: The reconstructed spectrum of Figure 49 through aacPlus.

The last problem is noise floor overflow due to interpolation mode. There are two tracks in TonalSignals set with serious noise floor overflow, `sin_600_19800_9div_m20_0db` and `sin_9kind_valious`. Take `sin_9kind_valious` for example, the frequency analysis is illustrated in Figure 52. The reconstruction signal by our design is showed in Figure 53. Therefore, in harmonic signal, non-interpolation should be used.

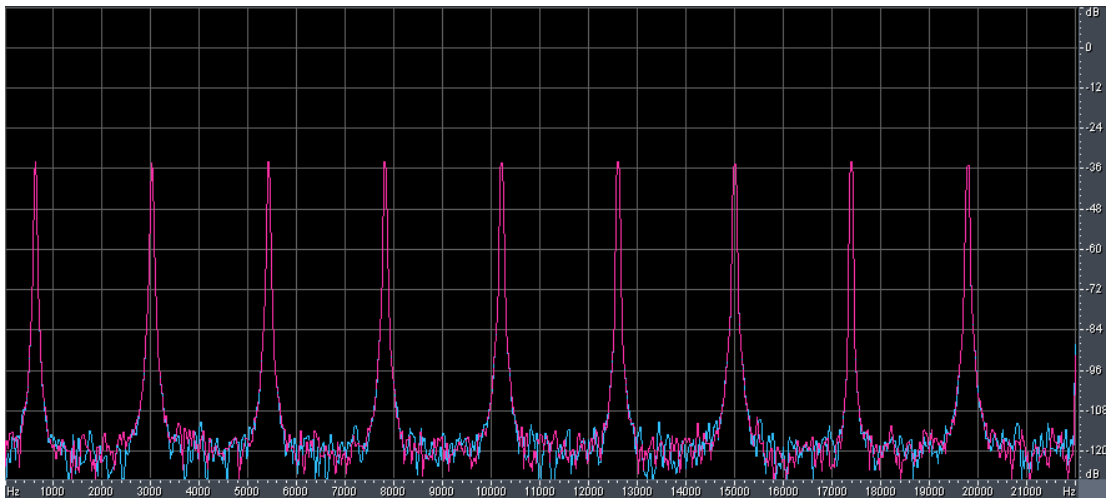


Figure 52: The frequency envelope of “sin_9kind_valious”.

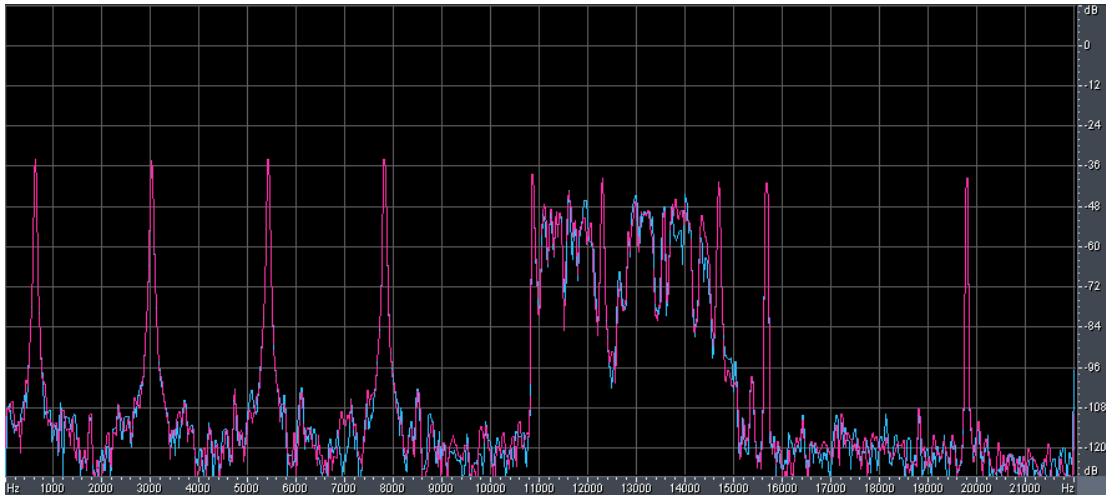


Figure 53: Noise floor overflow due to interpolation mode.

7.4 Subjective Quality Measurement

After the objective quality measurement, we perform subjective listening test to verify the quality improvement and possible risk of proposed methods in this thesis. The subjective quality measurement bases on MPEG test tracks, and use the tool called “MUSHRA” to be an assistant. There are three coding methods compared. The first one is NCTU HE-AAC with no any cuts in T/F grid, the second one is NCTU HE-AAC with uniform 7 cuts in T/F grid, and the final one is our proposed design. The testing bit rate is 80 kbps and the result is illustrated in Figure 54.

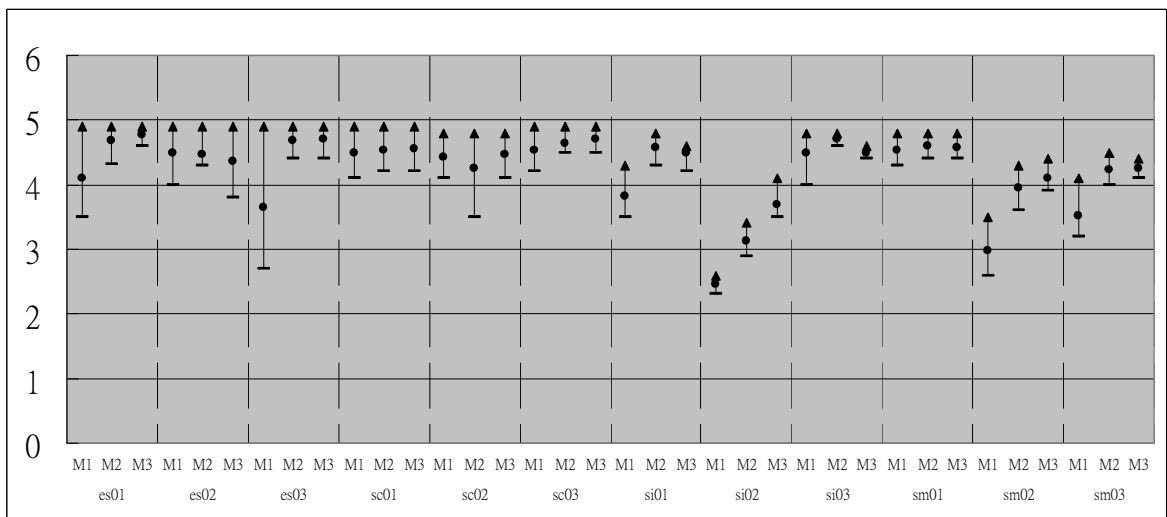


Figure 54: The result of subjective test at 80 kbps.

Summary

The result of subjective quality measurement is consistent with the result of objective quality measurement. In the speech signal (es01, es02, and es03), M1 is far

behind to M2 and M3, because this kind of signal needs finer resolution of T/F grid. On the other hand, the three methods have almost the same quality for sc01, sc02 and sc03, because in such signals, the contents of low bands are very similar to high bands, consequently, coarse resolution is enough. Especially, in sc02, M2 is worse than the other two methods due to the immoderate cutting and consumes too many bits. si02 is a critical signal to prove the advantages of our design. In si02, there are many “attacks” in time domain; therefore, the number and location of time borders in T/F grid are very important. Among three coding methods, M1 has no enough resolution, and M2 is lack of precision of time borders. However, our proposed DP design can handle this signal very well. si03 is a harmonic signal, and therefore, M3 is worse than M2 due to noise floor overflow by interpolation. To summarize, the proposed method perform well in subjective quality and conforms to the objective measurement.

7.5 Objective Quality Measurement with Existing Codecs

In this section, we compare NCTU HE-AAC with Coding Technologies 7.0.5 [12] and Nero 7 [13] at different bit rates.

| Bit Rate | 80 kbps | | |
|-----------------------|---------|---------------------------|-------------|
| Coding Methods | Nero 7 | Coding Technologies 7.0.5 | NCTU HE-AAC |
| Es01 | -1.88 | -0.78 | -0.67 |
| Es02 | -1.48 | -0.97 | -0.59 |
| Es03 | -2.07 | -0.8 | -0.68 |
| Sc01 | -2.29 | -1.06 | -0.96 |
| Sc02 | -2.57 | -1.54 | -1.08 |
| Sc03 | -2.87 | -1.17 | -1.11 |
| Si01 | -2.46 | -2.3 | -1.59 |
| Si02 | -2.37 | -1.02 | -1 |
| Si03 | -2.38 | -1.9 | -1.6 |
| Sm01 | -2.77 | -1.97 | -1.55 |
| Sm02 | -2.4 | -2.29 | -1.52 |
| Sm03 | -2.84 | -1.27 | -1.3 |
| Average | -2.365 | -1.4225 | -1.1375 |
| Sample Rate: 44100 Hz | | | |

Table 10: The objective quality measurement among different codecs at bit rate 80 kbps.

| Bit Rate | 64 kbps | | |
|-----------------------|---------|---------------------------|-------------|
| Coding Methods | Nero 7 | Coding Technologies 7.0.5 | NCTU HE-AAC |
| Es01 | -2.36 | -1.01 | -0.93 |
| Es02 | -1.76 | -1.49 | -0.82 |
| Es03 | -2.26 | -1.05 | -0.96 |
| Sc01 | -3.05 | -1.65 | -1.56 |
| Sc02 | -3.01 | -2.34 | -1.65 |
| Sc03 | -3.19 | -1.65 | -1.58 |
| Si01 | -3.02 | -2.84 | -1.93 |
| Si02 | -2.87 | -1.58 | -1.35 |
| Si03 | -2.94 | -2.19 | -2.04 |
| Sm01 | -3.21 | -2.73 | -2.12 |
| Sm02 | -2.74 | -2.74 | -2.18 |
| Sm03 | -3.25 | -1.71 | -1.64 |
| Average | -2.805 | -1.915 | -1.5633 |
| Sample Rate: 44100 Hz | | | |

Table 11: The objective quality measurement among different codecs at bit rate 64 kbps.

| Bit Rate | 48 kbps | | |
|----------------|---------|---------------------------|-------------|
| Coding Methods | Nero 7 | Coding Technologies 7.0.5 | NCTU HE-AAC |
| Es01 | -2.42 | -2.24 | -1.48 |
| Es02 | -2.42 | -2.65 | -1.28 |
| Es03 | -2.56 | -2.42 | -1.67 |
| Sc01 | -3.64 | -2.65 | -2.34 |
| Sc02 | -3.61 | -3.1 | -2.47 |
| Sc03 | -3.49 | -3.05 | -2.29 |
| Si01 | -3.55 | -3.42 | -2.68 |
| Si02 | -3.31 | -2.73 | -2.26 |

| | | | |
|-----------------------|--------|---------|---------|
| Si03 | -3.22 | -2.86 | -3.15 |
| Sm01 | -3.52 | -3.61 | -3.17 |
| Sm02 | -3.37 | -3.3 | -3.25 |
| Sm03 | -3.59 | -3.14 | -2.2 |
| Average | -3.225 | -2.9308 | -2.3533 |
| Sample Rate: 44100 Hz | | | |

Table 12: The objective quality measurement among different codecs at bit rate 48 kbps.

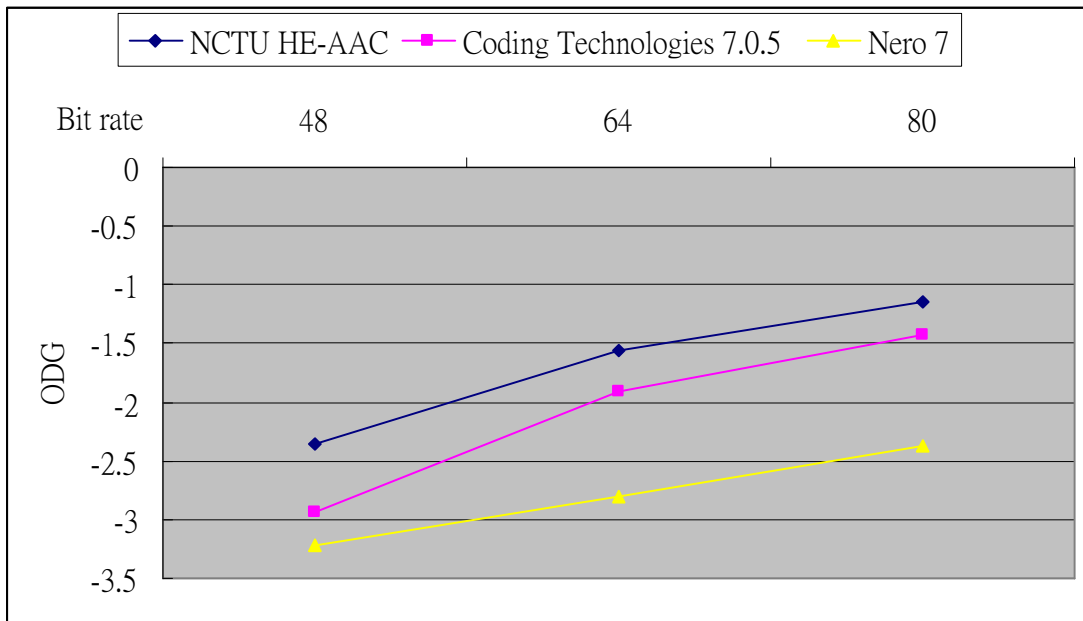


Figure 55: The ODG-bit rate comparison curve among different codecs.

Summary

Through observing the comparison results at different bit rates in Figure 55, our NCTU HE-AAC is superior to other encoders in average.

7.6 Objective Quality Measurement by SBR range with

Error Concealment

The purpose of the experiment is to compare the objective quality between HE-AAC with and without error concealment. The objective quality testing bases on

MPEG test tracks and three different bit rates: 80 kbps, 64 kbps and 48 kbps. The result is described below.

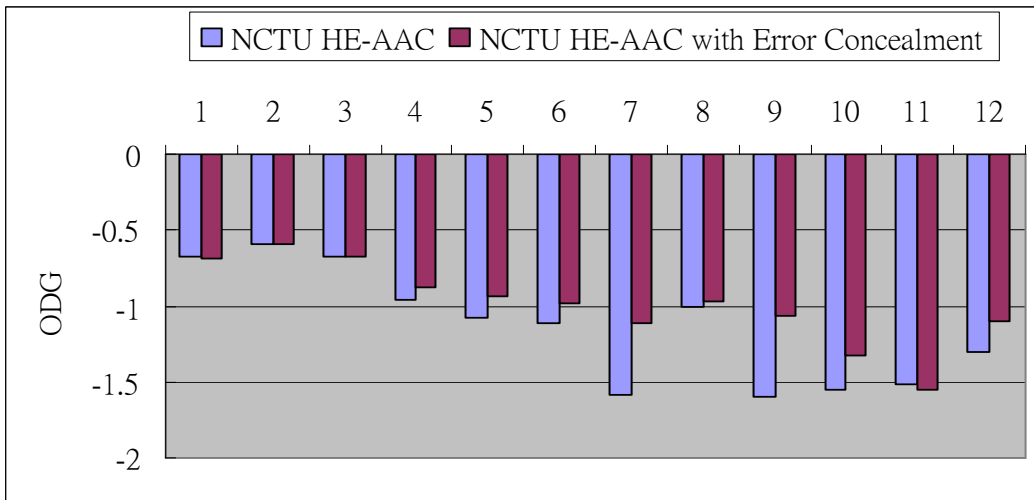


Figure 56: The result of objective quality measurement for error concealment based on MPEG test tracks at bit rate 80 kbps.

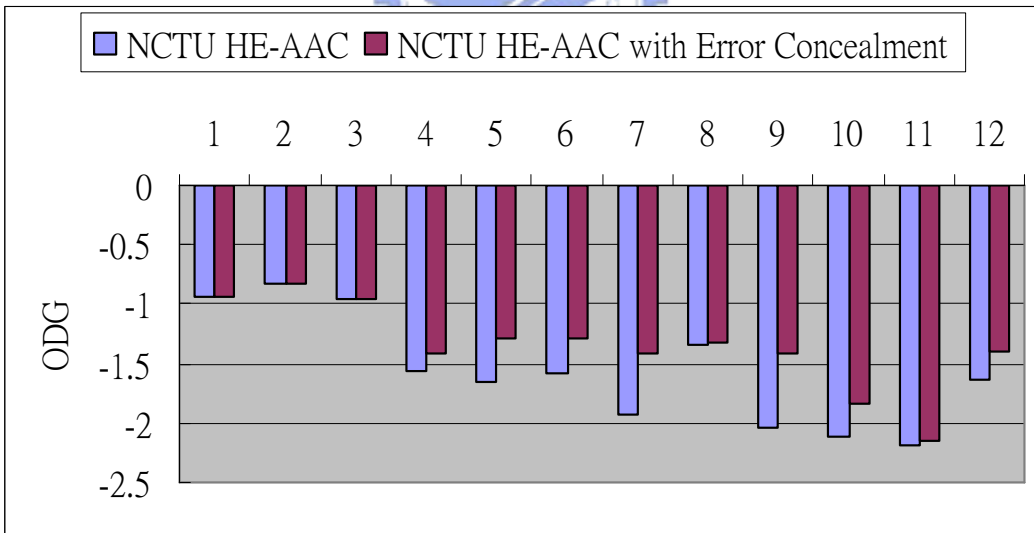


Figure 57: The result of objective quality measurement for error concealment based on MPEG test tracks at bit rate 64 kbps.

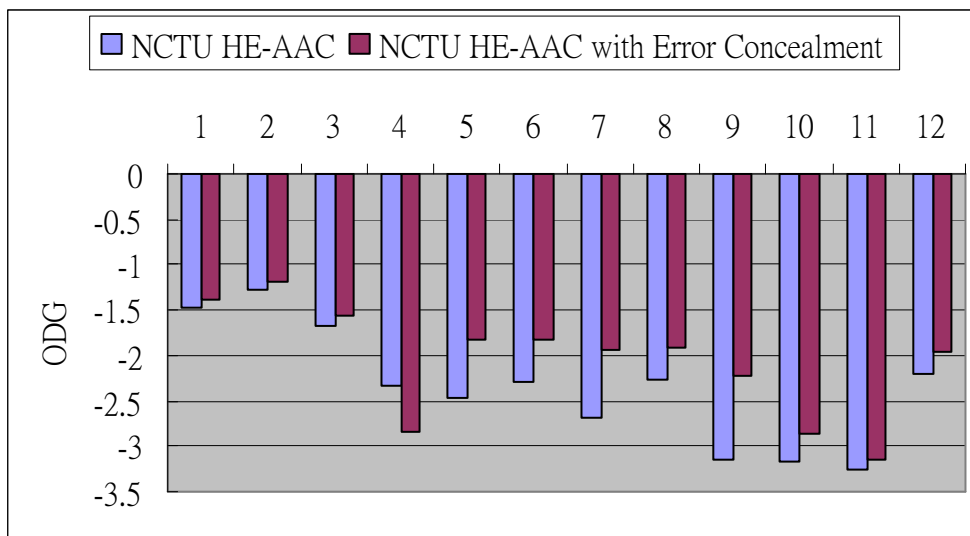


Figure 58: The result of objective quality measurement for error concealment based on MPEG test tracks at bit rate 48 kbps.

Summary

The ODG and bit rate comparison curve is illustrated in Figure 59. It is easy to see that with error concealment, the objective quality improves a lot at all bit rates. However, at bit rate 48 kbps, the ODG of sc01 with error concealment is worse than that without it. The reason is analyzed below. The frequency envelope of sc01 without error concealment at bit rate 48 kbps is described in Figure 60. Due to the insufficient bits, there is a wide spectral valley in AAC parts. On the other hand, error concealment repairs the spectral valley by estimating the energy of neighborhood. In this case, the reconstruction envelope from error concealment is different from the original one as illustrated in Figure 61.

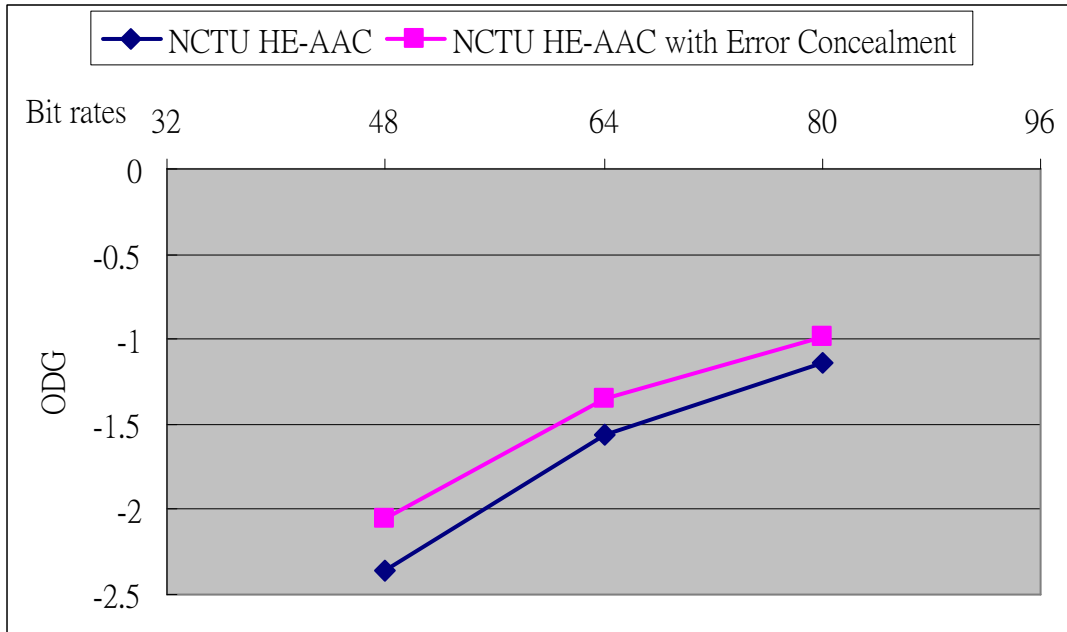


Figure 59: The ODG and bit rate comparison curve for NCTU HE-AAC with and without error concealment.

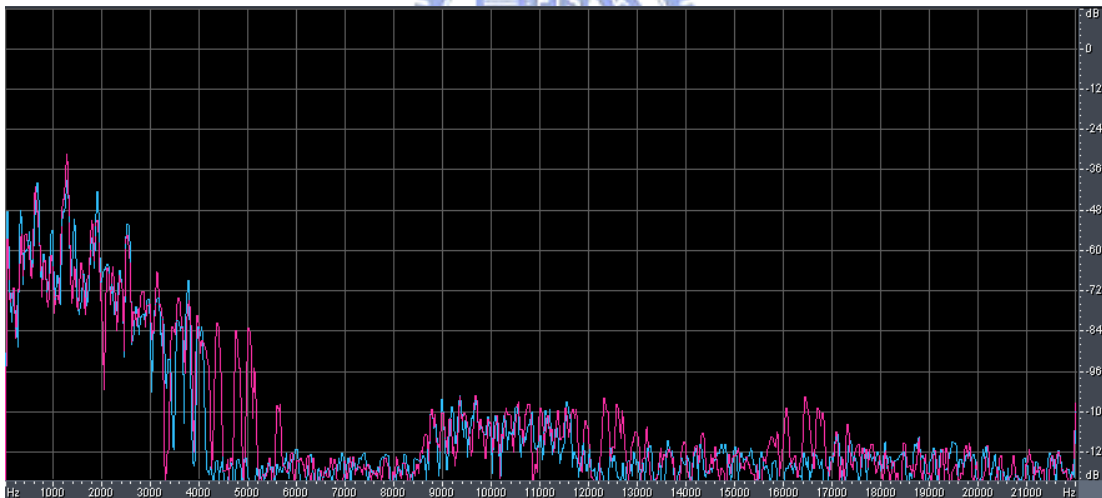


Figure 60: The frequency envelope of “sc01” by NCTU HE-AAC without error concealment.

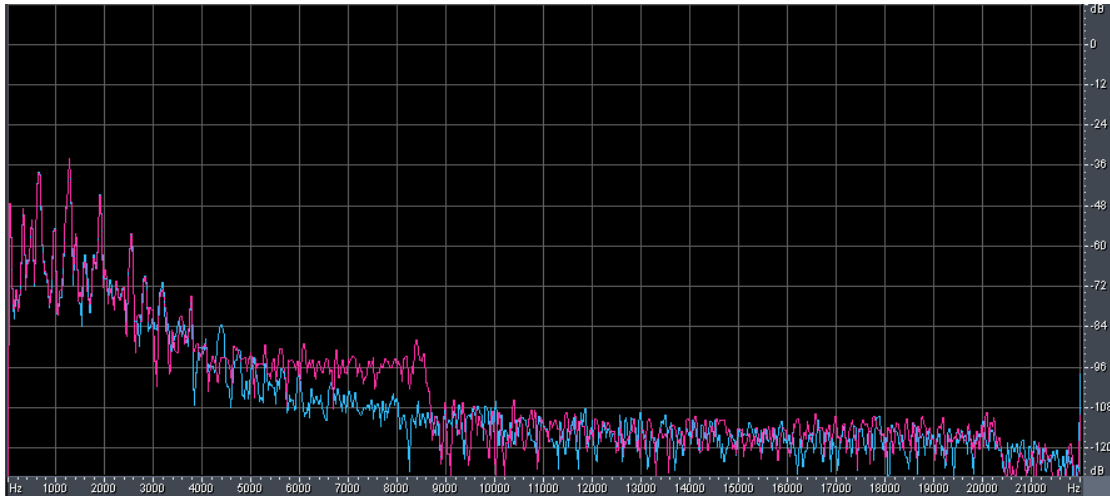


Figure 61: The frequency envelope of “sc01” by NCTU HE-AAC with error concealment. The blue line represents the original signal, and the red one is coded signal.



Chapter 8

Conclusion and Future Works

This thesis has proposed an efficient design of T/F grid in HE-AAC to enhance compression quality and maintain bit rate constraint. By introducing a reconstruction error measurement, DSR, to estimate the quality of T/F grid, the optimal format of T/F grid can be found out. Further, in order to reduce the searching complexity, we have proposed an efficient algorithm with dynamic programming. On the other hand, the factor of bit-consuming is also taken into consideration in our designs. Through taking account of bit-consuming and corresponding quality at the same time, our proposed T/F grid search algorithm not only can find the optimal format of T/F grid under the bit rates constraint, but be flexible at different bit rates.

The SBR range decision is another critical issue in the design of HE-AAC encoder. This paper proposes adaptive and fixed range methods, and discusses feasibility of both methods. Adaptive range method has better quality in SBR, but introduces some artifacts and bit-overhead. On the other hand, fixed range method lacks flexibility, and the spectral valley phenomenon may occur. Therefore, for preventing the spectral valley and preserving the flexibility, this thesis proposes a fixed range method with error concealment. With the help of error concealment, not only the bit-economy and bandwidth extension can be maintained, but also the spectral valley phenomenon can be eliminated.

Objective experiments based on the recommendation system by ITU-R Task Group 10/4 have been conducted on intensive tracks to prove the quality improvement, flexibility and efficiency of our new T/F grid design. Through subjective measure on tremendous music database, the quality in perceptual hearing of our proposed design is also verified. These experiments have shown that our T/F grid design in HE-AAC could well fit the encoders for various bit rates and preferred scenario.

Future Work

The estimated reconstruction error used in this paper is absolute error. If the masking of psychoacoustics model can be applied, and the estimated error can be more conformed to perceptual hearing and the resulting T/F grid can be more efficient.

This thesis proposes a T/F grid design through dynamic programming to reduce the complexity of searching. However, the complexity of dynamic programming is

still high. Therefore, we should take this T/F grid design as a model, and find another near optimal but more efficient search method. Finally, in the extensive experiments, there are still some problems and artifacts needs to be considered.

References

- [1] ISO/IEC JTC1/SC2/WGII MPEG, International Standard ISO 11172-3 “Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s.”
- [2] ISO/IEC 14496-3:1999, “Information Technology–Coding of Audiovisual objects, Part3: Audio.”
- [3] ISO/IEC, “Text of ISO/IEC 14496-3:2001/FDAM1, Bandwidth Extension,” ISO/IEC JTC1/SC29/WG11/N5570, March 2003, Pattaya, Thailand.
- [4] 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.
- [5] M. Wolters, K. Kjörling, D. Homm, H. Purnhagen, “A closer look into MPEG-4 High Efficiency AAC,” at the 115th AES Convention, New York, USA, October 10–13, 2003.
- [6] 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.
- [7] C.M. Liu, L.W. Chen, H.W. Hsu, and W.C. Lee, “Bit Reservoir Design for HE-AAC,” at the 118th AES Convention, Barcelona, Spain, May 28~31, 2005.
- [8] MP3Pro, website <http://www.mp3prozone.com/>
- [9] H. Purnhagen: “Low Complexity Parametric Stereo Coding in MPEG-4”, 7th International Conference on Audio Effects (DAFX-04), Naples, Italy, October 2004.
- [10] E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegård: “Low complexity parametric stereo coding”, Proc. 116th AES convention, Berlin, Germany, 2004, Preprint 6073.
- [11] 3GPP, website <http://www.3gpp.org/>.
- [12] Coding Technologies, aacPlusEval v2 Evaluation Package, Version 7.0.5, website <http://portal.codingtechnologies.de/eval/aacPlusEval/>.

- [13] Nero, website <http://www.nero.com>.
- [14] C.M. Liu, W.C. Lee, and H.W. Hsu, "High Frequency Reconstruction for Band-limited Audio Signals," Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03), London, UK, September 8-11, 2003.
- [15] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Audio Codec audio processing functions; "Enhanced aacPlus General Audio Codec General Description (Release 6)"
- [16] C.M. Liu, W.C. Lee, and H.W. Hsu, "High Frequency Reconstruction by Linear Extrapolation," presented at the 115th AES Convention, New York, October 10-13, 2003.
- [17] H.W. Hsu, C.M. Liu, W.C. Lee, and Z.W. Li, "Audio Patch Method in MPEG-4 HE-AAC Decoder," presented at the 117th AES Convention, San Francisco, October 28-31, 2004.
- [18] Coding Technologies, aacPlusEval v2 Evaluation Package, Version 7.0.5, website <http://portal.codingtechnologies.de/eval/aacPlusEval/>
- [19] Lars Gustaf Liljeryd, Kristofer Kjørling, Per Ekstrand, Fredrik Henn, "Efficient Spectral Envelope Coding Using variable Time/Frequency Resolution and Time/Frequency Switching", US Patent, US 2006/0031065 A1.
- [20] NCTU-AAC, website <http://psplab.csie.nctu.edu.tw/projects/index.pl/nctu-aac.html> .
- [21] NCTU-HEAAC, website <http://psplab.csie.nctu.edu.tw/projects/index.pl/nctu-heaac.html>
- [22] PSPLAB audio database, website <http://psplab.csie.nctu.edu.tw/projects/index.pl/testbitstreams.html>
- [23] ITU Radiocommunication Study Group 6, "Draft Revision to Recommendation ITU-R BS.1387- Method for objective measurements of perceived audio quality".
- [24] G. Stoll and F. Kozamernik, "EBU Listening Tests on Internet Audio Codecs", EBU TECHNICAL REVIEW, June. 2000.
- [25] International Telecommunications Union, Radiocommunication Sector BS.1534, "Method for the Subjective Assessment of Intermediate Quality Level Coding Systems – General requirements", Geneva 2001.