國立交通大學

資訊科學與工程研究所

碩 士 論 文

MPEG-4 HE-AAC 中之高頻調整模組設計

Design for High Frequency Adjustment Module in

MPEG-4 HE-AAC

研 究 生：楊詠成

指導教授：劉啟民 教授

李文傑 博士

中 華 民 國 九 十 五 年 六 月

MPEG-4 HE-AAC 中之高頻調整模組設計

Design for High Frequency Adjustment Module in MPEG-4 HE-AAC

研 究 生：楊詠成　　Student：Yung-Cheng Yang

指導教授：劉啟民　　Advisor：Dr. Chi-Min Liu

李文傑　　Dr. Wen-Chieh Lee

國 立 交 通 大 學

資 訊 科 學 與 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

June 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年六月

MPEG-4 HE-AAC 中之高頻調整模組設計

學生：楊詠成　　指導教授：劉啟民 博士

李文傑 博士

國 立 交 通 大 學

資 訊 科 學 與 工 程 研 究 所

中文論文摘要

高頻調整模組是 MPEG-4 HE-AAC 中的核心模組，其目標在於回復高頻重建訊號的音調度，使其如同原始訊號一般。兩個決定性的問題爲，如何準確的測量訊號的音調度，與如何決定共用的控制參數。控制參數之計算與音調度相關，其值將決定在解碼端的增益大小與額外補償之能量大小，換句話說，重建訊號的品質直接受到高頻調整模組的影響。

本論文提出一有效率的方法來測量子頻帶訊號的音調度，使用線性預測並基於 Levinson-Durbin 演算法以自動調整預測等級，使其符合不同的訊號內容。進而探討因控制參數之共用所造成的音訊失真，並提出如何有效地決定控制參數之方法。

Design for High Frequency Adjustment Module in MPEG-4 HE-AAC

Student: Yung-Cheng Yang Advisor: Dr. Chi-Min Liu

Dr. Wen-Chieh Lee

Institute of Computer Science and Engineering
College of Computer Science
National Chiao Tung University

ABSTRACT

High frequency adjustment module is the kernel module of spectral band replication (SBR) in MPEG-4 HE-AAC. The objective of high frequency adjustment is to recover the tonality of reconstructed high frequency signal as the original tonality. There are two crucial issues, the accurate measurement of tonality and the decision of shared control parameters. Control parameters, which are extracted according to signal tonality, will be used to determine gain control and energy level of additional components in decoder part. In other words, the quality of the reconstructed signal will be directly related to the high frequency adjustment module.

In this thesis, an efficient method based on Levinson-Durbin algorithm is proposed to measure the tonality by linear prediction approach with adaptive orders to fit the different subband contents. Furthermore, the artifact due to the sharing of control parameter is investigated and the efficient decision criterion of control parameter is proposed.

致謝

感謝劉啓民老師與李文傑博士兩年來的栽培與指導，使我逐步了解作研究的方法。感謝實驗室的楊宗翰、許瀚文學長，在實作方法與理論基礎上的協助與指導，讓我在專業知識及研究方法上獲得啓發。感謝李侃峻、唐守宏、張家銘同學，以及學弟曾信耀和胡正倫平常給予的協助與討論，使我獲益良多，讓我能夠如期完成論文。也感謝系上所有的老師與同學，在課業上給予的協助。

最後感謝我的父母家人與所有的朋友們，在研究所兩年生活中給予我的支持與鼓勵，使我能夠順利的完成學業，在此致上由衷的感謝。

Contents

Contents ..... **i**
Figure List ..... **ii**
Table List ..... **iv**
Chapter 1 Introduction **1**
Chapter 2 Backgrounds **2**
2.1 SBR Overview 2
2.2 High Frequency Adjustment in SBR Decoder 5
2.2.1 HF Adjustment for Noise-demand Grid 6
2.2.2 HF Adjustment for Tone-demand Grid 7
Chapter 3 Control Parameter Extraction **8**
3.1 Problem Definition 8
3.2 Reconstructed T/N dB Difference 10
3.2.1 T/N dB Difference for Noise-demand Grid 10
3.2.2 T/N dB Difference for Tone-demand Grid 11
3.3 Optimal Solution 12
Chapter 4 Tonality Measurement **14**
4.1 Scheme in 3GPP 14
4.2 Modified Levinson-Durbin Algorithm 15
4.3 Tonality Measurement of Inverse Filtered Signal 16
4.4 T/N dB Difference Measurement 17
Chapter 5 Artifacts **18**
5.1 Noise Overflow 18
5.2 Tonal Spike 19
5.3 Noise Floor Correction 20
Chapter 6 Experimental Results **22**
6.1 Experiment Environment 22
6.2 Objective Quality Measurement in MPEG Test Tracks 24
6.3 Objective Quality Measurement in Music Database 29
6.4 Subjective Quality Measurement 37
Chapter 7 Concluding Remark **38**
References **41**

Figure List

Figure 1: Basic principle of SBR. 2
Figure 2: Block diagram of HE-AAC encoder. 3
Figure 3: Block diagram of HE-AAC decoder. 4
Figure 4: Flow chart of three phases in HF adjustment process. 5
Figure 5: HF adjustment process of a noise-demand high resolution grid with three subbands b_0, b_1, b_2 6
Figure 6: HF adjustment process of a noise-demand high resolution grid with three subbands b_0, b_1, b_2. 7
Figure 7: Trade-off between tonal and noise compensation. 8
Figure 8: Flow chart of the MLD algorithm 16
Figure 9: Noise overflow phenomenon 18
Figure 10: Tonal spike phenomenon 19
Figure 11: ODG comparison of MPEG test tracks at bit rate 80kbps. 25
Figure 12: ODG comparison of MPEG test tracks at bit rate 64kbps. 26
Figure 13: ODG comparison of MPEG test tracks at bit rate 48kbps. 27
Figure 14: Average ODG in music database at bit rate 80kbps. 30
Figure 15: Average ODG in music database at bit rate 64kbps. 30
Figure 16: Average ODG in music database at bit rate 48kbps. 30
Figure 17: spectrum comparison of test track (c44.wav) in sjeng set. 31
Figure 18: spectrum comparison of test track (impulse_m20_0db.wav) in TonalSignals set. 32
Figure 19: spectrum comparison of test track (ms-test.wav) in TonalSignals set. 32
Figure 20: spectrum comparison of test track (sin_9kind_valious.wav) in TonalSignals set. 33
Figure 21: spectrum comparison of test track (sin_600_19800_9div_m20_0db.wav) in TonalSignals set. 33
Figure 22: spectrum comparison of test track (sq_300_625_1k_10k_15k_m20.wav) in TonalSignals set. 34
Figure 23: spectrum comparison of test track (test1152-lame.wav) in TonalSignals set. 34
Figure 24: spectrum comparison of test track (sin_300_625_1k_5k_10k_15K_20k_0db.wav) in TonalSignals set. . 35
Figure 25: spectrum comparison of test track (sin_300_625_1k_5k_10k_15K_20k_m6db.wav) in TonalSignals set.

.....35
Figure 26: spectrum comparison of test track (sin_sweep_30_20k_0db.wav) in TonalSignals set. 36
Figure 27: Results of listening test on MPEG test tracks at bit rate 80kbps. 37

Table List

Table 1: Summary of T/N energy analysis for noise-demand grid. 10
Table 2: Summary of T/N energy analysis for tone-demand grid. 11
Table 3: Relation between the given control parameter and the resultant T/N dB difference. 12
Table 4: Relation between the given T/N dB difference and the local optimal control parameter. 12
Table 5: Noise floor energy in HF reconstruction 20
Table 6: The twelve tracks recommended by MPEG 24
Table 7: ODG for proposed methods on MPEG test tracks at bit rate 80 kbps. 25
Table 8: ODG for proposed methods on MPEG test tracks at bit rate 64 kbps. 26
Table 9: ODG for proposed methods on MPEG test tracks at bit rate 48 kbps 27
Table 10: The PSPLab audio database [20] 29
Table 11: ODG comparison with existing codec at bit rate 80kbps 38
Table 12: ODG comparison with existing codec at bit rate 64kbps 39
Table 13: ODG comparison with existing codec at bit rate 48kbps 39

Chapter 1
Introduction

Spectral band replication (SBR) is a new audio coding enhancement tool to improve audio quality at low bit rates. The general advanced audio coding (AAC) combined with SBR is called high efficiency-AAC (HE-AAC) which has been adopted in MPEG-4 audio version 3. From the similarity between low frequency band and high frequency band, the main principle of SBR is to reconstruct the high frequency band signal by replicating from low frequency band and mixing with the generated tonal and noise signals, which only costs a few amount of side information provided by encoder. However, the prerequisite for perceptually consistent reconstruction is the accurate TNR measurement of the original signal. Some methods have been proposed to improve the accurateness of measurement. However, most of them need additional time-consuming calculation. For example, a direct estimation on frequency domain [1] is proposed to measure noise floor precisely, and an accurate method on time domain in [2] is given to capture tonal components. These methods are time-consuming due to the additional FFT and sinusoid analysis respectively. In this thesis, an efficient approach based on Levinson-Durbin algorithm [3] that can construct a lattice filter gradually, is proposed to decide the accurate prediction order adaptively. This thesis focuses on the two design issues of high frequency adjustment module, including tonality measurement and control parameters extraction.

The organization of this thesis is as follows: a brief overview of SBR tool and HF adjustment process in SBR decoder are given in Chapter 2. The affection of the accuracy of control parameter is discussed in Chapter 3, and a search method is also proposed to choose a shared optimal control parameter for a single noise-grid. Chapter 4 considers the artifacts due to the inaccurate tonality measurement, and proposes a modified Levinson-Durbin algorithm to measure TNR accurately. The two typical artifacts, noise overflow and tonal spike are discussed in Chapter 5, and a remedial method at the encoder end is proposed to prevent the noise overflow phenomenon. Finally, the experiments are conducted to check the quality performance in Chapter 6.

Chapter 2
Backgrounds

This chapter presents a brief overview of SBR and a closer look into the high frequency adjustment process in the SBR decoder.

2.1 SBR Overview

SBR (Spectral Band Replication) is a bandwidth extension tool used in combination with the AAC general audio codec. The significant improvement of performance is achieved by replicating the high frequency part of the spectrum. The replication is mainly a post-process performed in the decoder, which guided by a few amount of side information extracted from the encoder. The data rate is substantially reduced than the data rate required when using conventional AAC for coding the high frequency part. The basic principle of SBR is depicted in Figure 1.

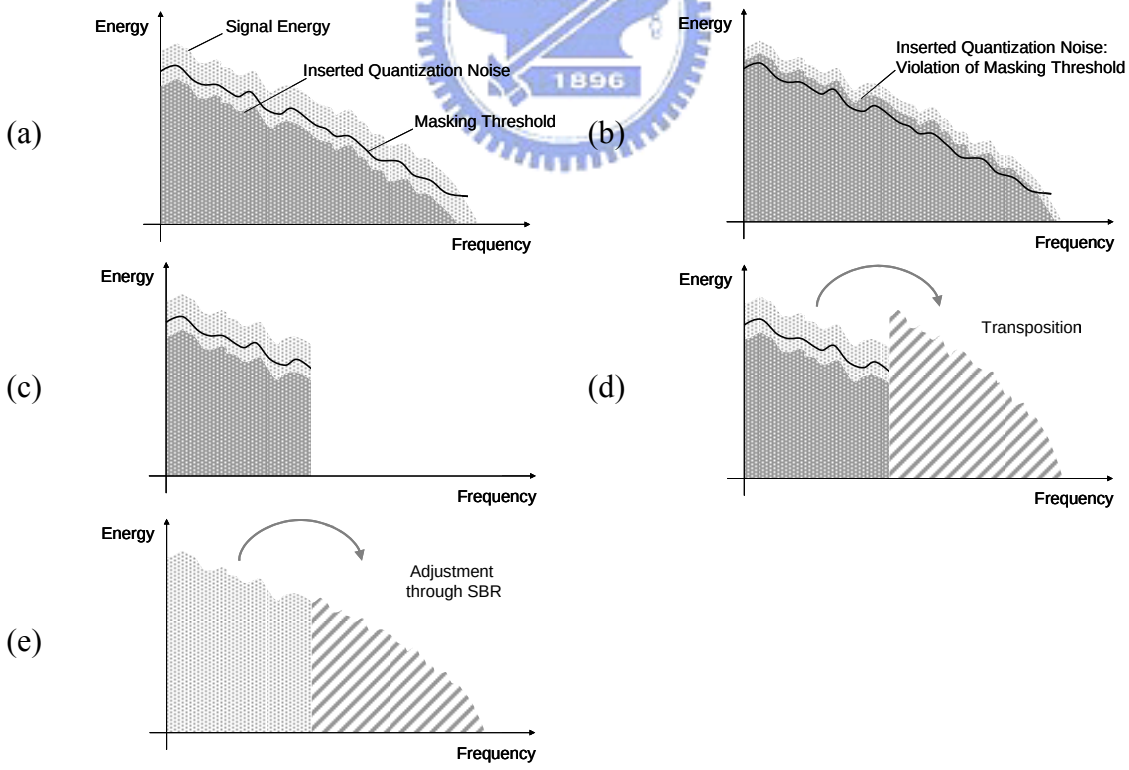


Figure 1: Basic principle of SBR.

In traditional perceptual audio coding, assuming a sufficient bit rate, the inserted quantisation noise will be kept under the masking threshold and therefore be inaudible (see Figure 1 a). At reduced bit rate, this masking threshold will be violated, the coding artifacts will be audible (see Figure 1 b). Thus, under the constraint of bit rate, usually the audio bandwidth is limited (see Figure 1 c). With SBR tool, the lower frequency part (from 0 to typically 5~13 kHz) is coded using a waveform coder (called 'core codec'; e.g. general AAC encoder). Additionally, a reconstruction of the high frequency part is done by transposition of the lower frequencies (see Figure 1 d), and a proper spectrum shaping which guided by side information is applied (see Figure 1 e).

In the encoder, the required side information which presents a parametric description of high frequency part of the spectrum is estimated given the original wide-band signal. The HE-AAC encoder can be roughly divided into modules as illustrated in Figure 2. At first, the time domain signal is fed to a down sampler, which provides a time domain signal at half sampling rate for the AAC core encoder. In parallel, the input signal is filtered by a 64-band QMF analyzer. All of the estimation in encoder is performed on QMF domain. The envelope extractor gives the time-frequency grid information and the average energy scale for each grid. The tonality estimator provides the tonality value of each QMF subband for further estimation. The inverse filter decides the inverse filtering mode which controls the high frequency whitening process. The parameter extractor delivers the control parameter to control the high frequency adjustment process.

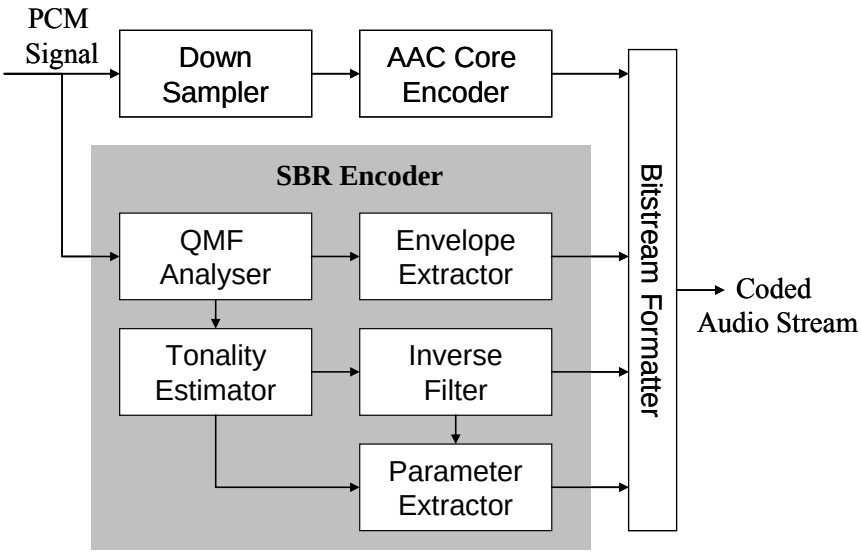


Figure 2: Block diagram of HE-AAC encoder.

In the decoder, the received bitstream is divided into the core coder bitstream and the SBR side information. The lower frequency band component is reconstructed by the underlying AAC decoder and filtered by a 32-band QMF analyzer to provide subband signal in the QMF domain for the SBR decoder. The SBR decoding process can be considered as replication and HF adjustment. The replication process generates high frequency component by patching consecutive QMF subbands from low frequency. The block diagram of the HE-AAC decoder is shown in Figure 3. A closer look into the high frequency adjustment process is presented in the next section.

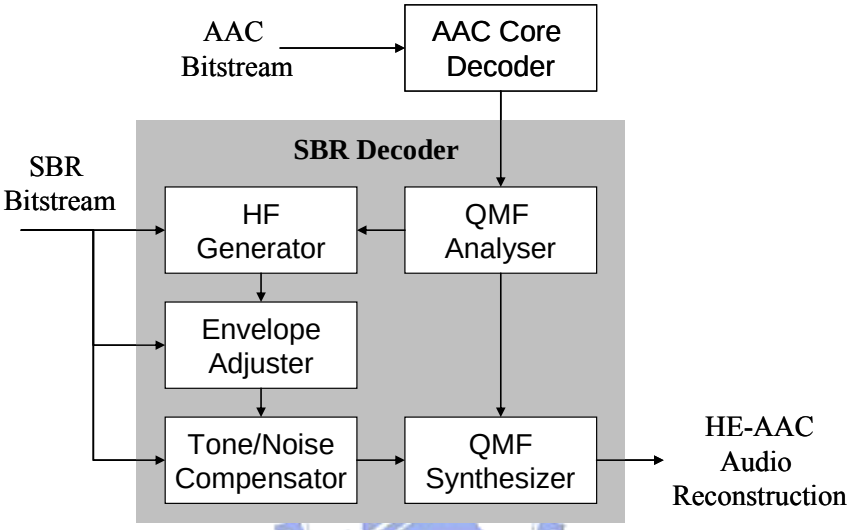


Figure 3: Block diagram of HE-AAC decoder.

2.2 High Frequency Adjustment in SBR Decoder

The mechanism of high frequency adjustment is to ensure that the reconstructed results perceptually as similar as possible through the proper spectrum shaping and additional tone/noise compensation. The whole HF adjustment process can be divided into three blocks as illustrated in Figure 3.

The HF generator generates high frequency signal by replicating the low band to high band and inverse filtered by a second order complex value predictor. The process in HF generator is referred to as generation phase and formulated in (1), assume the p^{th} subband in LF is mapped to b^{th} subband in HF, and n is the n^{th} sample in the QMF subband.

$$x_b[n] = x_p[n] - a_0 \cdot \alpha \cdot x_p[n-1] - a_1 \cdot \alpha^2 \cdot x_p[n-2], \quad (1)$$

where a_0 and a_1 are prediction coefficients, and α is the chirp factor determined by the inverse filtering mode. The output of HF generator is referred to as high frequency generation. After the generation phase, the envelope adjuster applies proper energy scaling on HF generation. This stage is referred to as scaling phase and the output of envelope adjuster is referred to as HF coordination. The last stage of HF adjustment is referred to as compensation phase, which the tone/noise compensator adds tonal or noise compensation to the HF coordination. The output of tone/noise compensator is referred to as HF reconstruction. The flow chart of three phases in HF adjustment process is illustrated in Figure 4. For the following explanation, we take a closer look into a high resolution grid which has the finest frequency resolution and a given time region.

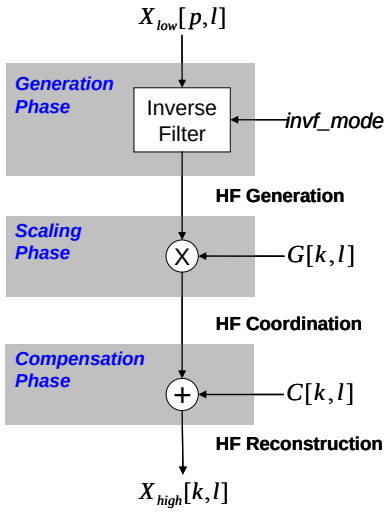


Figure 4: Flow chart of three phases in HF adjustment process.

2.2.1 HF Adjustment for Noise-demand Grid

If a high resolution grid does not have additional tone added, it is referred to as the noise-demand grid. In this mode, the magnitudes of each subband in the high resolution grid are adjusted by a gain control factor defined as

$$G_k^{ND} = \sqrt{\frac{E_k^o}{E_k^r} \cdot \frac{1}{1+Q_k}}, \tag{2}$$

where E_k^o and E_k^r are the averaged energies of the original high bands and the replicated low bands respectively in the k^{th} high resolution grid, and Q_k is the control parameter that determines the energy ratio between the replicated low bands and the decoded high bands. Also, the random noise is added to the HF bands with level defined as

$$C_k^n = E_k^o \cdot \frac{Q_k}{1+Q_k}. \tag{3}$$

Figure 5 illustrates the noise and the replicated signals allocated to the high resolution grid.

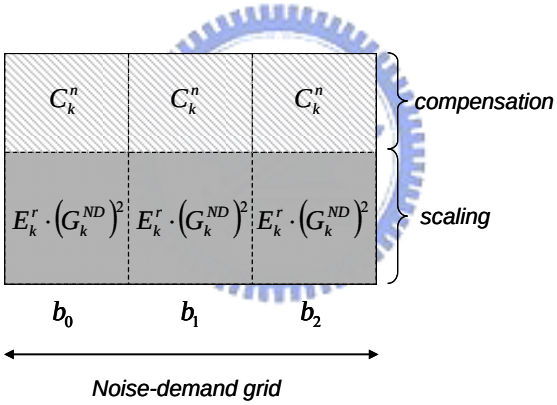


Figure 5: HF adjustment process of a noise-demand high resolution grid with three subbands b_0, b_1, b_2.

2.2.2 HF Adjustment for Tone-demand Grid

If a high resolution grid needs additional tone added, it will be referred to as the tone-demand grid and adapts the second mode. The main difference of the two modes is the definition of the gain factor and compensation amount, but they are both controlled by the control parameter. The gain factor for the tone-demand grid is defined as

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

For the compensated tone, the energy amount is defined as

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

Besides the compensated tone is added into the middle subband in the k^{th} high resolution grid, the remained subbands are compensated with random noise with energy level defined in (3). Figure 6 illustrates the reconstructed high bands in the tone-demand grid.

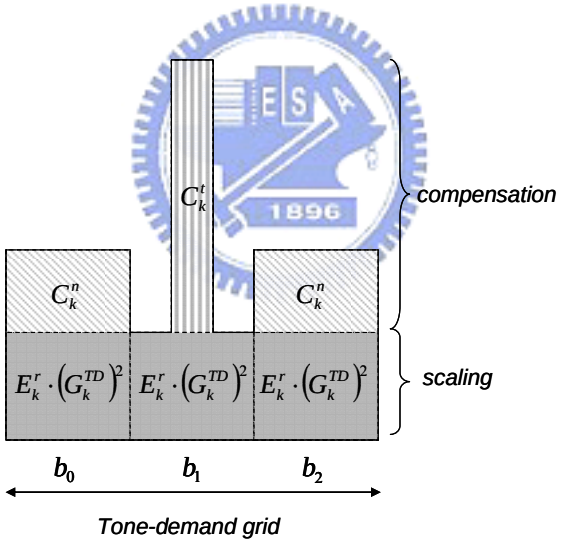


Figure 6: HF adjustment process of a noise-demand high resolution grid with three subbands b_0, b_1, b_2.

Chapter 3
Control Parameter Extraction

This chapter presents the analysis for the accuracy of control parameter extraction, and a search method is also proposed to choose an optimal control parameter for several high resolution grids to share.

3.1 Problem Definition

According to the definition of HF adjustment in section 2.2, the scaling and compensation process are both controlled by the control parameter. It is obvious that the perceptual quality of the SBR audio mainly depends on the reconstructed tone-noise content. However, both the magnitude of tonal and noise compensation are controlled by the same control parameter in an individual noise grid, which has frequency resolution four times coarser than a high resolution grid. In other words, the sharing problem of control parameter can be considered as a trade-off between tonal and noise compensation as illustrated in Figure 7. Besides, the number and the location of the added tones are constrained by the SBR syntax. Hence, it is difficult to recover the exact T/N content.

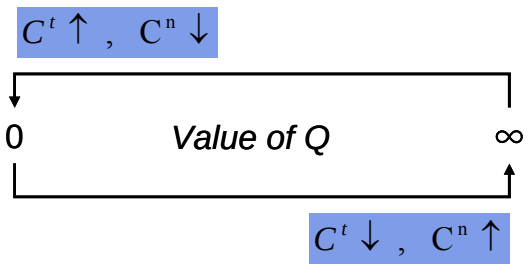


Figure 7: Trade-off between tonal and noise compensation.

The related works such as in [2], has demonstrated the suggested method in the standard [4] only provide a roughly approximation of original TNR. Also, the optimal solution of control parameter for a single subband is proposed, but the sharing problem between noise-demand and tone-demand is neglected. Under the two considerations, instead of maintaining the TNR of original signal, our proposed method determines an optimal parameter to maintain the minimum distortion of the differences between the tone and noise floor component in dB domain.

In the next subsection, the reconstructed difference of dB value between tonal component and noise floor of the reconstructed high resolution grid under a chosen control parameter is formulated, and then an optimal parameter is decided by minimizing the distortion function. For convenience, the notations referring to as the energies of noise floor and tone component in the three phases are defined as follows.

Generation Phase

E_k^n: Energy of noise floor in the k^{th} high resolution grid without any adjustment.

E_k^t: Energy of dominate tone component in the k^{th} high resolution grid without any adjustment.

Scaling Phase

$E_k'^n$: Energy of noise floor in the k^{th} high resolution grid after gain scaling.

$E_k'^t$: Energy of dominate tone component in the k^{th} high resolution grid after gain scaling.

Compensation Phase

$E_k''^n$: Energy of noise floor in the k^{th} high resolution grid after compensation.

$E_k''^t$: Energy of dominate tone component in the k^{th} high resolution grid after compensation.

3.2 Reconstructed T/N dB Difference

To measure the resultant dB difference of tonal component and noise floor in HF reconstruction, the associated changes in the three phases are simulated in this section.

3.2.1 T/N dB Difference for Noise-demand Grid

In the generation phase, E_k^n and E_k^t are estimated from the replicated low bands by linear prediction method that is explained in Chapter 4. After the scaling phase for the noise-demand grid, the energies of noise floor and tone component is modified as (6)(7) by the gain scaling factor.

$$E_k'^n = E_k^n \cdot \left(G_k^{ND}\right)^2. \tag{6}$$

$$E_k'^t = E_k^t \cdot \left(G_k^{ND}\right)^2. \tag{7}$$

Through the compensation phase, the T/N energies become

$$E_k''^n = E_k'^n + C_k^n. \tag{8}$$

$$E_k''^t = E_k'^t. \tag{9}$$

In (9), because of the added noise energy is relatively smaller compared to the tonal component energy $E_k'^t$, the effect of the adding compensation of noise can be ignored. At last, from (8) and (9) the resultant difference Δ_k^{ND} between tonal component and noise floor of the reconstructed high resolution grid in dB domain is derived as (10) under the given control parameter Q_k.

$$\Delta_k^{ND}(Q_k) = f_{dB}\left(E_k''^t\right) - f_{dB}\left(E_k''^n\right), \tag{10}$$

where $f_{dB}(\cdot)$ is the dB domain mapping. Substituting (2), (3), and (6)~(9) into (10) yields

$$\Delta_k^{ND}(Q_k) = f_{dB}\left(E_k^t\right) - f_{dB}\left(E_k^n + Q_k \cdot E_k^r\right). \tag{11}$$

The above T/N energy analysis through the three phases is summarized in Table 1.

Table 1: Summary of T/N energy analysis for noise-demand grid.

	3 phase		
	Generation	Scaling	Compensation
Tonal energy	E_k^t	$E_k'^t = E_k^t \cdot \left(G_k^{ND}\right)^2$	$E_k''^t = E_k'^t$
Noise floor energy	E_k^n	$E_k'^n = E_k^n \cdot \left(G_k^{ND}\right)^2$	$E_k''^n = E_k'^n + C_k^n$

3.2.2 T/N dB Difference for Tone-demand Grid

Similarly, by simulating the three phases, the resultant T/N difference for tone-demand grid can be estimated. After the gain scaling phase for the tone-demand grid, the energy of noise floor is modified as (12) by the gain factor.

$$E_k'^n = E_k^n \cdot \left(G_k^{TD}\right)^2 . \tag{12}$$

The tonal component obtained in the replicated low bands is ignored due to the assumption that the compensated tone is more dominated. After the compensation phase, the tonal component is revised as

$$E_k''^t = C_k^t . \tag{13}$$

As mentioned in section 2.2, the subbands without additional tone will be compensated with random noise. Hence, the noise floor becomes

$$E_k''^n = E_k'^n + C_k^n . \tag{14}$$

From (13) and (14), the resultant difference Δ_k^{TD} between tonal component and noise floor of the reconstructed high resolution grid in dB domain is derived as (15) under the given control parameter Q_k.

$$\Delta_k^{TD}(Q_k) = f_{dB}\left(E_k''^t\right) - f_{dB}\left(E_k''^n\right). \tag{15}$$

Substituting (4), (5) and (12)~(14) into (15) yields

$$\Delta_k^{TD}(Q_k) = f_{dB}\left(\frac{E_k^r}{E_k^n + E_k^r}\right) - f_{dB}(Q_k). \tag{16}$$

The above T/N energy analysis through the three phases is summarized in Table 2.

Table 2: Summary of T/N energy analysis for tone-demand grid.

	3 phase		
	Generation	Scaling	Compensation
Tonal energy	neglected	neglected	$E_k''^t = C_k^t$
Noise floor energy	E_k^n	$E_k'^n = E_k^n \cdot \left(G_k^{ND}\right)^2$	$E_k''^n = E_k'^n + C_k^n$

3.3 Optimal Solution

The measurement of the reconstructed T/N dB difference in a high resolution grid is formulated in the above section under the given control parameter Q_k. On the other hand, if the T/N dB difference of original signal in a high resolution grid is given, the local optimal control parameter Q_k can be derived from (17) or (18). The relation between control parameter and T/N dB difference is summarized in Table 3 and Table 4.

Table 3: Relation between the given control parameter and the resultant T/N dB difference.

	Given Q_k, the resultant T/N dB difference:	
Noise-demand	$\Delta_k^{ND}(Q_k) = f_{dB}(E_k^t) - f_{dB}(E_k^n + Q_k \cdot E_k^r)$	(11)
Tone-demand	$\Delta_k^{TD}(Q_k) = f_{dB}\left(\frac{E_k^r}{E_k^n + E_k^r}\right) - f_{dB}(Q_k)$	(16)

Table 4: Relation between the given T/N dB difference and the local optimal control parameter.

	Given Δ_k, the local optimal control parameter:	
Noise-demand	$Q_k^{ND} = \frac{1}{E_k^r}\left(\frac{E_k^t}{f_{dB}^{-1}(\Delta_k)} - E_k^n\right)$	(17)
Tone-demand	$Q_k^{TD} = \frac{E_k^r}{f_{dB}^{-1}(\Delta_k)\cdot(E_k^n + E_k^r)}$	(18)

In order to determine the optimal control parameter, the total distortion of the resultant dB differences in an individual noise grid is formulated as (19) under a common control parameter *Q*.

$$D(Q) = \sum_{k\in ND}\left(\Delta_k^{ND}(Q) - \Delta_k\right)^2 + \sum_{k\in TD}\left(\Delta_k^{TD}(Q) - \Delta_k\right)^2, \quad (19)$$

where Δ_k is the actual dB difference measured from the original HF subbands in the k^{th} high resolution grid. Hence, the optimal control parameter should be chosen to minimize the distortion function

$$Q^* = Arg\left\{\min_Q[D(Q)]\right\}. \quad (20)$$

According to the syntax of SBR, there are 31 candidates for the value of *Q*, and thus the optimal value can be found from the 31 possibilities. Substituting (11) and (16) into (19) results in

$$D(Q)=\sum_{k\in ND}\left\{f_{dB}\left(\frac{E_k^n}{E_k^r}+Q\right)-\left[f_{dB}\left(\frac{E_k^t}{E_k^r}\right)-\Delta_k\right]\right\}^2 + \sum_{k\in TD}\left\{f_{dB}(Q)-\left[f_{dB}\left(\frac{E_k^r}{E_k^n+E_k^r}\right)-\Delta_k\right]\right\}^2 . \quad (21)$$

From (22), it gives an approximation for $D(Q)$ in (21).

$$D(Q)\approx D'(Q)=\sum_{k\in ND}\left\{f_{dB}(R+Q)-\left[f_{dB}\left(\frac{E_k^t}{E_k^r}\right)-\Delta_k\right]\right\}^2 + \sum_{k\in TD}\left\{f_{dB}(R+Q)-\left[f_{dB}\left(\frac{E_k^r}{E_k^n+E_k^r}\right)-\Delta_k\right]\right\}^2 , \quad (22)$$

where R is defined as (23), and $|NG|$ is the number of all the high resolution grids in the noise grid.

$$R=\frac{\sum_{k\in ND}\left(\frac{E_k^n}{E_k^r}\right)+\sum_{k\in TD}(0)}{|NG|}=\frac{\sum_{k\in ND}\left(\frac{E_k^n}{E_k^r}\right)}{|NG|}. \quad (23)$$

Hence the optimal control parameter Q^* can be approximated by

$$Q^*\approx\tilde{Q}=Arg\left\{\min_Q\left[D'(Q)\right]\right\}, \quad (24)$$

that is

$$Q^*\approx\left\{f_{dB}^{-1}\left\{\frac{\sum_{k\in ND}\left[f_{dB}\left(\frac{E_k^t}{E_k^r}\right)-\Delta_k\right]+\sum_{k\in TD}\left[f_{dB}\left(\frac{E_k^r}{E_k^n+E_k^r}\right)-\Delta_k\right]}{|NG|}\right\}\right\}-R . \quad (25)$$

By searching the neighborhood of $\tilde{Q}$, it gives a faster method to find the optimal control parameter Q^* without covering the 31 candidates.

Chapter 4
Tonality Measurement

From Chapter 3, the extraction of the optimal control parameter significantly depends on the accurate estimation of the tonal component and noise-floor in the subbands. The improper measurement will result in unsuitable parameter and consequently may cause artifacts in the reconstructed HF spectrum. In this chapter, an efficient and accurate method for TNR measurement based on linear prediction approach is proposed to improve the standard second order predictor. Further more, considered the TNR will be affected by the inverse filtering process, an efficient method for the advanced TNR measurement is also proposed.

4.1 Scheme in 3GPP

A simple linear predictor of second order is suggested to measure the TNR of subband signals by the standard [4][5][6]. The prediction coefficients are calculated according to the covariance matrix. However, in order to precisely measure the actual TNR, the poles of the linear prediction filter must match the number of tones contained in the subband, and thus the prediction order should equal to the number of tones. If a subband contains more than two tones, it is obvious that the second order predictor would not be sufficient to capture all predictable components and cause missed detection. On the other hand, for a subband containing less than two tones, some noise components may be captured as predictable components by the second order predictor and start a false alarm. In the case of missed detection, the tonal energy is underestimated and the noise floor energy is overestimate, which will cause noise overflow in HF reconstruction. Oppositely, in the case of false alarm, the tonal energy is overestimated and the noise floor energy is underestimated, which may cause tonal spike. These typical artifacts in HE-AAC codec will be discussed in Chapter 5.

4.2 Modified Levinson-Durbin Algorithm

The proposed method is based on Levinson-Durbin algorithm that can construct a lattice filter incrementally until the demand prediction order is achieved. Because the tone component can be captured as the predictable components, the adaptive prediction order can be obtained when the prediction error does not decrease much with the increasing of the prediction order. The MLD algorithm is summarized as follows, and the flow chart of MLD algorithm is shown in Figure 8.

MLD Algorithm

Notation used in the algorithm:

(a) $\phi(i, j)$: autocorrelation of subband signal $x[n]$, $n=0\sim N$.

(b) p : the adaptive prediction order.

(c) $p_{\max}$: maximum number of prediction order.

(d) a_i^p : i^{th} prediction coefficient at order p.

(e) e^p : prediction error at order p.

(f) k^p : reflection coefficient at order p.

(g) Ψ : terminal threshold.

Input: $\phi(i, j)$.

Output: p and e^p.

The algorithm performs following steps for every subband:

Step 1: Initial conditions.

$$p = 1, \quad a_1^1 = \frac{\phi(0,1)}{\phi(0,0)},$$

$$e^1 = \phi(0,0)\left(1 - \frac{|\phi(0,1)|^2}{|\phi(0,0)|^2}\right).$$

Step 2: Increase order.

$$p \leftarrow p+1.$$

Step 3: Compute coefficients and prediction error.

I. $$k^p = \frac{1}{e^{p-1}}\left[\phi(0,p) - \sum_{i=1}^{p-1} a_i^{p-1}\phi(0,p-i)\right].$$

II. $$a_p^p = k^p,$$

III. $$a_i^p = a_i^{p-1} - k^p a_{p-1-i}^{*p-1}. \quad i = 1 \text{ to } p-1$$

IV. $$e^p = e^{p-1}\left(1 - |k^p|^2\right).$$

Step 4: Check terminal condition.

if $\frac{e^{p-1} - e^p}{e^{p-1}} > \Psi$ and $p < p_{\max}$, go to step 2.

Otherwise, $p \leftarrow p-1$, MLD is completed.

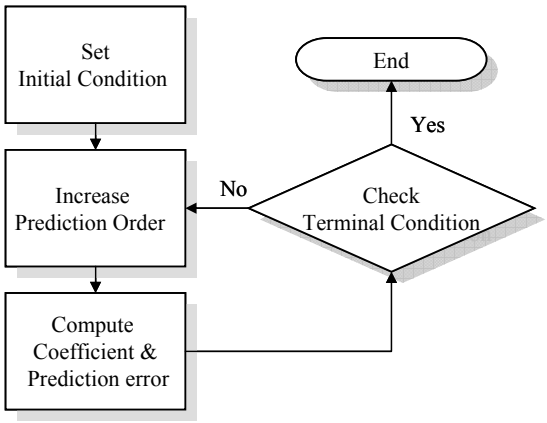


Figure 8: Flow chart of the MLD algorithm

4.3 Tonality Measurement of Inverse Filtered Signal

Inverse filtering is a HF whitening process in the decoder which is formulated in (1). The energy of tonal component in the replicated low bands will be eliminated with a certain degree that controlled by the chirp factor. Since the TNR will be changed after inverse filtering, the tonality measurement should be re-estimated to obtain the actual TNR. Otherwise, the unmatched TNR to the inverse filtered signal will consequently result in improper compensation. The following method provides an efficient approach to measure the TNR of inverse filtered signal without performing the inverse filtering process on time domain signal.

From previous section, given the autocorrelation matrix of subband signal, the MLD algorithm can measure the TNR and search the adaptive order automatically. Through the calculation for autocorrelation of inverse filtered signal, the inverse filtered TNR can be obtained by applying the MLD algorithm again. The autocorrelation of inverse filtered signal in b^{th} subband is defined as

$$\phi_b(i,j) = \sum_{n=0}^{N-1} x_b[n-i] \cdot x_b[n-j]^* . \tag{26}$$

By substituting (1) into (26) results in

$$\begin{aligned} \phi_b(i,j) &= \sum_{n=0}^{N-1} \left(x_p[n-i] - \sum_{k=1}^{2} \alpha^k a_{k-1} x_p[n-i-k] \right) \cdot \left(x_p[n-j] - \sum_{k=1}^{2} \alpha^k a_{k-1} x_p[n-j-k] \right)^* \\ &= \phi_p(i,j) - \sum_{k=1}^{2} \alpha^k a_{k-1} \phi_p(i+k,j) - \sum_{k=1}^{2} \alpha^k a^*_{k-1} \phi_p(i,j+k) + \sum_{k=1}^{2}\sum_{m=1}^{2} \alpha^{k+m} a_{k-1} a^*_{m-1} \phi_p(i+k,j+m) \end{aligned} , \tag{27}$$

where p^{th} subband in LF is mapped to b^{th} subband in HF. Since the autocorrelation of inverse filtered signal is obtained from (27), the tonality measurement of inverse filtered signal can be done by applying the MLD algorithm again, with $\phi_b(i,j)$ as its input.

4.4 T/N dB Difference Measurement

According to the result of the MLD algorithm, the tonal component can be measured from the predictable energy due to the stationary property. For the b^{th} subband in the scope of MLD, its tone component is defined as

$$e_b^t = \frac{\phi(0,0) - e^p}{p}, \tag{28}$$

where $\phi(0,0) - e^p$ is the predictable energy and p is the number of tones in the subband. In other words, (28) estimates the average energy of tones in the subband. Furthermore, the tone component for a high resolution grid is derived from the maximum among the tone component of the subbands.

$$E_k^t = \max\left\{e_b^t \mid b \in k\right\}. \tag{29}$$

On the other hand, the prediction error can be referred to as the noise content due to the random property. Hence, the noise floor for the bth subband in the scope of MLD can be obtained from

$$e_b^n = \frac{e^p}{N - p}, \tag{30}$$

where N is the number of the subband samples in the scope of MLD. Also, the noise floor for a high resolution grid is derived from the average noise floor among the subbands in (31).

$$E_k^n = \frac{1}{|k|} \cdot \sum_{b \in k} e_k^n. \tag{31}$$

Finally, the T/N dB difference used in Chapter 3 is considered as the dB value difference between the dominate tone and the average noise floor within a high resolution grid.

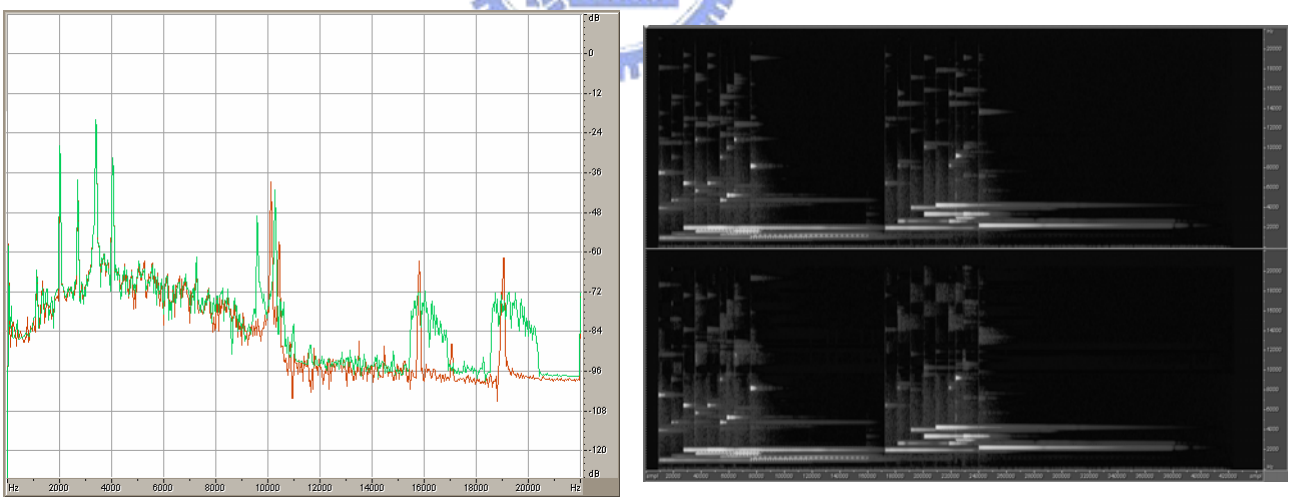
$$\Delta_k = f_{dB}\left(E_k^t\right) - f_{dB}\left(E_k^n\right). \tag{32}$$

Chapter 5
Artifacts

This chapter discussed about two typical artifacts in HE-AAC codec, noise overflow and tonal spike. The causes leading to the artifacts are pointed out, and a remedial method at the encoder end to prevent the noise overflow phenomenon in HF reconstruction is proposed.

5.1 Noise Overflow

As mentioned in Chapter 4, the missed detection of tonality measurement means the energy of tonal component is underestimated and the energy of noise component is overestimated. In this case, the inaccuracy of TNR measurement will consequently cause lesser tonal or overmuch noise compensation, which all leading to the noise overflow phenomenon in HF reconstruction that producing a fizzy sound perceptually. The noise overflow phenomenon is illustrated in Figure 9**.**



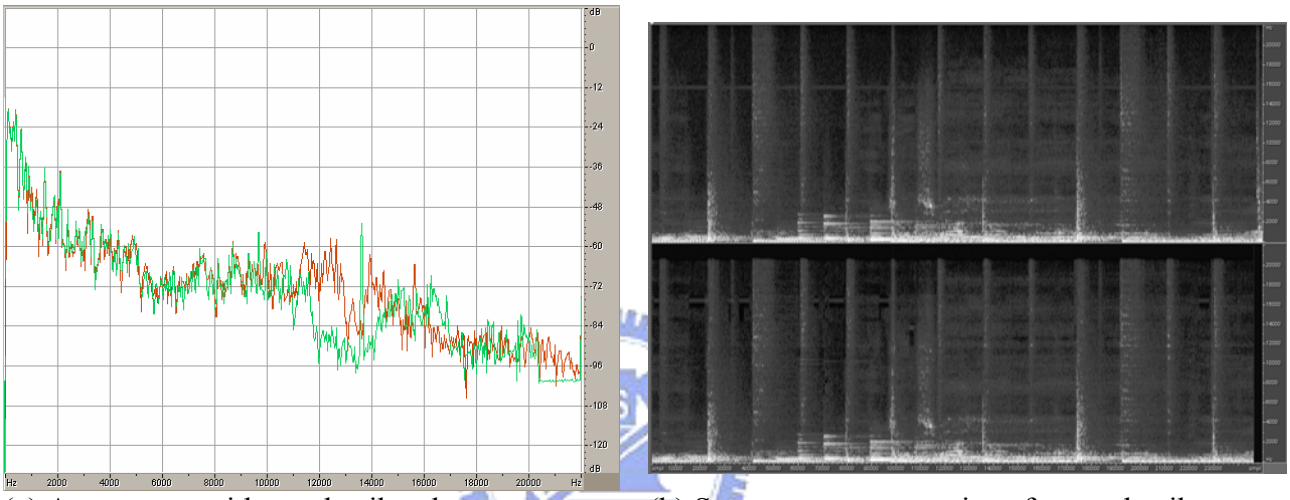
(a) A spectrum with noise overflow phenomenon.

(b) Spectrogram comparison for noise overflow phenomenon.

Figure 9: Noise overflow phenomenon

5.2 Tonal Spike

In the case for the false alarm of tonality measurement, the energy of tonal component is overestimated and the energy of noise component is underestimated. The inaccuracy of TNR measurement will consequently cause overmuch tonal or lesser noise compensation, which both leading to the tonal spike appearance in HF reconstruction that producing a metallic sound perceptually. The tonal spike phenomenon is illustrated in Figure 10.



(a) A spectrum with tonal spike phenomenon. (b) Spectrogram comparison for tonal spike phenomenon.

Figure 10: Tonal spike phenomenon

5.3 Noise Floor Correction

Noise floor correction is an alternative approach at the encoder end. It provides a remedial method to reduce the noise overflow phenomenon under the determined control parameter and compensation selection. Through the three phase simulation in Chapter 3, the noise floor energy N_k^r after HF adjustment can be obtained from (8) or (14) as shown in Table 5. Since the control parameter Q is determined, the noise floor energy in HF reconstruction can be considered as a function of E_k^o. The noise overflow detection can be done by comparing with the original noise floor. Also, from (11) and (16) in Table 3, it has shown that changing the averaged energy of original high bands in k^{th} high resolution grid (E_k^o) will not affect the reconstructed T/N dB difference. Hence, the noise floor correction can be done by modifying E_k^o. The following evaluation is based on energy grids.

Table 5: Noise floor energy in HF reconstruction.

	Noise floor energy in HF reconstruction	
Noise-demand	$N_k^{ND}\left(E_k^o\right)=\left(\frac{E_k^n}{E_k^r}+Q\right)\cdot\left(\frac{1}{1+Q}\right)\cdot E_k^o$	(33)
Tone-demand	$N_k^{TD}\left(E_k^o\right)=\left(\frac{E_k^n}{E_k^r}\cdot Q+Q\right)\cdot\left(\frac{1}{1+Q}\right)\cdot E_k^o$	(34)

To correct the noise floor energy in HF reconstruction, the averaged energy of original high bands E_k^o is modified by a revision factor R, which means replace E_k^o with $E_k^o \cdot R$ in (33) and (34). The resultant noise floor energy should equal to the original noise floor energy N^o according to

$$\frac{\sum_{k\in ND} N_k^{ND}\left(E_k^o\cdot R\right)+\sum_{k\in TD} N_k^{TD}\left(E_k^o\cdot R\right)}{\#_{hg}}=N^0, \tag{35}$$

where $\#_{hg}$ is the number of high resolution grids in the current energy grid.

The revision factor R can be obtained by substituting (33), (34) into (35) as following

$$\widetilde{R} = \frac{\#_{hg} \cdot N^0 \cdot (1+Q)}{\left[\left(\sum_{k\in ND} \Psi_k^n \cdot E_k^o\right) + Q \cdot \left(\sum_{k\in TD} \Psi_k^n \cdot E_k^o\right)\right] + Q \cdot \left[\sum_{k\in ND} E_k^o + \sum_{k\in TD} E_k^o\right]}, \tag{36}$$

where $\Psi_k^n = E_k^n / E_k^r$.

If $\widetilde{R} \cong 1$, it means that the reconstructed noise floor does not differ from the original noise floor largely. On the other hand, if $\widetilde{R} << 1$, it indicates the noise overflow phenomenon, hence the revision factor is applied.

Chapter 6
Experimental Results

In this chapter, a large amount of experiments are conducted for verifying our proposed approaches based on the MPEG test tracks and the music database collected in our lab. The experiments include both objective quality measurement and subjective measurement.

6.1 Experiment Environment

Computer Status:

Platform	Personal Computer
Operating System	Windows XP
CPU	Intel Pentium 4 2.4GHz
Memory	256MB DDR400 * 2
Mother Board	ASUS P4P800
Sound Card	ADI AD1985 AC' 97
Headphone	ALESSANDRO MUSIC SERIES PRO

Objective Quality Measurement Tool:

For objective quality evaluation, the thesis mainly adopts the PEAQ system (perceptual evaluation of audio quality) [16] which is the recommendation system by ITU-R Task Group 10/4. The system includes a subtle perceptual model to measure the difference between two tracks. The objective difference grade (ODG) is the output variable from the objective measurement method. The ODG values should range from 0 to −4, where 0 corresponds to an imperceptible impairment and −4 to impairment judged as very annoying. The improvement up to 0.1 is usually perceptually audible. The PEAQ has been widely used to measure the compression technique due to the capability to detect perceptual difference sensible by human hearing systems.

Subjective Quality Measurement Tool:

For subjective quality evaluation, the thesis mainly adopts the MUSHRA system [17]. The system allows the blind comparison of multiple audio files. Multi stimulus test with hidden reference and anchors has been designed to give a reliable and repeatable measure of the audio quality of intermediate-quality signals. MUSHRA has the advantage that it provides an absolute measure of the audio quality of a codec which can be compared directly with the reference. MUSHRA follows the test method and impairment scale recommended by ITU-R BS.1116 [18].

6.2 Objective Quality Measurement in MPEG Test Tracks

The twelve test tracks recommended by MPEG are shown in Table 6. These tracks include the critical music balancing on the percussion, string, wind instruments, and human vocal. In this section, the quality enhancement of proposed methods at different bit rates is verified based on these MPEG test tracks and NCTU-HEAAC [19] is adopted as the platform.

Table 6: The twelve tracks recommended by MPEG

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

Remarks:

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

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

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

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

Table 7: ODG for proposed methods on MPEG test tracks at bit rate 80 kbps.

Codec	NCTU-HEAAC			
Bit Rate	80 kbps			
Tracks	M0	M1	M2	M3
es01	-0.74	-0.74	-0.73	-0.73
es02	-0.65	-0.66	-0.65	-0.65
es03	-0.74	-0.74	-0.75	-0.75
sc01	-1.05	-1.05	-1.01	-1.01
sc02	-1.28	-1.28	-1.25	-1.25
sc03	-1.21	-1.21	-1.18	-1.19
si01	-1.67	-1.67	-1.65	-1.65
si02	-1.01	-1.01	-1.03	-1.03
si03	-1.69	-1.69	-1.64	-1.64
sm01	-1.63	-1.64	-1.6	-1.6
sm02	-1.71	-1.62	-1.6	-1.6
sm03	-1.41	-1.42	-1.34	-1.35
Max	-0.65	-0.66	-0.65	-0.65
Min	-1.71	-1.69	-1.65	-1.65
Average	-1.2325	-1.2275	-1.2025	-1.20417
M0: HF reconstruction without compensation				
M1: M0 + noise floor correction				
M2: HF reconstruction with tone/noise compensation				
M3: M2 + noise floor correction				

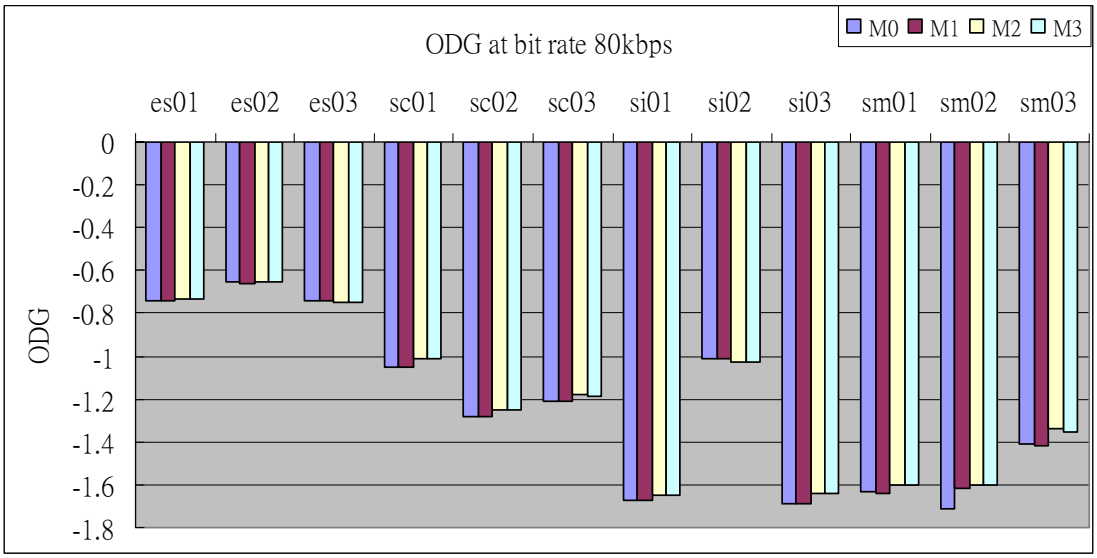


Figure 11: ODG comparison of MPEG test tracks at bit rate 80kbps.

Table 8: ODG for proposed methods on MPEG test tracks at bit rate 64 kbps.

Codec	NCTU-HEAAC			
Bit Rate	64 kbps			
Tracks	M0	M1	M2	M3
es01	-1.02	-1.03	-1	-1.01
es02	-0.89	-0.89	-0.87	-0.87
es03	-1.01	-1.03	-1.03	-1.03
sc01	-1.7	-1.7	-1.66	-1.66
sc02	-1.77	-1.78	-1.74	-1.74
sc03	-1.67	-1.68	-1.63	-1.63
si01	-2	-2	-2	-1.99
si02	-1.5	-1.5	-1.57	-1.57
si03	-2.16	-2.16	-2.09	-2.09
sm01	-2.18	-2.19	-2.17	-2.17
sm02	-2.45	-2.42	-2.51	-2.51
sm03	-1.8	-1.81	-1.73	-1.72
Max	-0.89	-0.89	-0.87	-0.87
Min	-2.45	-2.42	-2.51	-2.51
Average	-1.67917	-1.6825	-1.66667	-1.66583
M0: HF reconstruction without compensation				
M1: M0 + noise floor correction				
M2: HF reconstruction with tone/noise compensation				
M3: M2 + noise floor correction				

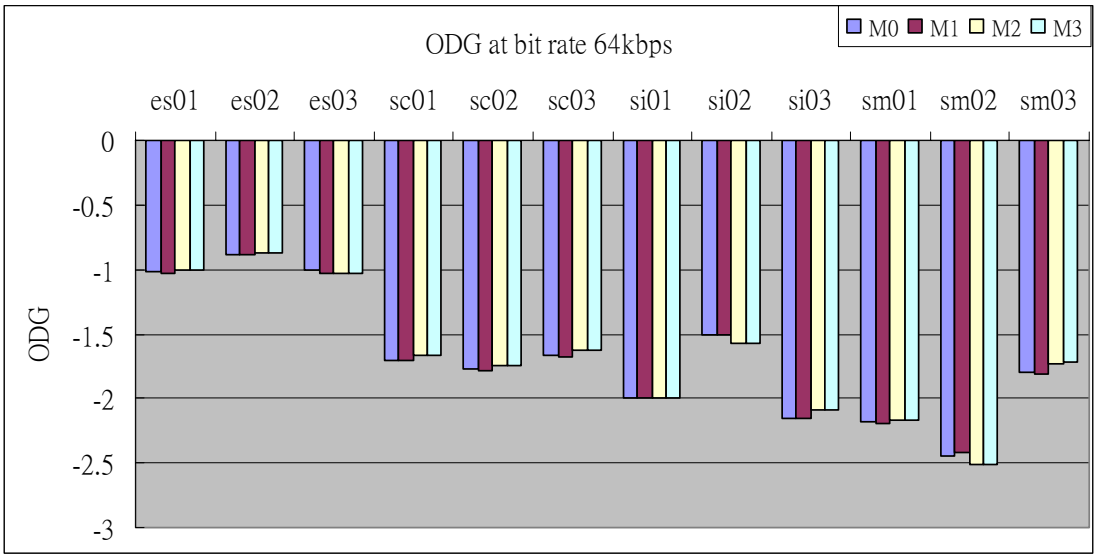


Figure 12: ODG comparison of MPEG test tracks at bit rate 64kbps.

Table 9: ODG for proposed methods on MPEG test tracks at bit rate 48 kbps

Codec	NCTU-HEAAC			
Bit Rate	48 kbps			
Tracks	M0	M1	M2	M3
es01	-1.58	-1.59	-1.56	-1.61
es02	-1.42	-1.4	-1.39	-1.42
es03	-1.71	-1.71	-1.71	-1.7
sc01	-2.46	-2.46	-2.4	-2.4
sc02	-2.64	-2.63	-2.59	-2.62
sc03	-2.41	-2.41	-2.33	-2.33
si01	-2.73	-2.73	-2.71	-2.74
si02	-2.4	-2.38	-2.48	-2.46
si03	-3.21	-3.21	-3.17	-3.17
sm01	-3.24	-3.24	-3.2	-3.2
sm02	-3.4	-3.41	-3.37	-3.38
sm03	-2.36	-2.37	-2.26	-2.26
Max	-1.42	-1.4	-1.39	-1.42
Min	-3.4	-3.41	-3.37	-3.38
Average	-2.46333	-2.46167	-2.43083	-2.44083
M0: HF reconstruction without compensation				
M1: M0 + noise floor correction				
M2: HF reconstruction with tone/noise compensation				
M3: M2 + noise floor correction				

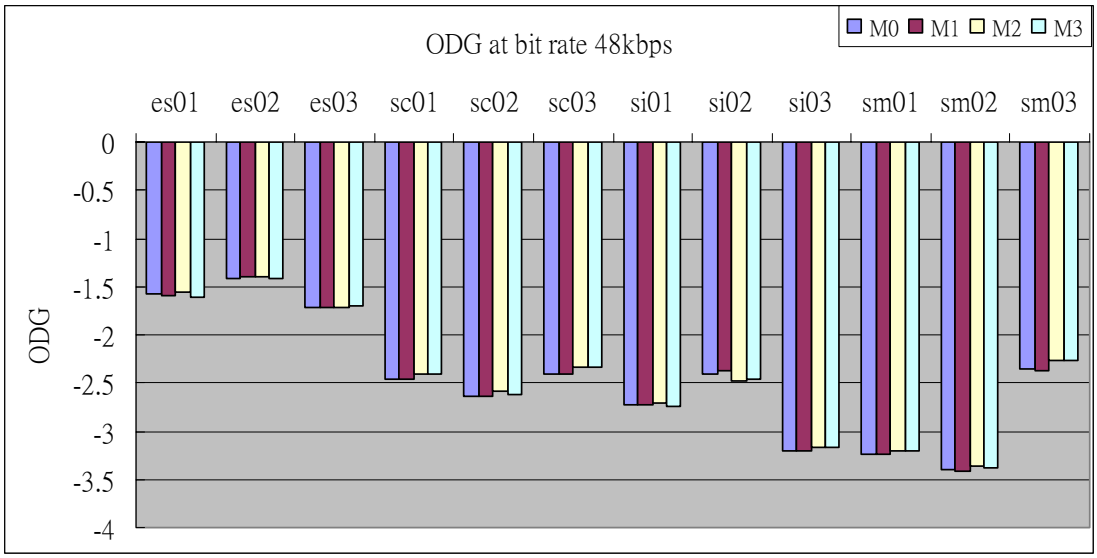


Figure 13: ODG comparison of MPEG test tracks at bit rate 48kbps.

As the results shows, the quality of HF reconstruction in SBR audio is improved by proper tone/noise compensation at different bit rate, except es03, si02 and sm02. The subjective test will be conducted later to check if the quality is degraded or improved. The comparison between noise floor correction turn on and turn off does not show much difference in the MPEG test tracks, because there are only a few amount of noise overflow phenomenon in this set. The improvement of noise floor correction is more distinct in music database which will be represented in next section.

6.3 Objective Quality Measurement in Music Database

From previous section, the proposed methods are verified on MPEG test set. However, there are only twelve tracks in MPEG set, but the proposed methods need great quantity of test tracks to prove its possible risk and robustness. Table 10 shows the audio database which is collected by our laboratory [20]. There are 15 categories and 319 test tracks in this audio database. Each category has its signal properties. With these large numbers of experiments, the quality of proposed methods can be assessed.

Table 10: The PSPLab audio database [20]

Bitstream Categories		# of tracks	Remark
1	ff123	101	Killer bitstream collection from ff123 [21].
2	gpsycho	23	LAME quality test bitstream [22].
3	HA64KTest	37	64 kbps test bitstream for multi-format in HA forum [23].
4	HA128KTestV2	12	128 kbps test bitstream for multi-format in HA forum [23].
5	horrible_song	16	Collections of critical songs among all bitstreams in PSPLab.
6	ingets1	5	Bitstream collection from the test of OGG Vorbis pre 1.0 listening test [24].
7	MPEG	12	MPEG test bitstream set for 48000Hz.
8	MPEG44100	12	MPEG test bitstream set for 44100 Hz.
9	Phong	8	Test bistream collection from Phong [25].
10	PSPLab	37	Collections of bitstream from early age of PSPLab. Some are good as killer.
11	sjeng	3	Small bitstream collection by sjeng.
12	SQAM	16	Sound quality assessment material recordings for subjective tests [26].
13	TestingSong14	14	Test bitstream collection from rshong, PSPLab.
14	TonalSignals	15	Artificial bitstream that contains sin wave etc.
15	VORBIS_TESTS_ Samples	8	Eight Vobis testing samples from HA [23].

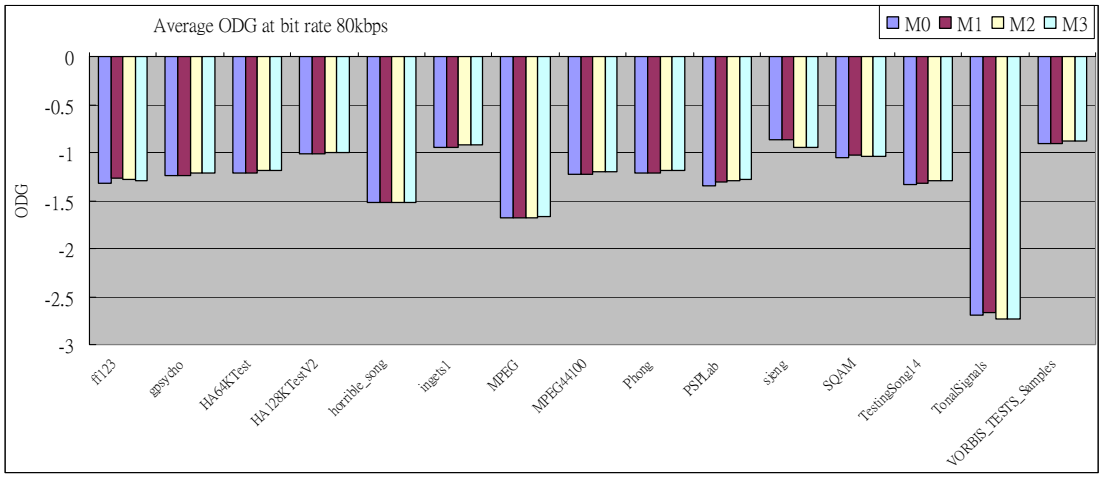


Figure 14: Average ODG in music database at bit rate 80kbps.

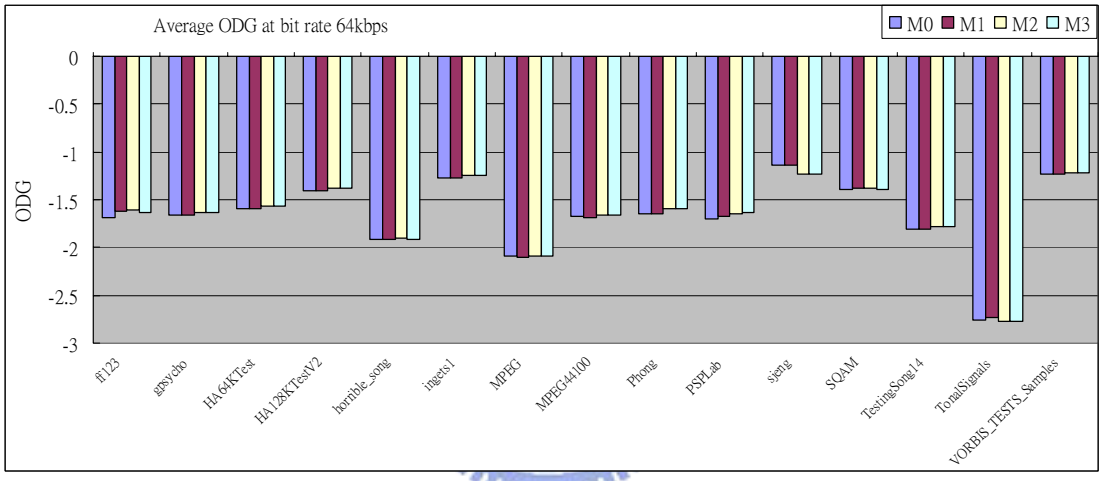


Figure 15: Average ODG in music database at bit rate 64kbps.

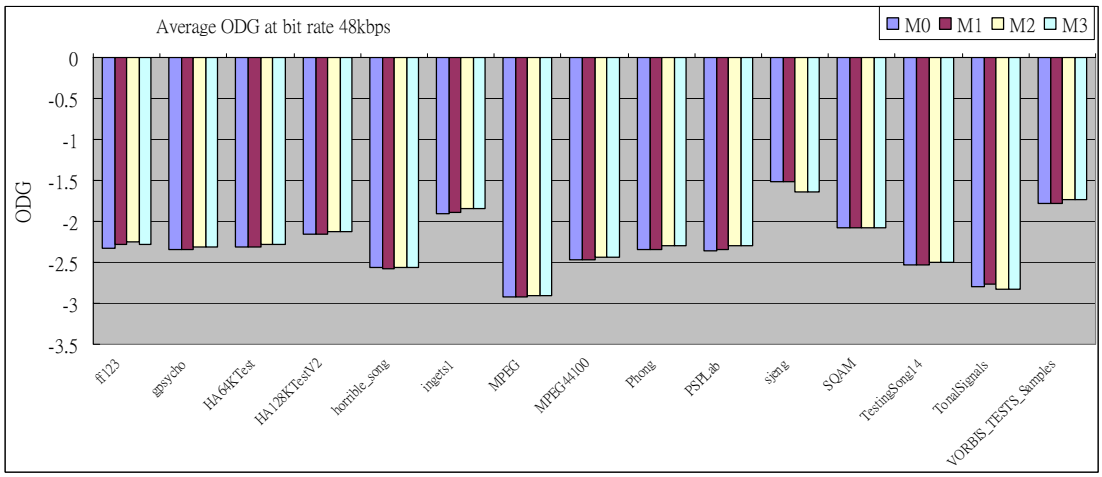


Figure 16: Average ODG in music database at bit rate 48kbps.

As the result shows, the quality of HF reconstruction is improved by tone/noise compensation in most categories except sjeng and TonalSignals. The sjeng set only contain three tracks, one of the tracks shows degradation and the other two show improvement. The degradation on TonalSignals set is caused by numerical error and tone shift in frequency domain. The spectrum comparison of degraded tracks and numerical error are given below.

Degraded tracks in sjeng set:

c44.wav (due to the inaccurate tonality measurement)

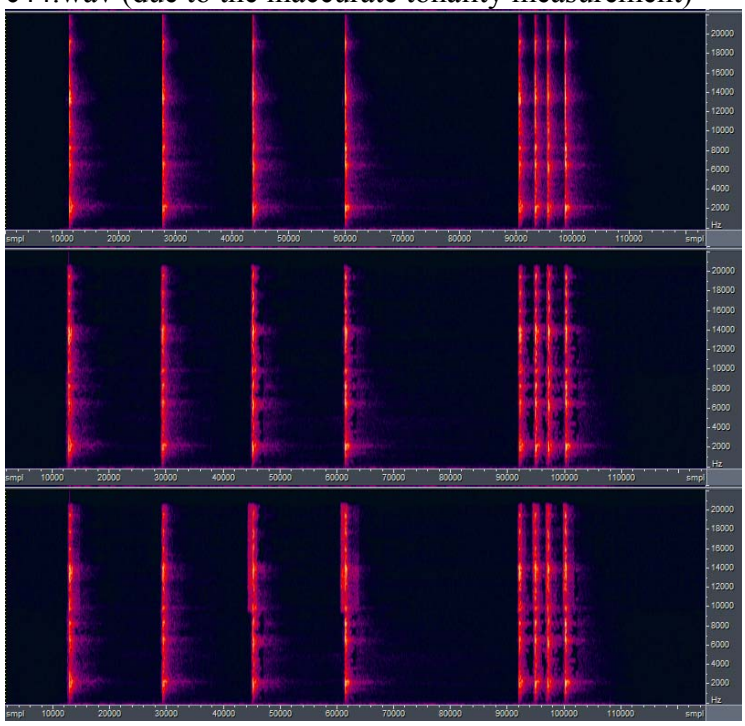
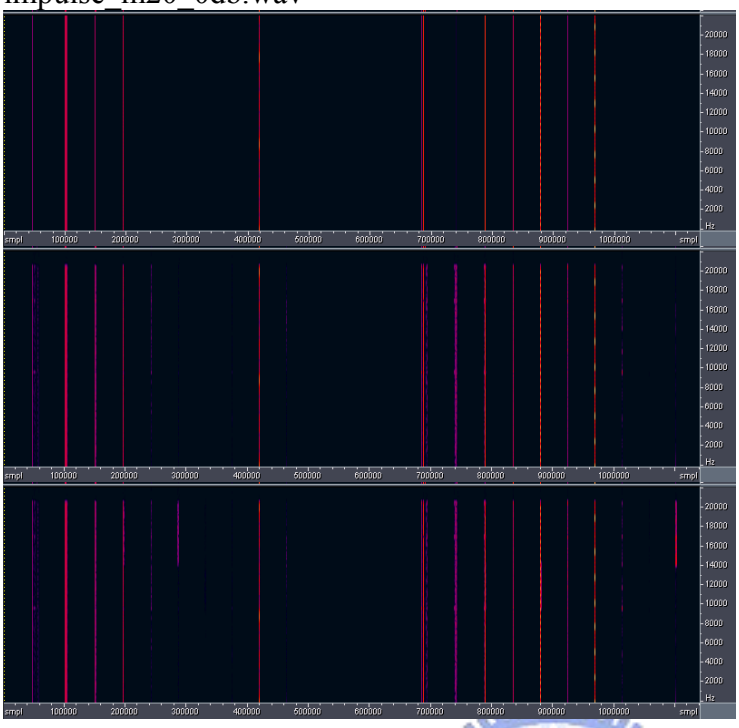


Figure 17: spectrum comparison of test track (c44.wav) in sjeng set.

Degraded tracks in TonalSignals set:

impulse_m20_0db.wav



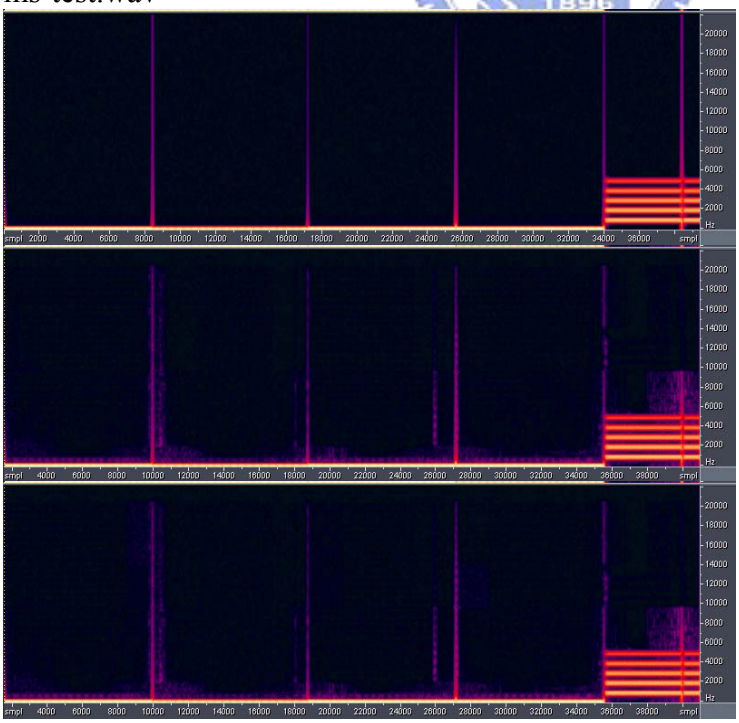
Original

M0

M2

Figure 18: spectrum comparison of test track (impulse_m20_0db.wav) in TonalSignals set.

ms-test.wav



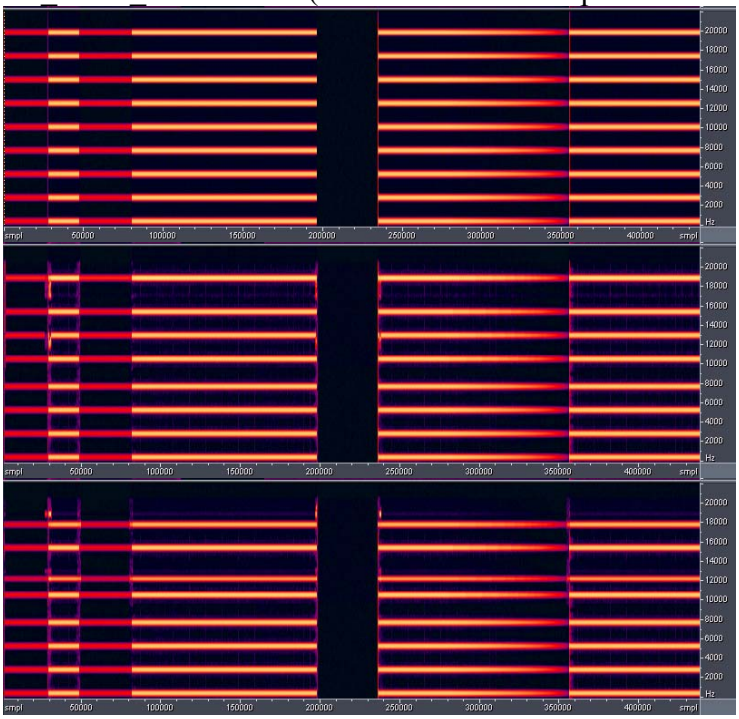
Original

M0

M2

Figure 19: spectrum comparison of test track (ms-test.wav) in TonalSignals set.

sin_9kind_valious.wav (due to the tone shift phenomenon)



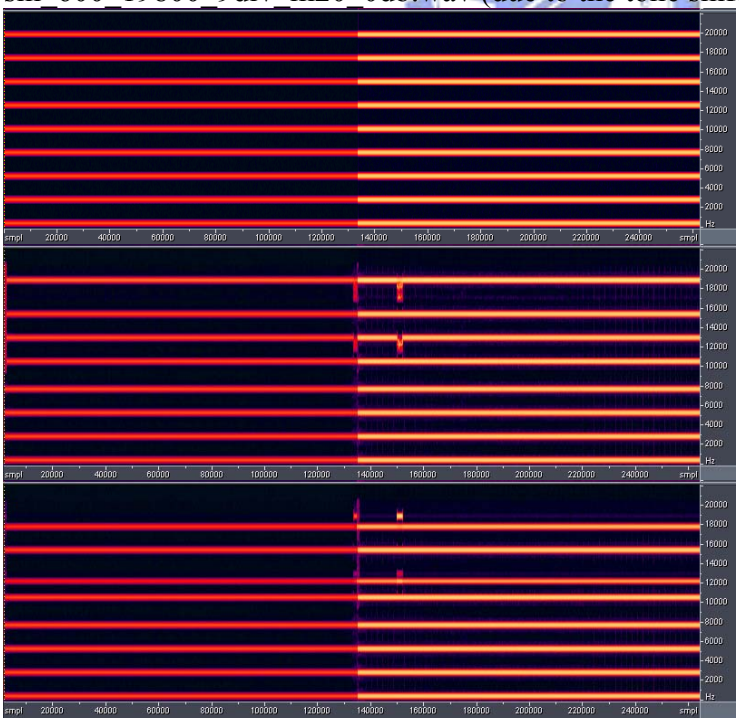
Original

M0

M2

Figure 20: spectrum comparison of test track (sin_9kind_valious.wav) in TonalSignals set.

sin_600_19800_9div_m20_0db.wav (due to the tone shift phenomenon)



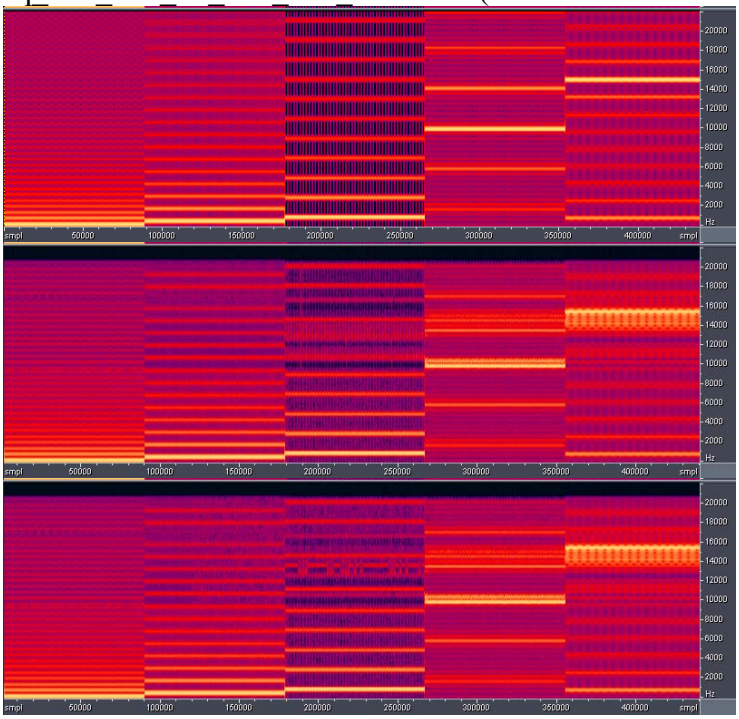
Original

M0

M2

Figure 21: spectrum comparison of test track (sin_600_19800_9div_m20_0db.wav) in TonalSignals set.

sq_300_625_1k_10k_15k_m20.wav (due to the tone shift phenomenon)



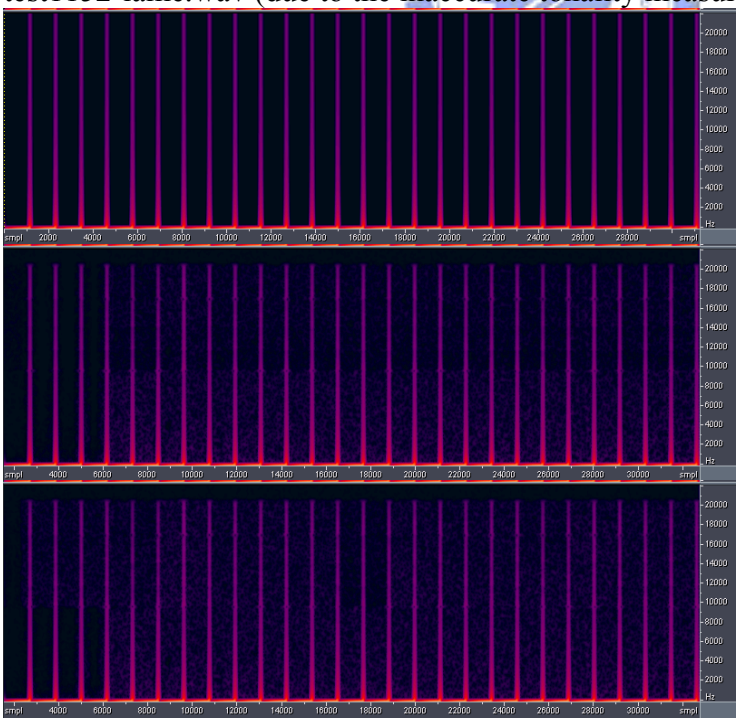
Original

M0

M2

Figure 22: spectrum comparison of test track (sq_300_625_1k_10k_15k_m20.wav) in TonalSignals set.

test1152-lame.wav (due to the inaccurate tonality measurement)



Original

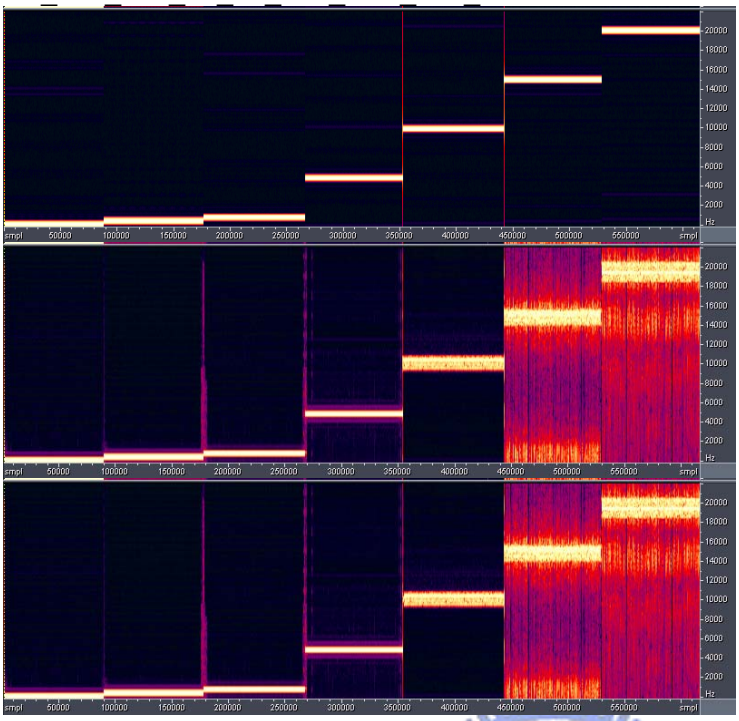
M0

M2

Figure 23: spectrum comparison of test track (test1152-lame.wav) in TonalSignals set.

Numerical error in TonalSignals set:

sin_300_625_1k_5k_10k_15K_20k_0db.wav



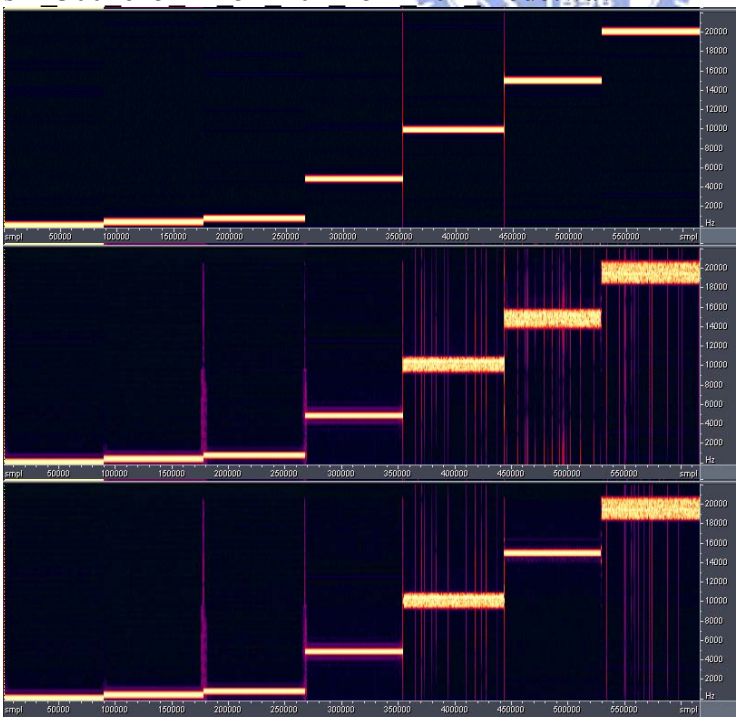
Original

M0

M2

Figure 24: spectrum comparison of test track (sin_300_625_1k_5k_10k_15K_20k_0db.wav) in TonalSignals set.

sin_300_625_1k_5k_10k_15K_20k_m6db.wav



Original

M0

M2

Figure 25: spectrum comparison of test track (sin_300_625_1k_5k_10k_15K_20k_m6db.wav) in TonalSignals set.

sin_sweep_30_20k_0db.wav

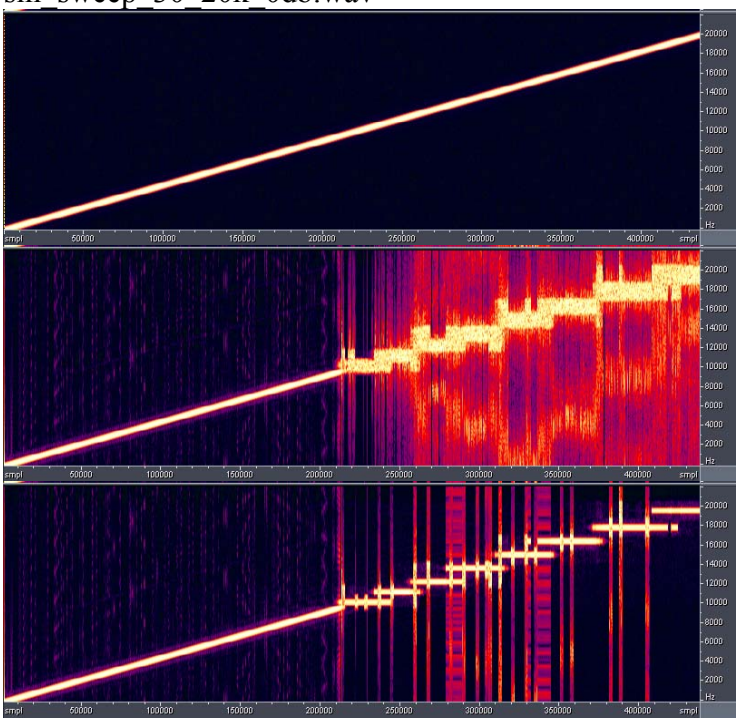


Figure 26: spectrum comparison of test track (sin_sweep_30_20k_0db.wav) in TonalSignals set.

6.4 Subjective Quality Measurement

In section 6.2, the ODG of some test tracks show degradation when comparing method M2 with M0 which is unexpected. The quality improvement will be supported by listening test in this section as shown in Figure 27. Most of the test tracks are improved by tone/noise compensation except si02, which is a transient signal. The reason of quality degradation is due to the assumption of MLD algorithm, that the input audio signal is stationary. Therefore if the input signal is transient, the tonality measurement will lose the accuracy and therefore results in improper compensation.

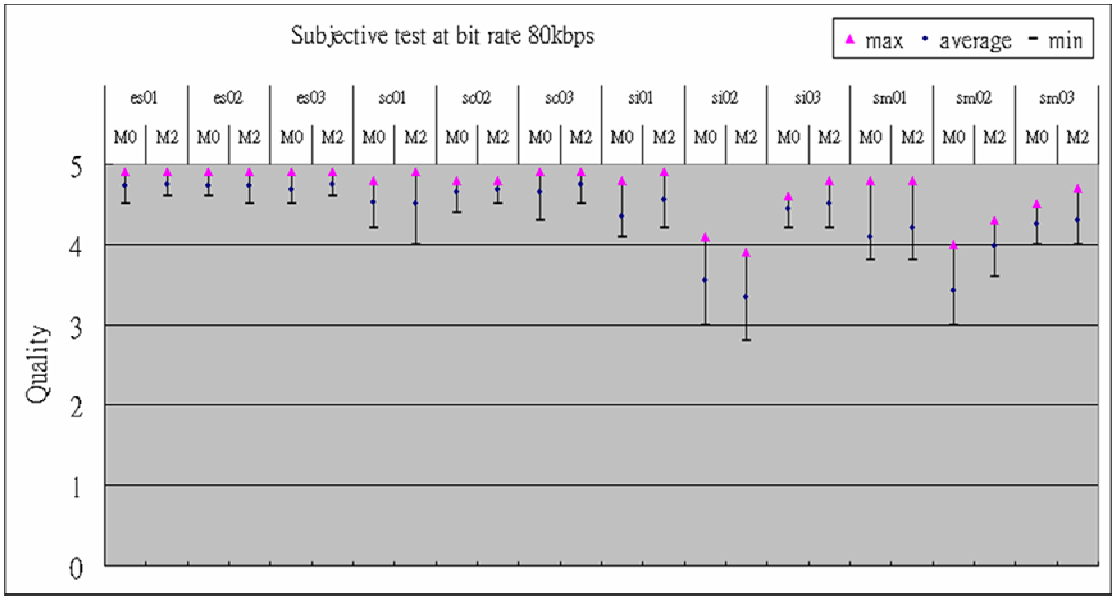


Figure 27: Results of listening test on MPEG test tracks at bit rate 80kbps.

Chapter 7
Concluding Remark

In this thesis, an efficient method based on Levinson-Durbin algorithm is proposed to measure the tonality by linear prediction approach, which can search the adaptive orders automatically to fit different subband contents. Further more, the MLD algorithm also provides an efficient method to measure the tonality of inverse filtered signal at the encoder end. Through the calculation for autocorrelation of inverse filtered signal, the inverse filtering process is avoided. The sharing of control parameter is also concerned, and an optimal decision criterion of control parameter is proposed. The causes leading to two typical artifacts in HE-AAC codec are pointed out and a remedial method to reduce the noise overflow phenomenon is proposed. Both objective and subjective tests are conducted to check the quality improvement. Other publication of this thesis is in [15].

Table 11~12 shows the ODG comparison with method M3 and existing codec (CodingTechnologies aacplus 7.0.5; NERO HE-AAC 3.0.0.0).

Table 11: ODG comparison with existing codec at bit rate 80kbps.

Bit Rate	80k		
Codec	NCTU-HEAAC	CT 7.0.5	NERO 3.0.0.0
es01	-0.73	-0.78	-1.46
es02	-0.65	-0.97	-2.14
es03	-0.75	-0.8	-2.4
sc01	-1.01	-1.05	-1.17
sc02	-1.25	-1.54	-1.22
sc03	-1.19	-1.17	-1.23
si01	-1.65	-2.3	-2.03
si02	-1.03	-1.03	-1.81
si03	-1.64	-1.9	-2.07
sm01	-1.6	-1.97	-2.14
sm02	-1.6	-2.27	-1.82
sm03	-1.35	-1.28	-1.39
Max	-0.65	-0.78	-1.17
Min	-1.65	-2.3	-2.4
Average	-1.20417	-1.4217	-1.74

Table 12: ODG comparison with existing codec at bit rate 64kbps.

Bit Rate	64k		
Codec	NCTU-HEAAC	CT 7.0.5	NERO 3.0.0.0
es01	-1.01	-1.01	-1.62
es02	-0.87	-1.48	-2.28
es03	-1.03	-1.04	-2.19
sc01	-1.66	-1.66	-1.96
sc02	-1.74	-2.35	-1.6
sc03	-1.63	-1.65	-1.73
si01	-1.99	-2.84	-2.77
si02	-1.57	-1.6	-2.24
si03	-2.09	-2.19	-2.86
sm01	-2.17	-2.73	-3.14
sm02	-2.51	-2.72	-2.25
sm03	-1.72	-1.71	-2.14
Max	-0.87	-1.01	-1.6
Min	-2.51	-2.84	-3.14
Average	-1.6658	-1.915	-2.2317

Table 13: ODG comparison with existing codec at bit rate 48kbps.

Bit Rate	48k		
Codec	NCTU-HEAAC	CT 7.0.5	NERO 3.0.0.0
es01	-1.61	-2.23	-1.97
es02	-1.42	-2.65	-2.49
es03	-1.7	-2.41	-2.7
sc01	-2.4	-2.65	-2.79
sc02	-2.62	-3.1	-2.64
sc03	-2.33	-3.05	-2.88
si01	-2.74	-3.42	-3.64
si02	-2.46	-2.73	-3.47
si03	-3.17	-2.86	-3.83
sm01	-3.2	-3.61	-3.85
sm02	-3.38	-3.32	-3.28
sm03	-2.26	-3.15	-3.08
Max	-1.42	-2.23	-1.97
Min	-3.38	-3.61	-3.85
Average	-2.4408	-2.9317	-3.0517

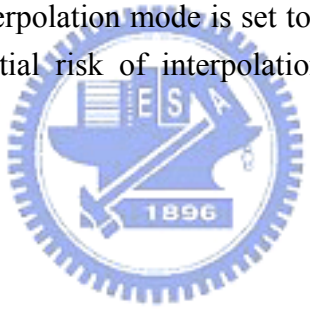
From the result of objective test, NCTU is better than CT than NERO. From subjective test, NCTU is better than CT, and CT is better than NERO On speech signal; CT is better than NCTU, and NCTU is better than NERO on attack signal.

Future work

1. The assumption of MLD algorithm is that the input audio signal is stationary. Therefore if the input signal is transient, the tonality measurement will lose the accuracy.

2. The inverse filtering mode decision in current design is a heuristic method. The inverse filtered tonal energy T_{inv} is approximated by $T_{inv} = T_r \cdot (1-\alpha)^2$, where T_r is the tonal energy before inverse filtering. A more accurate decision criterion should be considered.

3. The interpolation mode decision in current design is that if there is a tone exists in the original HF signal, the interpolation mode is set to 0. Otherwise, the interpolation mode is set to 1. The potential risk of interpolation mode switching need to be analyzed.



References

[1] US Patent, No.:US 6708145 B1, Mar. 16, 2004, “Enhancing Perceptual Performance of SBR and Related HFR Coding Methods by Adaptive Noise-Floor Addition and Noise Substitution Limiting”.

[2] San-Uk Ryu and Kenneth Roth, “Enhanced Accuracy of the Tonality Measure and Control Parameter Extraction Modules in MPEG-4 HE-AAC,” Proc. 119th AES Convention, Preprint 6586, Oct. 2005.

[3] N. Levinson, “The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction,” J. Math. Phys. 25, 261-278 (1947).

[4] 3GPP TS 26.404: “Enhanced aacPlus encoder SBR part,” June 2004.

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

[6] ISO/IEC JTC1/SC29 WG11 MPEG, “Text of ISO/IEC 14496-3:2001/AMD 1:2003, bandwidth extension,” Nov. 2003.

[7] M. Dietz, L. Liljeryd, K. Kjorling, and O. Kunz, "Spectral band replication, a novel approach in audio coding," Proc. 112th AES Convention, Preprint 5871, May 2002.

[8] P. Ekstrand, "Bandwidth extension of audio signals by spectral band replication," Proc. 1st IEEE Benelux Workshop on Model based Processing and Coding of Audio, Nov. 2002.

[9] A. Ehret, M. Dietz, and K. Kjorling, "State-of-the-art audio coding for broadcasting and mobile applications," Proc. 114th AES Convention, Preprint 5560, Mar. 2003.

[10] T. Ziegler, A. Ehret, P. Ekstrand, and M. Lutzky, "Enhancing mp3 with SBR: features and capabilities of the new mp3pro algorithm," Proc. 112th AES Convention, Preprint 5560, May 2002.

[11] M. Wolters, K. Kjorling, D. Homm and H. Purnhagen, "A closer look into MPEG-4 high efficiency AAC," Proc. 115th AES Convention, Preprint 5871, Oct. 2003.

[12] C.M. Liu, W. C. Lee, C. H. Yang, K. Y. Pang, T. Chiou, T. W. Chang, Y. H. Hsiao, H. W Hsu, C. T. Chien, "Design of MPEG-4 AAC Encoder," Proc. 117th AES Convention, Preprint 6201, Oct. 2004,.

[13] H.W. Hsu, C.M. Liu, and W.C. Lee, "Audio Patch Method in MPEG-4 HE AAC Decoder," Proc. 117th AES Convention, Preprint 6221, Oct. 2004.

[14] C.M. Liu, L.W. Chen, H.W. Hsu, and W.C. Lee, "Bit Reservoir Design for HE-AAC," Proc. 118th AES Convention, Preprint 6382, May 2005.

[15] Han-Wen Hsu, Yung-Cheng Yang, Chi-Min Liu, and Wen-Chieh Lee,"Design for High Frequency Adjustment Module in MPEG-4 HEAAC Encoder based on Linear Prediction Method," Proc. 120th AES Convention, Preprint 6755, May 2006.

[16] ITU Radiocommunication Study Group 6, "Draft Revision to Recommendation ITU-R BS.1387- Method for objective measurements of perceived audio quality".

[17] MUSHRA, website: http://ff123.net/abrhr/abchr.html.

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

[19] NCTU-HEAAC, website: http://psplab.csie.nctu.edu.tw/projects/index.pl/nctu-mp3.html.

[20] The Audio Database Collected in Perceptual Signal Processing Lab, website: http://psplab.csie.nctu.edu.tw/projects/index.pl/testbitstreams.html.

[21] Samples for Testing Audio Codecs from ff123, website: http://ff123.net/samples.html.

[22] Quality and Listening Test Information for LAME, website: http://lame.sourceforge.net/gpsycho/quality.html.

[23] Hydrogen Audio, website: http://www.hydrogenaudio.org.

[24] OGG Vorbis Pre 1.0 Listening Test, website: http://hem.passagen.se/ingets1/vorbis.htm.

[25] Phong's Audio Samples, website: http://www.phong.org/audio/samples.xhtml.

[26] Sound Quality Assessment Material, website: http://sound.media.mit.edu/mpeg4/audio/sqam/.