國立交通大學

多媒體工程研究所

碩 士 論 文

SBR, Parametric Coding 及 MPEG Surround 解碼器中高品質低功率四相鏡射濾波器組之設計

High Quality, Low Power QMF Bank Design for SBR, Parametric Coding and MPEG Surround Decoders

研 究 生：曾信耀

指導教授：劉啟民　教授

李文傑　博士

中 華 民 國　九 十 六　年 六 月

SBR, Parametric Coding 及 MPEG Surround 解碼器中高品質低功率四相鏡射濾波器組之設計

High Quality, Low Power QMF Bank Design for SBR, Parametric Coding and MPEG Surround Decoders

研 究 生：曾信耀　　Student：Hsin-Yao Tseng

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

多 媒 體 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Multimedia Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

SBR, Parametric Coding 及 MPEG Surround 解碼器中高品質低功率四相鏡射濾波器組之設計

學生：曾信耀　　指導教授：劉啓民 博士
李文傑 博士

國立交通大學多媒體工程研究所

中文論文摘要

複數的四相鏡射濾波器組由於其無效果失真的特性已經被廣泛的用在 SBR，Parametric Coding 及 Surround 等 MPEG-4 音訊壓縮的標準裡。但是由於利用複數的四相鏡射濾波器組及後續在複數域中處理所產生的高複雜度負擔導致了採用實數的四相鏡射濾波器組爲基本模組的低功率解碼器之發展。然而在實數的四相鏡射濾波器組產生的效果失真所導致的音訊缺陷是主要關心的議題。在此篇論文中採用實數的四相鏡射濾波器組所產生的音訊缺陷將被徹底地研究且提出一個全新的四相鏡射濾波器組之設計以達到高品質兼低功率的目的。而且此篇論文將會展示出將此種全新的四相鏡射濾波器組之設計應用在 SBR， Parametric Coding 及 MPEG Surround 解碼器中品質及複雜度上的優點

High Quality, Low Power QMF Bank Design for SBR, Parametric Coding and MPEG Surround Decoders

Student: Hsin-Yao Tseng Advisor: Dr. Chi-Min Liu

Dr. Wen-Chieh Lee

Institute of Multimedia Engineering College of Computer Science
National ChiaoTung University

ABSTRACT

Due to the aliasing-free properties, the complex quadrature mirror filter (QMF) bank has been used in MPEG-4 audio standard on SBR, parametric, and surrounds coding. The high complexity overhead from the complex QMF bank and the complex data processing in the decoder leads to the development of low power decoder which adopts the real QMF bank as the basic building module to reduce the complexity. However the artifacts from the aliasing in the real QMF bank are the major concern. This paper studies the artifacts from the real QMF bank and proposes a novel QMF bank design to achieve both low complexity and high quality. Also, this paper applies the novel QMF bank to develop the high-quality and low-power SBR, parametric, and MPEG surround decoders and shows the merits in complexity and quality.

致謝

感謝劉啓民老師兩年來的栽培及李文傑博士給予的指導，實驗室的楊宗翰學長、許瀚文學長、胡正倫同學以及學弟陳柏景的協助，在研究上提供我寶貴的意見，讓我在專業知識及研究方法獲得非常多的啓發。

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

Contents

Figure List

Table List

Chapter 1
Introduction

MFBs (modulated filterbanks) have been widely used as T/F mapping tools in a variety of perceptual audio codecs, such as MDCT for MPEG-2/4 AAC, PQMF for MPEG Layer I-III and MPEG-4 HE-AAC(SBR) Low Power version [1] and TDAC-variant FB for Dolby AC-3 which belong to the CMFB(Cosine modulated filterbanks) class and QMF bank for MPEG-4 HE-AAC(SBR) High Quality version [1], HE-AAC version 2(Parametric coding) [2] and MPEG Surround[3] which locate in Complex MFB class. A general form of analysis filters in cosine QMF bank as knows as pseudo QMF bank(PQMF) [4] can be represented as cosine modulated versions of a symmetric low-pass prototype filter $p(n)$ like (1).

$$h_k(n) = p(n)\cos\left(\frac{\pi}{2M}(k+0.5)(2n-N)+\theta_k\right). \quad (1)$$

Its matching synthesis filters can be represented as

$$f_k(n) = p(n)\cos\left(\frac{\pi}{2M}(k+0.5)(2n-N)-\theta_k\right). \quad (2)$$

One intrinsic property of cosine QMF bank illustrated in Figure 1 is that for one passband locates in positive frequency range; there is another corresponding passband in negative frequency range which is complex conjugated to its positive frequency one and this is why the signals are real-valued after passing through cosine QMF bank.

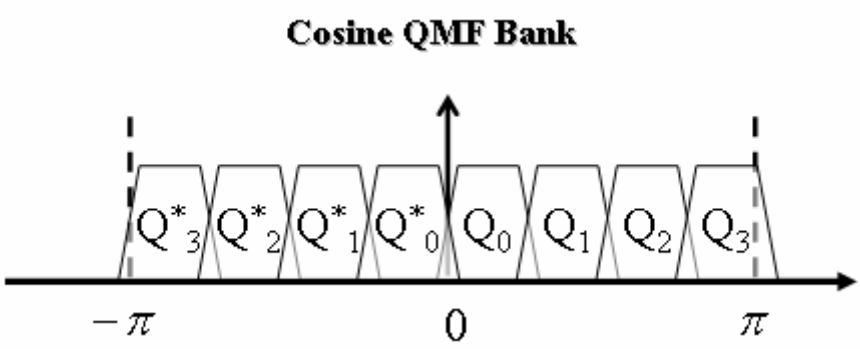

Figure 1: Illustration of 4-channel cosine QMF banks

Figure 2 shows the overlap between the desirable signal and the imaging-term in cosine QMF bank which comes from M factor decimation by decimator to make the system critically downsampled.

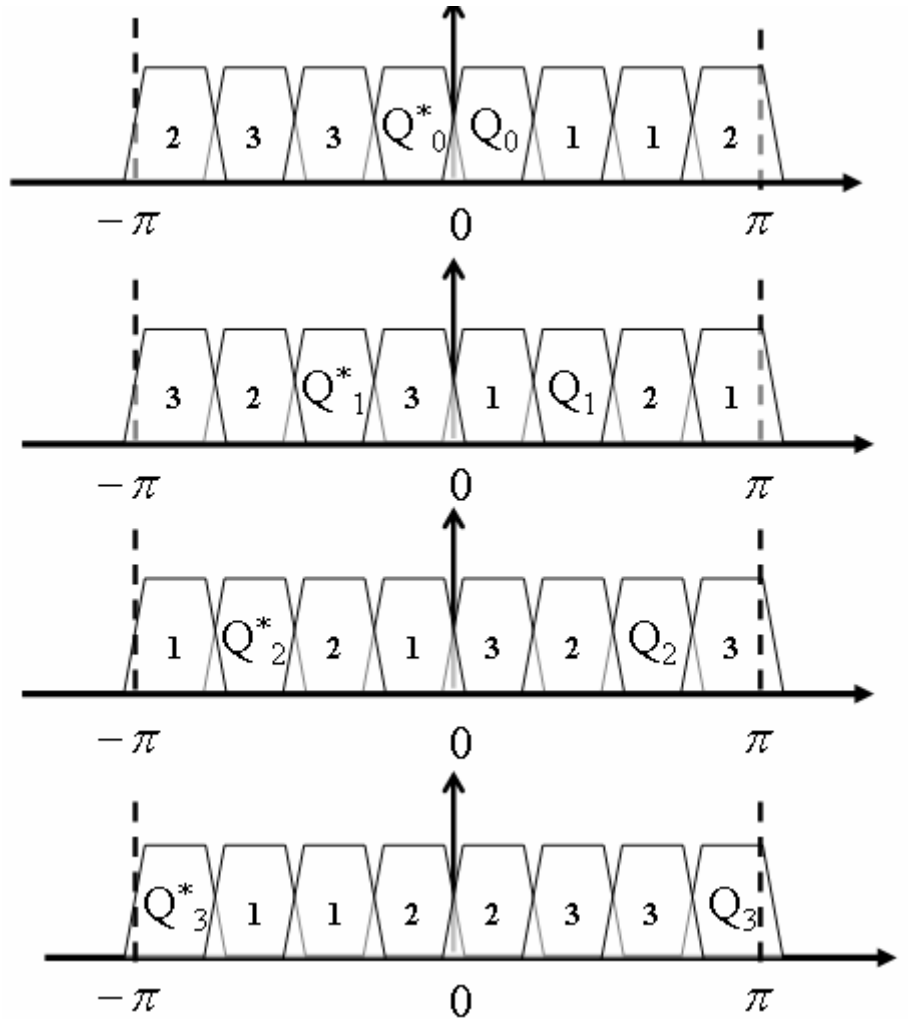

Figure 2: Illustration of overlap in cosine QMF banks

And the aliasing terms emerge from overlaps if the scaling gain between adjacent subbands is changed. On the other hand, if the scaling gain remains the same, the overlap in one passband can be eliminated by its adjacent subband where the aliasing term is alleviated.

A general form of analysis and synthesis filters in complex QMF bank[4] are represented in (3) and (4) respectively which can be viewed as adding imaginary part to the cosine QMF bank where the imaginary part is the sine modulated version of the same low-pass prototype filter used in cosine QMF bank.

$$h_k(n) = p(n)\exp\left(\frac{i\pi}{2M}(k+0.5)(2n-N)+\theta_k\right). \qquad (3)$$

$$h_k(n) = p(n)\exp\left(\frac{i\pi}{2M}(k+0.5)(2n-N)-\theta_k\right). \quad (4)$$

Figure 3 shows a 4-channel complex QMF bank, where there is no corresponding negative frequency component found in cosine QMF bank.

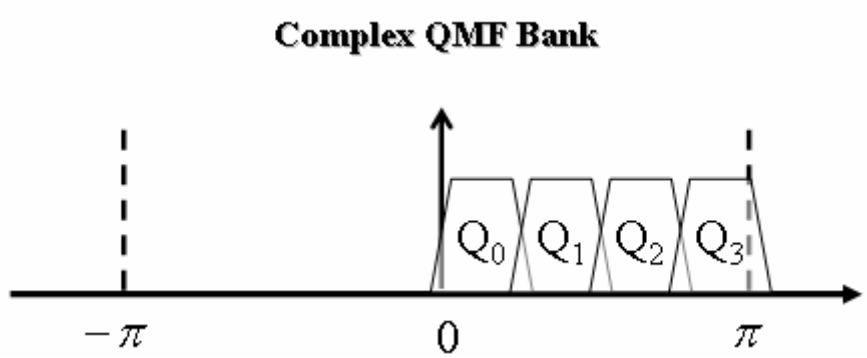

Figure 3: Illustration of 4-channel complex QMF banks

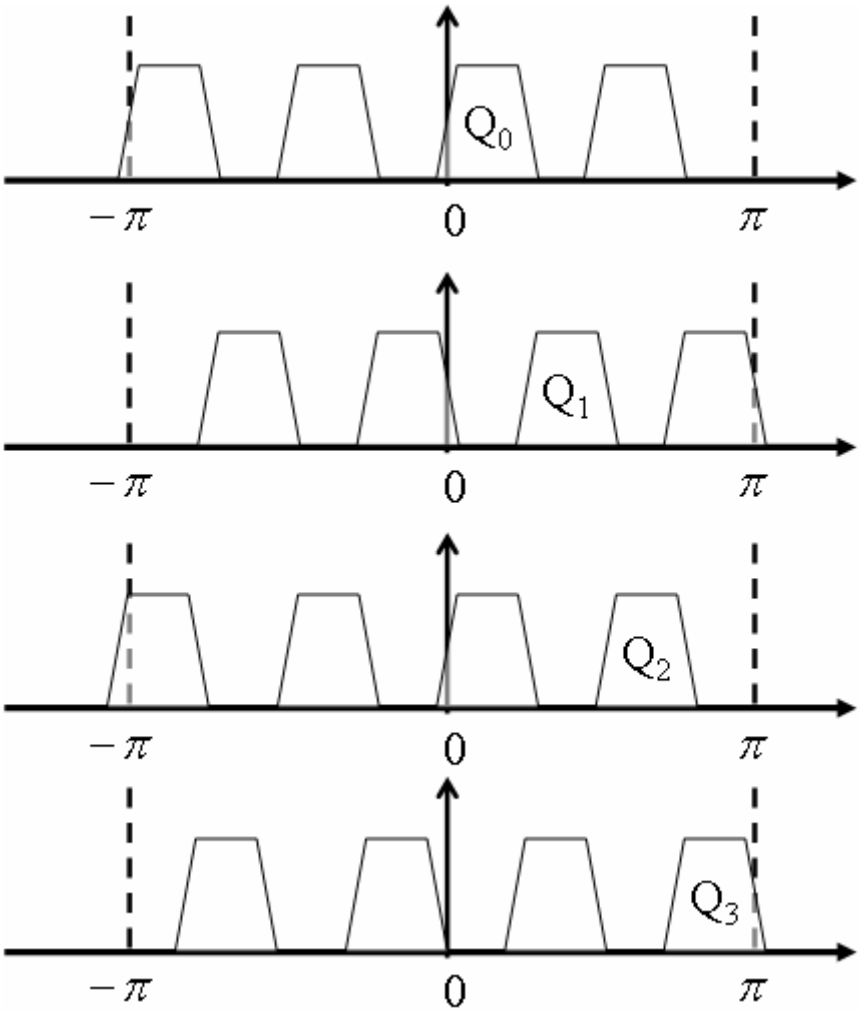

Figure 4: Illustration of no overlap in complex QMF banks

Figure 4 shows the situation after decimating by M to make the system critically down-sampled. As shown in Figure 4, because of the lack of negative frequency components, there is no overlap existing in cosine QMF banks which leads to aliasing term later. Therefore, any further modification on complex subband signal will not introduce any aliasing term which is opposed to cosine QMF bank and this feature makes it suitable to SBR where the high frequency components are synthesized from low frequency part.

Chapter 2
Aliasing Elimination Technique in HE-AAC Low Power Version

As mentioned in Chapter 1, the aliasing term exists if the cosine QMF bank is used and the scaling gain between adjacent subbands is altered. Figure 5 illustrates an example of aliasing term which is marked with black arrows. As shown in Figure 5, the aliasing term resulted from tone-like signal will cause the degradation of audio quality. Hence, in many decoders which adopts cosine QMF bank as essential T/F mapping tool will equip with some kind of auxiliary mechanism to avert aliasing term. Such mechanism employed in LP-SBR [1] is discussed below.

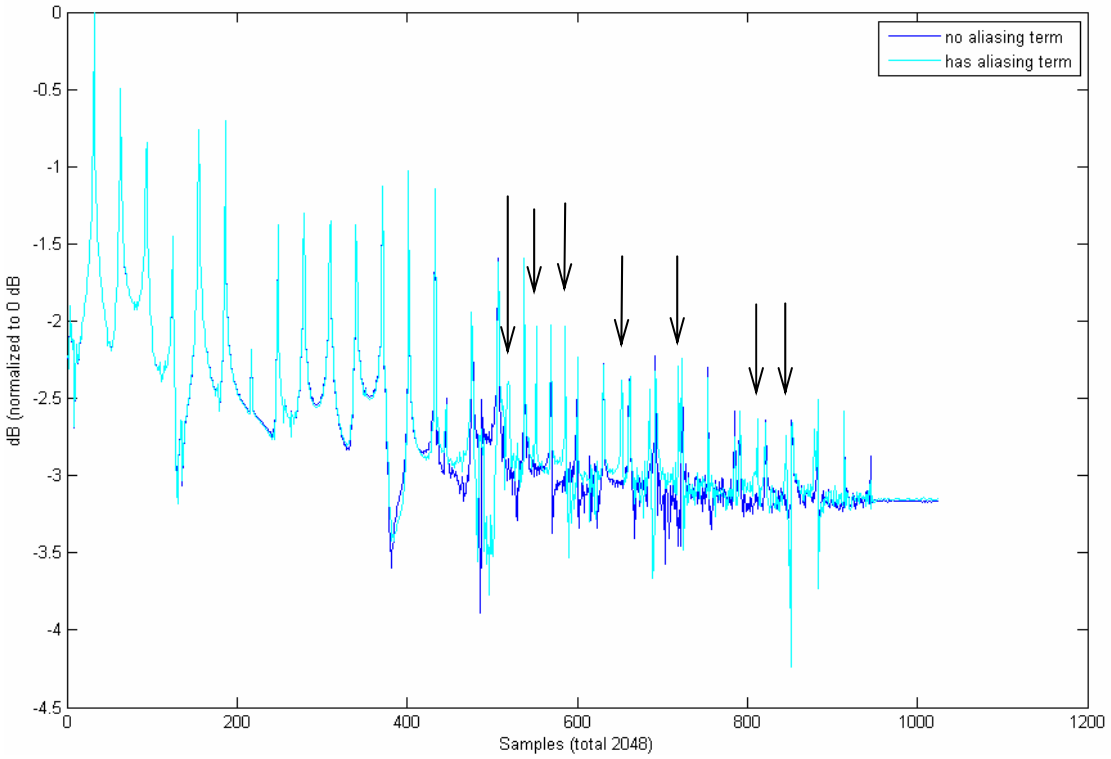

Figure 5: An example of aliasing term (marked with black arrow)

The block diagram of HQ-SBR (SBR High Quality version) and LP-SBR (SBR Low Power version) [1] are shown in Figure 6 and Figure 7. As shown in these two figures, not only the type of QMF bank is different but also there are two more modules in LP-SBR which are used to deal with the aliasing term problem [5]; one is

aliasing detector and the other is aliasing reducer. The purpose of aliasing detector is to detect subbands suffering from severe aliasing term due to strong tone-like components. The detection mechanism exploits 1st order reflection coefficients shown in (5) among three successive subbands, since the shape of QMF bank overlaps adjacent subbands on both sides. The complexity about calculating the coefficients costs little because the linear predictive coefficients are already calculated in HF generator module.

$$ref\ (k) = \begin{cases} \min\left(\max\left(-\frac{\phi_k(0,1)}{\phi_k(1,1)}, -1\right), 1\right) & if\quad \phi_k(1,1) \neq 0 \\ 0 & otherwise \end{cases} \tag{5}$$

In aliasing reducer, the gain of the consecutive detected subbands is equalized together in order to suppress the aliasing term but meanwhile the accuracy of energy envelope will be sacrificed. On the contrary, the gain of undetected subband will remain the same to keep the energy envelope correct.

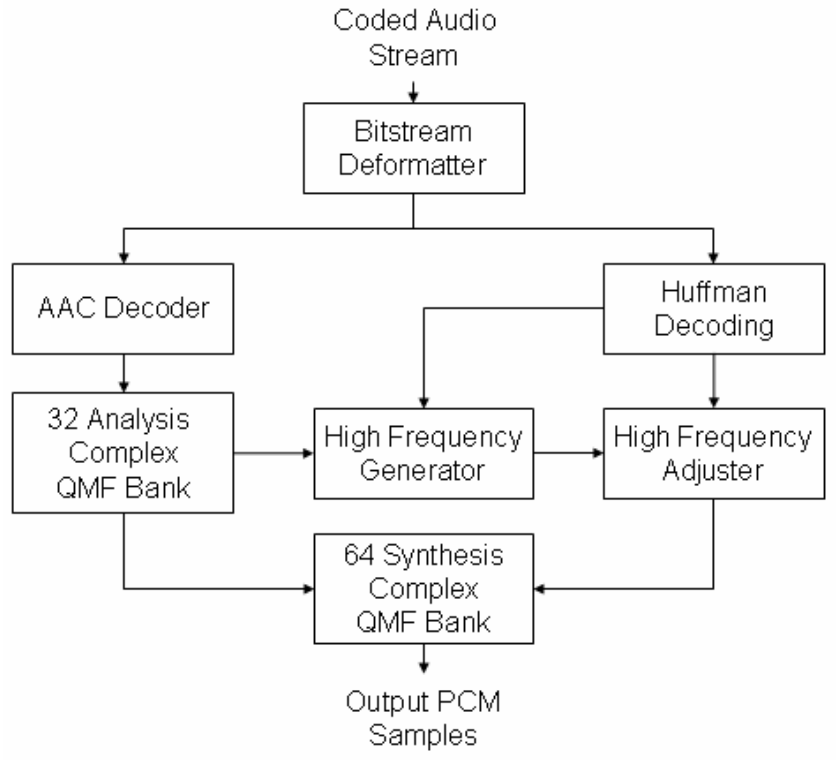

Figure 6: Block diagram of HE-AAC decoder High Quality version

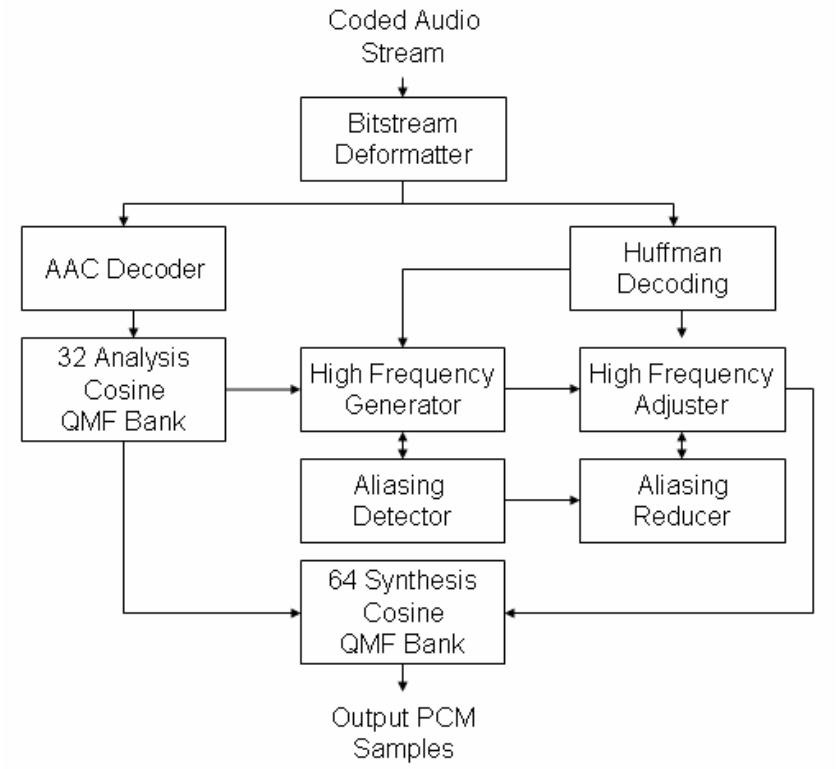

Figure 7: Block diagram of HE-AAC decoder Low Power version

Chapter 3
High Quality Low Power HE-AAC Decoder

As described in Chapter 2, the auxiliary mechanism used to reduce aliasing term in LP-SBR is defective. Forcing the scaling gain identical will lose the accuracy of energy envelope. The mechanism proposed in this chapter will not suffer such impairment. The concept behind proposed mechanism is the same as that in LP-SBR [5]. The concept is that if the subband signals are noise-like, its aliasing term can be ignored and then operated in real-value domain to save computation. On the other hand, if the subband signals are tone-like, its aliasing term needs to be averted to keep audio quality.

3.1 Framework of High Quality Low Power HE-AAC Decoder

The framework referred to as the HQ^{+}LP -SBR and is illustrated in Figure 8. The framework consists of three stages. The first is the “Type Decision of Filter Bank” which first determines the individual subband signal is tone-like or noise-like, and then chooses either complex or real real/synthesis QMF bank adaptively according to the amount of tone-like subbands. The second is the “complexification” or “realification” process applied to each subband to handle the aliasing terms. The third is “image part addition” that add the affection of the image part of complexified subband to the decoded signal.

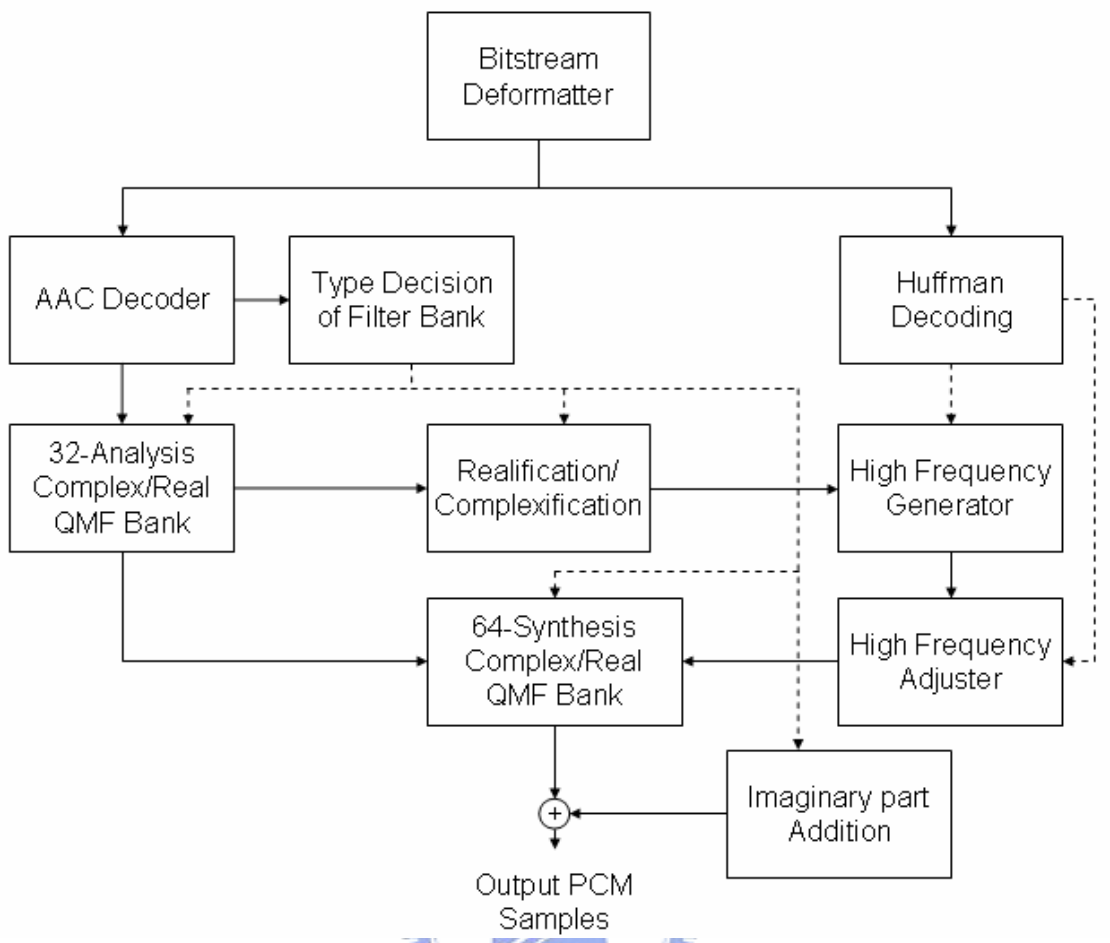

Figure 8: Block diagram of HQ^{+}LP -SBR

3.2 Type Decisioin of Filter Bank

The function of this module is to determine the type of essential filterbank and is accomplished by two steps. One is “subband type decision” that is to determine the type of subband signal as either being tone-like or noise-like. The other is “fiterbank type decision” that determines the type of essential analysis/synthesis fiterbank as either cosine or complex QMF.

Figure 9 shows the flowchart of “subband type decision”. First if the current frame is short window, the type of filterbank is set as real and all subbands are noise-like. This is because if the window sequence is short window, the HF (high frequency) will be unlikely to be tone-like. Otherwise, the magnitudes of 32 decoded MDCT spectral lines in one subband are sorted. Then, the noise-floor is estimated by

$$N = \frac{1}{16}\sum_{k=8}^{23} |X_s[k]|. \qquad (6)$$

where X_s means the sorted data, and the first 8 and the last 8 data are discarded to remove the affection of tonal component and zero-valued re-quantized data. On the other hand, the tonal component is estimated by the mean of the largest 4 data as

$$T = \frac{1}{4}\sum_{k=28}^{31} |X_s[k]| . \tag{7}$$

If the previous frame is not short window, the dB difference of *T* and *N* of current and previous frame are compared to a threshold (here it is 13dB and the reason of such choice will be discussed in 3.5) since the LF (low frequency) part of SBR comes from two AAC frames; if one of dB difference is larger than or equal to the threshold, this subband is marked as tone-like band, otherwise as noise-like band. On the other hand, if previous frame is short window, only the dB difference of *T* and *N* of current frame is compared.

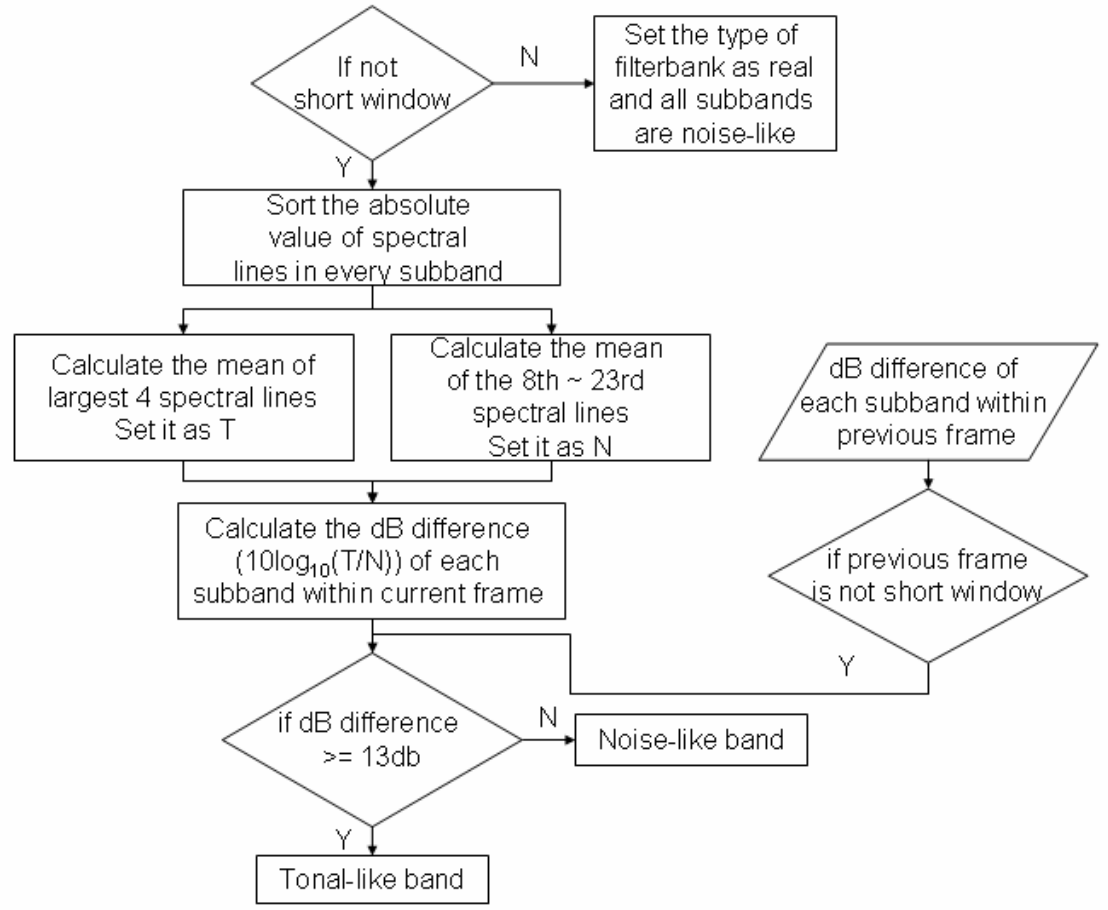

Figure 9: Flowchart of subband signal type decision

After the tone/noise-like types of all the subbands belonging to AAC part have been determined, the essential type of analysis/ synthesis QMF bank can be decided as illustrated in Figure 10. If there are more than 5 subbands determined as tone-like, the essential type of analysis/ synthesis QMF bank is complex QMF, otherwise cosine

QMF. The criterion about the value 5 is based on the consideration of whole computational complexity and will be discussed in Section 3.4.

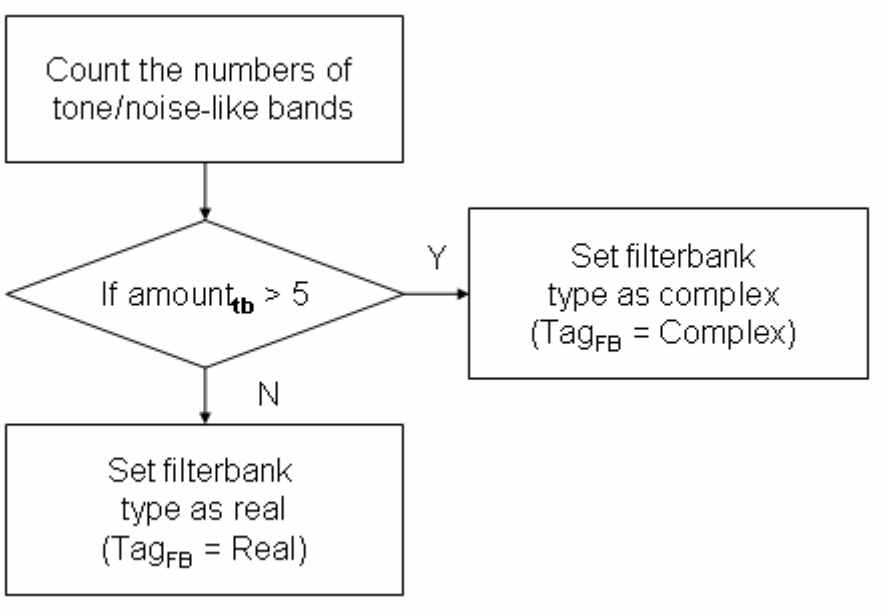

Figure 10: Flowchart of filterbank type decision

3.3 Complexification and Realification

This module is used to complement the deficiency of unitary QMF bank. If the QMF bank is cosine and some subbands are tone-like, the aliasing-term should be handled. The complexification process on such a subband is applied to eliminate the aliasing-term. On the other hand, if the QMF bank is complex and but the subband is noise-like, convert the QMF signals to real-value domain through realification process will lower the complexity. Figure 11 illustrates the flowchart of this module for each subband.

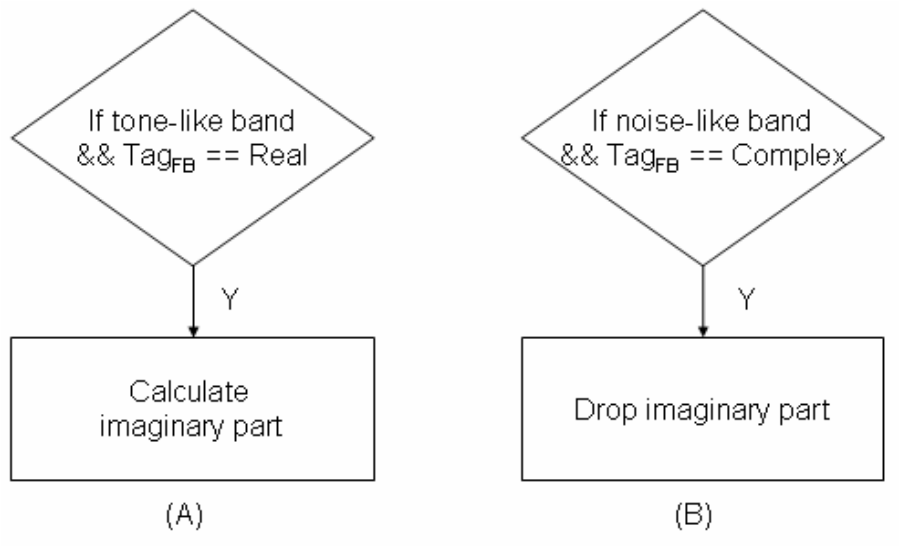

Figure 11: Flowchart of complexification (A) and realification (B)

The main matrix operations in cosine/complex analysis QMF are

$$X_r[k] = \sum_{n=0}^{63} x[n] \cos\left(\frac{\pi}{2M} (k+0.5)(2n-96) \right). \tag{8}$$

and

$$X_c[k] = \sum_{n=0}^{63} x[n] \exp\left(\frac{\pi}{2M} (k+0.5)(2n-96) \right). \tag{9}$$

for k = 0~31. Furthermore, the main matrix operations in cosine/complex synthesis QMF are

$$v_r[n] = \sum_{k=0}^{63} X_r[k] \cos\left(\frac{\pi}{2M} (k+0.5)(2n-64) \right). \tag{10}$$

and

$$v_c[n] = \mathrm{Re}\left\{ \sum_{k=0}^{63} X_c[k] \exp\left(\frac{\pi}{2M} (k+0.5)(2n-64) \right) \right\}. \tag{11}$$

for n = 0~127. From (8) and (9), the complexification and realification process on the analysis QMF band are to recalculate and drop the summation; that is

$$X_i[k] = \sum_{n=0}^{63} x[n] \sin\left(\frac{\pi}{2M} (k+0.5)(2n-96) \right). \tag{12}$$

In the synthesis part, if the essential QMF bank is complex and some subband signals are real, an energy adjustment on such subbands should be applied to compensate the absent of the image part. On the other hand, (10) and (11) can be represented as

$$v_c[n] = v_r[n] - \sum_{k=0}^{63} X_i[k] \sin\left(\frac{\pi}{2M} (k + 0.5)(2n - 64) \right). \quad (13)$$

Hence, if the essential QMF bank is cosine and the subband signals are complex, the mechanism of the block "Imaginary Part Addition" in Figure 8 is to add the summation of the products of the imaginary part of subband signals and sine coefficients. Then the decoded time-domain signals are obtained by subtracting the output of original cosine QMF bank with the output of "Imaginary Part Addition".

Additionally, there are three adjustments in the process. First, the time phase parameters of complex QMF should be the same as those of cosine QMF to achieve aliasing cancellation simultaneously. Second, the energy of HF real-valued subbands has to be divided by $\sqrt{2}$ as LP-SBR to avoid excessive energy problem. Third, the LF part that is not suffered from gain adjustment can be synthesized through cosine QMF bank to reduce the complex.

3.4 Complexity for Complexification and "Imaginary part Addition"

The direct computation of the summation (12) for the image part needs 64 multiplications. However, there are symmetric properties:

$$S[k, 32 - n] = S[k, n], \quad for \quad n = 0 \sim 31. \quad (14)$$

and $$S[k, 64 - n] = -S[k, 32 + n], \quad for \quad n = 0 \sim 31. \quad (15)$$

where $S[k, n]$ denotes the sine coefficients. There are thirty three different values in the sine coefficients. Hence, an image part needs only 33 multiplications by

$$\begin{aligned} \sum_{n=0}^{63} x[n] S[k, n] = & \sum_{n=0}^{15} (x[n] + x[32 - n]) S[k, n] \\ & + x[16] S[k, 16] + x[48] S[k, 48] \\ & + \sum_{n=1}^{15} (x[32 + n] - x[64 - n]) S[k, n] \end{aligned} \quad (16)$$

On the other hand, (13) can be formulated with the matrix form:

$$\vec{v} = \hat{C} \cdot \vec{X}_r - \hat{S} \cdot \vec{X}_i. \quad (17)$$

The form can be represented as

$$\varpi_c = \sum_{k=0}^{63} X_r[k] \cdot \hat{C}_k - \sum_{k=0}^{63} X_i[k] \cdot \hat{S}_k \,. \tag{18}$$

where $\hat{C}$ and $\hat{S}$ are cosine and sine matrix of size 128×64 defined by

$$\hat{C}[n,k] = \cos\left(\frac{\pi(k+0.5)(2n-64)}{128}\right). \tag{19}$$

and $$\hat{S}[n,k] = \sin\left(\frac{\pi(k+0.5)(2n-64)}{128}\right). \tag{20}$$

for k = 0~63 and n = 0~127. $\hat{C}_k$ and $\hat{S}_k$ mean the columns of $\hat{C}$ and $\hat{S}$. Therefore, an image component $X_i[k]$ needs 128 multiplications for $X_i[k] \cdot \hat{S}_k$ in (18). Similarly, by the facts

$$\hat{S}[k,64-n] = -\hat{S}[k,n] \,, \quad for \quad n = 0 \sim 63 \,. \tag{21}$$

and $$\hat{S}[k,128-n] = -\hat{S}[k,64+n] \,, \quad for \quad n = 0 \sim 63 \,. \tag{22}$$

there are only 65 different values in $\hat{S}_k$. Hence, an image component $X_i[k]$ needs only 65 multiplications for $X_i[k] \cdot \hat{S}_k$ in (18).

From the fast algorithms defined in [7] and [8], the total multiplications of the matrix operations needed in LP-SBR are 80 + 192 and is 256 + 512 in HQ-SBR. The multiplications required for one subband sample to be complexified and synthesized back to time-domain signal is 33 + 65. Hence, the maximum number of subbands that can be applied complexification is 5 under the constraint that the whole computational complexity of cosine QMF bank and complexification process can't outstrip the complex QMF bank

3.5 Threshold Evaluation

In this section, the reason of 13dB used in section 3.2 as a threshold to determine whether a subband is tone-like or noise-like is elaborated. Recalling the foreword given in the start of this chapter, this threshold-13dB can be considered as "quality and performance" threshold. Hence, this threshold should be able to correctly label all potential tone-like and noise-like subbands. The approach used here to prove this threshold is right is to statistically count up the dB difference of subbands belonging

to AAC part from tone and noise frames. Figure 12 illustrates the associated flowchart. In Figure 12, the key point in this approach is the way to properly decide tone frame and noise frame channel by channel. Here, the MOV (model output variable) -NMR (Noise-To-Mask-Ratio) - of PEAQ system (perceptual evaluation of audio quality) is used. In Figure 12, there are two kinds of frame NMR. One is NMR_c which comes from frame NMR of HQ-SBR and the other is NMR_r which comes from the frame NMR of SBR with cosine QMF bank equipped no auxiliary mechanism to avert aliasing term. Obviously, the NMR_c will be smaller than NMR_r because of lack of aliasing term and the value of NMR is always negative.

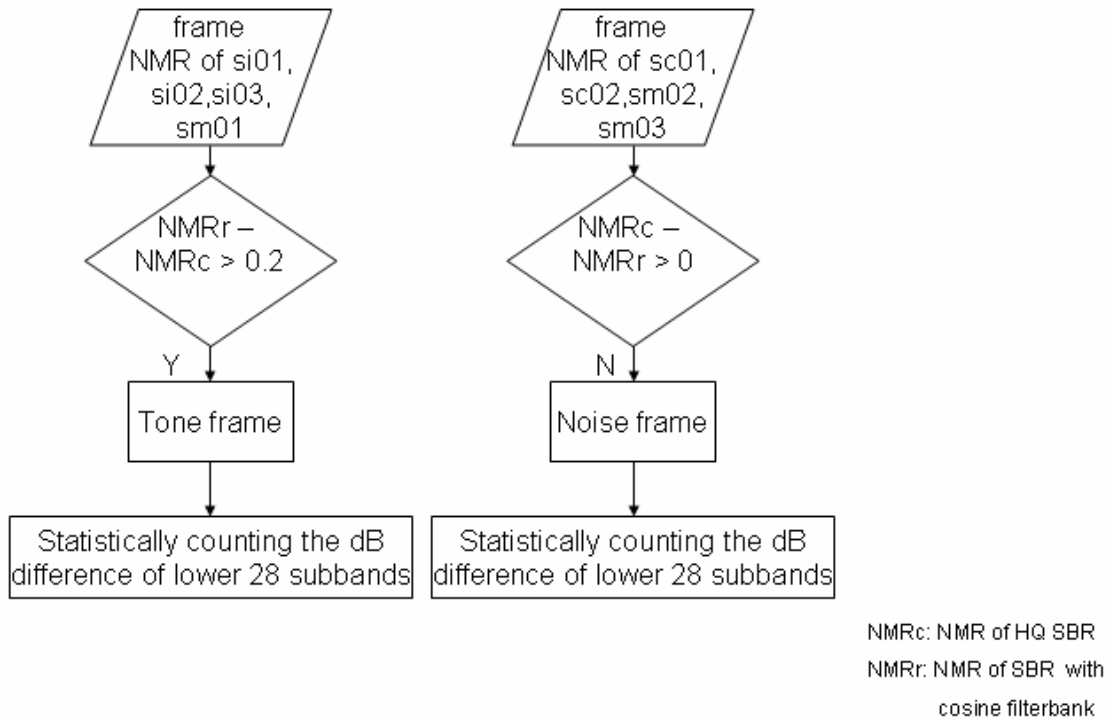

Figure 12: Flowchart of threshold evaluation

For now the mapping from AAC frame to NMR frame needs to explain ue to the one-to-many mapping by the upsampling in SBR. Figure 13 shows the relationship between AAC frame and NMR frame. In Figure 13, A_x denotes one AAC frame and N_x denotes one NMR frame. As shown in Figure 13, one AAC frame is crossing three NMR frames. Therefore, the NMR_c or NMR_r of one AAC frame is averaging from three NMR frames.

Finally, choosing si01, si02, si03 and sm01 as test songs for tone frame and sc01, sc02, sm02 and sm03 for noise frame is because they are representatives in their own category.

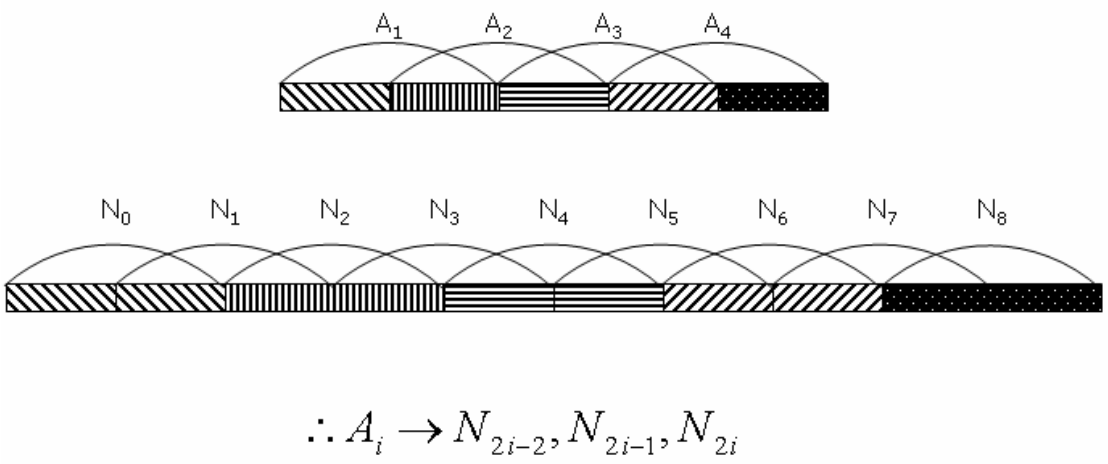

Figure 13: Relationship between AAC frame and NMR frame

Figure 14 illustrates the probability distribution of dB difference of tone and noise frames where the blue line represents for tone frame and red line represents for noise frame. As shown in Figure 14, the majority of dB difference of tone frame surrounds between 50~80 dB and the minority is between 0~30 dB which is because the approach used here not only counts in tone-like subbands but also noise-like subbands while there is a potential tonal components appearing in tone frame. And the minority of dB difference of noise frame surrounding between 50~80 dB can be neglected because the peak is around 0.02 and far smaller while comparing to the peak of majority.

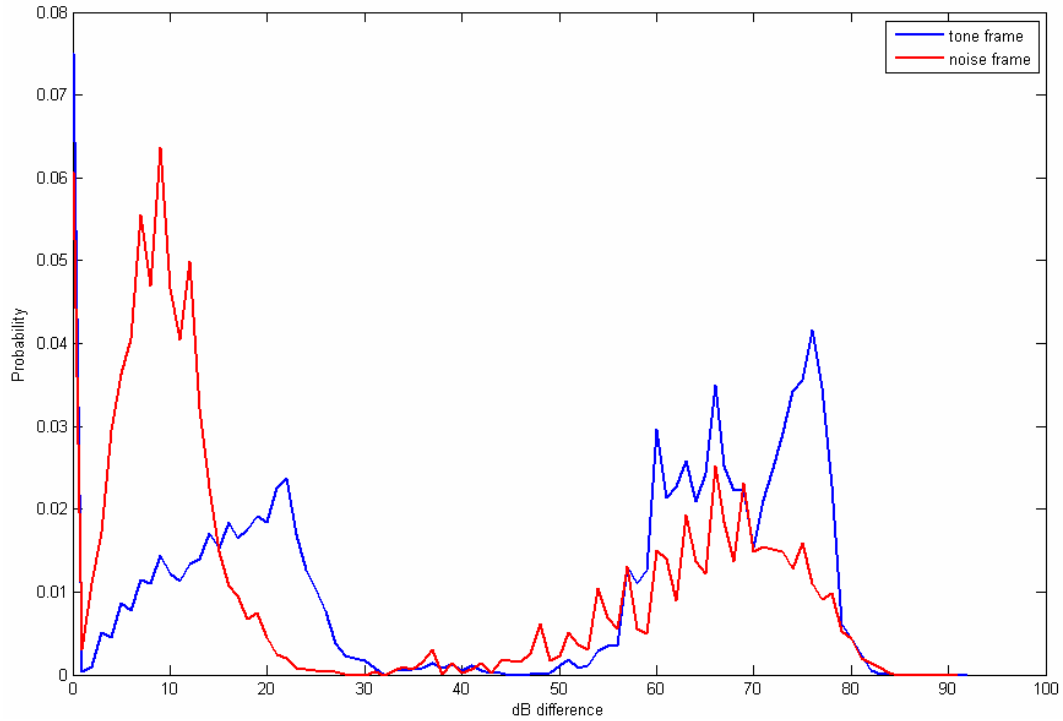

Figure 14: Probability distribution of dB difference of tone and noise frame

Figure 14 also indicates that subbands around 0~20 dB difference are harmless and have not to be complex. On the other hand, subbands around 55~80 dB difference are important to quality and needed to be complex to avert aliasing term. And Figure 15 also shows that the average ODG is increasing while threshold increases from 10 to 70. Although the average ODG seems good at 80 dB difference, individual ODG of tonal signal degenerates extensively. Hence, a receivable threshold may lie between 10 ~ 20 dB difference. Therefore, from Figure 16, the choice of 13 dB as threshold used in Section 3.2 is the best balance between quality and performance.

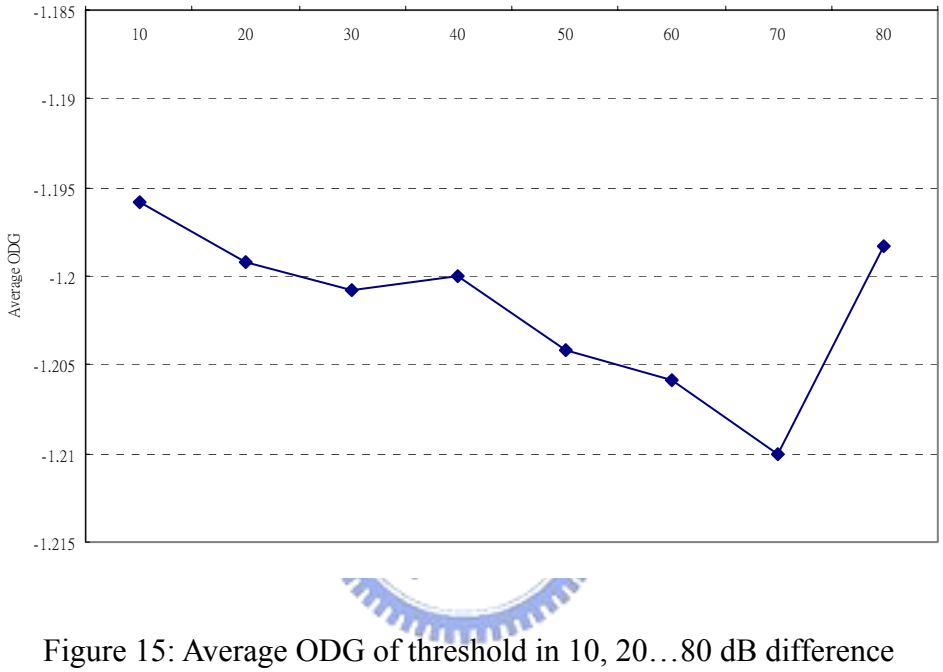

Figure 15: Average ODG of threshold in 10, 20…80 dB difference

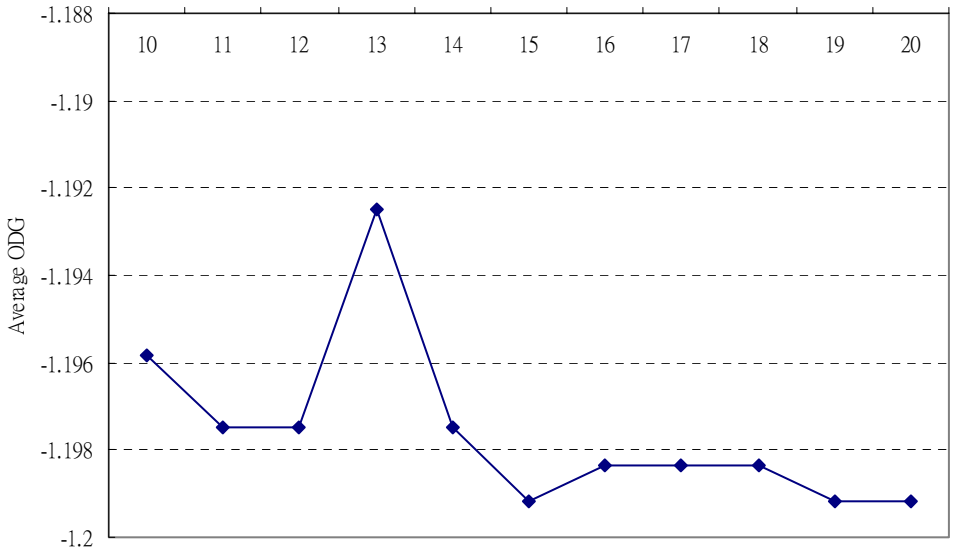

Figure 16: Average ODG of threshold between 10 ~ 20 dB difference

3.6 Merits of HQ$^+$LP-SBR

From the perspective of quality, because the aliasing-free property fulfilled in HQ$^+$LP-SBR is based on the intrinsic non-aliasing attribute which complex QMF bank has, the expected audio quality will approach HQ-SBR.

As far as computational complexity is concerned, because applying realification to some noise-like subband leads to real-valued operation in the following module when the type of essential QMF bank is complex, the total computational complexity will be smaller than HQ-SBR. And if the type of the essential QMF bank is cosine, due to the chosen criterion described in Section 3.4 plus some noise-like subband which will operate in real-valued domain in the subsequent procedures guaranteeing the total computational complexity which will be smaller than HQ-SBR.

Chapter 4
Low Power HE-AAC v2 Decoder

Chapter 3 describes the the built-up of a high quality but low power HE-AAC decoder. In this chapter, the parametric stereo coding tool [2] is fitted into the framework of HQ$^+$LP-SBR to create a low power HE-AAC version 2 decoder. The brand new low power HE-AAC version 2 decoder is referred to as the HQ$^+$LP-PS. Because HQ$^+$LP-PS is extended from HQ$^+$LP-SBR, it inherits the merits of HQ$^+$LP-SBR which means possessing the property of low power and high quality at the same time. But since in original PS, gain adjustment occurs in the LF part as well as HF part. Therefore, in HQ$^+$LP-PS, if the subband is tone-like no matter it locates in LF part or HF part, it should be complex to avert aliasing term which means the LF part can not just be synthesized through cosine QMF bank. And in HQ$^+$LP-PS, as long as the subband is real-valued the energy of it has to be divided by $\sqrt{2}$ to avoid excessive energy problem.

4.1 Framework of Low Power HE-AAC v2 Decoder

The frameworks of original PS and HQ$^+$LP-PS are illustrated in Figure 17 and Figure 18 respectively. By comparing these two figures, the new added blocks in HQ$^+$LP-PS are “Subband Type Decision” and “Realification”. Except for “Subband Type Decision”, the implementation of ”Realification” is equal to that in HQ$^+$LP-SBR. The necessary modifications in “Subband Type Decision” are described in Section 4.2.

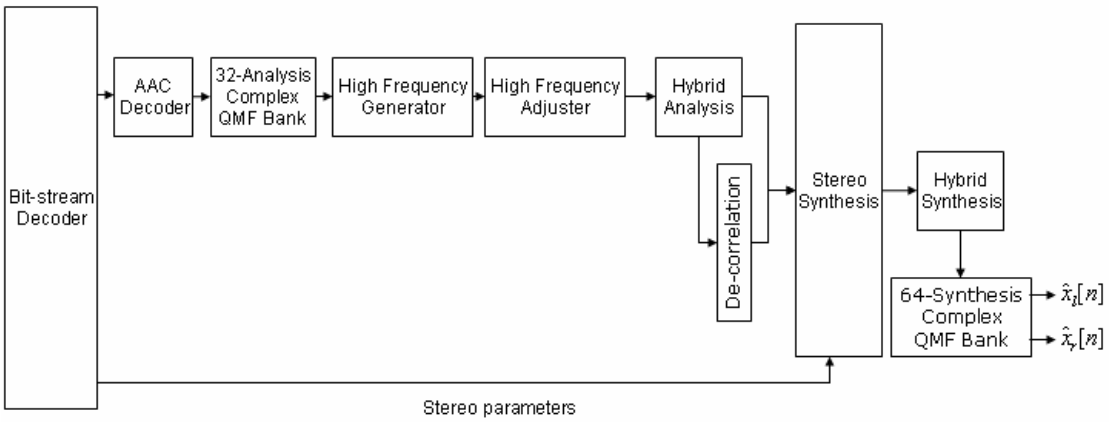

Figure 17: Block diagram of original PS

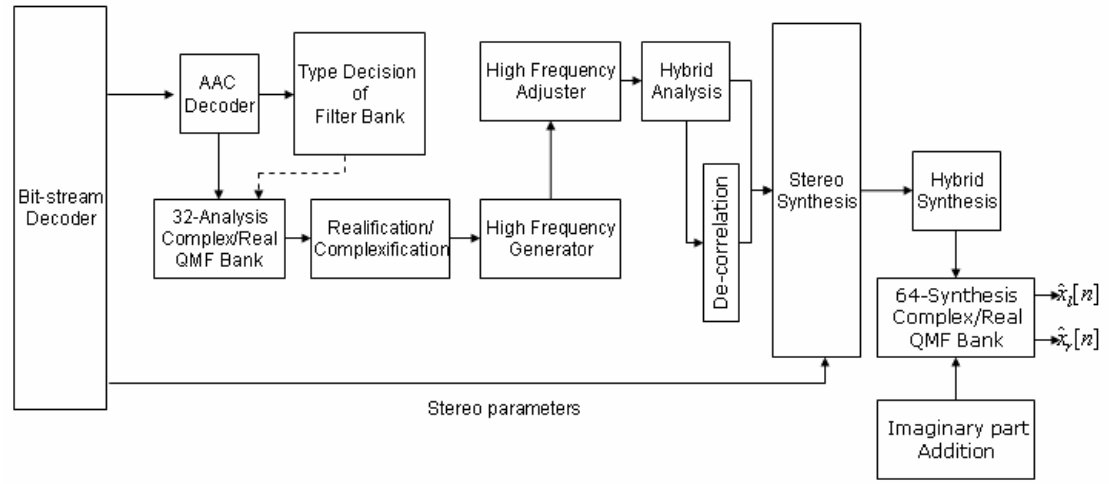

Figure 18: Block diagram of HQ$^+$LP -PS

4.2 Modification in Subband Type Decision

Figure 19 illustrates the flowchart of "Subband Type Decision". As shown in Figure 19, the way of calculating *T* and *N* and determining the type of subband signal is the same as the way used in Section 3.2. The differences between Figure 9 and Figure 19 are the short window and threshold. In HQ$^+$LP-SBR, there is another path if the current frame is short window which is because if the window sequence is short window, the HF will be unlikely to be tone-like. But in HQ$^+$LP-PS, the gain of LF part will be adjusted to keep the accuracy of energy envelope and even the window sequence is short window, there could be some tone-like subband in LF part. Therefore, in HQ$^+$LP-PS, no matter current frame is short window or not, the dB

difference of *T* and *N* of current and previous frame are compared to a threshold and if previous frame is short window, only the dB difference of *T* and *N* of current frame is compared. Moreover since the LF part is more sensitive to human ear than HF part, the threshold should be more restricted to keep quality.

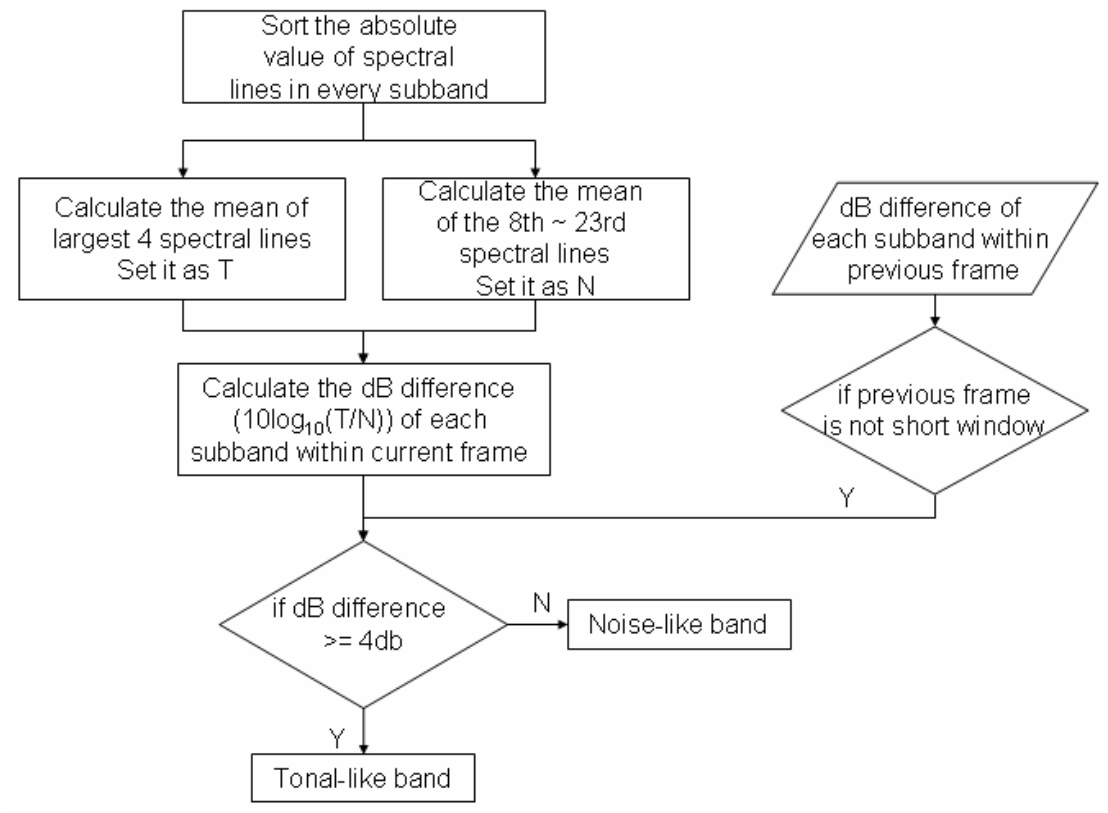

Figure 19: Flowchart of subband signal type decision

.

Chapter 5
Quality Assessment

5.1 Objective Quality Measurement Environment

For objective quality evaluation, the PEAQ system [9] (perceptual evaluation of audio quality) which is the recommendation system by ITU-R Task Group 10/4 is adopted. 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. Following experiments are based on this PEAQ system. The twelve test tracks recommended by MPEG are shown in Table 1.These tracks include the critical music balancing on the percussion, string, wind instruments, and human vocal.

Table 1: 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.

5.2 Objective Quality Measurement in HQ+LP-SBR

In order to verify the decoder adopting the new approach can reach the objective sound quality as the HQ-SBR and is better than LP-SBR, the ODG results of HQ-SBR, HQ+LP-SBR and LP-SBR are compared at the same time and NCTU-HEAAC[10] is adopted as platform.

Table 2: ODG of HQ-SBR, HQ+LP-SBR and LP-SBR at 96kbps

Codec	NCTU-HEAAC		
Bit Rate	96kbps		
Tracks	HQ-SBR	HQ+LP-SBR	LP-SBR
es01	-0.6	-0.61	-0.75
es02	-0.5	-0.53	-0.62
es03	-0.72	-0.73	-0.87
sc01	-0.72	-0.74	-0.78
sc02	-0.93	-0.96	-1.01
sc03	-0.91	-0.93	-1
si01	-1.2	-1.23	-1.26
si02	-0.81	-0.81	-0.94
si03	-1.22	-1.27	-1.36
sm01	-1.27	-1.35	-1.35
sm02	-1.32	-1.29	-1.34
sm03	-1.1	-1.13	-1.2
Max	-0.5	-0.53	-0.62
Min	-1.32	-1.35	-1.36
Average	-0.941666667	-0.965	-1.04

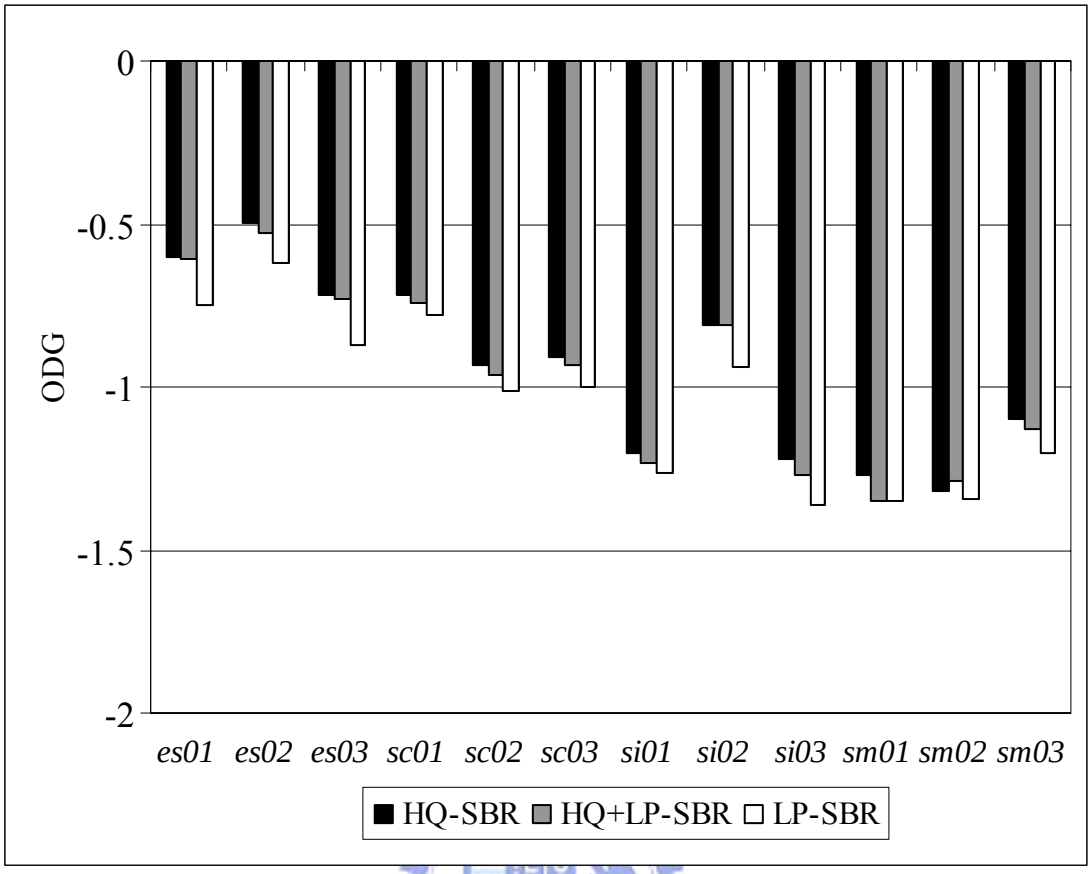

Figure 20: ODG of HQ-SBR, HQ^{+}LP-SBR and LP-SBR at 96kbps

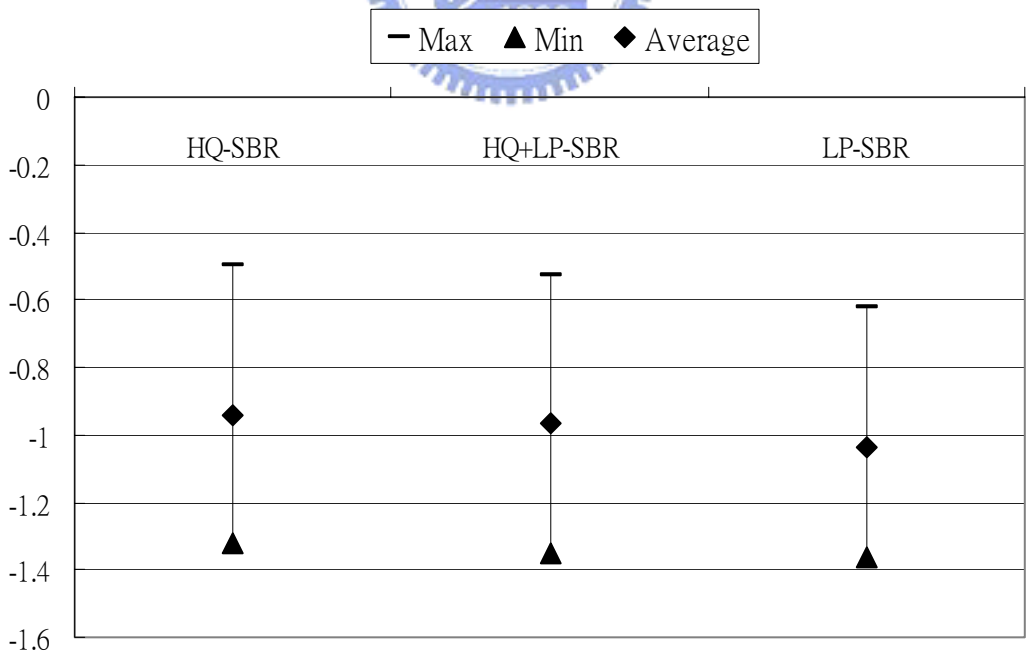

Figure 21: The variance in the ODG of HQ-SBR, HQ^{+}LP-SBR and LP-SBR at 96kbps

Table 3: ODG of HQ-SBR, HQ+LP-SBR and LP-SBR at 80kbps

Codec	NCTU-HEAAC		
Bit Rate	80kbps		
Tracks	HQ-SBR	HQ+LP-SBR	LP-SBR
es01	-0.68	-0.69	-0.81
es02	-0.59	-0.6	-0.7
es03	-0.79	-0.8	-0.94
sc01	-0.97	-1	-1.07
sc02	-1.24	-1.27	-1.32
sc03	-1.17	-1.18	-1.25
si01	-1.6	-1.59	-1.63
si02	-1	-1.01	-1.15
si03	-1.6	-1.64	-1.75
sm01	-1.58	-1.62	-1.64
sm02	-1.61	-1.57	-1.67
sm03	-1.32	-1.35	-1.42
Max	-0.59	-0.6	-0.7
Min	-1.61	-1.64	-1.75
Average	-1.179166667	-1.193333333	-1.279166667

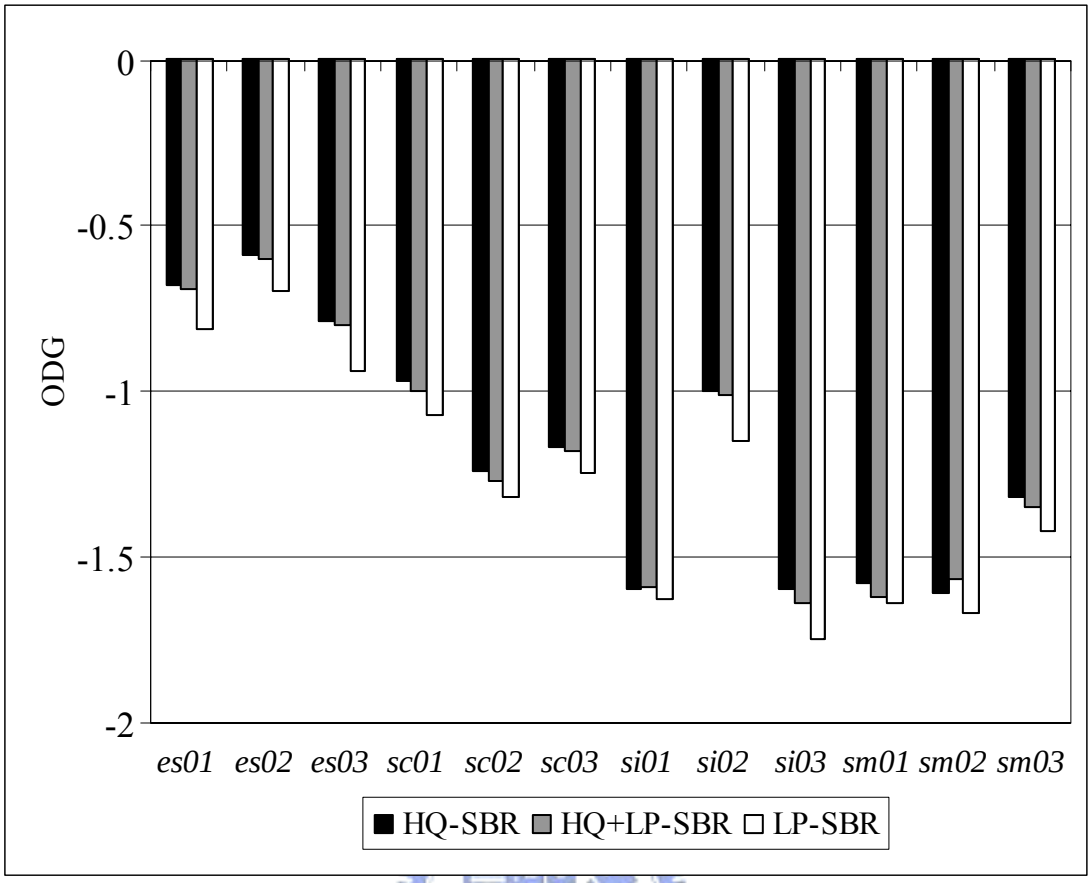

Figure 22: ODG of HQ-SBR, HQ+LP-SBR and LP-SBR at 80kbps

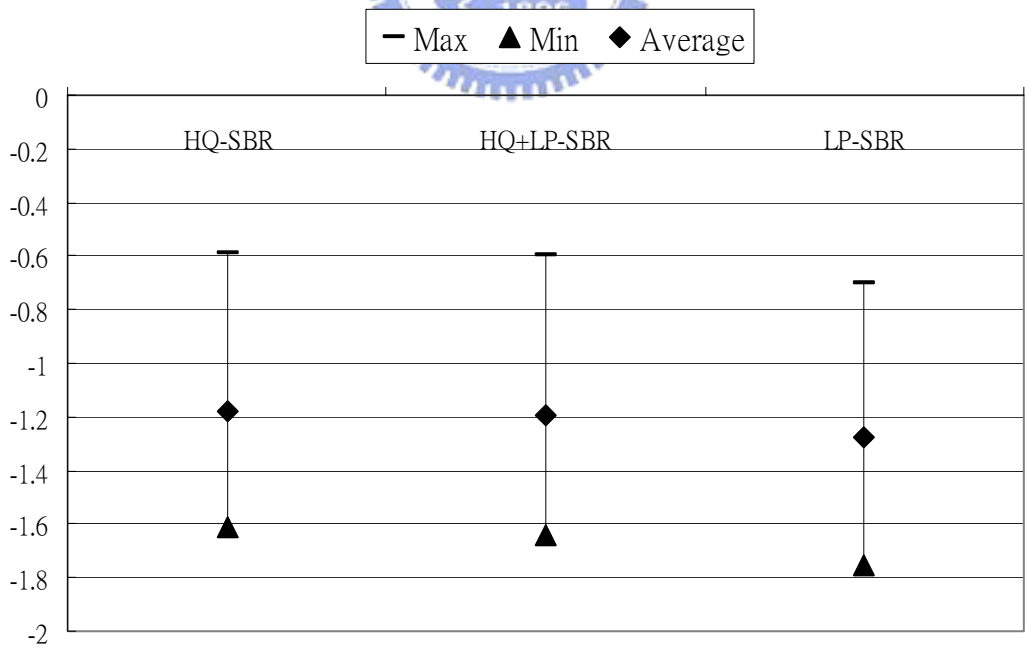

Figure 23: The variance in the ODG of HQ-SBR, HQ+LP-SBR and LP-SBR at 80kbps

Table 4: ODG of HQ-SBR, HQ$^+$LP-SBR and LP-SBR at 64kbps

Codec	NCTU-HEAAC		
Bit Rate	64kbps		
Tracks	HQ-SBR	HQ$^+$LP-SBR	LP-SBR
es01	-0.96	-0.98	-1.07
es02	-0.84	-0.86	-1
es03	-1.03	-1.04	-1.16
sc01	-1.57	-1.65	-1.75
sc02	-1.71	-1.75	-1.77
sc03	-1.61	-1.63	-1.71
si01	-1.95	-1.97	-2.02
si02	-1.39	-1.42	-1.61
si03	-2.06	-2.15	-2.11
sm01	-2.13	-2.14	-2.24
sm02	-2.19	-2.22	-2.28
sm03	-1.71	-1.73	-1.77
Max	-0.84	-0.86	-1
Min	-2.19	-2.22	-2.28
Average	-1.595833333	-1.628333333	-1.7075

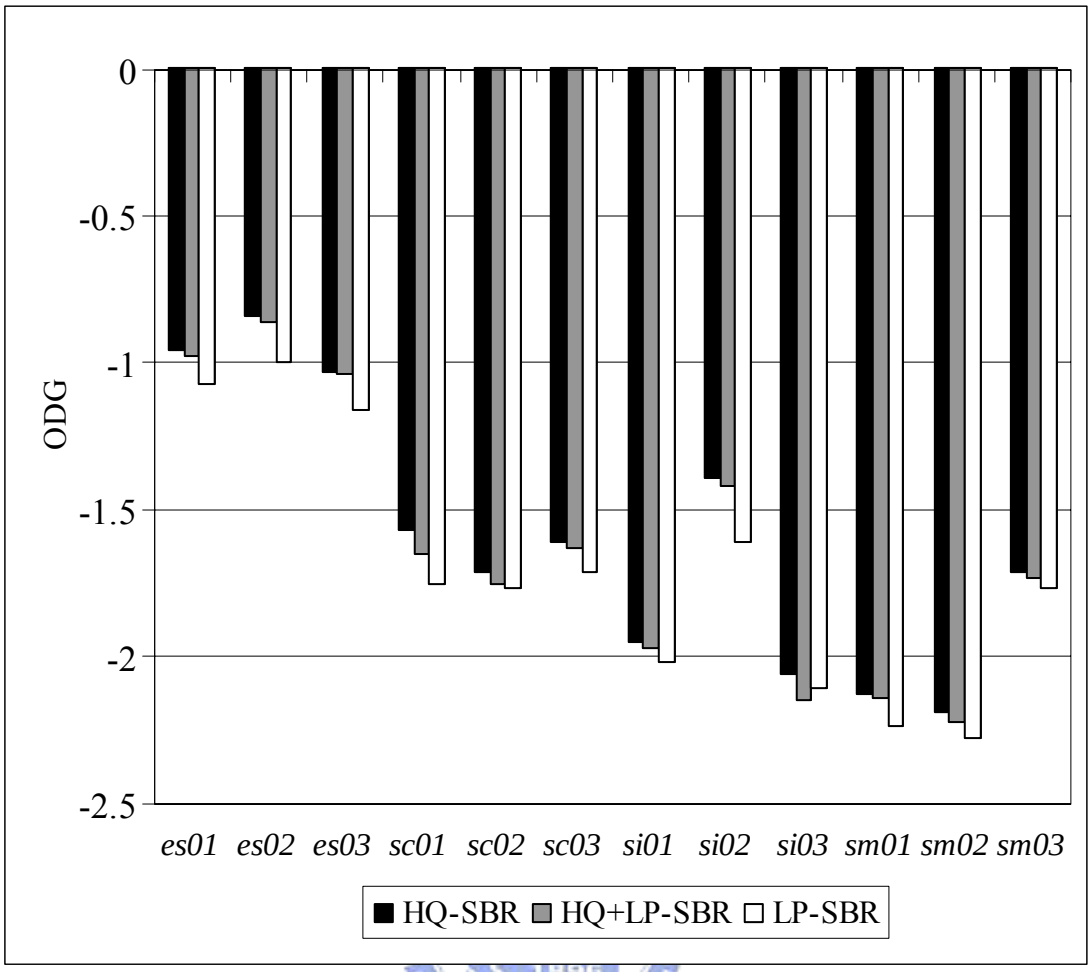

Figure 24: ODG of HQ-SBR, HQ^{+}LP-SBR and LP-SBR at 64kbps

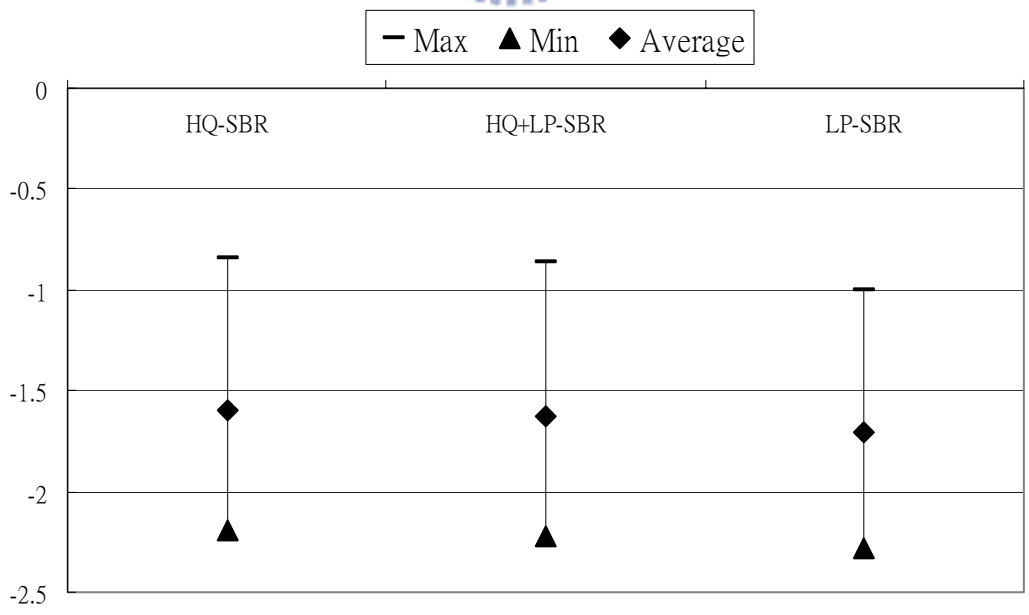

Figure 25: The variance in the ODG of HQ-SBR, HQ^{+}LP-SBR and LP-SBR at 64kbps

HQ+LP-SBR is verified under the target bit-rate of SBR. Under each bit-rate, the resulted ODG of HQ-SBR, HQ+LP-SBR and LP-SBR are compared.The presented experiment data shows that the performance of HQ+LP-SBR meets our expectation which means that in the best case, its objective sound quality equals to that of HQ-SBR and in the worst case equals to LP-SBR. Generally speaking, in most cases the ODG of HQ+LP-SBR will lie between that of HQ-SBR and LP-SBR and the average ODG of HQ+LP-SBR is almost the same as HQ-SBR.

On the other hand, Figure 26 ~ Figure 28 illustrates the spectrum of one frame decoded by HQ-SBR, LP-SBR and HQ+LP-SBR. Figure 29 ~ Figure 31 are another example reflecting the same result obtained in ODG from the perspective on frequency domain. In Figure 27 and Figure 30, the spectrum of HQ-SBR and HQ+LP-SBR are illustrated and compared with each other to display that the energy floor of the two is almost the same. Moreover, Figure 26 and Figure 29 show the spectrum of HQ-SBR and LP-SBR at the same time to demonstrate that the energy floor of LP-SBR is sometimes higher or lower than that of HQ-SBR which is resulted from the constraint-consistent scaling gain- LP-SBR must satisfy. In Figure 28 and Figure 31, the spectrum of LP-SBR and HQ+LP-SBR is compared to manifest that HQ+LP-SBR will not suffer from the consistent scaling gain problem.

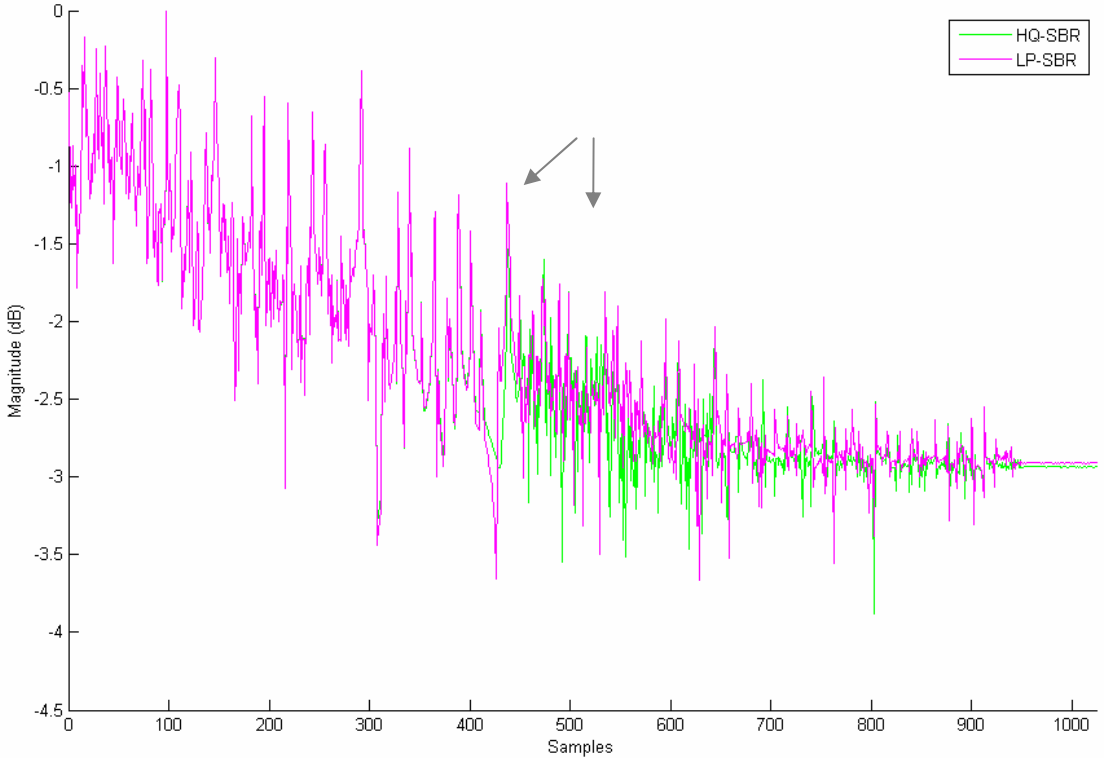

Figure 26: The spectrum of signal decoded by HQ-SBR and LP-SBR

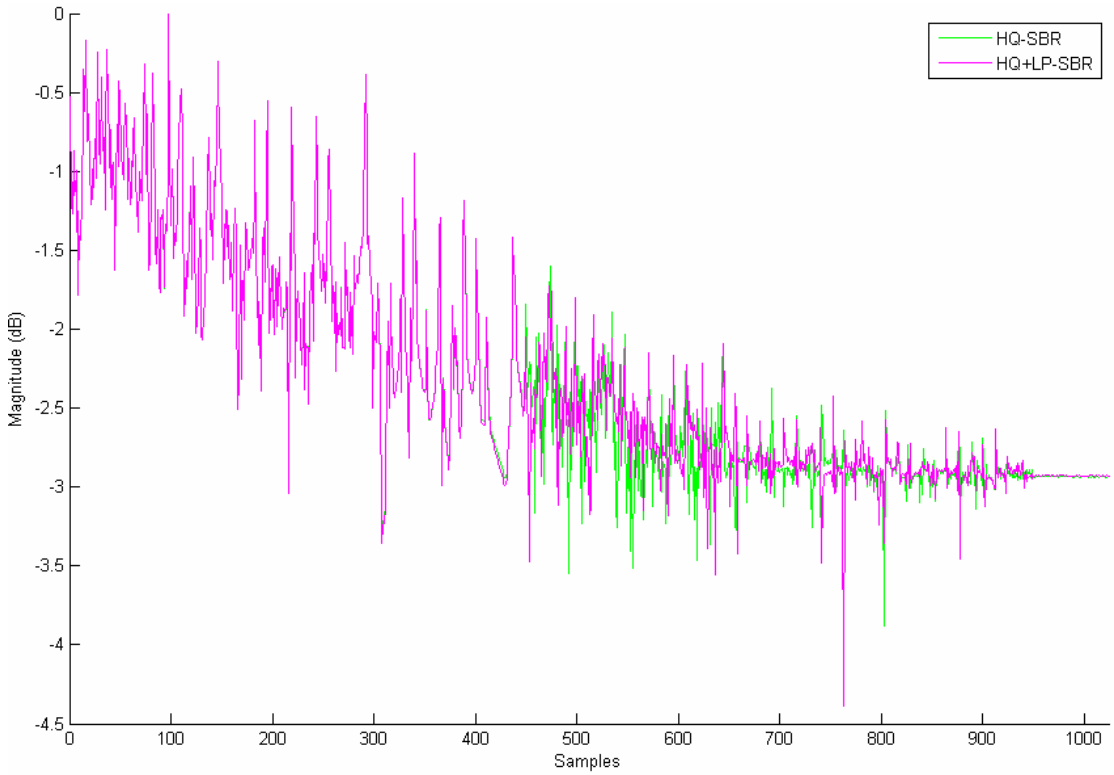

Figure 27: The spectrum of signal decoded by HQ-SBR and HQ$^+$LP-SBR

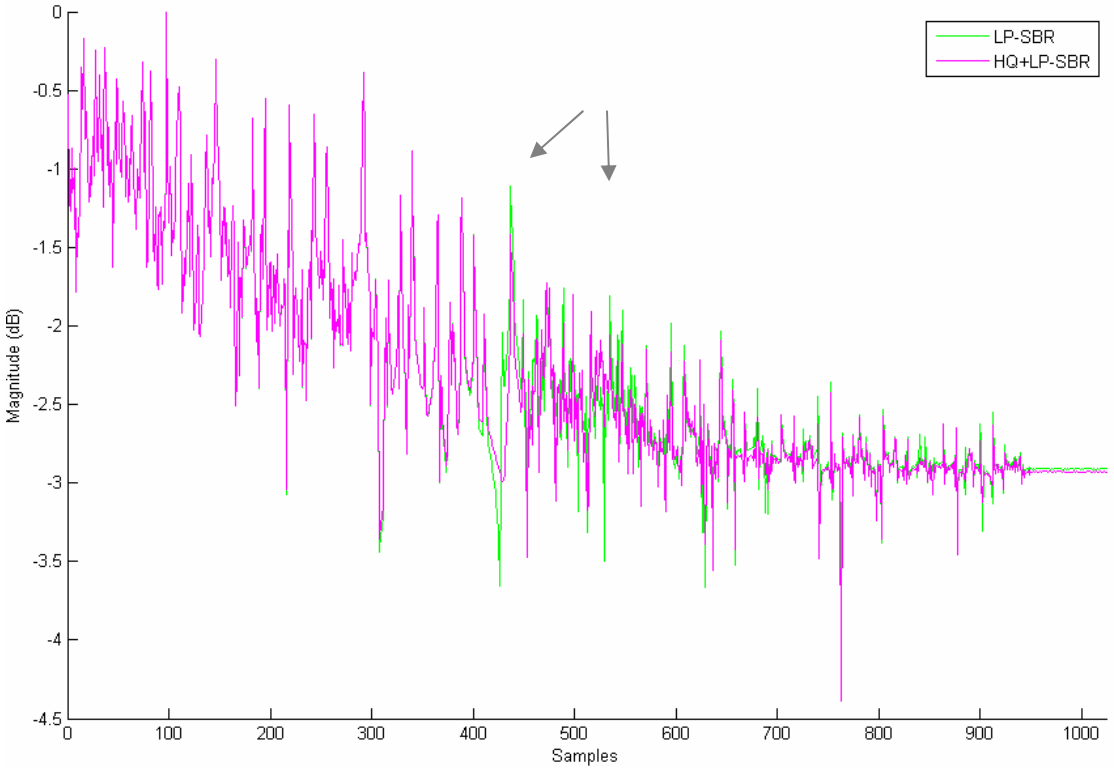

Figure 28: The spectrum of signal decoded by LP-SBR and HQ$^+$LP-SBR

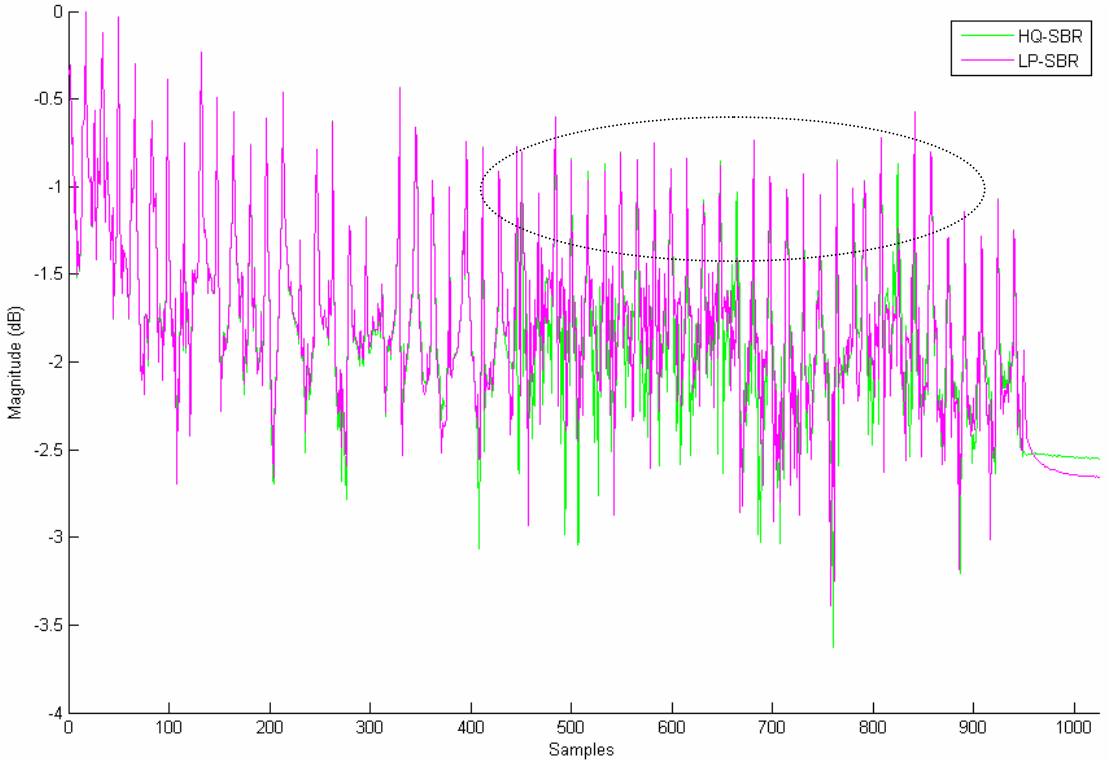

Figure 29: The spectrum of signal decoded by HQ-SBR and LP-SBR

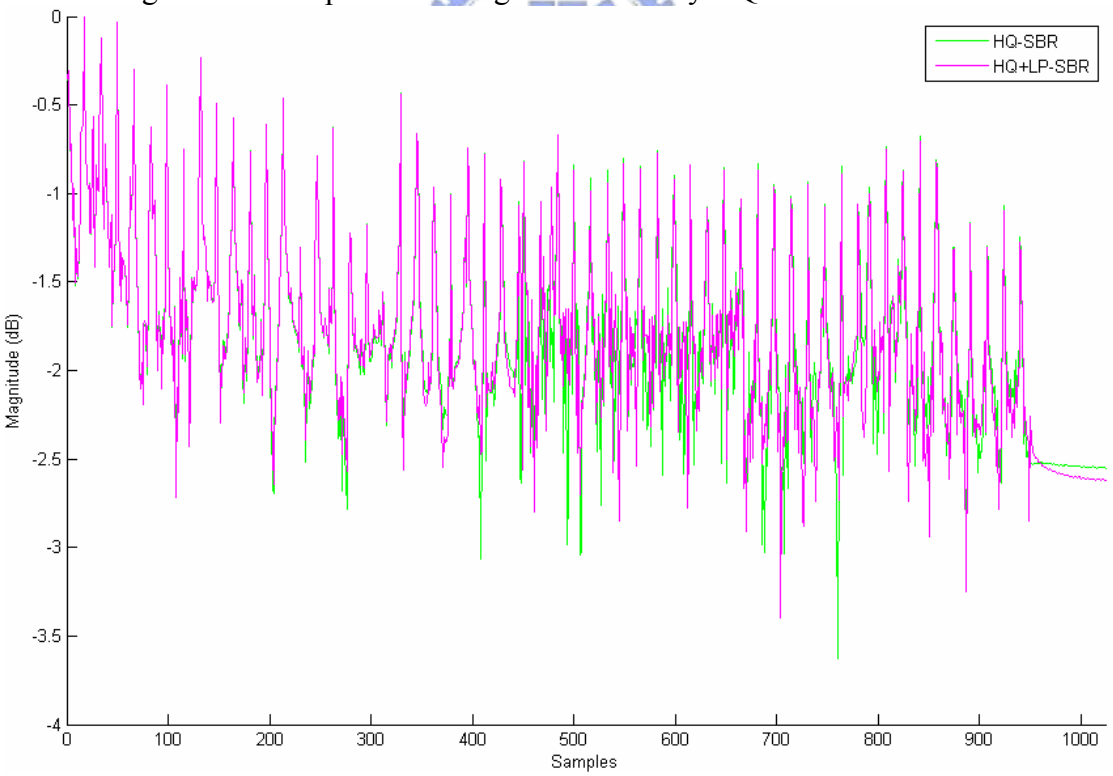

Figure 30: The spectrum of signal decoded by HQ-SBR and HQ^{+}LP-SBR

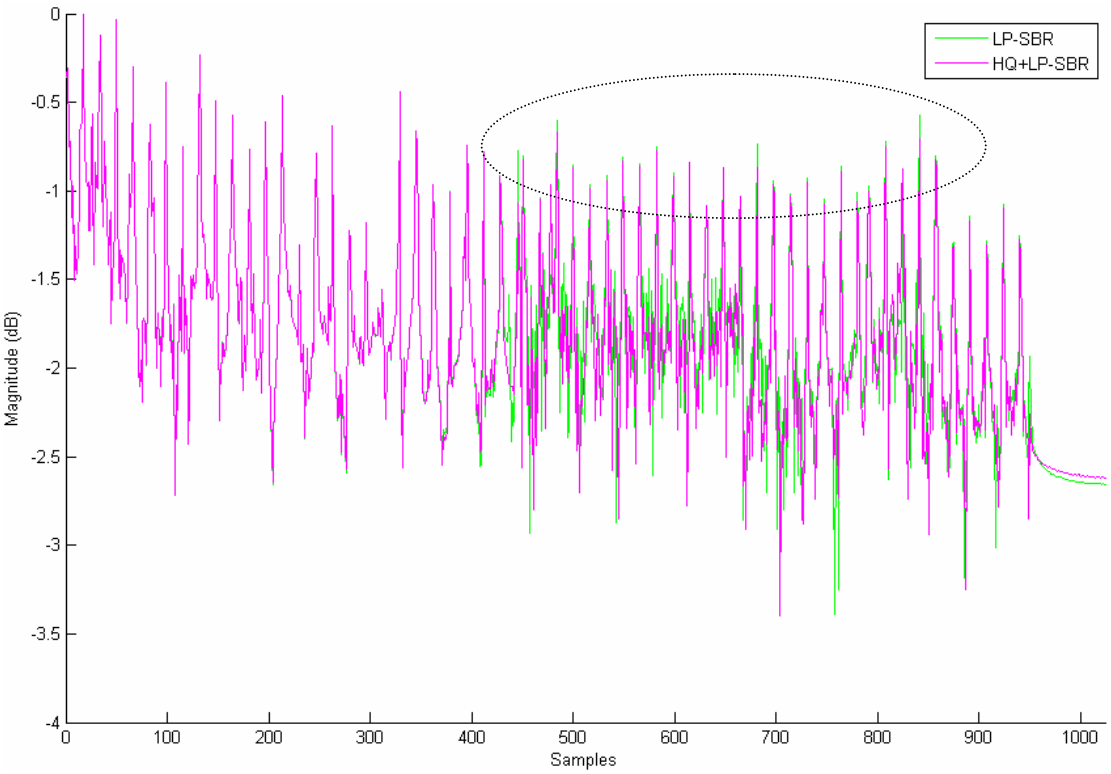

Figure 31: The spectrum of signal decoded by LP-SBR and HQ^{+}LP-SBR

5.3 Objective Quality Measurement in HQ+LP-PS

The ODG result of the tracks encoded by adopting fixed stereo parameters sets with averaging downmix approach and adaptive T/F stereo parameter extraction with KLT downmix approach and then decoded by original PS and HQ+LP-PS are compared at the same time in order to verify the decoder adopting the new approach can reach the objective sound quality as original PS.

Table 5: ODG of HQ-PS and HQ+LP-PS at 48kbps

Codec	NCTU-HEAAC			
Bit Rate	48kbps			
Tracks	M0	M1	M3	M4
es01	-1.58	-1.62	-1.36	-1.39
es02	-1.45	-1.46	-1.43	-1.45
es03	-1.63	-1.65	-1.57	-1.59
sc01	-3.35	-3.36	-3.08	-3.09
sc02	-3.11	-3.1	-2.92	-2.92
sc03	-2.52	-2.52	-2.35	-2.37
si01	-2.72	-2.72	-2.61	-2.53
si02	-2.45	-2.5	-2.18	-2.22
si03	-1.67	-1.78	-1.64	-1.62
sm01	-2.85	-2.85	-2.81	-2.81
sm02	-2.98	-2.95	-2.62	-2.64
sm03	-2.64	-2.63	-2.39	-2.38
Max	-1.45	-1.46	-1.36	-1.39
Min	-3.35	-3.36	-3.08	-3.09
Average	-2.4125	-2.42833333	-2.24667	-2.25083333
M0: Fixed stereo parameter sets with averaging downmix approach and original PS				
M1: Fixed stereo parameter sets with averaging downmix approach and HQ+LP-PS				
M3: Adaptive stereo parameters sets with KLT downmix approach and original PS				
M4: Adaptive stereo parameters sets with KLT downmix approach and HQ+LP-PS				

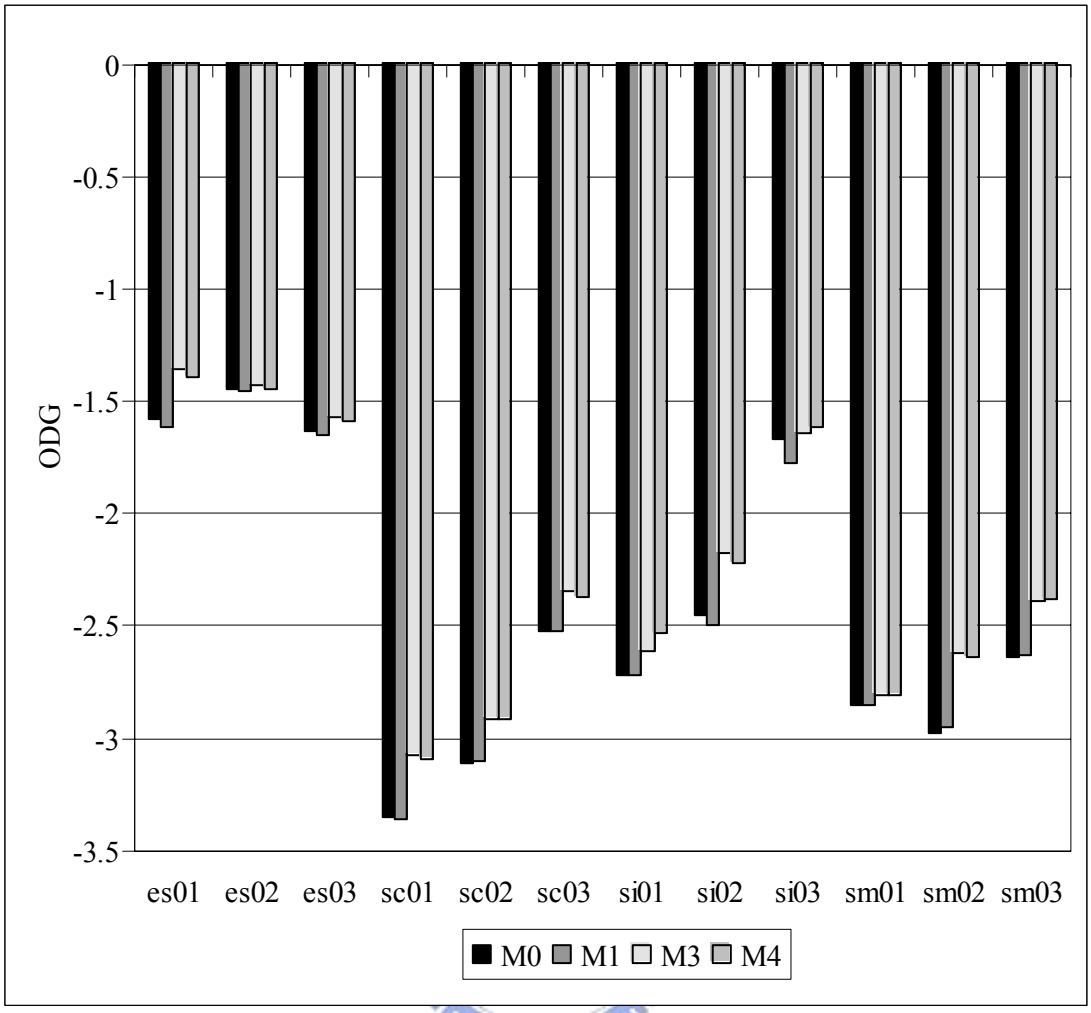

Figure 32: ODG of HQ-PS and HQ+LP-PS at 48kbps

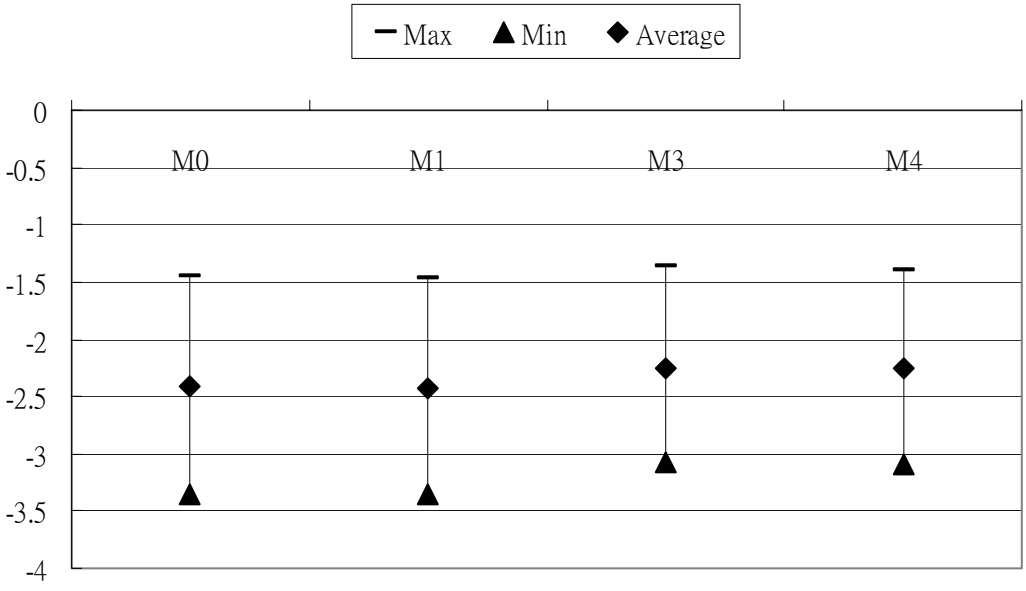

Figure 33: The variance in the ODG of HQ-PS and HQ+LP-PS at 48kbp

Table 6: ODG of HQ-PS and HQ$^+$LP-PS at 36kbps

Codec	NCTU-HEAAC			
Bit Rate	36kbps			
Tracks	M0	M1	M3	M4
es01	-2.4	-2.41	-2.2	-2.19
es02	-2.72	-2.72	-2.7	-2.72
es03	-2.81	-2.83	-2.84	-2.87
sc01	-3.4	-3.4	-3.24	-3.21
sc02	-3.25	-3.25	-3.1	-3.11
sc03	-2.79	-2.77	-2.65	-2.64
si01	-2.9	-2.87	-3.01	-2.82
si02	-2.59	-2.61	-2.44	-2.46
si03	-2.16	-2.17	-2.26	-2.18
sm01	-3.12	-3.12	-3.13	-3.12
sm02	-3.2	-3.23	-3.05	-3.04
sm03	-2.74	-2.74	-2.54	-2.55
Max	-2.4	-2.41	-2.2	-2.19
Min	-3.4	-3.4	-3.24	-3.21
Average	-2.84	-2.84333	-2.76333	-2.7425
M0: Fixed stereo parameter sets with averaging downmix approach and original PS				
M1: Fixed stereo parameter sets with averaging downmix approach and HQ$^+$LP-PS				
M3: Adaptive stereo parameters sets with KLT downmix approach and original PS				
M4: Adaptive stereo parameters sets with KLT downmix approach and HQ$^+$LP-PS				

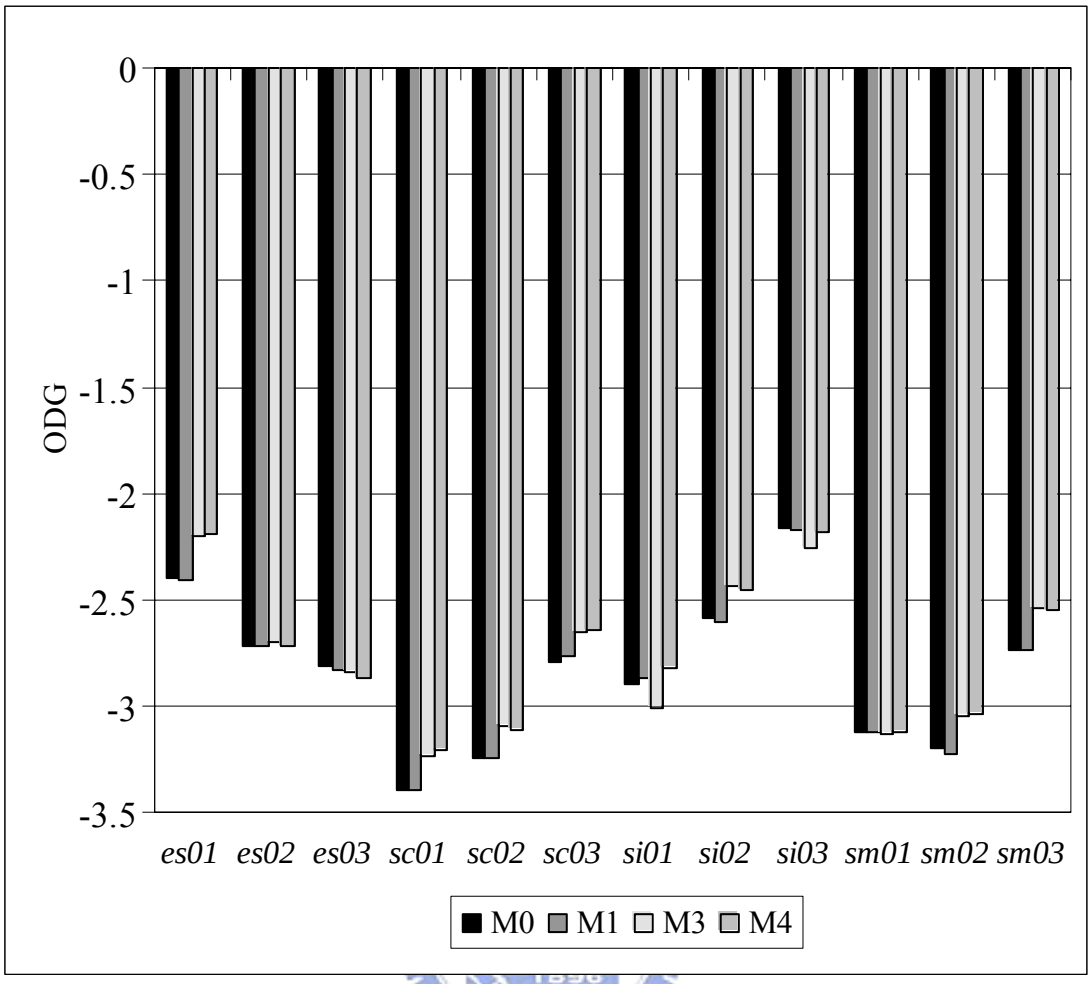

Figure 34: ODG of HQ-PS and HQ+LP-PS at 36kbps

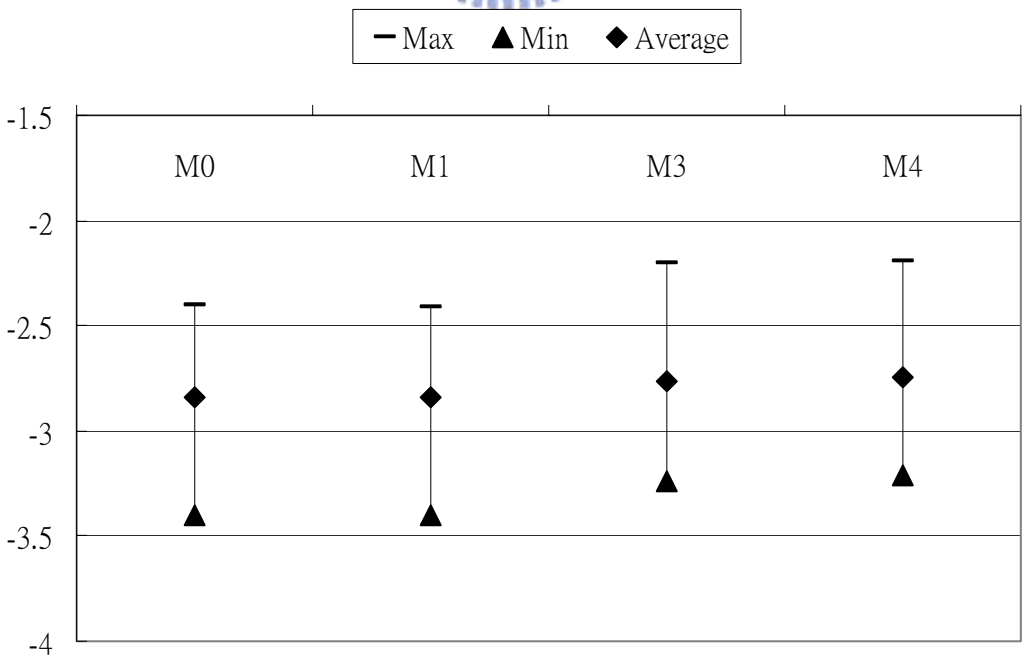

Figure 35: The variance in the ODG of HQ-PS and HQ+LP-PS at 36kbp

Table 7: ODG of HQ-PS and HQ+LP-PS at 24kbps

Codec	NCTU-HEAAC			
Bit Rate	24kbps			
Tracks	M0	M1	M3	M4
es01	-3.41	-3.4	-3.21	-3.2
es02	-3.57	-3.57	-3.45	-3.43
es03	-3.74	-3.76	-3.66	-3.67
sc01	-3.46	-3.47	-3.29	-3.3
sc02	-3.26	-3.26	-3.24	-3.24
sc03	-3.12	-3.19	-3.05	-3.06
si01	-3.34	-3.38	-3.25	-3.21
si02	-3.25	-3.28	-3.16	-3.16
si03	-3.62	-3.64	-3.28	-3.21
sm01	-3.79	-3.78	-3.66	-3.59
sm02	-3.56	-3.57	-3.6	-3.58
sm03	-3.07	-3.11	-2.88	-2.9
Max	-3.07	-3.11	-2.88	-2.9
Min	-3.79	-3.78	-3.66	-3.67
Average	-3.4325	-3.45083	-3.31083	-3.29583
M0: Fixed stereo parameter sets with averaging downmix approach and original PS				
M1: Fixed stereo parameter sets with averaging downmix approach and HQ+LP-PS				
M3: Adaptive stereo parameters sets with KLT downmix approach and original PS				
M4: Adaptive stereo parameters sets with KLT downmix approach and HQ+LP-PS				

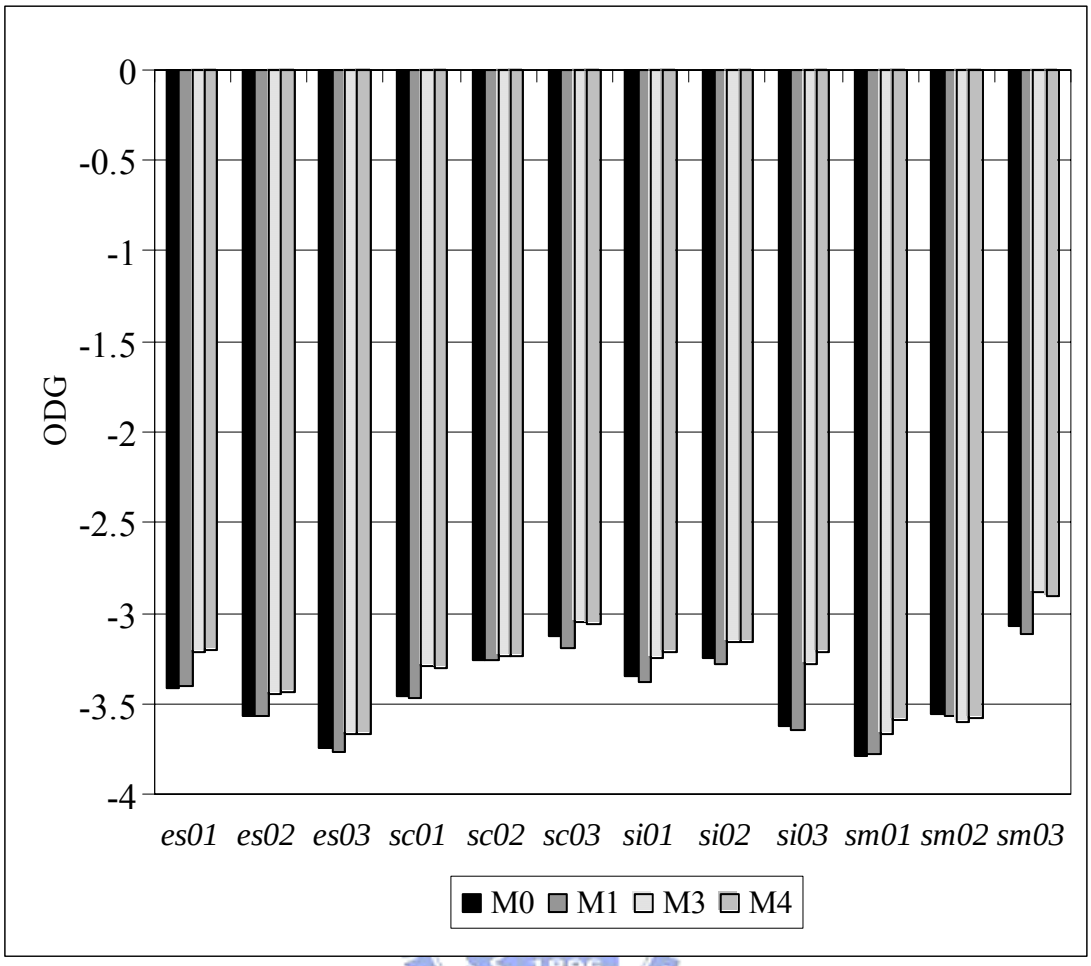

Figure 36: ODG of HQ-PS and HQ+LP-PS at 24kbps

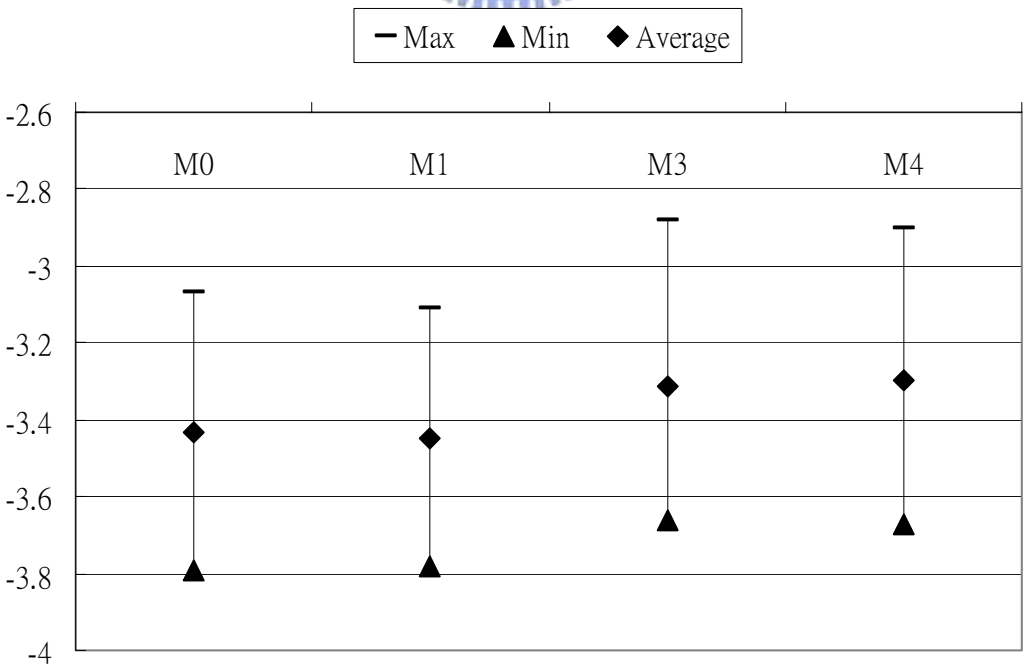

Figure 37: The variance in the ODG of HQ-PS and HQ+LP-PS at 24kbp

HQ+LP-PS is the same verified under the target bit-rate of PS coding. The presented experiment data shows that the performance of HQ+LP-PS meets our expectation which means that in most cases the objective quality of HQ+LP-PS is equal to original PS. The average ODGs of original PS and HQ+LP-PS reflect this situation. Figure 38 shows the spectrum of one frame decoded by original PS and PS using cosine QMF bank without aliasing reduction mechanism and Figure 39 shows the spectrum of the same frame decoded by original PS and HQ+LP-PS. As shown in Figure 38, because of lack of aliasing reduction mechanism, the aliasing terms exist even in low frequency part which deteriorate audio quality severely. But Figure 39 indicates that successful aliasing reduction mechanism in HQ+LP-PS results in no aliasing term in HQ+LP-PS which causes the average ODGs of HQ+LP-PS almost equal to average ODGs of original PS.

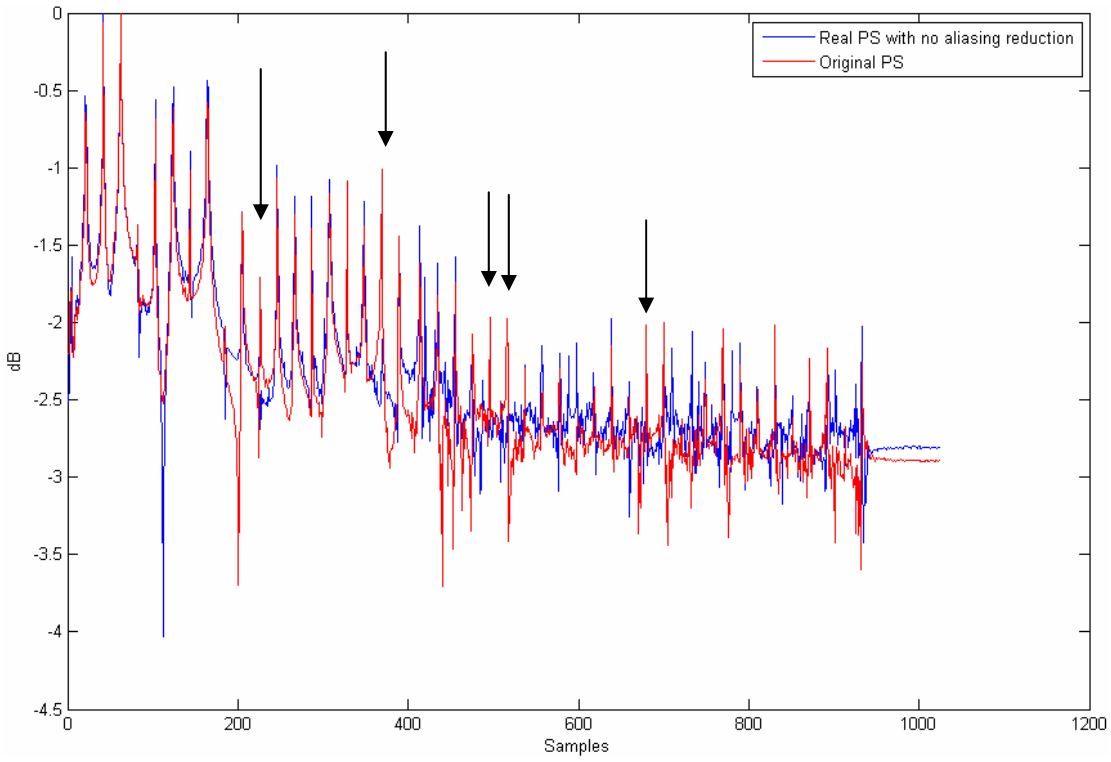

Figure 38: The spectrum of signal decoded by original PS and PS using cosine QMF bank without aliasing reduction mechanism

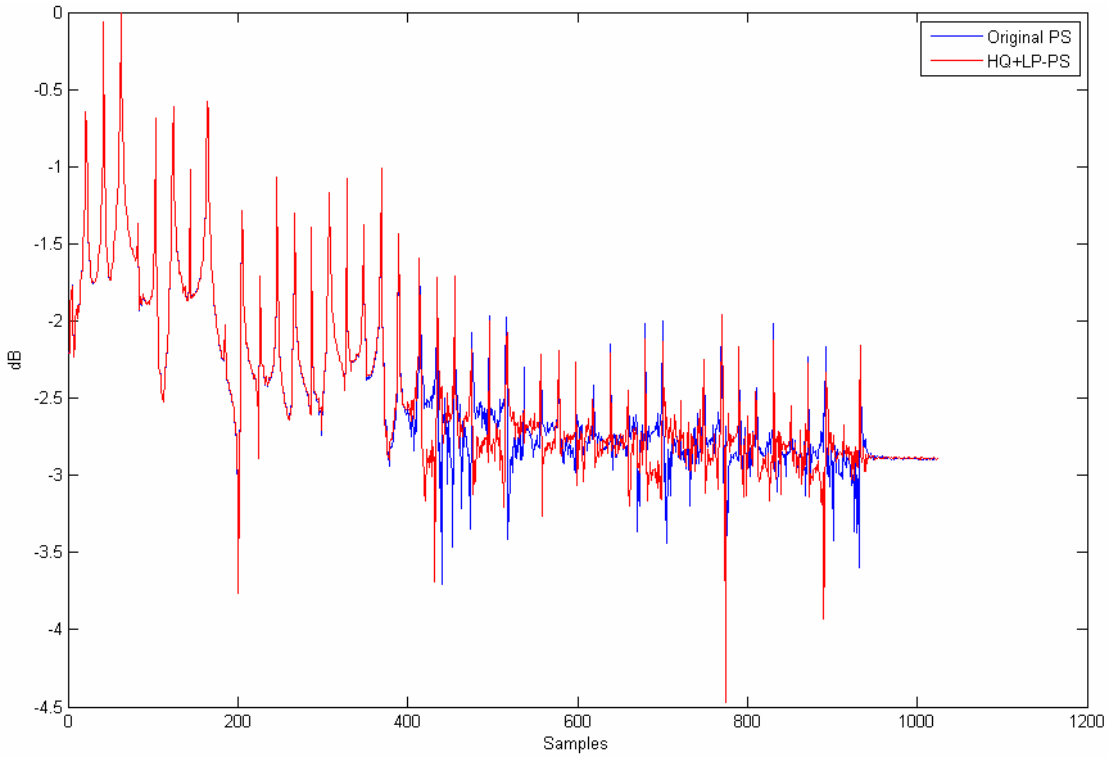

Figure 39: The spectrum of signal decoded by original PS and HQ+LP-PS

5.4 Subband Types Ratio

As described in chapter 3, the more real-valued subbands the more reduction in computational complexity of HQ$^+$LP-SBR. Therefore, if there are more real-valued subbands, the whole computational complexity will incline to LP-SBR. On the other hand, more complex-valued subbands will lead the tendency of whole complexity toward HQ-SBR. Hence, the ratio of complex-valued and real-valued subbands in each test tracks listed in Table 1 is presented below.

Table 8: Ratio of complex and real subbands in HQ$^+$LP-PS at 96kbps

Codec	NCTU-HEAAC			
Bit Rate	96kbps			
Tracks	Left Real	Left Complex	Right Real	Right Complex
es01	89.98%	10.02%	90.00%	10.00%
es02	97.32%	2.68%	97.33%	2.67%
es03	95.90%	4.10%	95.92%	4.08%
sc01	65.50%	34.50%	65.57%	34.43%
sc02	64.91%	35.09%	64.98%	35.02%
sc03	67.51%	32.49%	67.58%	32.42%
si01	47.53%	52.47%	47.68%	52.32%
si02	91.55%	8.45%	91.58%	8.42%
si03	29.80%	70.20%	29.86%	70.14%
sm01	20.89%	79.11%	21.06%	78.94%
sm02	65.91%	34.09%	65.99%	34.01%
sm03	55.27%	44.73%	55.35%	44.65%

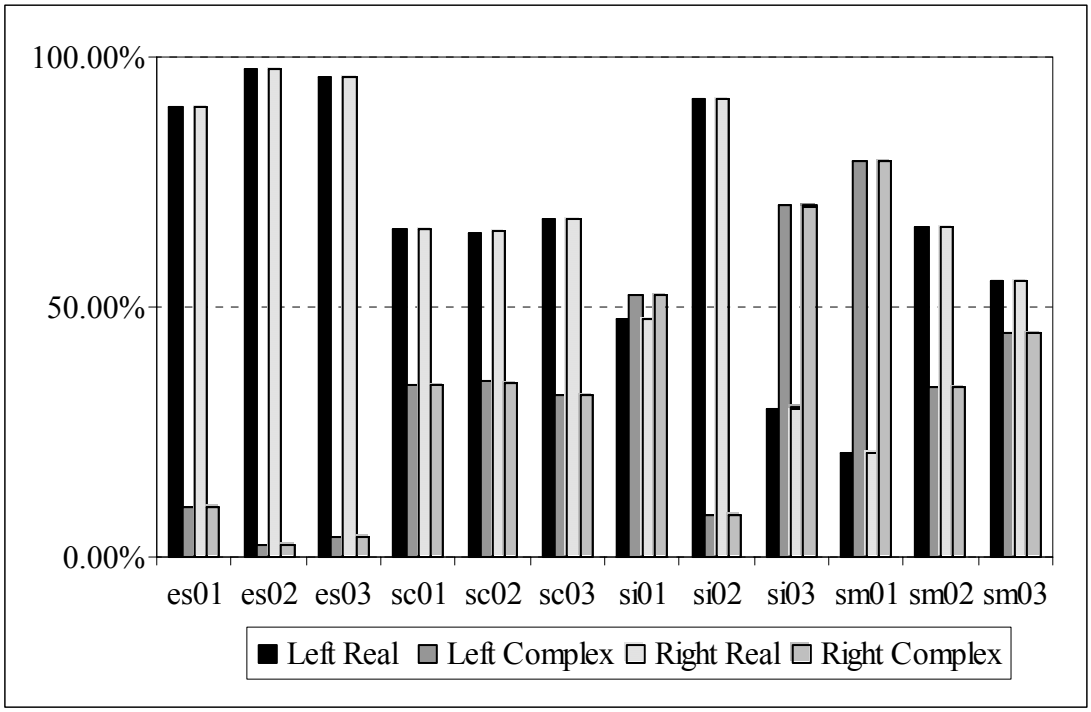

Figure 40: Ratio of complex and real subbands in HQ^{+}LP-PS at 96kbps

Table 9: Ratio of complex and real subbands in HQ^{+}LP-PS at 80kbps

Codec	NCTU-HEAAC			
Bit Rate	80kbps			
Tracks	Left Real	Left Complex	Right Real	Right Complex
es01	90.94%	9.06%	90.96%	9.04%
es02	97.42%	2.58%	97.42%	2.58%
es03	95.67%	4.33%	95.69%	4.31%
sc01	57.89%	42.11%	57.98%	42.02%
sc02	54.89%	45.11%	54.97%	45.03%
sc03	54.28%	45.72%	54.37%	45.63%
si01	42.58%	57.42%	42.74%	57.26%
si02	86.90%	13.10%	86.94%	13.06%
si03	27.02%	72.98%	27.08%	72.92%
sm01	15.90%	84.10%	16.07%	83.93%
sm02	50.28%	49.72%	50.39%	49.61%
sm03	46.64%	53.36%	46.73%	53.27%

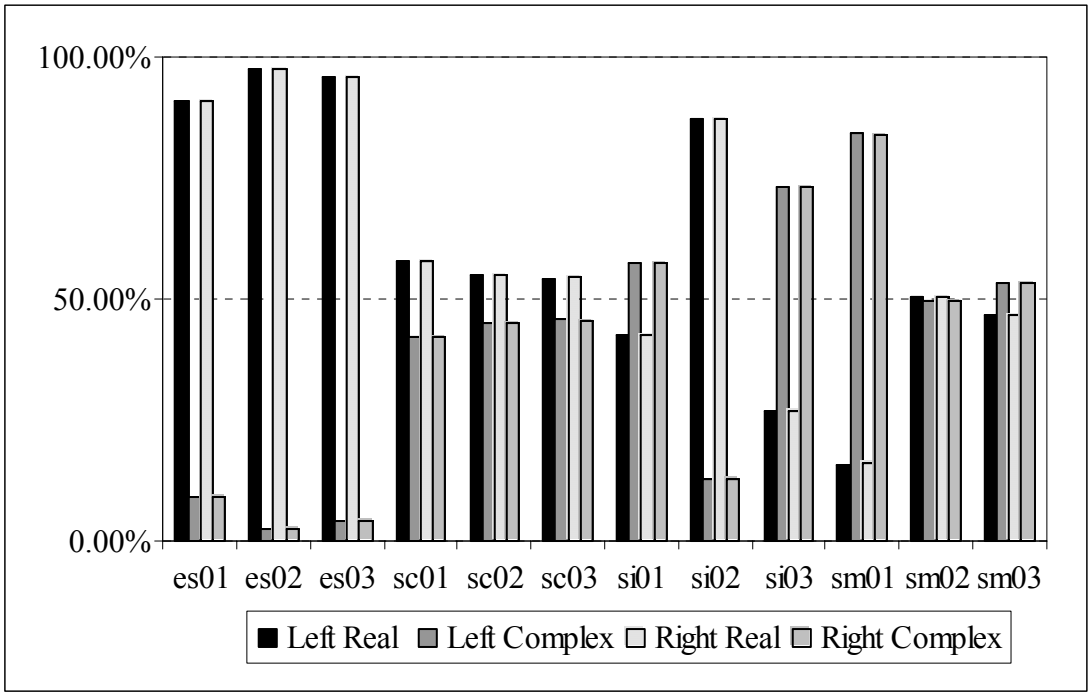

Figure 41: Ratio of complex and real subbands in HQ^{+}LP-PS at 80kbps

Table 10: Ratio of complex and real subbands in HQ^{+}LP-PS at 64kbps

Codec	NCTU-HEAAC			
Bit Rate	64kbps			
Tracks	Left Real	Left Complex	Right Real	Right Complex
es01	90.83%	9.17%	90.85%	9.15%
es02	96.86%	3.14%	96.86%	3.14%
es03	95.07%	4.93%	95.09%	4.91%
sc01	57.50%	42.50%	57.59%	42.41%
sc02	46.02%	53.98%	46.12%	53.88%
sc03	40.21%	59.79%	40.33%	59.67%
si01	41.19%	58.81%	41.36%	58.64%
si02	80.13%	19.87%	80.19%	19.81%
si03	25.32%	74.68%	25.38%	74.62%
sm01	13.25%	86.75%	13.43%	86.57%
sm02	42.04%	57.96%	42.18%	57.82%
sm03	36.65%	63.35%	36.75%	63.25%

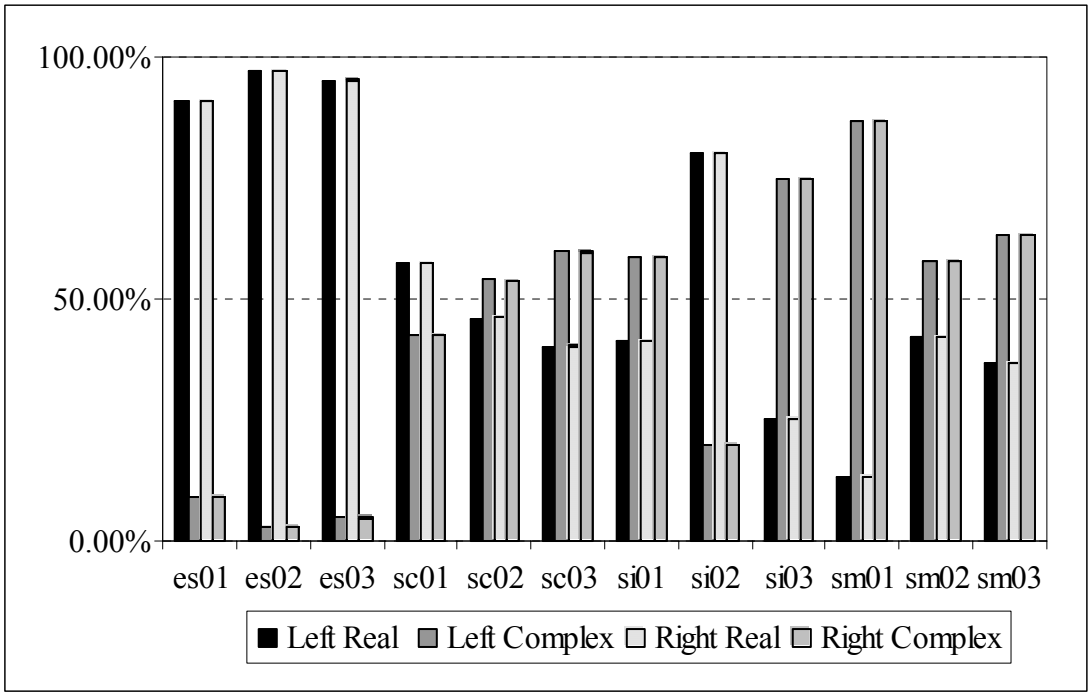

Figure 42: Ratio of complex and real subbands in HQ$^+$LP-PS at 64kbps

As shown in above figures, the ratio of real-valued subbands of es01, es02, es03 and si02 tracks under each bit-rate is higher than 80%. Therefore, the whole complexity of HQ$^+$LP-SBR while decoding such tracks will be close to LP-SBR. And the overall complexity of HQ$^+$LP-SBR while decoding sc01, sc02, sc03, sm02 and si01 will be between HQ-SBR and LP-SBR since the ratio of complex-valued and real-valued subbands is almost 50% to 50%. Then because of the intrinsic tone-rich property in si03, sm01 and sm03, the ratio of complex-valued subbands will exceed real-valued subbands. Therefore, the overall complexity of HQ$^+$LP-SBR while decoding si03, sm01 and sm03 will approach to HQ-SBR. Moreover, Figure 20, Figure 22, Figure 24 and Figure 40 ~ Figure 42 also proof that HQ$^+$LP-SBR is not only reaching quality as HQ-SBR but also reducing computational complexity through adaptively switching type of QMF bank based on the characters of subband signal.

Chapter 6
Conclusion and Future Work

A decoder with low complexity as well as high quality based on the adaptation to signal characters has been proposed. In this architecture of decoder, the high quality is reached by using the intrinsic aliasing-free property complex QMF bank has without suffering the problem of consistent scaling gain in LP-SBR and costing down the computational complexity by transforming the subband signal into real-value domain. The objective measurement by the perceptual evaluation of audio quality system has illustrated that the quality is closed to HQ-SBR and original PS.

Furthermore, the QMF bank used in MPEG Surround is equal to that in SBR and MPEG Surround can be viewed as an extension to PS plus a successful extension of HQ$^+$LP-SBR to PS. Therefore, the framework of HQ$^+$LP-SBR could fit into MPEG Surround to create a low power as well as high quality MPEG Surround decoder and Figure 43 shows its block diagram.

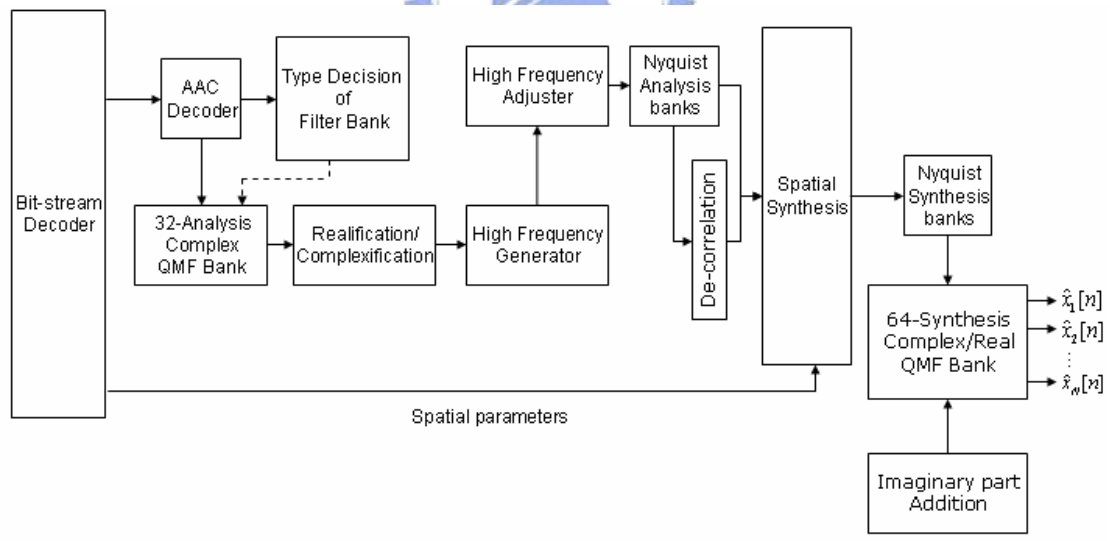

Figure 43: Block diagram of HQ$^+$LP-SBR plus MPEG Surround

References

[1] ISO/IEC 14496-3:2001/Amd.3:2004 “Information technology-- Coding of audiovisual objects - Part 3: Audio”.

[2] Draft ISO/IEC 14496-3 (Audio 3rd Edition), “Coding of Moving Pictures and Audio, Subpart 8: Technical description of parametric coding for high quality audio.”

[3] ISO/IEC 23003-1:2006/FDIS “Information technology—MPEG audio technologies - Part 1: MPEG Surround”.

[4] P.P. Vaidyanathan, “Multirate Systems and Filter Banks,” Englewood Cliffs, NJ: Prentice-Hall, 1993.

[5] O. Shimada, T. Nomura, Y. Takamizawa, M. Serizawa, N. Tanaka, M. Tsushima, T. Norimatsu, C.K. Seng, K.K. Hann and N.S. Hong, “A Low Power SBR Algorithm for the MPEG-4 Audio Standard and its DSP Implementation,” AES 116th Convention, Berlin, Germany, 2004 May 8-11.

[6] Jürgen Herre, Kristofer Kjörling, Jeroen Breebaart, Christof Faller, Sascha Disch, Heiko Purnhagen, Jeroen Koppens, Johannes Hilpert, Jonas Rödén, Werner Oomen, Karsten Linzmeier and Kok Seng Chong, “MPEG Surround – The ISO/MPEG Standard for Efficient and Compatible Multi-Channel Audio Coding,” AES 122nd Convention, Vienna, Austria, 2007 May 5-8.

[7] H.W. Hsu, C.M. Liu, and W.C. Lee, “Fast Complex Quadrature Mirror Filterbanks for MPEG-4 HE-AAC,” AES 121st Convention, San Francisco, CA, USA, 2006 October 5 – 8.

[8] S.W. Huang, T.H. Tsai, and L.G. Chen, “Fast Decomposition of Filterbanks for the State-of-the-Art Audio Coding,” IEEE Signal Processing Letter, VOL12, NO. 10, October 2005.

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

[10] NCTU-HEAAC, website: http://psplab.csie.nctu.edu.tw