國立交通大學

資訊工程學系

碩士論文

AAC 中訊窗轉換方法之設計

Design of Window Switch Method in AAC

研究生：彭康硯

指導教授：劉啟民　教授

李文傑　博士

中華民國 九十三 年 六 月

AAC 中訊窗轉換方法之設計

Design of Window Switch Method in AAC

研 究 生：彭康硯　　Student：Kang-Yan Peng

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 工程 系

碩 士 論 文

A Thesis

Submitted to Institute of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National ChiaoTung University

in partial Fulfillment of the Requirements

for the Degree of Master in

Computer Science and Information Engineering

June 2004

HsinChu, Taiwan, Republic of China

中華民國 九十三 年 六 月

AAC 中訊窗轉換方法之設計

學生：彭康硯　　指導教授：劉啓民 博士
李文傑 博士

國立交通大學資訊工程所碩士班

中文論文摘要

訊窗轉換方法是爲了得到更好的時間或頻率解析度而改變訊窗的大小。當我們對瞬變的訊號進行編碼時，量化誤差會散佈在整個訊窗當中並且無法被遮蔽。訊窗轉換方法會利用較短的訊窗來限制量化誤差所散佈的範圍。本篇論文對於 MPEG 2/4 AAC 提出一個完整的訊窗轉換設計。

如何成功的設計訊窗轉換方法?主要分爲四個設計，分別爲訊窗切換的時機，短訊窗的心理聲學模式，短訊窗的編組方法以及與其他 AAC 模組配合等議題。在論文中，我們將對這四個議題作深入的研究與探討，並且與最佳的位元分配方法結合以達到更好的編碼效率。在品質的量測上，除了主觀的聆聽評量之外，我們還採用了 ITU 所發展的 PEAQ 測試系統來評估音訊壓縮後的誤差程度。

Design of Window Switch Method in AAC

Student: Kang-Yan Peng Advisor: Dr. Chi-Min Liu

Dr. Wen-Chieh Lee

Institute of Computer Science and Information Engineering
National ChiaoTung University

ABSTRACT

Window switch method is to change the window size of the filterbank and achieve the better time/frequency resolution. When a transient signal is being coded the quantization noise is spread out over the entire window length of the filterbank and in this way is not masked by the signal. Window switch method is to use short window length to restrain the spreading of quantization noise. This thesis proposes an integrated design of window switch in ISO/IEC MPEG-2/4 Advanced Audio Coding (AAC).

However, the success of AAC window switch is approached by four design aspects: window decision, psychoacoustic model of short window, short windows grouping and combination with other AAC modules. This thesis presents the design method with well concerns on the four issues and jointly integrated with the optimal bit allocation method to have good computing efficiency. Both subjective and objective tests have been conducted to demonstrate the improved quality over the existing method. The objective system adopted is the PEAQ which is the recommendation system by ITU-R Task Group 10/4.

致謝

感謝劉啓民老師的栽培及李文傑博士給予的指導，實驗室的楊宗翰、簡鉅庭學長，同學許瀚文、蕭又華、張子文和邱挺，以及學弟陳立偉和蘇明堂的協助，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

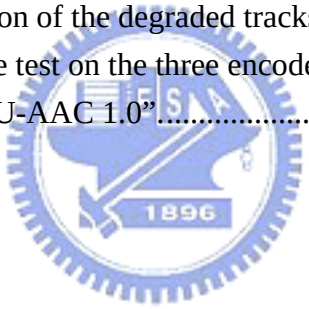

Table List

Chapter 1

Introduction

The principle of the Window Switch is to change the window size of filterbank and achieve the better time/frequency resolution. The artifacts so-called “pre-echo” occur when a transient signal is being coded. Because transient signals need high coding precision to control the signal change in time, the lack of bits makes the quantization error spread out over the entire window length of the filterbank.

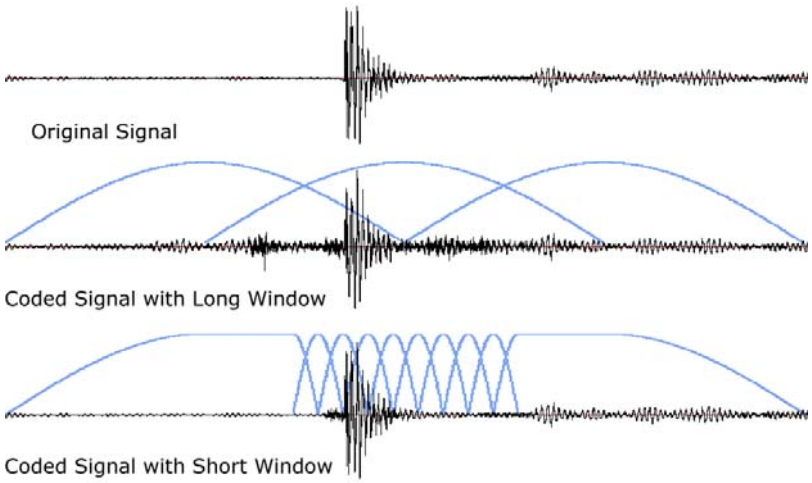

Figure 1: A transient signal coded by different window sizes.

Figure 1 is an example of a transient signal coding. The spreading of quantization error is visible in the signal which is coded by long window and it can’t be masked by the signal. Comparing with the long window coded signal, the spreading of quantization error is constrained by the shorter window length in the short window coded signal. Therefore, the objective of window switch is to control the spreading of quantization error. ISO/IEC MPEG-2/4 Advanced Audio Coding (AAC) [1][2] also includes the window switch mechanism. This thesis proposes new designs to integrate other AAC modules with the window switch.

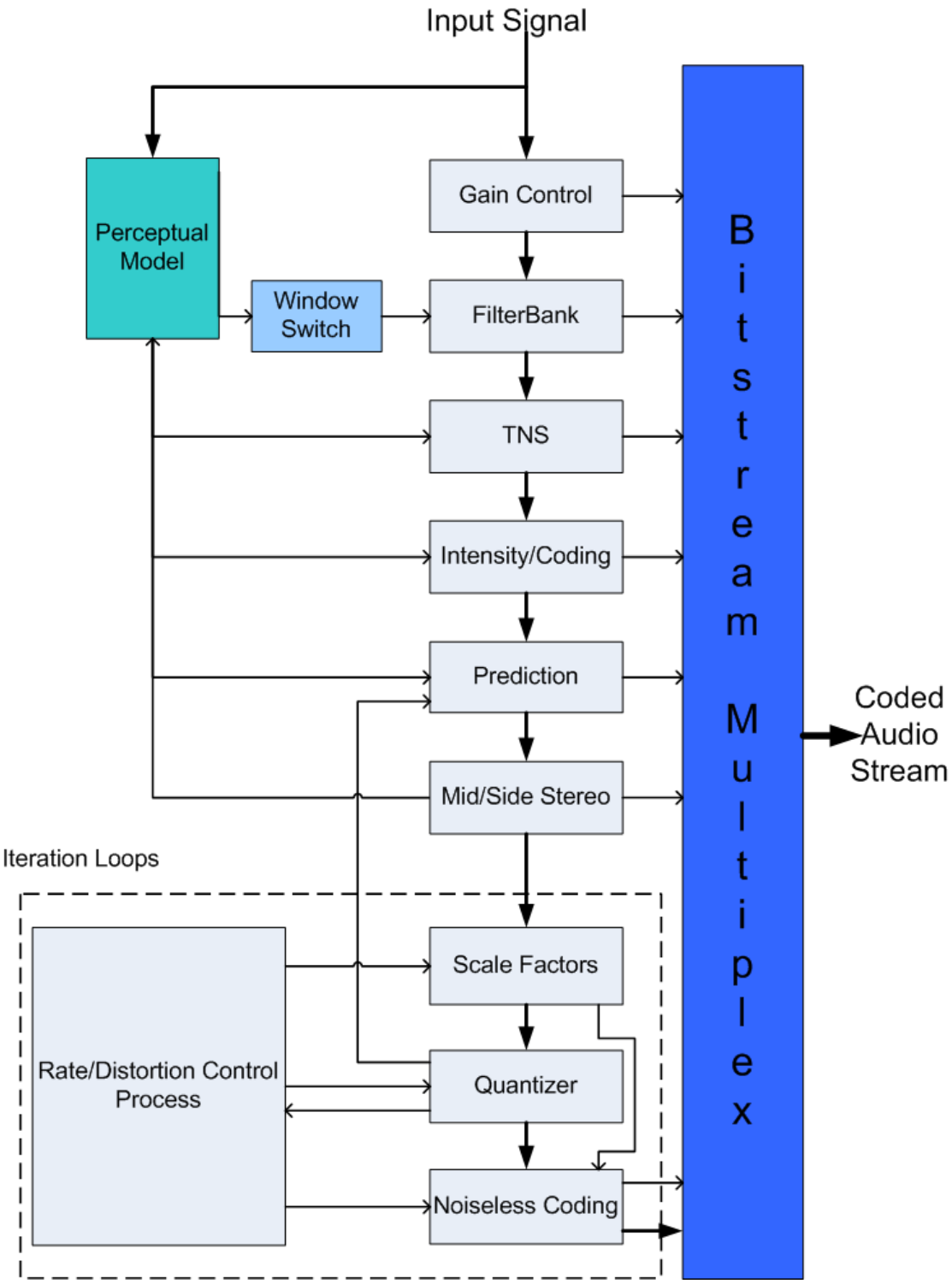

Figure 2: AAC encoder block diagram.

Figure 2 illustrates the block diagram of AAC encoder. Based on this block diagram, this thesis considers the window switch through four design issues. The first issue, window decision policy, which is also discussed widely [3]-[7] in window switch. Window decision is based on the transient signal detection and should be

finished before the filterbank. In the general perceptual encoder, different window sizes require different psychoacoustic models. Reducing the complexity of short window psychoacoustic model is the second issue. In the AAC, short windows can be grouped. And the short windows in the same group share the same scale factors. Therefore, the appropriated grouping method is the third design issue. The last issue is the combination with other AAC modules (e.g. TNS, M/S).

This thesis is organized as follows. In Chapter 2, the brief backgrounds on the fundamental knowledge of pre-echo phenomenon, the window switch in AAC and related works of window decision will be introduced. In Chapter 3, the four design issues of the window switching in AAC will be investigated, which include window decision, psychoacoustic model for short windows, grouping method of short windows and combining with other encoding modules. Then, the indication of the performance of our window switch design is presented in Chapter 4. Chapter 5 summarizes the analysis and experimental results.

Chapter 2
Backgrounds

This Chapter explains the “pre-echo” phenomenon and introduces the window switch mechanism in AAC.

2.1 Pre-echo Phenomenon

Temporal masking includes simultaneous masking, pre-masking and post-masking. The effects of these types of masking are shown in Figure 3. From Figure 3, the duration of effective masker of pre-masking and post-masking are approximately 20 ms and 100 ms respectively.

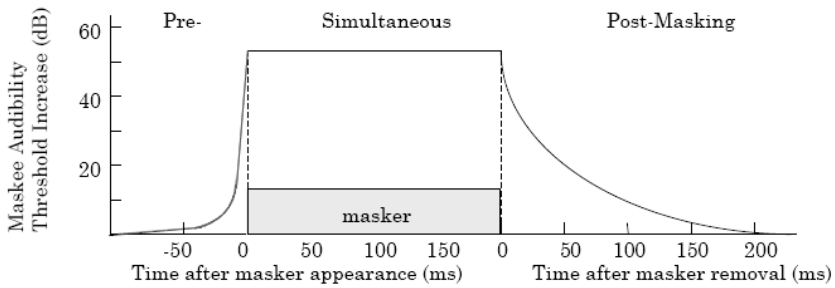

Figure 3: The effect of pre-masking, simultaneous masking and post-masking [8].

When a transient signal or an audio attack is coded in the frequency domain, the quantization error is spread throughout the entire signal block in the time domain [9]. Since the part of the signal prior to the attack is relatively small, the attack contributes most of the energy to the signal block, and thus controls the generation of the masking threshold [10]. Then the threshold is too high in the quite region of the block. Long window size is 2048 samples in AAC and represents about 46 ms when the sample rate is 44.1kHz. Because the pre-masking lasts for no more than 20 ms, the spreading of quantization error is easy to be heard when using long window to encode transient signal. It is called pre-echo phenomenon. On the other hand, short window size is 256 samples in AAC and represents about 5.8 ms when the sample rate is 44.1kHz. Using short window to encode transient signal can control the spreading of quantization error and provide more fine time resolution.

2.2 Window Switch Mechanisms in AAC

This section firstly introduces different filterbank transforms in AAC. Then, it introduces the grouping and interleaving methods which can group short windows and reduce the number of quantization bands to increase the coding gain.

2.2.1 Transform Window Switch

The adaptation of the time-frequency resolution of the filterbank to the characteristics of the input signal is done by shifting between transforms that input lengths are either 2048 or 256 samples. The 256 sample length for transient signal coding was selected as the best compromise between frequency selectivity and pre-echo suppression at a data rate 64 kbit/s per channel [11]. During the transitions between long and short transforms, a start or stop bridged window is used that preserves the time-domain aliasing cancellation properties of the MDCT and IMDCT transforms and maintains block aliment. These bridged transforms are designated “start” and “stop” sequences, respectively. The conventional long transform with the 2048-sample length is termed a “long” sequence, while the short transforms occur in groups called “short” sequence. The short sequence is composed of eight short windows transforms which are arranged to overlap 50% with each other and have the half transforms at the boundaries to overlap with the start and stop window shapes. This overlap sequence groups transform windows into start, stop, long and short sequences is show in Figure 4.

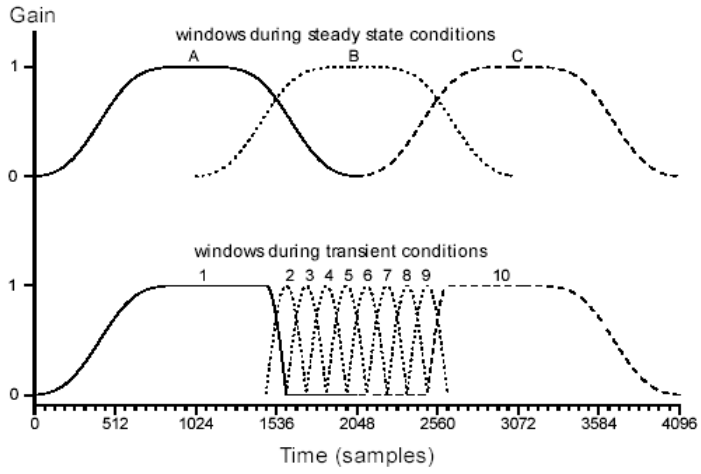

Figure 4: Comparison of window overlap for steady and transient conditions [11].

Figure 4 displays the window overlap and process appropriate for both steady-state and transient conditions. Curves A, B and C represent this process when window switching is not employed. All transforms have 2048 samples and they are composed of long sequences only. The lower part of the figure shows the use of window switching to smooth transition to and from the shorter 256-sample transforms (#2-#9). The start (#1) and stop (#10) sequences allow a smooth transition between short and long transforms.

The AAC window switching method allows the flexibility of encoding transients with eight or more 256-sample transforms, while they preserve the window alignment of the channels. For transients which are closely spaced, a single eight-short-window sequence can be extended by adding more consecutive short windows, subject to the restriction that short windows must be added in integral multiples of eight.

2.2.2 Short Window Grouping and Interleaving

If the window sequence is composed of eight short windows, then the set of 1024 coefficients is actually a matrix of 8×128 frequency coefficients, it represents the time-frequency resolution of the signal over the duration of the eight short windows. The coefficients associated with contiguous short windows can be grouped such that they share scale factors among all scale factor bands within the group. In addition, the coefficients within a group are interleaved by interchanging the order of scale factor bands and windows. To be specific, the AAC standard assumes the set of 1024 coefficients c before interleaving is indexed as follows:

$$c[g][w][b][k]$$

where

g = index on groups
w = index on windows within a group
b = index on scale factor bands within a window
k = index on coefficients within a scale factor band

and the rightmost index varies most rapidly. After interleaving, the coefficients are indexed as

$$c[g][b][w][k]$$

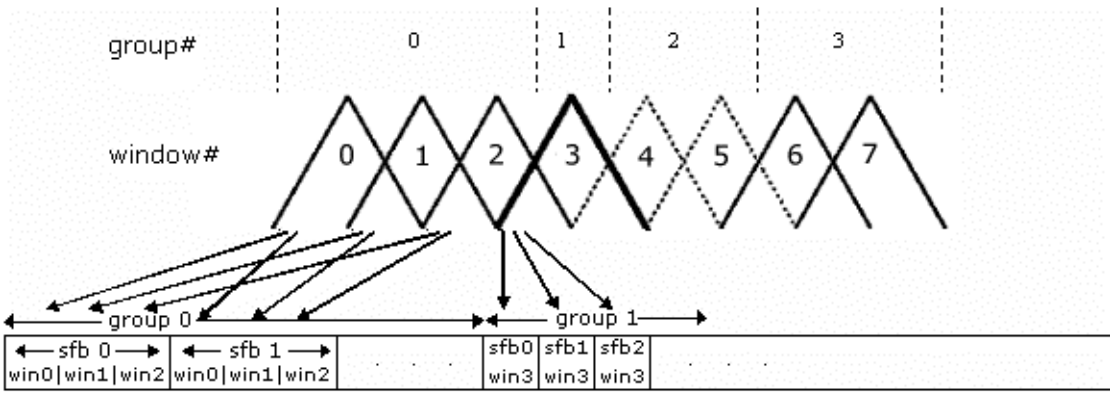

Figure 5: An Example of short windows grouping and interleaving

In Figure 5, the group 0 includes short window 0, 1 and 2. After interleaving, the first band of these three short windows form a "big" scale factor band (sfb 0). The grouping manners provide the flexibility on the number of scale factor bands for different coding considerations.

2.3 Related Works

Window decision is a critical module important issue in window switch. In this section, three methods on window decision will be discussed. First method which is used in MPEG standard sets a perceptual entropy threshold to decide when to switch window. Energy ratio method [4] was proposed to determine whether the signal is transient. Frequency energy ratio method in [7] is also a method of transient signal determination.

2.3.1 Perceptual Entropy Method

The perceptual entropy (PE) [12] is defined in MPEG standard References [1] as:

$$PE = \sum_{b} BW_b * \log\left(\frac{E_b + 1}{Masking_b}\right) \qquad (1)$$

where b is the index of the threshold calculation partition, E_b is the sum of energy in partition b, BW_b is the number of frequency lines in partition b, and $Masking_b$ is the masking in partition b.

To perform the pre-echo control, the $Masking_b$ is modified to:

$$Masking_b = \max(qthr_b, \min(nb_b, nb\_l_b * repelev)) \quad (2)$$

where $qthr_b$ is the threshold in quiet, nb_b and nb_l_b are thresholds of partition for the this and last block, and *repelev* is defined as a constant 2.

When the signal bursts into higher energy, the thresholds from nb_l_b to nb_b become higher as a result of the increase of the signal energy. Then the $Masking_b$ will be small and the value of *PE* will be larger. When the frame *PE* is higher than the predefined threshold, *PE_SWITCH*, the encoder will change the window type into short window in order to increase the time resolution and decrease the pre-echo effect.

But the frame PE is not directly related with signal change. The sudden increase of signal energy is not the only factor that brings on the increase of *PE*. The distribution of the signal energy and the tonality characteristic of the signal are also important for influencing the *PE*. If the signal has a lot frequency lines that have been detected as pure tone (e.g. harmonic), it will increase *PE* which is even higher than the window switch threshold. In this case, the lower frequency resolution of short window will lead to audible frequency distortion [6].

2.3.2 Energy Method

Traditionally, transient are regarded as signal segments, which have time-domain energy function rapidly change from low to high value [4]. Transient detection algorithm bases on an energy-based criterion which uses the signal energy within two sliding short windows. The energy function (*ef*) is defined as,

$$ef(n) = \frac{1}{L} \sum_{k=n-\frac{L}{2}}^{n+\frac{L}{2}} x^2(k) \quad (3)$$

where *n* is the center point of length-*L* short window and x(*k*) is the input signal. The transient character of the signal is determined by the energy ratio as,

$$C(n) = \log(\frac{ef(n)}{ef(n-L)}), n = 0 \text{ to N -1} \quad (4)$$

where N is the analysis of window length, and $L < N$.

This method is easy to implement. But the signal which has segments with rapid changes in spectral content can’t be detected by this method.

2.3.3 Frequency Energy Method

In order to detect signal change both in time domain and frequency domain, [7] has proposed a transient detection algorithm which is operated in the spectral domain to capture both time domain and frequency domain transients. Each band energy *en(b)* is calculated. Consequently, band energy function of window *n* ($f_n(b)$) is defined as,

$$f_n(b) = \alpha \cdot en(b) + (1-\alpha) \cdot f_{n-1}(b) \tag{5}$$

For each band *b*, a transient measurement *G(b)* is define as,

$$G(b) = \frac{en(b)}{f_n(b)} \tag{6}$$

When the transient measurement *G(b)* is greater than a threshold *T*, the transient band flag d(b) will be set as 1. The total number of transient band *F* is calculated as,

$$F = \sum_{b=0}^{N-1} d(b), \text{ where } d(b) = \begin{cases} 1 & G(b) > T \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

This method detecting the change of each band can detect the transient in both time domain and frequency domain. However, the complexity of transforming signals from time domain into frequency domain for each short window is higher than the other two methods. Besides, the characteristics of each frequency are different. Accordingly, the transient measurement thresholds for each band are hard to decide.

Chapter 3
Design of Window Switch Method in AAC

In Chapter 2, we have introduced the pre-echo phenomenon and window switch mechanisms in AAC. Besides, the related works of window decision are discussed. In this Chapter, an integrated design of window switch in AAC will be proposed. The window decision is the most important and also the first issue in this Chapter. Then the method of omitting the psychoacoustic model of short window to reduce the complexity will be proposed. The third issue is the design of grouping method which can lead an appropriated bit allocation to reach a well quality. The last issue is the design of combination with TNS and M/S coding module in AAC.

3.1 Window Decision

The design of window decision is the most important part of window switch. Because the short window has a higher time resolution, and the long window has a higher frequency resolution. The transient signal needs short windows to control the pre-echo effect and the stationary signal needs long window to resolve the lines in the signal spectrum in order to extract the redundancy. If the transient signal uses long window, the pre-echo phenomenon will happen. If the stationary signal uses short window, the low frequency resolution will make the encoded signal not precise enough in the frequency domain.

This section proposes a design of window decision by three kinds of information: the global energy ratio, the zero-crossing ratio and the tonal attack. Window decision decides the window type of next frame. After deciding next window type, the current window type will be switched by comparing with next and prior window type. Therefore, in the last subsection, the window type switch method will also be discussed.

3.1.1 Global Energy Ratio

Transient signals usually occur when the time domain energy has rapid change. Therefore, the energy ratio is a kind of important information to detect transient signal. Traditionally, the energy ratio detection method [4] only considers the energy ratio between two sliding short windows. Generally, the pre-echo effect is generated by the signal with global max energy. But, the energy ratio between two sliding windows will ignore the gradually increasing signal. Figure 6 is an example of speech signal. Figure 6 (a) represents a transient signal which is increasing gradually. Figure 6 (b) is the value of traditional energy ratio. The max energy ratio in Figure 6 (b) is about 2.1. However, if the transient threshold is set at 2, the misjudgment will happen easily. Figure 6 (c) illustrates the variation of global max ratio. The global energy ratio method can provide a noticeable value of ratio and overcome the problem in traditional energy ratio method.

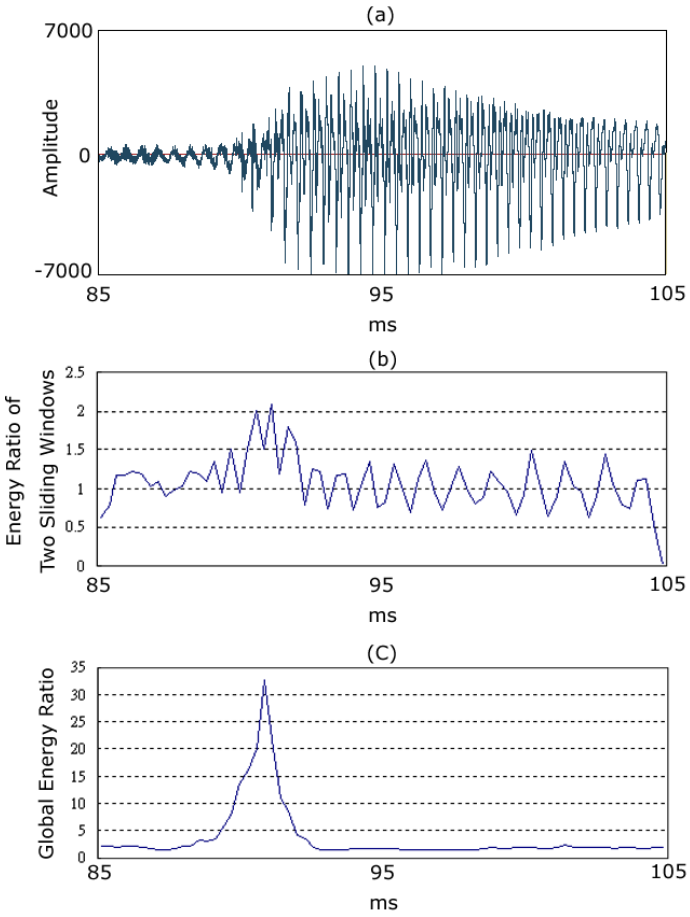

Figure 6: (a) Transient signal segment, (b) energy ratio of two sliding short windows, (c) values of global energy ratio.

In common with the traditional energy ratio method, we calculate a energy function as,

$$En(i) = \frac{1}{256} \sum_{k \in w_i} x_k^2 \tag{8}$$

Then the maximum energy *Max_En* and minimum energy *Min_En* in a set of short windows' energy *En(i)* are found. The global energy ratio is defined as,

$$Global\_En\_Ratio = \frac{Max\_En}{Min\_En} \tag{9}$$

When the *Global_En_Ratio* is greater than a threshold T_e, the signal is regard as a transient signal. The implement of this method is as easy as the traditional energy ratio method. However, this method is more general and it also can prevent the post-echo problem.

3.1.2 Zero-Crossing Ratio

As traditional energy ratio method, the global energy ratio can't detect the signal which has segments with rapid changes in spectral content. However, zero-crossing rates can represent the main frequency content of signal. Figure 7 shows a transient signal with stable global energy ratio, but this signal has rapid change in spectral content. Zero-crossing ratio can detect this kind of transient signal.

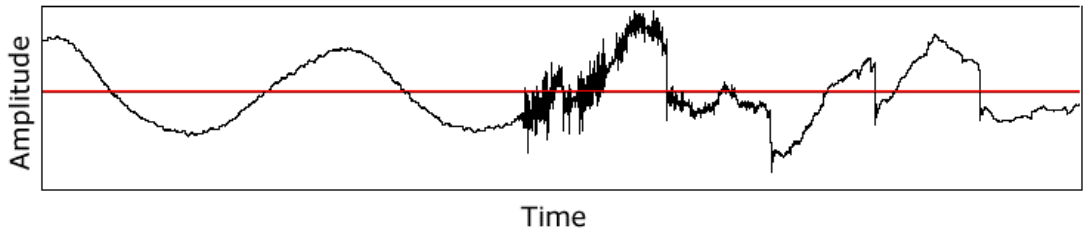

Figure 7: A transient signal with rapid changes in spectral content.

The zero-crossing rate of each window is defined as,

$$Ze(i) = \frac{\text{Number of Zero Crossing in } w_i}{256} \tag{10}$$

Then the maximum zero-crossing rate *Max_Ze* and minimum zero-crossing rate *Min_Ze* in a set of short windows' zero-crossing rate are found. The zero-crossing ratio is defined as,

$$Ze\_Ratio = \frac{Max\_Ze}{Min\_Ze} \tag{11}$$

When the *Ze_Ratio* is greater than a threshold T_z, the signal is regard as a transient signal. This method has lower complexity than the method introduced in subsection 2.3.3. This method can detect the transient in violin and speech signal.

3.1.3 Tonal Attack

The short window has lower frequency resolution than that of the long window. Figure 8 (a) is an example of pure tone signal, and this signal will be regard as a transient signal by the global energy ratio. In Figure 8 (c), transforming the tonal signal by a shorter transform will make the side band energy increase. We define the tonal attack effect when the signal has a tonal band which is analyzed by the psychoacoustic model of long window. In other words, if there is a band with tonality greater than a threshold *T*, the encoder doesn't use short windows in this frame to keep the frequency resolution.

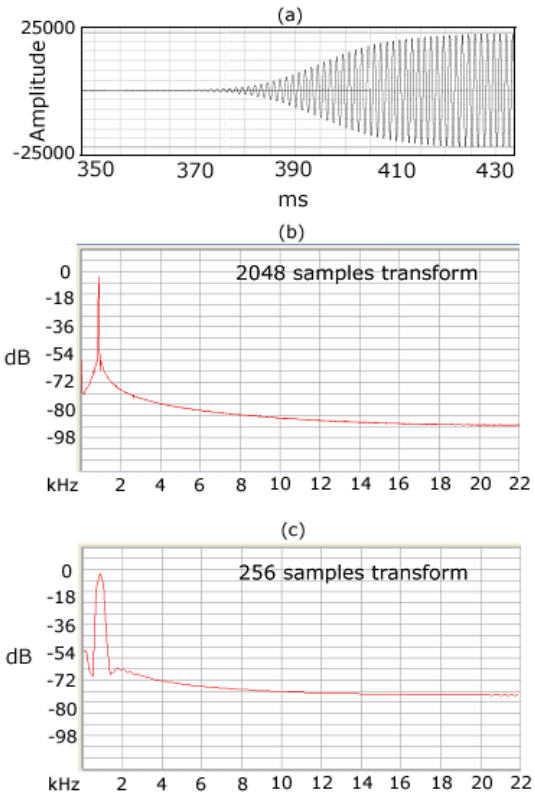

Figure 8: (a) A pure tone signal, (b) the frequency transformed by 2048-sample transform, (c) the frequency transformed by 256-sample transform.

Window decision method is composed of above three kinds of information. Figure 9 illustrates the window decision execution which uses global energy ratio and zero-crossing ratio to detect transient signal and then avoid the erroneous detection of tonal attack.

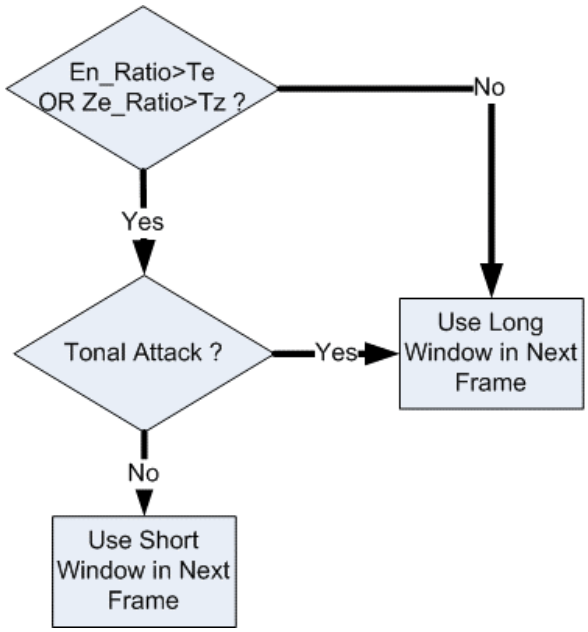

Figure 9: Window Decision Flowchart.

3.1.4 Window Type Switch Method

The start window should be used to bridge long and short window types. But window decision only decides the usage of long or short window type. Therefore, window decision should decide the window type of next frame in advance. The switch of current window type considers both prior and next window types. If the next frame is different from the prior one, the current frame must switch to the start or stop window type.

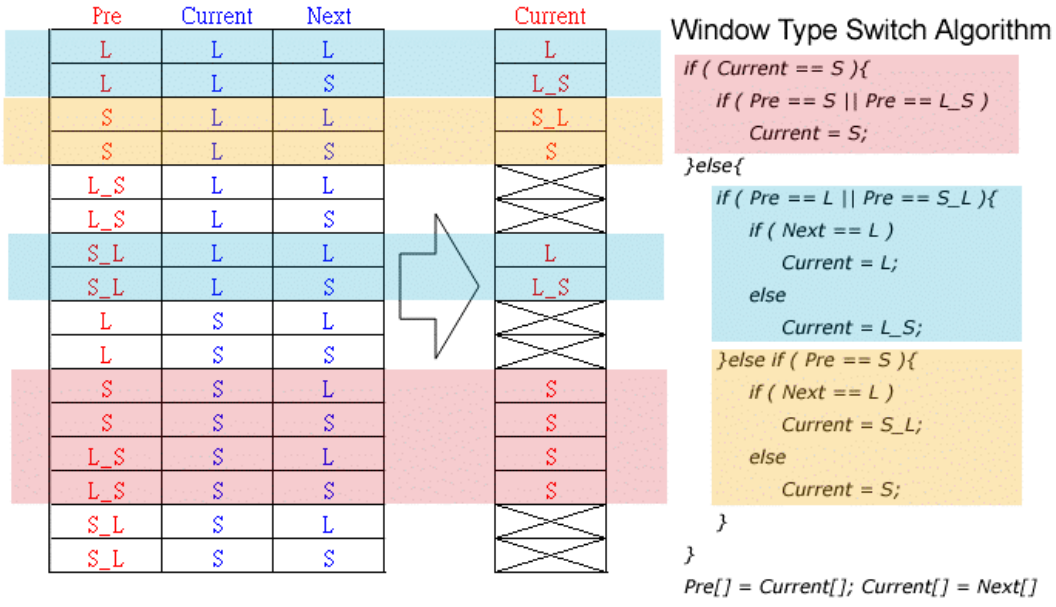

Figure 10: Window type switch analysis table and algorithm.

Figure 10 shows the analysis of all possible situations of the window type switch. Long window, short windows, start window and stop window are represented by L, S, L_S, and S_L respectively. By removing some impossible situations, we can get a simple switching algorithm.

3.2 Psychoacoustic Model of Short Windows

By [13], one major disadvantage of the adaptive window switching technique is that it introduces additional complexity into coder. Since different window sizes require different interpretations and normalizations of the psychoacoustic model. In AAC, if the window sequence is composed of eight short windows, the coder needs to execute short window psychoacoustic model for eight times. Long window psychoacoustic model has higher frequency resolution and more precise masking threshold information than that of short window psychoacoustic model. Thus, this section presents a method to omit the execution of short window psychoacoustic model and replaces it by long window psychoacoustic model.

The psychoacoustic model calculates the minimum masking threshold which is necessary to determine the just-noticeable noise-level for each band in the filterbank [14]. The signal-to-masking ratio (SMR) is used in the bit allocation to determine the actual quantizer level in each subband of the block. Because the short window psychoacoustic is omitted, the band SMR of short window should be estimated from the mapping bands of long window.

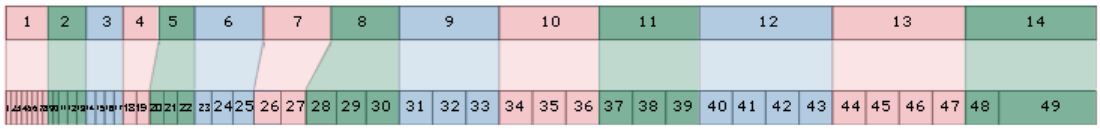

Figure 11: Band SMR mapping example from long window to short window in AAC when the sample rate is 44.1kHz.

Figure 11 is an example which shows the mapping result of the 49 bands in the long window corresponding to the 14 bands in the short window. If the frame uses short window type, it needs to take SMRs from long window. Therefore, each band of the short window takes the maximum SMR of mapped bands of long window as its SMR. Selecting the maximum SMR is in order to reduce the risk of quantization error.

3.3 Window Grouping of Short Windows

In subsection 2.2.2, grouping and interleaving method have been introduced. The short window can handle the transient signal well by controlling the spreading of quantization noise to be within the short windows. However, when the AAC coder decides to use short windows, the total number of scale factor bands is twice larger than when it uses only one long window. By the grouping method, the short windows in the same group share the same scale factors. Therefore, this method reduces the total number of scale factor bands. Every time when a group is added, it would increase a set of scale factors. The more scale factor bands exists, the more bits that side information needs. Therefore, it would make inadequate bits exist and then problems of quantization errors are produced. Every time when one group is decreased, short windows which sharing the same scale factors increase. And the sum of errors between shared scale factors and estimated scale factors of each short window also increases. For the reasons given above, this section proposes a design of grouping method to improve the quality and the computing complexity.

It is intuitive to design the grouping method by using the estimated scale factors of eight short windows. Therefore, if the scale factors can be estimated earlier in the encoder the grouping method can be applied more flexible with other codec module (e.g. like M/S coding). In the following subsections, this thesis attempts to introduce an efficient estimation method of scale factors and then present the grouping method by these estimated scale factors.

3.3.1 Scale Factor Estimation

[15] and [16] present a noise estimation method. The expectation of quantization error of the non-uniform quantizer e_i is

$$E[e_i^2] = \frac{4}{27}\Delta_q{}^2 \cdot E[|X_i|^{\frac{1}{2}}] \tag{12}$$

where Δ_q is the quantization step size which is defined as

$$\Delta_q = 2^{\frac{3}{8}(g-c_q)} \tag{13}$$

where g is global gain independent of the scale factor band q. c_q is scaling factor in each scale factor band.

Scale factor estimation of bit allocation is based on the bandwidth proportional noise-shaping criterion. From [17], the noise level for the scale factor bands should be proportional to the effective bandwidth $B(q)$.

$$\sigma_{N(q)}^2 = \kappa \cdot \sigma_{M(q)}^2 \cdot B(q) \tag{14}$$

where $\sigma_{N(q)}^2$ and $\sigma_{M(q)}^2$ are the noise energy and the masking energy associated with the scale factor band q.

With (12) relating the scale factor with the noise power, it is straightforward to combine (12) and (14). Let $E[e_i^2] = \sigma_{N(q)}^2$ and define $T_q^2 = \sigma_{M(q)}^2 \cdot B(q)$. The expectation of the quantization error for bit allocation is

$$E[e_i^2] = \kappa \cdot T_q^2 \approx \frac{4}{27} \Delta_q{}^2 E[|X_i|^{\frac{1}{2}}] \tag{15}$$

The square of quantization step size $\Delta_q{}^2$ is

$$\Delta_q{}^2 = 2^{\frac{3}{8}(g-c_q)} = \frac{27}{4} \kappa \cdot T_q^2 / E[|X_q|^{0.5}] \tag{16}$$

The difference between global gain and scale factor can be evaluated by

$$g - c_q = \frac{8}{3} \cdot \left(\log_2 \left(\frac{27}{4} \kappa \cdot T_q \right) - \log_2 E[|X_q|^{0.5}] \right) \tag{17}$$

From (17), the global gain can be evaluated from

$$g = \underset{q}{Max} \{ g - c_q \} \tag{18}$$

and the scale factors for all sub-bands are obtained.

3.3.2 Grouping Method

Since that the short windows in the same group share scale factors among all scale factor bands within the group, the differences between the shared scale factors ($sharesfb_{g,b}$) and estimated scale factors ($sf_{b,w}$) of short windows in the same group should be bounded. In addition to the difference of scale factors, the influence of this difference is proportional to the bandwidth ($bandwidth_b$). So, the scale factor error of group g can be estimated by the following equation.

$$E_g = \sum_b \sum_{w \in g} \left| sf_{b,w} - sharedsf_{g,b} \right| \times bandwidth_b \tag{19}$$

The criterion of grouping method minimizes the grouping number, and the scale factor error E_g of each group should be smaller than a threshold M. By the criterion, we design an algorithm and display it in the Figure 12. Firstly, scale factors estimation will be executed. After that, the grouping method starts with the first short window. Because short windows in one group should be continuous, every short window in the beginning tries to join the group that the last short window belongs to. If the scale factor error of the new group is smaller than threshold M, joining the short window will be successful; or creating a new group for the short window.

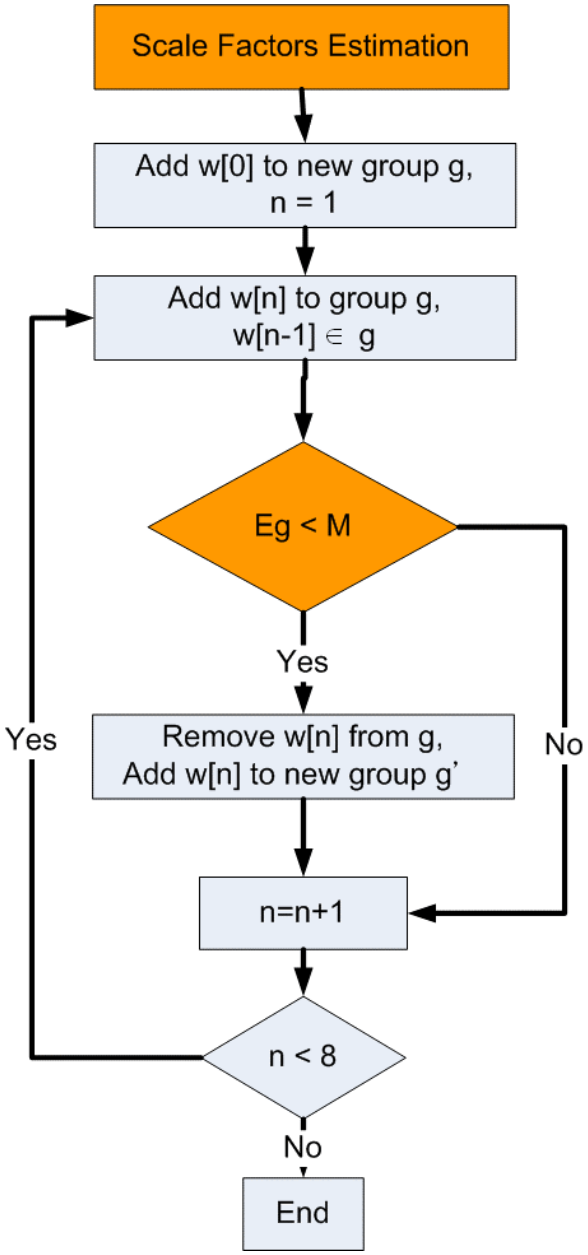

Figure 12: The flowchart of grouping method of short windows.

3.4 Joint Design with Other AAC Modules

After the filterbank, the design of each encoding module should consider two modes, long window mode and short window mode. In AAC, if short window type is used, the scale factor bands will be determined after grouping method. Therefore, long window and short windows can use the same bit allocation policy. Between the filterbank and bit allocation in the Figure 2, there are four modules: TNS, intensity coding, prediction and M/S coding. But, in the NCTU_AAC [18], there are only TNS and M/S coding modules. This section explains the relationship between window switch and these two coding modules.

3.4.1 TNS and Window Switch

TNS is also a technique to prevent the pre-echo phenomena. Consequently, TNS also has problems to decide whether the signal is transient. In [19], TNS is applied according to the perceptual entropy and the location of attacks. In addition, section 3.1 describes that window decision is decided by three kinds of information. Therefore, the decision methods of TNS and window switch are not the same; TNS and window switch can work together and achieve better quality.

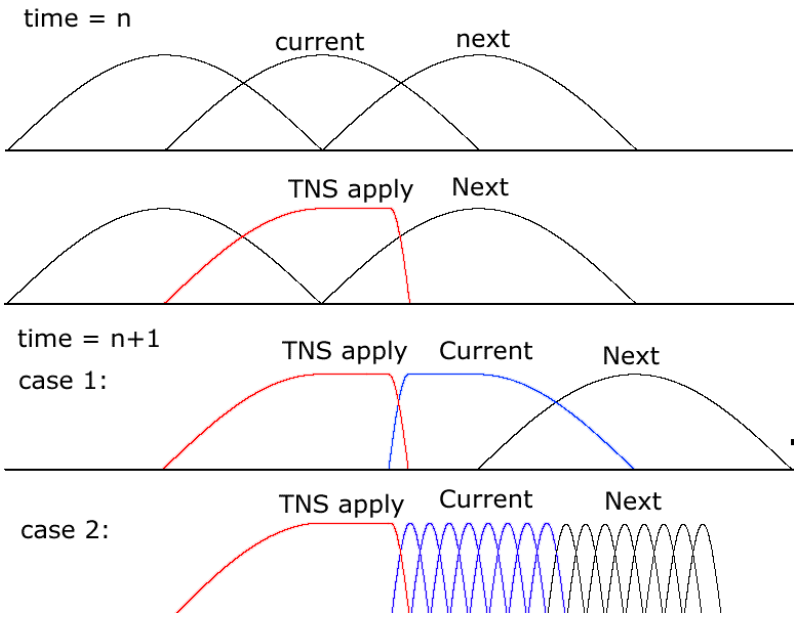

Figure 13: Window type switch when TNS is applied and attempts to ease aliasing.

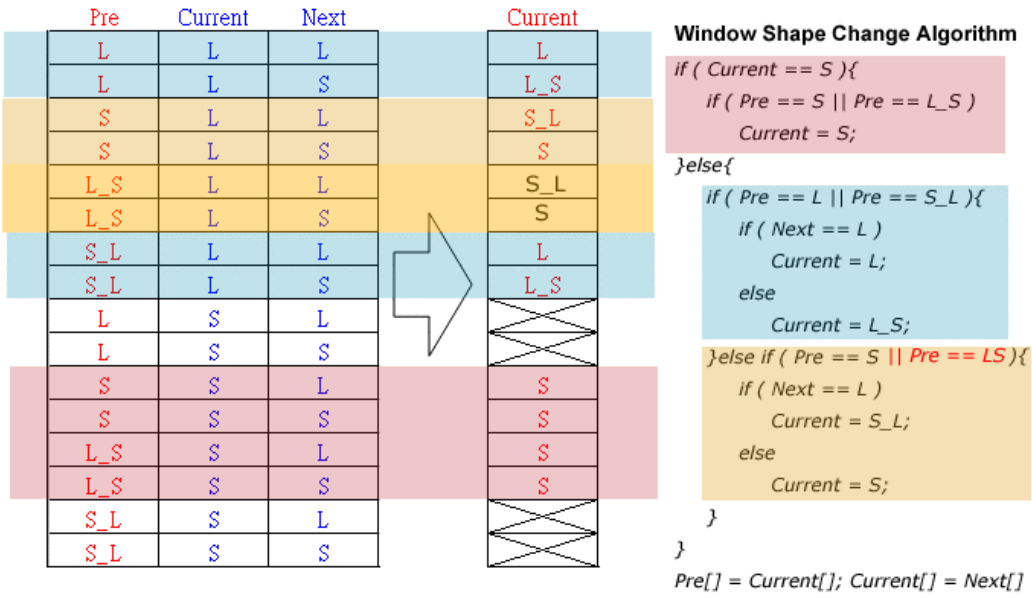

Figure 14: The modified window type switch algorithm.

Chang's method [19] has proposed a window switch among start, stop, and long windows to ease aliasing. Figure 13 illustrates that if the current window type is long window, it will be switched to the start window when TNS is applied. In the next time (n+1), the new situation (when prior window type is start, current window type is long, and next window type is also long) should be considered. Figure 14 is the modification of the window type switch in subsection 3.1.4. Accordingly, Figure 14 considers the new situation caused by TNS.

3.4.2 M/S Coding and Window Switch

In the stereo coding of AAC, M/S mechanism is applicable when both window type and the same grouping manner in the two stereo channels are the same. This subsection proposes the window coupling and group coupling method to have good coding efficiency under the constraint.

Window Coupling

When one channel is short windows type and another is long window type, we check the similarity of these two channels first. If they are similar, we have to decide using long or short window type simultaneously. The perceptual entropy (PE) can assist us to judge the similarity and window decision. Figure 15 illustrates the flowchart of window coupling. It shows that the difference of PEs, *T1*, is used to judge the similarity. Then, we set another PE threshold *T2* to decide the window type.

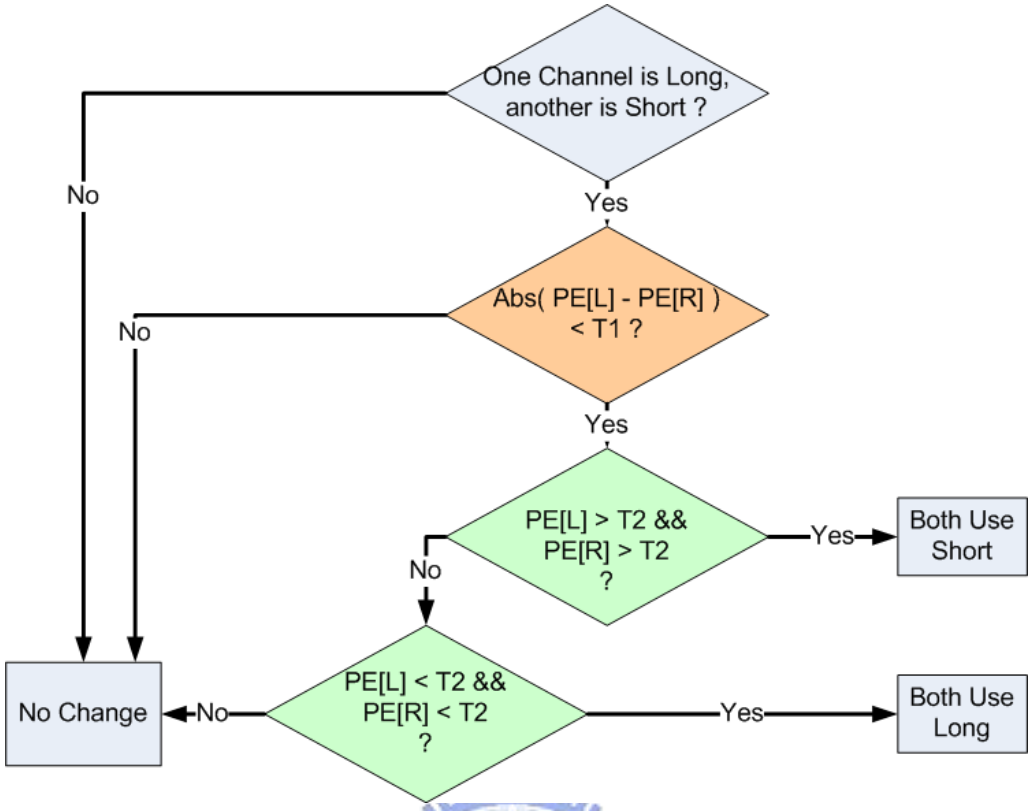

Figure 15: Flowchart of window coupling method.

Group Coupling

As the grouping method discussed in section 3.3, we calculate the sum of scale factor error in both channel and group two channels simultaneously. In the left portion of Figure 16, the grouping method is used individually in two channels. The purpose of group coupling method is to keep the same the grouping manner in both channels as illustrated in Figure 16.

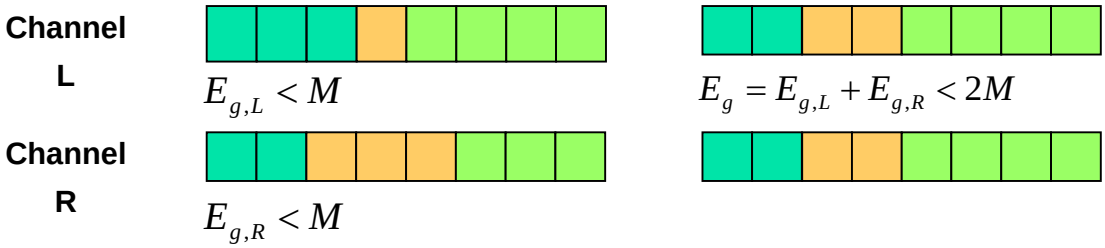

Figure 16: Example of grouping individually and simultaneously.

The criterion of grouping method minimizes the grouping number, and the total scale factor error E_g of each group in both channels to be smaller than a new threshold

2M.

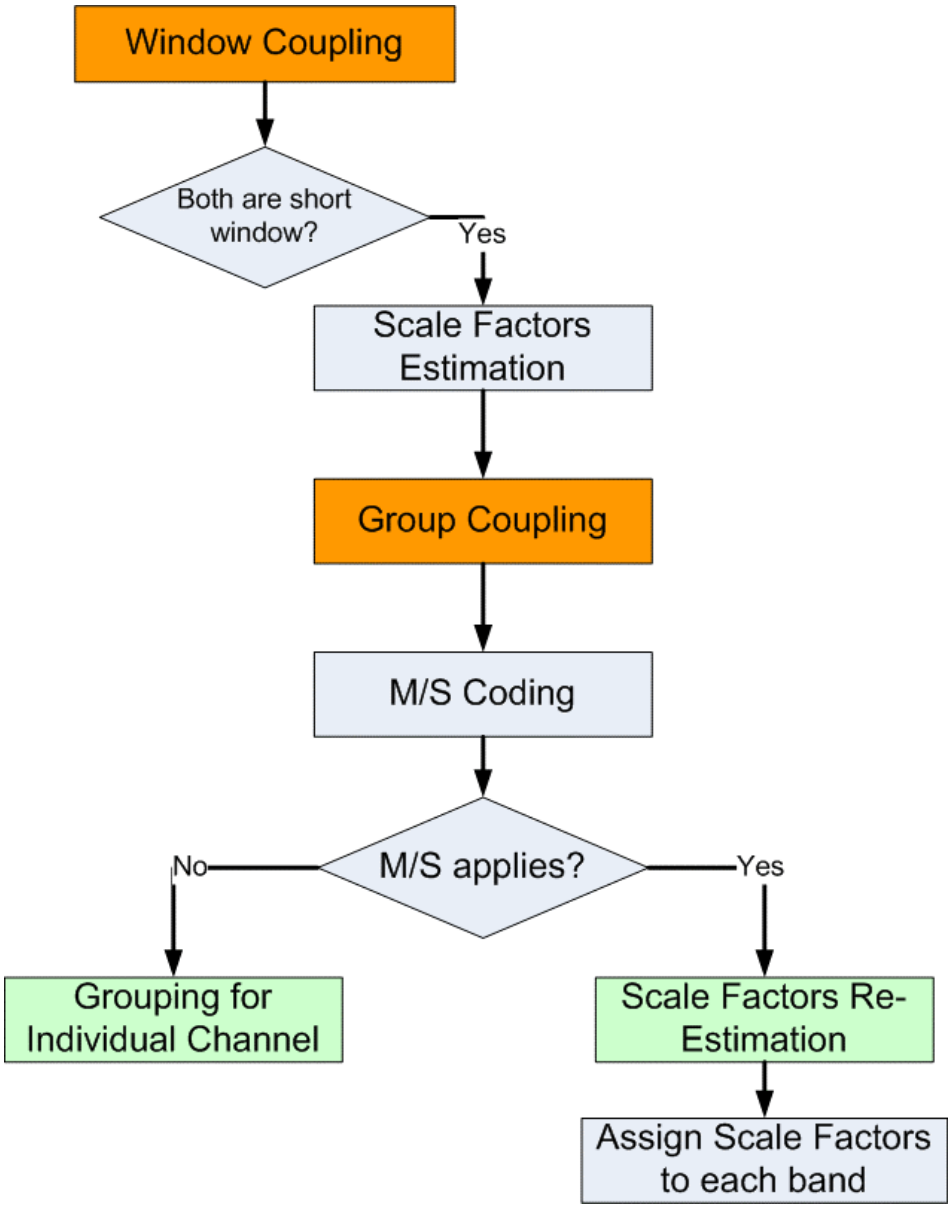

Figure 17: Flowchart of two coupling methods.

Figure 17 explains the relationship with the M/S coding. When the M/S is switched on, the energies of two channels will be modified and the scale factor associated with each scale factor band will be re-estimated. When the M/S doesn't apply, the grouping can be applied individually to two stereo channels. The new NCTU_AAC flowchart is illustrated in Figure 18 and Figure 19.

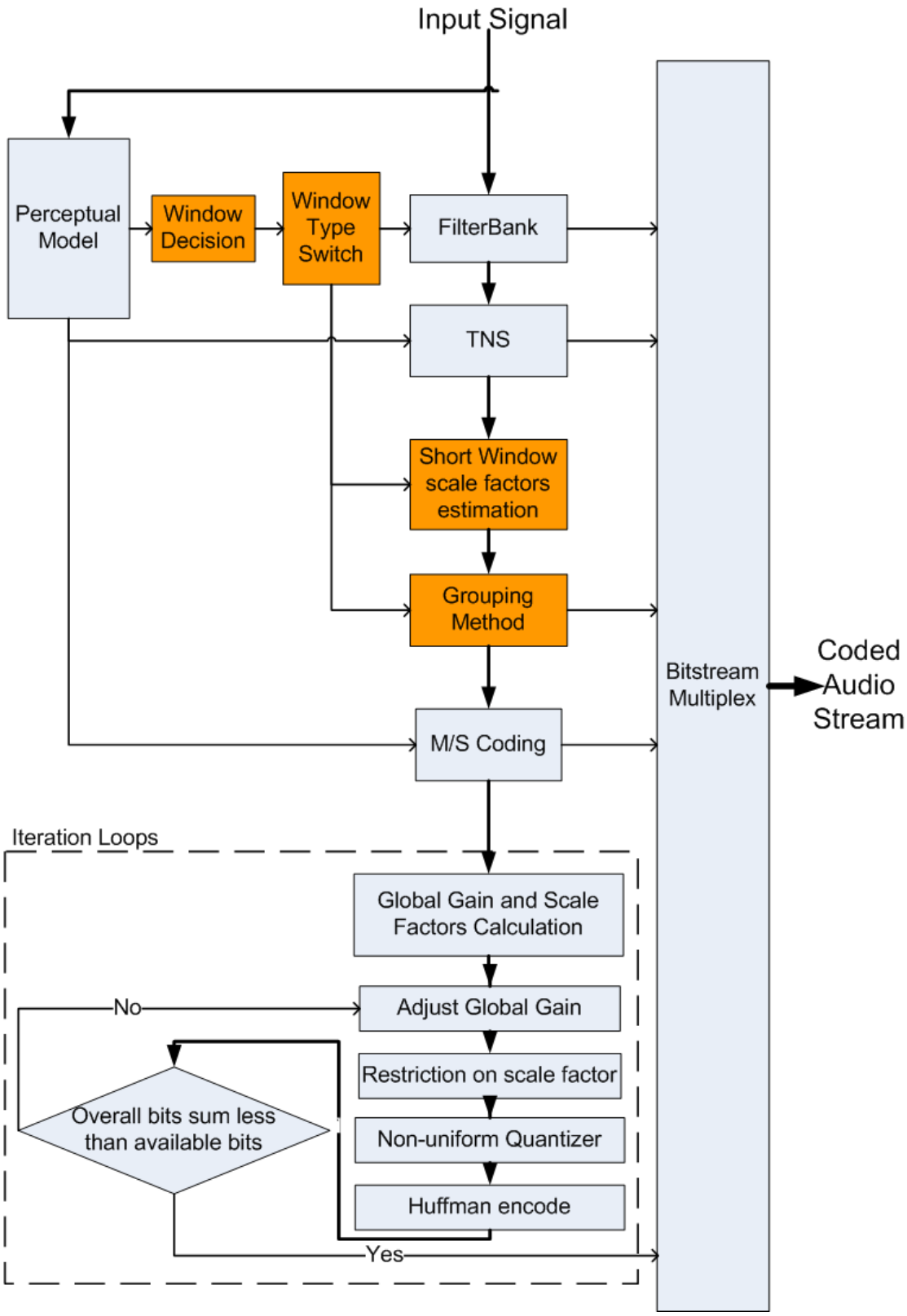

Figure 18: NCTU_AAC block diagram without two coupling methods.

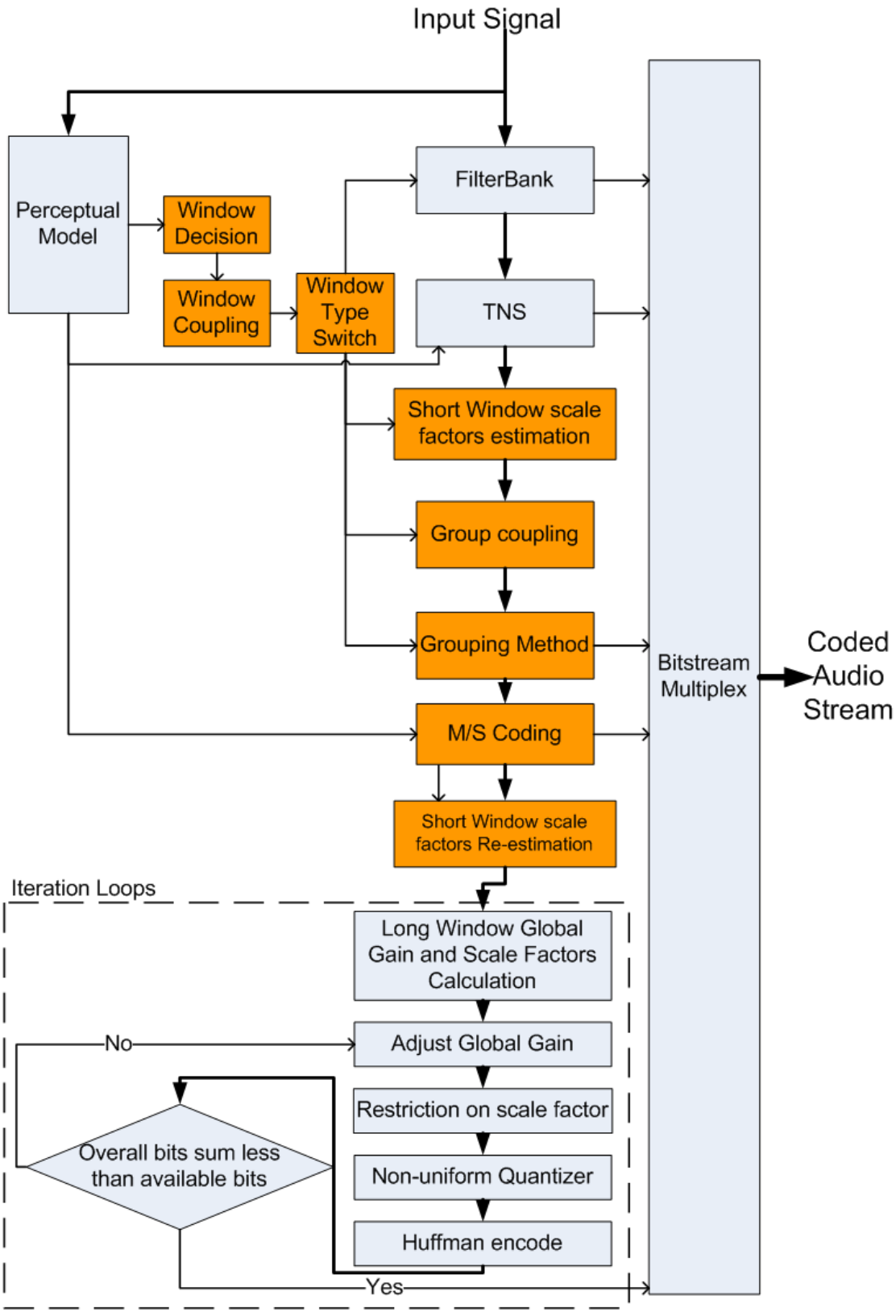

Figure 19: NCTU_AAC block diagram with two coupling methods.

Chapter 4
Experiments

This Chapter focuses on the quality measurement in NCTU_AAC platform. There are five primary experimental aspects. The first one is for the window decision method. The second one is for grouping method. The third one is for two coupling methods. The fourth one is the report of 327 tracks test. The last one is for the comparison of objective quality with other AAC encoders.

Table 1: MPEG testing track set.

Track		Signal Description			
		Signal	Mode	Time(sec)	Remark
1	es01	vocal (Suzan Vega)	Stereo	10	(c)
2	es02	German speech	Stereo	8	(c)
3	es03	English speech	Stereo	7	(c)
4	sc01	Trumpet solo and orchestra	Stereo	10	(d)
5	sc02	Orchestral piece	Stereo	12	(d)
6	sc03	Contemporary pop music	Stereo	11	(d)
7	si01	Harpsichord	Stereo	7	
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)
12	sm03	Plucked strings	Stereo	13	

Remark:
(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.
(e) High bandwidth: stresses the Device Under Test, loss of high frequencies, program-modulated high frequency noise.
(f) Low volume testing.

All experiments in this paper are based on different psychoacoustic models, and the new M/S coding module is proposed by [20] and [21] respectively.

4.1 Experiments of Window Decision

As mentioned in the section 3.1, a new window decision is proposed. The new decision method consists of three kinds of information, energy ratio, zero-crossing ratio and tonal attack. This section explains that the energy threshold and the zero-crossing threshold will be firstly calibrated. Figure 20 is the energy threshold calibration based on zero-crossing threshold 5.0. After obtaining the energy threshold, Figure 21 shows the calibration of zero-crossing threshold. Then the energy threshold needs to calibrate again (Figure 22). Finally, the energy threshold is 6.0 and the zero-crossing threshold is 4.5.

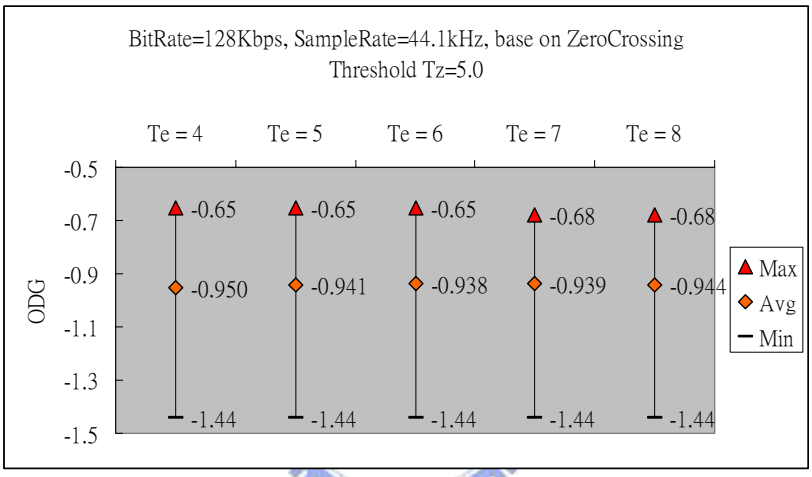

Figure 20: ODG for different Energy Threshold based on the Zero-Crossing Threshold Tz=5.0. The horizontal line is the average ODG among all the tested tracks in Table 1. The best ODG and the worst ODG in the tested tracks are marked with the triangle and "—" around the horizontal line.

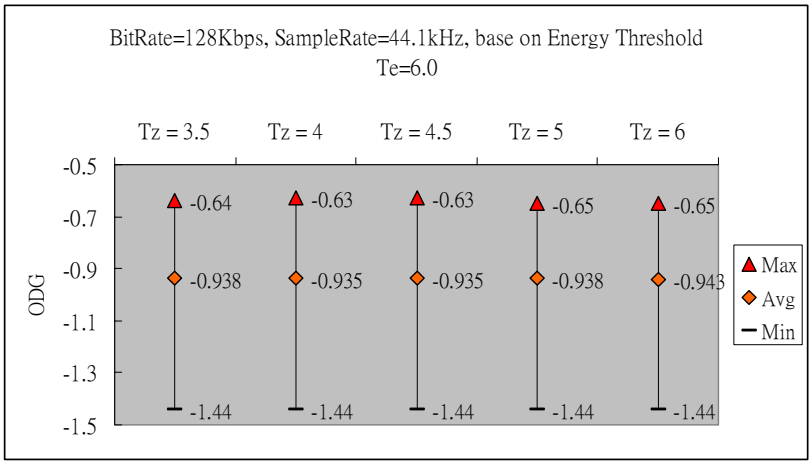

Figure 21: ODG for different Zero-Crossing Threshold based on the Energy Threshold Te=6.0.

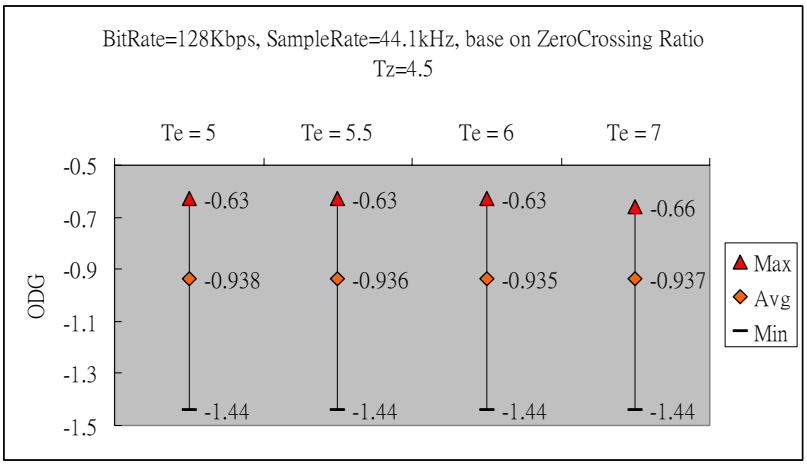

Figure 22: ODG for different Energy Threshold based on the Zero-Crossing Threshold Tz=4.5.

After calibrating these thresholds, the new window decision method is finished. Figure 23 and Table 2 are the comparison results of different window decision methods, showing that the new decision method is better than the other two methods, only long window method and *PE* decision method. The speech voice songs (e.g. es01, es02, and es03) and the attack songs (e.g. si02) have an outstanding improvement.

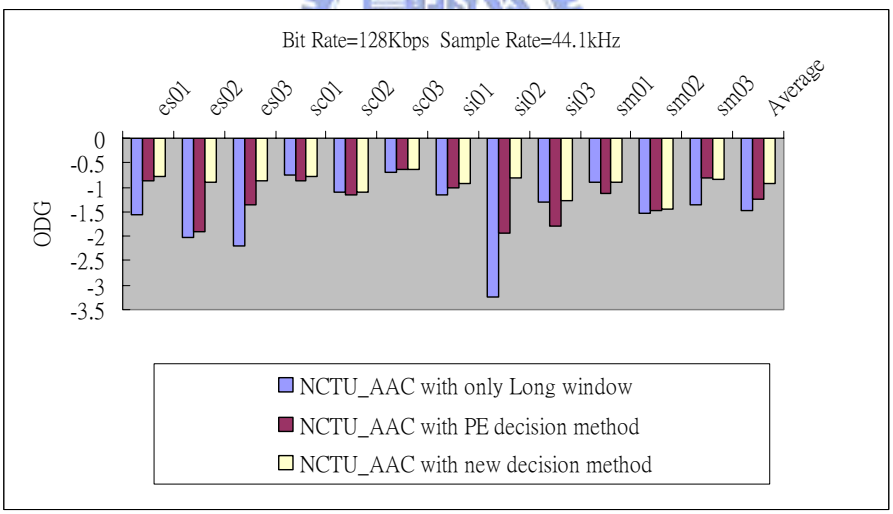

Figure 23: Objective test using P4 on the three decision methods: “NCTU-AAC with only Long Window”, “NCTU-AAC with *PE* decision method” and “NCTU-AAC with new decision method”

Table 2: Detail ODG values of objective test on the three decision methods.

Psychoacoustic Model	P1	P1	P1	P4	P4	P4
use Short	No	Yes	Yes	No	Yes	Yes
M/S	L/R	L/R	L/R	L/R	L/R	L/R
Coupling	No	No	No	No	No	No
es01	-1.78	-1.02	-0.9	-1.57	-0.87	-0.77
es02	-2.16	-2.03	-0.93	-2.03	-1.9	-0.89
es03	-2.52	-1.6	-0.88	-2.21	-1.37	-0.86
sc01	-0.81	-0.84	-0.88	-0.75	-0.87	-0.77
sc02	-1.06	-1.06	-1.06	-1.11	-1.15	-1.11
sc03	-0.86	-0.77	-0.78	-0.7	-0.64	-0.63
si01	-1.36	-1.05	-1.04	-1.16	-1.02	-0.94
si02	-3.45	-2.1	-0.91	-3.24	-1.93	-0.81
si03	-2.42	-2.43	-2.42	-1.29	-1.79	-1.27
sm01	-1.49	-1.49	-1.48	-0.9	-1.14	-0.9
sm02	-1.79	-1.97	-1.7	-1.54	-1.47	-1.44
sm03	-1.37	-0.72	-0.74	-1.37	-0.81	-0.83
Average	-1.75583	-1.42333	-1.14333	-1.48917	-1.24667	-0.935

Bit Rate : 128kbps (CBR)

Sample Rate : 44100 Hz

P1 : ISO Standard Psychoacoustic Model 2

P4 : AM/GM Psychoacoustic Model

4.2 Experiments on Grouping Threshold

This section focuses on calibrating the grouping threshold *M*. The purpose of the grouping threshold is to control the scale factor errors in one group. If threshold *M* is large, the number of group decreases and it will enlarge the error of scale factor. Conversely, if the threshold *M* is small, the number of group and the side information will increase simultaneously.

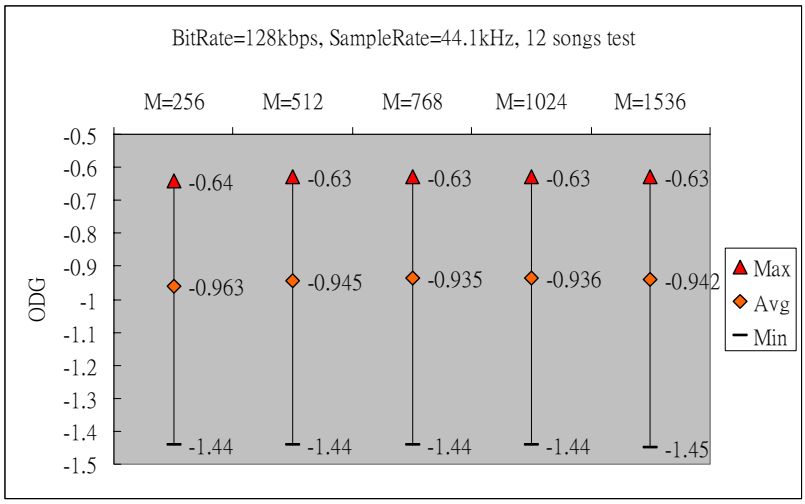

Figure 24: ODG for different Grouping Threshold based on the new window decision method.

Figure 24 summarizes the effect of the grouping threshold. Threshold around 768 has led to best quality.

4.3 Experiments on Coupling Method

Coupling keeps window types and grouping manners consistent in both stereo channels. Therefore, coupling methods play an important role in the join design between window switch and M/S coding. The window coupling method checks the similarity of two channels by a *PE* threshold *T1*. If the two channels are similar, another *PE* threshold *T2* is used to re-decide the window type in both channels. The first part of this section is to calibrate these two thresholds, *T1* and *T2*. Figure 25, Figure 26, and Figure 27 are the calibration of thresholds *T*1 and *T2*. As thresholds measurement in section 4.1, thresholds *T1* and *T2* are also measured repeatedly. After calibrating these thresholds, we measure the quality improvement and show it in the Figure 28 and Table 3.

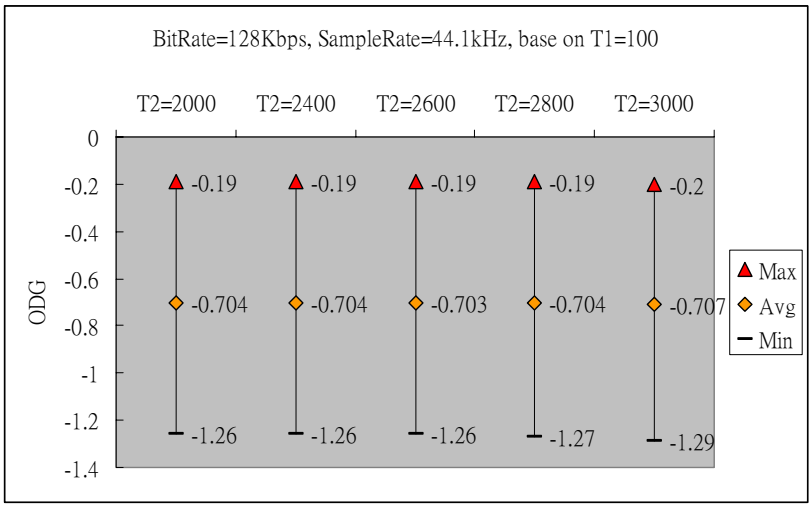

Figure 25: ODG for different *PE* Threshold T2 based on T1=100.

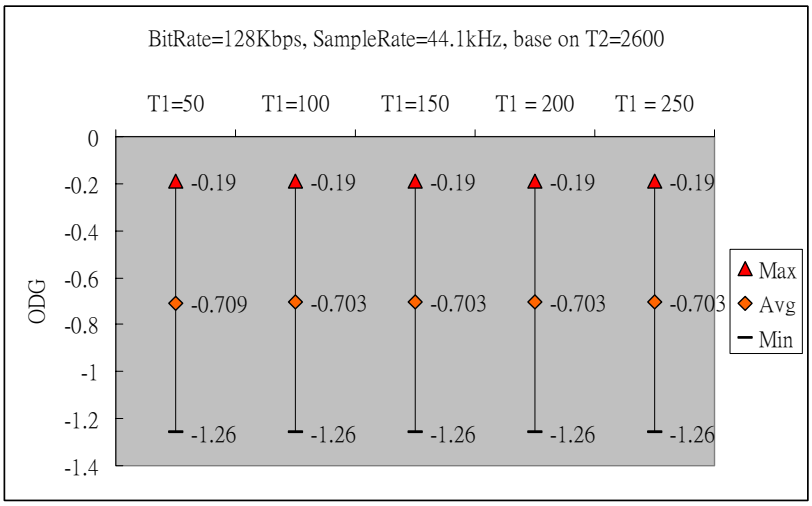

Figure 26: ODG for different *PE* Threshold T1 based on T2=2600.

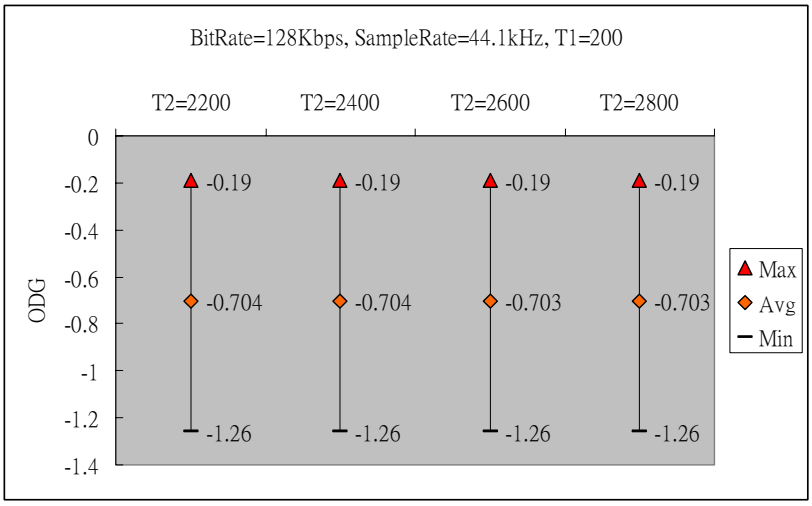

Figure 27: ODG for different *PE* Threshold T2 based on T1=200.

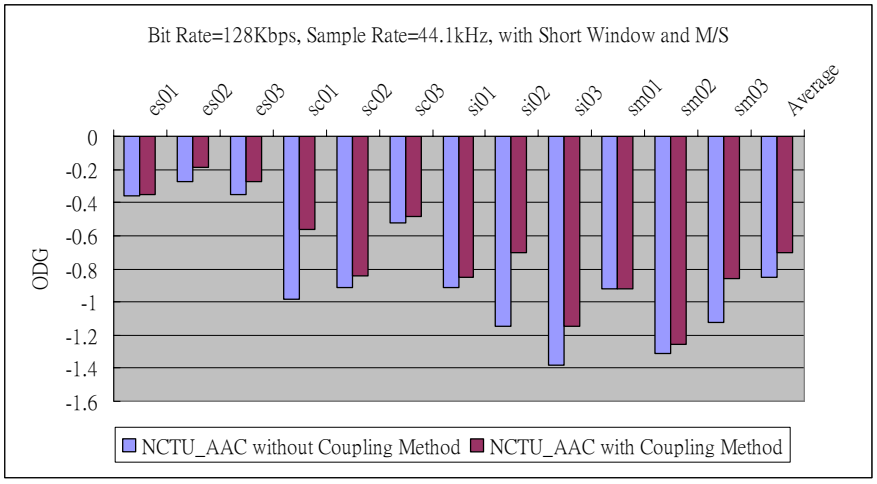

Figure 28: Objective test using P4 on the two methods: "NCTU_AAC without Coupling Method" and "NCTU_AAC with Coupling method".

Table 3: Detail ODG values of objective test on using coupling methods or not.

Psychoacoustic Model	P1	P1	P4	P4
Allow Short	YES	YES	YES	YES
M/S	M/S	M/S	M/S	M/S
Coordination	No	YES	No	YES
es01	-0.38	-0.35	-0.36	-0.35
es02	-0.22	-0.18	-0.27	-0.19
es03	-0.26	-0.27	-0.35	-0.27
sc01	-1.15	-0.62	-0.98	-0.56
sc02	-0.96	-0.87	-0.91	-0.84
sc03	-0.62	-0.57	-0.52	-0.48
si01	-2.01	-1	-0.91	-0.85
si02	-1.2	-0.75	-1.15	-0.7
si03	-2.57	-2.34	-1.38	-1.15
sm01	-1.48	-1.48	-0.92	-0.92
sm02	-1.45	-1.45	-1.31	-1.26
sm03	-0.9	-0.76	-1.12	-0.86
Average	-1.1	-0.88667	-0.84833	-0.7025

Bit Rate : 128kbps (CBR)

Sample Rate : 44100 Hz

P1 : ISO Standard Psychoacoustic Model 2

P4 : AM/GM Psychoacoustic Model

Figure 28 and Table 3 are the experiment results of coupling method. Table 3 shows that whether in P1 or P4 psychoacoustic model, coupling methods can improve the quality.

4.4 327 Tracks Test

In order to measure window switch quality, a large number of test bitstreams are needed. In PSPLab audio database [22], there are 16 sets and 327 tracks. For each bitstream set, they are briefly described in Table 4.

Table 4: The description for each bitstream set

Bitstreams categories	Number of Tracks	Remark
ff123	103	Killer bitstream collection from ff123.
Gpsycho	24	LAME quality test bitstream.
HA64KTest	39	64 Kbps test bitstream for multi-format in HA forum.
HA128KTestV2	12	128 Kbps test bitstream for multi-format in HA forum.
horrible_song	16	Collections of killer songs among all bitstream in PSPLab.
ingets1	5	Bitstream collection from the test of OGG Vorbis pre 1.0 listening test.
Mono	3	Mono test bitstream.
MPEG	12	MPEG test bitstream set for 48KHz.
MPEG44100	12	MPEG test bitstream set for 44100 Hz.
Phong	8	Test bitstream collection from Phong.
PSPLab	37	Collections of bitstream from early age of PSPLab. Some are good as killer.
Sjeng	3	Small bitstream collection by sjeng.
SQAM	16	Sound quality assessment material recordings for subjective tests.
TestingSong14	14	Test bitstream collection from rshong.
TonalSignals	15	Artificial bitstream that contains sin wave etc.
VORBIS_TESTS_Samples	8	
Total	327	

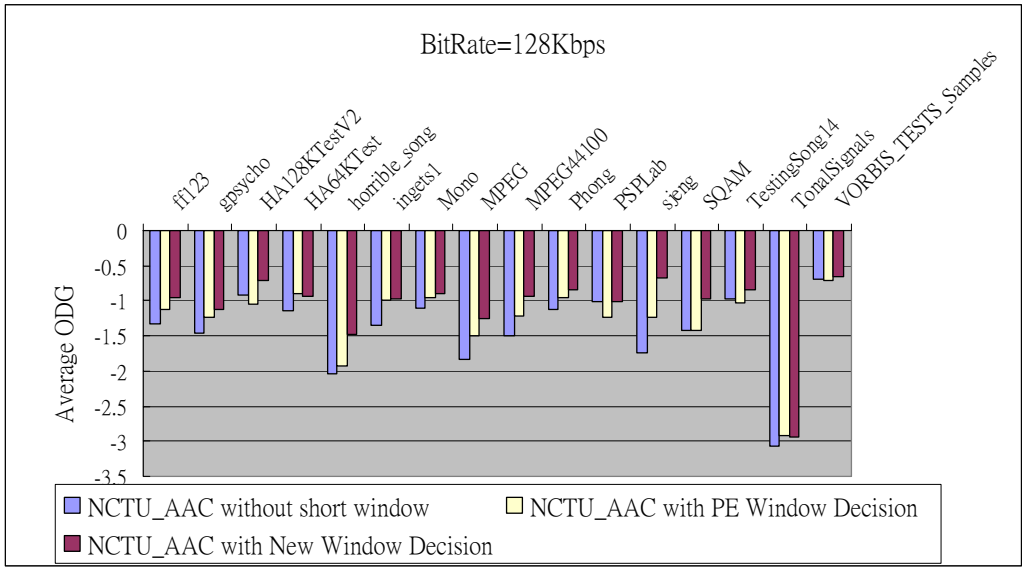

Figure 29: For 16 bitstream sets, objective test on the three methods: "NCTU-AAC 1.0 without short window", "NCTU-AAC 1.0 with *PE* Window Decision" and "NCTU-AAC 1.0 with New Window Decision".

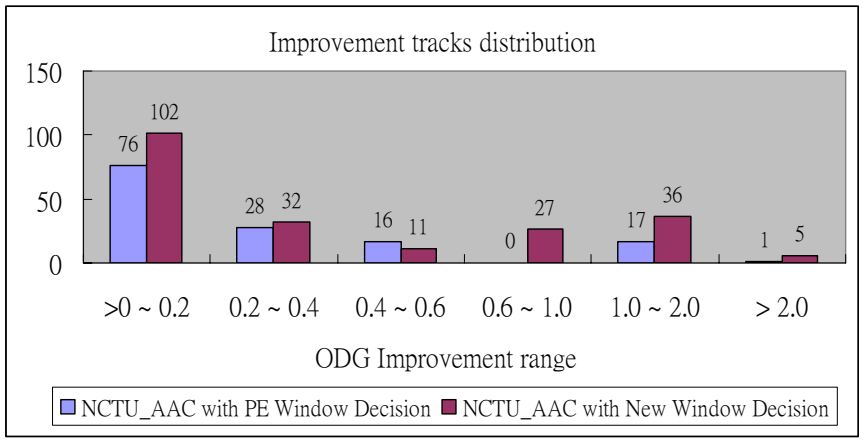

Figure 30: Dstribution of the improved tracks.

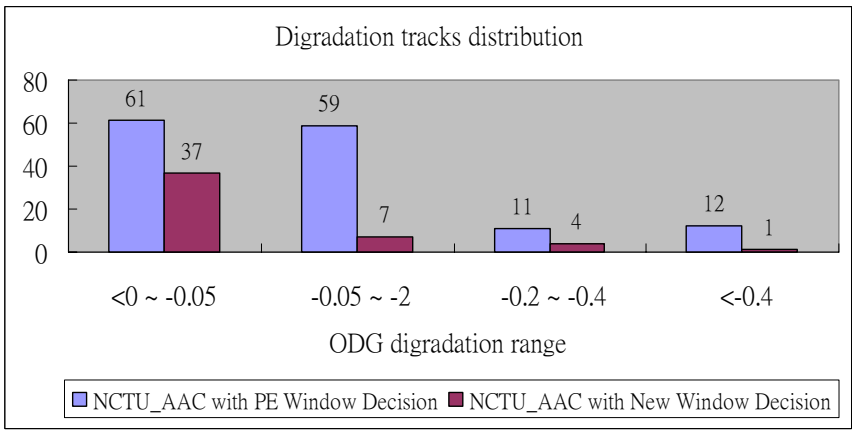

Figure 31: Dstribution of the degraded tracks..

Figure 29 illustrates the three experiments for the 16 bitstream sets, where each bar denotes the average ODG of each bitstream set. Generally, the new window decision method has a better quality than that of the *PE* decision method. In detail, for 327 tracks, Figure 30 illustrates the distribution of the improved tracks for the two different methods: window switch based on *PE* and window switch based on new method. The x-axis represents six different improvement ranges and the y-axis means the number of tracks improved. The different range is the difference between methods using and not using short window. It is clear that new window decision method has better improvement than that of the method based on *PE* in quantity and quality. Besides, Figure 31 illustrates the distribution of the degraded tracks distribution for the two different methods like Figure 30. Window decision method is still better than *PE* decision in the degradation tracks distribution. Most of the degradation tracks degrade below -0.05. There are only 7 tracks degrading by ODG value larger than 0.1. The reason causing the bad quality is that the tone in low frequency can't be precisely detected. Therefore, bad frequency resolution in short window induces bad encoded quality.

4.5 Experiments of Quality Comparison

In this section, we compare NCTU_AAC 1.0 and other two commercial AAC encoders, QuickTime [23] and Nero [24]. The experiments in this section focus on the quality comparison. The complexity comparison can be referred to [20].

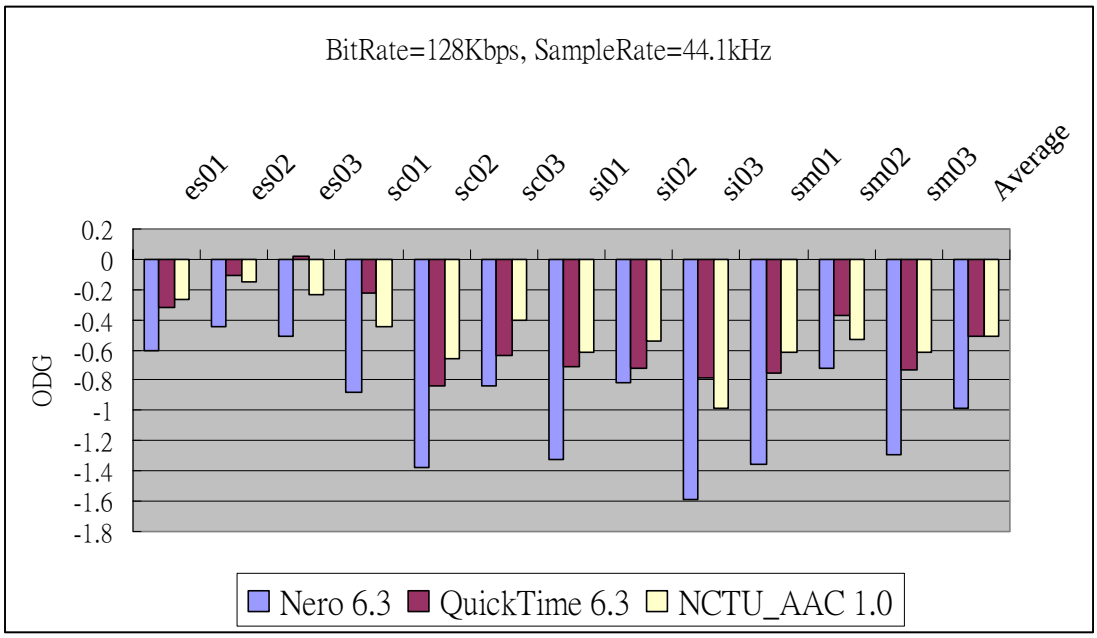

Figure 32: Objective test on the three encoders: "Nero 6.3", "QuickTime 6.3" and "NCTU-AAC 1.0".

Figure 32 illustrates three experiments for MPEG 12 tracks. Each bar denotes the ODG value of each track. All tracks encoded by NCTU_AAC 1.0 are better than that encoded by Nero 6.3. NCTU_AAC 1.0 has better encoding quality than QuickTime 6.3 in 7 tracks. In average, NCTU_AAC 1.0 performs better than other two encoders. Table 5 shows the detail result of the three encoders.

Table 5: Detail ODG result of quality comparison of three encoders.

	Nero 6.3	QuickTime 6.3	NCTU_AAC 1.0
es01	-0.6	-0.32	-0.27
es02	-0.45	-0.11	-0.15
es03	-0.51	0.02	-0.23
sc01	-0.88	-0.22	-0.45
sc02	-1.38	-0.84	-0.66
sc03	-0.84	-0.64	-0.4
si01	-1.32	-0.71	-0.62
si02	-0.82	-0.72	-0.54
si03	-1.59	-0.78	-0.98
sm01	-1.36	-0.75	-0.61
sm02	-0.72	-0.37	-0.53
sm03	-1.29	-0.73	-0.62
Average	-0.98	-0.51417	-0.505

Bit Rate : 128kbps

Sample Rate : 44100 Hz

NCTU_AAC uses AM/GM Psychoacoustic Model, Window Switch, TNS, M/S, and Bit Reservoir.

Chapter 5
Conclusion

This thesis has proposed a new window switch method to improve the quality. Firstly, window decision based on energy ratio, zero-crossing ratio, and tonal attack has been proposed. Secondly, short window psychoacoustic model is replaced by long window psychoacoustic model, and it aligns the masking by short window energy. Thirdly, grouping method has been proposed. Then, for the combination with TNS, the window type switch algorithm has been modified. Finally, for M/S coding, coupling methods on groups and window types has been proposed.

References

[1] ISO/IEC, “Coding of moving pictures and audio –IS 13818-7 (MPEG-2 advanced audio coding, AAC)”, Doc. ISO/IEC JTCI/SC29/WG11 n1650, Apr. 1997.

[2] ISO/IEC, “Information technology- coding of audiovisual objects”— ISO/IEC.D 4496 (Part 3, Audio), 1999.

[3] P. Masri and A. Bateman, “Improved modeling of attack transients in music analysis-resynthesis,” ICMC, 1996.

[4] J. Kliewer and A. Mertins, “Audio subband coding with improved representation of transient signal segments,” EUSIPCO-98, Sept. 1998.

[5] R. Vafin, R. Heusdens, S. Par and B. Kleijn, ”Improved modeling of audio signals by modifying transient locations,” IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001.

[6] Z. Hou, W. Dou and Z. Dong, “New window-switching criterion of audio compression,” Multimedia Signal Processing, 2001 IEEE Fourth Workshop on, 3-5 Oct. 2001.

[7] S. B. Venkata, K. M. Ashich, V. M. Vijayachandran and M. K. Vinay, ”Transient detection for transform domain coders,” AES 116th Convention, Germany, 8-11 May 2004.

[8] E. Zwicker and H. Fastl, “Psychoacoustics: facts and models,” Springer-Verlag, Berlin Heidelberg, 1990.

[9] K. Brandenburg and J. Johnston, “Second generation perceptual audio coding: the hybrid coder,” AES 88th Convention, Montreux, 13-16 Mar. 1990.

[10] Y. Mahieux and J. P. Petit, “High-quality audio transform coding at 64 kbps,” IEEE Transactions on Communications, Vol. 42, pp. 3010-3019, Nov. 1994.

[11] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson and Y. Oikawa, “ISO/IEC MPEG-2 advanced audio coding,” Journal of AES, Vol 45, no. 10, pp 789-814, October 1997.

[12] J. D. Johnston, “Transform coding of audio signals using perceptual noise criteria,” IEEE Journal on Selected Area in Communications. Vol. 6, No. 2, Feb. 1998.

[13] J. Herre and J. D. Johnston, “Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS),” AES 101st Convention, Los Angeles, 8-11 Nov. 1996.

[14] K. Brandenburg and G. Stoll, “The ISO/MPEG-cudio codec: a generic standard for coding of high quality digital audio,” AES 92nd Convention, Vienna, 24-27 Mar. 1992.

[15] C. M. Liu, W. J. Lee and R. S. Hong, “A new criterion and associated bit allocation method for current audio coding standards,” Proceedings of the 5th international Conference on Digital Audio Effects (DAFX), 2002.

[16] Chu-Ting Chien, “Bit allocation for MPEG-4 advanced audio coding,” CSIE Master Thesis of NCTU, 2003.

[17] A. Dueñas, R.Pérez, B. Rivas, E. Alexandre and A. Pena, “A robust and efficient implementation of MPEG-2/4 AAC natural audio coders,” AES 112th Convention, Munich, 10-13 May 2002.

[18] NCTU_AAC,
website http://psplab.csie.nctu.edu.tw/projects/index.pl/nctu-aac.html .

[19] Tzu-Wen Chang, “Efficient temporal noise shaping for MPEG 4 advanced audio coding,” CSIE Master Thesis of NCTU, 2004.

[20] Ting Chiou, “Efficient psychoacoustic model for MPEG-4 audio coding based on filterbank,” CSIE Master Thesis of NCTU, 2004.

[21] Yo-Hua Hsiao, “M/S coding enhancement in MP3 and AAC,” CSIE Master Thesis of NCTU, 2004.

[22] PSPLab audio database
website http://psplab.csie.nctu.edu.tw/projects/index.pl/testbitstreams.html .

[23] Apple, QuickTime,
website http://www.apple.com/quicktime/.

[24] Nero,
website http://www.nero.com/ .