國 立 交 通 大 學

多媒體工程研究所

碩 士 論 文

用於 MPEG 環場音訊差餘編碼之快速

QMF-to-MDCT 轉換

Fast QMF-to-MDCT Transform for Residual Coding in MPEG Surround

研 究 生：潘俊宇

指導教授：劉啓民　教授

蕭旭峯　教授

中 華 民 國　九 十 九　年　十 月

用於 MPEG 環場音訊差餘編碼之快速 QMF-to-MDCT 轉換

Fast QMF-to-MDCT Transform for Residual Coding in MPEG Surround

研 究 生：潘俊宇　　Student: Chun-Yu Pan

指導教授：劉啟民　　Advisor: Dr. Chi-Min Liu

蕭旭峯　　Dr. Hsu-Feng Hsiao

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

October 2010

Hsinchu, Taiwan, Republic of China

中 華 民 國　九 十 九　年　十　月

用於 MPEG 環場音訊差餘編碼之快速 QMF-to-MDCT 轉換

學生：潘俊宇　　指導教授：劉啓民 博士
蕭旭峯 博士
國立交通大學多媒體工程研究所碩士班

中文論文摘要

MPEG Surround (MPS)爲一低位元率多聲道音訊壓縮標準，其壓縮原理是透過降混音處理將多聲道合成雙聲道或是單聲道訊號。在此低位元率的壓縮標準中，差餘編碼提供了一個有彈性的切換機制，藉由傳輸降混音壓縮過程中的預測誤差，在解碼端達到低頻帶或是全頻帶高品質音訊的重建。因此，如何產生並解碼出正確的差餘訊號，成爲一個相當重要的議題。差餘訊號藉由 AAC 編碼器傳輸並解碼出 MDCT(Modified Discrete Cosine Transform)頻率係數提供給 MPS 差餘解碼器重建，爲了降低計算上的複雜度，此解碼器並未使用原本的解析/分析架構，而提供了一個從 MDCT 到 QMF 係數的近似快速轉換方法。

此篇論文，探討此快速轉換所造成的瑕疵，並提出新的快速編碼器架構，藉由改變差餘訊號的編碼架構，修正上述的問題。改進的效果將藉由頻譜和誤差計算呈現並驗證。

Fast QMF-to-MDCT Transform for Residual Coding in MPEG Surround

Student：Chin-Yu Pan　　Advisor：Dr. Chi-Min Liu

Dr. Hsu-Feng Hsiao

Institute of Multimedia and Engineering

College of Computer Science

National Chiao Tung University

Abstract

The MPEG Surround (MPS) is an important standard which defines the method to compress the multi-channel audio signals. In MPS, the residual coding with flexible rate control is an optional tool that encodes the prediction error between the downmix signal and the original signal for the target bit-rate in order to enhance the quality of the reconstructed signal. In the decoder defined in MPS standard, the residue can be decoded by the fast algorithm that maps the coefficients in the MDCT (Modified Discrete Cosine Transform) domain used in AAC to the QMF subband domain used in MPS to reconstruct the multi-channel signals. This coefficient mapping between different domains has greatly simplified the computation complexity required in the AAC encoder and the QMF analysis filterbank. However, the fast mapping algorithm cannot losslessly reconstruct the residue signal. In this thesis, the reconstructed loss caused by the fast mapping are investigated and two new fast mapping algorithms from the QMF subband domain to the MDCT domain are proposed for the encoder. From the simulation results, the proposed mapping algorithms show merits in terms of audio loss and computation complexity.

致謝

感謝劉啓民老師兩年來的栽培及蕭旭峯老師給予的指導，實驗室的許瀚文、高永軒學長以及同學陳德沛、董昀修的協助，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 MPEG Surround (MPS) standard [1] has achieved high compression efficiency by downmixing the multi-channel signals into the signal with a less number of channels. The downmixing transform is not invertible and can have a lot of audio quality degradation for some audio tracks. To achieve a better tradeoff between the quality and the bit rates, the residual coding has defined to encode the difference between the original signals and the signals reconstructed from the downmix signals [2][3]. In the decoder, to reduce the computation complexity in mapping the coefficients from the MDCT domain to the complex-valued QMF subband domain, a fast mapping has been adopted in the standard.

In the encoder, as suggested by the MPS standard, the forward transform from the QMF domain to the MDCT domain can be realized by concatenating a QMF synthesis filterbank and a forward MDCT. However, the structure of the forward transform in the encoder and the backward MDCT-to-QMF transform in the decoder is asymmetric. Accordingly, the suggested forward transform in the encoder cannot match the fast backward transform in the decoder which leads to the audio loss in the audio reconstruction. This thesis proposes the new encoding method to reduce the audio loss. We consider the design from two approaches. The first one proposes an inverse mapping matching with the decoding process. The second one reduces the loss through a least square criterion.

This thesis is organized as follows. Chapter 2 provides an overview of MPS coder. Chapter 3 introduces MPS residual coding and the decoding. Chapter 4 proposes two new encoding methods for MPS residual coding. Chapter 5 considers the method through experiments. Chapter 6 summarizes the thesis.

Chapter 2 Background

2.1 MPEG Surround Basic Concept

The concept of MPS is shown in Figure 1 [4]. A multichannel input is converted to a mono, stereo or five channel downmixed signal and a side information bit stream with spatial parameters. The differences between multichannel input signals and the reconstructed signals are called residual signals which are coded through the residual coding. The downmixed signals are processed by a legacy downmix encoder and combined with the spatial parameters and residual signals in the encoder. .

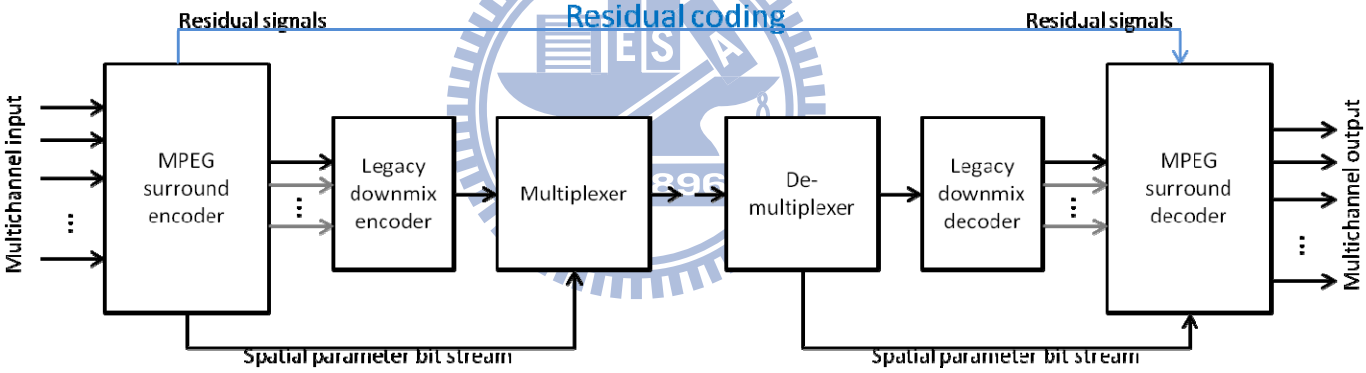


Figure 1. MPEG surround basic concept

2.2 MPEG Surround Encoder

In MPS encoder, the multichannel signal is downmixed through a tree structure which is cascaded by reverse Two-to-One (R-TTO) and reverse Three-to-Two (R-TTT) coding blocks [2]-[7]. Figure 2 shows the 5-1-5-1 and 5-2-5 tree structure, and the corresponding spatial parameters.

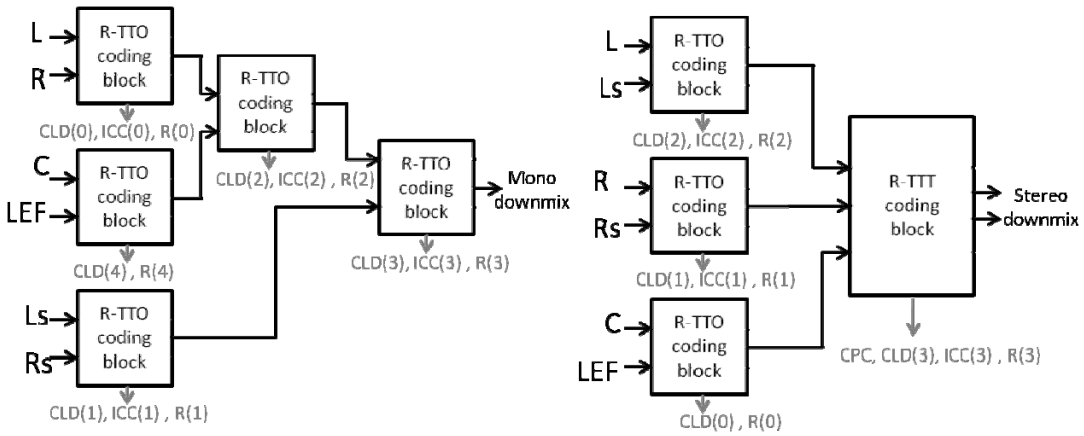


Figure 2. Tree structure of MPS encoder

Each TTO box converts two input channels into a mono output by downmix matrix H^{-1} which composes of spatial parameter ICC (interchannel correlation) and CLD (channel level difference). The downmixed operations are shown below

$$S(n) = H^{-1}X(n) = \begin{bmatrix} H11 & H12 \\ H21 & H22 \end{bmatrix}^{-1} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \begin{bmatrix} m(n) \\ d(n) \end{bmatrix} \tag{1}$$

$$H^{-1} = \begin{cases} \dfrac{1}{-c_1p_1 - c_2p_2} \begin{bmatrix} -1 & -1 \\ -c_2p_2 & c_1p_1 \end{bmatrix}, & \text{b} \le \text{residual bands} \\ \dfrac{1}{-c_1p_1c_2q_2 - c_2p_2c_1q_1} \begin{bmatrix} c_2q_2 & -c_1p_1 \\ -c_2p_2 & c_1q_1 \end{bmatrix}, & \textit{otherwise} \end{cases} \tag{2}$$

$$c_1 = \sqrt{\frac{10^{ICLD/10}}{1+10^{ICLD/10}}} \tag{3}$$

$$c_2 = \sqrt{\frac{1}{1+10^{ICLD/10}}} \tag{4}$$

$$p_1 = \cos(\alpha + \beta) \tag{5}$$

$$p_2 = \cos(-\alpha + \beta) \tag{6}$$

$$q_1 = \sin(\alpha + \beta) \tag{7}$$

$$q_2 = \sin(-\alpha + \beta) \tag{8}$$

$$\alpha = \frac{1}{2}\arccos(ICC) \tag{9}$$

$$\beta = \tan\left(\frac{c_2 - c_1}{c_2 + c_1}\arctan(\alpha)\right) \tag{10}$$

where X(n) is the input signals, the downmixed signal m(n) and residual signal d(n) are presented in S(n). There are two downmix methods in downmix matrix H^{-1}, depending on whether the residual coding is used in present band.

In analogy to the OTT encoder box, the TTT encoder box mixes three audio signals into two output channel. The downmix operation is shown below

$$\begin{aligned} S(n) &= N(k)TX(n) \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -r_1 & -r_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_l(n) \\ x_r(n) \\ \frac{1}{\sqrt{2}}x_c(n) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -r_1+1 & -r_2+1 & -r_1-r_2-1 \end{bmatrix} \begin{bmatrix} x_l(n) \\ x_r(n) \\ \frac{1}{\sqrt{2}}x_c(n) \end{bmatrix} \\ &= \begin{bmatrix} s_l(n) \\ s_r(n) \\ d(n) \end{bmatrix} \end{aligned} \tag{11}$$

It's assumes that the output channels S_l and S_r is a linear combination of three input channels x_l, x_r and x_c. This downmix process accomplished by downmixed matrix T and predicted matrix N with two channels predict coefficients (CPC) r_1 and r_2. The prediction error d(n) called residual signals, which can be transmitted to the MPS decoder to enable prefect waveform reconstruction.

2.3 MPEG Surround Decoder

In MPS decoder, the corresponding decode process is achieved. Figure 3 show

5-1-5-1 and 5-2-5 cascade decoder.

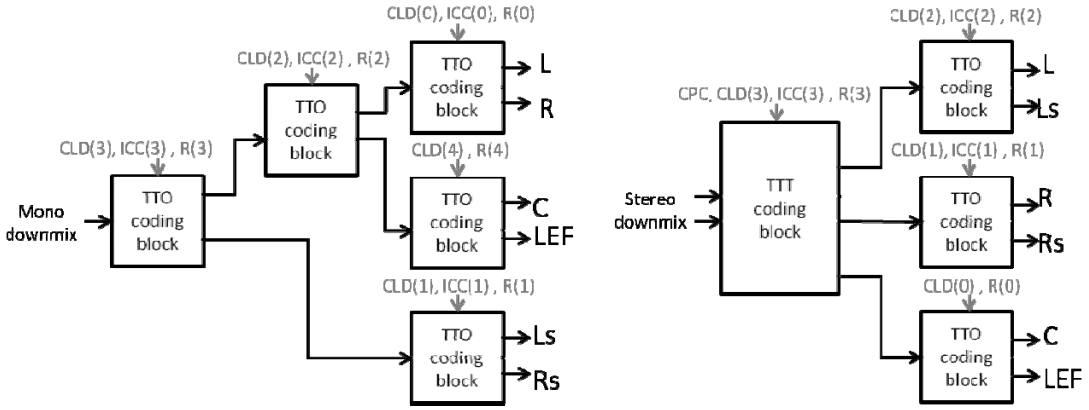


Figure 3. Tree structure of MPS decoder

Each One-to-Two (OTT) decoding box reconstructs the mono signals to the stereo.

$$X(n) = HS(n) = \begin{bmatrix} H11 & H12 \\ H21 & H22 \end{bmatrix} \begin{bmatrix} m(n) \\ d(n) \end{bmatrix} = \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} \tag{22}$$

$$H = \begin{cases} \begin{bmatrix} c_1 p_1 & 1 \\ c_2 p_2 & -1 \end{bmatrix}, & \text{b} \le \text{residual bands} \\ \begin{bmatrix} c_1 p_1 & c_1 q_1 \\ c_2 p_2 & c_2 q_2 \end{bmatrix}, & otherwise \end{cases} \tag{13}$$

The definitions of these coefficients c_1, c_2, p_1, p_2, q_1 and q_2 are same as the encoder. It's satisfy that

$$HH^{-1} = I \tag{14}$$

Also, the Two-to-Three decoding box recovers the downmixed signals to three channels by invert the encoding process with

the predict coefficient r_1 and r_2 which transmitted from encoder. The upmix operation is shown below

$$
\begin{aligned}
X(n) &= (NT)^{-1} S(n) \\
&= \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -r_1+1 & -r_2+1 & -r_1-r_2-1 \end{bmatrix}^{-1} \begin{bmatrix} s_l(n) \\ s_r(n) \\ d(n) \end{bmatrix} \\
&= \frac{1}{3} \begin{bmatrix} r_1+2 & r_2-1 & 1 \\ r_1-1 & r_2+2 & 1 \\ 1-r_1 & 1-r_2 & -1 \end{bmatrix} \begin{bmatrix} s_l(n) \\ s_r(n) \\ d(n) \end{bmatrix} \\
&= \begin{bmatrix} x_l(n) \\ x_r(n) \\ \frac{1}{\sqrt{2}} x_c(n) \end{bmatrix}
\end{aligned}
\tag{35}
$$

If the residual signal is not transmitted to decoder, the decorrelator will generate the needed decorrelated signal to goal upmix process.

To speed up these decoding processes, MPS decoder modifies the cascade tree structure to parallel matrix structure. Instead of cascade several coding boxes and process them separately, MPS decoder generates all needed decorrelated signals in advance and upmixes in one step.

The changes of the architecture are schematized in Figure 4.

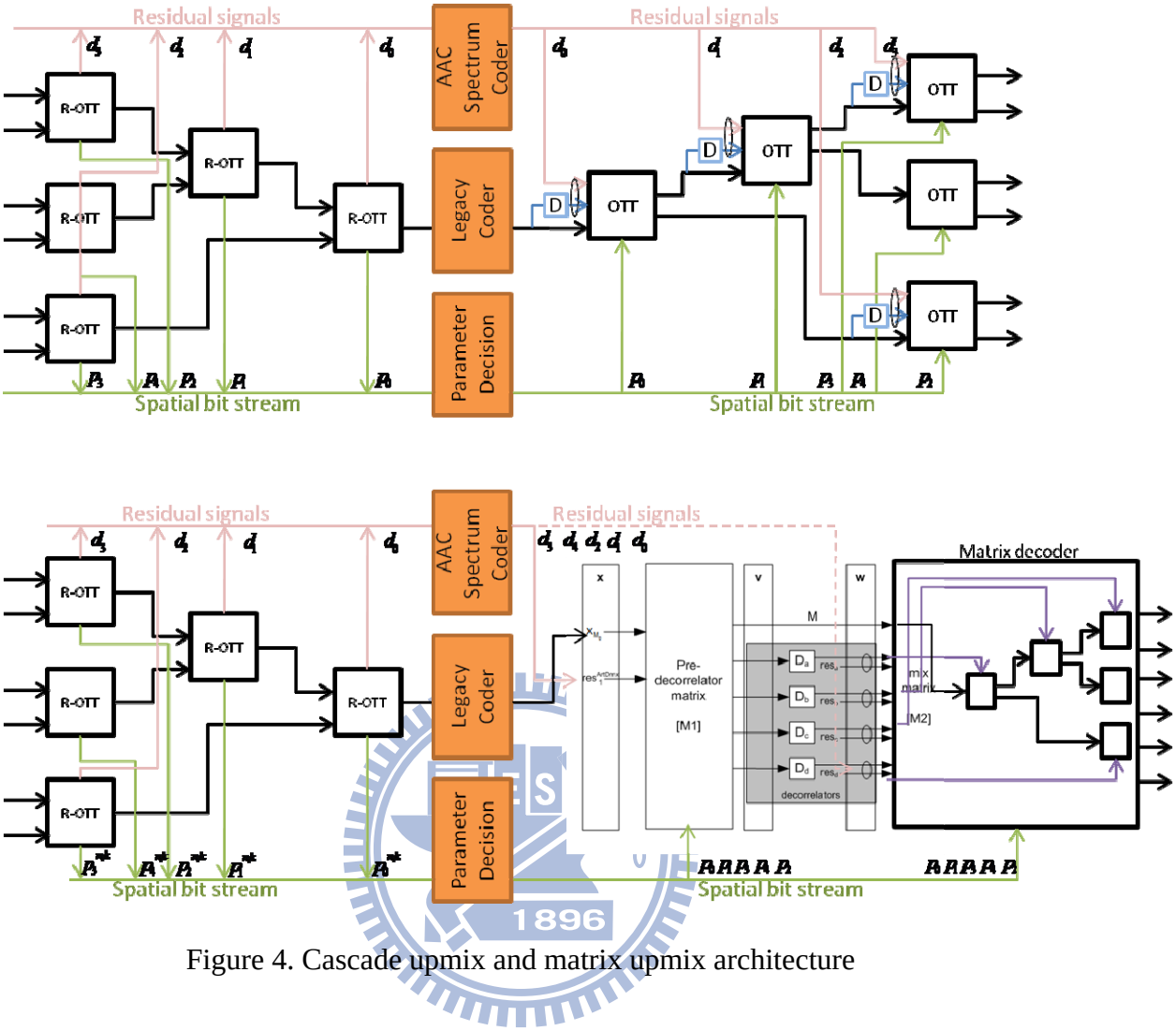


Figure 4. Cascade upmix and matrix upmix architecture

Chapter 3 MPS Residual Decoder

3.1 MPS Residual Coding

The residual signals generated by MPS encoder will be sent to MPS residual encoder. It transforms the QMF domain signals to MDCT domain coefficients and compressed by AAC spectrum encoder. The MPS decoder decoded the MDCT coefficients and mapping them back to QMF domain. These residual signals will be added to the up-mixed signals directly to achieve high quality reconstruction. The block diagram is shown in Figure 5.

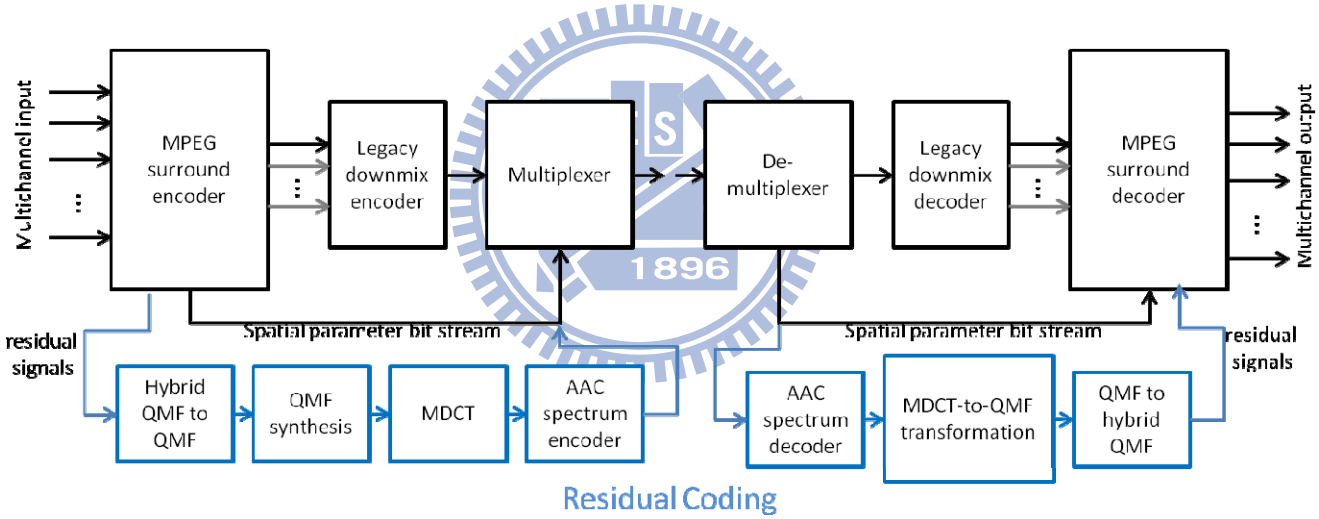


Figure 5. Block diagram of MPS residual coding

3.2 Basic Concept

The flow chart of MPS residual decoder can be shown in Figure 5. It decodes the MDCT coefficients which transmitted from encoder to QMF domain, complex residual signals. This process can simply decompose to two parts, pass and stop band reconstruction and positive complexfication. The more detail we will show in next section.

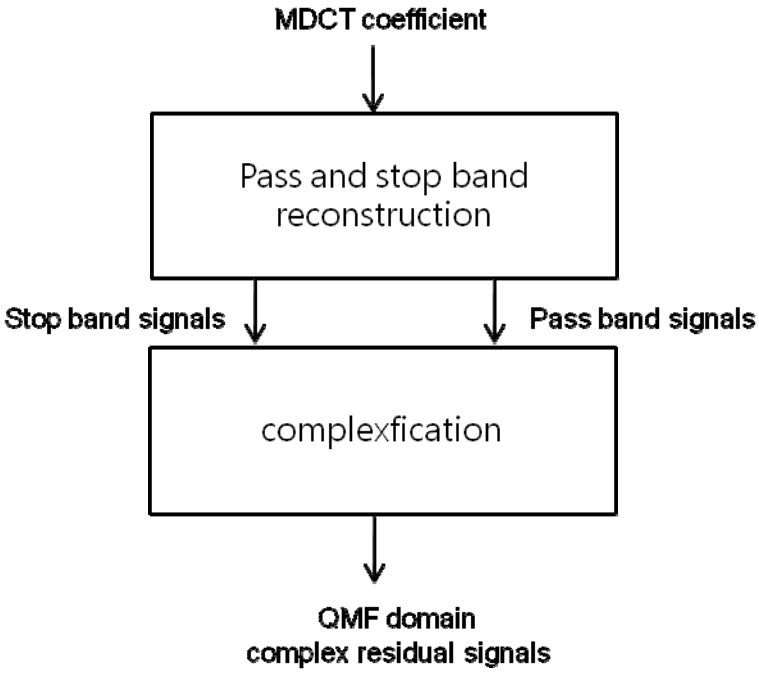


Figure 6. MPS residual decoder

3.3 Pass and Stop Band Reconstruction

To reconstruct the complex residual signals from real MDCT coefficients, MPS residual decoder proposes a fast method to rebuild the original signal. This process is shown in upper part of Figure 6.

In this rebuild step, MPS residual decoder generates two signals to simulate the real part and imaginary part of complex signals from real input.

The first step of complexficated process is permuting the MDCT coefficients S to QMF-domain form, which composes of time domain QMF coefficients on each QMF band. Next, copy the reordering signals $S_{permute}$ to generate stop band signals S_{stop} and pass band signals S_{pass}. The operations are shown below

$$S_{pass}(k,r) = S(kN + r) \tag{16}$$

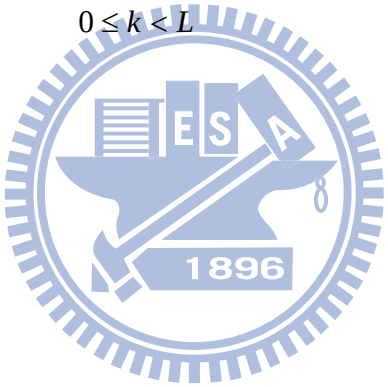
$$S_{pass}(k,r) = W_{pass}(r) S_{permute}(k,r) \tag{17}$$

$$S_{stop}(k,r)=\begin{cases} W_{stop}(r)S_{permute}(k,N-1-r) & ,\frac{N}{2}\le r<N \quad ,k=0 \\ W_{stop}(r)S_{permute}(k-1,r) & ,\frac{N}{2}\le r<N \quad ,1\le k<L \\ W_{stop}(r)S_{permute}(k+1,r) & ,0\le r<\frac{N}{2} \quad ,0\le k<L-1 \\ W_{stop}(r)S_{permute}(L,r) & ,0\le r<\frac{N}{2} \quad ,k=L-1 \quad ,L<64 \\ W_{stop}(r)S_{permute}(L-1,N-1-r) & ,0\le r<\frac{N}{2} \quad ,k=L-1 \quad ,L=64 \end{cases} \tag{18}$$

where N is half of frame length, L is how many residual bands used. r is QMF coefficient index on each QMF band k. W_{stop} and W_{pass} are MDCT windows. The ranges of r and k are

$$0\le r<N \tag{19}$$

$$0\le k<L \tag{20}$$



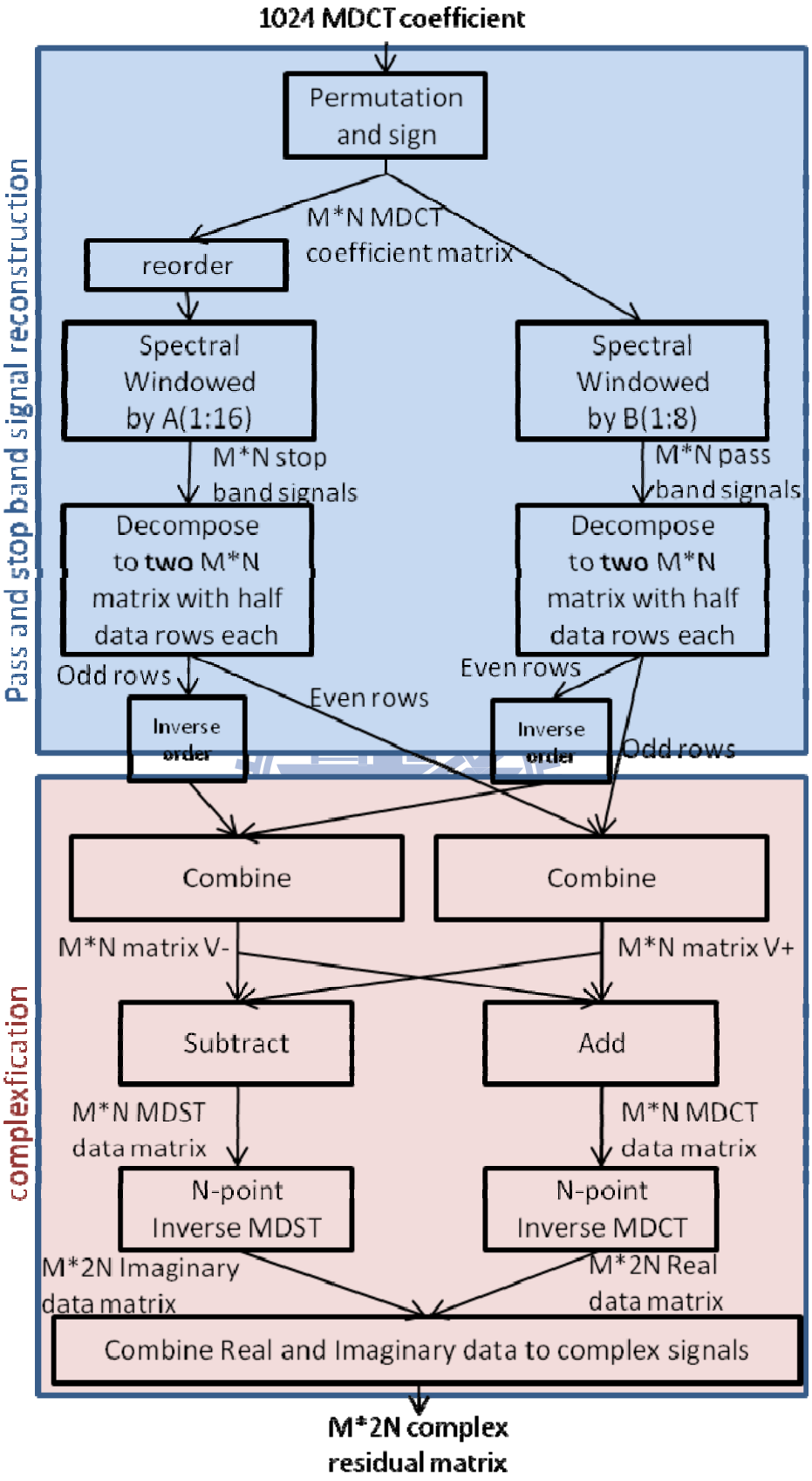


Figure 7.Flow chart of MPS residual decoder

After the reorganization step, the stop band and pass signals are determined. They prepare to simulate the positive frequency and negative frequency signals on the next reconstruction step. Figure 7 shows the FFT result of reconstructed signals of original residual signals, pass band and stop signals separately.

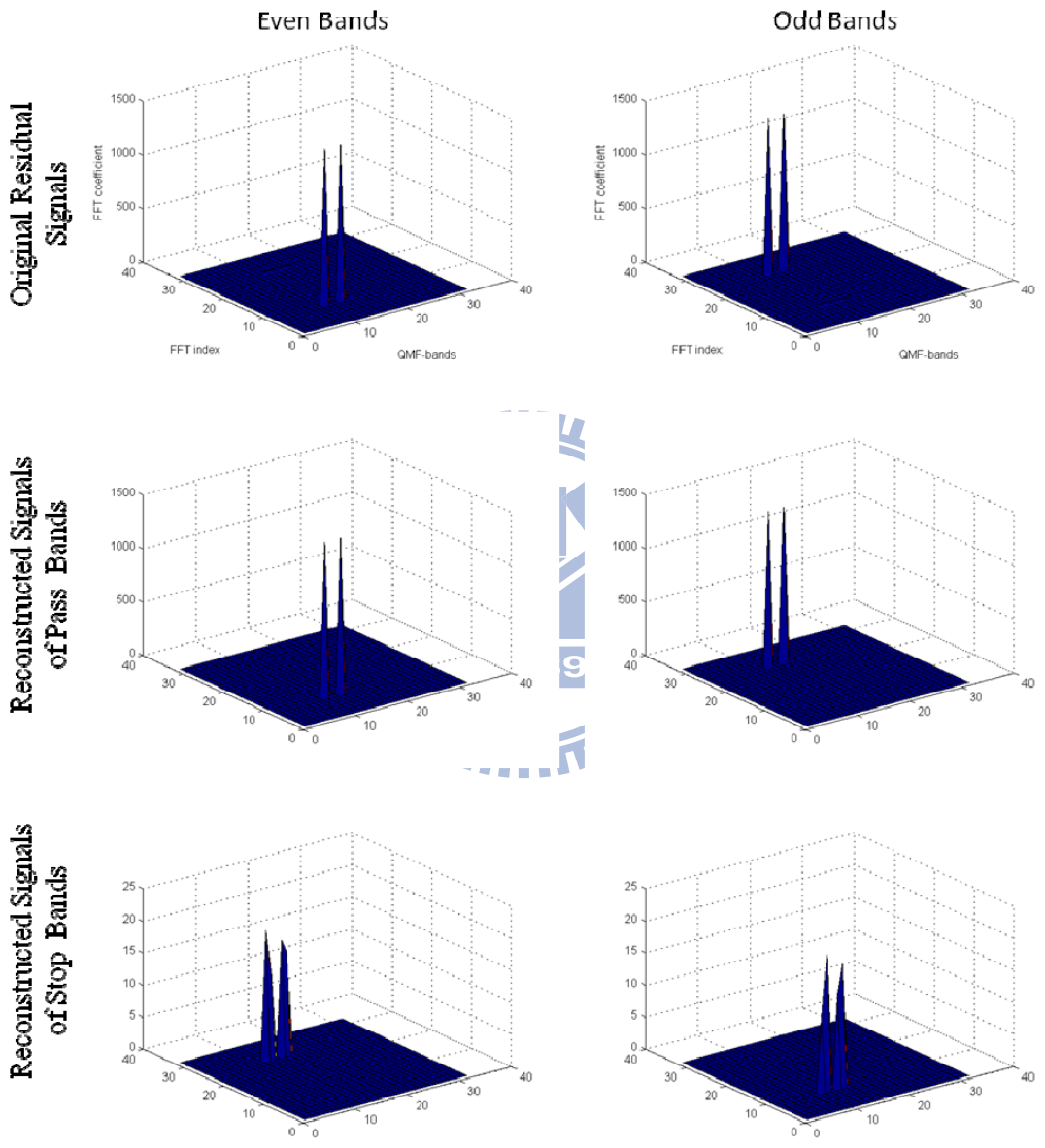


Figure 8. Complexfication process in MPS residual decoder

In each graph, x-axis indicates the QMF bands of even or odd index, y-axis and z-axis show the 32-point FFT index and magnitude of FFT coefficients. The complexfication process is illustrated in next section.

3.4 Complexfication

The second part of the MPS residual decoder is frequency reconstruction. It uses the pass band and stop band signals to reconstruct the complex QMF signals.

In the lower part of figure 6, we can see that the first step is collecting the needed signals which should be recover in positive and negative frequency from stop band and pass band signals to generate the temp signals V_{plus} and V_{minus}. The operations are

$$V_{plus} = \begin{cases} S_{pass}(k,r), & \text{k} \in \text{even} \\ S_{stop}(k,r), & \text{k} \in \text{odd} \end{cases} \tag{21}$$

$$V_{minus} = \begin{cases} S_{pass}(k,N-1-r), & \text{k} \in odd \\ S_{stop}(k,N-1-r), & \text{k} \in \text{even} \end{cases} \tag{22}$$

and these temp signals will be sent to MDCT and MDST to achieve frequency reconstruction. We can decompose the reconstruction of positive and negative frequency separately:

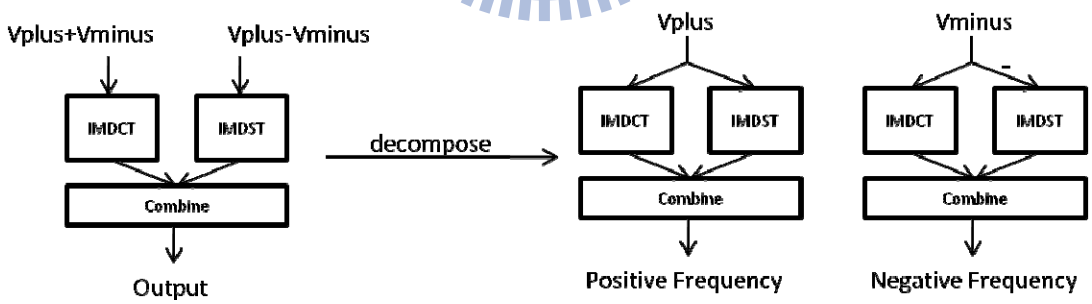


Figure 9. Decompose the frequency reconstruction

and we can analysis these processes using extended MDFT. The upper half rows of each matrix show the positive frequency and the lower half rows show the negative frequency parts.

$$M'_{EMDFT} = \begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDCT} - j \begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDST} \tag{23}$$

$$\begin{aligned}
&M'_{EMDFT}\, S_{plus} \\
&= \left(\begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDCT} - j \begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDST} \right) \left[(M_{MDCT})^T V_{plus} + j (M_{MDST})^T V_{plus} \right] \\
&= \begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDCT} (M_{MDCT})^T V_{plus} + \begin{bmatrix} I_N \\ J_N \end{bmatrix} M_{MDST} (M_{MDST})^T V_{plus} \\
&\quad + j \left[\begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDCT} (M_{MDST})^T V_{plus} + \begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDST} (M_{MDCT})^T V_{plus} \right] \\
&= \begin{bmatrix} I_N \\ -J_N \end{bmatrix} V_{plus} + \begin{bmatrix} I_N \\ J_N \end{bmatrix} V_{plus} \\
&= \begin{bmatrix} 2V_{plus} \\ 0_N \end{bmatrix}
\end{aligned} \tag{24}$$

$$\begin{aligned}
&M'_{EMDFT}\, S_{minus} \\
&= \left(\begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDCT} - j \begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDST} \right) \left[(M_{MDCT})^T J_N V_{minus} - j (M_{MDST})^T J_N V_{minus} \right] \\
&= \begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDCT} (M_{MDCT})^T J_N V_{minus} - \begin{bmatrix} I_N \\ J_N \end{bmatrix} M_{MDST} (M_{MDST})^T J_N V_{minus} \\
&\quad + j \left[\begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDCT} (M_{MDST})^T J_N V_{minus} - \begin{bmatrix} I_N \\ -J_N \end{bmatrix} M_{MDST} (M_{MDCT})^T J_N V_{minus} \right] \\
&= \begin{bmatrix} I_N \\ -J_N \end{bmatrix} J_N V_{minus} - \begin{bmatrix} I_N \\ J_N \end{bmatrix} J_N V_{minus} \\
&= \begin{bmatrix} J_N \\ -I_N \end{bmatrix} V_{minus} - \begin{bmatrix} J_N \\ I_N \end{bmatrix} V_{minus} \\
&= \begin{bmatrix} 0_N \\ -2V_{minus} \end{bmatrix}
\end{aligned} \tag{25}$$

with these properties

$$M_{MDCT} M_{MDST}{}^T = 0_N \tag{26}$$

$$M_{MDST} M_{MDCT}{}^T = 0_N \tag{27}$$

After these processes, we can get the complex signals from real coefficient to goal complexfication and signal reconstruction.

Chapter 4 MPS Residual Encoder

4.1 MPS Suggested Coding

To encode the residual signals from complex to real MDCT coefficients, MPS suggested an immediate method to realize the transformation from QMF to MDCT. It concatenates a QMF synthesis filter bank and a forward MDCT. The upper part of Figure 10 shows the architecture.

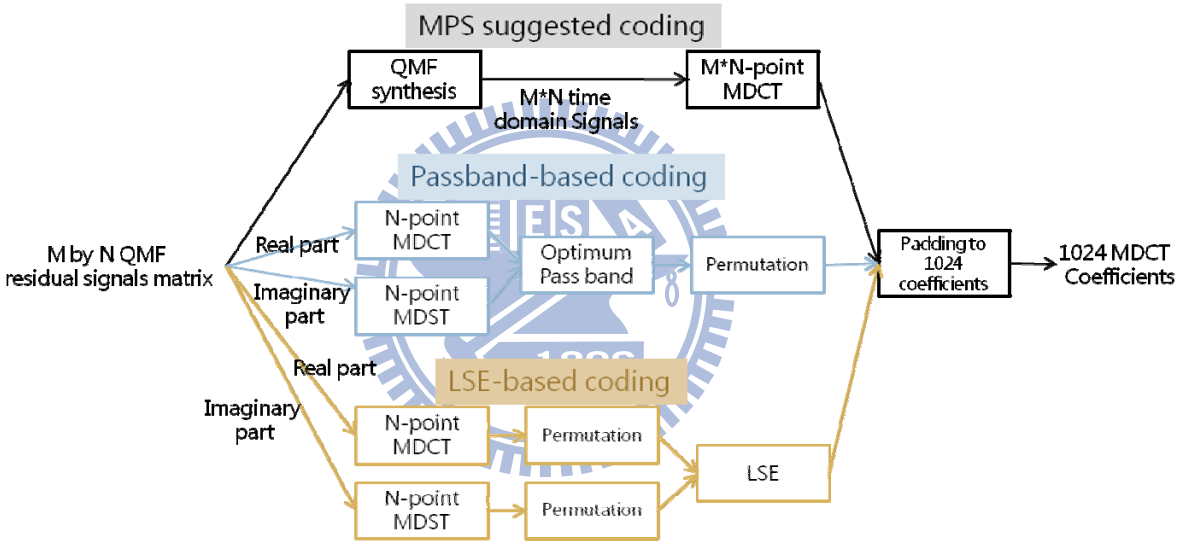


Figure 10. Residual encoder

4.2 Passband-Based Coding

However, the MPS proposed encoded method does not match with the fast mapping decoder. The MPS residual decoder mapping the real MDCT to complex QMF coefficients directly without synthesize back to time domain. It will cause two issues: First, the mismatch of frequency resolution and scale coefficient. To normalize the transformation scale, the MDCT and IMDCT scale coefficient pair should satisfy

$$Scale_{MDCT} * Scale_{IMDCT} = 1/N \tag{28}$$

where N is MDCT frame length.

However, the IMDCT length is 16-point in decoder and 1024-point in MPS suggested encoder is unequal. The mismatch will cause incorrect energy of reconstructed signals. Second, instead of perform a complete synthesis/analysis process to realize the complexfication. The MPS residual decoder uses a fast MDCT to QMF domain transform to approximate the original process. It uses the MDCT coefficient to simulate the real and imaginary signals directly. This approximate method is lossy. Therefore, the corresponding encoded method should be modified. We propose a fast algorithm achieve the QMF to MDCT mapping by perform the short-time MDCT in each QMF band. And permute the MDCT coefficients to satisfy the AAC spectrum coder which deal with the residual signals. The flow chart shows in middle part of Figure 10.

4.3 Least-Square–Estimation-Based Coding

Furthermore, to design a suitable encoder to match the residual decoder and minimize the reconstructed error of real to complex transformation, we impose the least square error algorithm on fast mapping in encoder. Let the objective function be

$$\begin{aligned} O(s_l) &= \left\| D_{\beta,l} D_s s_l - b_l \right\|^2 + \left\| D_{\alpha,l} P_l D_s s_l - a_l \right\|^2 \\ &= \left(D_{\beta,l} s_l - b_l \right)^T \left(D_{\beta,l} s_l - b_l \right) + \left(D_{\alpha,l} P_l s_l - a_l \right)^T \left(D_{\alpha,l} P_l s_l - a_l \right) \\ &= s_l{}^T \left(D_{\beta,l} D_{\beta,l} \right) s_l - b_l{}^T D_{\beta,l} s_l - s_l{}^T D_{\beta,l} b_l + b_l{}^T b_l \\ &\quad + s_l{}^T \left(\left(D_{\alpha,l} P_l \right)^T D_{\alpha,l} P_l \right) s_l - a_l{}^T D_{\alpha,l} P_l s_l - s_l{}^T D_{\alpha,l} P_l a_l + a_l{}^T a_l \\ &= s_l{}^T (D_{\beta,l}{}^2 + P_l{}^T D_{\alpha,l}{}^2 P_l) s_l - 2(b_l{}^T D_{\beta,l} + a_l{}^T D_{\alpha,l} P_l) s_l + b_l{}^T b_l + a_l{}^T a \end{aligned} \tag{29}$$

$$b_{16*k+r} = \begin{cases} C(\tilde{x}_{R,k}) + S(\tilde{x}_{I,k}), if\ k \in even \\ J_N C(\tilde{x}_{R,k}) - J_N S(\tilde{x}_{I,k}), if\ k \in odd \end{cases} \tag{30}$$

$$a_{16*k+r} = \begin{cases} C(\tilde{x}_{R,k}) - S(\tilde{x}_{I,k}), if\ k \in even \\ J_N C(\tilde{x}_{R,k}) + J_N S(\tilde{x}_{I,k}), if\ k \in odd \end{cases} \quad (31)$$

Find an optimal MDCT coefficient vector s_l such that the reconstructed error is minimum, where D_α and D_β are diagonal window matrices for pass band and stop band signal. D_s is the sign matrix, b_l and a_l are objective pass band and stop band signals compose of real and imaginary part of residual signal x. C and S are MDCT and MDST matrices. We can find the minimum error solution by differentiating

$$s_l^T (D_{\beta,l}^2 + P_l^T D_{\alpha,l}^2 P_l) s_l - 2(b_l^T D_{\beta,l} + a_l^T D_{\alpha,l} P_l) s_l + b_l^T b_l + a_l^T a \quad (32)$$

with respect to MDCT coefficient vector s_j to get the solution

$$s_l = (D_{\beta,l}^2 + P_l^T D_{\alpha,l}^2 P_l)^{-1} \left(D_{\beta,l} b_l + P_l^T D_{\alpha,l} a_l \right) \quad (33)$$

The LSE fast mapping is shown in lower part of Figure 10, and the experiment results will be shown in next chapter.

Chapter 5 Experiment Results

5.1 Experiment Method

The experiment is conducted in the following ways:

1. Use three different residual encoders, MPS suggested coding, passband-based coding and LSE-based coding, to covert the residual signals from QMF to MDCT domain.
2. Decode the MDCT coefficients to QMF domain residual signals and compare the reconstructed results for each encoder.

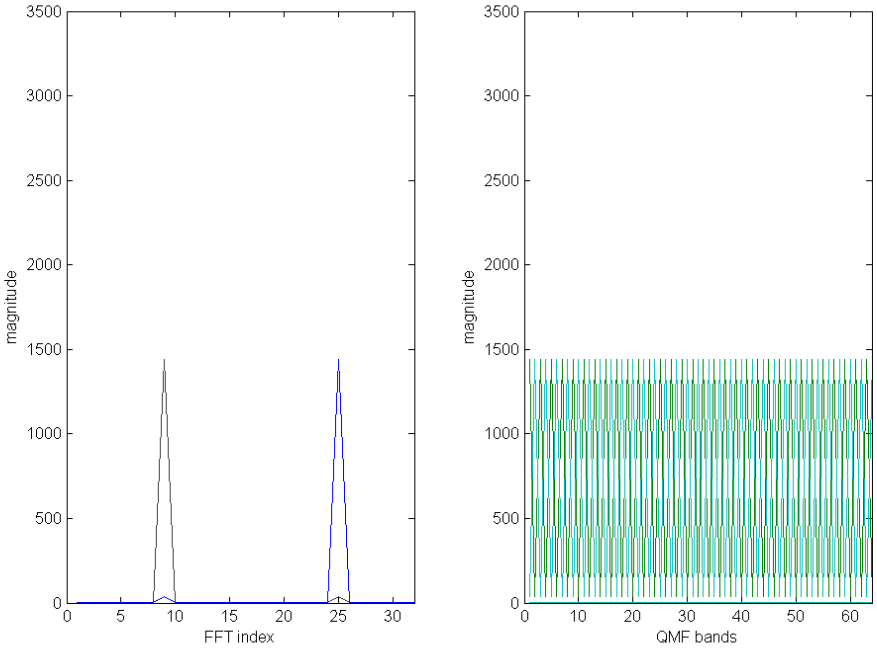
5.2 Experiment on Sinusoidal Signals

In the first experiment, we use a tonal signal consisting of 64 sinusoidal components to confirm the reconstruction situations on the 64 QMF subbands. For evaluating the effects of the different coding methods, we compare the reconstructed residual signals with the original one through DFT (discrete Fourier transform) spectra.

The error is computed as follows.

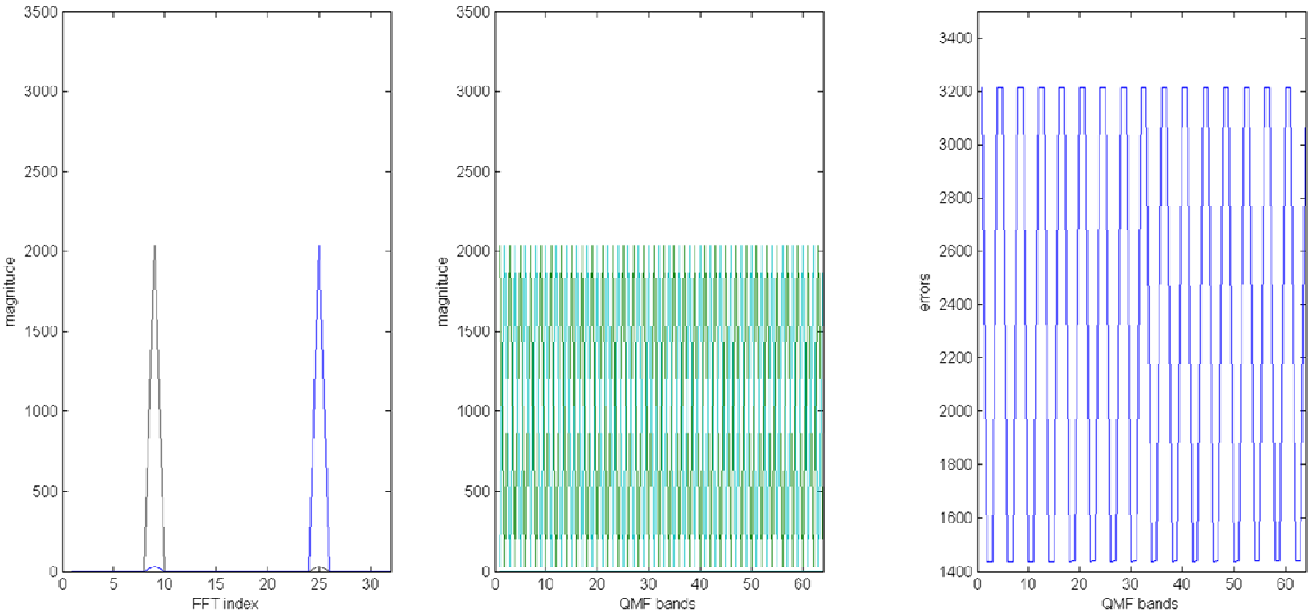
$$\sum_{k=0}^{63}\sum_{n=0}^{N-1}(x_k[n]-y_k[n])(x_k[n]-y_k[n])^* \tag{34}$$

where $x_k[n]$ is the original residual subband signal, $y_k[n]$ is the decoded residual subband signal; k is the subband index, n is the time index, and N is the frame length.



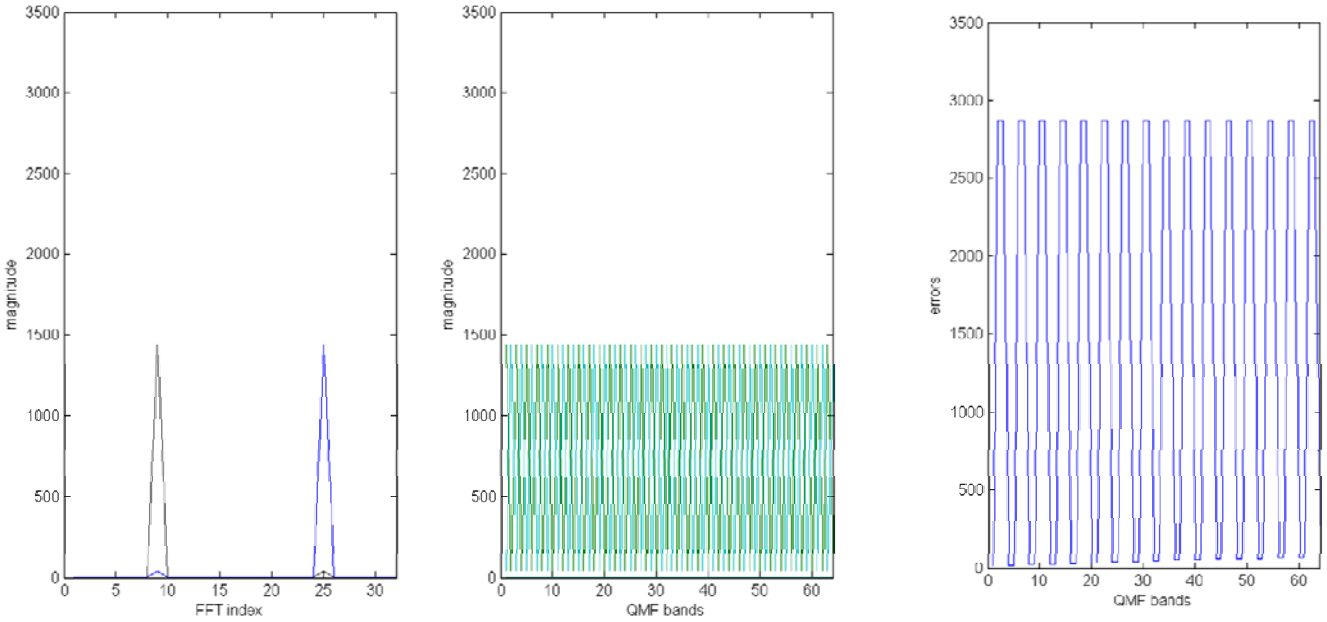
(a) **(b)**

Figure 11. DFT spectra of the original residual signal simulated by a tonal signal consisting of 64 sinusoild components. (a) DFT (magnitude) spetrcrum in each QMF band, where the spectra of 64 subbands are overlapped. (b) DFT responses of 64subbands, where all DFT frequency lines are overlapped into one line in each QMF subband. (The DFT spectrum is computed by 32-point FFT.)



(a) **(b)** **(c)**

Figure 12. DFT spectra of the reconstructed signal using MPS suggested coding and the corresponding reconstructed error. (a) DFT (magnitude) spetrcrum in each QMF band, where the spectra of 64 subbands are overlapped. (b) DFT responses of 64subbands, where all DFT frequency lines are overlapped into one line in each QMF subband. (c) error in each QMF band.



(a) **(b)** **(c)**

Figure 13. DFT spectra of the reconstructed signal using passband-based coding and the corresponding reconstructed error. (a) DFT (magnitude) spetrcrum in each QMF band, where the spectra of 64 subbands are overlapped. (b) DFT responses of 64subbands, where all DFT frequency lines are overlapped into one line in each QMF subband. (c) error in each QMF band.

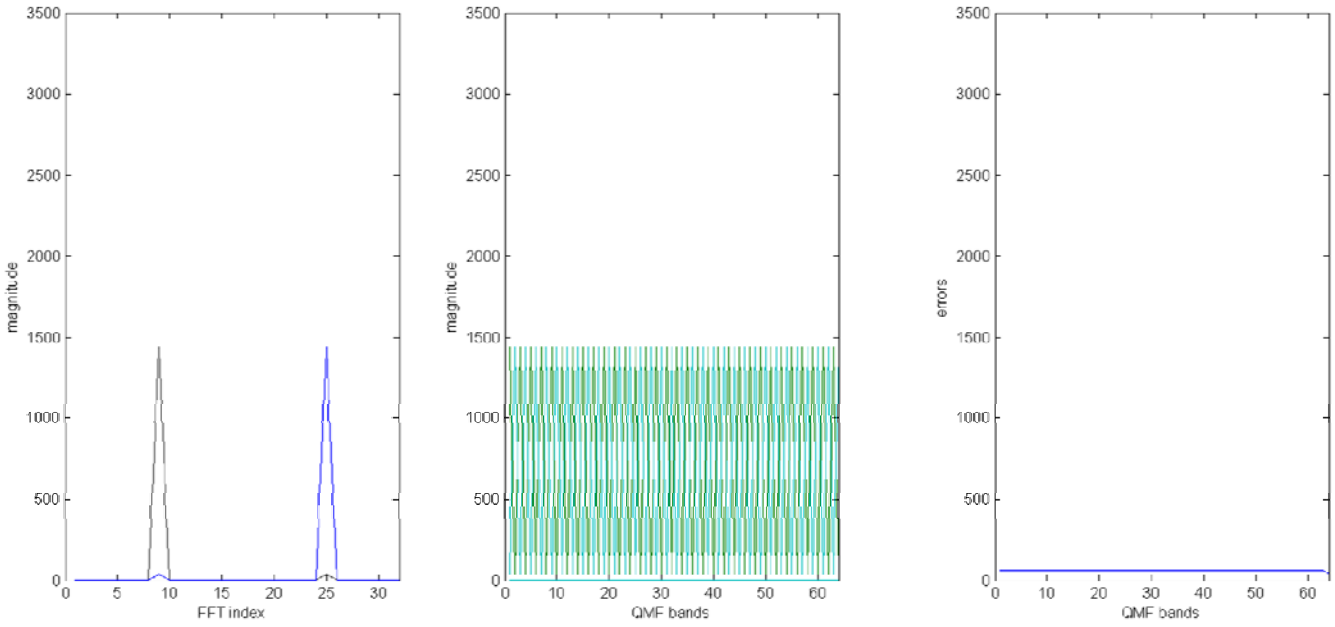


Figure 14. DFT spectra of the reconstructed signal using LSE-based coding. (a) DFT (magnitude) spetrcrum in each QMF band, where the spectra of 64 subbands are overlapped. (b) DFT responses of 64subbands, where all DFT frequency lines are overlapped into one line in each QMF subband. (c) error in each QMF band.

Table 1. The energy ratio between the original and reconstructed residual signals

	Energy(reconstructed signal) / Energy(original residual signal)
MPS Suggested Coding	1.41
Passband-Based Coding	1
LSE-Based Coding	1

Table 2. Signal to noise ratio of original signal and errors per frame

	SNR=(10* $\log_{10}$ (Energy(original)/ Energy(error))
MPS Suggested Coding	3.87
Passband-Based Coding	4.92
LSE-Based Coding	33.86

The energy ratio between the original and reconstructed signals and the signal to noise ratio between original signals and errors per frame are shown in Table 1 and Table 2, respectively. In Table 1, the energy of an audio frame is computed by

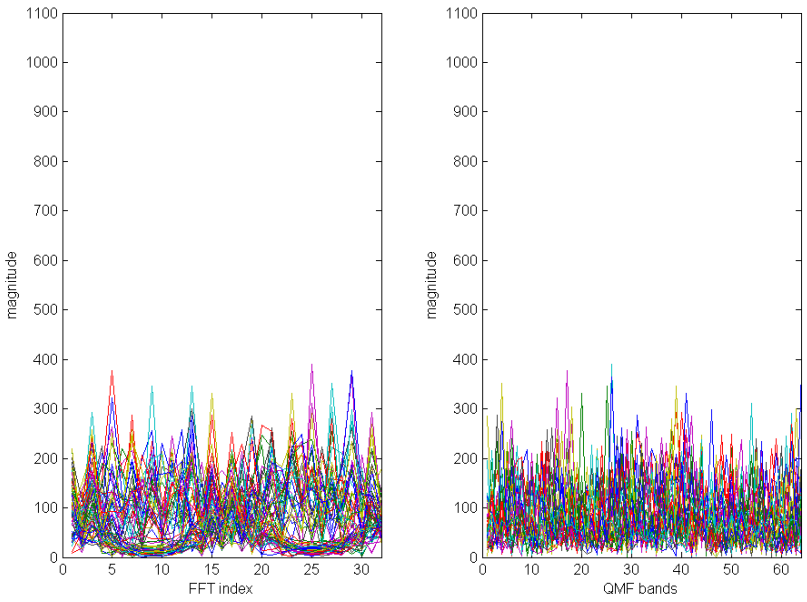
$$\sum_{k=0}^{63}\sum_{n=0}^{N-1}(x_k[n])(x_k[n])^* \tag{34}$$

where $x_k[n]$ is the residual subband signal; k is the subband index, n is the time index, and N is the frame length. The superscript (*) denotes the conjugate operation. As can been seen in Table 1, the MPS suggested coding cannot preserve the energy. This is because that the MDCT in encoder is inconsist with the IMDCT length in decoder, leading to the mismatch of the MDCT/IMDCT scale. Therefore, thanks to using the MDCT length harmonizing with the MPS residual decoder, both the passband-based and LSE-based coding methods preserve the energy of the original residual signal well.

On the other hand, the signal to noise ratio between the original residual signal and errors per frame is shown in Table 2. Note that the complexfication from real MDCT coefficients to complex QMF residue is lossy, and none of the three methods has the perfect reconstruction property. As expected, the LSE-based coding method arrives the minimum error among the three methods.

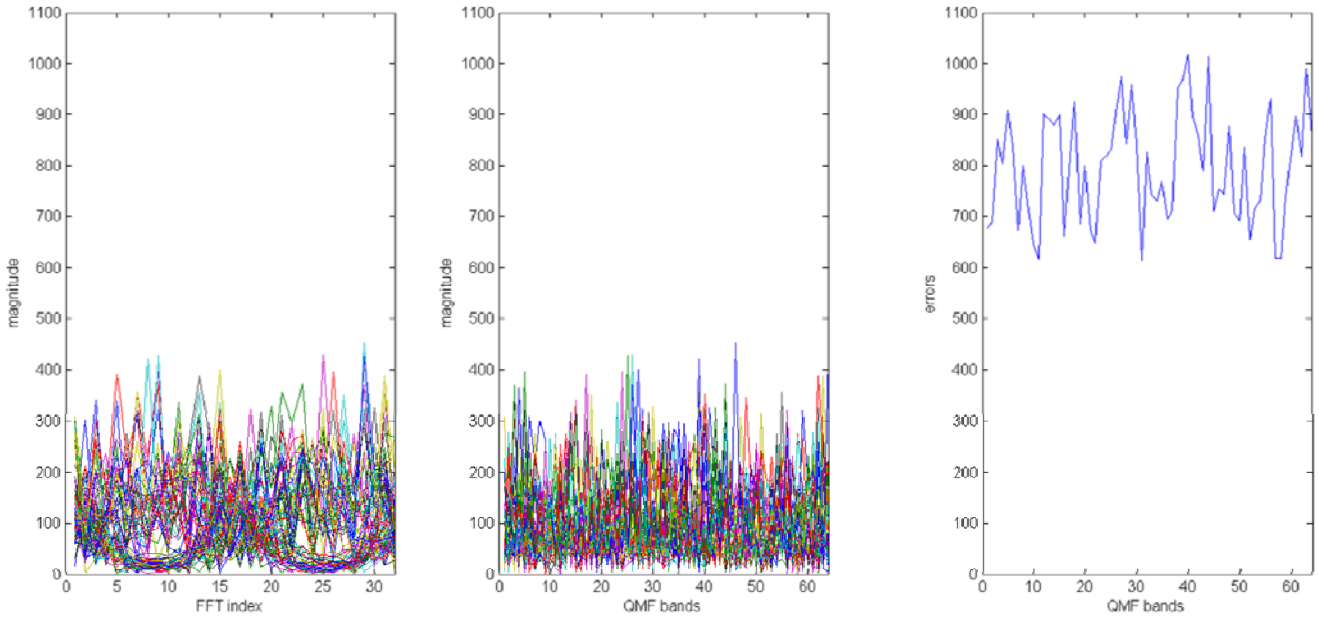
5.3 Experiment on White Noise

In the second experiment, white noise is taken as the original residual signal to check the effect. Likewise, the reconstructed result is compared with the original residual signal in DFT frequency domain.



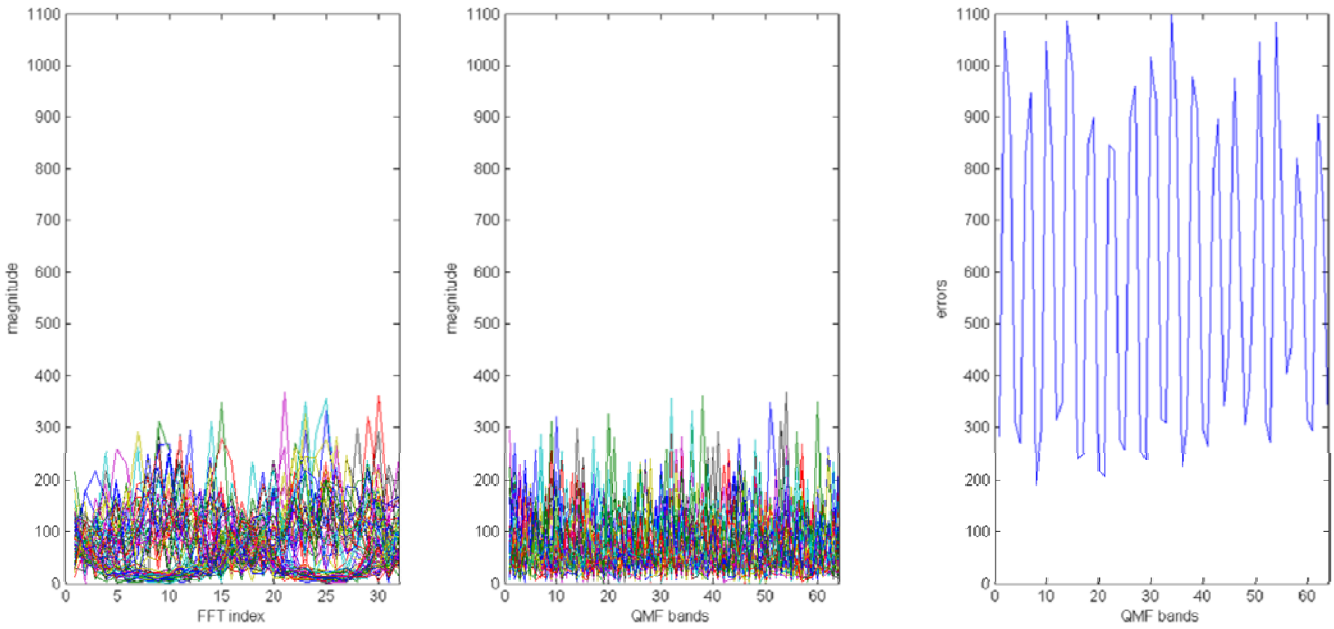
(a) **(b)**

Figure 15. DFT spectra of the original residual signal simulated by a white noise. (a) DFT (magnitude) spetrcrum in each QMF band, where the spectra of 64 subbands are overlapped. (b) DFT responses of 64subbands, where all DFT frequency lines are overlapped into one line in each QMF subband.



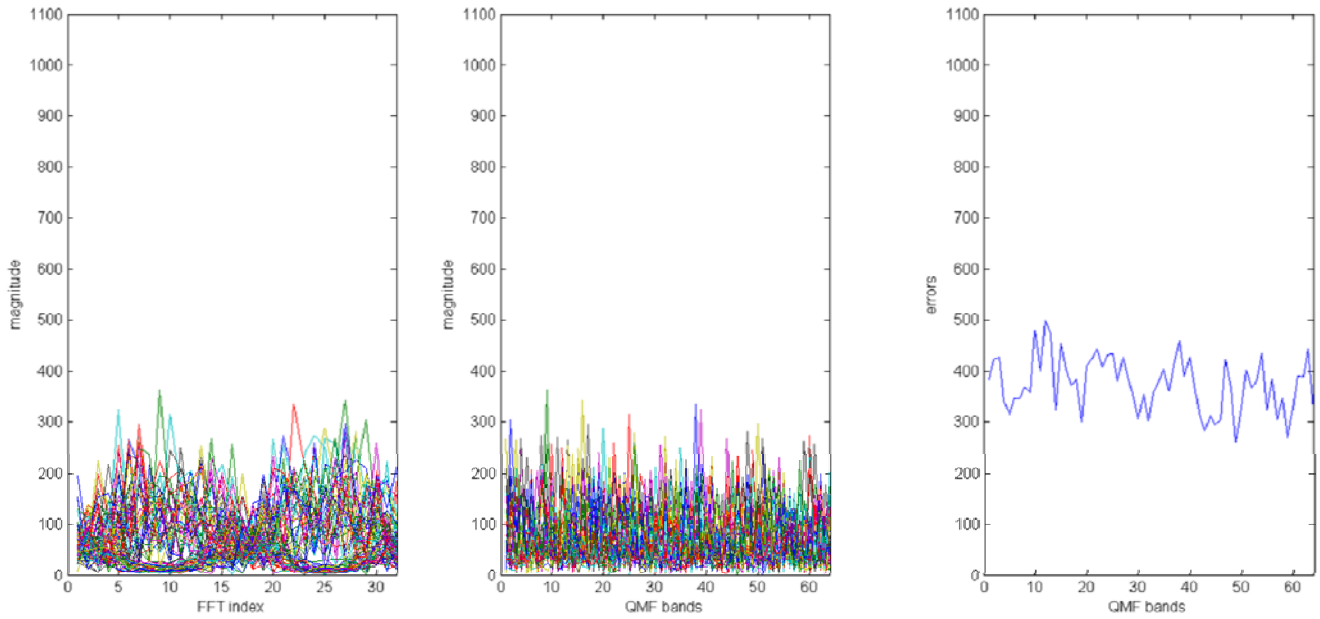
(a) **(b)** **(c)**

Figure 16. DFT spectra of the reconstructed signal using MPS suggested coding and the corresponding reconstructed error. (a) DFT (magnitude) spetrcrum in each QMF band, where the spectra of 64 subbands are overlapped. (b) DFT responses of 64subbands, where all DFT frequency lines are overlapped into one line in each QMF subband. (c) error in each QMF band.



(a) **(b)** **(c)**

Figure 17. DFT spectra of the reconstructed signal using MPS suggested coding and the corresponding reconstructed error. (a) DFT (magnitude) spetrcrum in each QMF band, where the spectra of 64 subbands are overlapped. (b) DFT responses of 64subbands, where all DFT frequency lines are overlapped into one line in each QMF subband. (c) error in each QMF band.



(a) (b) (c)

Figure 18. DFT spectra of the reconstructed signal using MPS suggested coding and the corresponding reconstructed error. (a) DFT (magnitude) spetrcrum in each QMF band, where the spectra of 64 subbands are overlapped. (b) DFT responses of 64subbands, where all DFT frequency lines are overlapped into one line in each QMF subband. (c) error in each QMF band.

Table 3. The energy ratio between the original and reconstructed residual signals

	Energy(Reconstructed Residual Signal) / Energy(Original Residual Signal)
MPS Suggested Coding	1.21
Passband-Based Coding	0.91
LSE-Based Coding	0.94

Table 4. Signal to noise ratio of original signal and errors per frame

	SNR=(10* $\log_{10}$ (Energy(original)/ Energy(error))
MPS Suggested Coding	6.776
Passband-Based Coding	10.45
LSE-Based Coding	12.22

Table 3 shows the energy ratio between the original and reconstructed residual signals. Similar to the result in Section 5.2, the reconstructed residual signal encoded by MPS suggested coding has a larger energy than that of the original one due to the mismatch of the MDCT/MDST scale in encoder/decoder side. The reconstructed energies by the passband-based coding and LSE-based coding are preserved better.

The signal to noise ratio between the original residual signals and errors per frame is shown in Table 4. The LSE-based coding method obtains the minimum error among the three methods. Thus, the LSE-based coding method is the best solution among these methods from the aspect of complexfication error minimization.

Chapter 6 Conclusion and Future Works

The thesis brought out a new design of MPS residual encoder which includes a fast QMF-to-MDCT transformation method and a least square error method for minimize the complexfication error. The original MPS residual encoder is coding the residual signal through the QMF synthesis and long time MDCT. We have propose a fast QMF-to-MDCT transformation method which can decrease the computation complexity and fix the problem of mismatch of MDCT scale. Furthermore, the losses of MPS residual decoder can be corrected by applying the LSE method on new residual encoder.

The new residual encoder can decrease the losses of complexficated procedure. But the effects on subjective test should be verified.

Reference

[1] *Information Technology MPEG Audio Technologies Part 1: MPEG Surround*, ISO/IEC 23003-1

[2] J. Roden, J. Breebaart, J. Hilpert, H. Purnhagen, E. Schuijers, J. Koppens, K. Linzmeier, A. Holzer, "A Study of the MPEG Surround quality versus bit-rate curve", *123rd AES Convention*, Oct 5-8, 2007

[3] J. Breebaart, J. Herre, C. Faller, J. Roden, F. Myburg, S. Disch, H. Purnhagen, G. Hotho, M. Neusinger, K. Kjorling, W. OOmen, "MPEG Spatial Audio Coding / MPEG Surround: Overview and Current Status", *119th AES Convention*, Oct 7-10, 2005

[4] J. Breebaart, G. Hotho, J. Koppens, E. Schuijers, W. Oomen, S. Van De Par, "Background, Concept, and Architecture for the Recent MPEG Surround Standard on Multichannel Audio Compression", *J. Audio Eng. Soc.*, Vol. 55, No. May 5, 2007

[5] L. Villemoes, J. Herrre, J. Breebarrt, G. Hotho, S. Disch, H. Purnhagen, K. Kjorling, "MPEG Surround: The Forthcoming ISO Standard for Spatial Audio Coding", *AES 28th International Conference*, Pitea, Sweden, June 30 to July 2, 2006.

[6] J. Breebaart, C.Faller, *Spatial Audio Processing, MPEG Surround and Other Application*, John Wiley & Sons, Ltd, New York, 2007

[7] J. Herre, K. Kjorling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Roden, W. Oomen, K. Linzmeier, K. Seng Chong, "MPEG Surround – The ISO/MPEG Standard for efficient and Compatible Multi-Channel Audio Coding", *112nd AES Convention*, May 5-8, 2007

[8] *Information Technology – Coding of Audiovisual objects Part3: Audio*, ISO/IEC 14496-3:2009