HIGH FREQUENCY RECONSTRUCTION FOR BAND-LIMITED AUDIO SIGNALS

Chi-Min Liu, Wen-Chieh Lee, and Han-Wen Hsu

Department of Computer Science and Information Engineering
National Chiao Tung University, Hsinchu, Taiwan, ROC
cmliu@csie.nctu.edu.tw

ABSTRACT

Current existing digital audio signals are always restricted by sampling rates and bandwidth fit for the various storages and communication bandwidth. Take for example the widely spread mp3 tracks encoded by the standard MPEG1 layer 3. The audio bandwidth in MP3 is restricted to 16 kHz due to the protocols constraints defined. This paper presents the method to reconstruct the lost high frequency components from the band-limited signals. Both the subjective and objective measures have been conducted and shown the better quality. Especially, the important objective measurement by the perceptual evaluation of audio quality system, which is the recommendation system by ITU-R Task Group 10/4 has proven a significant quality improvement.

1. INTRODUCTION

There are two main reasons losing high frequency components of audio signals. One is the reduction of sampling in audio signals. To avoid aliasing effects, a wideband signal should be processed to be band-limited to a narrowband signal when the sampling rate is not enough. Furthermore, under restriction of limited bit rate, to get the best quality of hearing, most audio compression CODEC's scarify the high frequency components of signals, and put all available bits to the low frequency component that is more important for human hearing. As shown in Figure 1, the goal of audio bandwidth extension is to reconstruct the high frequency component lost.

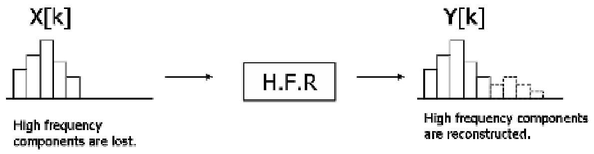

Figure 1: *Block diagram of audio bandwidth extension*

The topic is not entirely new; some attempts have been made to extrapolate a wideband signal from its narrowband frequency components [1]-[4]. However, for most of them were limited to speech, instead of a general audio signal. This paper presents a novel method to reconstruct the high frequency components for general audio signals. Also, a challenge objective of the paper is to reconstruct high frequency signals of the original signals instead of the pseudo signals. Based on the challenge objective, the measurement on the improvement has been planned through the perceptual evaluation of audio quality system [5], which is the recommendation system by ITU-R Task Group 10/4 to measure the perceptual difference of the artifact.

An advanced scheme referred to as "spectral band replication (SBR)" [9]-[12] has become the reference model of the MPEG-4 version 3 audio standard to compress high frequency contents. The SBR is different from the method presented in this paper in that it needs side information on the frequency contents extracted in encoder to help the reconstruction of the high frequency contents. From the aspects of compression, the method presented in this paper does not need additional information from either encoders or decoders. All the encoded music with limited bandwidth can be reconstructed to improve the perceptual quality.

2. RECONSTRUCTION METHOD

The method is a frequency-domain approach since we reconstruct the high frequency signals in the frequency domain. Let $X[k]$ be the spectrum signals at some time frame. The method reconstructs the high frequency signals with a linear extrapolation on the magnitude with logarithm scale. We adopt the logarithm scale due to the fitting with the magnitude absorption model [7]. On the other hand, the frequency scale will be in linear model due to the harmonic extension in linear scale. On the assumption, we reconstruct the signals from the aspects of envelope and fine detail. We try to find the envelope of the high frequency through the linear extrapolation of signals with frequencies lower than the reconstructed point, say k_c. On the detailed spectrum, we try to find the unit spectrum from the low frequency signals and then used to replicate to the high frequency fitting the envelope defined. Figure 2 illustrates the concept.

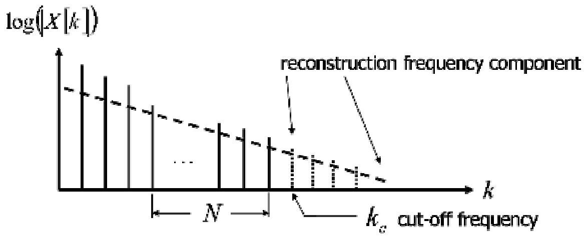

Figure 2: *Linear extrapolation on the magnitude with logarithm scale*

2.1. Least Squares Method by Linear Model

The envelope is basically evaluated by the following theorem:

Theorem 1 Given a set M consists of N frequency lines with logarithm magnitude; that is

$$M = \{ \ln(|X[k_c - N]|), \ln(|X[k_c - (N-1)]|), \ldots, \ln(|X[k_c - 1]|) \} \quad (1)$$

Assume *L:* $\ln|X[k]| = a_{opt} \cdot k + b_{opt}$ is the linear approximation with the least-square method on the *N* frequency lines. Then

$$a_{opt} = \frac{12}{(N-1)N(N+1)} \cdot \ln\left\{ \prod_{i=1}^{\frac{N-1}{2}} \left[\frac{|X[k_c - i]|}{|X[k_c - (N+1-i)]|} \right]^{\left(\frac{N+1}{2} - i\right)} \right\} \quad (2)$$

and

$$b_{opt} = \frac{\ln\left(\prod_{i=1}^{N} |X[k_c - i]| \right)}{N} - \left(k_c - \frac{N+1}{2} \right) a_{opt} \quad (3)$$

<*Proof*> We need to find b_{opt} and a_{opt} such that the summation

$$\sum_{i=1}^{N} [b + (k_c - i)a - \ln(|X[k_c - i]|)]^2 = \sum_{i=1}^{N} [b + (k_c - i)a - X'[k_c - i]]^2 = \left\| \begin{bmatrix} 1 & k_c - 1 \\ 1 & k_c - 2 \\ \vdots & \vdots \\ 1 & k_c - N \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} - \begin{bmatrix} X'[k_c - 1] \\ X'[k_c - 2] \\ \vdots \\ X'[k_c - N] \end{bmatrix} \right\|^2 \quad (4)$$

has the minimum value, where $X'[k_c - i] = \ln(|X[k_c - i]|)$ .The problem can be solved by solving a normal equation; that is,

$$\begin{bmatrix} 1 & 1 & ,\ldots, & 1 \\ k_c - 1 & k_c - 2 & ,\ldots, & k_c - N \end{bmatrix} \begin{bmatrix} 1 & k_c - 1 \\ 1 & k_c - 2 \\ \vdots & \vdots \\ 1 & k_c - N \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 1 & 1 & ,\ldots, & 1 \\ k_c - 1 & k_c - 2 & ,\ldots, & k_c - N \end{bmatrix} \begin{bmatrix} X'[k_c - 1] \\ X'[k_c - 2] \\ \vdots \\ X'[k_c - N] \end{bmatrix} \quad (5)$$

This is equivalent to solve the equation (6).

$$\begin{bmatrix} N & Nk_c - \frac{N(N+1)}{2} \\ Nk_c - \frac{N(N+1)}{2} & Nk_c^2 - N(N+1)k_c + \frac{N(N+1)(2N+1)}{6} \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} X'[k_c - i] \\ k_c \cdot \sum_{i=1}^{N} X'[k_c - i] - \sum_{i=1}^{N} i \cdot X'[k_c - i] \end{bmatrix} \quad (6)$$

By Gaussian-Jordan elimination method [13], (6) can be reduced to

$$\begin{bmatrix} 1 & k_c - \frac{N+1}{2} \\ 0 & \frac{(N-1)N(N+1)}{12} \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=1}^{N} X'[k_c - i]}{N} \\ \sum_{i=1}^{N} \left(\frac{N+1}{2} - i \right) X'[k_c - i] \end{bmatrix} \quad (7)$$

The optimum solution b_{opt} and a_{opt} can be found by solving (7). ■

The complexity to calculate a_{opt} is $O(N^2)$, where *N* is the number of frequency lines to predict the envelope. In next subsection, a fast computing method is proposed.

2.2. Fast Computing a_{opt}

Assume *N* is positive integer and *N*>1. We denote Y_i and W_i in (2) according to

$$Y_i = X[k_c - i]; \quad for\ i = 1, 2, \ldots, \frac{N-1}{2}. \quad (8)$$

and

$$W_i = X[k_c - (N+1-i)]; \quad for\ i = 1, 2, \ldots, \frac{N-1}{2}. \quad (9)$$

Substituting (8) and (9) to (2) yields

$$a_{opt} = \frac{12}{(N-1)N(N+1)} \cdot \left\{ \ln\left[\prod_{i=1}^{\frac{N-1}{2}} |Y_i|^{\left(\frac{N+1}{2} - i\right)} \right] - \ln\left[\prod_{i=1}^{\frac{N-1}{2}} |W_i|^{\left(\frac{N+1}{2} - i\right)} \right] \right\}. \quad (10)$$

That is

$$a_{opt} = \frac{12}{(N-1)N(N+1)} \cdot \left\{ \ln\left\| \prod_{i=1}^{\frac{N-1}{2}} Y_i^{\left(\frac{N+1}{2} - i\right)} \right\| - \ln\left\| \prod_{i=1}^{\frac{N-1}{2}} W_i^{\left(\frac{N+1}{2} - i\right)} \right\| \right\} \quad (11)$$

.

Furthermore, we define the product of a series of Y_j as Z_i, that is

$$Z_i = \prod_{j=1}^{i} Y_j; \quad for\ i = 1, 2, \ldots, \frac{N-1}{2}. \quad (12)$$

Taking a recursive way to calculate Z_i leads to

$$Z_i = Z_{i-1} \cdot Y_i; \quad for\ i = 1, 2, \ldots, \frac{N-1}{2}. \quad (13)$$

$Z_0 = 1$. Similarly, we define the product of a series of W_j as V_i,

$$V_i = \prod_{j=1}^{i} W_j; \quad for\ i = 1, 2, \ldots, \frac{N-1}{2}. \quad (14)$$

Taking a recursive way to calculate V_i leads to

$$V_i = V_{i-1} \cdot W_i; \quad for\ i = 1, 2, \ldots, \frac{N-1}{2}. \quad (15)$$

$V_0 = 1$. The recursive forms in (13) and (15) can be derived as

$$\prod_{i=1}^{\frac{N-1}{2}} Y_i^{\left(\frac{N-1}{2} - i\right)} = \prod_{i=1}^{\frac{N-1}{2}} Z_i \quad . \quad (16)$$

and

$$\prod_{i=1}^{\frac{N-1}{2}} W_i^{\left(\frac{N-1}{2} - i\right)} = \prod_{i=1}^{\frac{N-1}{2}} V_i \quad . \quad (17)$$

Substituting (16) and (17) to (11) yields

$$\mathrm{a}_{\mathrm{opt}} = \frac{12}{(N-1)N(N+1)} \cdot \ln\left(\left| \frac{\prod_{i=1}^{\frac{N-1}{2}} Z_i}{\prod_{i=1}^{\frac{N-1}{2}} V_i} \right| \right) . \quad (18)$$

From (18), the complexity of computing the values of Z_i needs $\frac{N-3}{2}$ multiplications. To compute the product of Z_i, it also needs $\frac{N-3}{2}$ multiplications. Hence, computing $\prod_{i=1}^{\frac{N-1}{2}} Z_i$ totally needs

$N-3$ multiplications. Similarly, to compute the value of $\prod_{i=1}^{\frac{N-1}{2}} V_i$ needs $N-3$ multiplications. Using (18) to calculate a_{opt} needs totally $2N-6$ multiplications. Thus, computing (18) leads to a linear complexity and needs only one logarithm, division and absolute operation, respectively. On the other hand, computing b_{opt} needs a constant complexity due to

$$Z_{\frac{N-1}{2}} \cdot V_{\frac{N-1}{2}} \cdot X\left(k_c - \frac{N+1}{2}\right) = \prod_{i=1}^{N} X\left(k_c - i\right) \tag{19}$$

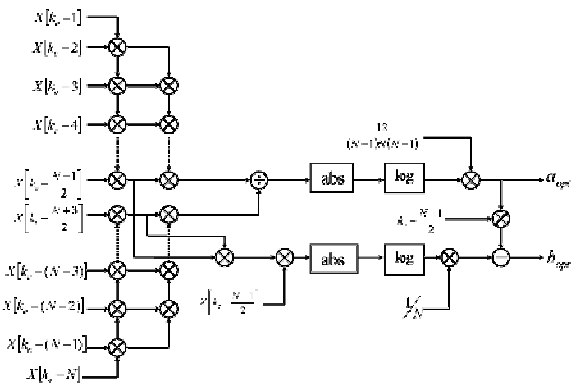

Figure 3: *Signal flow diagram of the fast computing method.*

2.3. Construction on Detail Spectrum

The detail spectrum is reconstructed by taking and duplicating a segment of low frequency components from $X[k_c - 1]$ to $X[k_c - U]$. For any nonnegative integer β, $X[k_c + \beta]$ is defined as

$$X[k_c + \beta] = \frac{X[k_c + (\beta(\mathrm{mod}U)) - U]}{b_{opt} \cdot \exp^{a_{opt}(k_c + (\beta(\mathrm{mod}U)) - U)}} \cdot b_{opt} \cdot \exp^{a_{opt}(k_c + \beta)} \tag{20}$$

That is,

$$X[k_c + \beta] = X[k_c + (\beta(\mathrm{mod}\,U)) - U] \cdot \exp^{a_{opt}(\beta - (\beta(\mathrm{mod}\,U)) + U)} \tag{21}$$

Representing (21) as the recursive equation leads to

$$X[k_c + \beta] = X[k_c + \beta - U] \cdot \exp^{a_{opt} \cdot U} \quad \forall \mathrm{int}\ \beta \geq 0 \tag{22}$$

To sum up, (18) and (22) constitute the frequency extension technique. There are three calibrations required for the algorithm. The first calibration is on the dithering of the zero magnitude to avoid the undefined problem of the logarithm of zero, the zero magnitudes of frequency lines are replaced with a small positive real number ε. ε needs to adapt with the audio frames. A too large or small ε will affect the evaluation of envelope slope. This paper calculates the average magnitude of the N frequency lines and multiplies the value by 0.001 to have ε.

The second calibration is on the envelope parameter a_{opt}. a_{opt} should be constrained to be non-positive. Hence, we set those positive a_{opt} to -0.01 to avoid the increasing envelope.

The third calibration is on the selection of the reconstruction basis. The method extends to high frequency by duplicate the low frequency contents recursively to high frequency contents based on a reconstruction unit. Once the content of the reconstruction unit is abnormal, the extension of high frequency components from low frequency part may not be applicable. Figure 4 illustrates the phenomenon. In Figure 4, there is a huge prominence that is exactly our reconstructed unit. When the reconstruction unit is used to extend for the high frequency signals, the resultant spectrum is illustrated in Figure 5. A criterion should be used to skip the reconstruction method when there is no qualified reconstruction units found.

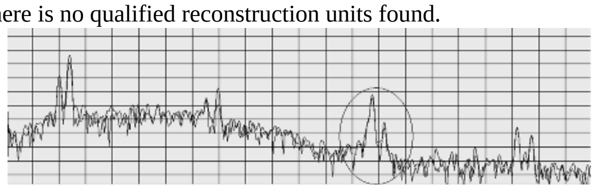

Figure 4: *Spectrum of the original audio signal.*

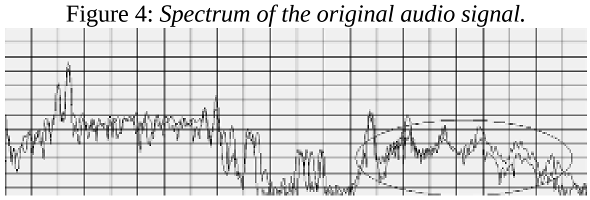

Figure 5: *Spectrum of the compressed audio signal with bandwidth extension.*

A simple way on the detection mechanism is to monitor the ratio of the summation of the frequency magnitudes on the reconstructed unit and the relative summation of estimated pseudo magnitudes.

$$Detecion\ Ratio\ \varphi = \frac{\sum_{i=1}^{U} X_P[k_c - i]}{\sum_{i=1}^{U} |X[k_c - i]|} \tag{23}$$

where

$$\sum_{i=1}^{U} X_P[k_c - i] = \sum_{i=1}^{U} \exp^{b_{opt} + a_{opt}(k_c - i)}, \tag{24}$$

If the ratio is lower than a threshold, the reconstruction method is skipped. Substituting (24) into (23) leads to

$$\varphi = \begin{cases} \dfrac{\exp^{b_{opt} + a_{opt}k_c} \dfrac{\left(1 - \exp^{-a_{opt}U}\right)}{\exp^{a_{opt}} - 1}}{\sum_{i=1}^{U} |X[k_c - i]|} & if\ a_{opt} \neq 0 \\ \dfrac{U \exp^{b_{opt}}}{\sum_{i=1}^{U} |X[k_c - i]|} & if\ a_{opt} = 0 \end{cases} \tag{25}$$

The algorithm can be summarized as follows:

Input data: The basic sources to extend bandwidth are described below.

(a) $M : \{ X[k_c - N], X[k_c - (N-1)], \ldots, X[k_c - 1] \}$

(b) k_c: cut-off frequency

(c) k_e : reconstruction-ended frequency

(d) N : the size of the set M

(e) U : reconstructed unit length

There are total nine steps of the algorithm expressed as follow:

Step1: Replace $X[k_c - i]$ of zero value with a small real number ε , for i=1 to N

Step2: Calculate Z_i and V_i recursively

(a) Let $Z_0 = 1$ and $V_0 = 1$

(b) Let $Z_i = Z_{i-1} \cdot X[k_c - i]$ and $V_i = V_{i-1} \cdot X[k_c - (N+1-i)]$ for *i=1* to *N*.

Step3: Calculate $\prod_{i=1}^{\frac{N-1}{2}} Z_i$ and $\prod_{i=1}^{\frac{N-1}{2}} V_i$ respectively.

Step4: Calculate a_{opt} according to (18)

Step5: If a_{opt} >0, let a_{opt} =0.

Step6: Calculate b_{opt} according to (3).

Step7: Calculate Unit Decay Ratio ρ , $\rho = \exp(a_{opt} \cdot U)$

Step8: Calculate Detection Ratio φ

If φ < threshold, the algorithm stops. Otherwise, go to Step 9 .

Step9: Duplicate the spectrums recursively

Make $X[k] = \rho \cdot X[k-U]$ for $k = k_c$ to k_e .

The block diagram and the associated flow chart of the algorithm are illustrated by Figure 6 and Figure 7 respectively.

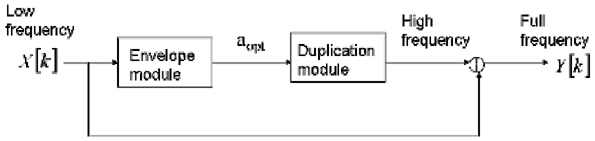

Figure 6: *A block diagram of audio bandwidth extension.*

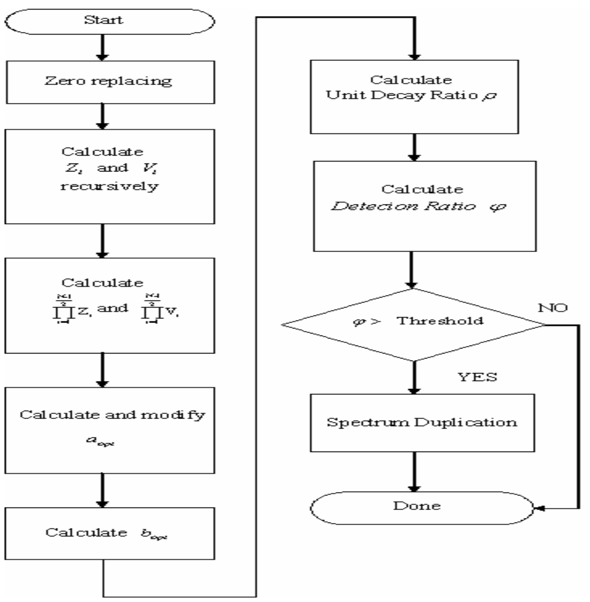

Figure 7: *The flow chart of audio bandwidth extension.*

3. EXPERIMENTS

For the high frequency reconstruction, it is difficult to measure the improvement on perceptual quality. This paper verifies the quality improvement by comparing the reconstructed audio with the original CD quality audio. The perceptual quality is measured through the PEAQ (perceptual evaluation of audio quality) system [5]. 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 ideally range from 0 to -4, where 0 corresponds to an imperceptible impairment and -4 to an 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 system. The mp3 tracks are prepared for bit rates at 128 kbps and 96kbps. The music tracks include the 50 test tracks in [5] and other critical music balancing on the percussion, string, wind instruments, and human vocal was prepared. Also, the mp3 encoder used to prepare the music tracks are the Lame version 3.88 [6], which can have a better quality than other commercial mp3 encoders. The mp3, due to the protocol defined, has always scarified the signal quality above 16k. Also, as illustrated in Figure 11, the algorithm illustrated in Section 2 can be directly implemented in the spectrum lines in the reconstruction of mp3 decoder to save the complexity.

The experiments have been conducted also for various cut-off frequencies including 16k, 14k and 12k Hz. In general, during reconstruction, from the cut-off frequency to 17k Hz the reconstructed spectrum envelope is followed by the prediction envelope of the algorithm. Above the range, the slope of envelope is modified to zero to adapt with the behavior of the higher frequency spectrum that usually decays slowly.

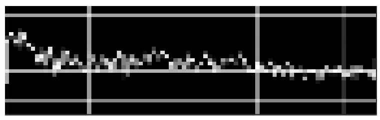

Figure 8: *The spectrum of the original audio signal.*

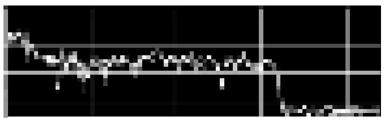

Figure 9: *The spectrum of the compression audio signal by Lame.*

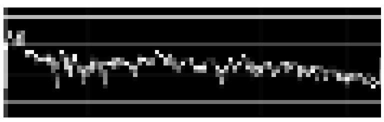

Figure 10: *The spectrum of the audio signal with bandwidth extension.*

In this paper, for bit rate 128 kbps, the reconstruction unit length is chosen as 1 or 2 kHz and for 96 kbps, the length is chosen as 4 or 5 kHz. This is because at 96 kbps, the long length can lead to a better robustness when the music is not compressed well.

Figure 12, 13, 15, and 17 have the average ODG, minimum ODG, and maximum ODG among the fifty tracks under different bit rates as well as different cut-off frequencies. Figure12 and 15 are the mp3 tracks compressed with M/S coding turned off while Figure 14 and 17 M/S coding turned on. In those figures, the top bar and the down bar represents the minimum ODG and the maximum ODG respectively and the middle square represents average ODG. Furthermore, in each figure for each pair of statistics lines in order from left to right, the first shows the ODG comparing the mp3 tracks with the original CD and the second is the ODG generated from the high frequency reconstruction presented in this paper. On the other hand, Figure12, 14, 16, and 18 illustrate the average gains from the fifty tracks. The gain is defined as the difference of the ODG of a pair. Similarly, the top bar and the down bar represent the minimum and maximum gains respectively and the middle square represents average gain.

From the test data of the fifty tracks, when the cut-off frequency is 16k and bit rate is 128k, we found that no track losses the quality by more than 0.08 in ODG but can gain improvement up to 1.21. The result indicates that the technique can have almost no risk in improving the quality in the most widely adopted compression case at present. Also, the subjective test indicates the tracks after high frequency reconstruction is “brighter” than the original mp3 tracks. In general, as the cut-off frequency decreases, the prediction is more difficult. Thus, the gain also decreases. Nevertheless, the average gains in the different cases are still more than 0.19. Especially, at the bit rate 96k, the method offers the average gain even up to 0.54.

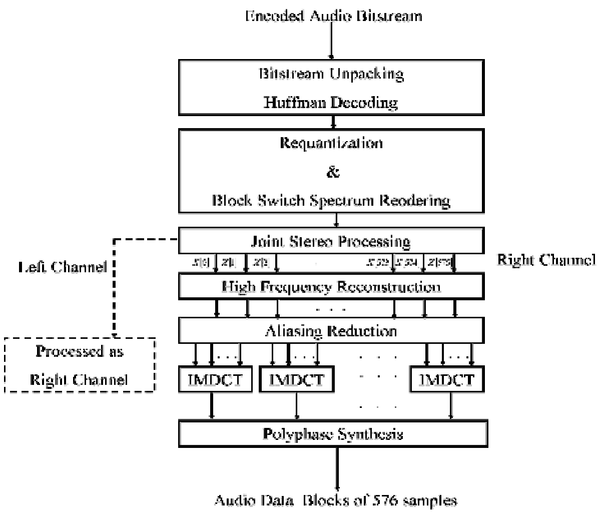

Figure 11: *The diagram of High Frequency Reconstruction incorporated into MP3 decoder.*

4. CONCLUSION

This paper has presented a method to reconstruct the high frequency signals to have better sound quality. The reconstruction consists of the envelope reconstruction, fine-detail duplication, and the algorithm calibration. Various mp3 tracks with different bit rates and cut-off frequencies have been conducted and showed the better performance compared to the original music.

This paper develops not only the algorithm but also the fast computing method for the envelope, fine-detail, and the calibration techniques. Through both the subjective and objective measure, the method is verified to be able to improve the perceptive quality of band-limited audio signals. Especially, the objective measurement by the perceptual evaluation of audio quality system, which is the recommendation system by ITU-R Task Group 10/4 has proven a significant quality improvement.

5. REFERENCE

[1] C. Avendaro, et. al., "Beyond Nyquist: towards to recovery of broad-bandwidth speech from Narrow-Bandwidth Speech", in Proc. Eurospeech, (Madrid), pp. 165-168, 1995.

[2] Y.M. Cheng, et. al., "Statistical recovery of wideband speech from narrowband speech," IEEE Trans. Speech Audio Process., vol. 2, pp. 544-548, 1994.

[3] J. Epps et. al., "A new technique for wideband enhancement of coded narrowband speech," Proc. IEEE Speech Coding Workshop (Porvoo Finland), 1999, pp. 174-176.

[4] J. Valin at. al., "Bandwidth extension of narrowband speech for low bit-rate wideband coding,” IEEE Speech Coding Workshop, 17-20 Sep. 2000, pp. 130-132.

[5] ITU Radiocommunication Study Group 6, “DRAFT REVISION TO RECOMMENDDATION ITU-R BS.1387 - Method for objective measurements of perceived audio quality”.

[6] Lame, website http://www.mp3dev.org/mp3

[7] Sabine W. C., Collected Papers on Acoustics (originally 1921), Peninsula Publishing, Los Altos, CA, USA, 1993.

[8] Erik Larsen, et. al., "Efficient high-frequency bandwidth extension of music and speech" In 112th AES Convention, Munich, May 10-May 13, 2002.

[9] ISO/IEC, “WD Text for Backward Compatible Bandwidth Extension for General Audio Coding”, ISO/IEC JTC1/SC29/WG11, MPEG2002/N4611 March 2002.

[10] Per Ekstrand et. al., "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm." In 112th AES Convention, Munich, May 10-May 13, 2002.

[11] Martin Dietz et. al., "Spectral Band Replication, a novel approach in audio coding." In 112th AES Convention, Munich, May 10 - May 13, 2002.

[12] Jeongil Seo et. al., "A Simple Method for Reproducing High Frequency Components at Low-Bit Rate Audio Coding" In 113th AES Convention, Los Angeles, October 5–8, 2002.

[13] Steven J. Leon, “Linear Algebra with applications”, Macmillan, 1990.

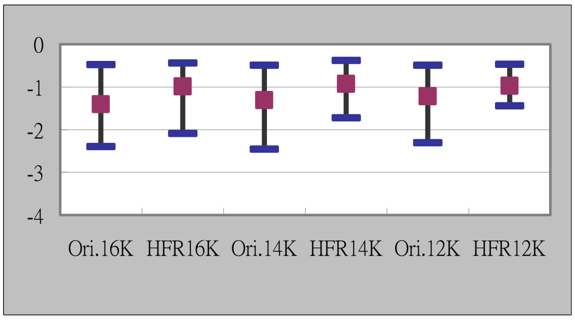

Figure 12: *The ODG range of the mp3 and H.F.R audio with different cut-off frequencies under 128k bit rate.*

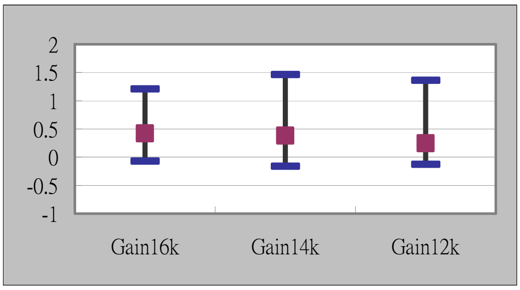

Figure 13: *The gain of High Frequency Reconstruction corresponding to Figure12.*

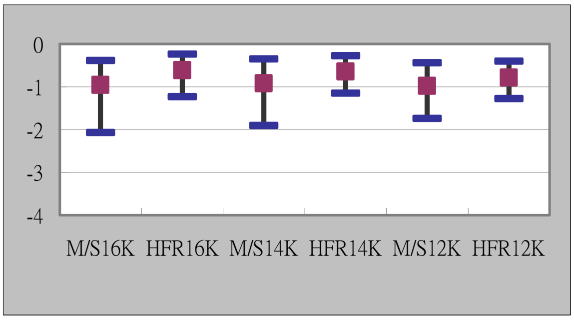

Figure 14: *The ODG range of the mp3 with M/S coding and H.F.R audio with different cut-off frequencies under 128k bit rate.*

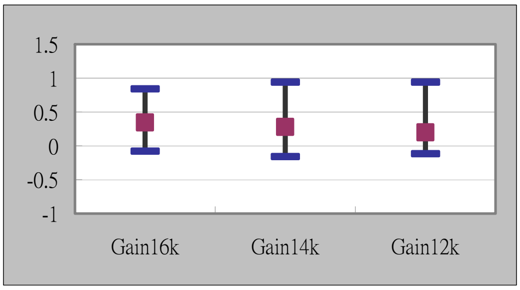

Figure 15: *The gain of High Frequency Reconstruction corresponding to Figure14.*

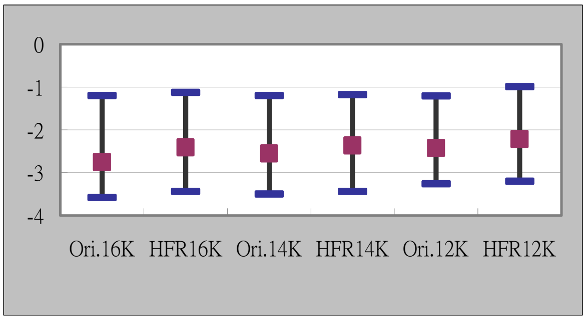

Figure 16: *The ODG range of the mp3 and H.F.R audio with different cut-off frequencies under 96k bit rate.*

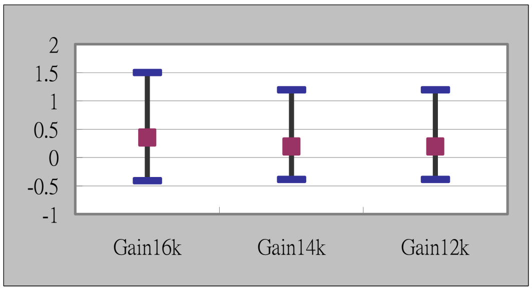

Figure 17: *The gain of High Frequency Reconstruction corresponding to Figure16..*

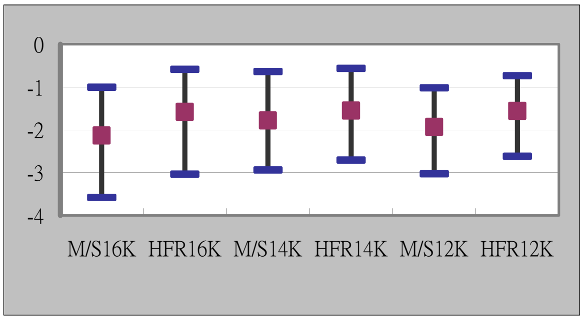

Figure 18: *The ODG range of the mp3 with M/S coding and H.F.R audio with different cut-off frequencies under 96k bit rate.*

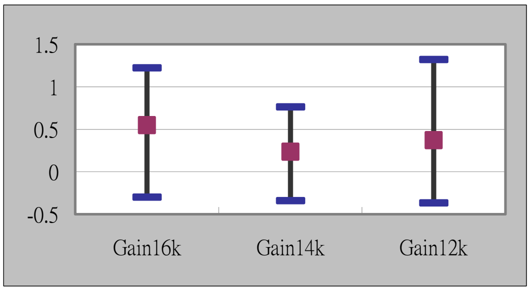

Figure 19: *The gain of High Frequency Reconstruction corresponding to Figure18..*