Chapter 5 Experiments

This thesis verifies the perceptual quality improvement by comparing the patched audio with the original CD quality audio. The perceptual quality is measured through the PEAQ (perceptual evaluation of audio quality) system [21]. 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.

5.1 Verification of Quality Enhancement by Audio Patch Method based on MPEG Test Tracks

Both the MP3 and AAC tracks are prepared for bit rates at 128 kbps and 96kbps and for sample rate at 44.1 kHz. The twelve test tracks recommended by MPEG, as shown in Table1, include the critical music balancing on the percussion, string, wind instruments, and human vocal. Also, the MP3 encoder and the AAC encoder used to prepare the music tracks are the Lame version 3.88 [25] and QuickTime version 6.3 [26], respectively. They can have a better quality than other commercial MP3 and AAC encoders. The MP3, due to the protocol defined, has always scarified the signal quality above 16k. QuickTime also scarifies the signal quality above 16k for AAC tracks. As illustrated in Figure 33 and Figure 34, the algorithms illustrated in Chapter 2 and Chapter 3 can be directly implemented on the spectrum lines in the reconstruction of MP3 and AAC decoders.

Table 1: The twelve test tracks recommended by MPEG.

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.

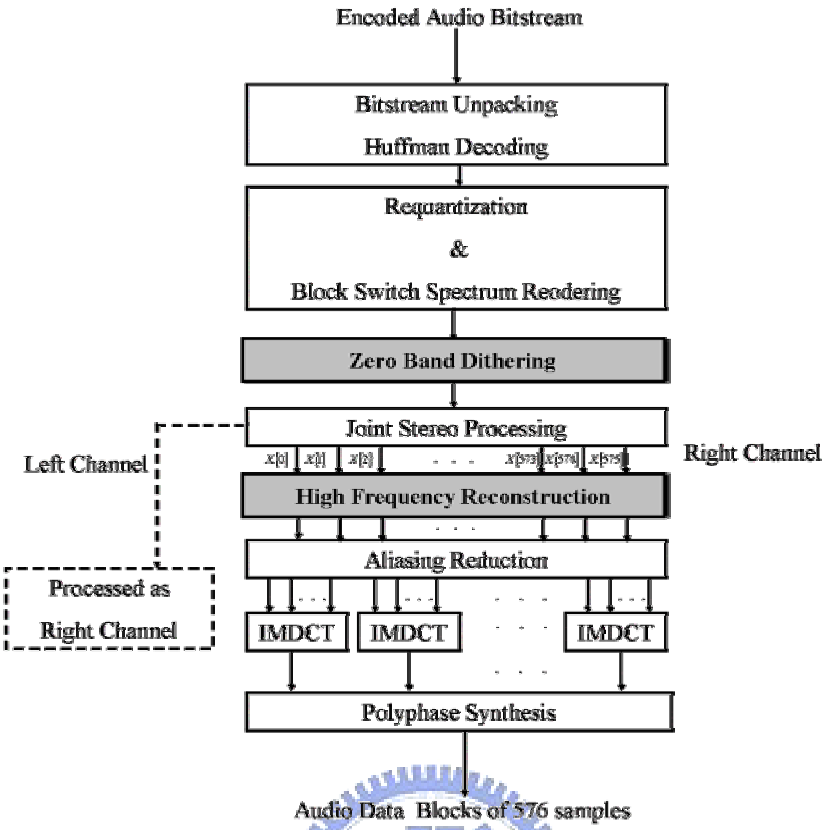


Figure 33: The diagram of Audio Patch Method incorporated into MP3 decoder.

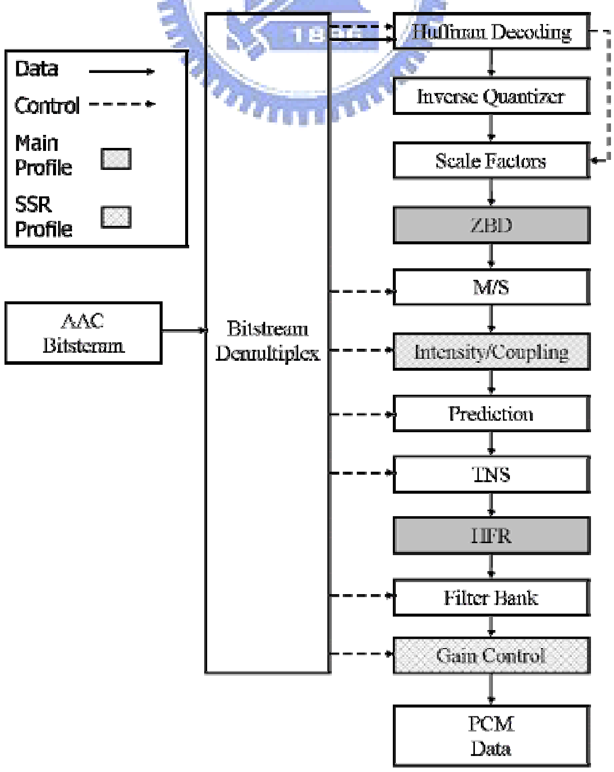


Figure 34: The diagram of Audio Patch Method incorporated into AAC decoder.

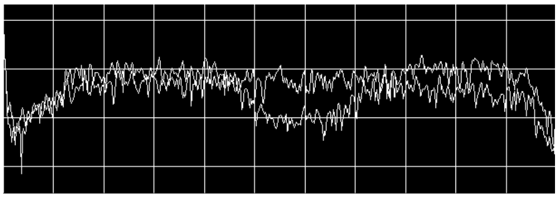


Figure 35: The spectrum of the original audio signal.

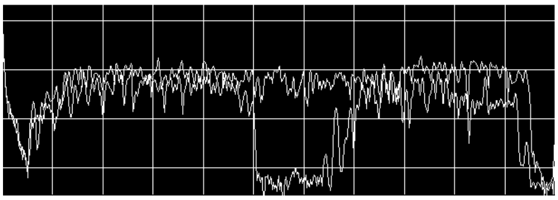


Figure 36: The spectrum of the compression audio signal with two zero bands.

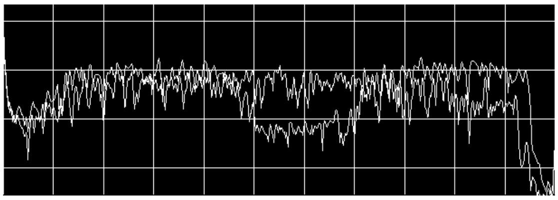


Figure 37: The spectrum of the audio signal with zero band dithering.

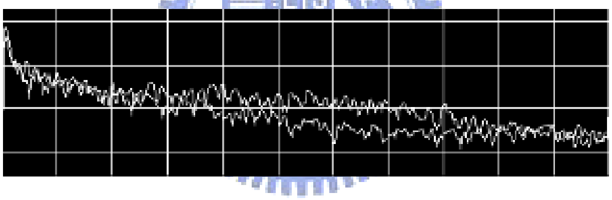


Figure 38: The spectrum of the original audio signal.

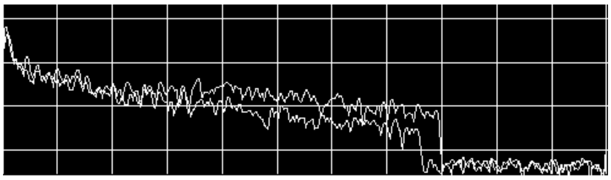


Figure 39: The spectrum of the compression audio signal with narrow bandwidth.

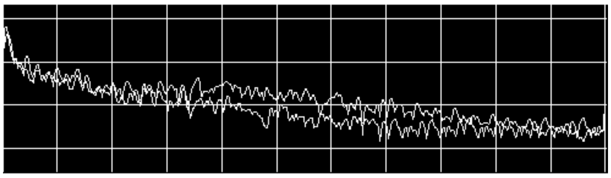


Figure 40: The spectrum of the audio signal with high frequency reconstruction.

Figures 41, 43, 46, and 48 illustrate the ODGs for the twelve tracks under different decoding processing as well as different bit rates. In those figures, every four bars combined as a data set for a track and they represents the ODG of the original decoded music, the decoded music with zero band dithering, the decoded music with high frequency reconstruction and the decoded music with overall audio patch method that includes zero band dithering and high frequency reconstruction, respectively. On the other hand, Figure 42, 44, 47, and 49 illustrate the gains from the twelve tracks. The gain is defined as the difference of the ODG of a pair. Figure 45, 50 have the average gain, minimum gain, and maximum gain among the twelve tracks under different decoding processing as well as different bit rates. In those figures, the top bar represents the minimum ODG gain, the down cross represents the maximum ODG gain and the middle square represents average ODG gain. On the other hand, Figure 51 illustrates the ODG range comparison of the AAC tracks, the MP3 tracks, ZBD audio, HFR audio and APM audio under 128k and 96k bit rate. The order of the statistics lines follow as the descendant order of the average gains.

From the test data of the twelve tracks, we found that no MP3 track losses the quality in ODG from the zero band dithering, the high frequency reconstruction or the overall audio patch method, and can gain improvement even up to 0.78 at 128 k bit rate and 1.45 at 96k bit rate. Similarly, no AAC track losses the quality by more than 0.07 but can gain improvement up to 0.41 at 128 k bit rate and 1.14 at 96k bit rate. The result indicates that the audio patch techniques can have almost no risk in improving the quality in the most widely adopted compression case at present. Especially, the effect of zero band dithering is also confirmed. As shown in Figure 35-36, the two zero bands in Figure 36, one is located at the left side and the second is located at the middle of the compression spectrum, has been patched well. For MP3 and AAC, zero band dithering can only gain improvement in average 0.08 and 0.06 at 128k bit rate respectively. This is because the compressed tracks under high bit rate have high quality and the opportunity to run the process of zero band dithering decreases largely. However, it even can gain improvement up to 0.24 and 0.27 at 128k bit rate respectively. Other the hand, for lower 96k bit rate, both the average gains is more than 0.19. Especially, the dithering method offers the gain even up to 0.4 at 96k bit rate.

The subjective test also indicates the tracks after the audio patch are “brighter” than the original tracks and the artifact “fishy noise”, especially for the track “velvet”, is eased effectively. On the other hand, from Figure 51, it shows the objective quality of Lame 3.88 at 128k bit rate with the audio patch method approaches to QuickTime 6.3 at 128k bit rate, the objective quality of QuickTime 6.3 at 96k bit rate with the audio patch method is better than Lame 3.88 at 128k bit rate, and the objective quality of Lame 3.88 at 96k bit rate with the audio patch method is better than QuickTime 6.3 at 96k bit rate. In other words, the audio patch technique can enhance the compression audio quality.

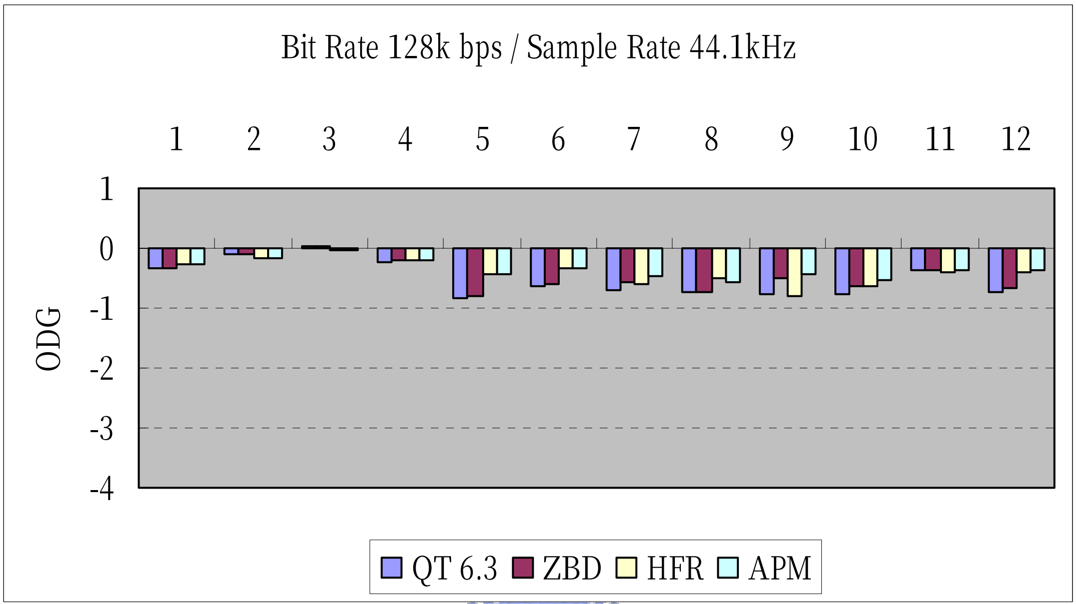


Figure 41: The ODG of the QT 6.3 tracks, ZBD audio, HFR audio and APM audio under 128k bit rate.

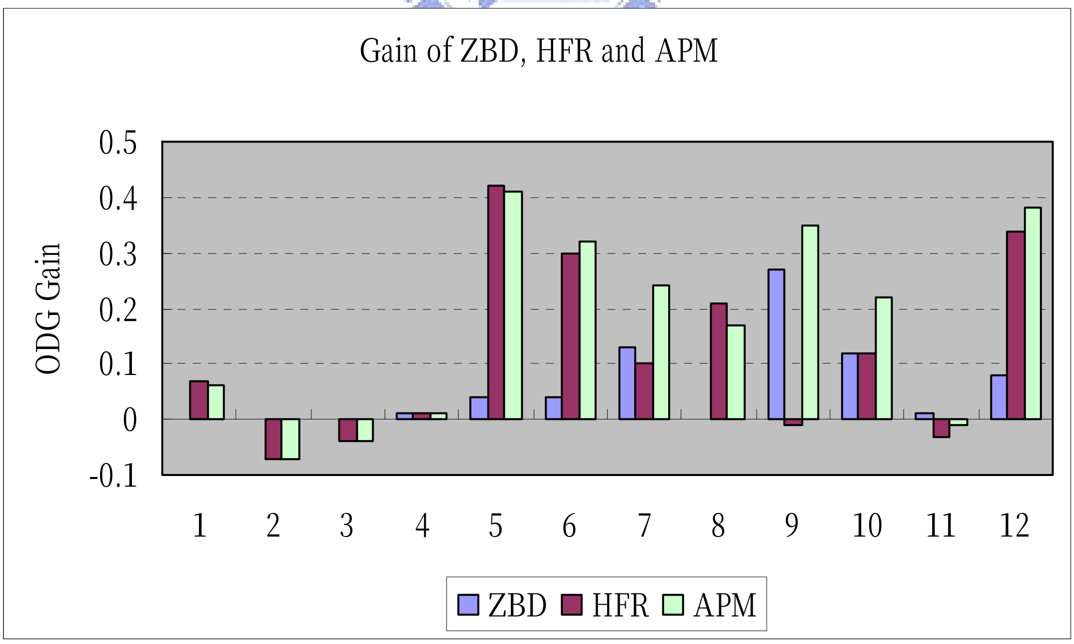


Figure 42: The gain of Zero Band Dithering , High Frequency Reconstruction and Audio Patch Method corresponding to Figure41.

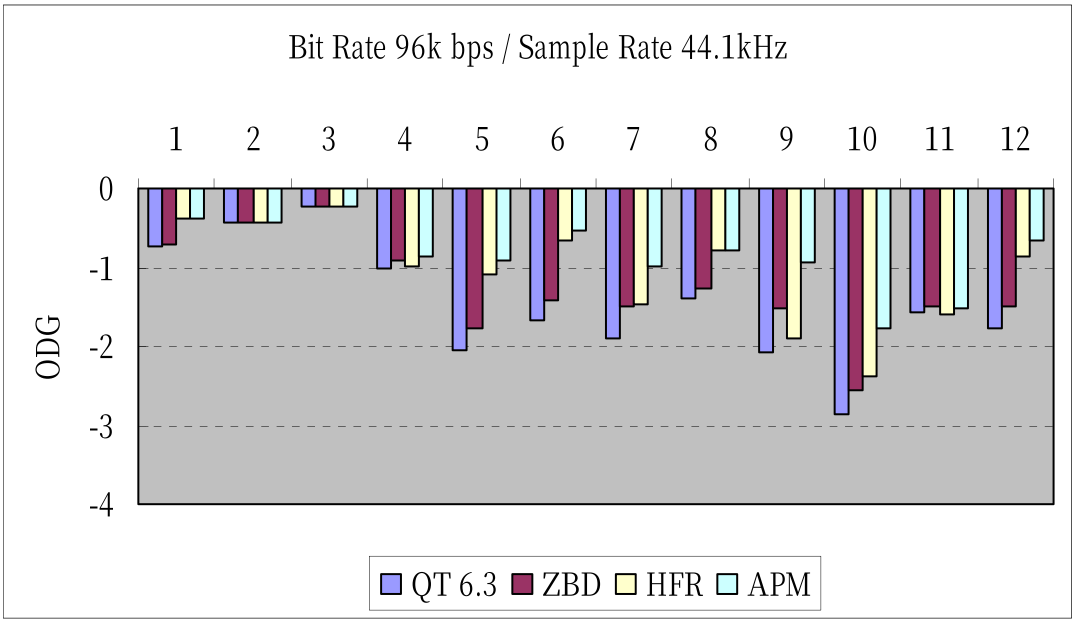


Figure 43: The ODG of the QT 6.3 tracks, ZBD audio, HFR audio and APM audio under 96k bit rate.

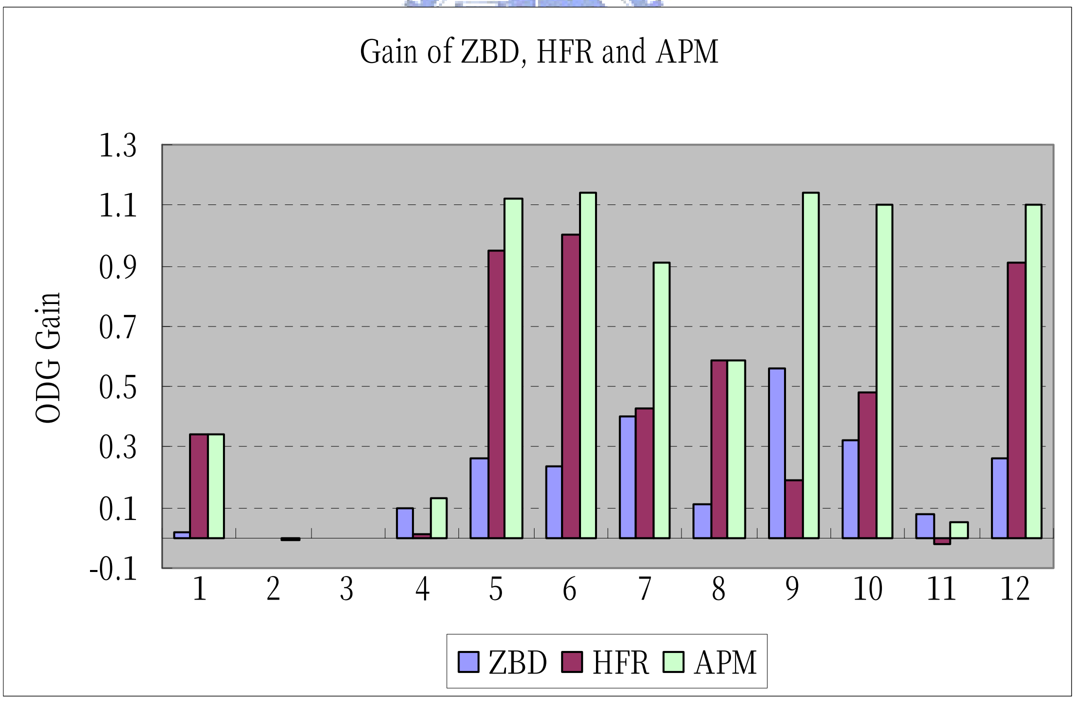


Figure 44: The gain of Zero Band Dithering, High Frequency Reconstruction and Audio Patch Method corresponding to Figure43.

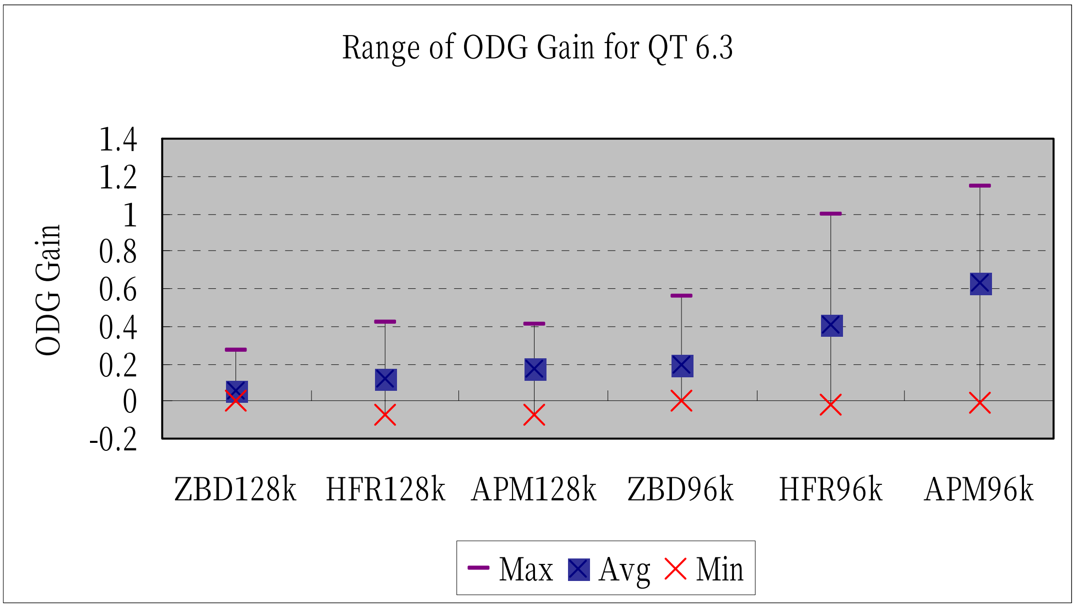


Figure 45: The range of the gain of ZBD, HFR and APM corresponding to QuickTime 6.3 at 128k and 96k bit rate.

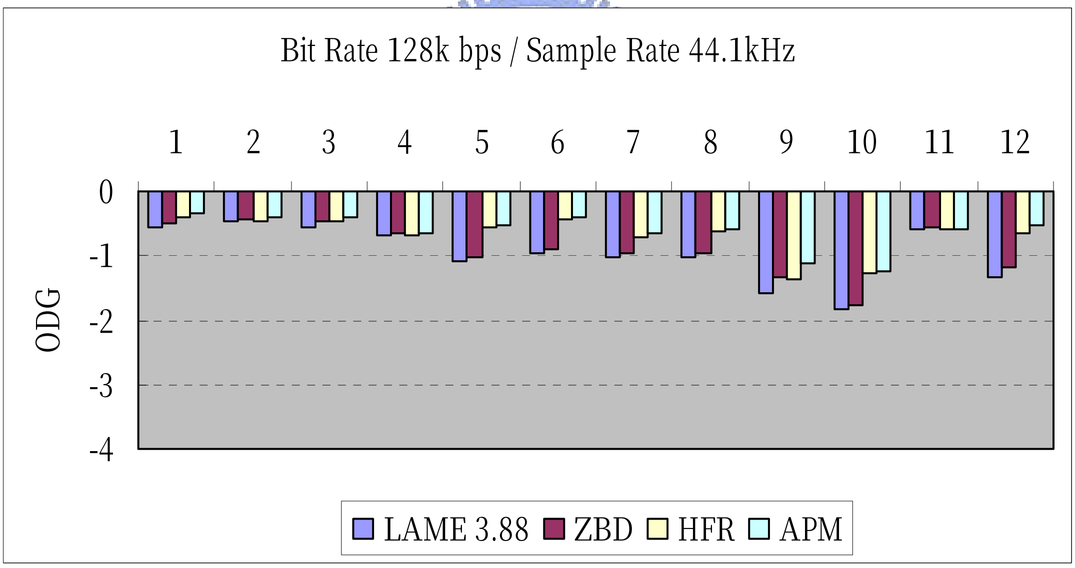


Figure 46: The ODG of the LAME 3.88 tracks, ZBD audio, HFR audio and APM audio under 128k bit rate.

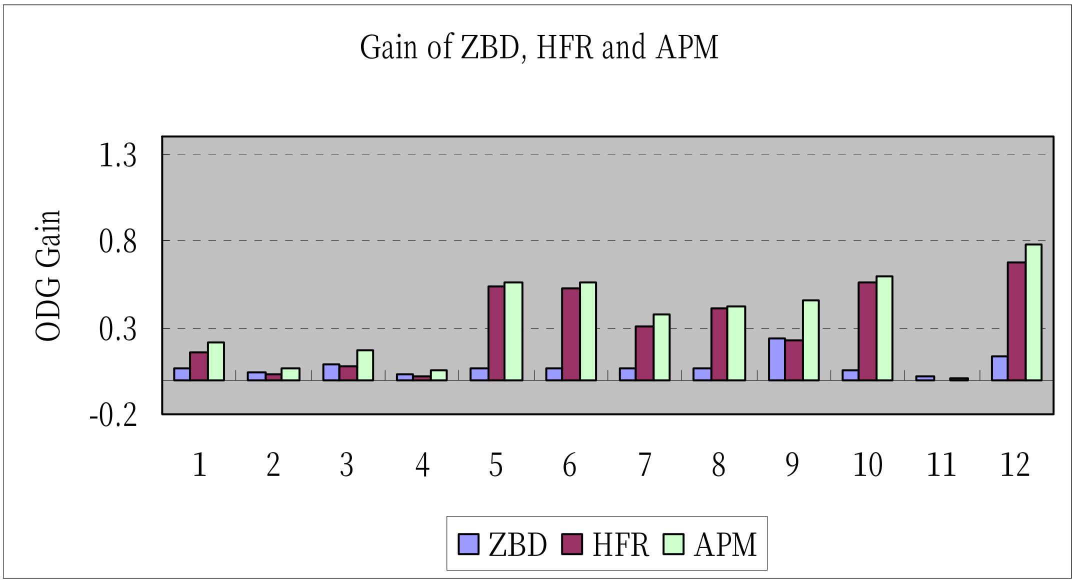


Figure 47: The gain of Zero Band Dithering, High Frequency Reconstruction and Audio Patch Method corresponding to Figure46.

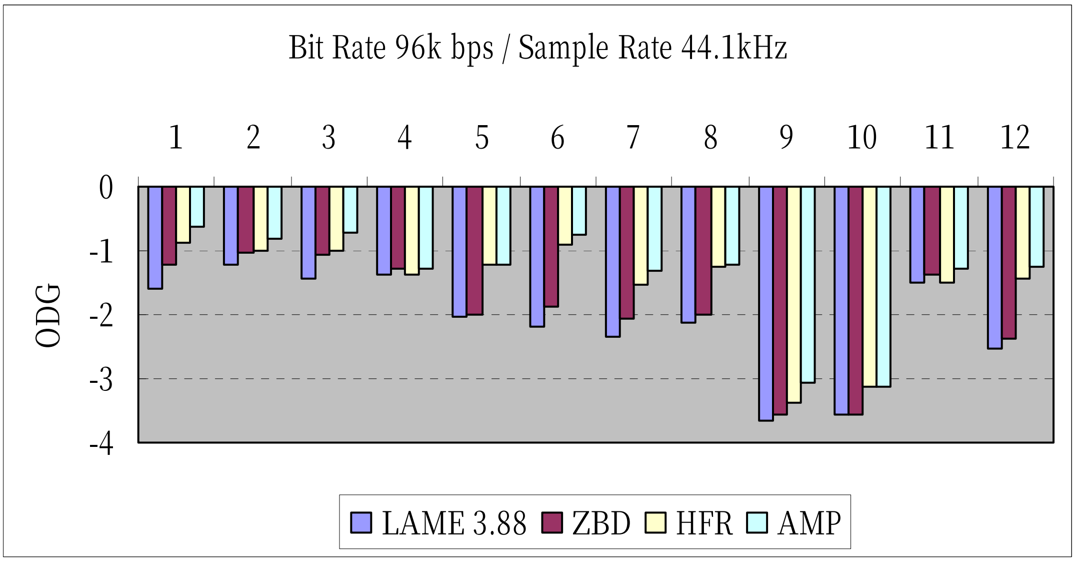


Figure 48: The ODG of the MP3 tracks, ZBD audio, HFR audio and APM audio under 96k bit rate.

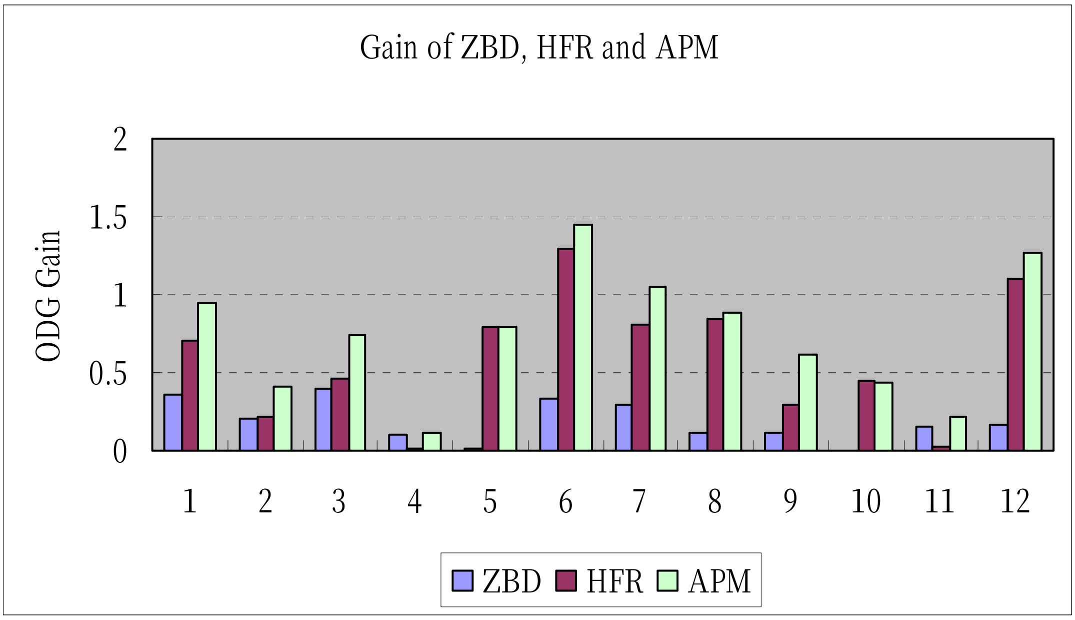


Figure 49: The gain of Zero Band Dithering, High Frequency Reconstruction and Audio Patch Method corresponding to Figure 48.

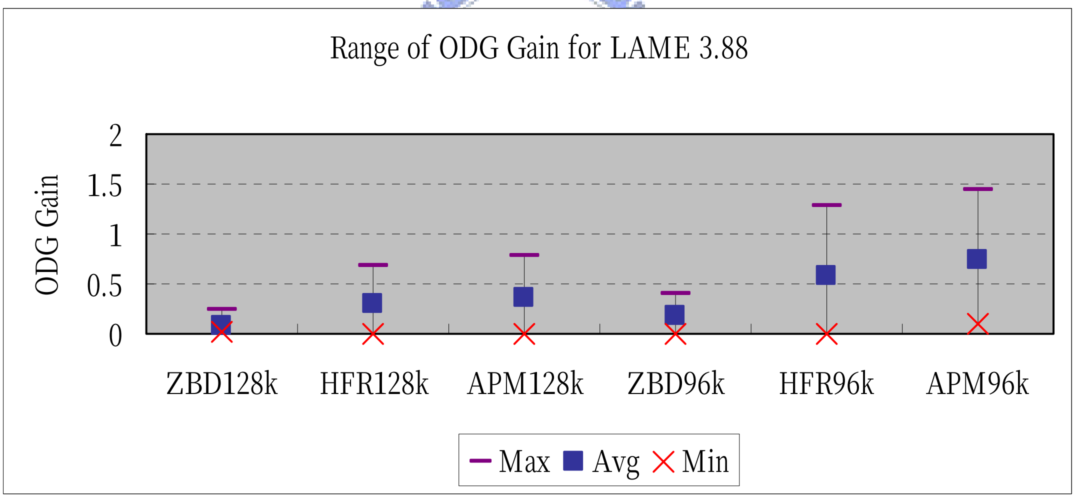
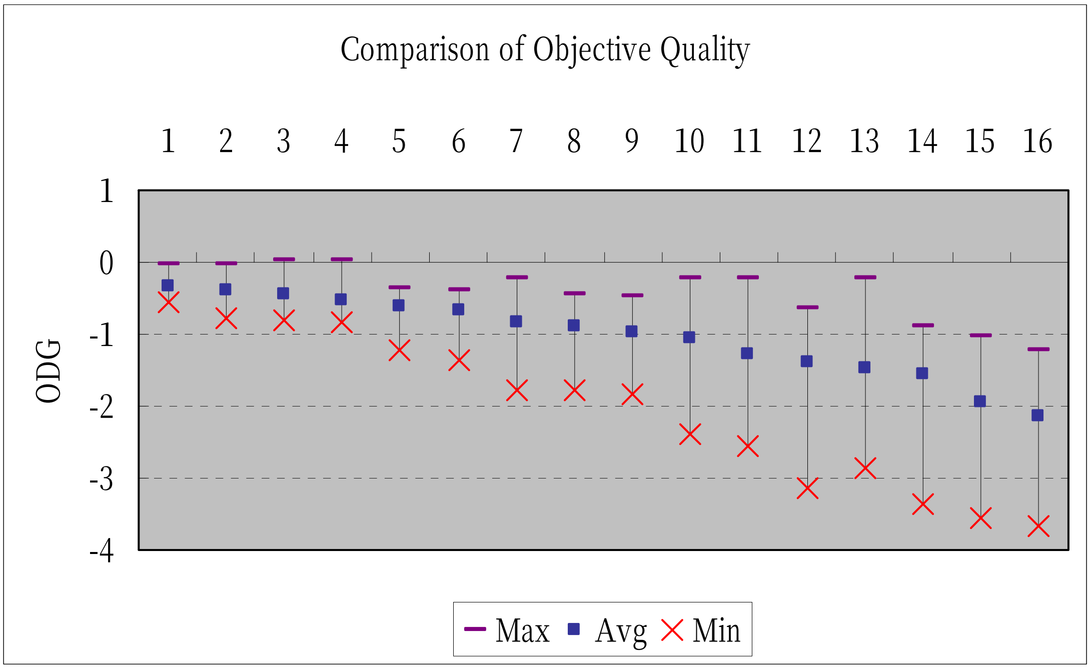


Figure 50: The range of the gain of ZBD, HFR and APM corresponding to Lame 3.88 at 128k and 96k bit rate.



- QT 6.3 with ASP under 128k bit rate.
- QT 6.3 with HFR under 128k bit rate.
- QT 6.3 with ZBD under 128k bit rate.
- QT 6.3 under 128k bit rate.
- Lame 3.88 with APM under 128k bit rate.
- Lame 3.88 with HFR under 128k bit rate.
- QT 6.3 with APM under 96k bit rate.
- Lame 3.88 with ZBD under 128k bit rate.
- Lame 3.88 under 128k bit rate.
- QT 6.3 with HFR under 96k bit rate.
- QT 6.3 with ZBD under 96k bit rate.
- Lame 3.88 with APM under 96k bit rate.
- QT 6.3 under 96k bit rate.
- Lame 3.88 with HFR under 96k bit rate.
- Lame 3.88 with ZBD under 96k bit rate.
- Lame 3.88 under 96k bit rate.

Figure 51: The ODG range comparison of the QT 6.3 tracks, the LAME 3.88 tracks, ZBD audio, HFR audio and ASP audio under 128k and 96k bit rate.

5.2 Verification of Quality Enhancement by Audio Patch Method based on Music Database

To confirm the effect of quality improvement and evaluate the possible risk for a variety of music category, the audio database [27] of our laboratory, NCTU PSPLab, is adapted to verify the audio patch method further. The database includes 327 tracks that are separated into sixteen sets with different signal properties as shown in Table2. Also, the Lame version 3.88 is used to prepare the music tracks. Because there are five tracks that are 24 bit PCM format unsuitable for LAME encoder, only 322 tracks are used to test in practice.

Figure 52 and 54 illustrate the average ODGs for the sixteen track set in the database under different decoding processing as well as different bit rates. In those figures, every two bars combined as a data set for a track set and they represent the ODG of the original decoded music and the decoded music with audio patch method, respectively. Figure 53 and 55 illustrate the ODG gains by audio patch method from the sixteen set. From the experiments, it shows that the objective quality is improved for all the sets except the 15^{th} set at 128k bit rate. The improvements for most sets are at least 0.2 and 0.6 at 128k and 96k bit rate, respectively. For the 322 tracks in the database, the average ODGs are form -1.05 up to -0.70 and from -2.10 up to -1.23 at 128k and 96k bit rate, respectively. The result exerted on a great deal of tracks confirms the patch method can enhance the decoded audio quality indeed, not restrained for the special music category.

On the other hand, to evaluate the possible risk, Figure 56-61 illustrate the distributions of ODG gains from the 322 tracks at 128 k and 96k bit rate, respectively. Figure 56 and Figure 59 are the enhancement track distributions, and Figure 57 and Figure 60 are the degradation track distributions. However, there are some tracks that are band-limited originally, not due to compression. Those tracks get worse ODGs after high frequency reconstruction, because the measurement of PEAQ system is the difference between the original source track and the decoded track. Hence, those should not be subsumed the statistics distribution. Figure 58 and Figure 61 exclude those original band-limited tracks. From Figure 58 and Figure 61, the degradation tracks are centralized on the range "-0.05 ~ 0". The high frequency of the tracks is not rich, and the high frequency reconstruction usually causes the unnoticeable slight quality degradation that is not over 0.1. Figure 62 illustrates an example for the type tracks. The remainder almost belongs to the 15^{th} set consisting of artificial bitstreams that contains sin waves. The artificial structure containing strong tone components causes a wide gap between the energies of the tone component and the noise-like component. This makes the processing of zero band dithering usually generate unsuitable noise floor for the unusual phoneme illustrated in Figure 63-65.The tracks in the degradation range that is over -0.1 are almost due to the reason. To decrease the patch gain can improve the problem effectively. Furthermore, the music category is usually used for testing, not a common music. The reasons mentioned above almost cover the all causes of the quality degradation due to the audio patch method. However, finally, the special exquisite harmonic structures as illustrated in Figure 66 give another example of quality degradation. The quality of this kind of signals mainly depends on the keeping of the exquisite harmonic structure. Once the harmonic structure is destroyed, the ODG will degrade very largely. For example, the ODG for the 128k bps decoded audio in Figure 67 is -3.21. However, if we filter the same decoded audio to discard the frequency component over than 10k Hz, the ODG is up to -2.41. This shows that we would rather discard the destroyed high frequency component than keeping the defective harmonic structure. Because the regenerated high frequency component by the high frequency reconstruction method is extended by duplicating the low frequency component, therefore the ODG degradation after the processing of the audio patch method can be expected. However, the range of ODG degradation is still lower than -0.16 for the kind of signals after audio patch.

In conclusion, the risk of the patch method is mainly due to artificial factors. For common music, the quality degradation is almost not over than 0.07 from the great deal of experiments.

Table 2: The PSPLAB audio database.

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

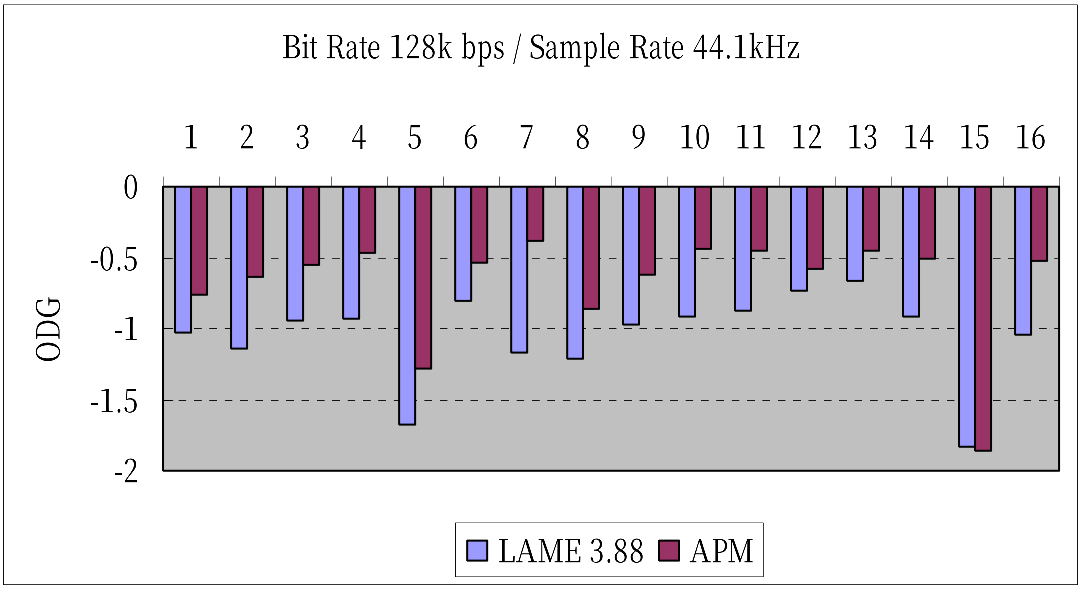


Figure 52: The average ODG of the LAME 3.88 tracks and APM audio under 128k bit rate for the sixteen track sets of the PSPLab database.

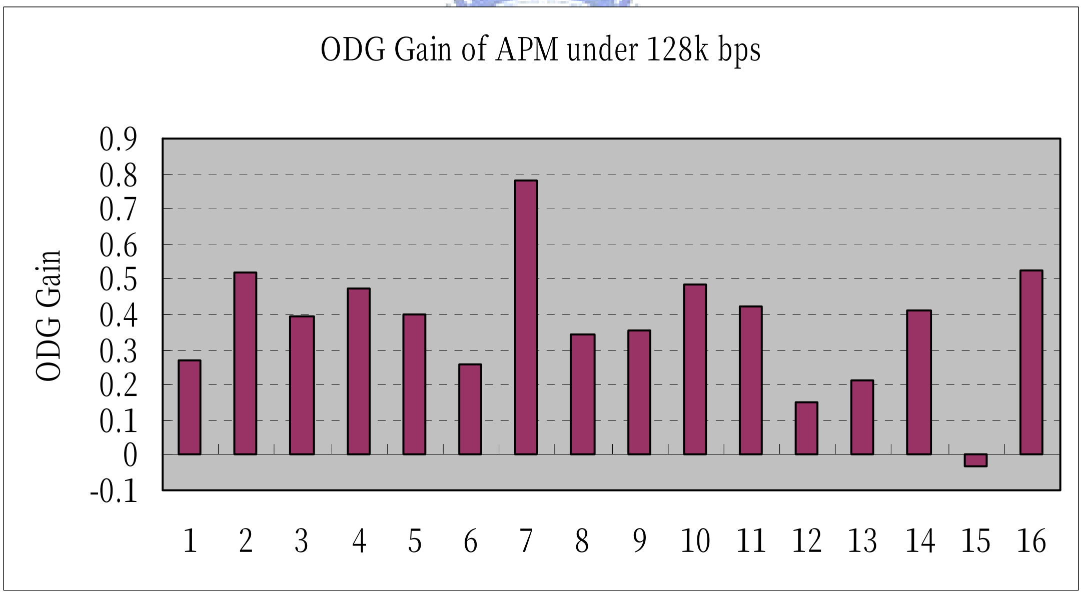


Figure 53: The gain of Audio Patch Method for the sixteen track sets corresponding to Figure 52.

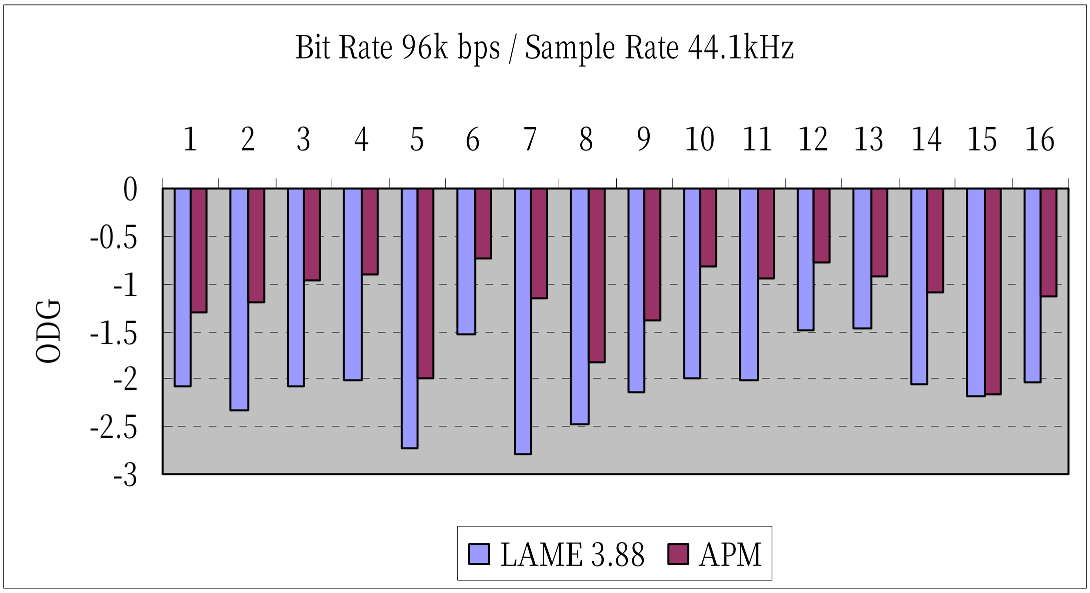


Figure 54: The average ODG of the LAME 3.88 tracks and APM audio under 96k bit rate for the sixteen track sets of the PSPLab database.

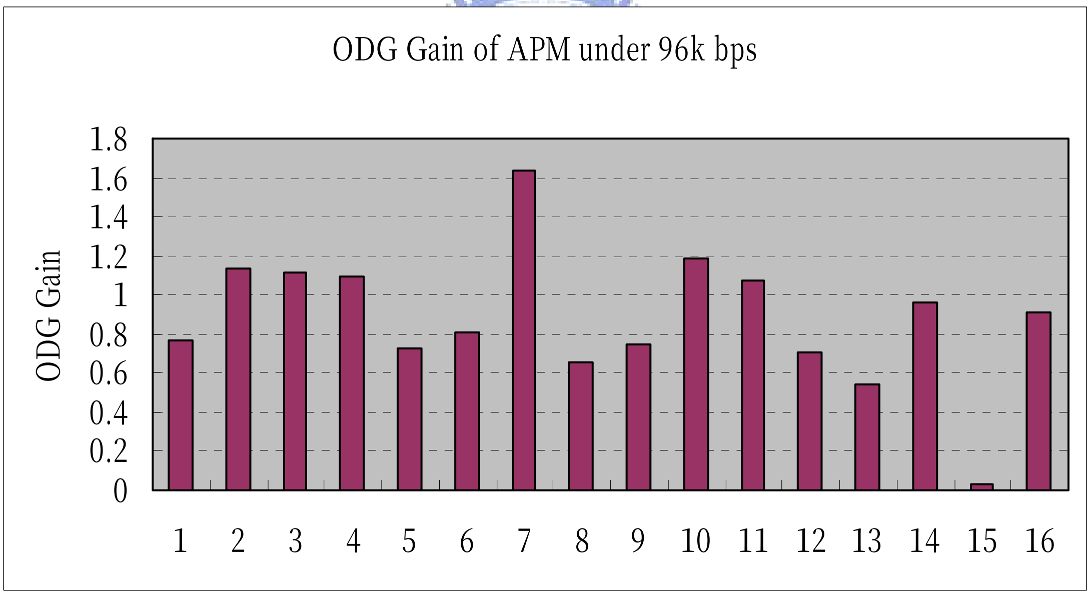


Figure 55: The gain of Audio Patch Method for the sixteen track sets corresponding to Figure 54.

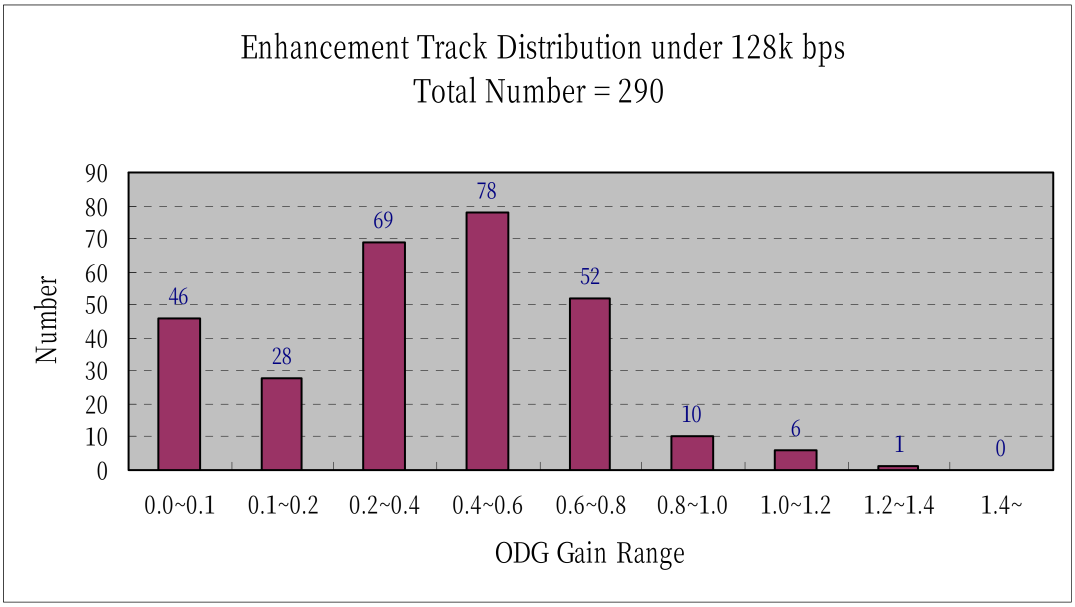


Figure 56: The enhancement track distribution among the 322 tracks contained in the sixteen track sets of the PSPLab database under 128k bps.

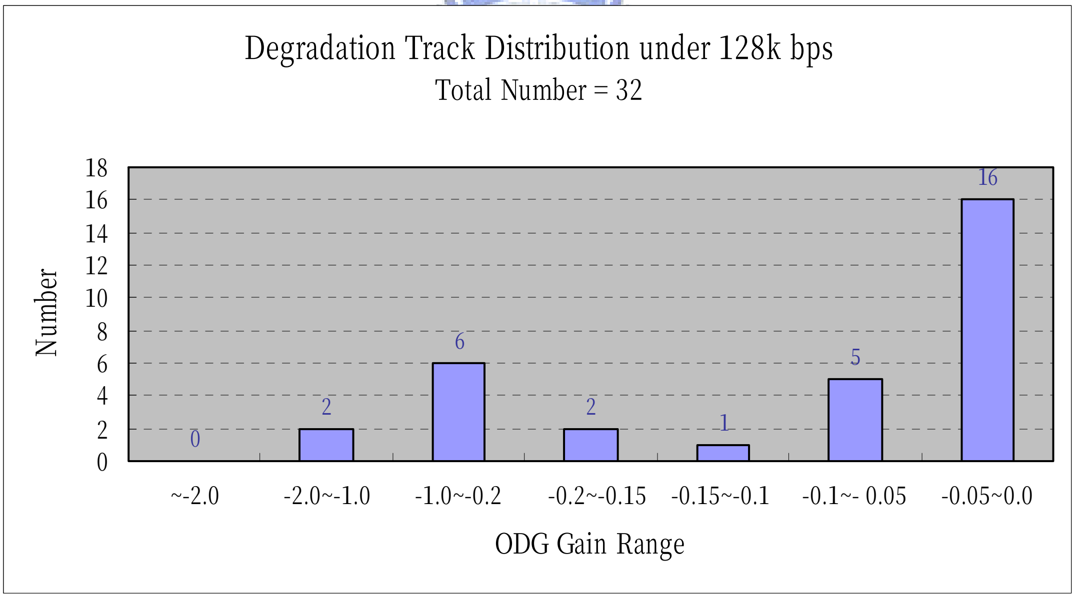


Figure 57: The enhancement track distribution among the 322 tracks contained in the sixteen track sets of the PSPLab database under 128k bps.

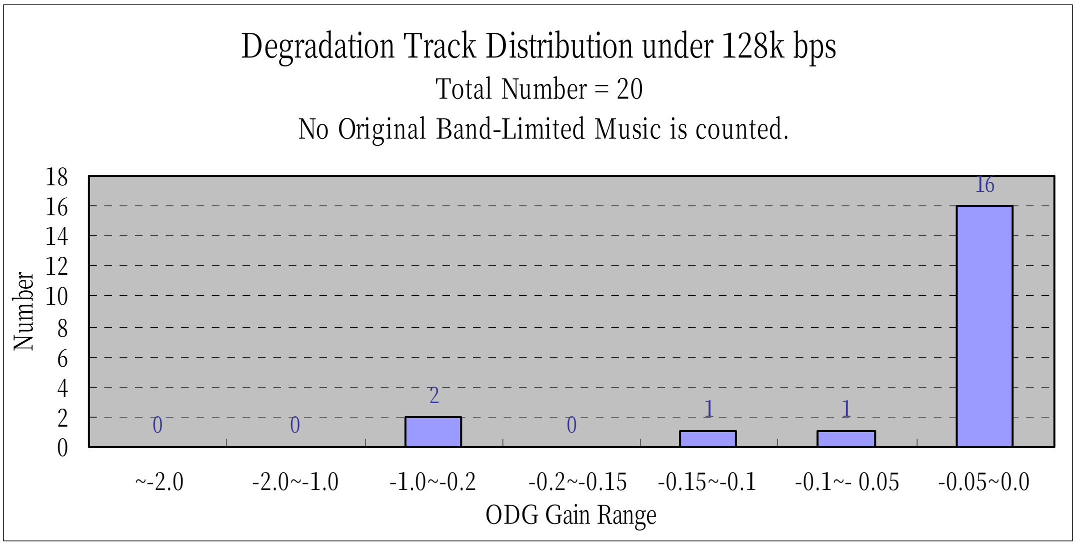


Figure 58: The degradation track distribution among the 322 tracks contained in the sixteen track sets of the PSPLab database. The source music band-limited originally is not counted under 128k bps.

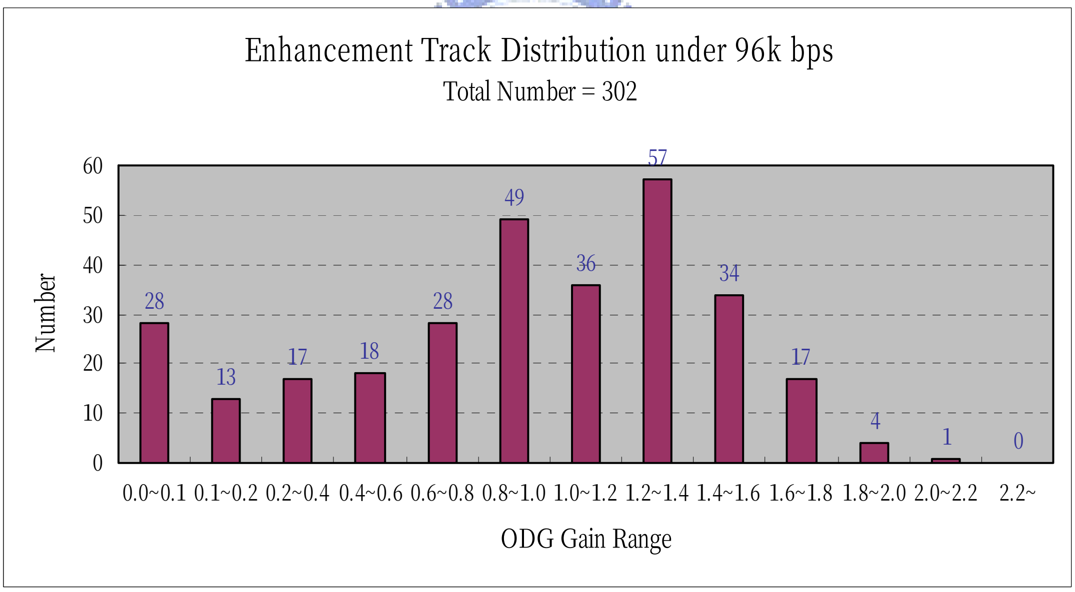


Figure 59: The enhancement track distribution among the 322 tracks contained in the sixteen track sets of the PSPLab database under 96k bps.

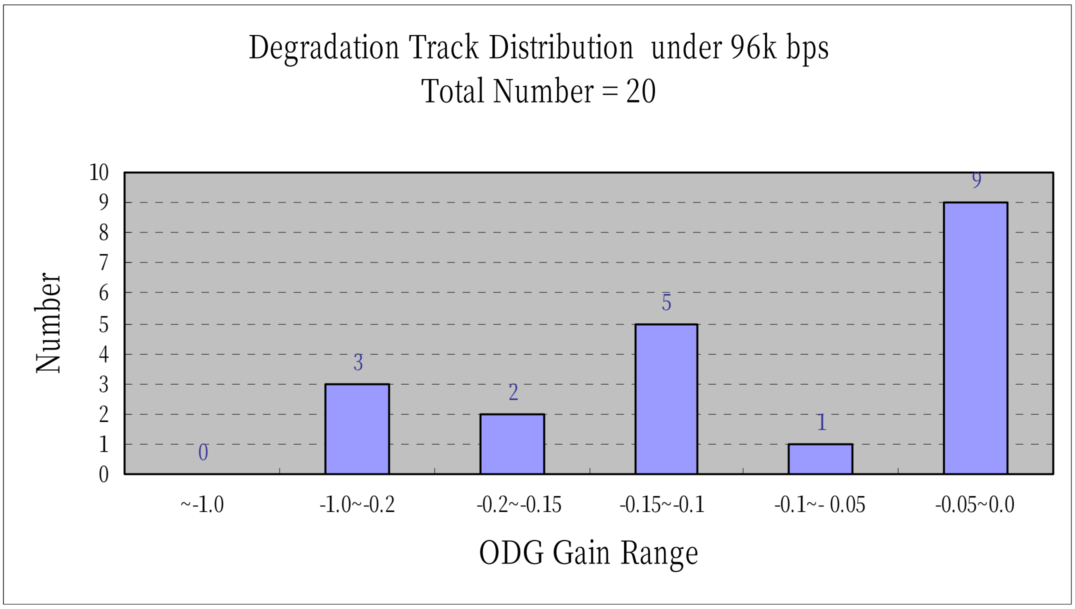


Figure 60: The enhancement track distribution among the 322 tracks contained in the sixteen track sets of the PSPLab database under 96k bps.

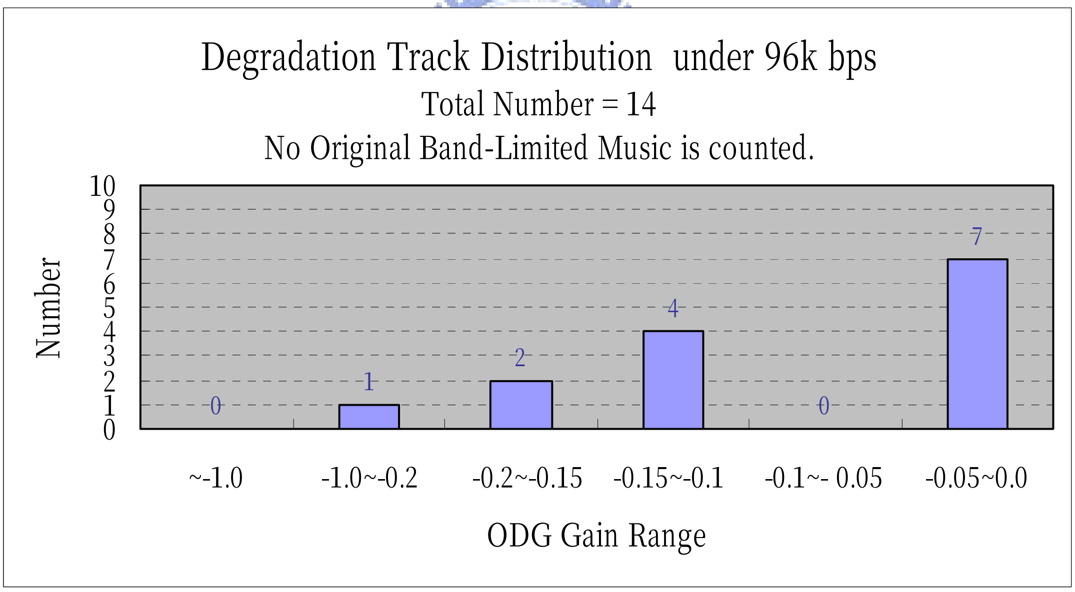


Figure 61: The degradation track distribution among the 322 tracks contained in the sixteen track sets of the PSPLab database. The source music band-limited originally is not counted under 96k bps.

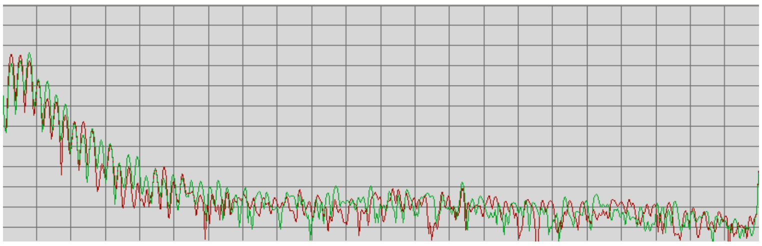


Figure 62: The spectrum of horn track. The high frequency content is not rich.

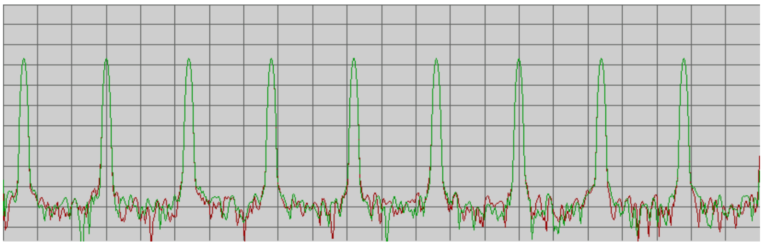


Figure 63: The spectrum of the original audio that is an artificial signal containing a wide gap between the tone component and noise-like component.

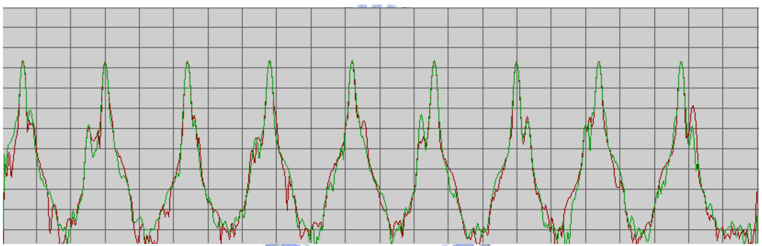


Figure 64: The spectrum of decoded audio corresponding to **Figure 63**.

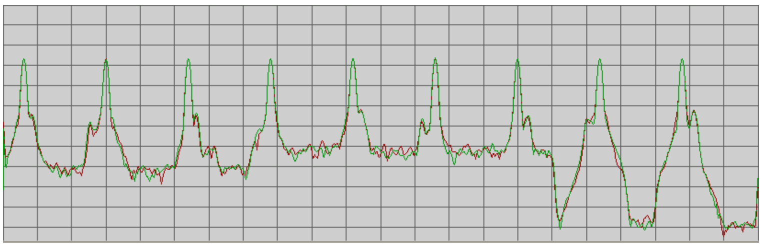


Figure 65: The spectrum of decoded audio with audio patch method corresponding to Figure 64.The zero band dithering generates the unsuitable noise floor with an excessive energy.

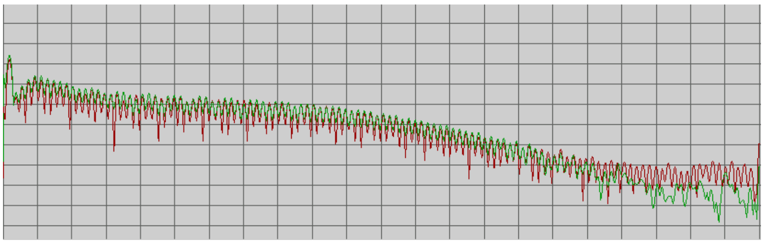


Figure 66: The spectrum of the original audio that contains exquisite harmonic structures.

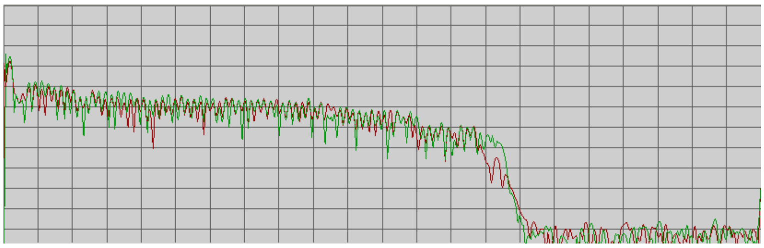


Figure 67: The spectrum of decoded audio corresponding to Figure 66.

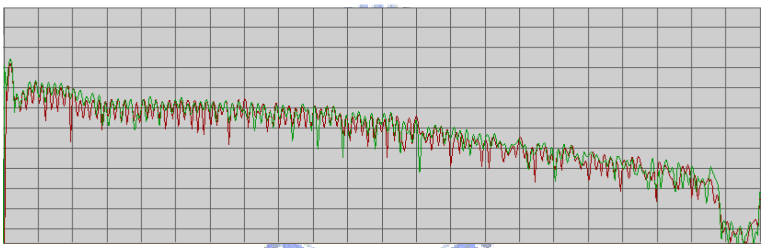


Figure 68: The spectrum of decoded audio with audio patch method corresponding to Figure 67.