Chapter 1 Introduction

Audio compression greatly decreases the burden of storage and transmission. Unfortunately, current audio encoders like MP3 or AAC inevitably leads to some artifacts due to the bit rate constraint. This paper focuses on two artifacts—spectral valley and spectral clipping. Spectral valley is from the improper bit allocation or the encoder scheme which leads to the unusual hollow which easily perceived as the annoying “fishy” noise. Spectral clipping is from the clipping of the high frequency contents in audio encoders to have bits reserved for the low frequency contents which is more sensitive to the human hearing systems. This thesis proposes a spectrum patch method to handle these artifacts in the decoders. The method consists of two individual parts, zero band dithering and high frequency reconstruction.

Formally, the spectral valley, as shown in Figure 1, is defined as a zero band that is a spectral band containing zero energy in the middle of the spectrum. In other words, the distortion energy of a zero band is equal to the energy of the original signal on the zero band. Zero band phenomenons are mainly due to unsuitable bit allocation policies or excessive masking energy estimation.

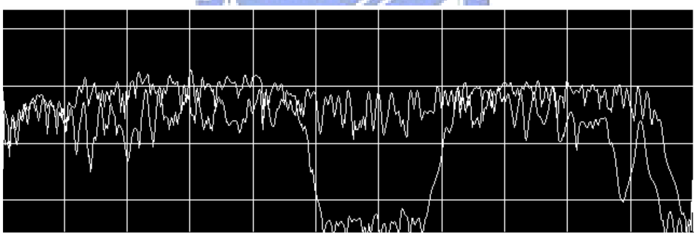

Figure 1: Audio spectrum containing a zero band.

Human hearing is very sensitive to the “fishy” noise caused by zero bands. As shown in Figure 2 , the objective of zero band dithering is to patch the valley in the spectrum to conceal the artifact.

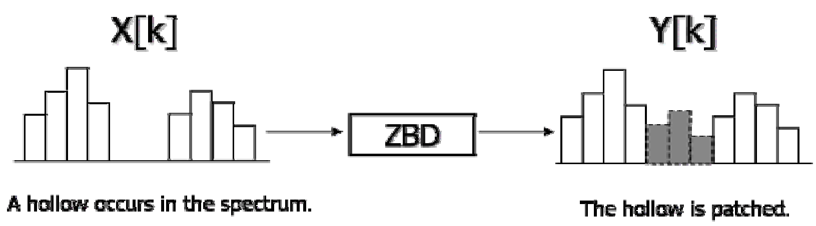

Figure 2: Block diagram of zero band dithering.

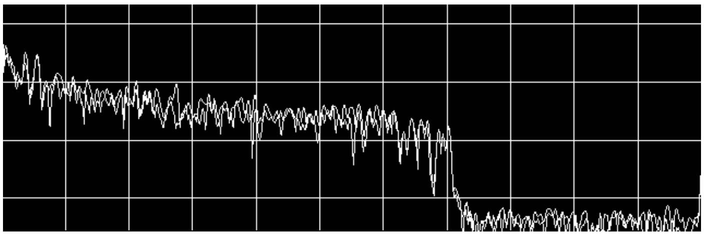

Figure 3: Spectrum of a band-limited audio signal.

Under restriction of limited bit rate for compression, most audio codec's scarify the bits required for high frequency and put all available bits to the low frequency component that is more relevant for human hearing. Figure 3 illustrates the strategy. The objective of high frequency reconstruction, as shown in Figure 4, is to reconstruct the high frequency components lost of the band-limited signals to make audio sound brighter.

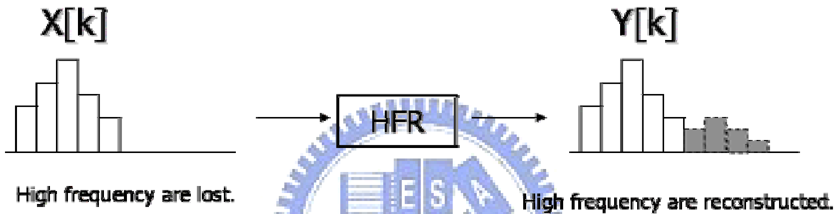

Figure 4: Block diagram of high frequency reconstruction.

Many attempts have been made to extrapolate a wideband signal from its narrowband frequency components [1]-[12]. For most of them were limited to speech, instead of a general audio signal. Recently, some methods for a general audio signal have been proposed progressively [13] [14]. However, they were almost either time-domain approaches, or frequency-domain approaches that needs priori information. Because the process of audio decoding is mainly based on frequency-domain, the time-domain approaches are not effective to apply to compressed audio. Furthermore, an advanced scheme referred to as "spectral band replication (SBR)" [15]-[19] has become the reference model of the MPEG-4 version 3 audio standard to compress high frequency contents. The SBR needs side information on the frequency contents extracted in encoder to help the reconstruction of the high frequency contents in decoder. Hence, the SBR can be merely applied to the only special audio format to improve the perceptual quality, not all the encoded music with limited bandwidth. On the other hand, in MPEG-4 AAC encoder, perceptual noise substitution (PNS) [20] that is used to code the noise-like components efficiently can improve the zero band phenomenons. The priori information of the signal energies in the scalefactor bands for which PNS is active also needs to be transmitted to decoder to reconstruct the noise-like components. Hence, PNS can be only applied to MPEG-4 AAC tracks, not all the encoded music such as MP3 tracks. 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 can be dithered or reconstructed to improve the perceptual quality. Both the subjective and objective measures have been conducted and shown the better quality. Especially, the objective test is through the perceptual evaluation of audio quality system [21], which is the recommendation system by ITU-R Task Group 10/4, to measure the perceptual difference of the artifacts.

The rest of this thesis is organized as follow: Chapter 2 gives an overview of common time-frequency mapping tools. Chapter 3 discusses the cause of zero bands and proposes a dithering method. In Chapter 4, both a high frequency reconstruction method and a fast computing method are proposed. Chapter 5 conducts experiment to check out the objective quality of the patch method.