Audio Codec Using Psychoacoustic Masking

Yiwen Chen
Department of Electrical and Computer Engineering
Drexel University
Philadelphia, PA
yc625@drexel.edu

***Abstract* — The Audio Codec Developed in this paper is using the psychoacoustic masking model to compress the data in both time domain and frequency domain.**

Keywords—Audio Codec, Psychoacoustic Masking Model

I. Introduction

The audio codec is an audio data compression tools, with or without loss of information. The audio signal would be encoded to a one-dimensional array. The array could be decoded and restore the audio data. In order to evaluate the performance of the codec, the SNR and perception test was used.

II. Encoding

A. Masking Effect in Frequency Domain

The implementation of the encoding program of the codec is based on the masking effect. First, the audio signal was processed by the short-time Fourier transform. The result is the Fourier series in multiple time frames.

Masking is the process by which the threshold of hearing for one sound is raised by the presence of another sound. [1] In frequency masking, the sound with a high amplitude would mask the nearby frequency if the corresponding frequency is lower than the masking threshold in the critical band. (Fig. 1)

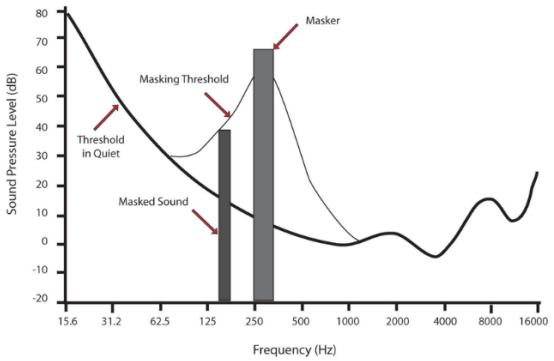

Fig. 1. Masking Effect of a Pure Tone (Borelli, D 2016) [2]

The Mel Filter Bank is used to calculate the threshold function in different frequency. The Fig 2 shows the Mel-spaced filter bank with unit amplitude. The critical band is increasing with the central frequency. However, this increment is not a linear relationship. [3]

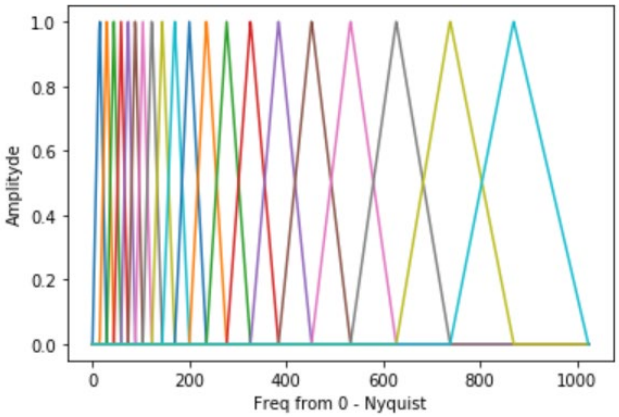

Fig. 2. Mel Spaced Filter Bank with Unit Amplitude

B. Masking Effect in Time Domain

In time domain masking, loud sounds would mask the sound before. The Fig. 3 shows the time domain masking. In this codec, the sounds appear after a loud sound within 200 microseconds would be masked.

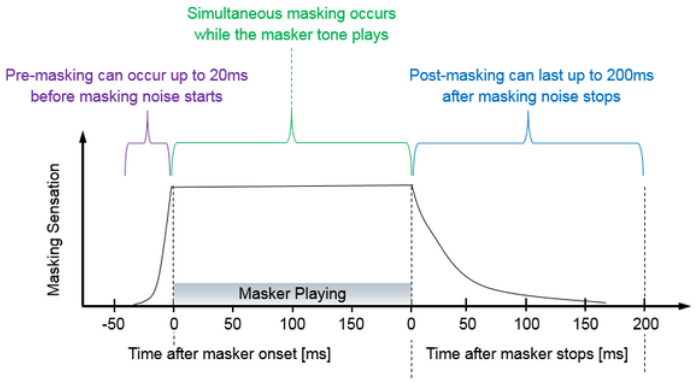

Fig. 3. Time Domain Masking (The Music Telegraph,2019) [4]

C. Quantization

After performing the masking in both frequency domain and time domain, the result was stored in an array. Each row in the array is the resulting Fourier series in each time frame. Due to the output of codec function should be one-dimensional array, those data need to be further processed. As

Identify applicable funding agency here. If none, delete this text box.

there are several values of zero found in the resulting array, there's a need to develop a data store technic for spare arrays. The method of the coordinate list was used to perform this optimization. In this method, the index would be stored at the beginning of the resulting array. And the value would be stored after the index. In order to restore the data, the size of the index part and max value should be stored. The data of Fourier series in each frame, which is stored in the format of floating-point, also needs to be transformed to 16-bits integer.

Quantization method was used to store the floating-point number to 16bits integer. In this project, there are two quantization method was used uniform quantization and ununiform quantization method called A-Law Quantization, which is used in the Europe telephone network. The maximum number was stored at the beginning cell for restoring the data. Then all the flourier series are normalized to 1 and fed to the uniform quantizer. The fig. 4. shows the comparison between uniform quantization and ununiform one, and the uniform quantization method is used in this project after evaluating the quantization error using the SNR model.

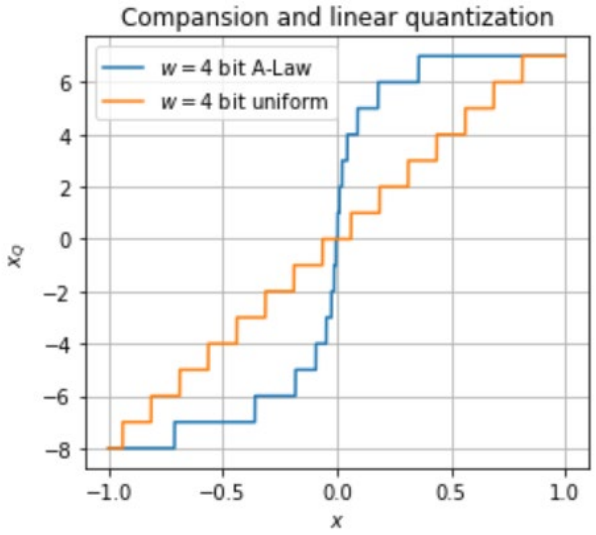

Fig. 4. Comparation of Quantization (dsp-nbsphinx) [5]

III. Decode

A. *Restore Short-Time Fourier Series*

In the decoding process, the first two cell store the size of short-time Fourier Series and follow by two cells store the amplitude information to restore the data from quantization. Then the index list was store. The index data would be copied to an array and then the data of short-time Fourier series by frame would be restored to a separate array. The inverse short-time Fourier transformation was used to restore the audio data. Then this data would be quantized to 16-bits integer and scale by the max amplitude of the original audio signal. An additional test, aiming at change made by restored phase data, was also performed. Due to the little difference between the output signal with and without phase data, the all phase data is dropped.

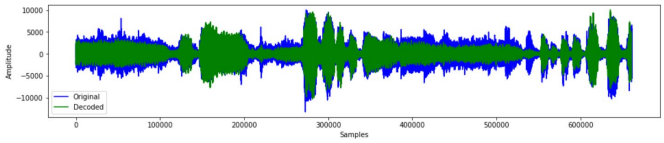

Fig. 5. Comparation with Original Signal

References

[1] Masking - masking of sounds. (n.d.). Retrieved from https://www.hear-it.org/Masking.

[2] Borelli, D., Gaggero, T., Rizzuto, E., & Schenone, C. (2016). Holistic control of ship noise emissions. *Noise Mapping*, *3*(1). doi: 10.1515/noise-2016-0008

[3] Crypto. (n.d.). Mel Frequency Cepstral Coefficient (MFCC) tutorial, Retrieved from http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/.

[4] The Music Telegraph MP3 Encoding & Masking. (2019, March 25). Retrieved from https://m.themusictelegraph.com/373.

[5] Non-Linear Requantization of a Speech Signal. (n.d.). Retrieved from https://dsp-nbsphinx.readthedocs.io/en/nbsphinx-experiment/quantization/nonlinear_quantization_speech_signal.html