The Use of FFT and MDCT in MP3 Audio Compression

John (Jake) Guckert
Math 2270
Spring 2012

Overview

- Overview of MP3 codec
- MP3 Encoding Algorithm
- What is the Fast Fourier Transform?
- How is the FFT implemented in MP3 encoding?
- What is the Modified Discrete Cosine Transform?
- How is the MDCT implemented in MP3 encoding?
- Summary

Overview of MP3 Codec

- MP3 stands for MPEG-1 Audio Layer III.
- Created in 1994 by the Moving Pictures Expert Group (MPEG).
- It is a lossy compression, meaning that not all of the original data is preserved after the compression algorithm is finished.
- MP3 is also based on ideas from the field of psychoacoustics. The idea is that the the human ear can only discern sounds from 20Hz to 20KHz, so any data outside of this threshold can be discarded to make the file smaller.

MP3 Encoding Algorithm

- The overall algorithm is broken up into 4 main parts.
- Part 1 divides the audio signal into smaller pieces, these are called frames. An MDCT filter is then performed on the output.
- Part 2 passes the sample into a 1024-point FFT, and then the psychoacoustic model is applied. Another MDCT filter is performed on the output.

MP3 Encoding Algorithm Cont.

- Part 3 quantifies and encodes each sample. This is also known as noise allocation. The noise allocation adjusts itself in order to meet the bit rate and sound masking requirements.
- Part 4 formats the bitstream, called an audio frame. An audio frame is made up of 4 parts, The Header, Error Check, Audio Data, and Ancillary Data.

The Fast Fourier Transform

- The FFT is an algorithm that computes the Discrete Fourier Transform and its inverse.
- The FFT produces the exact same result as evaluating the DFT directly, but the FFT produces an answer much faster.
- In general the DFT is found by using the equation:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}}$$

Where $X_0 \ldots X_{N-1}$ are complex numbers and $k = 0 \ldots N-1$

The FFT Applied to MP3 Encoding

- The FFT is used as a filter bank on an audio sample. It is used to filter out unwanted or unneeded data from the sample.
- First, incoming audio samples, s(n) , are normalized based the following equation x(n):

$$x(n) = \frac{s(n)}{N(2^{b-1})}$$

Where N is the FFT length of the sample and b is the number of bits in the sample.

The FFT continued

- Second, the masking threshold of the sample is found by using an estimate of the power density spectrum, P(k). P(k) is computed by using a 1024-point FFT.

$$P(k) = PN + 10\log\left[\sum_{n=0}^{N-1} h(n)x(n)\exp\left(-j\frac{2\pi kn}{N}\right)\right]^2, 0 \leqslant k \leqslant N-1$$

- h(n) is a Hann Window denoted by: $h(n) = 0.5\left(1 - \cos\frac{2\pi n}{N-1}\right), 0 \leqslant i \leqslant N-1$

- PN is the power normalization term, it is usually around 96 decibels.

What is the Modified Discrete Cosine Transform?

- The MDCT is a Fourier related transform based on type-IV DCT. It has an additional property of being “lapped.”
- In general it was designed to be performed on larger, consecutive blocks of datasets where parts of the blocks overlapped.
- The MDCT is a linear function that has half as many outputs as inputs.

MDCT cont.

- This linear function transforms 2N real numbers to N real numbers according to the equation:

$$F : R^{2\mathrm{N}} \Rightarrow R^{N}$$

$$X_k = \sum_{n=0}^{2\mathrm{N}-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right]$$

The MDCT in MP3 Encoding

- The MDCT limits the sources of output distortion at the quantization stage.
- It is also used as and analysis filter given by:

$$h_k(n) = w(n)\sqrt{\frac{2}{M}}\cos\left[\frac{(2\mathrm{n}+M+1)(2\mathrm{k}+1)\pi}{4\mathrm{M}}\right]$$

- This is a block function and it extends across two input blocks at a time.

The MDCT in MP3 Encoding Cont.

- The MDCT performs a series of inner products between the input data x(n), and the analysis filter $h_k(n)$.
- This eliminates the blocking artifacts that would cause a problem during the reconstruction of the sample.
- The inverse MDCT reconstructs the samples without the blocking artifacts.

$$x(n) = \sum_{k=0}^{M-1} \left[X(k) h_k(n) + X^P(k) h_k(n+M) \right]$$

Summary

- The MP3 encoding algorithm has numerous complex parts.
- The FFT, DFT, and MDCT play a key role in encoding audio samples.
- These three transformations also play a role in other media compression formats, not just MP3.