ECE462 – Lecture 26

MPEG Audio and Psychoacoustics

- Frequency Masking: When an audio signal consists of multiple frequencies the sensitivity of the ear changes with the relative amplitude of the signals. If the frequencies are close and the amplitude of one is less than the other then the second (weakest) frequency may not be heard. The range of closeness for frequency masking is frequency and relative magnitude dependent (Critical Bands)

Critical Band Bandwidth

- For masking frequencies < 500 Hz. Remains approximately constant in width (~ 100 Hz)
- For masking frequencies > 500 Hz: Increases approximately linearly with frequencies (that is ~ 100 Hz for each additional 500 Hz)

- Ex: 1 Given a frequency of 300 Hz what is the next highest (integer) frequency signal that is distinguishable by human ear?

  (assuming the later signal is substantially weaker)

Answer: Critical band for 300 Hz is 100 Hz centered at 300. thus it is 250-350 Hz long. So the next highest audible frequency is 351 Hz.

- Ex: 2 Given a frequency fo 5000 Hz. What is the next highest (integer) frequency that is distinguishable human ear? (assuming again the later is weaker)

At 5000 Hz the critical band is 5000/500 * 100 = 1000 Hz

Centered at 5000 Hz: So the next highest audible frequency is 5501 Hz.

- Temporal Masking: After the ear hears loud sound it takes a further short while before it can hear a quieter sound.
- The cause for both types of masking is the tiny hair cells within the human ear. These are excited by air pressure variations. Different hair cells respond to different ranges of frequencies.
- Question: What is the level of compression needed?

For CD quality audio we use 44.1 KHz (44,100 samples/sec, 16 bits/sample and stereo (2 channels). This requires approximately 1.4 Mbps. Recall that MPEG 1 encodes video + audio at 1.5 Mbps (~1.25 Mbps video + 250 Kbps audio)

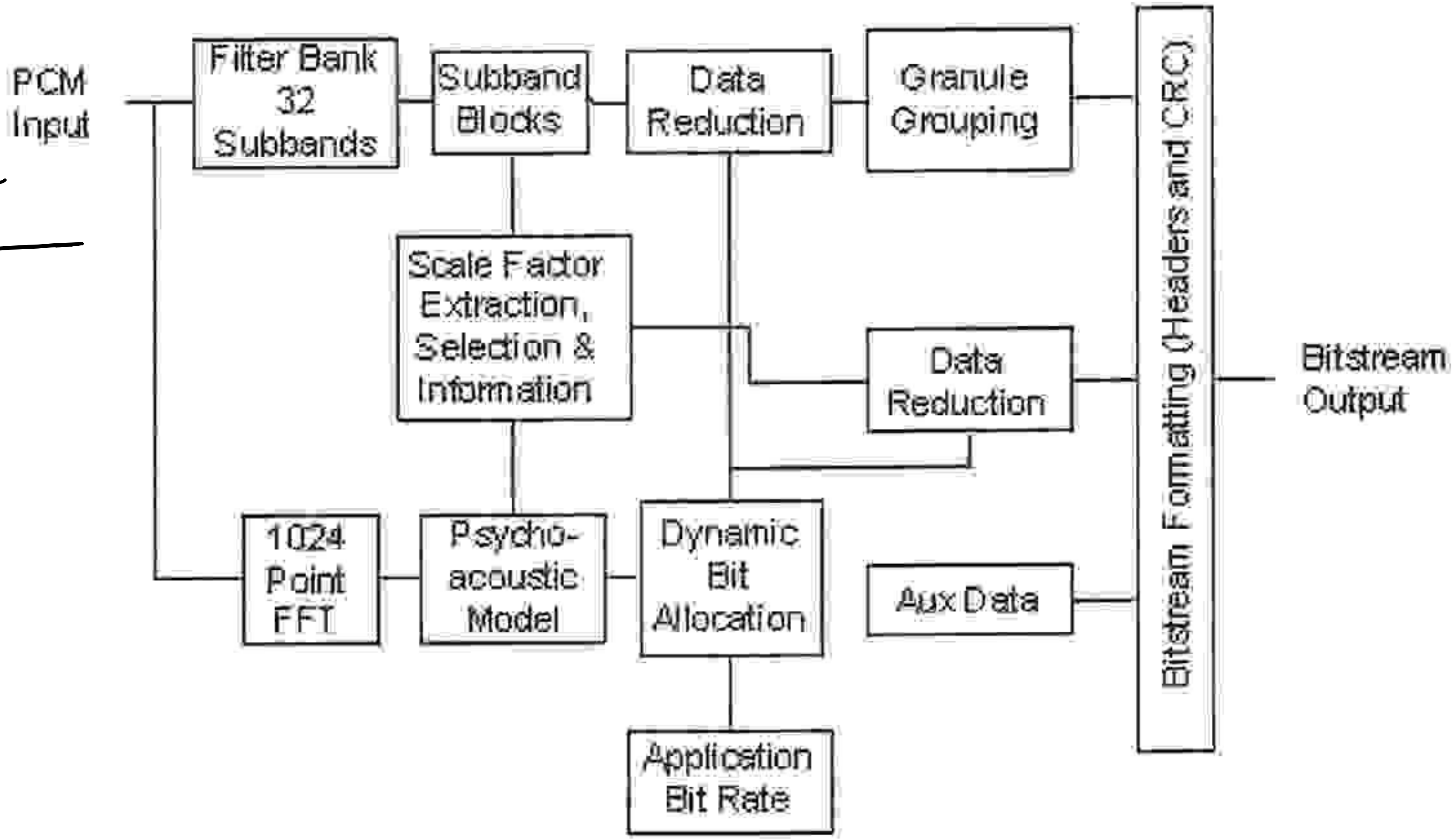
- Question: How does MPEG audio compression exploits frequency and temporal masking?

In MPEG Audio Compression

- Bandwidth is divided into frequency sub-bands using a bank of analysis filters – critical band filters.
- Each analysis filter is using a scaling factor of sub-band max amplitudes for psychoacoustic modeling.
- FFT (DFT) used. Signal to mask ratios used for frequencies below a certain audible threshold.

16 bit

Basic Mpeg Audio



MPEG Audio Layers

Layer I (128m192 Kb/s)

Layer II (96,128 Kb/s)

Layer III (64,96,128 Kb/s)

- Layer I/II are often treated together
- Layer III is the most advanced and also known as MP3
- Complexity and sound quality increases with layer
- Higher layer can decode all lower layers
- Sampling frequency, 32,44.1 or 48 KHz
- Two channels (MPEG 1) (5.1 channels in MPEG-2)

Sub-band Mapping

- In all the three layers a 16 bit linearly quantized PCM input signal is mapped from the time domain to 32 frequency bands (sub-bands).
- 32 frequency bands (sub-bands)
- Masking levels are calculated via psychoacoustic analysis to find the magnitude of the allowed quantization errors. Each sub-band is treated separately.

The modified DCT

- Designed to be performed on consecutive blocks of data, where subsequent blocks are overlapping by 50%.
- Good energy – compaction as in original DCT.
- Less blocking artifacts compared to original DCT.
- The MDCT maps 2N samples to N samples

$MDCT\colon F(u) = \sum_{i=0}^{2N-1} f(i)\cos[\frac{\pi}{N}(u+\frac{1}{2})(i+\frac{1}{2}+\frac{N}{2})]$, u=0,1,…,N-1

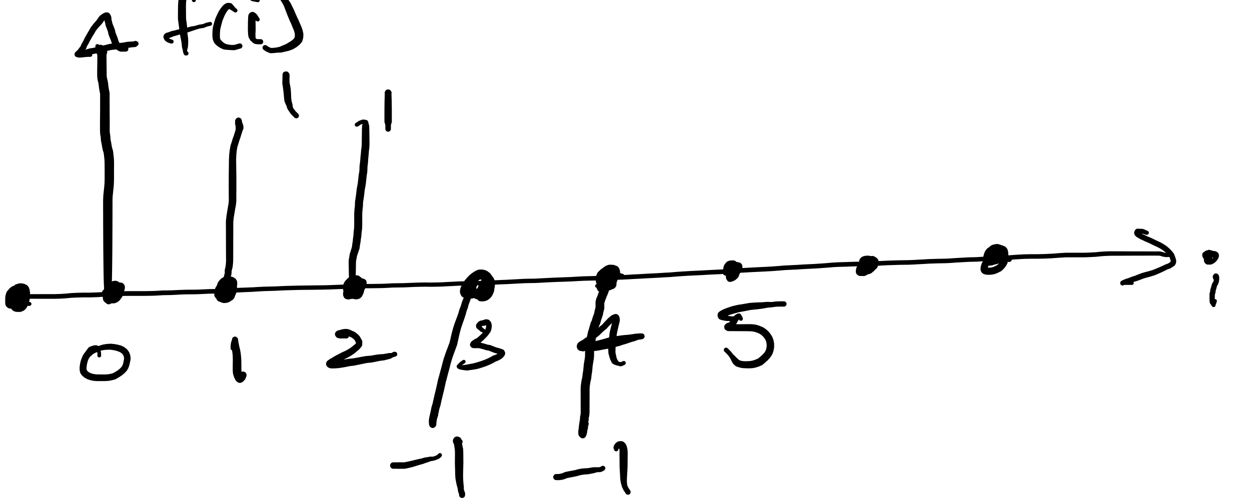
$$IMDCT\colon Y(i) = \frac{1}{N}\sum_{u=0}^{N-1} F(u)\cos\left[\frac{\pi}{N}\left(u+\frac{1}{2}\right)\left(i+\frac{1}{2}+\frac{N}{2}\right)\right], \qquad i = 0,1,\cdots,2N-1$$

Note that Y(i) is a time aliased (50%) version of F(i) ………

The modified DCT

- Ex:

Let



- Let us calculate the MDCT for N = 2. This means that we take segments of length 2N=4 overlapped by 50% that is two samples.

0 0 0 1 1 -1 -1 0 0 0

The modified DCT

- The MDCT for N=2 is based on the 4X2 matrix

$$A = \begin{bmatrix} 0.3827 & -0.9239 \\ -0.3827 & 0.9239 \\ -0.9239 & 0.3827 \\ -0.9239 & 0.3827 \end{bmatrix} : 4 \times 2$$

Thus, given X: 1 x 4

MDCT: X.A 1MDCT: ½ (XA). A^T

In our example:

MDCT 1: [0 0 0 1]. A 1

IMDCT1: ½ [0 0 0 1]. A. A^T = 0 0 | 0.5 0.5

MDCT2: [0 1 1 1].A

IMDCT2: ½ [0 1 1 -1]. A. A^T = | -0.5 0.5 | 0 0

MDCT3: [1 -1 -1 0] A

IMDCT3: ½ [1 -1 -1 0]. A. A^T = | 1 -1 | -0.5 -0.5

MDCT4: [-1 0 0 0] A

IMDCT2: ½ [-1 0 0 0]. A. A^T = -0.5 0.5 0 0

= 0 1 1 -1 -1 0 0

Note that in this case: A. A^T = $\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$