Additional Notes on Audio Compression

1. Simple Audio Compression Methods

Traditional lossless compression methods (Huffman, LZW, etc.) usually don't work well on audio compression (the same reason as in image compression).

The following are some of the Lossy methods:

- Silence Compression - detect the "silence", similar to run-length coding
- Adaptive Differential Pulse Code Modulation (ADPCM)

  e.g., in CCITT G.721 -- 16 or 32 Kbits/sec.

  (a) Encodes the difference between two or more consecutive signals; the difference is then quantized --> hence the loss
  (b) Fewer bits are needed when the difference is smaller.

  - It is necessary to predict where the waveform is headed --> difficult
- Linear Predictive Coding (LPC) fits signal to speech model and then transmits parameters of model --> sounds like a computer talking, 2.4 kbits/sec.
- Code Excited Linear Predictor (CELP) does LPC, but also transmits error term --> audio conferencing quality at 4.8 kbits/sec.

2. MPEG Audio Compression

Some facts

- MPEG-1: 1.5 Mbits/sec for audio and video

  About 1.25 Mbits/sec for video, no more than 256 kbits/sec for audio

  (Uncompressed CD audio is 44,100 samples/sec * 16 bits/sample * 2 channels > 1.4 Mbits/sec)
- MPEG audio supports sampling frequencies of 32, 44.1 and 48 KHz.
- Compression ratio ranging from 2.7 to 24.

- With Compression ratio 6:1 (16 bits stereo sampled at 48 KHz is reduced to 256 kbits/sec) and optimal listening conditions, expert listeners could not distinguish between coded and original audio clips.

- Supports one or two audio channels in one of the four modes:

    1. Monophonic -- single audio channel

    2. Dual-monophonic -- two independent channels, e.g., English and French

    3. Stereo -- for stereo channels that share bits, but not using Joint-stereo coding

    4. Joint-stereo -- takes advantage of the correlations between stereo channels

Bit Allocation Algorithm:

1. From the psychoacoustic model, calculate the signal-to-mask ratio (SMR) for each subband
    - SMR = 20 log_10 (signal / min_masking_threshold)

      This determines the quantization, i.e. the minimum number of bits that is needed, if available. (The amounts of the signals above the threshold, i.e. SMR, need to be coded. Signals below the threshold do not.)

2. Calculate signal-to-(quantization)-noise ratio (SNR) for all signals
    - A lookup table provides an estimate of SNR assuming a given number of quantizer levels.

3. Mask-to-(quantization)-noise ratio: MNR = SNR - SMR (in dBs)

4. Iterate until no bits left to allocate
    - Allocate bits to the subband with the lowest MNR
    - Look up new estimate of SNR for the subband allocated more bits, and re-calculate MNR

** Masking effect means we can raise the quantization noise floor around a strong sound because the noise will be masked off anyway. Must ensure that all the quantization noises are inaudible, i.e. below the masking thresholds.

** If more bits are allowed, allocate them so to further increase SNR. For each additional bit, we get 6 dB better SNR.

MPEG Layers

- MPEG defines 3 layers for audio. Basic model is same, but codec complexity increases with each layer.

- Layer 1: DCT type filter with one frame and equal frequency spread per band. Psychoacoustic model only uses frequency masking. It uses the strongest (tonal) frequency in each band to calculate the masking effects.

- Layer 2: Use three frames in filter (before, current, next, a total of 1152 samples). This models a little bit of the temporal masking.

- Layer 3: Better critical band filter is used (non-equal frequencies), psychoacoustic model includes temporal masking effects, takes into account stereo redundancy, and uses Huffman coder.

  Stereo Redundancy Coding:

  - Intensity stereo coding -- at upper-frequency subbands, encode summed signals instead of independent signals from left and right channels.
  - Middle/Side (MS) stereo coding -- encode middle (sum of left and right) and side (difference of left and right) channels.

Effectiveness of MPEG audio

Layer	Target Bit-rate	Quality at 64 kb/s	Quality at 128 kb/s	Theoretical Min. Delay
Layer 1	192 kb/s	---	---	19 ms
Layer 2	128 kb/s	2.1 to 2.6	4+	35 ms
Layer 3	64 kb/s	3.6 to 3.8	4+	59 ms

- Quality factor: 5 - perfect, 4 - just noticeable, 3 - slightly annoying, 2 - annoying, 1 - very annoying
- Real delay is about 3 times of the theoretical delay