**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 applied to audio compression:

* 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 consecutive signals,

(b) adapts at quantization so fewer bits are used when the value is smaller.

* + It is necessary to predict where the waveform is headed -> difficult
  + Apple has proprietary scheme called ACE/MACE. Lossy scheme that tries to predict where wave will go in next sample. About 2:1 compression.
* 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.

## Psychoacoustics

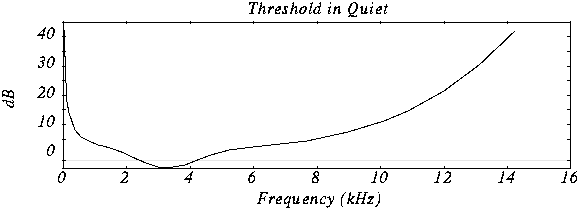
These methods are related to how humans actually hear sounds:

### Human hearing and voice

* Range is about 20 Hz to 20 kHz, most sensitive at 2 to 4 KHz.
* Dynamic range (quietest to loudest) is about 96 dB
* Normal voice range is about 500 Hz to 2 kHz
  + Low frequencies are vowels and bass
  + High frequencies are consonants

**Question: How sensitive is human hearing?**

* Experiment: Put a person in a quiet room. Raise level of 1 kHz tone until just barely audible. Vary the frequency and plot



### Critical Bands

* Perceptually uniform measure of frequency, non-proportional to width of masking curve

About 100 Hz for masking frequency < 500 Hz, grow larger and larger above 500 Hz.

* The width is called the size of the *critical band*

**Barks**

* Introduce new unit for frequency called a *bark*(after Barkhausen)

1 Bark = width of one critical band

For frequency < 500 Hz,

For frequency > 500 Hz,

* Masking Thresholds on critical band scale:

## MPEG Audio Compression

### Some facts

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

About 1.2 Mbits/sec for video, 0.3 Mbits/sec for audio

(Uncompressed CD audio is 44,100 samples/sec \* 16 bits/sample \* 2 channels > 1.4 Mbits/sec)

* Compression factor ranging from 2.7 to 24.
* With Compression rate 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.
* MPEG audio supports sampling frequencies of 32, 44.1 and 48 KHz.
* Supports one or two audio channels in one of the four modes:

**1.**

Monophonic - single audio channel

**2.**

Dual-monophonic - two independent channels (similar to stereo)

**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

### Steps in algorithm:

**1.**

Use convolution filters to divide the audio signal (e.g., 48 kHz sound) into frequency subbands that approximate the 32 critical bands -> *sub-band filtering*.

**2.**

Determine amount of masking for each band caused by nearby band using the results shown above (this is called the *psychoacoustic model*).

**3.**

If the power in a band is below the masking threshold, don't encode it.

**4.**

Otherwise, determine number of bits needed to represent the coefficient such that noise introduced by quantization is below the masking effect (Recall that 1 bit of quantization introduces about 6 dB of noise).

**5.**

Format bitstream

### MPEG Layers

* MPEG defines 3 layers for audio. Basic model is same, but codec complexity increases with each layer.
* Divides data into frames, each of them contains 384 samples, 12 samples from each of the 32 filtered subbands as shown below.
* Layer 1: DCT type filter with one frame and equal frequency spread per band. Psychoacoustic model only uses frequency masking.
* 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.

### Effectiveness of MPEG audio

--------------------------------------------------------------------

Layer Target Ratio Quality @ Quality @ Theoretical

bitrate 64 kbits 128 kbits Min. Delay

--------------------------------------------------------------------

Layer 1 192 kbit 4:1 --- --- 19 ms

Layer 2 128 kbit 6:1 2.1 to 2.6 4+ 35 ms

Layer 3 64 kbit 12:1 3.6 to 3.8 4+ 59 ms

--------------------------------------------------------------------

* 5 = perfect, 4 = just noticeable, 3 = slightly annoying, 2 = annoying, 1 = very annoying
* Real delay is about 3 times theoretical delay