Special Topics in Multimedia System

Indian Institute of Technology Delhi
(IITD)
New Delhi

SIL801

Audio Compression

Recap

- Audio Sampling and Quantization
- Pulse Code Modulation (PCM)
- Differential PCM (DPCM)
- Adaptive DPCM
- Psychoacoustic
    - Frequency masking
    - Temporal masking

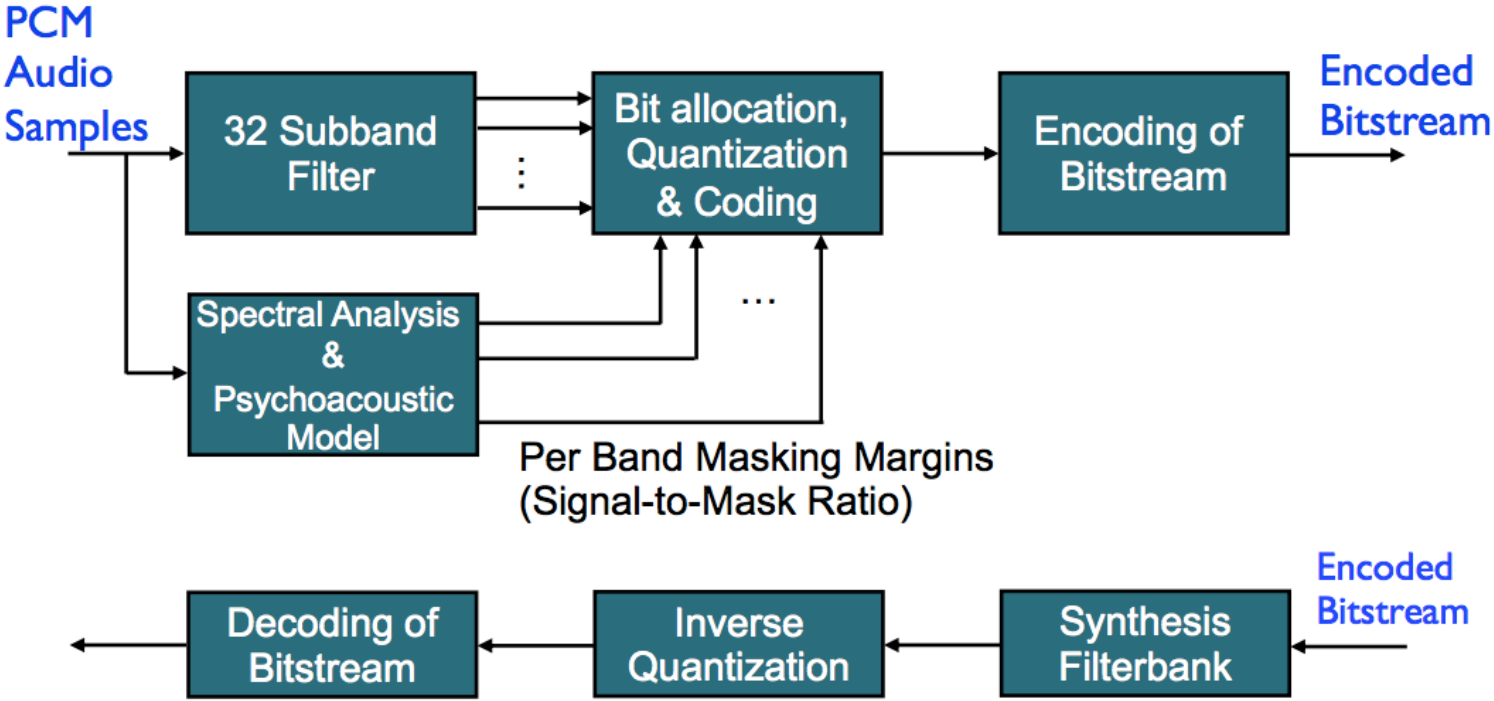
MPEG-1

Layer 1
- Psychoacoustic model only uses simultaneous frequency masking.

Layer 2
- This models a little bit of the temporal masking.

Layer 3
- Psychoacoustic model includes temporal masking effects, and takes into account stereo redundancy.
- Huffman coder.
- Known as MP3

MPEG-1: Overview Block Diagram



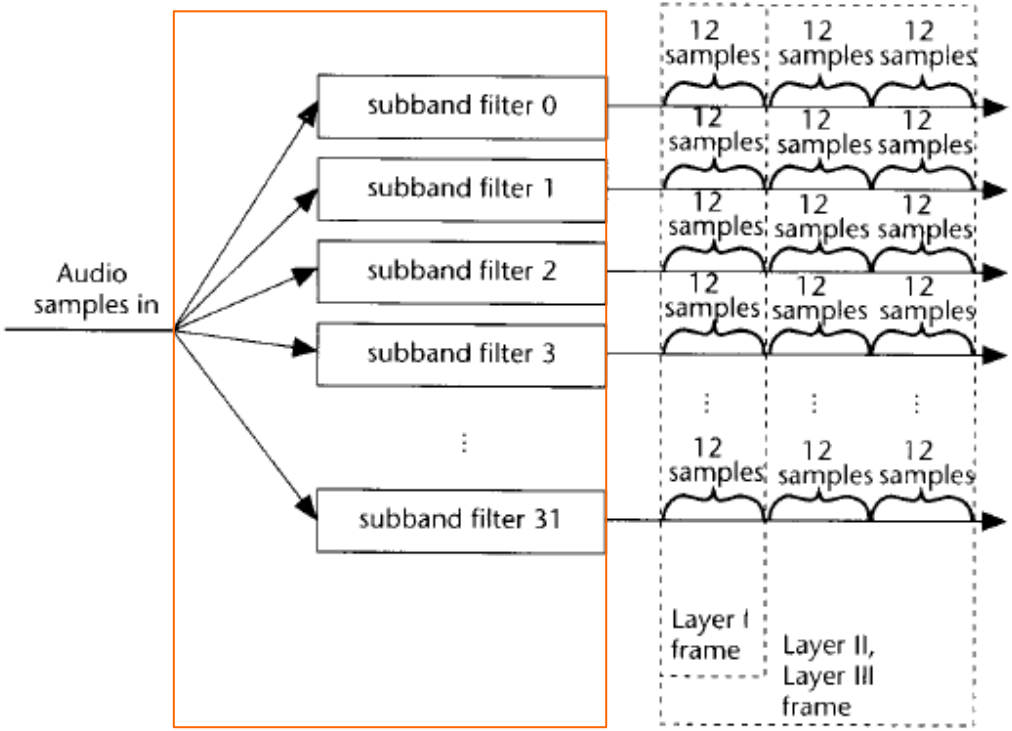
Source http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/

Sub Band Filtering

32 PCM samples yields 32 subband samples:

◦ Each sub-band corresponds to a frequency band evenly spaced from 0 to Nyquist frequency.
e.g., @48 kHz sampling rate, each sub-band is 24kHz/32 = 750 Hz wide.

NOTE: 32 constant-width subbands do not accurately reflect the ear's critical bands

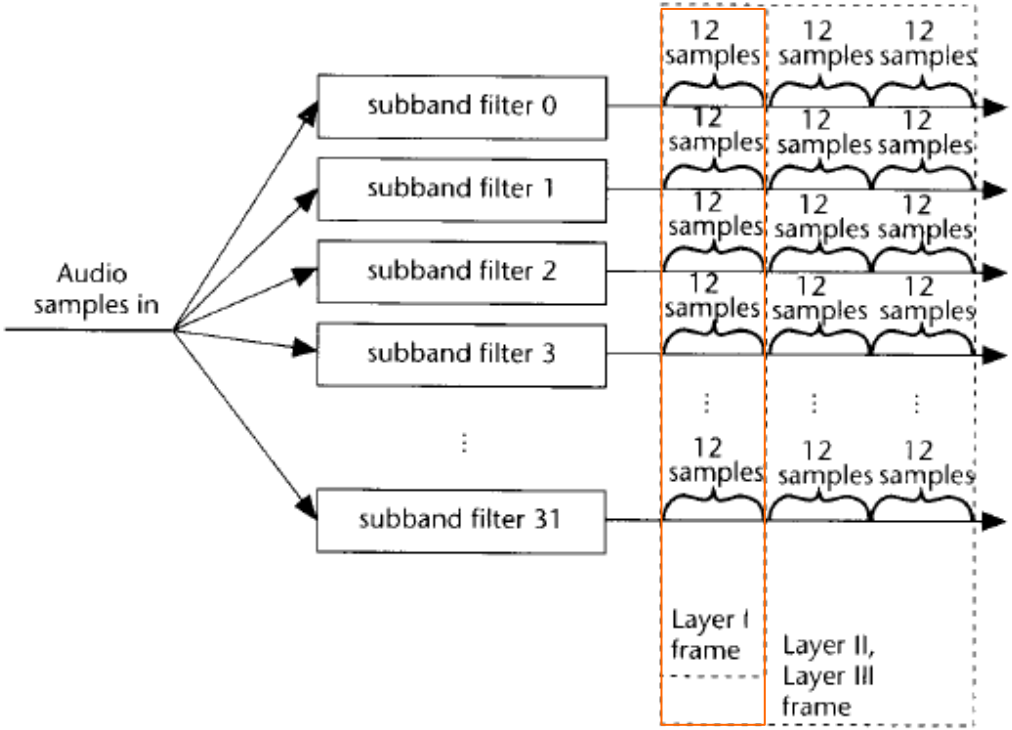


Sub Band Filtering

Samples out of each filter are grouped into blocks, called frames:

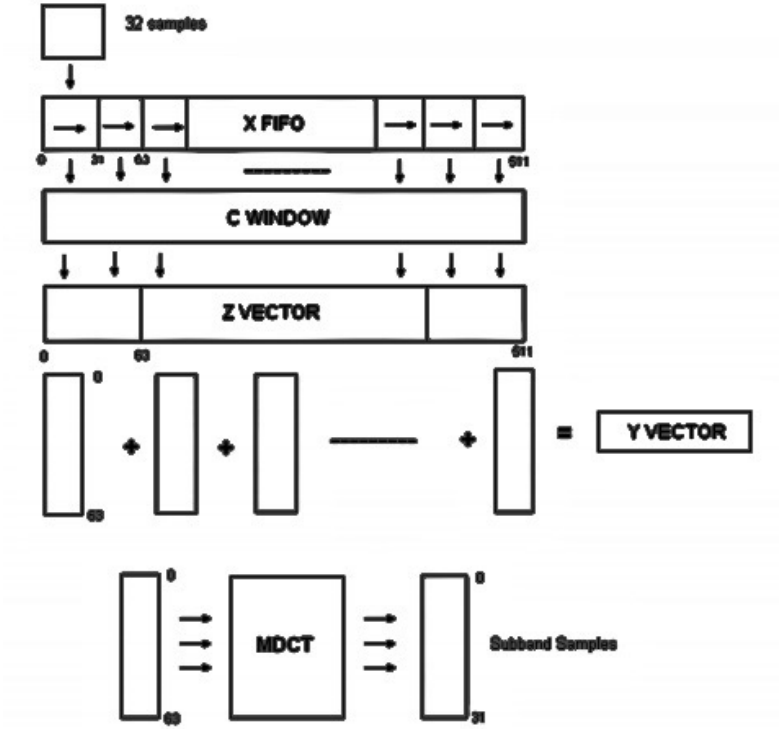
- Blocks of 12 for Layer 1 (384 samples).
- Blocks of 36 for Layers 2 and 3 (1152 samples)

For Layer 1, 32×12 = 384 samples @48 kHz represents 32×20.8µs×12 = 8 ms of audio



Sub Band Filtering: Polyphase Filter

Analysis



1. Shift in 32 input samples into FIFO X buffer

2. Window samples $Z_i = C_i \times X_i$

3. Partial computation Y_i

4. Compute 32 signals S_i

Psychoacoustic Model: Implementation

1. Fast Fourier Transform (FFT) is used to get detailed spectral information about the signal:
    - 512-point FFT for Layer 1
    - 1024-point FFT for Layer 2 and 3
2. From each subband's samples, separate tonal (sinusodial) and nontonal (noise) maskers in the signal.
3. Determine masking threshold for each sub-band.
4. Determine the global masking threshold.
5. Determine the minimal masking threshold in each subband.
6. Calculate the signal-to-mask ratio (SMR) in each sub band, which is the ratio of signal energy to minimum masking threshold.

Psychoacoustic Model: Implementation

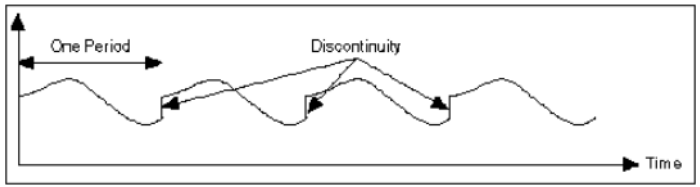
FFT computes DFT:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi nk}{N}} \qquad k = 0,\ldots,N-1$$

$$e^{-jx} = \cos x + j \sin x \quad \text{Euler's formula}$$

DFT inherently assumes that data is a single period of a periodically repeating waveform. Sampled values of the signal are multiplied by a (window) function that tapers toward zero at either end.

Data Windowing



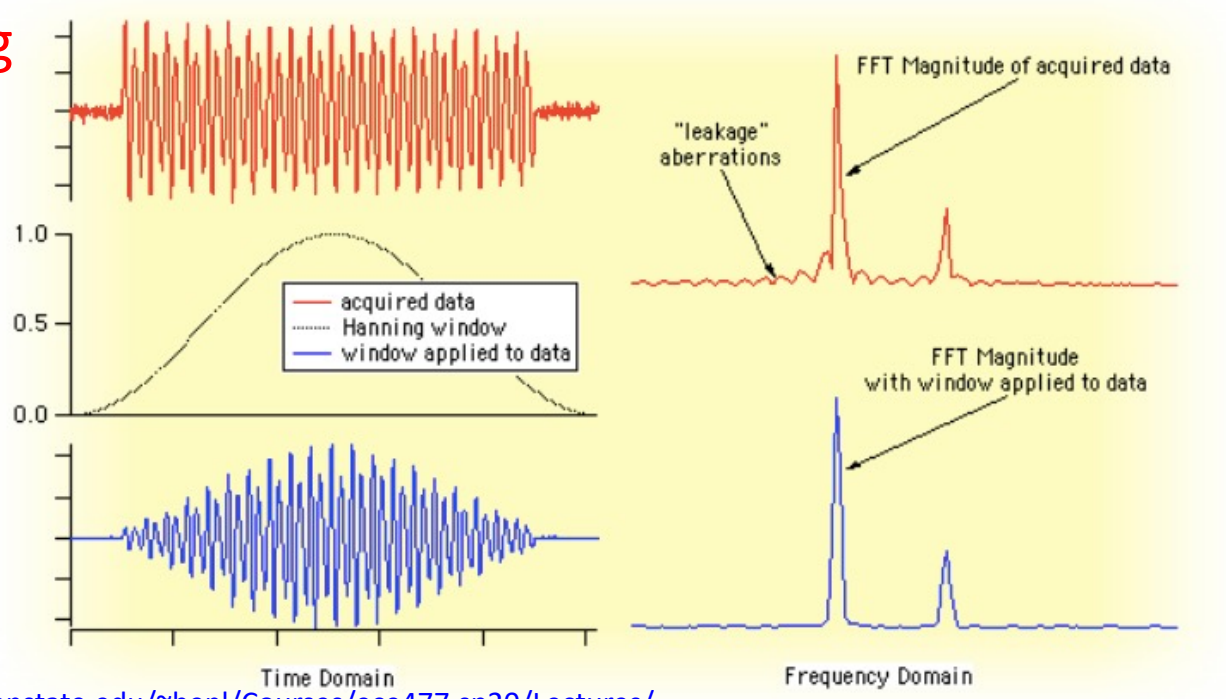
Psychoacoustic Model: Implementation

Data Windowing

Hann Window

$$w[n] = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N}\right)\right]$$

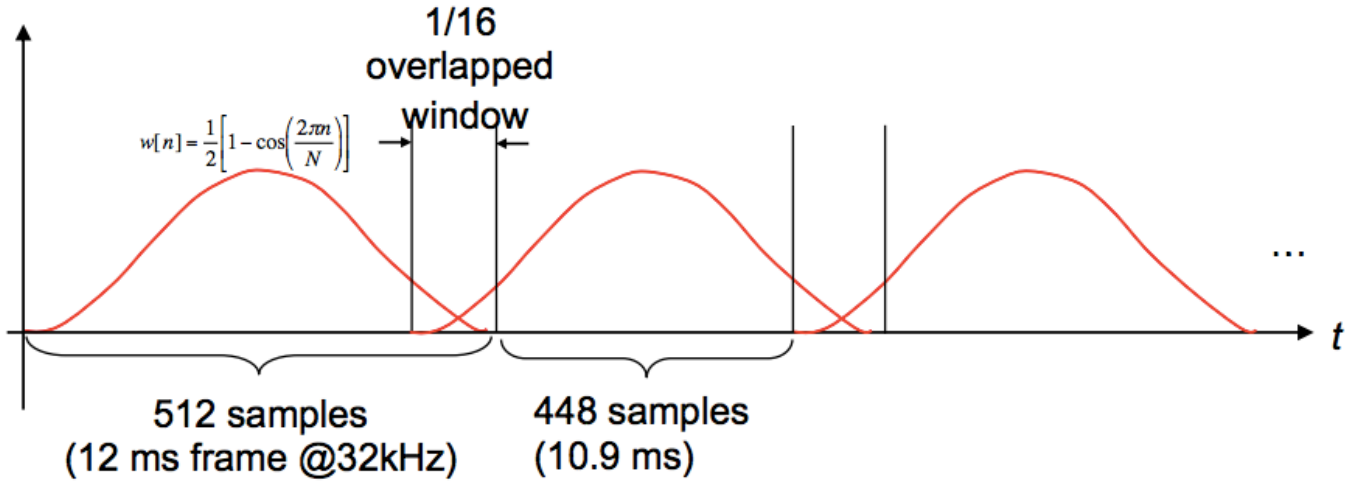
Fade-in fade-out



Source http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/

Psychoacoustic Model: Implementation

Data Windowing



Source http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/

Psychoacoustic Model: Implementation

Power Spectral Density

$$P(k) = \underbrace{90.302\ \text{dB}}_{\text{Power Normalization Term}} + \underbrace{10 \log \left| \sum_{n=0}^{N-1} \underbrace{w(n)}_{\text{Hann Window}} \underbrace{x(n) e^{-j\frac{2\pi kn}{N}}}_{\text{FFT}} \right|^2}_{\text{PSD}}\ \text{dB} \qquad 0 \le k \le \frac{N}{2}$$

Conversion to dB

Source http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/

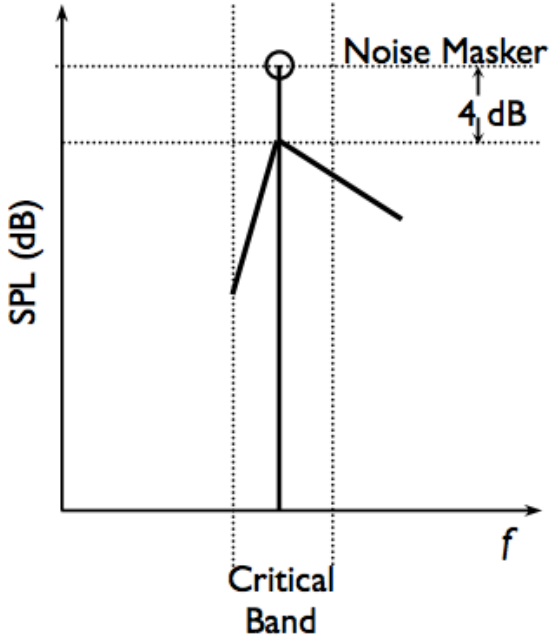
Psychoacoustic Model: Implementation

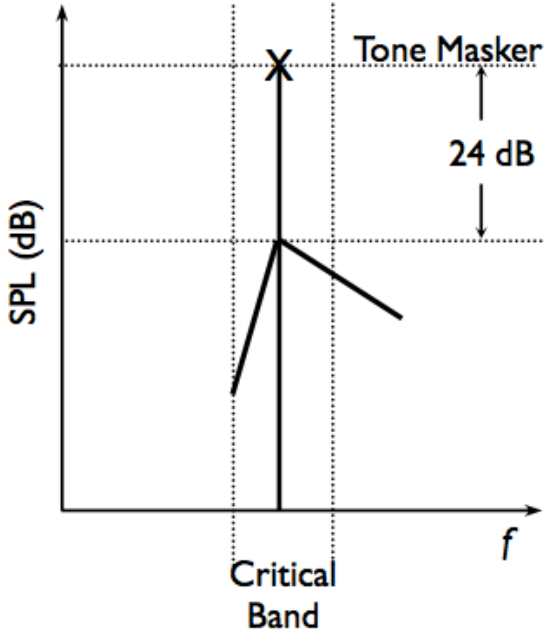
Tonal and Non-Tonal (Noise) Maskers

- Tonal maskers are signals that generate pure tone, i.e., harmonically rich.
- Noise maskers have no single dominant frequency, i.e., more noise like.
- Tonal and noise maskers have different masking characteristics.

Psychoacoustic Model: Implementation

Tonal and Non-Tonal (Noise) Maskers





Psychoacoustic Model: Implementation

Tonal and Non-Tonal (Noise) Maskers

A component is considered tonal if P(k)-P(k±Δk) ≥ 7 dB, where Δk is Bark range.

Tonal maskers, $P_{TM}(k)$

$$P_{TM}(k) = 10 \log \sum_{j=-1}^{1} 10^{0.1P(k+j)}$$

Energies from 3 adjacent spectral components centered around the peak are combined.

Noise maskers, $P_{TM}(k)$

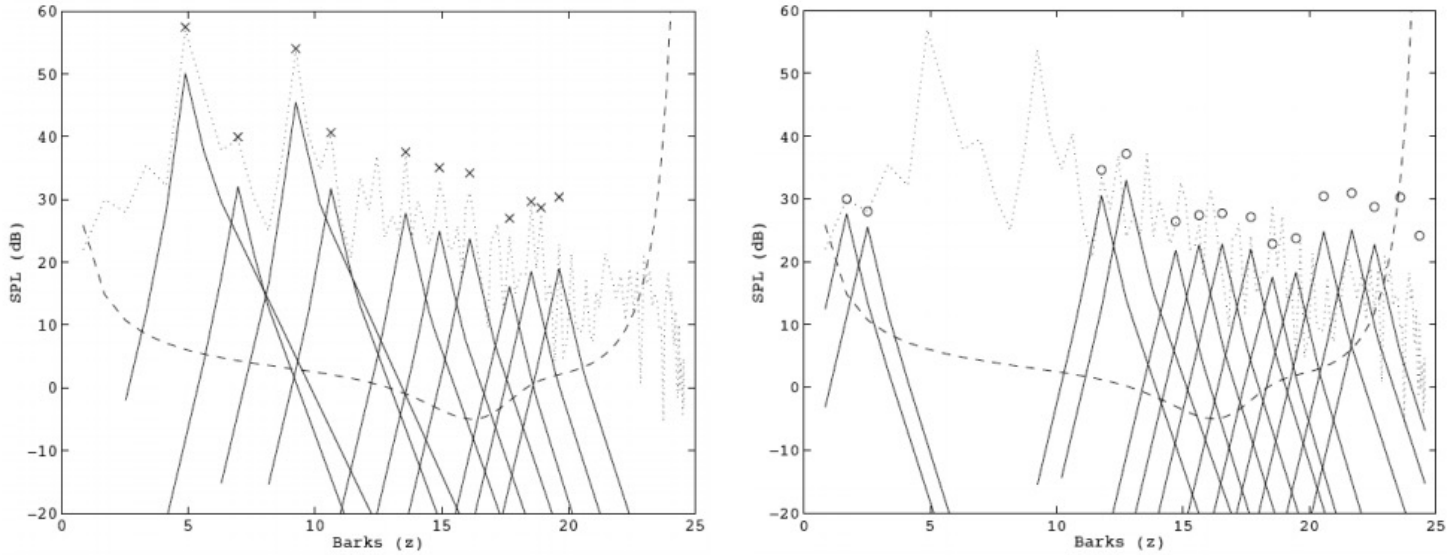
$$P_{NM}(k) = 10 \log \sum_{j} 10^{0.1P(j)}$$

Energies from all spectral components within each critical band are combined.

$P_{TM, NM}(k) \geq T_q(k)$, for the final thresholds the spread function is also considered

Psychoacoustic Model: Implementation

Tonal and Non-Tonal (Noise) Maskers



Tonal maskers

Noise maskers

Source http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/

Psychoacoustic Model: Implementation

Tonal and Non-Tonal (Noise) Maskers

Suppose the levels of the first 16 of the 32 sub-bands are:

Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Level (dB)	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1

The level of the 8th band is 60 dB, the pre-computed masking model specifies a masking of 12 dB in the 7^{th} band and 15 dB in the 9^{th}

- The signal level in 7th band is 10 (< 12 dB), so ignore it.
- The signal level in 9th band is 35 (> 15 dB), so send it.

Only the signals above the masking level need to be sent

Psychoacoustic Model: Implementation

Scaling

Block of 12 samples for each subband is scaled to normalize the peak signal level within a subband:

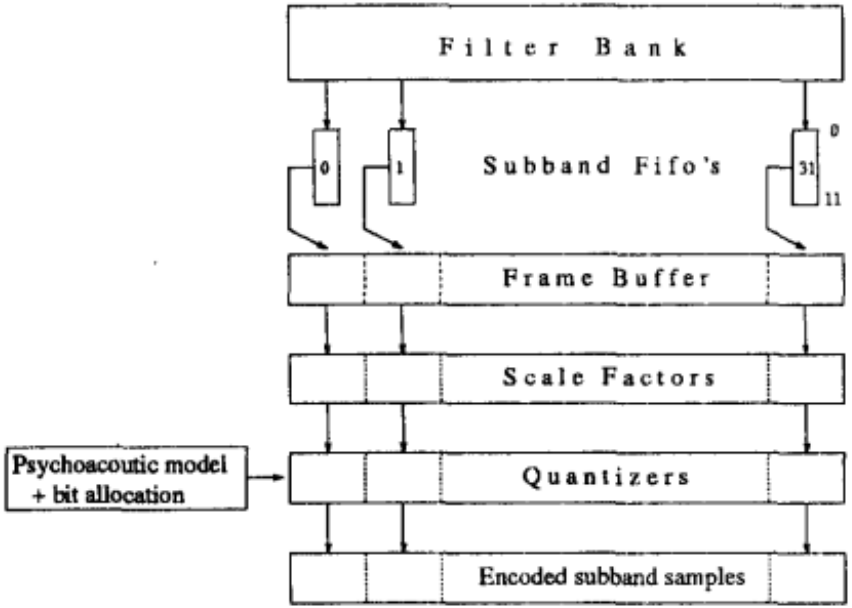
- Largest signal quantized using 6-bit scale-factor.

The receiver needs to know the scale factor and quantisation levels used:

- Information included along with the samples

Psychoacoustic Model: Implementation

Scaling



Psychoacoustic Model: Implementation

Bit Allocation

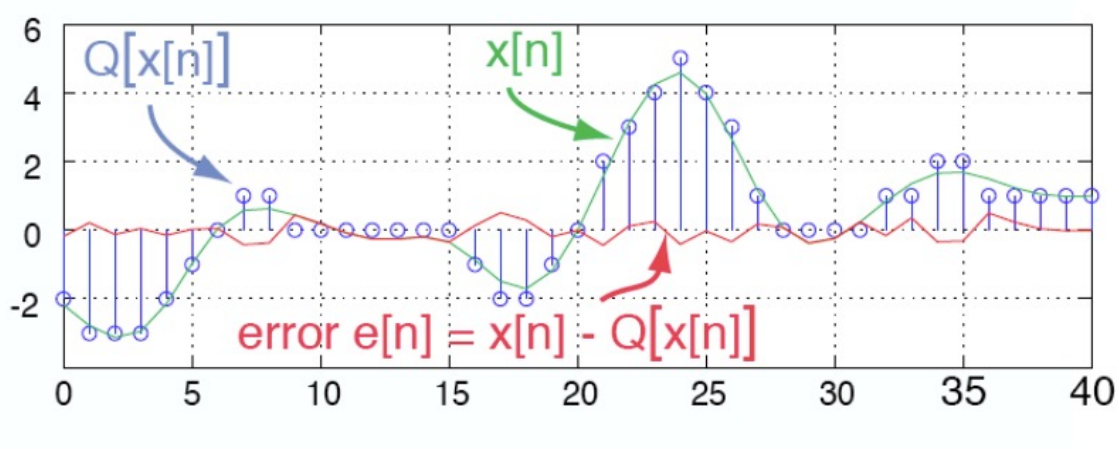
For each audio frame, bits must be distributed across the sub-bands from a predetermined number of bits defined by the target bit rate:

- 192 kbps target rate => 8 ms/frame @48kHz => 1.54 kbits/frame.

Objective is to minimize NMR, or maximize MNR, over all sub-bands, i.e., minimize quantization noise.

Psychoacoustic Model: Implementation

Bit Allocation



x[n] = Q[x[n]] + e[n]

SNR= 20log(x[n]/e[n])

- More bits allocated to quantization, smaller e[n]

Source http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/

Psychoacoustic Model: Implementation

Bitstream Organisation

Encoded Samples
- 384 for Layer 1, 1152 for Layers 2 and 3

Side Info
- Encoding parameters (Layer specific)

Frame Header (32 bits)
- Syncword
- Bit-rate and sampling information.
- Number of channels (1 or 2)
- Misc. other stuff.

Source http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/

MPEG-1: Sum up

- Layer 1
  - 384 samples per frame
  - Up to 448 kbps
- Layer 2
  - 1152 samples per frame
  - Up to 384 kbps
- Layer 3 (MP3)
  - Range of bit-rates from 8 kbps to 320 kbps
  - Better filterbank
  - Improved psychoacoustic model
  - Better bit-allocation process
  - Entropy coding
  - ...

References

1. NPTEL course on Multimedia Processing by Prof S Sengupta.
2. http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/
3. Guide to MPEG-1 Audio Standard, Seymore Shlien, IEEE Transactions on Broadcasting, Vol 40, No 4, Dec 1994