Audio compression

- Lossless compression (~1:3)
  - Predictive coding
- Lossy compression
  - Speech compression (~1:8)
  - Sound compression (~1:6)

Lossless audio compression

- Quality does not degrade
  - Predictive coding
  - Encoding (entropy/dictionary)
- Uncompressed audio (e.g. WAV)
  - Linear pulse-code modulation
  - Large size

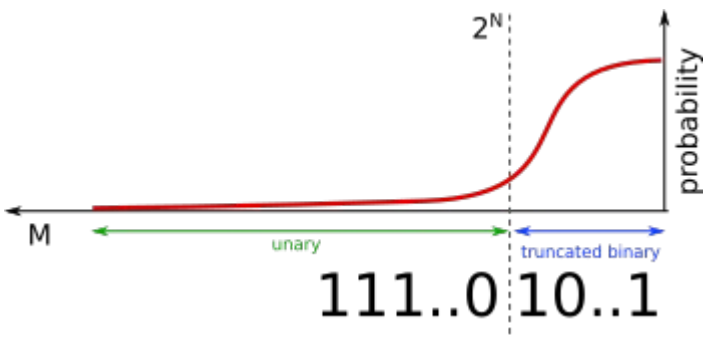
FLAC compression

- Free Lossless Audio Codec
- Reduction from 50% up to 80% (in some cases)
  - Linear prediction
  - Error coding (Golomb-Rice)
  - RLE
  - Stereo (inter-channel correlation)
- Data hierarchy
  - Before/after encoded: block, sub-block
  - Encoded: frame, sub-frame
  - Sub-frames share some encoding parameters

Prediction and residual

- Zero
  - Digital silence, constant value
  - RLE
- Verbatim
  - Zero-order predictor
  - Residual is signal itself
- Fixed Linear
  - Fitting p-order polynomial to p points
  - Efficient algorithm
- FIR Linear
  - Linear combination of previous samples
  - Slower, diminishing returns

- Golomb codes
  - Split value in two parts (divide by M)
  - Quotient – unary coding
  - Remainder – (truncated) binary coding
  - Efficient if small values dominate distribution
  - Rice coding – M is 2N

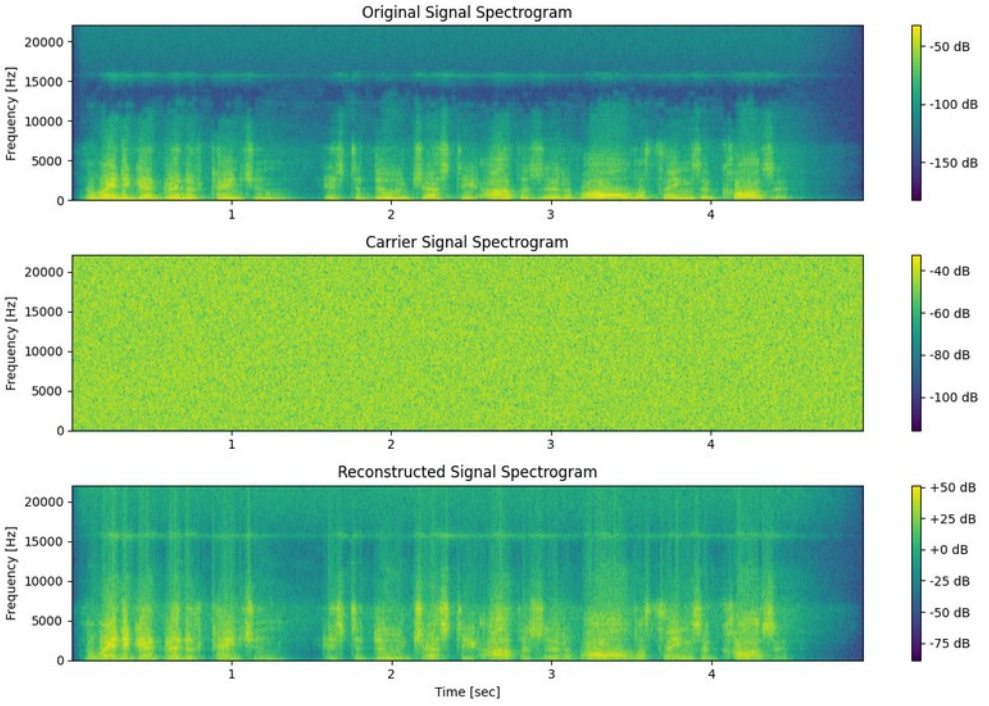


Speech compression

- Speech is formed by air traveling from lungs to mouth
  - Excitation signal formed by vocal folds opening and closing
- During speech the vocal tract shape is changing
  - Mouth opening and closing, tongue moving
  - This gives the excitation signal its spectral shape
- Speech coding approaches
  - Non-linear quantization
  - Multi-channel encoding (bands with different resolution)
  - Modeling speech (Source-filter model)

Vocoder

- Analyze spectral characteristics of one sound and applying them to another sound
  - Modulator – typically voice
  - Carrier - typically a synthesized waveform
- Common carrier options
  - White noise
  - Sine wave – periodic



Vocoders for compression

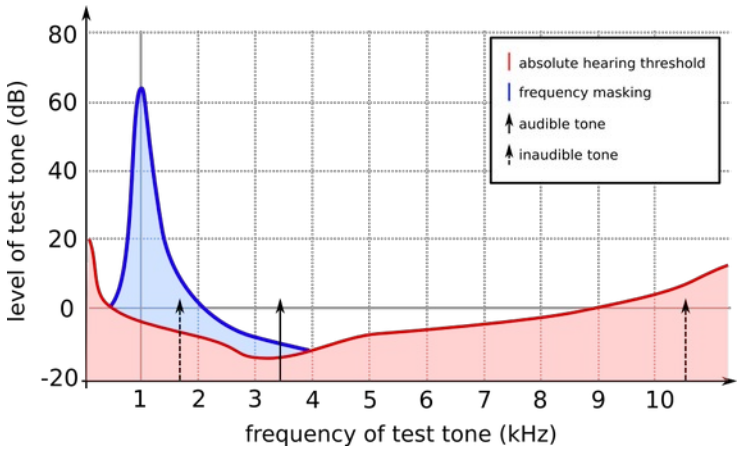
- LPC-10 vocoder
  - Linear prediction based on past model coefficients
  - Two modes (voiced - periodic waveform, unvoiced – noise generator)
  - Gain, pitch estimation
  - Sound processed in frames (30-50 frames per second)
  - Intended bandwidth: 2.4 kbps (GPS)
- CELP vocoder
  - Analysis by synthesis (optimizing resulting signal perceptual error in closed loop)
  - Short-term prediction (LPC analysis) + Long-term prediction (codebooks)
  - Intended bandwidth: 4.8 kbps
  - MPEG-4 Audio

MPEG-1 Audio Compression

- Three layers of compression
  - Downwards compatibility
  - Each more complex (encoder)
  - Quality depends on available space (bitrate)
- Main concepts
  - Frequency domain (DCT)
  - Non-uniform quantization - bands
  - Imperfections in human auditory system

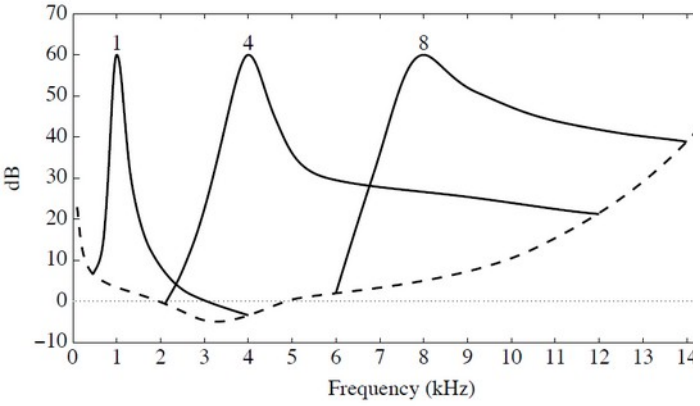
Frequency masking

- Core compression mechanism in MPEG-1
- Louder tone masks silent tones nearby
  - Silent tones are not perceived
  - Lower frequencies mask higher ones better
  - Louder the sound, more tones it masks



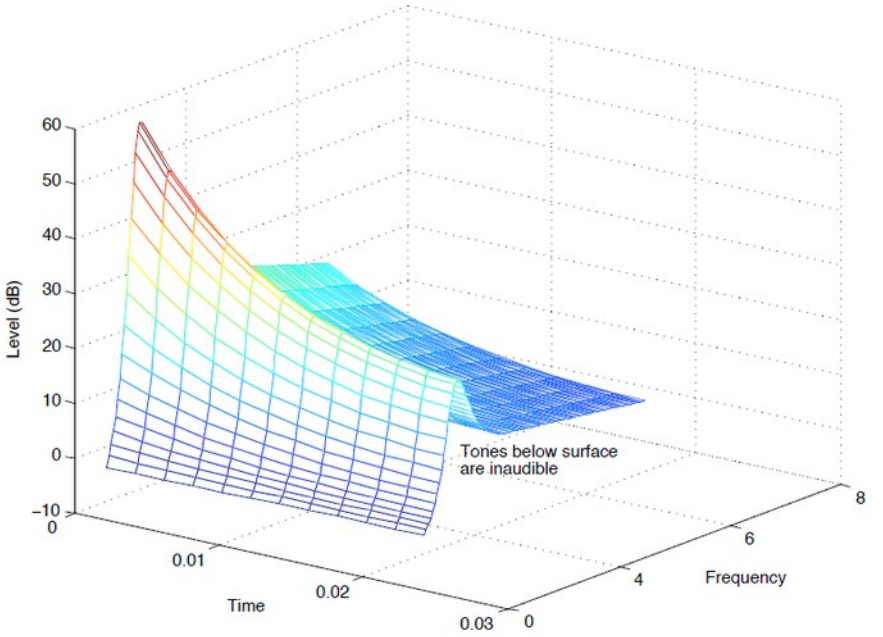
Critical bands

- Groups of hair cells respond to frequency range
  - Within the band a strong frequency overwhelms cells
  - Other frequencies are not detected
- About 24-25 critical bands
  - Sound will seem louder if it spans two bands
- Perceptual non-uniformity
  - Bandwidth constant below 500Hz (100Hz)
  - Bandwidth linearly increases above 500Hz



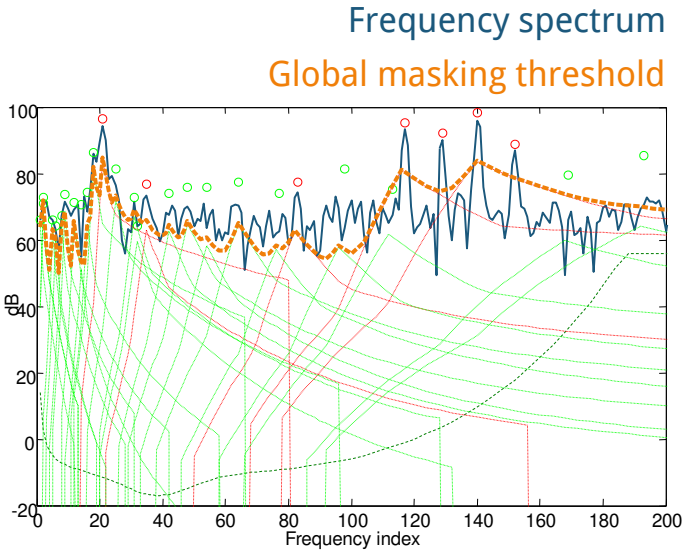
Temporal masking

- After a loud sound it takes time to hear quiet sound
  - Hair cells need a time-out
  - Duration depends on time and frequency similarity



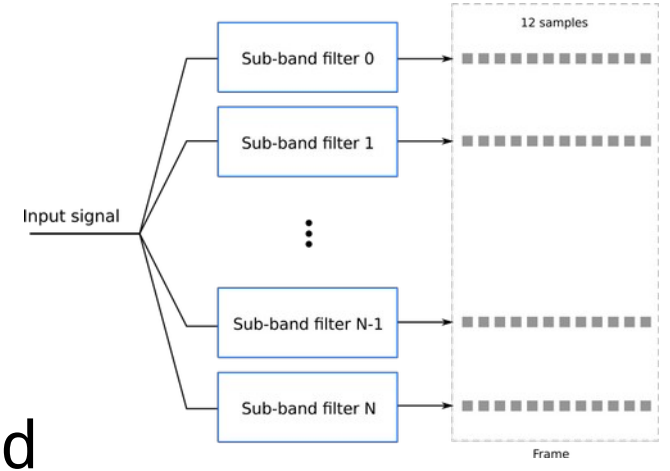
Psycho-acoustical model

- Compute masking levels
  - Frequency masking + absolute hearing threshold = global masking threshold (GMT)
- Signal-to-Mask Ratio, Mask-to-Noise Ratio
  - SMR = Signal – GMT (how much is signal louder than mask)
  - MNR = SNR – SMR (SNR based on quantization levels)
- Bit allocation
  - Quantization noise below GMR
  - Not always possible (low bitrate)
  - Distribute bits across bands based on MNR



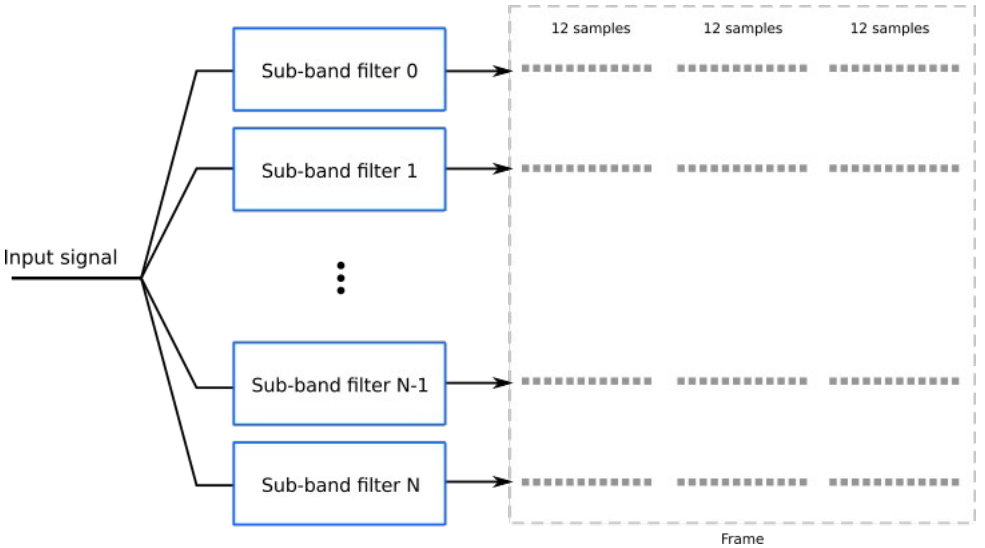
MPEG-1 Layer 1

- Intention: stored audio
- Critical band analysis
  - 32 filters, using FFT
  - Equal frequency spread per critical band
  - 12 samples in frame (scaled and quantized)
- Perceptual model
  - Only frequency masking



MPEG-1 Layer 2

- Intended use: digital audio broadcast
- Improvements
  - Use three groups together (3x12 samples)
  - Groups can share scaling factors
  - Supports frame skipping and sharing
  - Basic temporal masking



MPEG-1 Layer 3

- Intention: audio transmission over ISDN
- Improvements
  - Using MDCT instead of FFT
  - Non-uniform critical bands
  - Analysis-by-Synthesis noise allocation
  - Temporal masking
  - Stereo redundancy
  - Huffman coding

Stereo signal

- Intensity stereo coding:
  - Single summed signal and scale factors
  - Same signal, different magnitudes
- Middle/Side Stereo Mode:
  - Middle (sum of L and R) - M
  - Difference between channels - S

Decoding MPEG1 audio

- Psychoacoustic model is not required
- Quantization levels and scaling factors are used to reconstruct frequency bands
- Inverse frequency domain transform gives us waveform for decoded segment

MPEG-1 Overview

Layer	Target bitrate	Ratio	Quality @ 64 kbit	Quality @ 128 kbit	Theoretical min. delay
1	192 kbit	4:1	/	/	19 ms
2	128 kbit	6:1	2.1 – 2.6	4+	35 ms
3	64 kbit	12:1	3.6 – 3.8	4+	59 ms

Perceptual quality: 5 – perfect … 1 – very annoying

MPEG-2 Part 3 audio coding

- Backwards compatible
- Extensions
  - Multi-channel coding – 5.1 channel audio
  - Multilingual coding - 7 multilingual channels
  - Lower sampling frequencies
  - Optional Low Frequency Enhancement

MPEG-2 Part 7

- Advanced Audio Coding
- Not backwards compatible
- Increased complexity
- Up to 48 channels
- Used on DVD