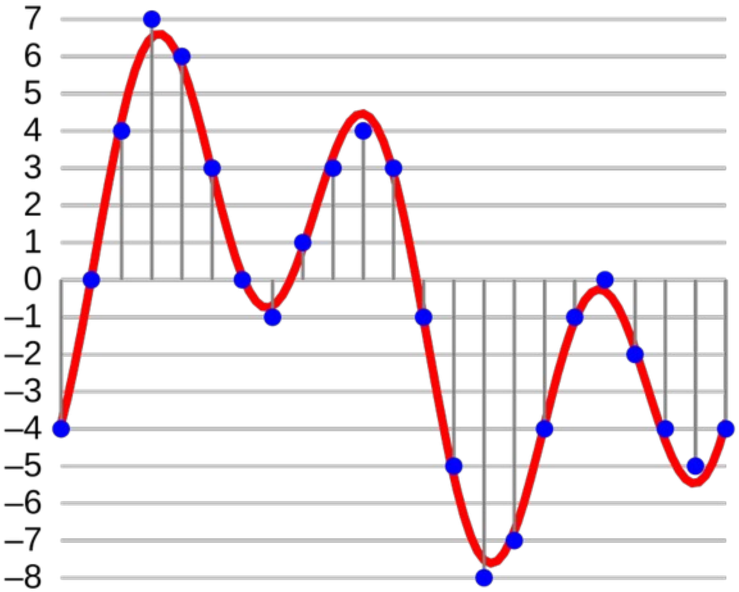
Audio Processing

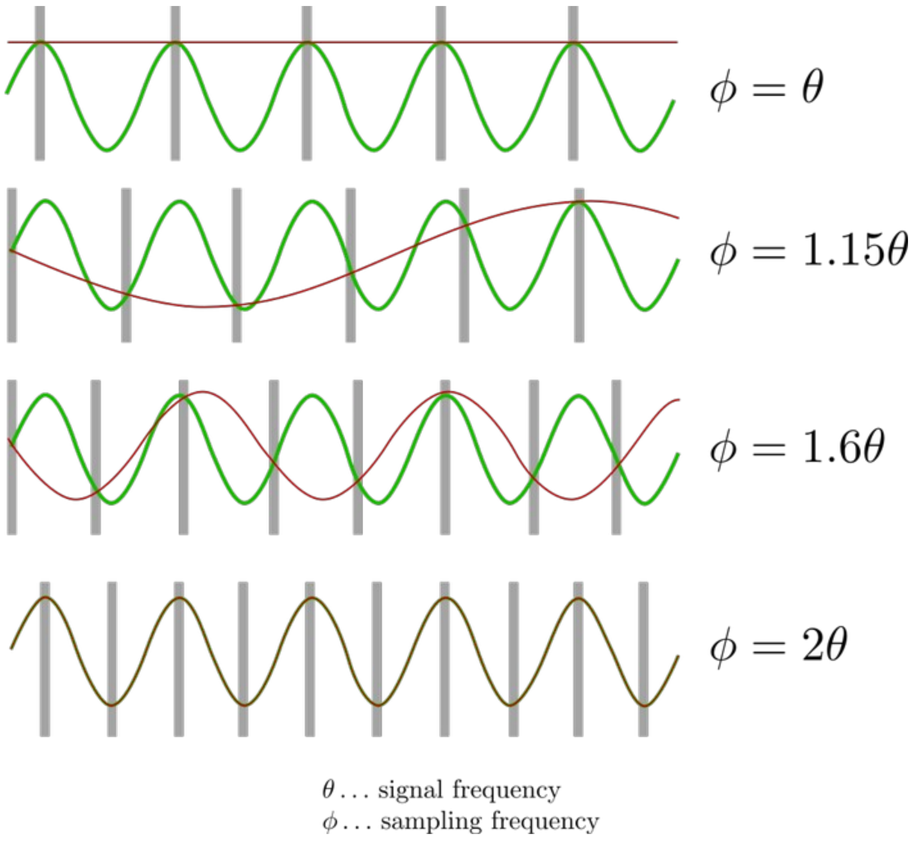
Digital audio

- Digital processing and storage
- Source is an analog signal
- Sampling – observe pressure at fixed intervals
  - Aliasing – artifacts due to low sampling
- Quantization – encode values with fixed interval of integers
  - Quantization noise – loss due to rounding



Signal sampling

- Uniform sampling (Dirac comb)
- Nyquist-Shannon theorem
  - Avoid aliasing
  - Band-limited signal
  - Sample rate twice the maximum frequency
  - Low pass filter (< f/2) + Sampling with frequency f



Signal-to-noise (SNR)

- Random fluctuations in signal (noise)
  - Ratio between power of signal and noise (voltage)
  - Measured with decibels (tenth of a *bel*)

$$SNR = 20log_{10}\frac{V_{signal}}{V_{noise}}$$

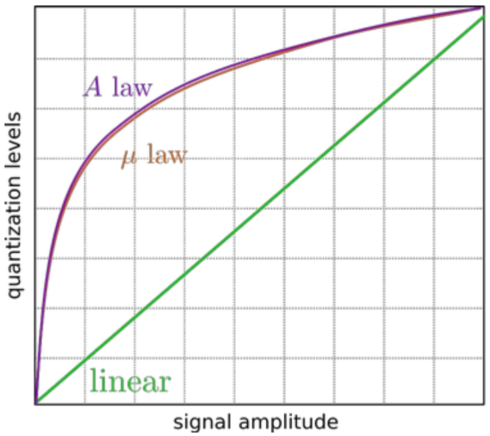
- Everyday usage
  - Comparison to just-audible sound of 1kHz
  - Conversation: 60 dB
  - Train: 90 dB
  - Pain: 140 dB

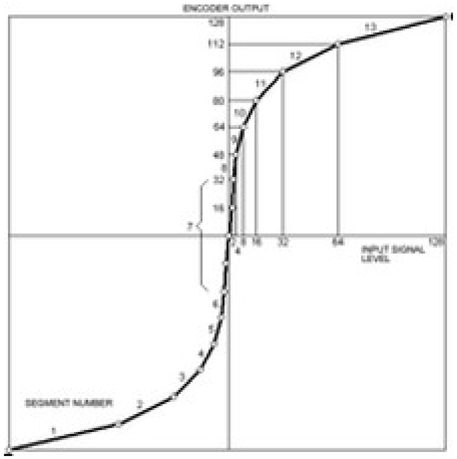
Signal quantization

- Assign integer values to measured ones
  - [-V … V] → [0 … N]
  - Quantization error (rounding)
- Signal-to-quantization noise (SQNR)
  - Higher is better (more signal vs. noise)
  - Worst case (peak signal) $SQNR = 20\log_{10} 2^Q = 6.02 \times Q(dB)$
  - Statistical independence $SQNR = 6.02 \times Q + 1.76(dB)$
- 12 bit fine for adequate reproduction
- Dithering
  - Small amount of noise added to the signal
  - Quantization errors are more random and less noticeable

Pulse-code modulation

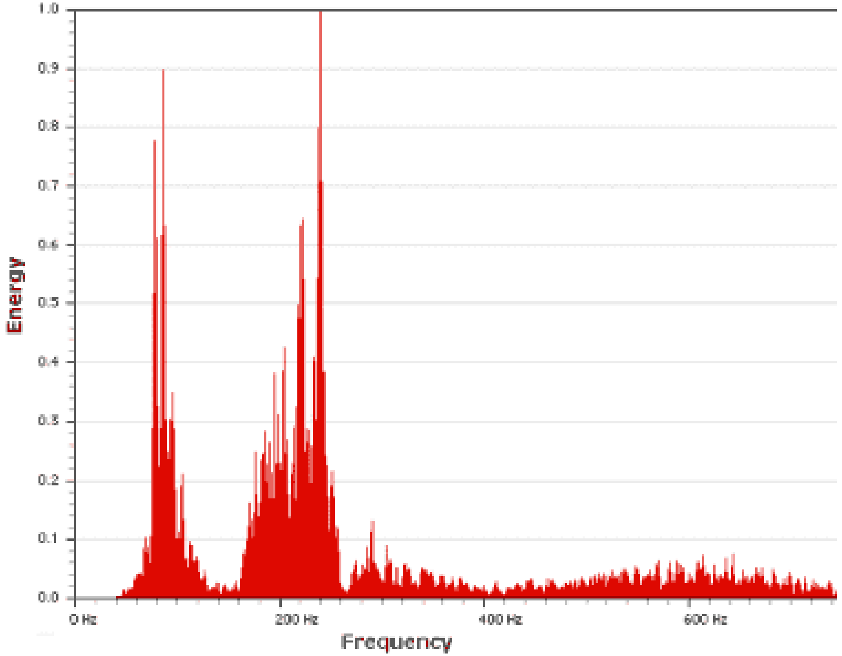
- Formal term for sampling + quantization
- Linear quantization
  - Uniform levels
- Non-linear quantization
  - Better signal-to-noise ratio for low-amplitude signals
  - A-law, μ-law algorithm
- (Adaptive) Differential Pulse Code Modulation
  - Encode difference to previous value
  - Encode difference to predicted value





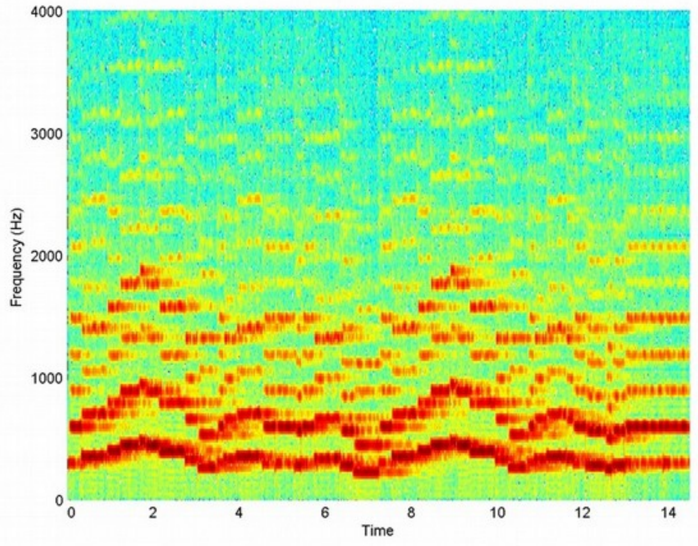
Frequency spectrum

- Linear combination of basis functions
  - Sinusoidal (sine and cosine) – repeatability
  - Coefficients – presence of individual basis functions



Short-term Fourier Transform

- Computed for a time window
  - FT – captures overall signal properties
  - STFT - Captures changes in the signal
- Windowing function
  - Type, size, hop
- Visualization
  - Spectrogram
  - Log-scale, colormap



Analog vs. digital audio processing

- Analog processing
  - Continuous signal – electrical current or voltage
  - Processing done via electronic components
- Digital processing
  - Sampled signal
  - Processing done on general purpose computers
  - More powerful and efficient

Common digital sound parameters

Use case	Sampling rate	Bits per sample	Frequency band (Hz)
Telephony (GSM)	8.000 Hz	8	200-3400
G.722 (voice over IP)	16.000 Hz	14	50–7000
Audio CD	44.100 Hz	16	5-20k
Digital Audio Tape, Dolby AC-3	48.000 Hz	16	5-20k
High-quality acquisition and reproduction. Supported in DVD-Audio , Dolby TrueHD, DTS-HD ...	96.000 192.000 Hz	24 (max)	0-96k (max)

Audio filter

- Analog audio filter
  - Medium that transmits and modifies audio signal
  - Electronic components
  - Speakers – cannot cover entire spectrum
  - Mouth cavity – changing shape
- Digital audio filter
  - Algorithm that operates on digital signals
  - Approximation of analog filters
  - Better SNR

Filter taxonomy

- Linear filters / Non-linear filters
  - Is the output result of a linear difference equation?
  - Non-linear filters create additional frequency components, not present in the original signal
- Causal / Non-causal
  - Is the output result only of past values?
- Time-invariant / Time-variant
  - Is the output the same if we send it to the filter a bit later?

Linear filters

$$\underbrace{y(n)}_{\text{Output}} = \underbrace{b_0x(n) + b_1x(n-1) + b_2x(n-2)}_{\text{Input}} + \cdots - \underbrace{a_1y(n-1) - a_2y(n-2) - \cdots}_{\text{Past output}}$$

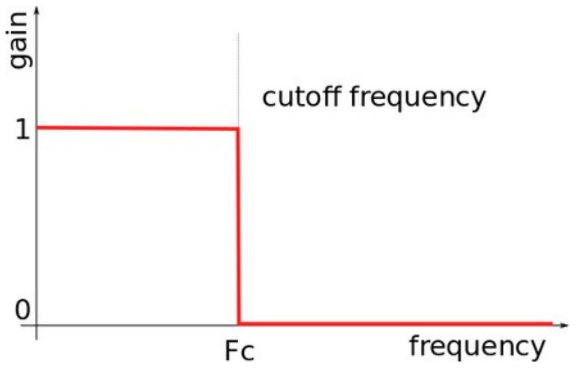
- Non-recursive filters (FIR):
  - Finite response $a_n = 0; \forall n > 1$
- Recursive filters (IIR)
  - Potentially infinite response
  - Implementations more compact

Frequency-response analysis

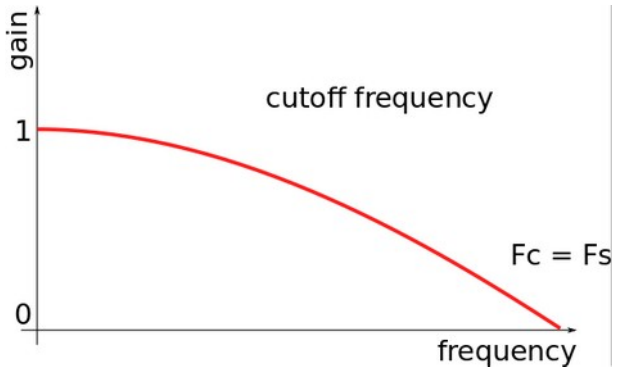
- Compare input and output frequency spectrum
- Only possible for LTI filters
  - Retain sinusoidal signal
  - Observe parametric properties of elementary inputs

Low-pass filter

- Ideal low-pass filter
  - Requires infinite signal
- Simple low-pass filter
  - Cutoff is Fs / 2
  - Transition is very gradual
- Low-pass filter design
  - Delay (see into the future)
  - Computational complexity

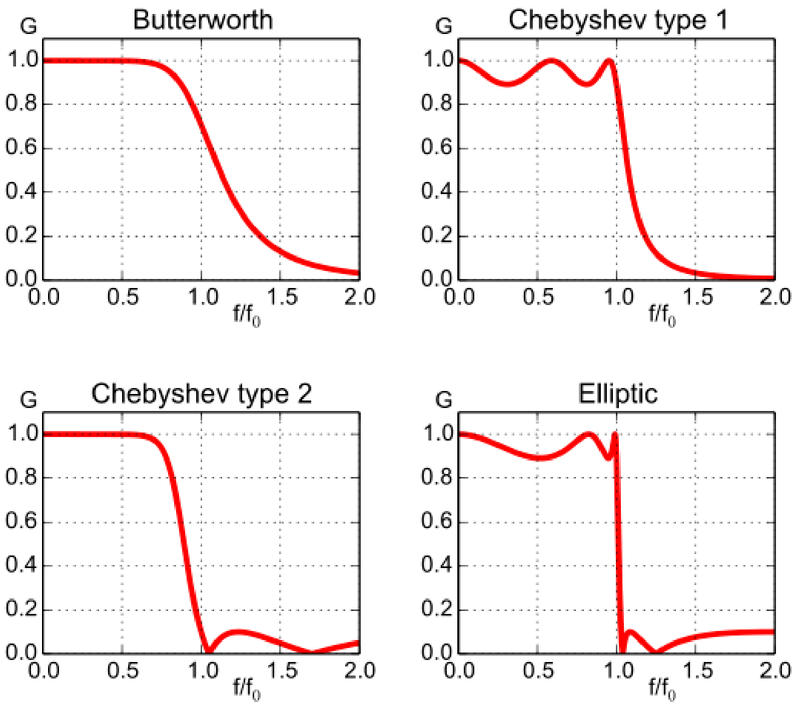


$$y(n) = x(x) + x(n-1)$$



Low-pass filter implementations

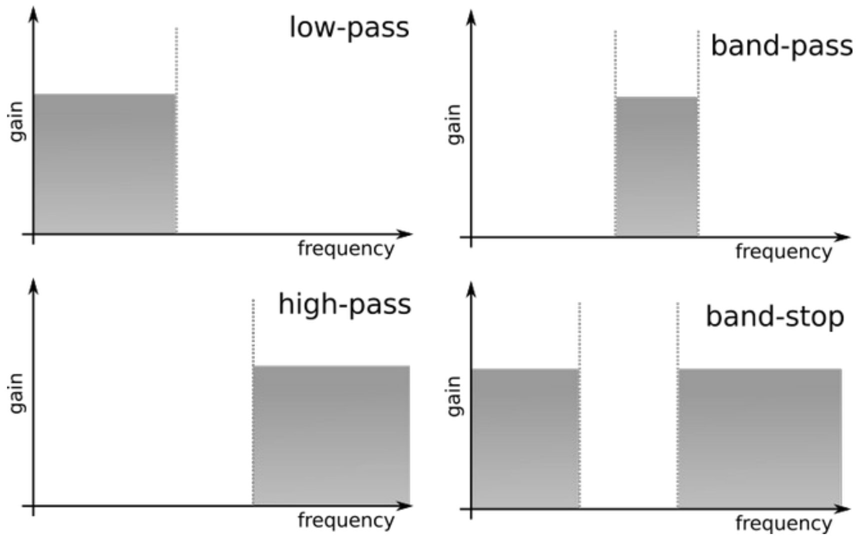
- Butterworth
- Chebyshev
- Elliptic



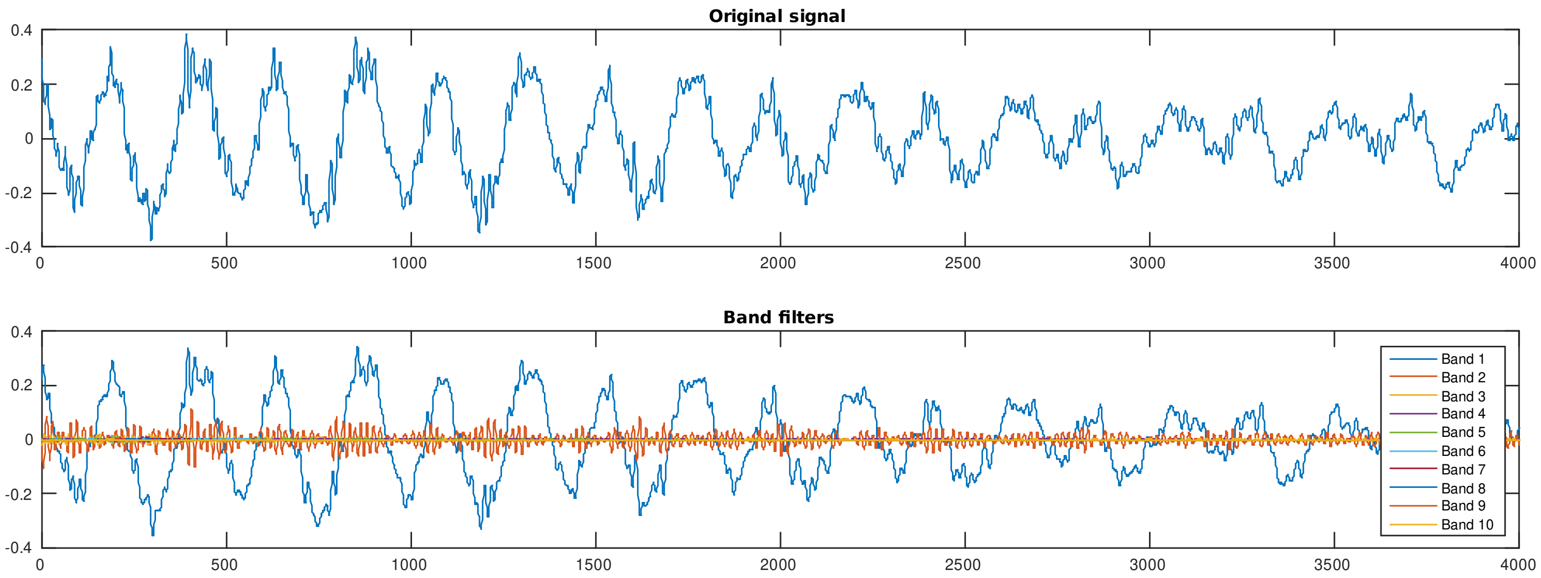
Source: Wikipedia (Geek3)

Frequency filtering

- Combinations of prototype low-pass filter
- Pass only frequencies in passband
  - Low-pass
  - High-pass
  - Band-pass
  - Band-stop

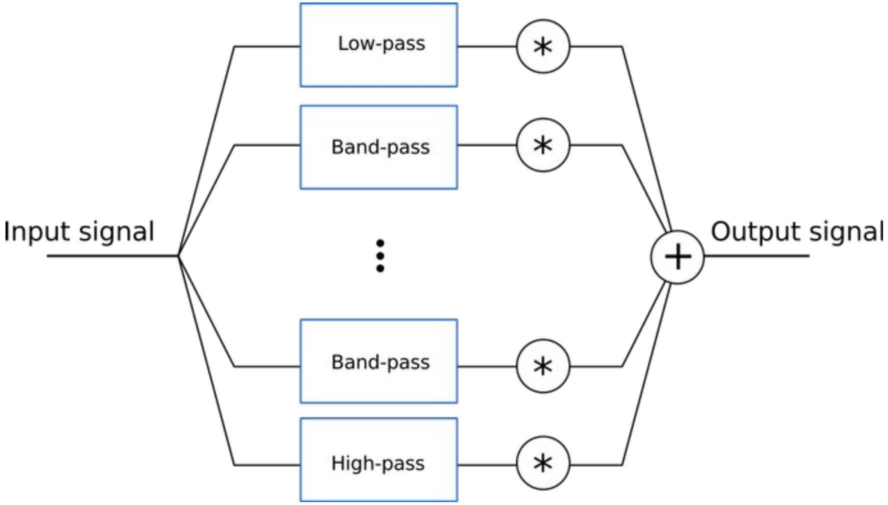


Splitting frequency bands



Audio equalization

- Multi-band signal can be combined back
  - Split signal into multiple bands
  - Weight individual bands
  - Combine signal again

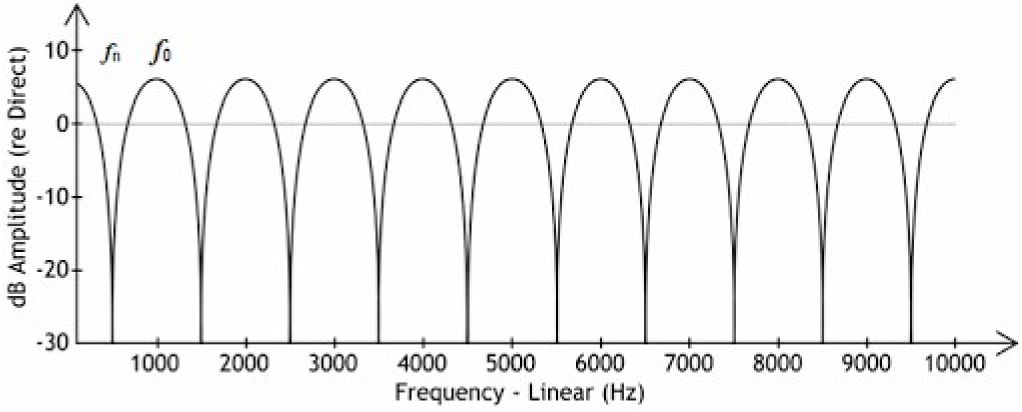


Comb filter

- Basic building block of many effects

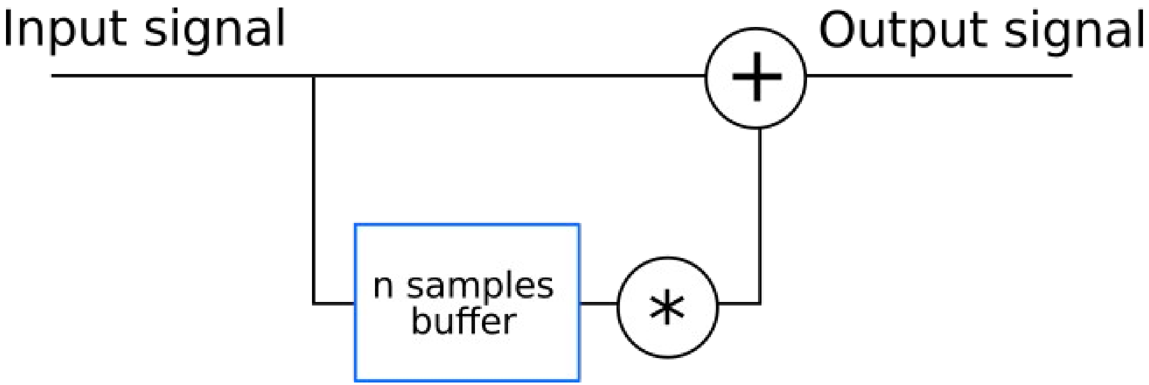
$$y(n) = x(n) + \alpha_1 x(n - t_1) + \alpha_2 y(n - t_2)$$

- Constructive, destructive interference
- Feedforward (FIR)
- Feedback (IIR)



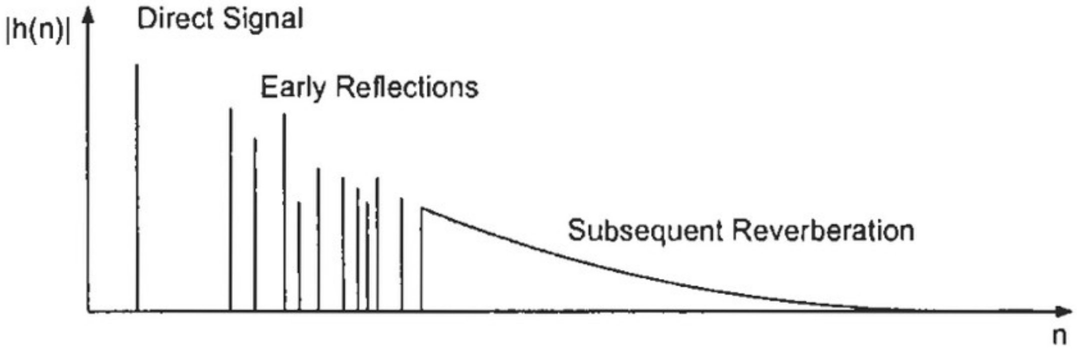
Delay

- A time-shifted signal is added to the original
- Delay below 50-100ms is not perceived as delay



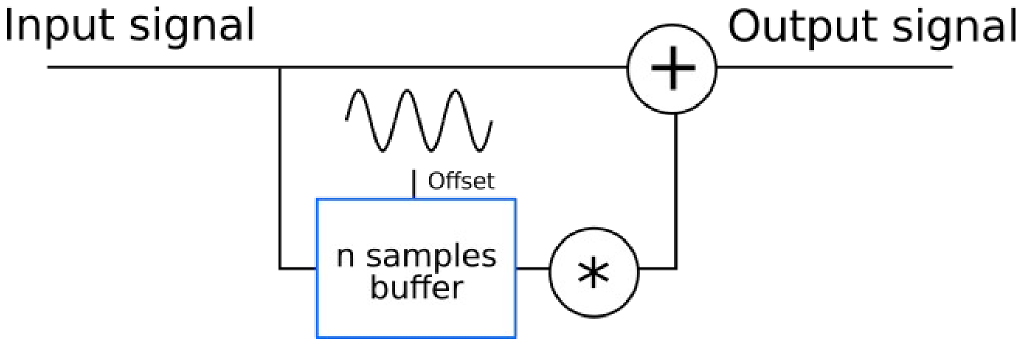
Echo

- Echo filter – simulate acoustics of a room
  - Multiple decaying delays – early reflections
  - Subsequent reverberation – random signal



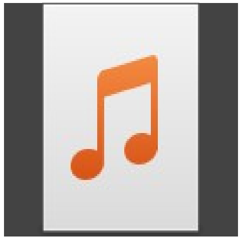
Flanger and phaser

- Delay varies with low frequency
- Feedback loop
- Time-variant filter



Chorus

- Sounds with almost same time and similar pitch
- Naturally: choir, string orchestra
- Similar to flanger
  - Longer delay times
  - Different combing effect



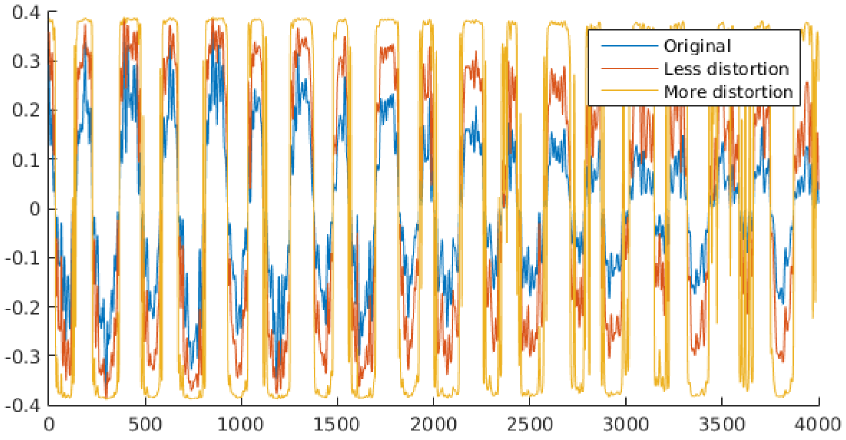
Non-linear filters

- Not describable by a LDE
- New frequencies in the signal
- Noise removal
  - Median filter
- Dynamic range compression
  - Companding – handling dynamic range
  - Clipping – amplitude clipping
  - Distortion – non-linear amplitude modificaton

Distortion

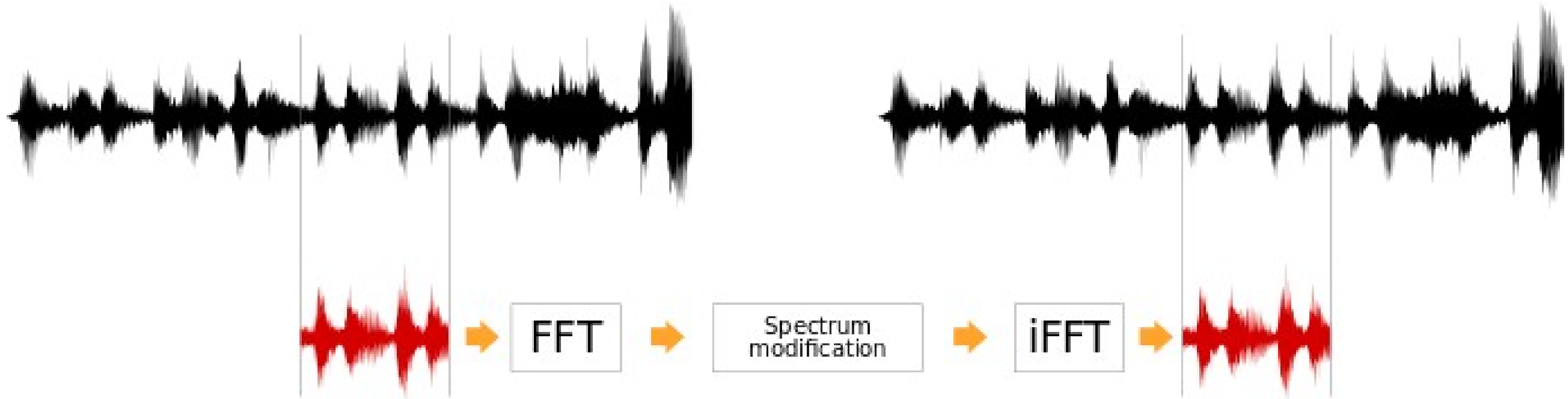
- Overdriven guitar effect
- Clipping of high energy frequencies
  - Soft clipping
  - Hard clipping

$$y(n) = \frac{(1+k)x(n)}{1+k|x(n)|}$$



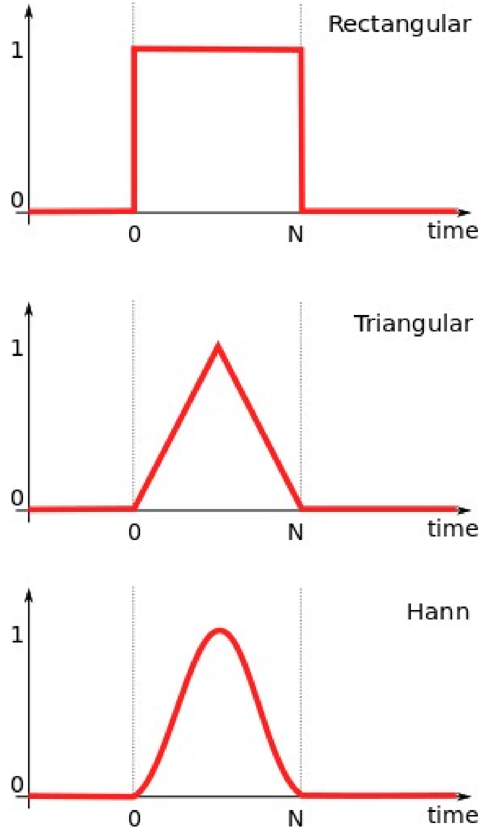
Spectrum Analysis

- Process signal in small segments
  - Simulates human perception (10-20ms segments)
  - Computationally efficient (STFT)
  - Transient events (e.g. percussion)
  - Resource efficient, near real-time



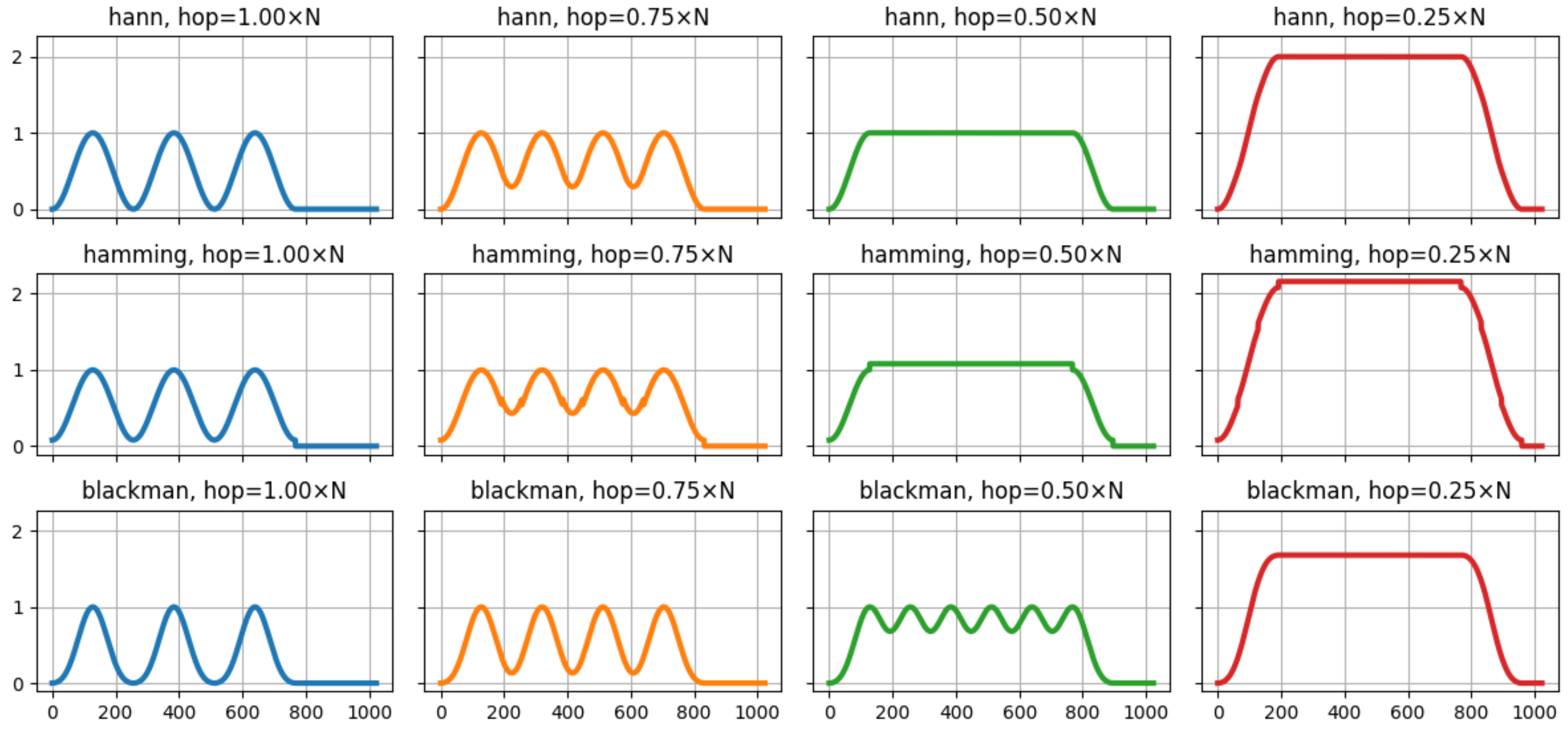
Extracting a segment

- Window function
  - Filters out a short segment of signal
  - Zero outside the interval
  - Symmetric (typically)
- Spectral leakage
  - New frequency components
  - Time-variant functions (windows, Dirac)
  - Side-lobes, cross-talk
  - Windows distribute leakage differently



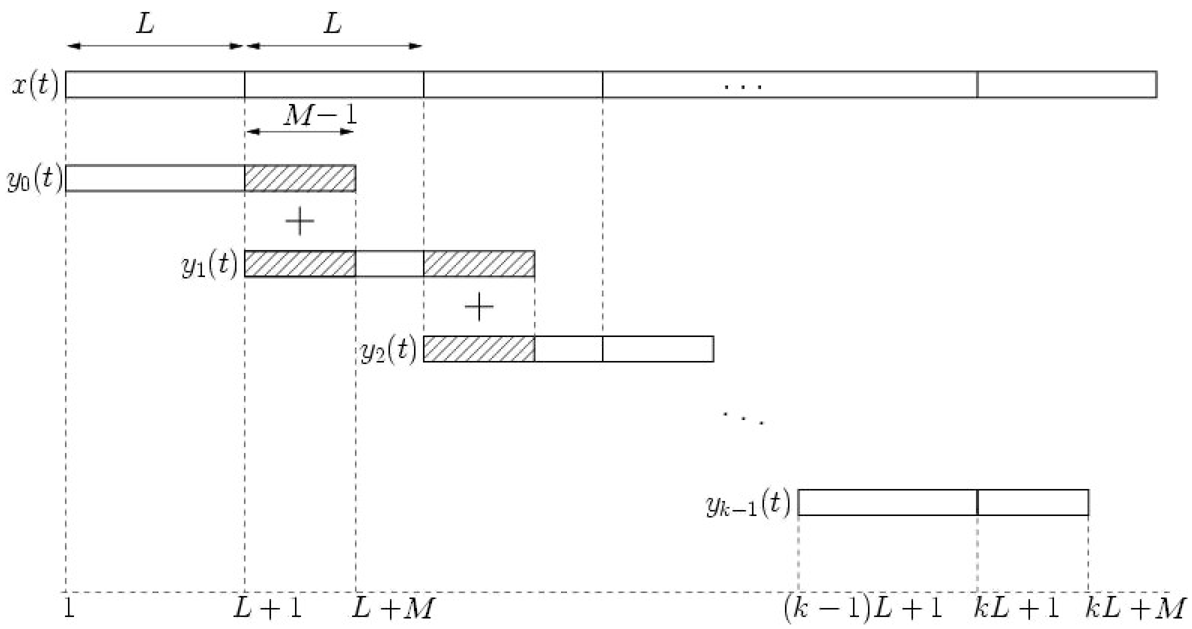
Distributing leakage

- Preserve amlitude envelope



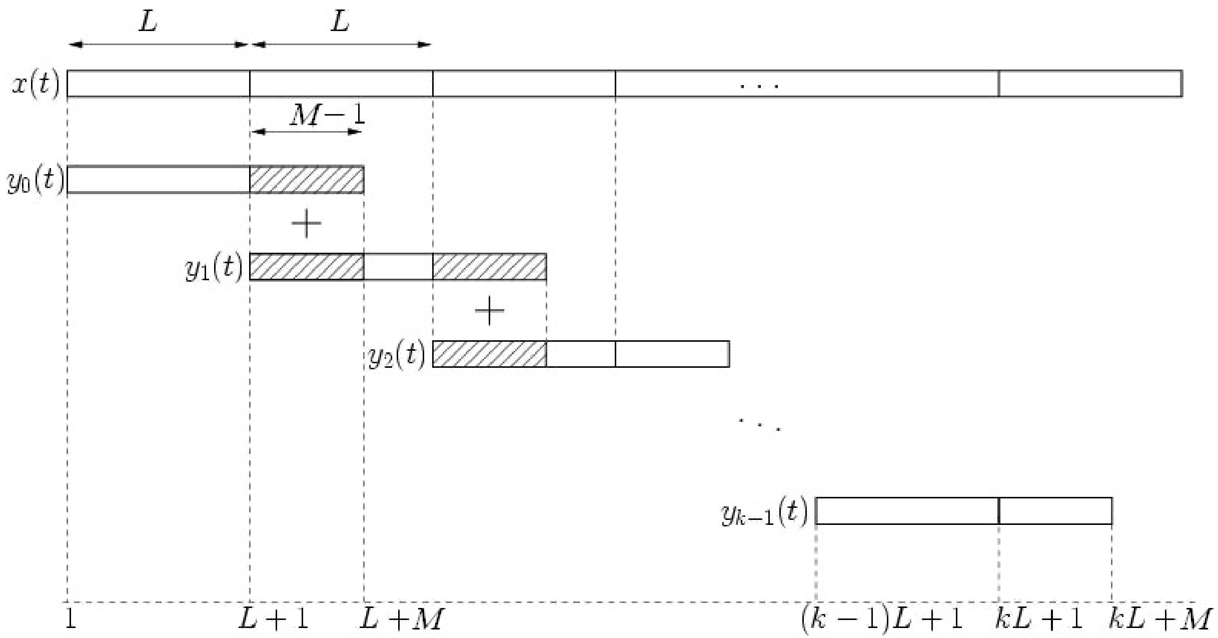
Overlap-Add

- Combine frames back to a signal
- Constant Overlap-Add
  - Hop size
  - Overlap size



Convolution with Overlap-Add

- Efficient convolution implementation
  - Interval size L
  - Filter size M
- Convolution is multiplication in Fourier domain

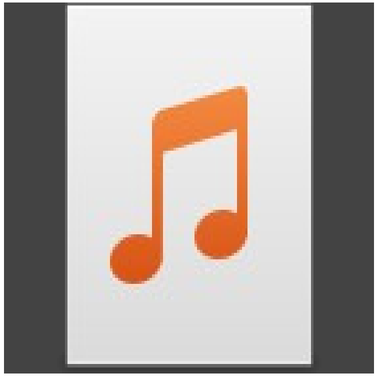


Filter banks

- Set of filters that decomposes signal into components
  - Efficiently implemented using STFT
  - Shared computation
- Vocoders
- Compression (MPEG)

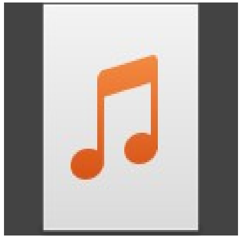
Changing playback speed/pitch

- Sample rate conversion
  - Timing is changed
  - Transposed pitch (chipmunk effect)
- Overlap-Add
- Process windows (STFT)



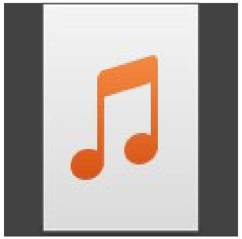
Time stretch

- Short, smoothly windowed block of samples
  - Select windows that can be combined with larger/smaller overlap
  - Combine blocks (overlap-add)



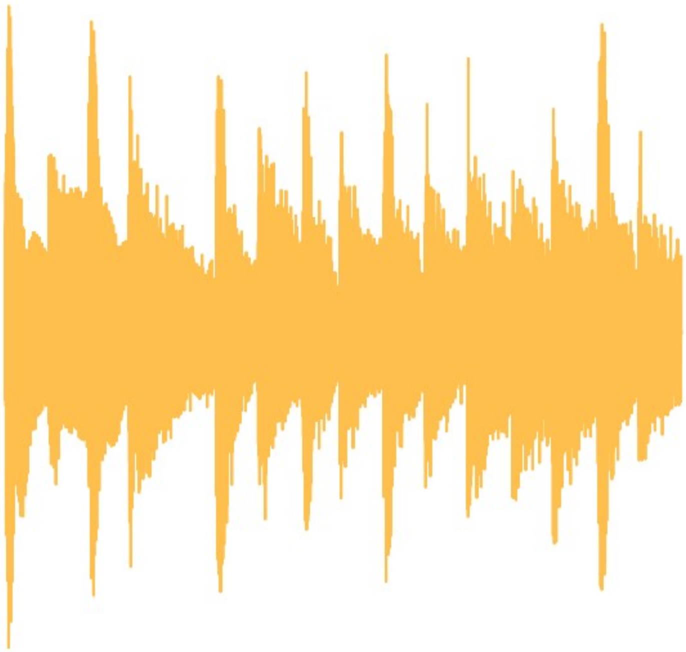
Pitch shift

- Preserve time, shift pitch (frequencies)
- Using time stretching
  - Stretch time
  - Re-sample to original length



Transformation Comparison

Audio



Image

Summary

- Sensory data is represented as a signal
  - Shared underlying concepts
  - Data with respect to space and/or time
- Take into account human perception
  - Some changes are more noticeable, pleasing …
- Next time: **Compression in Multimedia**