Audio

What is sound?

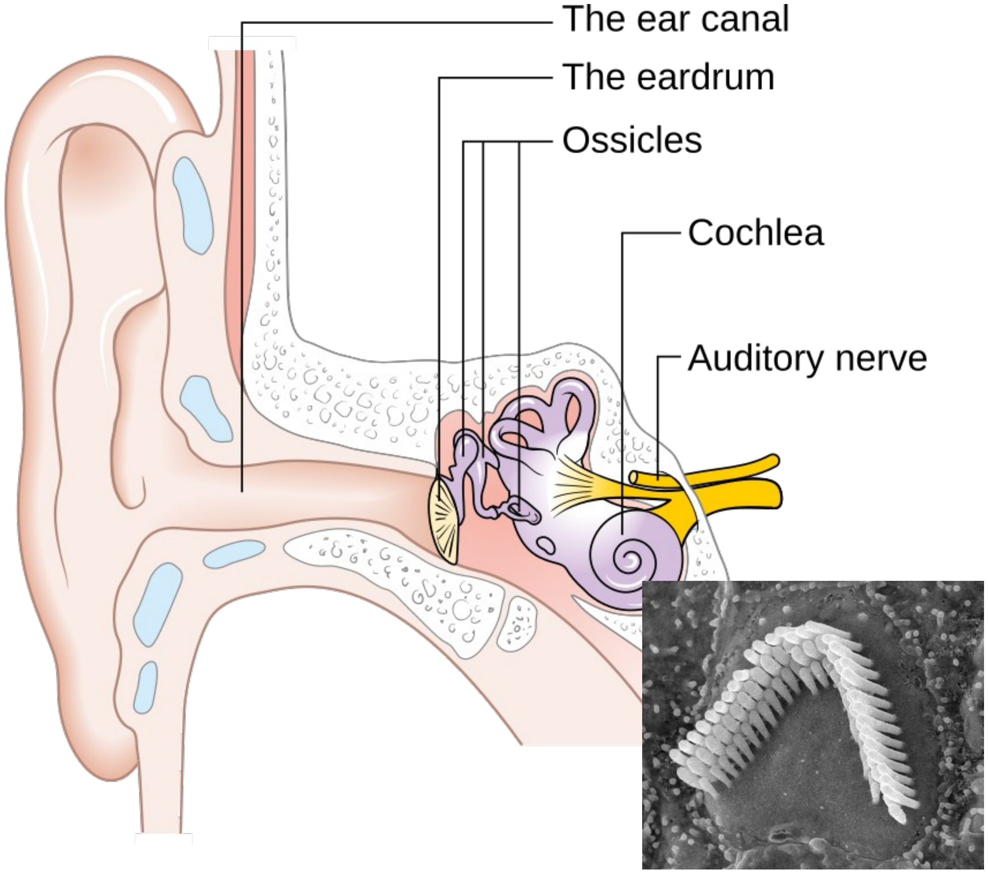
- Wave of pressure in medium
  - Particles repeatedly compressed and expanded
  - Longitudinal waves
  - Requires medium (air, water)
  - Electronic representation - audio
- Wave phenomenon
  - Reflection – bouncing
  - Refraction – angle change when entering different medium
  - Diffraction – bending around obstacle

Measurable sound characteristics

- Frequency (Hz)
  - Number of occurrences of a repeating event per unit of time
- Amplitude, pressure, intensity (W/m^2)
  - Amount of change over a period
- Duration (seconds)
- Direction
- Speed
  - Speed based on medium
  - Air: ~331 m/s

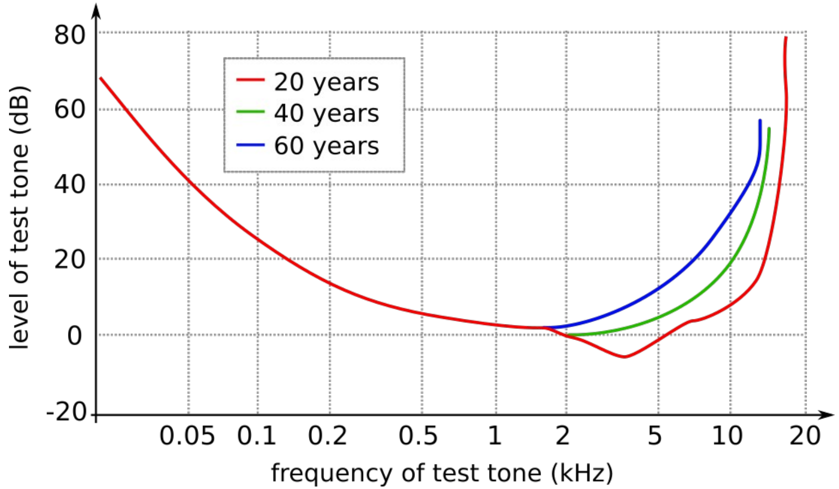
Human auditory perception

- Sound travels the ear canal to the eardrum that vibrates
- Ossicles amplify the vibration
- Cochlea contains liquid that vibrates
- Liquid shakes hair cells
- Hair cells are sensitive to different frequencies
- Responses are transmitted via auditory nerve



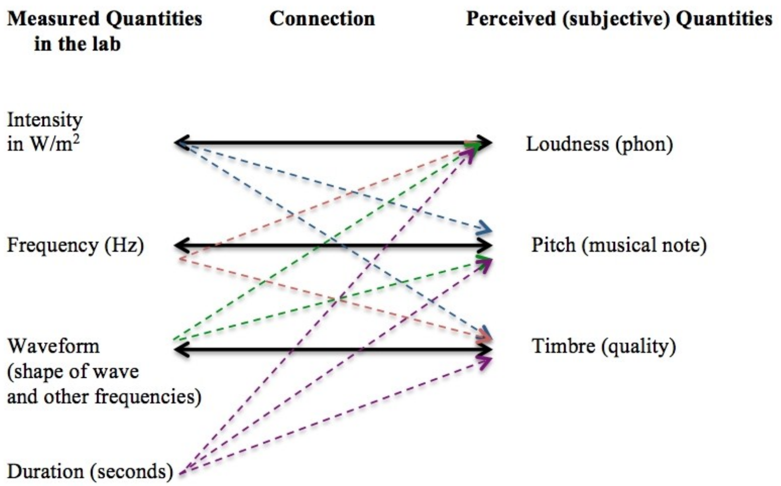
Human ear sensitivity

- Frequencies between 20Hz and 20kHz
  - Some have to be louder than other
- Threshold of hearing
  - Amplitude where a pure tone is detected with 50% accuracy



Perception of sound

- Pitch (low/high)
- Loudness (loud/soft)
- Timbre, tone color
  - Combination of multiple frequencies
  - Change over time
- Sonic texture
  - Multiple sources
  - Unison, polyphony, homophony, cacophony
- Spatial location



Town S., Bizley J. "Neural and behavioral investigations into timbre perception", Frontiers in Systems Neuroscience, 2013

Signal-to-noise (SNR)

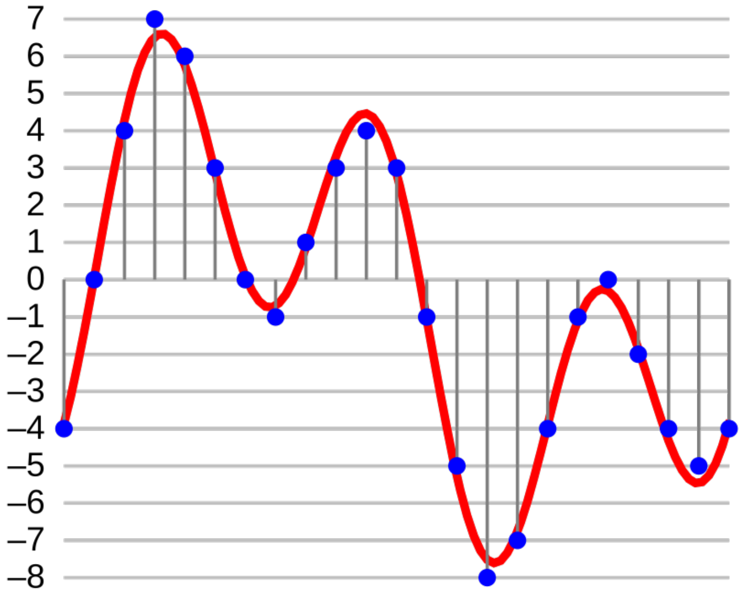
- Random fluctuations in signal (noise)
  - Ratio between power of signal and noise (voltage)
  - Measured with decibels (tenth of a *bel*)

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

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

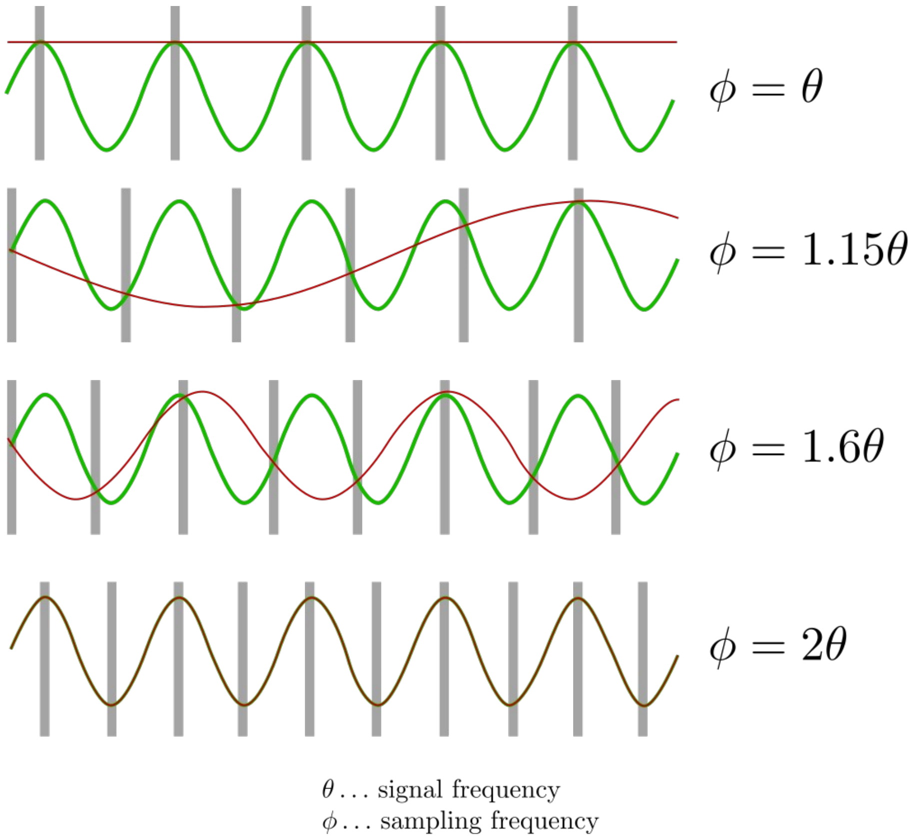
Digital audio

- 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

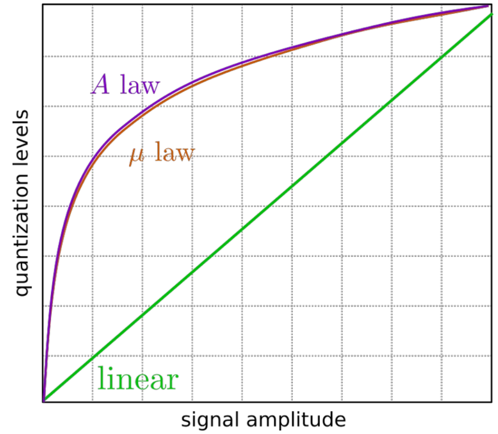


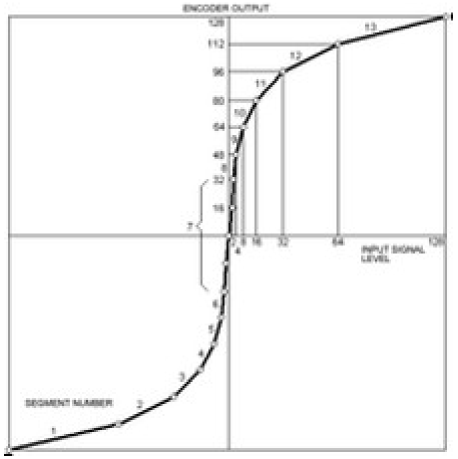
Signal quantization

- 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) $SQNR = 20 \log_{10} 2^N = 6.02 \times N(dB)$
  - Statistical independence $SQNR = 6.02 \times N + 1.76(dB)$
- 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

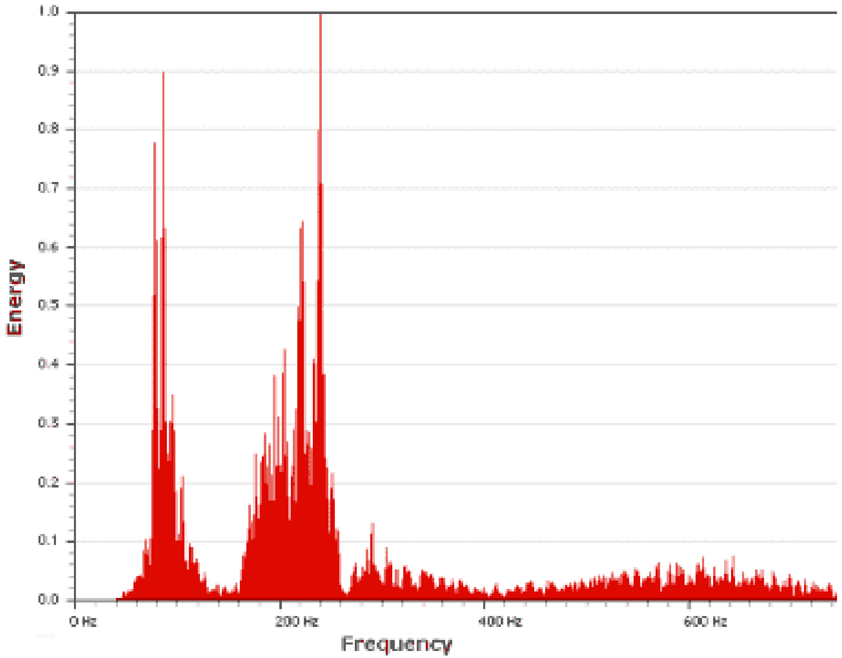
- Formal term for sampling + quantization
- Linear quantization
  - Uniform levels
- Non-linear quantization
  - Better signal-to-noise ratio for low-amplitude signals
  - A-law, μ-law algorithm
- (Adaptive) Differential Pulse Code Modulation
  - Encode difference to previous value
  - Encode difference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





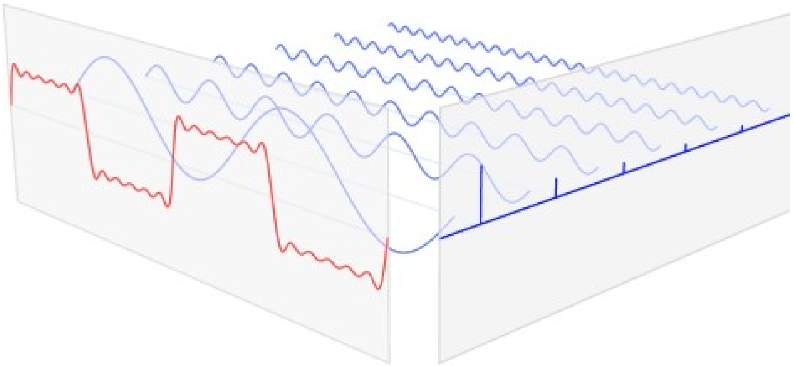
Frequency spectrum

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



Fourier analysis

- General functions represented/approximated by sums of simpler trigonometric functions
- Decomposing signal into base sine waves
  - Frequency distribution
  - Simplifies certain operations
  - Fourier transform
  - Inverse transform

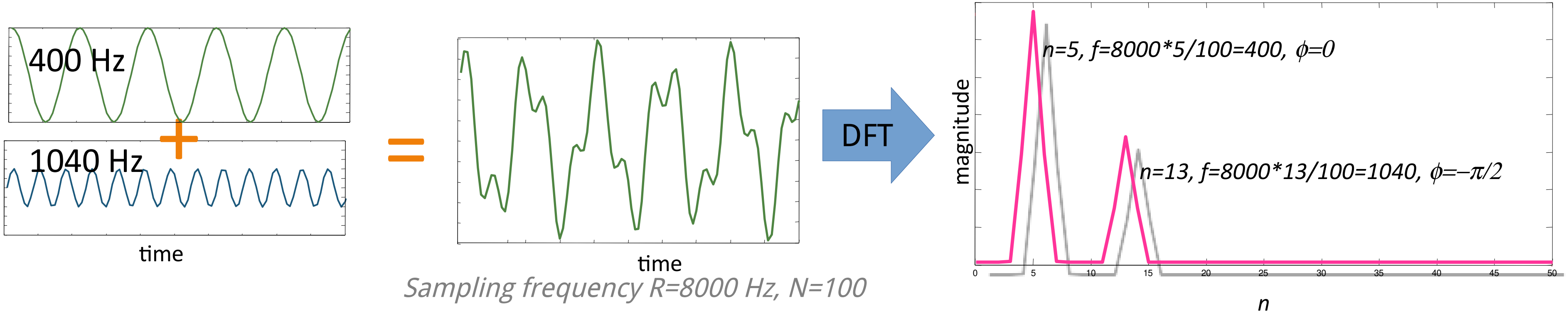


Discrete Fourier Transform

- N point signal described with N coefficients

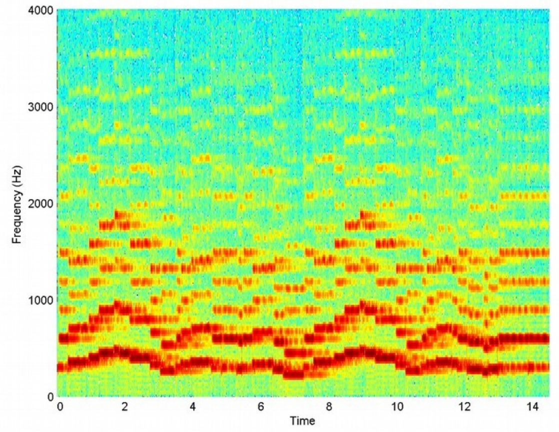
$$\{x_0, x_1, \ldots, x_{N-1}\} \rightarrow \{X_0, X_1, \ldots, X_{N-1}\} \qquad X_k = \sum_{n=0}^{N-1} x_n \cdot [\cos(2\pi kn/N) - i \cdot \sin(2\pi kn/N)]$$

- Fast implementation (FFT) $O(n^2) \rightarrow O(n \log n)$



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

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

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

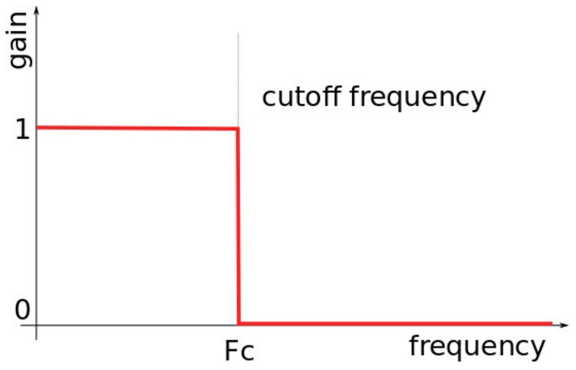
- Non-recursive filters (FIR): $a_n = 0; \forall n > 1$
  - Finite response
- Recursive filters (IIR)
  - Potentially infinite response
  - Implementations more compact

Frequency-response analysis

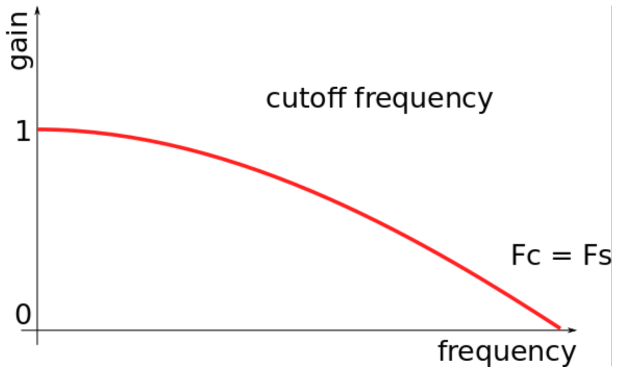
- Compare input and output frequency spectrum
- Only possible for LTI filters
  - Retain sinusoidal signal
  - Observe parametric properties of elementary inputs

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

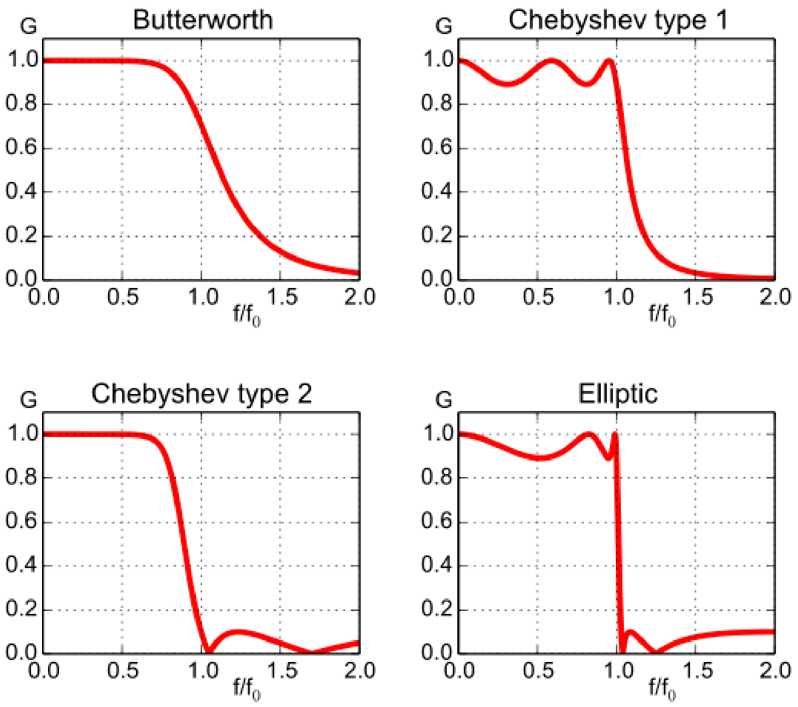


$$y(n) = x(x) + x(n-1)$$



Low-pass filter implementations

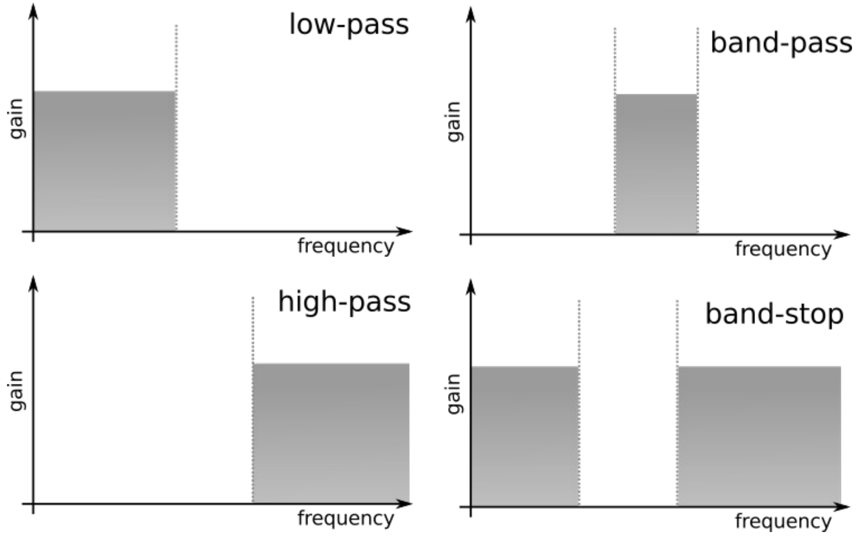
- Butterworth
- Chebyshev
- Elliptic



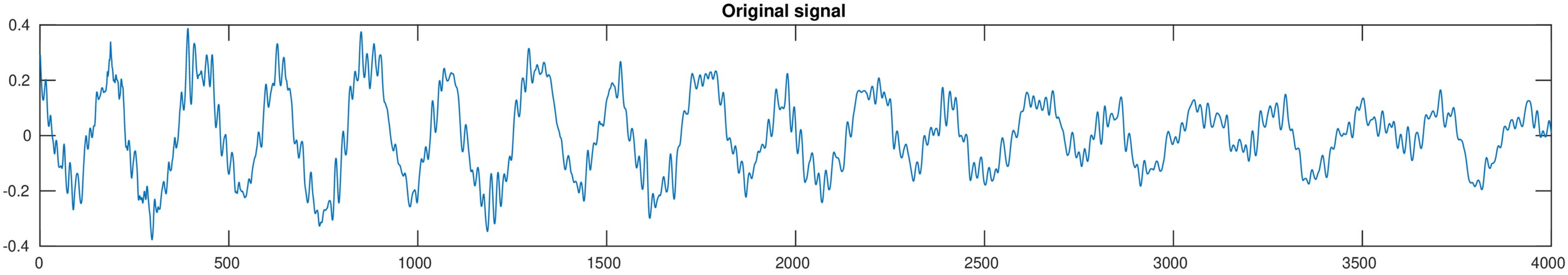
Source: Wikipedia (Geek3)

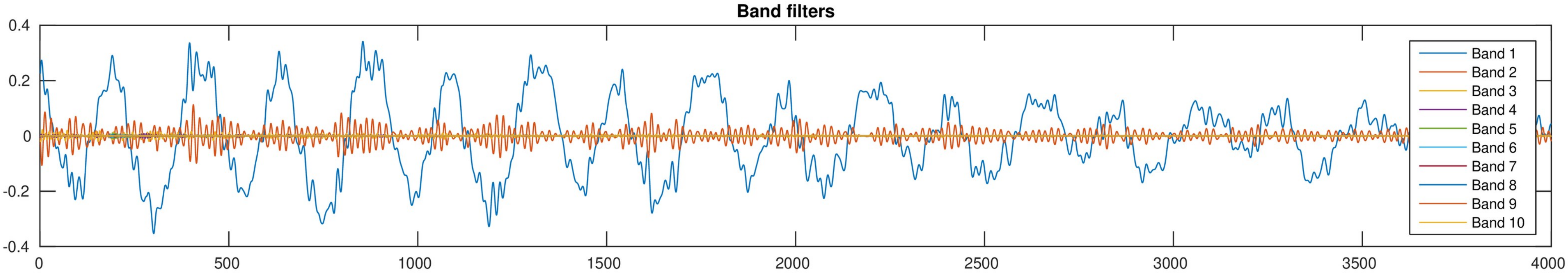
Frequency filtering

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



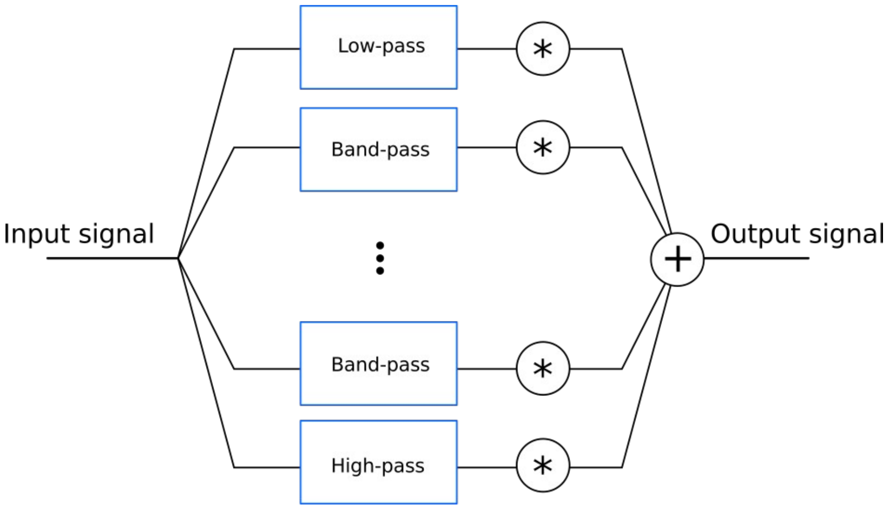
Splitting frequency bands





Audio equalization

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

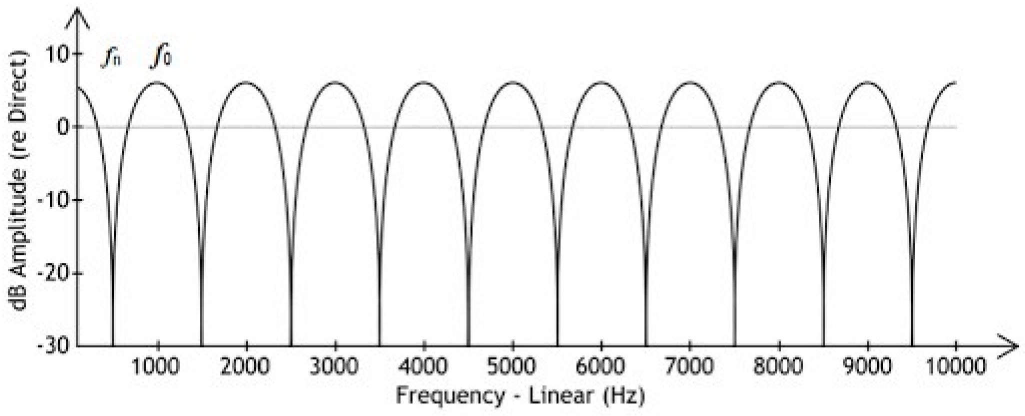


Comb filter

- Basic building block of many effects

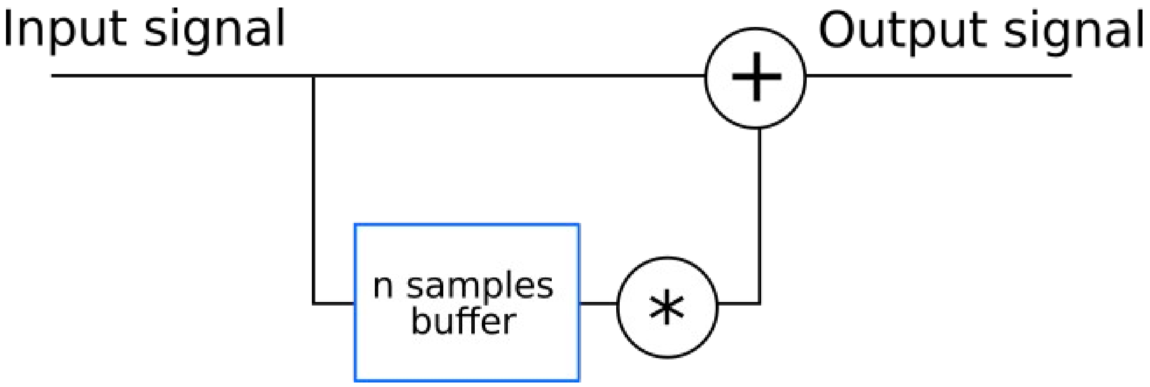
$$y(n) = x(n) + \alpha_1 x(n - t_1) + \alpha_2 y(n - t_2)$$

- Feedforward
- Feedback



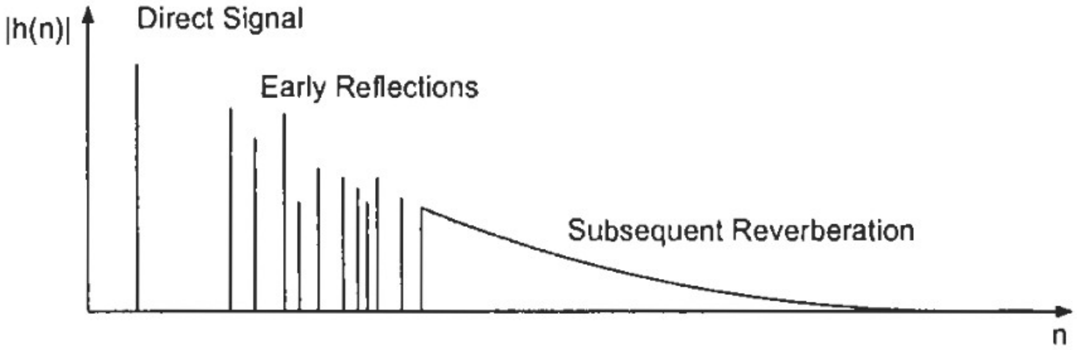
Delay

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



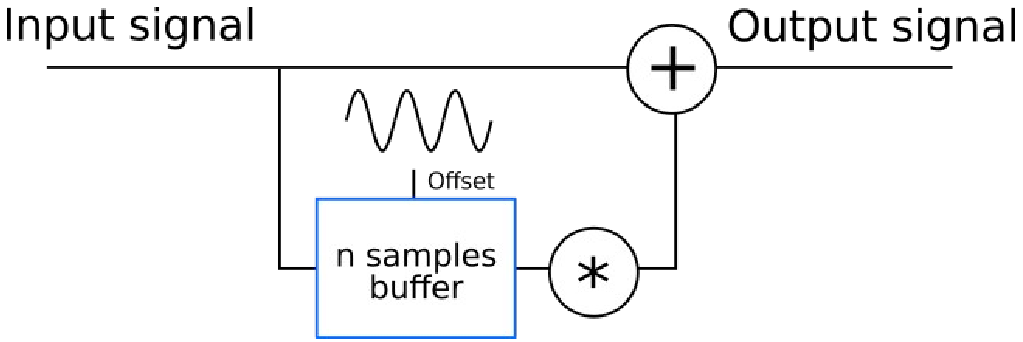
Echo

- 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

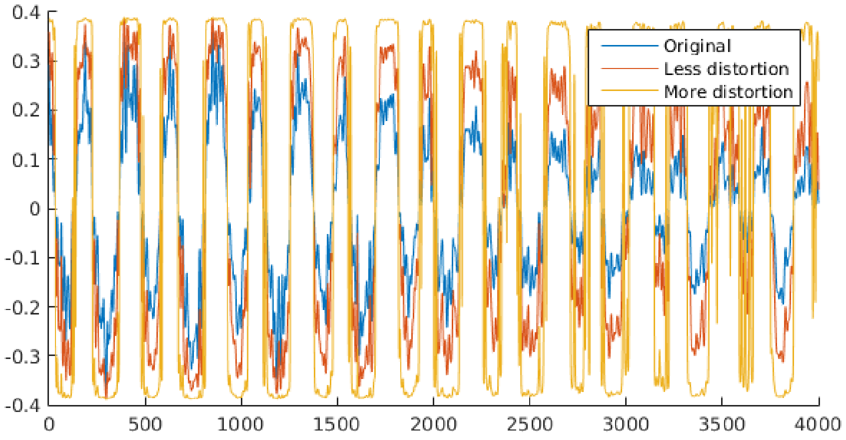
Non-linear filters

- 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

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

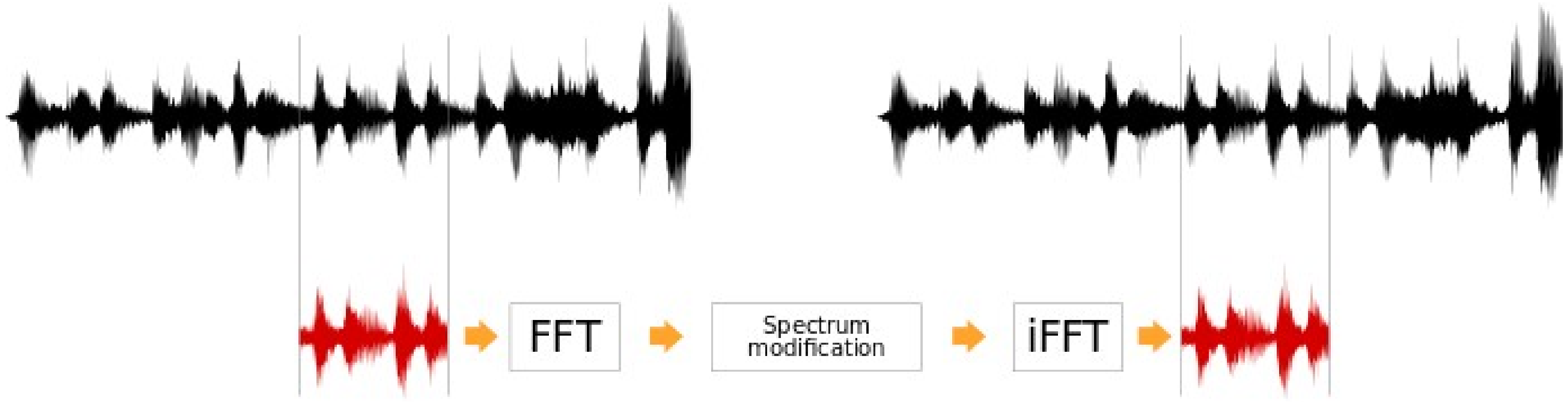


Changing playback speed/pitch

- Sample rate conversion
  - Timing is changed
  - Transposed pitch (chipmunk effect)
- Frequency domain
  - Process windows (STFT)
- Time domain

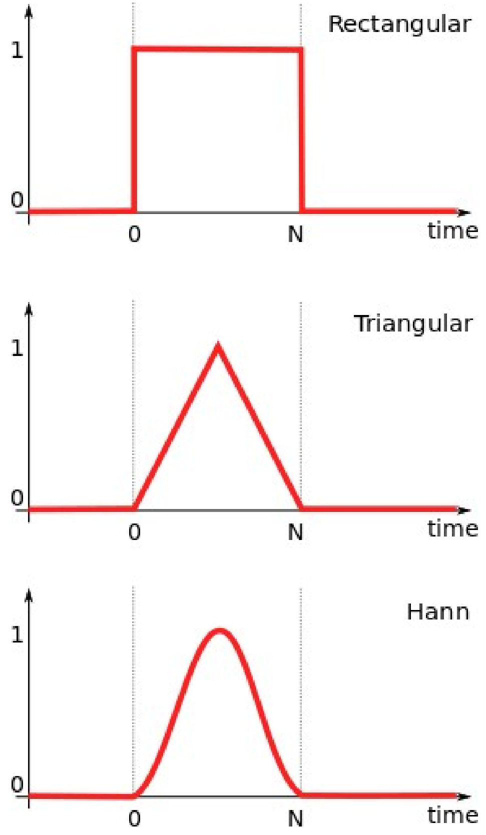
Spectrum Analysis

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



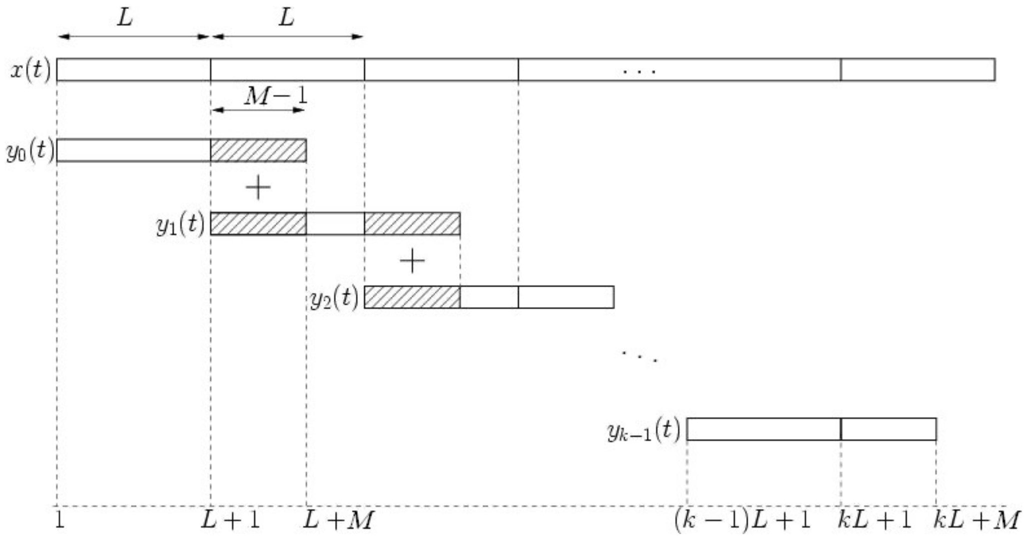
Extracting a segment

- 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



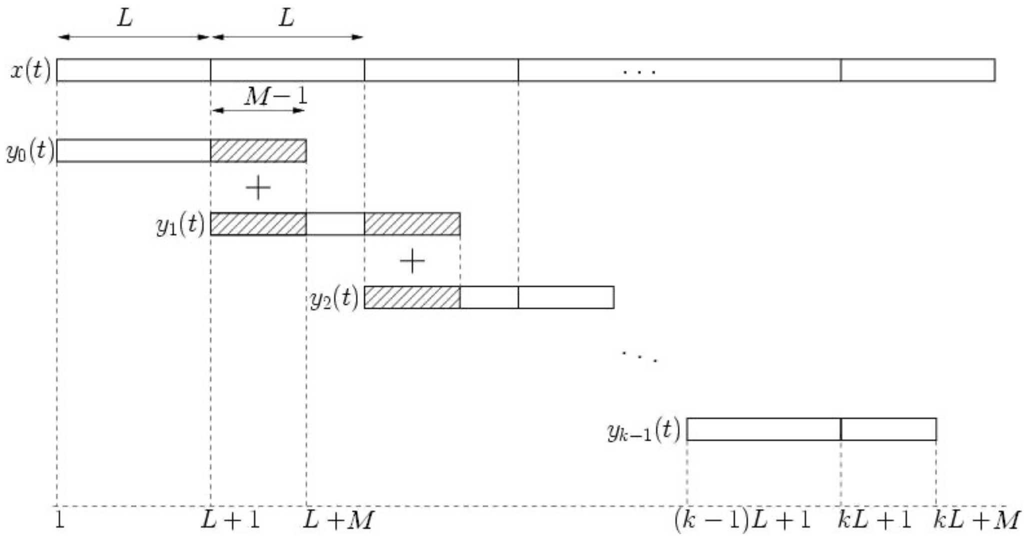
Overlap-Add

- 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



Time stretch

- Short, smoothly windowed block of samples
- FFT transform
- Processing to the Fourier transform
  - Resampling the FFT blocks
- Inverse STFT
- Combine blocks (overlap-add)

Pitch shift

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

Filter banks

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

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)