Digital Audio

Denis Helic

- Changes in air pressure - sounds
- Sound is a continuous wave, or vibrations in air
- Human audio perception system
    - Human ear senses sounds
    - Transmits signals via nerves to the brain
    - Brain percieves sounds

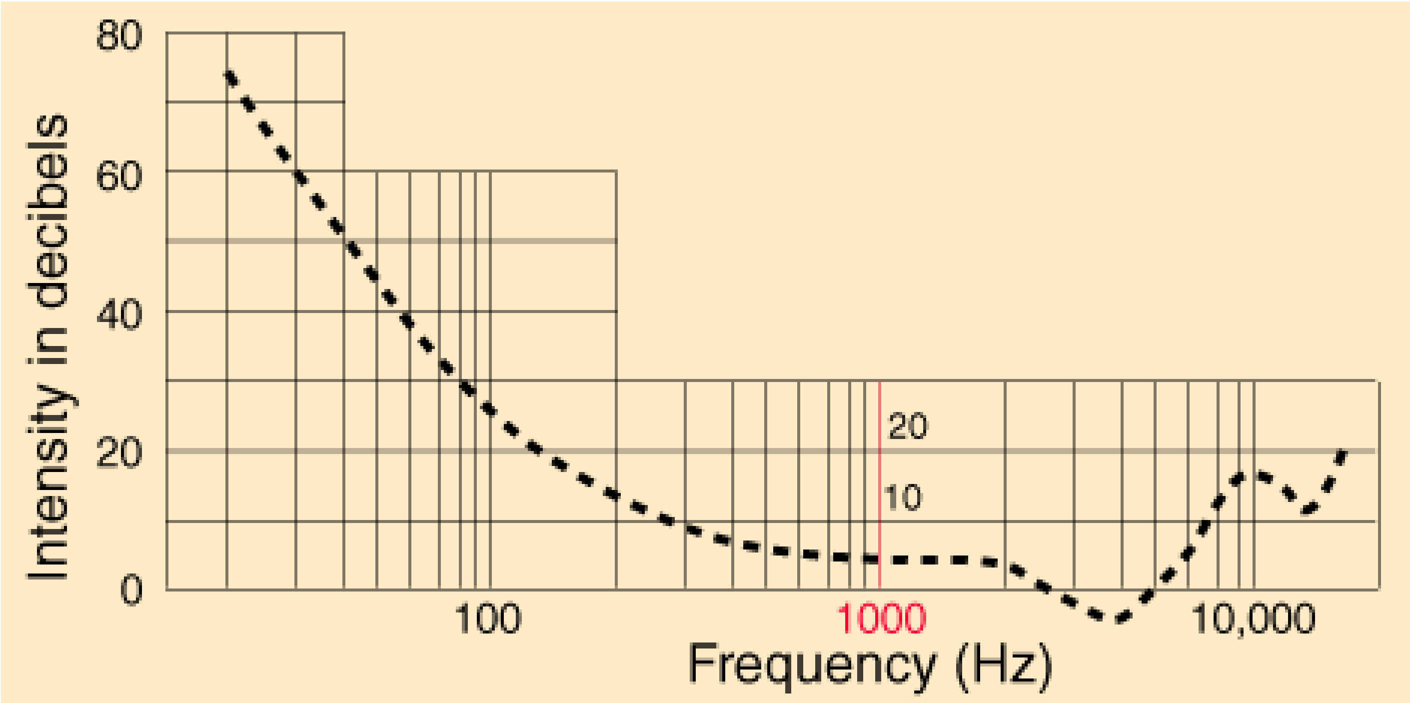
- Properties of sound that we percieve
  - Intensity or loudness
  - Frequency or pitch (how often the pressure changes)

- Perception of loudness
  - Proportional to the pressure
  - Measured in decibels (dB)
  - 0dB hard to percieve
  - 120dB very loud $\rightarrow$ causes pain

- Perception logarithmic
  - 10dB increase (10x increase in the pressure)
  - Twice as much loud as before

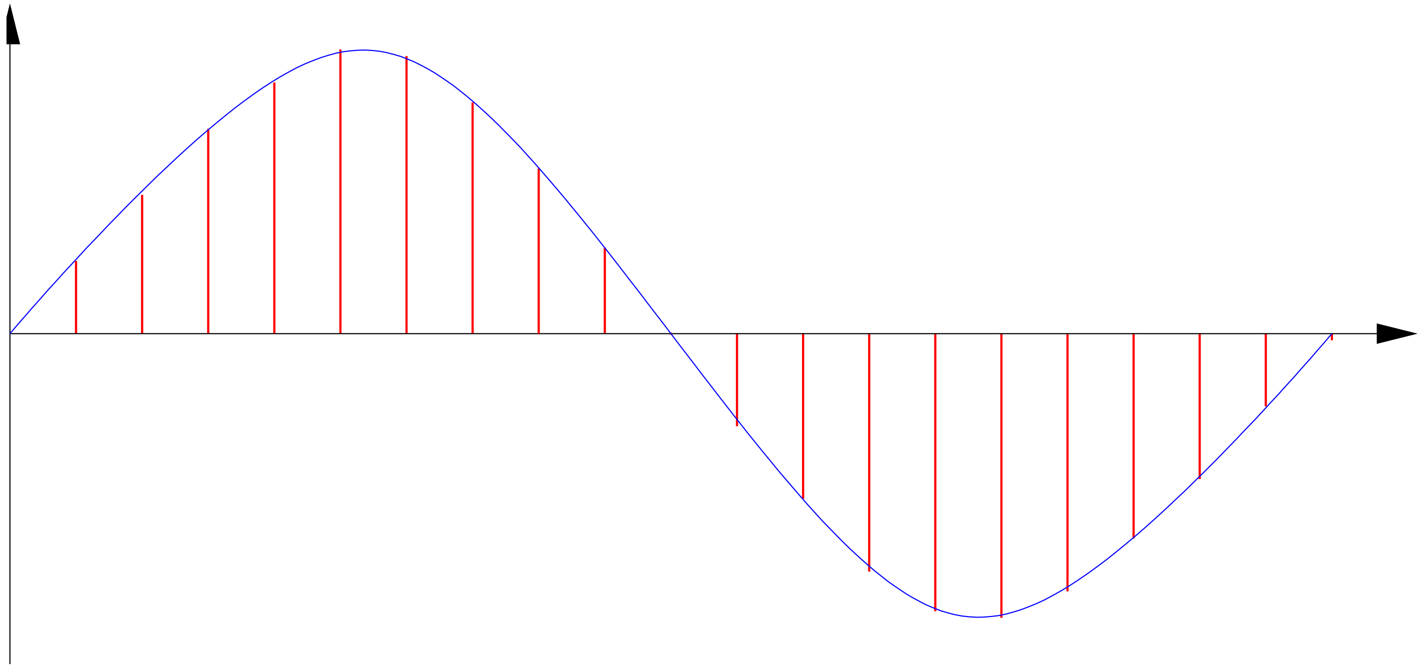
- Perception of frequency
  - Humans can percieve frequencies from 16Hz up to 20kHz
  - Perception logarithmic
  - 2x increase in frequency percieved as an octave
  - e.g. difference between 200Hz and 400Hz percieved same as difference between 5kHz and 10kHz

- Loudness and frequency depend on each other
  - Different frequencies with same loudness percieved with different loudness
  - Human ear is most sensitive at 3kHz, i.e. we can hear even very quiet sounds
  - The largest loudness range is up to 1kHz, i.e. the sound must be loud at lower frequencies to be percieved

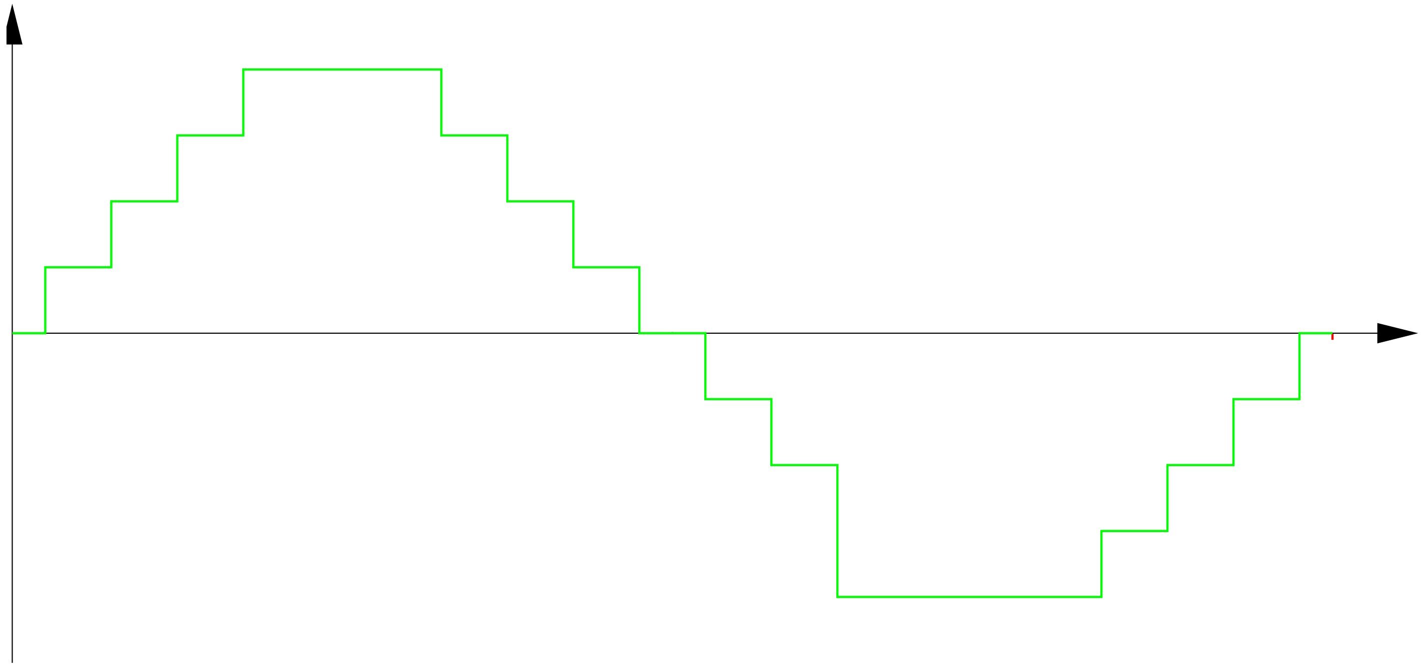

80
60
40
20
0
Intensity in decibels
20
10
100
1000
10,000
Frequency (Hz)

- Properties of sound waves
  - Single sinus-wave
  - Basewave (the lowest frequency) and some overtones
  - Overtones: harmonic - basewave frequency multipled by an integer (2, 3, ...)
  - Overtones: partial - basewave frequency multiplied by a non-integer
  - Different instuments, same tone, different sounds

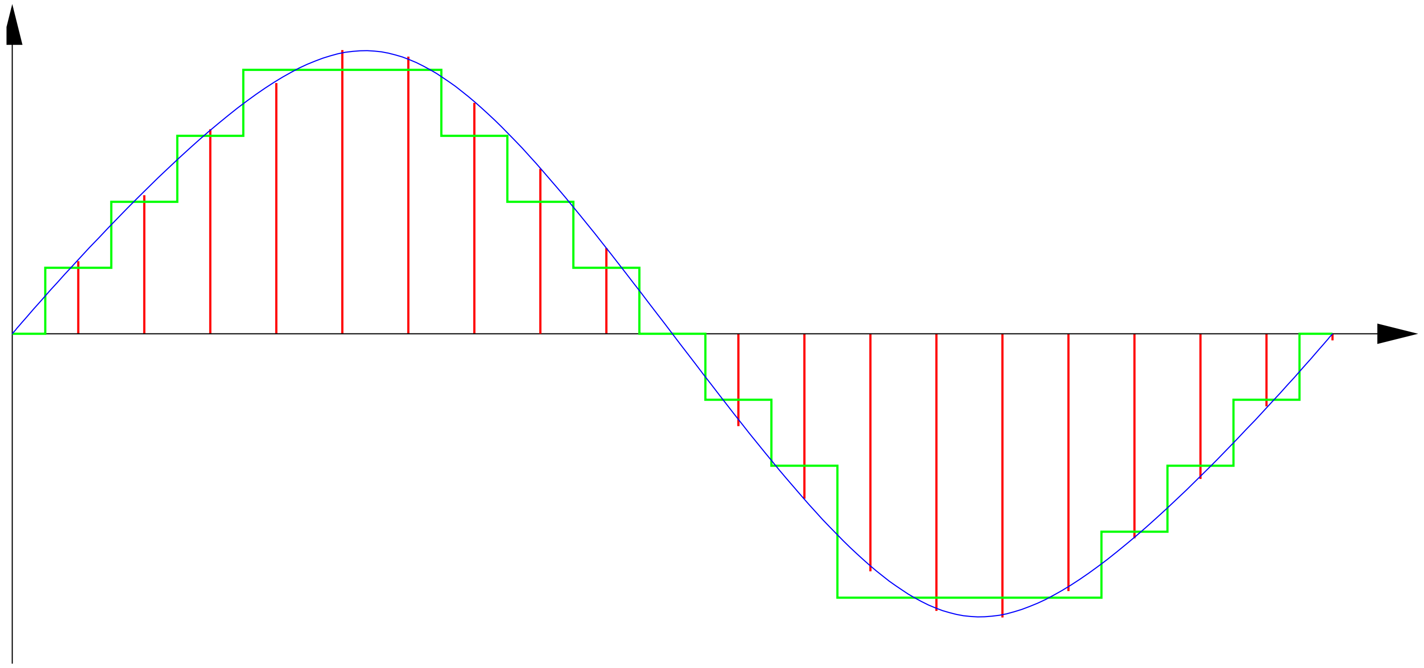
- samples
  - discrete times



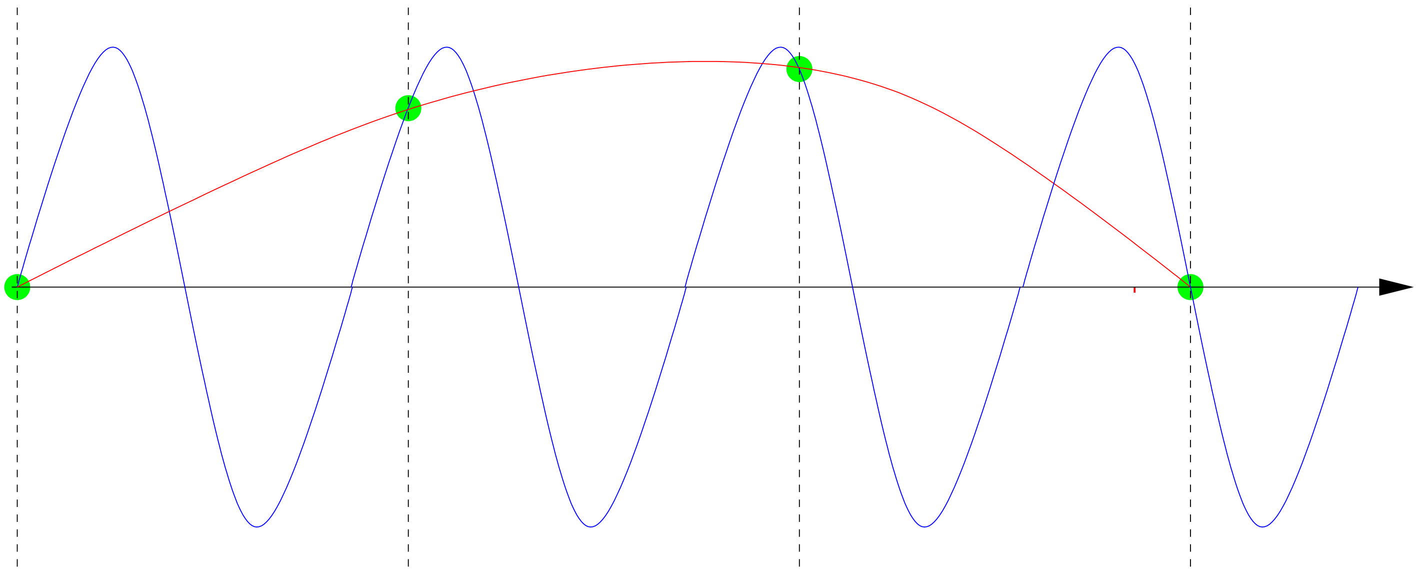
- discrete times and values



- quantization errors



- parameters of A/D-converter
  - samplingrate
  - samplingsize (dynamic, resolution between low/high volume)
- Shannon/Nyquist's theorem: sampling rate must be twice the highest frequency of the signal



- typical values

	Sampling Rate	Resolution	Frequency Range	Data/sec
Telephone	8kHz	8bit	200-3400Hz	8kB/sec
CD	44.1kHz	16bit	20-20000Hz	176kB/sec
DAT	48kHz	16bit	20-20000Hz	192kB/sec
DVD Audio	max 192kHz	max 24bit	0-96000Hz (max)	1152kB/sec

- Entropy coding
- runtime length encoding (instead 1,5,5,5,5 $\rightarrow$ 1,4*5)
- silence compression
- huffman encoding: statistics $\rightarrow$ different code length
- LZW (Lempel, Ziv & Welch): repeating patterns

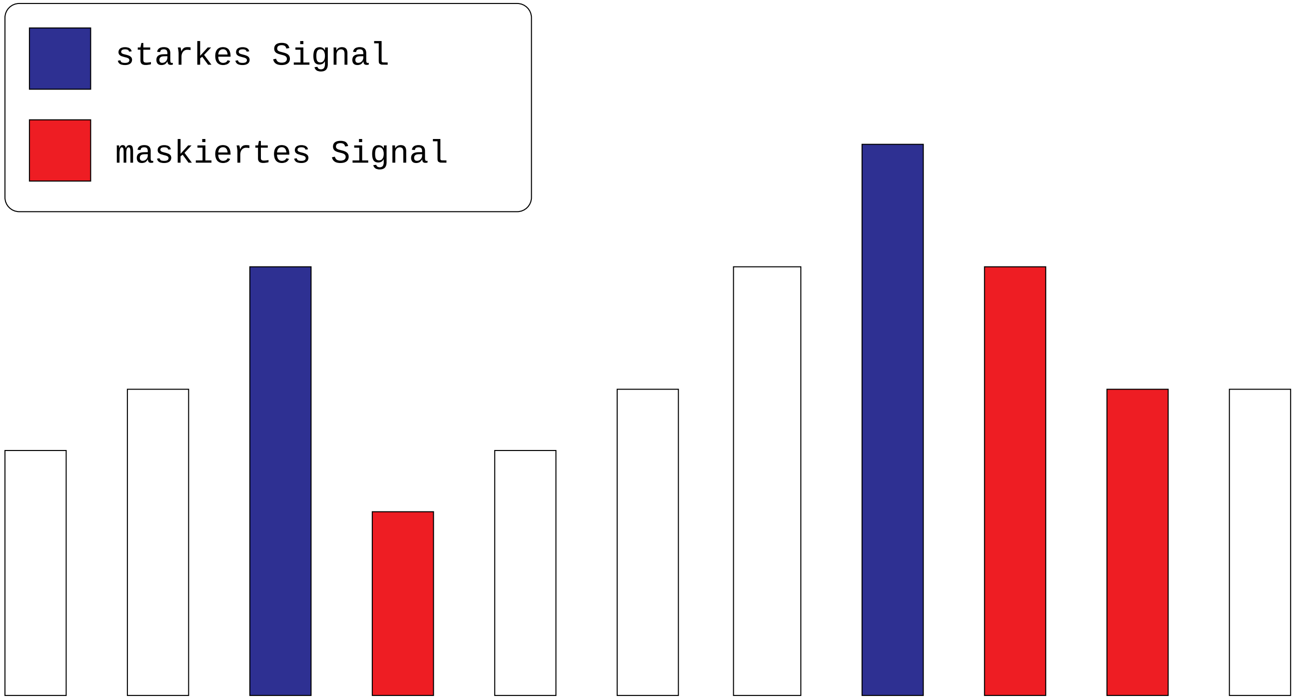
- Source Coding (Prediction)
- Differential Puls Code Modulation (DPCM)
  - only difference between two samples is stored, needs less resolution (number of bits), problem when large differences
- Adaptive Differential Puls Code Modulation: smaller values
  $\longrightarrow$ less resolution (bits)

- μ-Law, a-Law: non-linear mapping of samplevalue to loudness
  $\rightarrow$ higher dynamic
- all lossless compressions for audio only useful in a limited way

- principle: do not store unimportant information
- psychoacoustic model decides, which information is not hearable
  - brain does the same
- algorithms:
  - predictive coding: store only (predicted) difference from current to predicted next sample
  - sub band coding: divide frequencies in sub-bands, loudest sub-band most important
  - spectral or transform coding: fourier-transformation $\longrightarrow$ changes slower (easier to code)

- most important part of audio compression
- most secret in most companies
- GPL model PSYCHO (`http://lame.sourceforge.net/gpsycho/gpsycho.html`)
- adaptive threshold value
  - dependent not only on human being, also on frequency
  - depends on noise around (ticking of alarm clock unhearable during ringing alarm)

- sounds are masqueraded
  - simultaneous, after, and before (!)



  - the closer the frequencies match, the louder the masqueraded sound may be

- Dolby AC-3
  - more lawyers than engineering staff
  - cinema, DVD
- ATRAC (Adaptive Transform Accoustic Coding) (Sony): minidisc
  - compression to one fifth (20%)
- MPEG-1
  - 32kHz, 44.1kHz, 48kHz
  - 2 channels
- MPEG-2
  - additional 16kHz, 22.05kHz, 24kHz
  - 5 channels (left, right, center, 2 surround)

- originally video-compression
- MPEG audio layers (1,2,3) (lower bitrates)
- aim: not only compression
  - cheap hardware-encoder
  - fast-forward/backward
  - resistent against cascading (no new artefacts)
  - guaranteed bandwidth
  - results in "best usage of given bandwidth for given signal"
- fixed filterbase (better implementation in hardware)
- no knowledge of past signal needed (lots of memory needed)
- variable bitrate (better compression)

- depending on layer (1,2,3) frequencies are divided into sub-bands (FFT)
  - layer 1,2: 32 sub-bands of equal size
  - layer 3: unequal size
  - problems: the ear does not use sub-bands of equal size, adjacent bands overlap
- psychoacustic model → different importance of different sub-bands
- joint-stereo (layer 3): different bandwidth for sum and differential signal (70/30)
- MPEG-2 AAC
  - 4 sub-bands, complex coding

- Ogg Vorbis
  - open source, no patent, license fees
  - very good quality (cut off freq.)
  - multi channel
- TwinVQ
  - uses pattern of samples in form of vectors
  - standard pattern
  - sounds like parts of signal are lost
  - good for speech compression (MPEG 4)

- Wave (`*.wav`): a-Law, μ-Law und ADPCM
- MSAudio (ASF/WMA): no detailed information available, streamable, faster than mp3
- RealAudio: streaming format, unknown coding
- MP3Pro: better than mp3 at low bitrates (up to 96kBit)

- Musical Instrument Digital Interface
- serial interface standard for interchange of music data (synthesizers, keyboards, ...)
- describes sound
  - pitch
  - volume
  - timbre (waveform)
  - dynamic of stroke
  - instrument
  - etc.