Introduction to Media Informatics: Audio

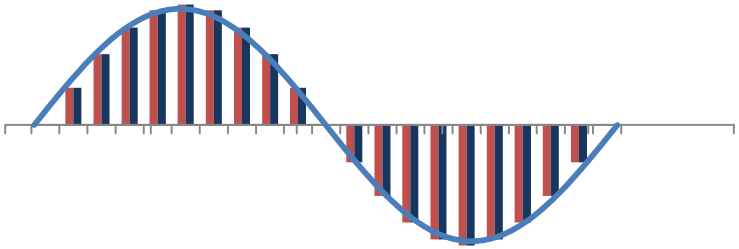
Dr. Mathias Lux
Associate Professor
Alpen-Adria Universität Klagenfurt

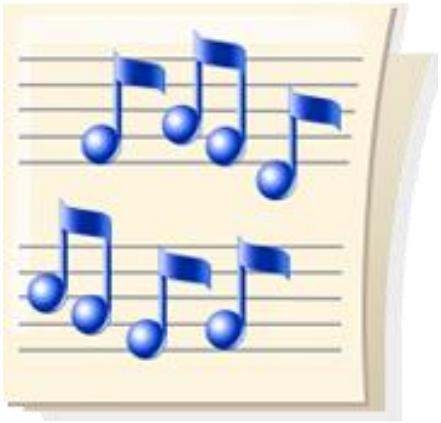
Agenda

- Audio Basics
- Audio Visualization
- http://www.sonicvisualiser.org/

What is digital sound?

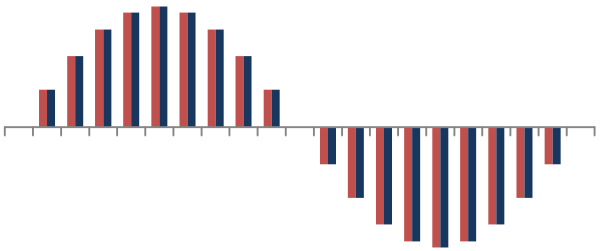
- A digitization of the wave.
  - Either a recipe for reconstruction
  - Or a discrete approximation





Sampled sound

- Wave gets sampled x times a second
  - E.g. 48.000 times -> 48 kHz sampling rate
- Obtained values are stored
  - E.g. 256, 240, 13, -7, -12, -44, ….
  - Quantization to e.g. 2^8 levels -> 8 Bit
- Possibly from different sensors
  - Stereo -> 2 channels



Sampled sound

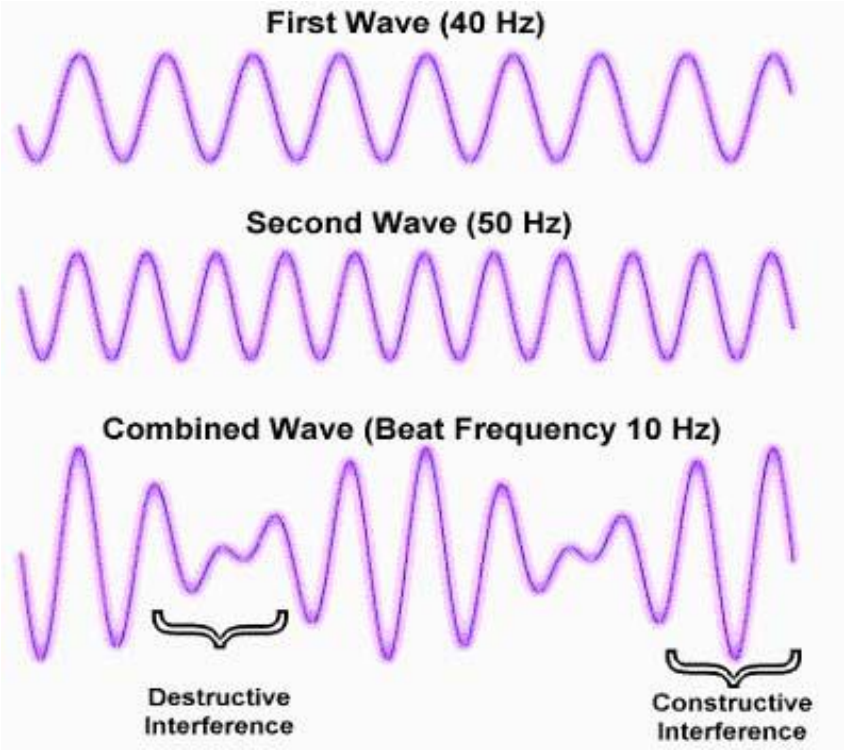
- Example: 8 kHz, 16 bit Stereo
  - Sound wave is sampled 8.000 times a second
  - Samples are stored in 16 bit numbers
- That's *Pulse Code Modulation* (PCM)
  - Often used in WAV files …
  - Also as input from microphone or line in

What do we want to capture?

- Humans can hear
  - From around 16 – 21 Hz
  - To around 16 kHz – 19kHz
  - 16 bit is enough (CD), 32 bit even better

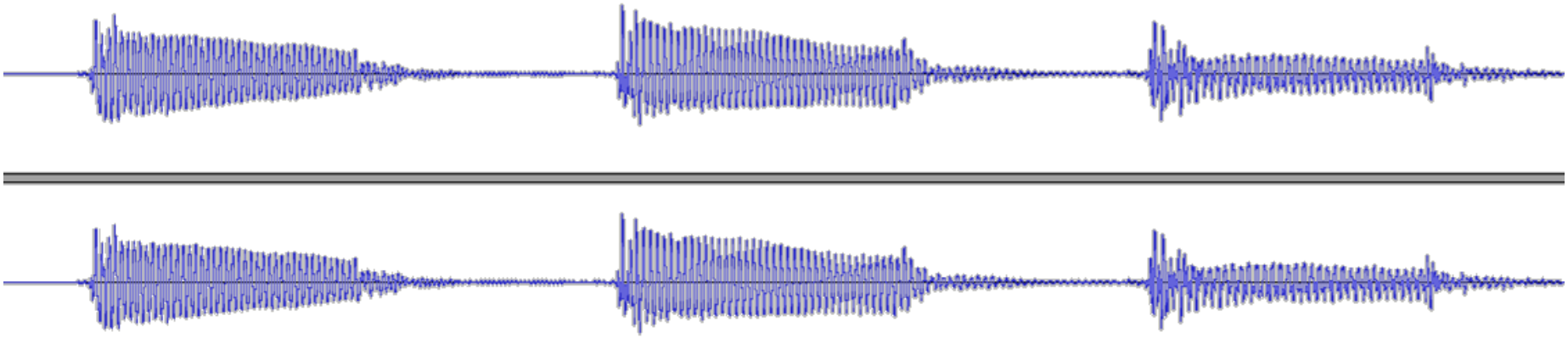
Sound Mixing

- Multiple sounds at the same time?



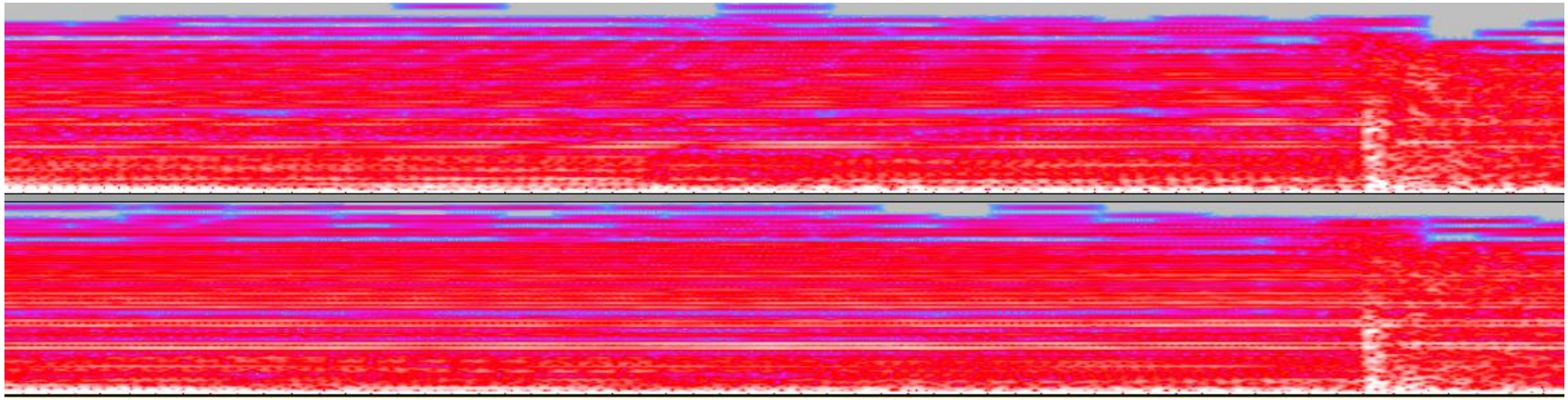
Audio Representation: Waveform

- Gives the power of a signal of a given time
- Visualization of pulse code modulation



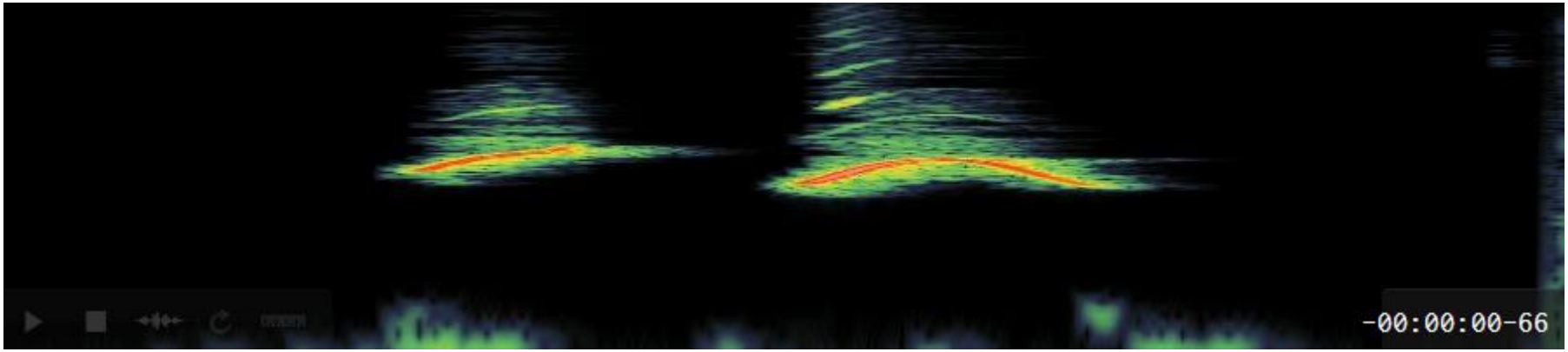
Audio Representation: Spectral

- Gives power in frequency bands as color
  - gray -> blue -> red -> white
- Shows if there are high or low frequencies present.



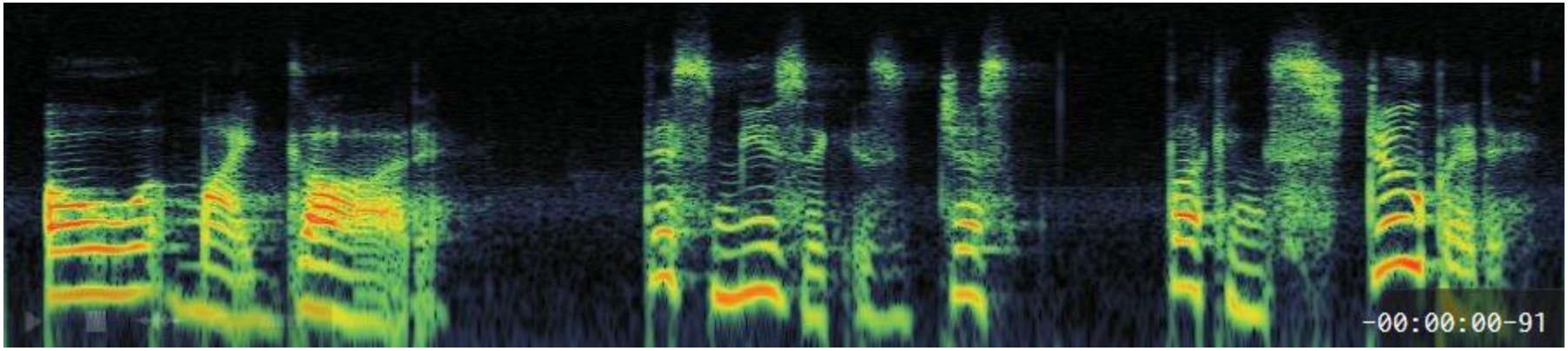
Audio Representation: Spectral

- Whistle
  - http://www.freesound.org/people/THE_bizniss/sounds/39548/



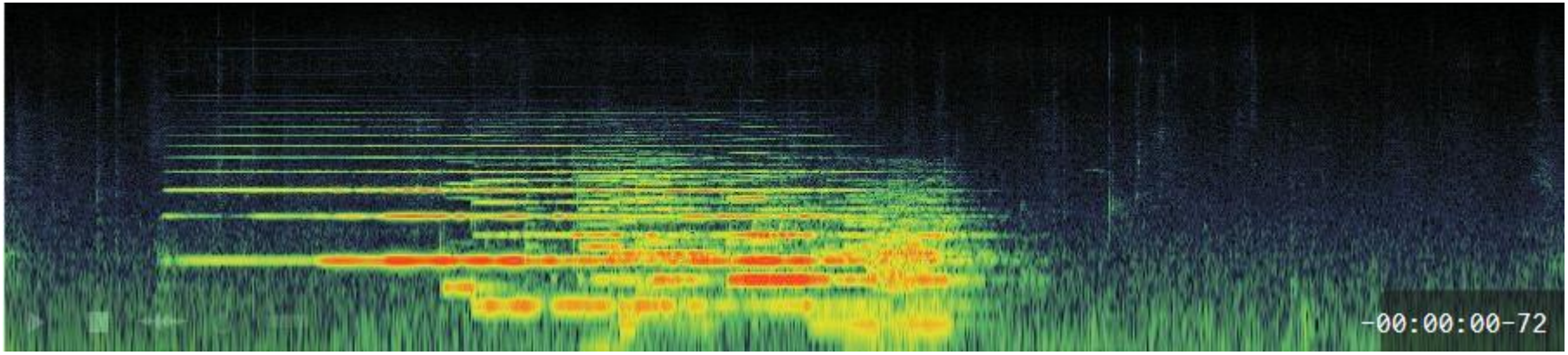
Audio Representation: Spectral

- Voice
  - http://www.freesound.org/people/epanody/sounds/107720/



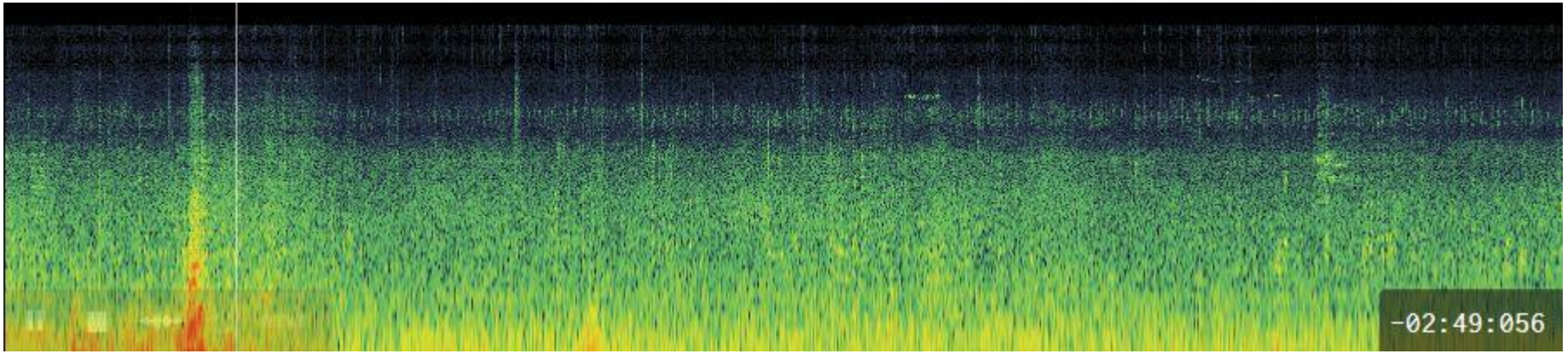
Audio Representation: Spectral

- Music
  - http://www.freesound.org/people/IEDlabs/sounds/82323/



Audio Representation: Spectral

- Environmental
  - http://www.freesound.org/people/sagetyrtle/sounds/36734/



Audio Processing

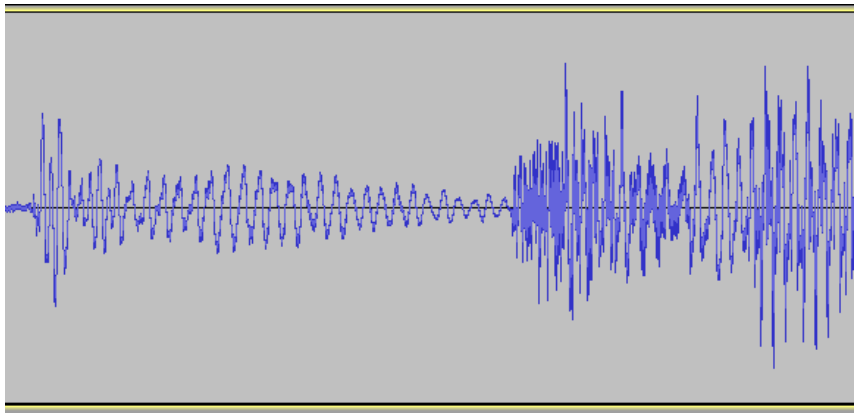
- High Pass Filter
  - Removes low frequencies
- Low Pass Filter
  - Removes high frequencies

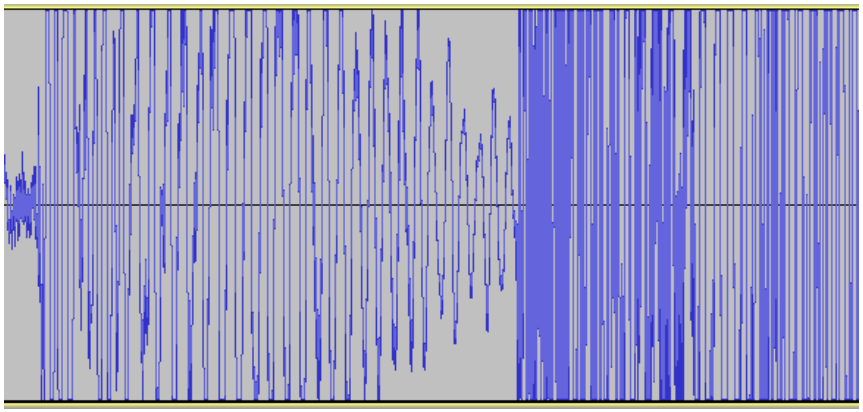
- Demo: Audacity

Audio Processing

- Amplify
  - increase power of samples, ie. sample * x
  - may result in clipping

- Demo: Audacity





Audio Processing

- Change sample rate
  - makes sound faster or slower
  - changes frequencies

- Demo: Audacity, change speed filter.

Audio Processing

- Change pitch or tempo
  - sampling rate change does both
  - doing just one is complicated
- First compute frequency domain representation
  - ie. with Short Time Fourier Transform
- Then re-compute frequencies based with different time basis
- Apply inverse Fourier transform

Audio Processing

- Adding Echo
  - Add the same signal in a new track and reduce amplitude
  - Mix with original track
- Can be done online
  - ie. computer games, 3D sound

Sound Formats

- Waveform Audio Format
  - Container for several compression formats
  - Includes PCM, MP3, GSM, □-Law
- Musical Instrument Digital Interface
  - Control codes for instruments
  - Instruments can be "emulated"
- Compressed Audio Formats
  - MP3, OGG, AAC, ...

Audio Compression

- Lossless compression enough?
  - does not work with audio too well
- Multiple options for compression
  - MPEG-1 Layer 3 (MP3), Ogg Vorbis
  - MPEG-4 AAC, Dolby Digital (AC3)
- In General: Psycho-Acoustic Modeling

MP3 Compression

- Developed by MPEG for Audio Compression (1992)
  - Part of MPEG-1 (Layer-3: Audio)
- Based on Psychoacoustic modeling
  - Developed on an empirical basis
  - Masking effects on 27 frequency bands
  - Model details which frequencies mask which

Critical Bandwidth

- If two sounds share the same critical bandwidth, the ear cannot distinguish between them.
- Sound separation is possible if the frequencies are "different enough"

Temporal Masking

- Sudden stimulus makes other sound inaudible.
- Backward masking
  - obscures a sound immediately preceding the masker
- Forward masking
  - obscures a sound immediately following the masker

MP3-Kompression

- Masking leads to actual compression
- Redundancies in stereo channels are exploited too
- Compression about 10 : 1