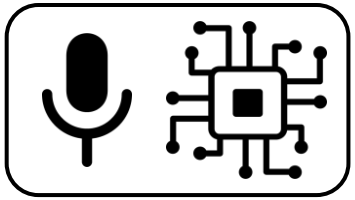
Computational Analysis of Sound and Music

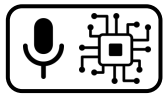


Sound Perception

Dr.-Ing. Jakob Abeßer

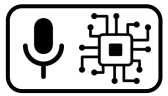
Fraunhofer IDMT

jakob.abesser@idmt.fraunhofer.de



Outline

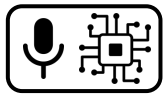
- **Auditory Scene**
- Sound Intensity & Loudness
- Frequency & Pitch & Chroma
- Timbre
    - Harmonics / Transients / Noise
    - Mel Frequency Cepstral Coefficients (MFCC)



Auditory Scene

- Everything a listener perceives in a particular acoustic environment

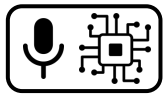
Own



Auditory Scene

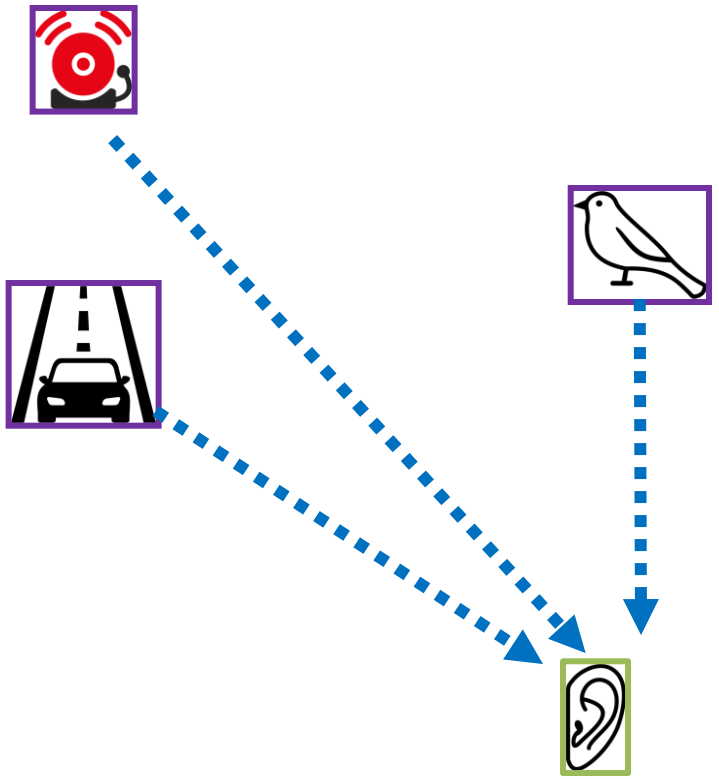
- Everything a listener perceives in a particular acoustic environment
- Type & spatial alignment of sound sources

Own



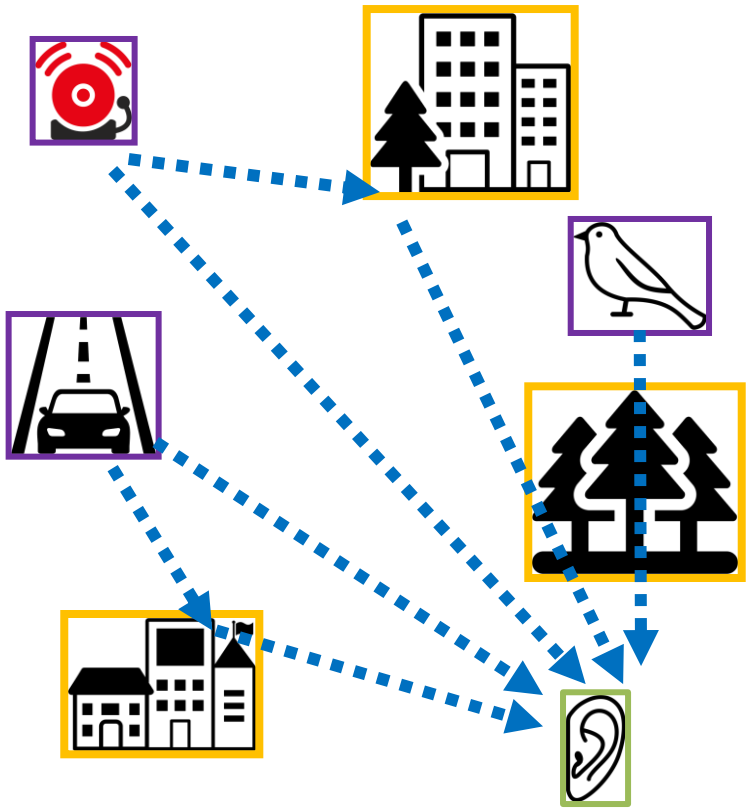
Auditory Scene

- Everything a listener perceives in a particular acoustic environment
- Type & spatial alignment of sound sources
- Physical characteristics of acoustic environment (sound transmission)



Auditory Scene

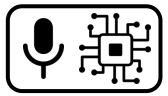
- Everything a listener perceives in a particular acoustic environment
- Type & spatial alignment of sound sources
- Physical characteristics of acoustic environment (sound transmission)
  - Room shape & size
  - Ambient noise
  - Reflective & absorbent surfaces & objects



Own

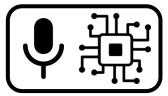
Auditory Scene

- Perceptual organization (listener)
    - Spatial hearing for sound **localization**
    - **Auditory streaming** (grouping sounds into separate streams)
- Soundscape → qualitative, cultural, and aesthetic relationship between listeners and their acoustic environment



Outline

- Auditory Scene
- **Sound Intensity & Loudness**
- Frequency & Pitch & Chroma
- Timbre
    - Harmonics / Transients / Noise
    - Mel Frequency Cepstral Coefficients (MFCC)

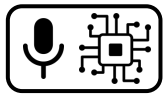


Sound Intensity & Loudness

Intensity

- Sound pressure p (PA)
  - How much the air pressure is changed by a sound wave
- Sound power P (W)
  - Amount of energy emitted by sound source per time
- Sound Intensity I $(\mathrm{W/m^2})$
  - Ratio between the power of an acoustic wave and the area it traverses
  - Example: Omnidirectional sound source (spherical wave)

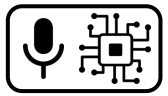
$$I := \frac{P}{4\pi r^2}$$



Sound Intensity & Loudness

Sound Pressure Level (SPL)

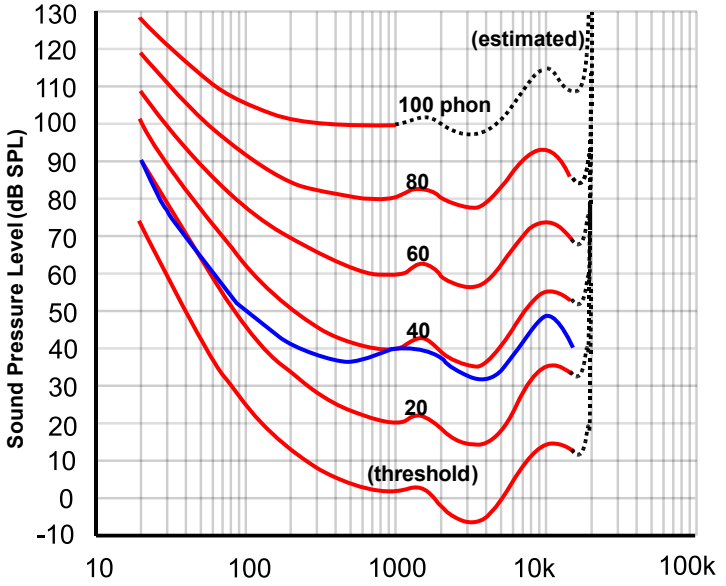
- Level (in dB)
  - Logarithm of the ratio of a physical quantity to a reference value
- Sound pressure level (SPL)
  - $L_p \coloneqq 10 \cdot \log_{10}\left(\frac{p}{p_0}\right)^2 = 20 \cdot \log_{10}\left(\frac{p}{p_0}\right)$
    - Human hearing threshold: $p_0 = 20\ \mu\text{Pa}$
  - Sound pressure doubles → SPL increases by 6 dB
  - Range: barely audible sounds (0 dB) → very loud sounds (120 dB)



Sound Intensity & Loudness

Loudness

- Loudness → Magnitude of auditory sensation
- Loudness level
  - Linear scale (sone) → 1 sone = sinusoidal signal (1 kHz, 40 dB SPL)
  - Logarithmic scale (phon) → 1 phon = sinusoidal signal (1 kHz, 50 dB SPL)
- Frequency dependent → Equal-loudness curves

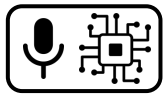


Equal-loudness contours (red) (from ISO 226:2003 revision)
Original ISO standard shown (blue) for 40-phons

Fig-A3-2

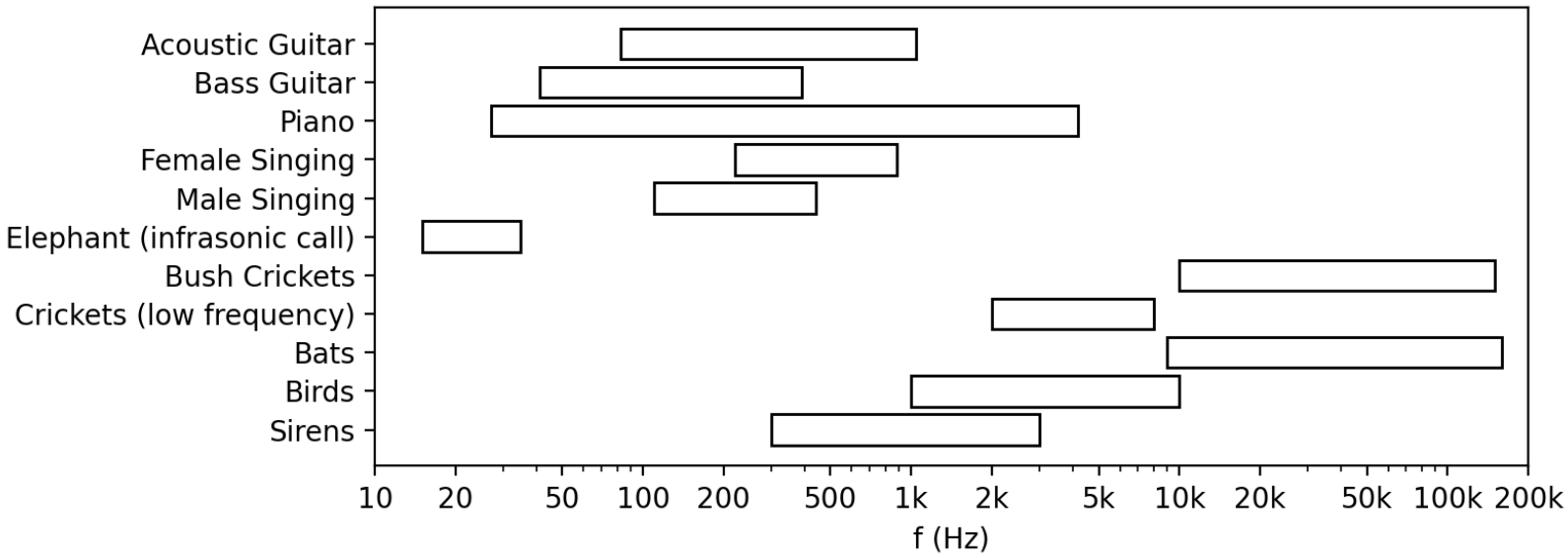
Outline

- Auditory Scene
- Sound Intensity & Loudness
- **Frequency & Pitch & Chroma**
- Timbre
    - Harmonics / Transients / Noise
    - Mel Frequency Cepstral Coefficients (MFCC)



Frequency & Pitch & Chroma

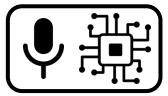
- Frequency ranges
  - Audible frequencies (20 Hz $\leq f <$ 20 kHz)
    - Human voice, musical instruments, everyday sounds
    - Most animal vocalizations



Own

Frequency & Pitch & Chroma

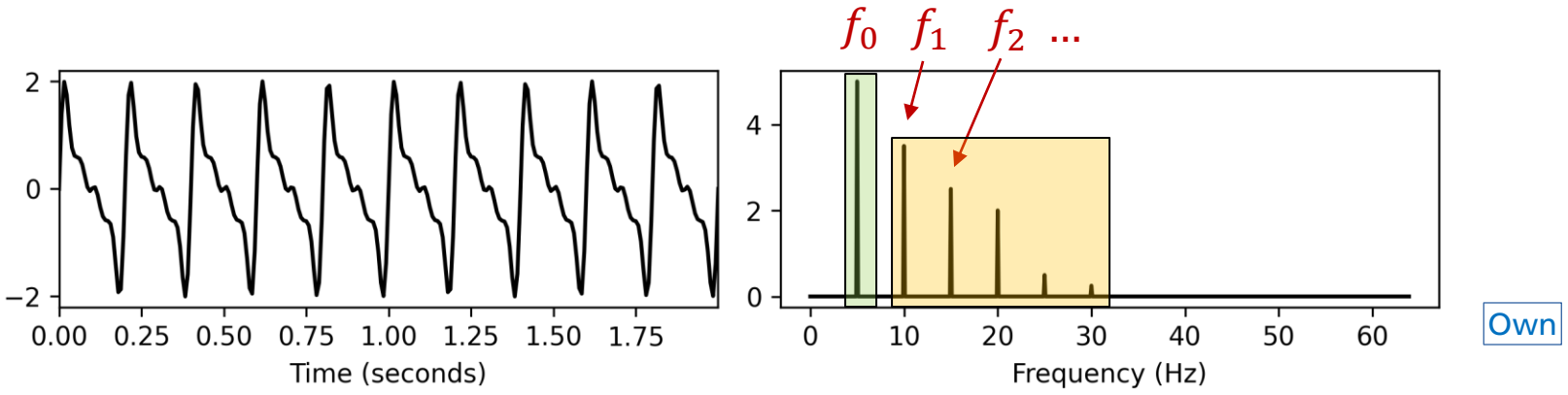
- Frequency ranges
  - Infrasonic ($f < 20$ Hz)
    - Climate & seismic events
    - Perceivable by elephants & whales
  - Ultrasonic ($f \geq 20$ kHz)
    - Medicine (sonography), navigation (radar)
    - Bioacoustic (e.g. bat echolocation)



Frequency & Pitch & Chroma

Periodic Signals

- Sum of pure tones (partials)
  - Fundamental frequency f_0
  - Harmonic frequencies $f_k \approx (k+1)f_0$



Own

Frequency & Pitch & Chroma

Pitch

- Perceptual property (sort sounds from low to high pitch)
- Closely related to frequency
  - MIDI pitch p → Frequency (Hz): $f = 440 \cdot 2^{\frac{p-69}{12}}$

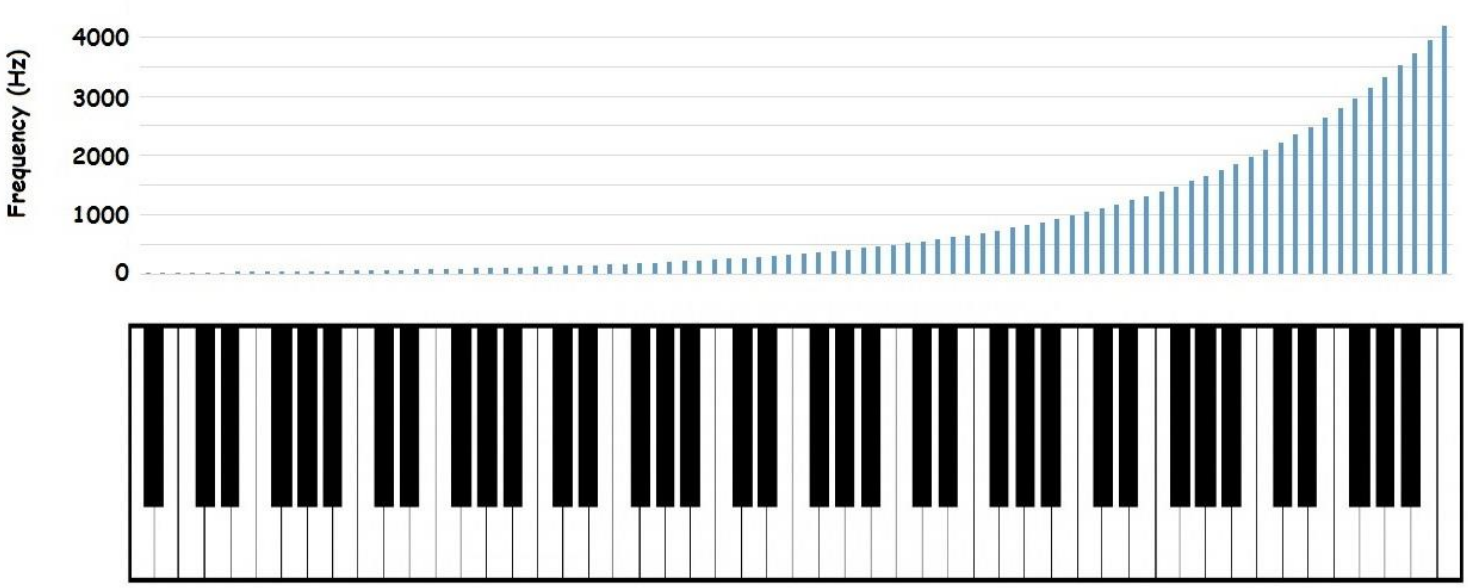


Fig-A3-3

Frequency & Pitch & Chroma

Pitch Intervals

- Depends on frequency ratio
- Example

Note	**Pitch p**	**Frequency f**
A3	57	220 Hz
A4	69	440 Hz
A5	81	880 Hz

Octave intervals

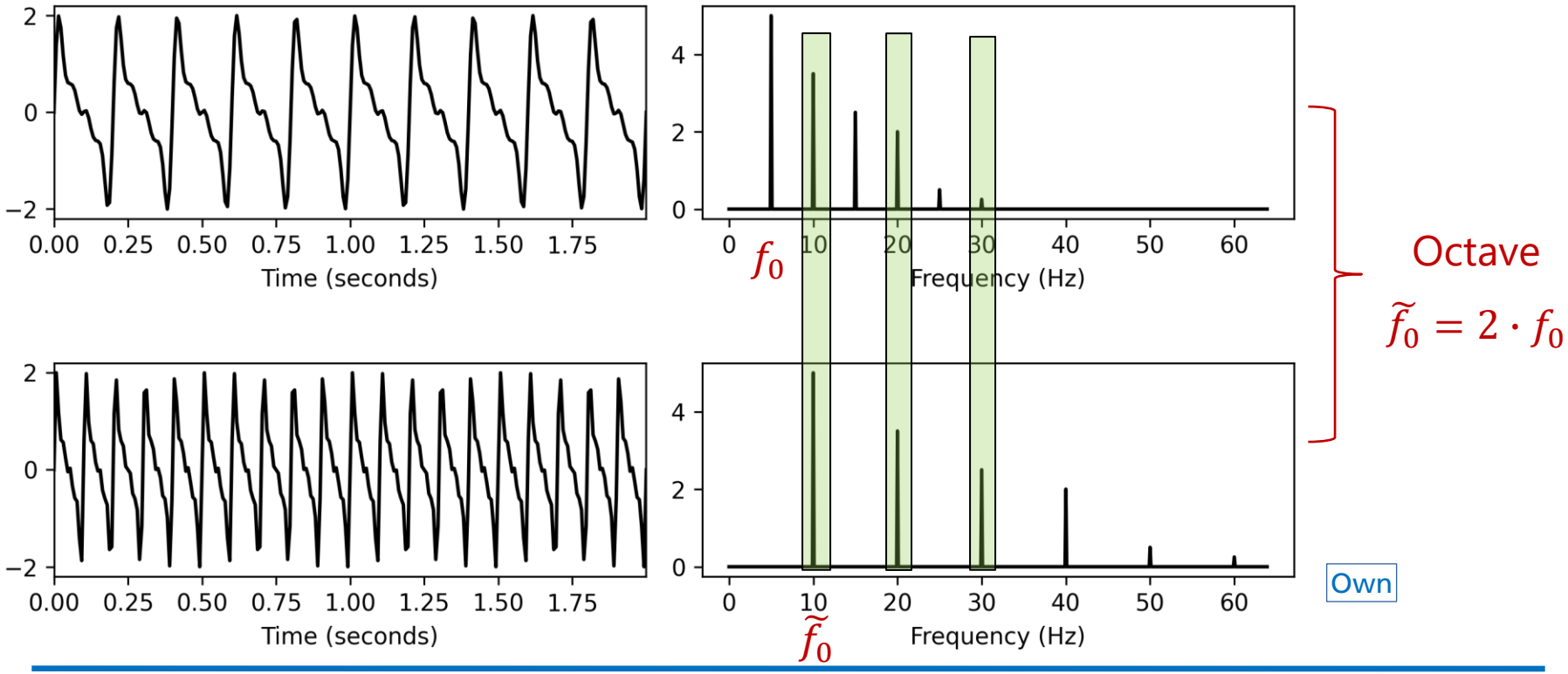
$f(\mathrm{A4}) = 2 \cdot f(\mathrm{A3})$

$f(\mathrm{A5}) = 2 \cdot f(\mathrm{A4})$

Frequency & Pitch & Chroma

Pitch Intervals

- Consonant intervals
  - Shared partial frequencies



Frequency & Pitch & Chroma

Chroma

- Human pitch perception is periodic
  - 2 pitches one octave apart are perceived as similar
  - Pitch = chroma + tone height (octave number)
- Chroma → C, C#, D, D#, …, B (12)

Figure 3.3a from [Müller, FMP, Springer 2015]

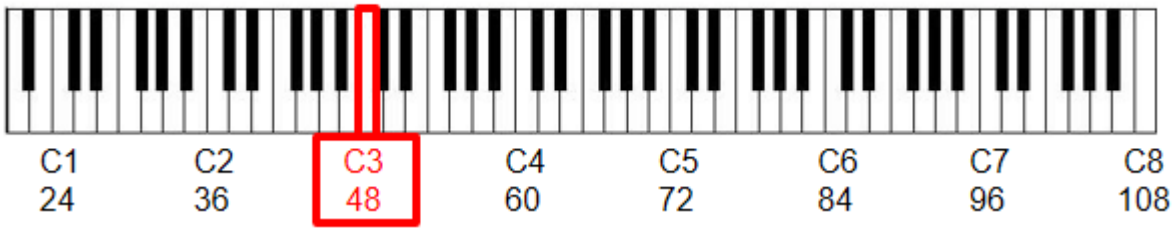
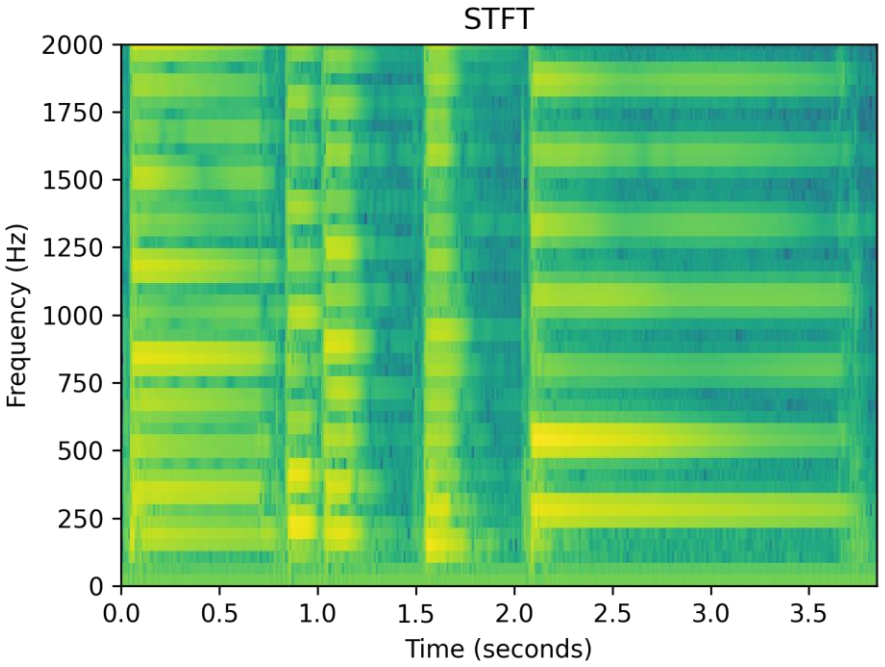


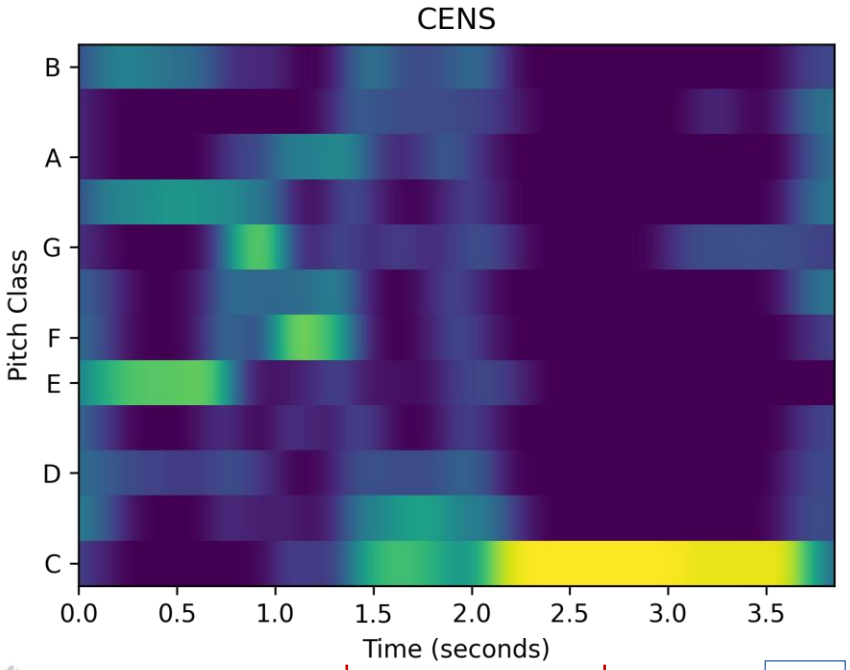
Fig-A3-4

Frequency & Pitch & Chroma

Chromagram

- Chromagram → Energy per pitch-class over time
  - Similar to CQT (semi-tone resolution) without octaves





`Librosa: librosa.feature.chroma_stft`

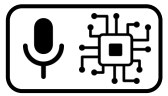
Aud-A3-1

Octave

Own

Outline

- Auditory Scene
- Sound Intensity & Loudness
- Frequency & Pitch & Chroma
- **Timbre**
    - Harmonics / Transients / Noise
    - Mel Frequency Cepstral Coefficients (MFCC)

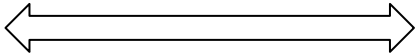


Timbre

- Multi-dimensional perceptual attribute
- Timbre = Difference between musical tones of same pitch & loudness

(Subjective) perceptual attributes

Correlations?



(Objective) sound characteristics

- Temporal / spectral envelope
- Tonal / noise-like components
- Partial (frequency) energies …

Timbre

Harmonics / Transients / Noise

- “Building blocks” of sounds
- Harmonics → horizontal lines in spectrogram (see “Pitch” section)
- Transients → vertical lines in spectrogram
    - High amplitude
    - Short duration
    - Wide-band energy distribution

Timbre

Harmonics / Transients / Noise

- String instruments

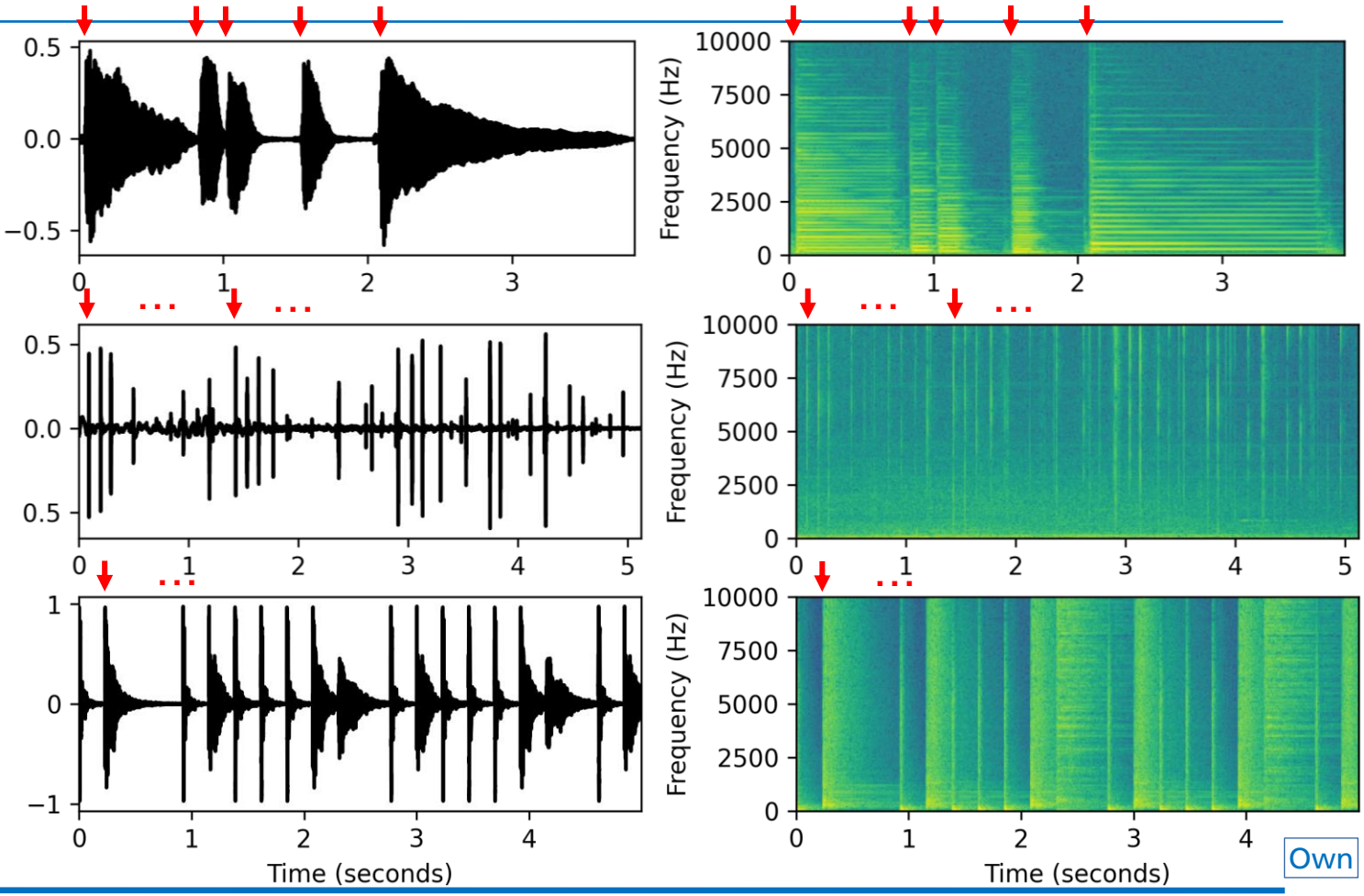
Aud-A3-1

- Bat vocalizations

Aud-A3-2

- Drum instruments

Aud-A3-3



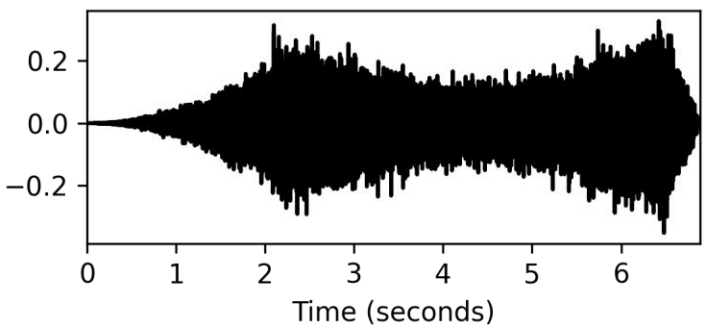
Own

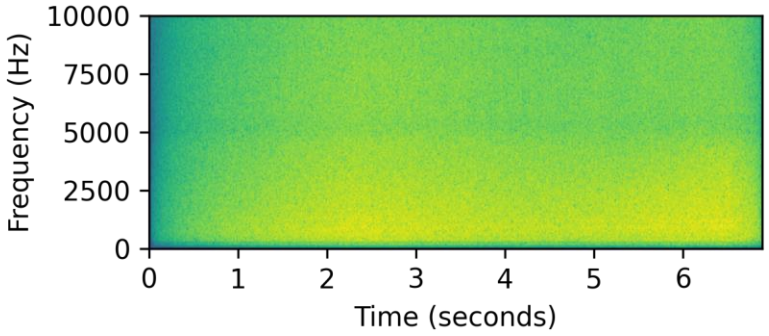
Timbre

Harmonics / Transients / Noise

- Noise
  - Non-periodic, texture-like
  - Random fluctuations of air pressure
- Examples
  - Consonants (speech)
  - Wind (random aerodynamic turbulences)
  - Waves (ocean)

Aud-A3-4





Own

Timbre

Temporal Envelope

- Smooth curve outlining the signal extreme points
- ADSR envelope model (also used for audio synthesis)
    - Attack, Decay, Sustain, Release

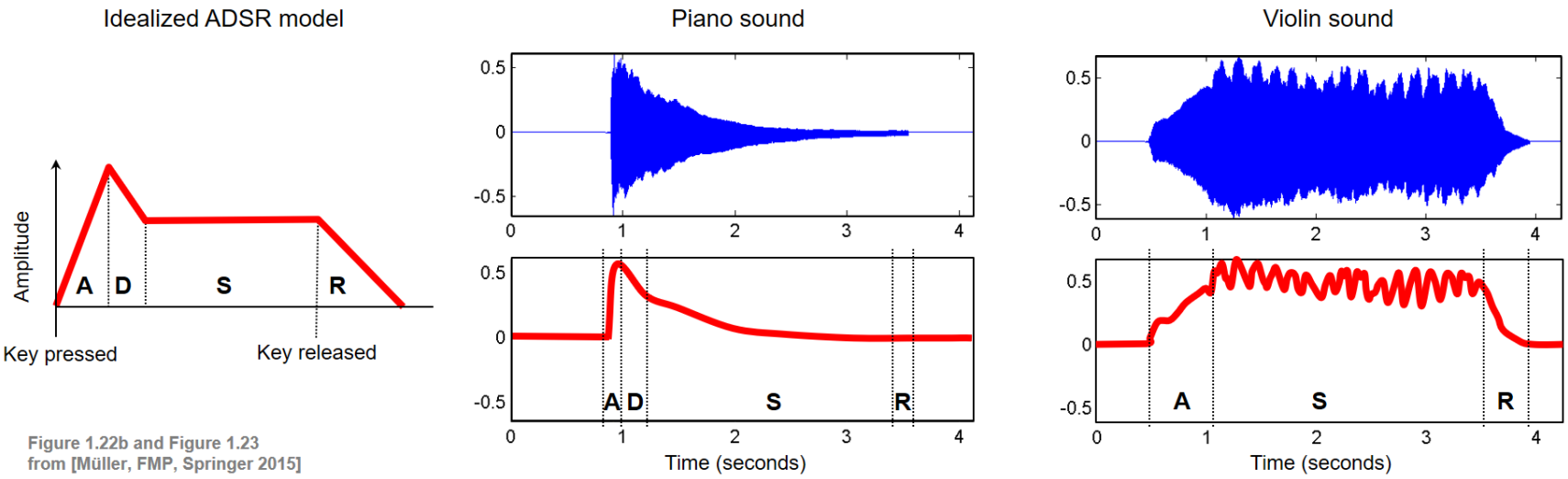


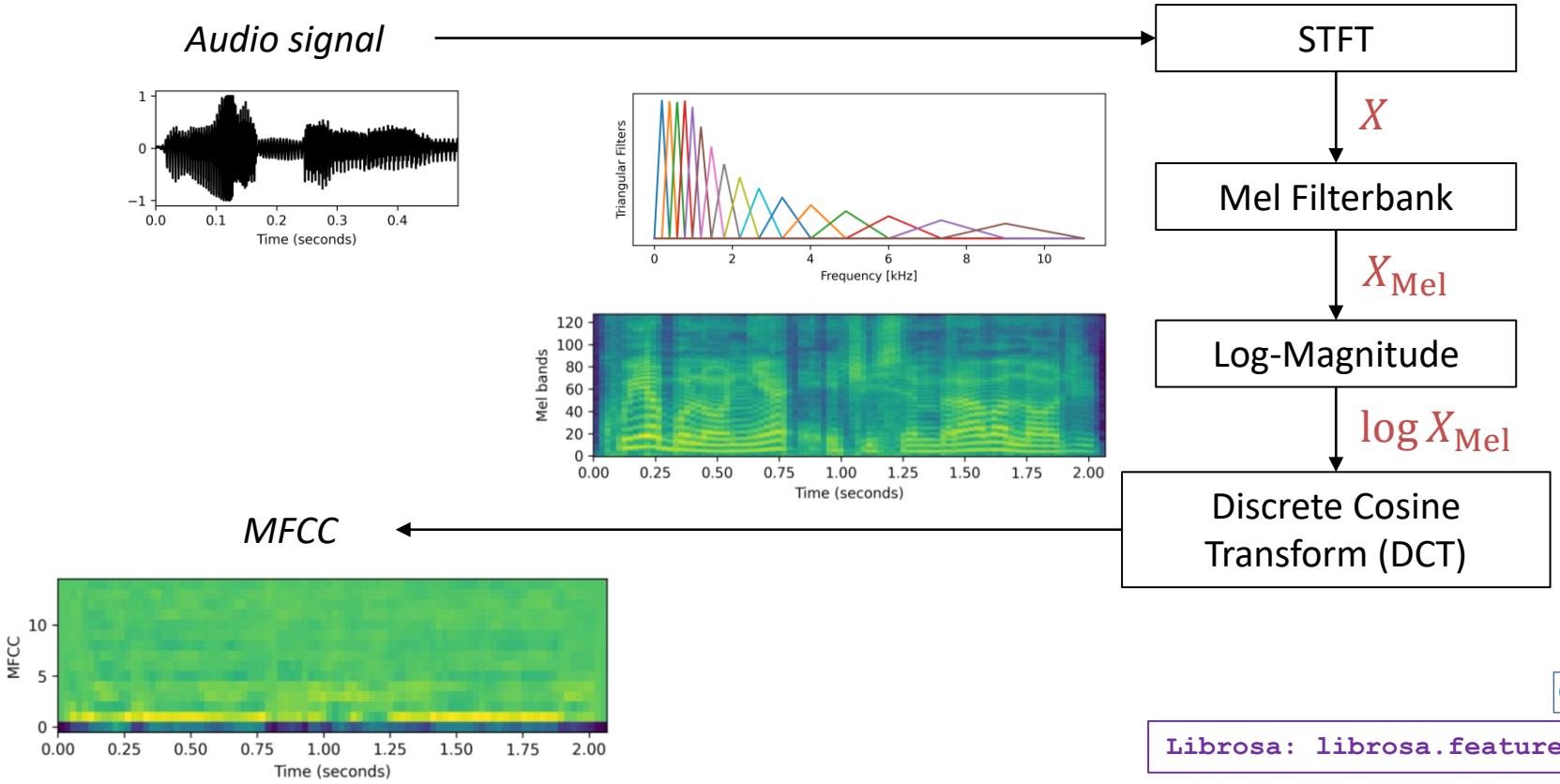
Figure 1.22b and Figure 1.23
from [Müller, FMP, Springer 2015]

Fig-A3-5

Timbre

Mel-Frequency Cepstral Coefficients (MFCCs)

- Compact representation of spectral envelope



Own

`Librosa: librosa.feature.mfcc`

Programming session

Fig-A2-13

References

Images

Fig-A3-1: Iain at English Wikipedia, “Ear-anatomy-text-small-en.svg,” Website https://en.m.wikipedia.org/wiki/File:Ear-anatomy-text-small-en.svg, CC BY-SA 3.0 licence, 2013.

Fig-A3-2: By Lindosland at en.wikipedia - This drawing was created with LibreOffice Draw., Public Domain, https://commons.wikimedia.org/w/index.php?curid=16477782

Fig-A3-3: https://pressbooks.pub/app/uploads/sites/140/2022/07/Piano_to_F.jpg

Fig-A3-4: : M. Müller (2015): Fundamentals of Music Processing (FMP), Springer, 2015, Fig. 3.3a

Fig-A3-5: M. Müller (2015): Fundamentals of Music Processing (FMP), Springer, 2015, Fig. 1.22b & Fig. 1.23

References

Audio

Aud-A3-1: xserra, “piano-phrase.wav,” Website https://freesound.org/people/xserra/sounds/196765/, CC BY 4.0 licence, 2013.

Aud-A3-2: IliasFlou, "bat house.wav,", Website https://freesound.org/people/IliasFlou/sounds/498058/, CC0 licence, 2019.

Aud-A3-3: Daniel Lucas, “Drum beat loop 3,” Website https://freesound.org/people/danlucaz/sounds/517860/, CC0 1.0 licence, 2020.

Aud-A3-4: IENBA, "Seashore", Website https://freesound.org/people/IENBA/sounds/489398/, CC0 licence, 2019.