Computational Analysis of Sound and Music

Audio & Time-Frequency Representations

Dr.-Ing. Jakob Abeßer

Fraunhofer IDMT

jakob.abesser@idmt.fraunhofer.de

Outline

- Sound & Waveform
- Signal Discretization (Sampling & Quantization)
- Short-Time Fourier Transform (STFT)
- Mel Spectrogram
- Constant-Q Spectrogram

Sound & Waveform

Acoustic wave

- Pressure fluctuation
- Emitted from vibrating object (vocal cord, membrane, string, etc.)
- Propagates through transmission medium (air, water)
- Perceived (ear) or recorded (microphone)

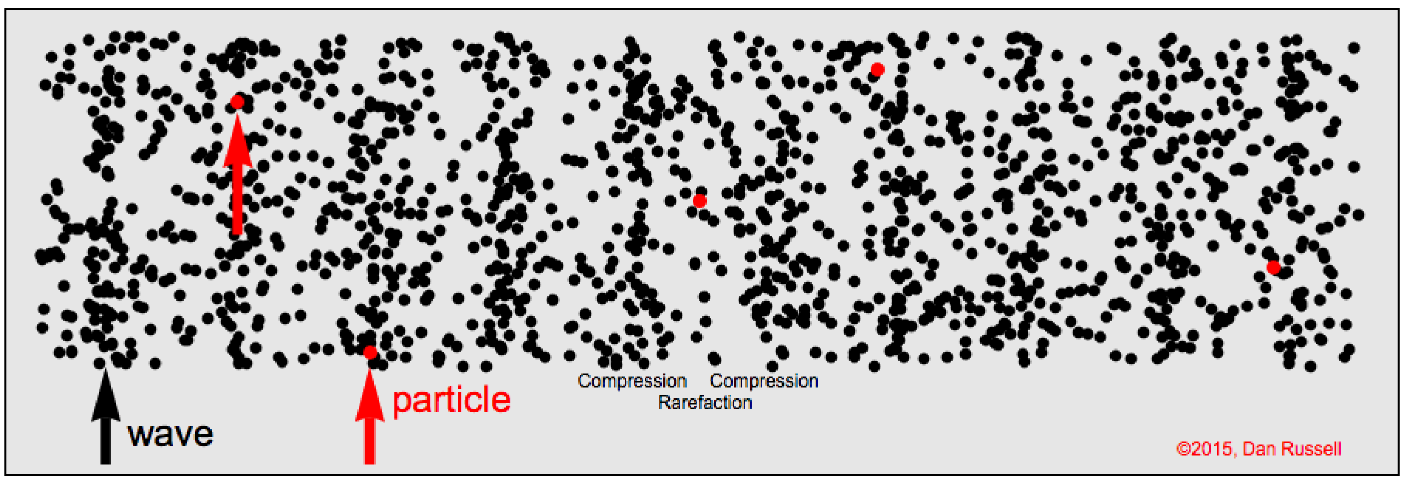
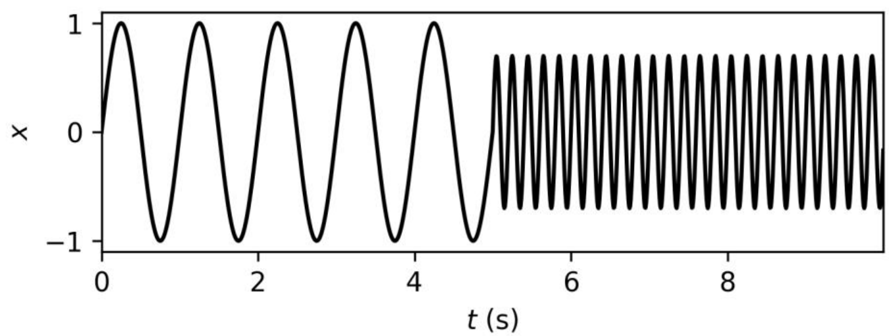


Fig-A2-1

Sound & Waveform

Waveform

- Waveform → amplitude displacement over time (at fixed position)



Own

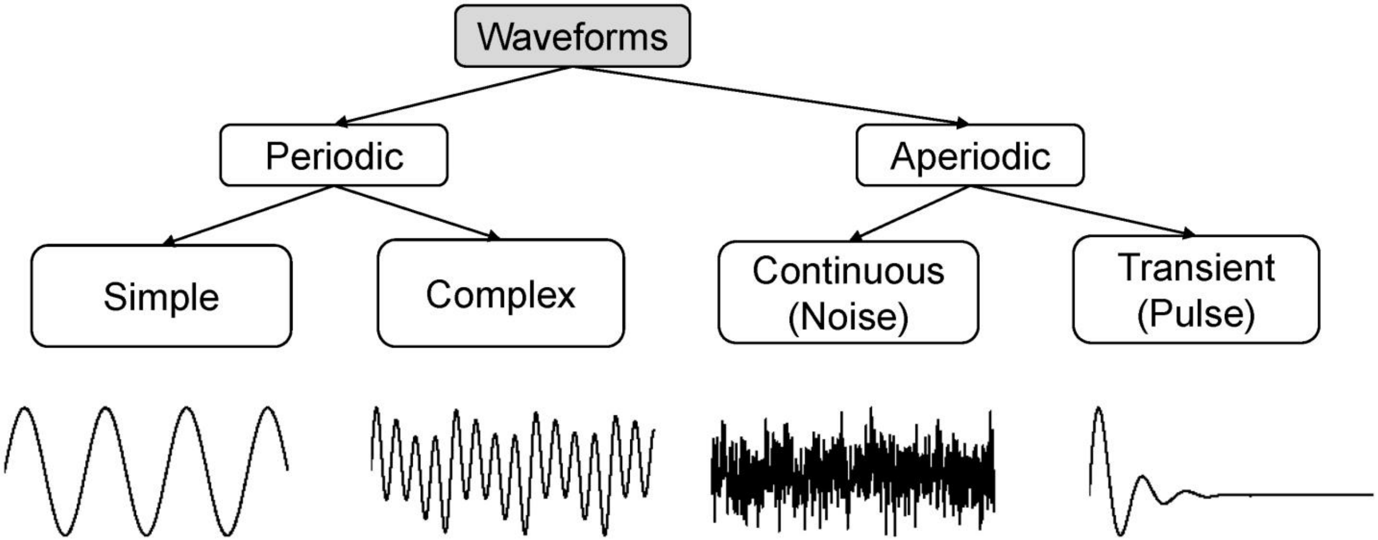
- Periodic signals → wave cycle repeating after period T

$$x(t) = x(t+T) = x(t+2T) = \cdots$$

Sound & Waveform

Waveform

- Categorization

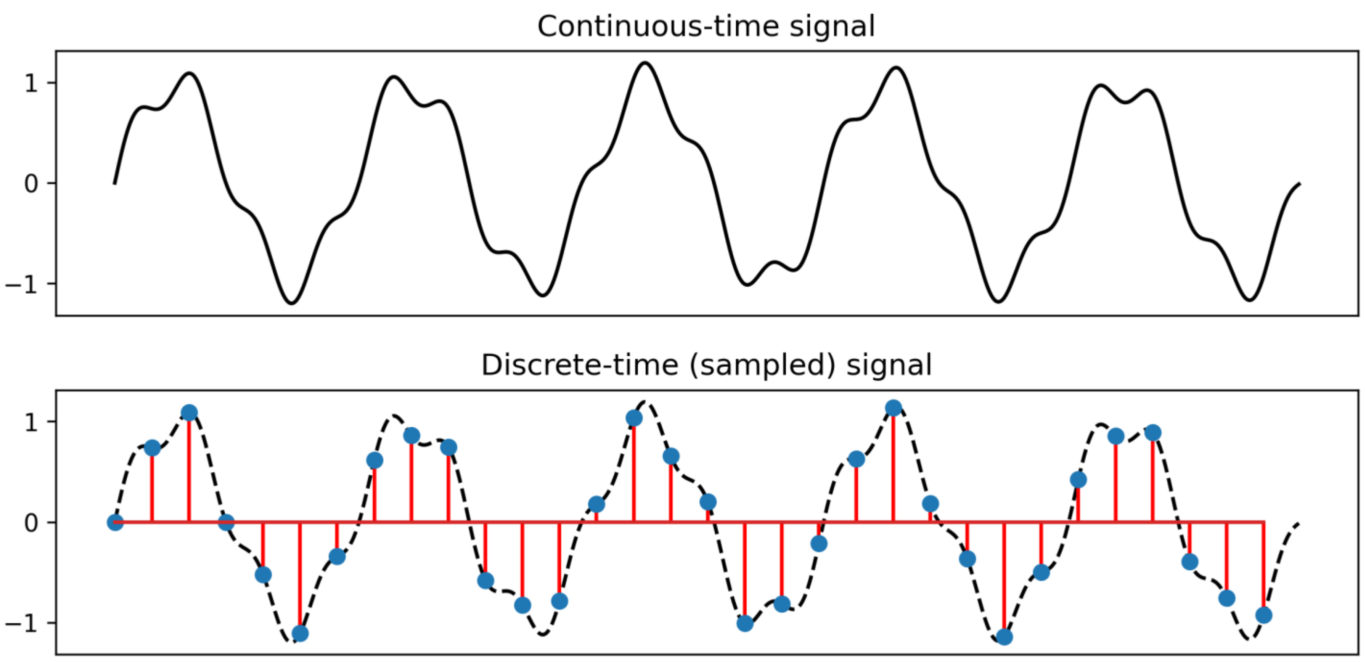


Own

Signal Discretization

Sampling

- Analog waveforms → digital sound signals
- Discrete in time (sampling) $x(n) \coloneqq f(n \cdot T_s) = f(\frac{n}{f_s})$



Own

Signal Discretization

Sampling

- Sampling frequency f_s
    - Commonly 44.1 / 48 / 96 kHz
- Nyquist-Shannon theorem
    - Sampling of signals with limited bandwidth

$$f_s \geq 2 \cdot f_+$$

Signal Discretization

Quantization

- Continuous waveform amplitudes → discrete set of amplitudes
- Binary encoding using b bits → 2^b amplitude values
- Quantization step size
    - $\Delta_\mathrm{q} = \frac{x_+ - x_-}{2^b}$
- Example
    - $x_- = -1$
    - $x_+ = 1$
    - $b = 16$
    - $\Delta_\mathrm{q} \sim 0.00003$

Short-Time Fourier Transform (STFT)

Discrete Fourier Transform (DFT)

- Discrete Fourier Transform (DFT)

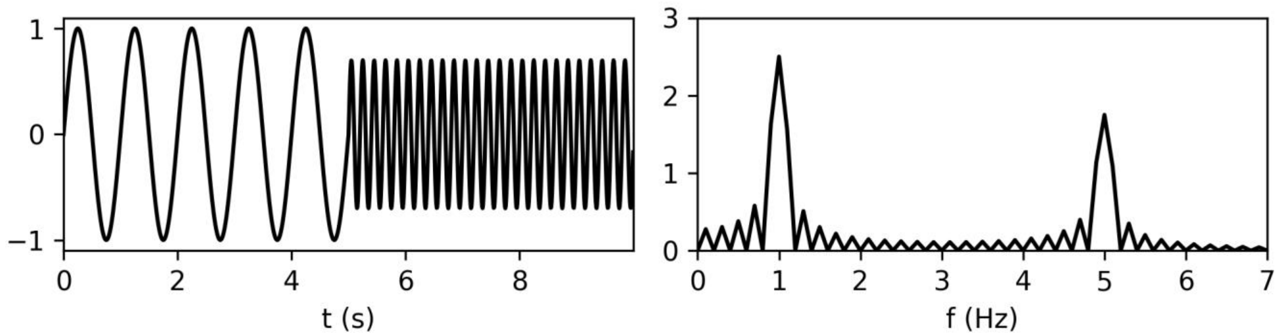
$$\mathcal{X}(k) := \sum_{n=0}^{N-1} x(n) e^{-2\pi i k n / N}$$

  - N number of samples
  - K frequency bands ($k \in [0, K-1]$)
    - Corresponding frequency: $\frac{k \cdot f_s}{N}$
    - Frequency resolution increases with increasing N
  - Efficient implemented as Fast Fourier Transform (FFT)
    - N must be power of 2
  - Magnitude spectrum: $X(k) := |\mathcal{X}(k)|$

Short-Time Fourier Transform (STFT)

Discrete Fourier Transform (DFT)

- Signal approximation using multiple sinusoidal functions
    - Only “global” view on signal
- Example
    - Two consecutive sine signals with frequencies $f = 1\ \mathrm{Hz}$ and $f = 5\ \mathrm{Hz}$



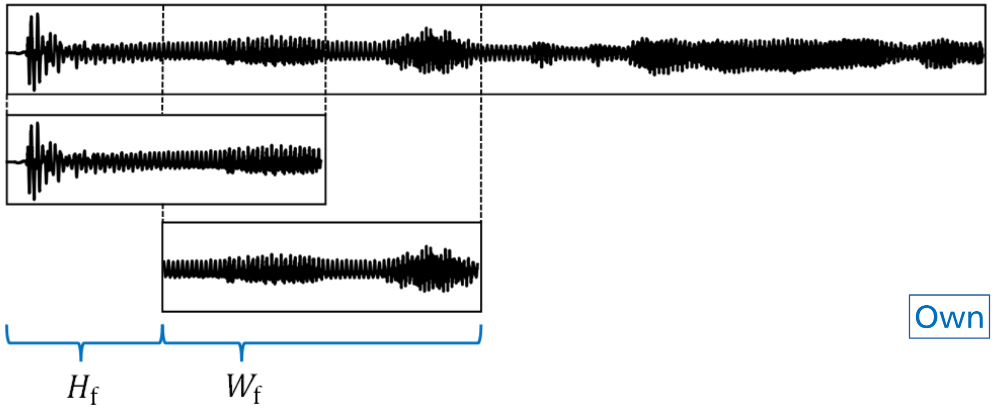
Own

- Time of frequency change cannot be detected in spectrum!

Short-Time Fourier Transform (STFT)

Windowed Signal Analysis

- Windowed signal analysis for "local" view



  - Hopsize H_f
  - Blocksize W_f

- Number of frames $N_\mathrm{f} = \left\lfloor \frac{N - W_\mathrm{f}}{H_\mathrm{f}} \right\rfloor + 1$

Short-Time Fourier Transform (STFT)

STFT

- Short-Time Fourier Transform (STFT)

$$\mathcal{X}(m,k) \coloneqq \sum_{n=0}^{W_\mathrm{f}-1} x(n+mH_\mathrm{f})w(n)e^{-2\pi ikn/W_\mathrm{f}}$$

- Time frame: $m \in [0, N_\mathrm{f}-1] \rightarrow$ Physical time (s): $T_\mathrm{coef} \coloneqq \frac{m \cdot H_\mathrm{f}}{f_\mathrm{s}}$
- Frequency index $k \in [0, K-1]) \rightarrow$ Frequency (Hz): $F_\mathrm{coef} \coloneqq \frac{k \cdot f_\mathrm{s}}{W_\mathrm{f}}$
- Magnitude & phase spectrogram

$$X(m,k) \coloneqq |\mathcal{X}(m,k)|$$

$$\Phi(m,k) \coloneqq \angle\mathcal{X}(m,k)$$

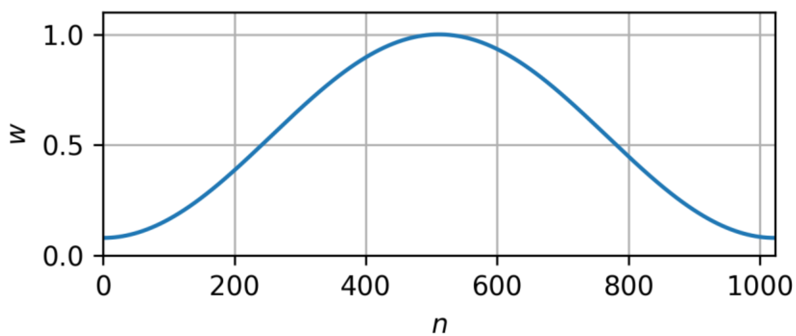
`Librosa: librosa.stft`

Short-Time Fourier Transform (STFT)

STFT

- Window function
    - Reduce spectral leakage
- Example: Hamming window ($N = 1024$)

$$w(n) \coloneqq 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right)$$

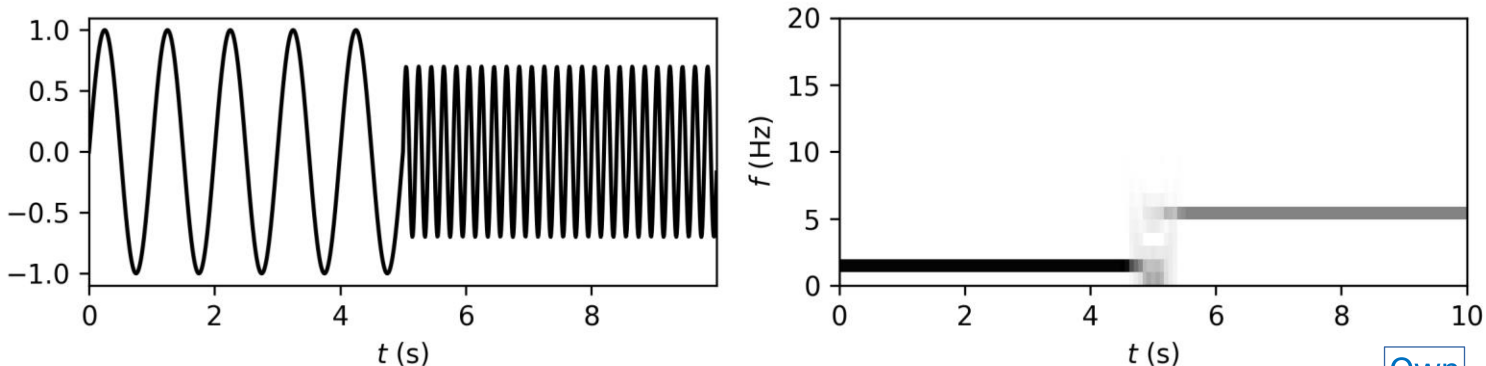


Own

Short-Time Fourier Transform (STFT)

STFT

- Example
  - Two consecutive sine signals with frequencies $f = 1\ \mathrm{Hz}$ and $f = 5\ \mathrm{Hz}$

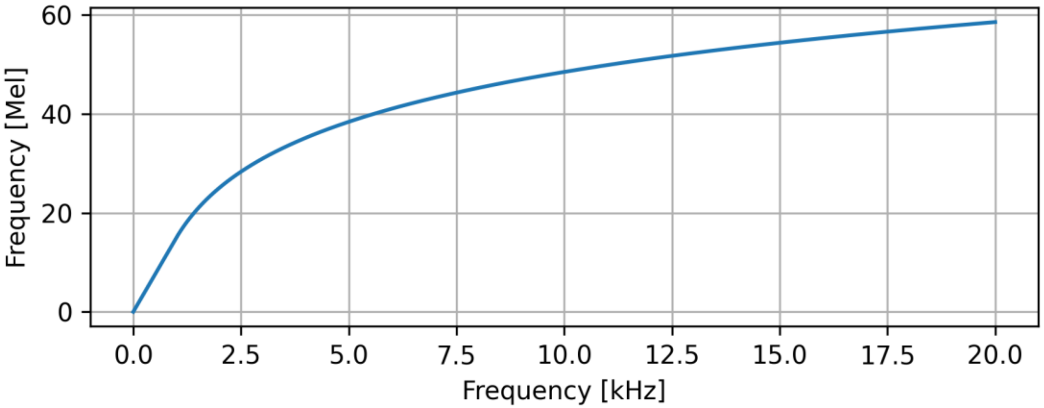


Own

  - Time of frequency change can be detected in the spectrogram!

Mel Spectrogram

- Mel frequency scale
  - Perceptually linear scale to represent pitch perception
  - Two-piece approximation
  - $$f_{\mathrm{Mel}} = \begin{cases} \frac{3 \cdot f}{200} \text{ for } f < 1000 \\ 15 + 27 \log_{6.4}\left(\frac{f}{1000}\right) \text{ for } f \geq 1000 \end{cases}$$



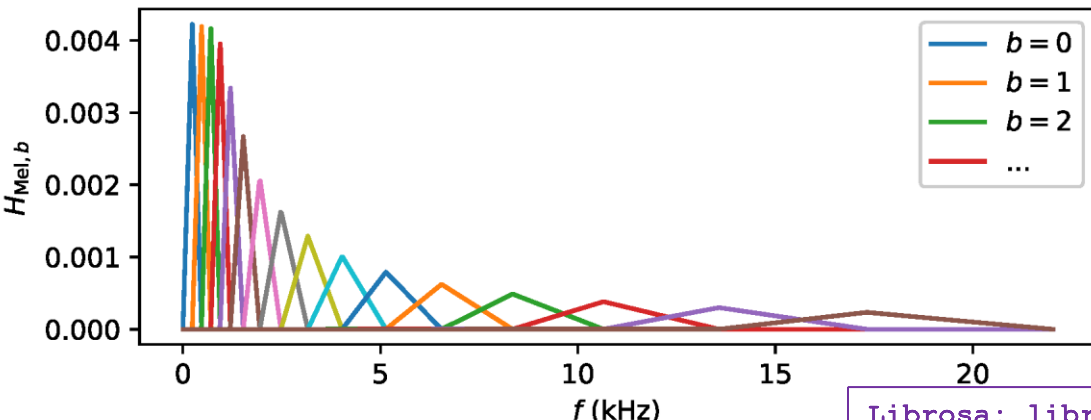
Own

Mel Spectrogram

- Mel spectrogram
  - Local energy distributed along K_{Mel} Mel frequency bands
  - Computed efficiently using triangular filterbank applied to power spectrogram

$$X_{\mathrm{Mel}} \coloneqq H_{\mathrm{Mel}} \times X^2$$

  - Example ($K_{\mathrm{Mel}} = 16$, $f_s = 16\ \mathrm{kHz}$)

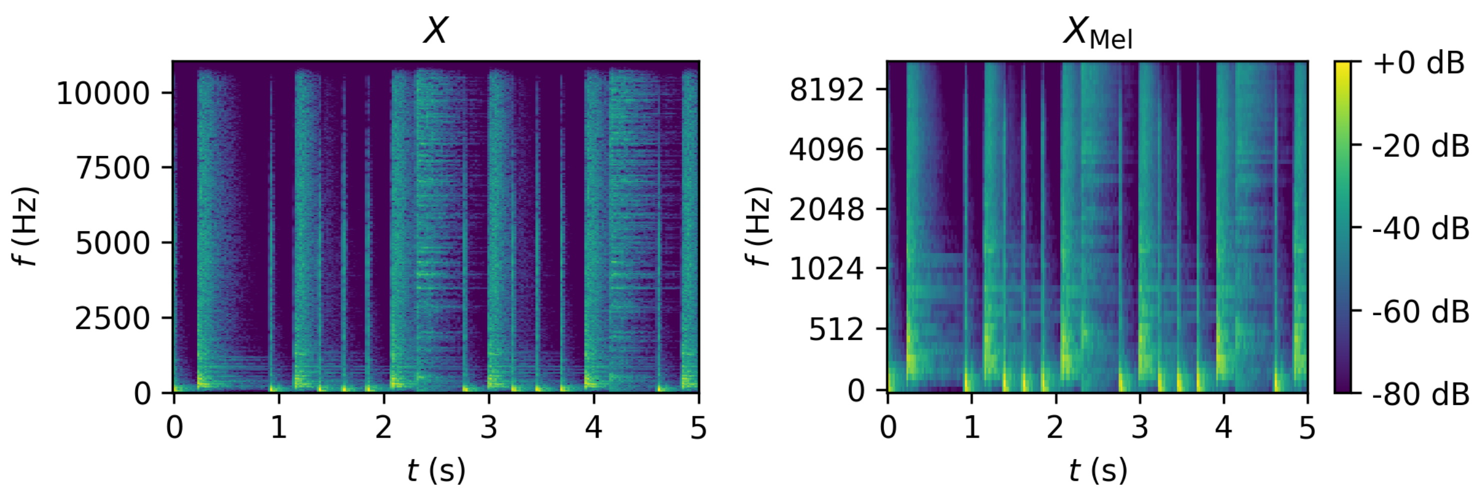


Own

`Librosa: librosa.feature.melspectrogram`

Mel Spectrogram

- Example
  - $K_{\mathrm{Mel}} = 16$, $f_s = 16\ \mathrm{kHz}$
  - Drum beat with kick drum, snare drum, and open hi-hat



Own

Aud-A2-1

  - 513 frequency bands → 16 Mel bands
    - Compression by 96.9% !

Constant-Q Transform (CQT)

- Geometrically spaced center frequencies

$$f_{\mathrm{CQT}}(i) \coloneqq f_{\mathrm{ref}} \cdot 2^{i/b}$$

- Increasing filter bandwidth

$$\Delta_{CQT}(i) \coloneqq f_{i+1} - f_i = f_i \left(2^{1/b} - 1\right)$$

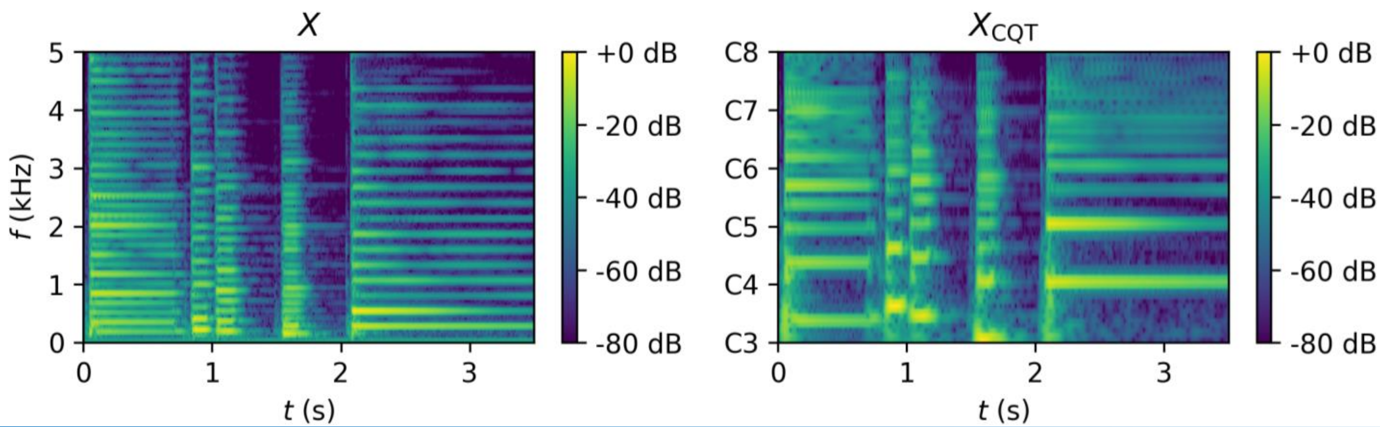
- Constant Q-factor

$$Q(i) \coloneqq \frac{f_{\mathrm{CQT}}(i)}{\Delta_{CQT}(i)} = \frac{1}{2^{1/b} - 1}$$

`Librosa: librosa.cqt`

Constant-Q Transform (CQT)

- Properties
  - Logarithmic frequency binning (harmonic frequencies → fixed shifted pattern)
  - Variable time resolution → Longer analysis windows for low frequencies
- Example
  - Piano melody, semitone resolution ($b = 12$)

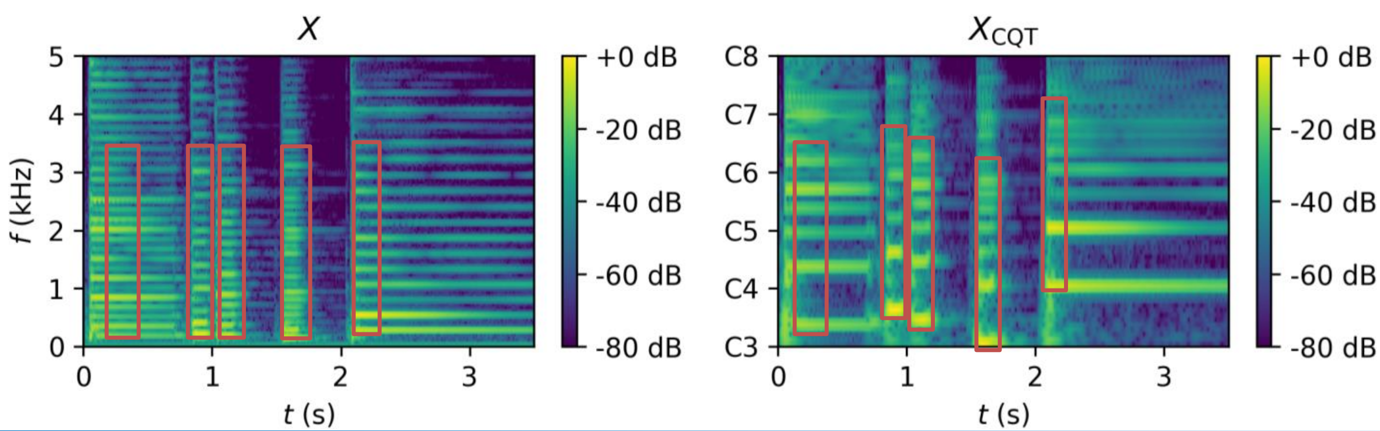


Own

Aud-A2-2

Constant-Q Transform (CQT)

- Properties
  - Logarithmic frequency binning (harmonic frequencies → fixed shifted pattern)
  - Variable time resolution → Longer analysis windows for low frequencies
- Example
  - Piano melody, semitone resolution ($b = 12$)



Own

Aud-A2-2

Programming session

Fig-A2-2

Programming session

References

Images

Fig-A2-1: D. A. Russell: *Acoustics and Vibration Animations* (https://www.acs.psu.edu/drussell/Demos/waves-intro/Lwave-Red-2.gif)

Fig-A2-2: *Jupyter logo* (https://upload.wikimedia.org/wikipedia/commons/thumb/3/38/Jupyter_logo.svg/1200px-Jupyter_logo.svg.png)

References

Audio

Aud-A2-1: Daniel Lucas, "Drum beat loop 3," Website https://freesound.org/people/danlucaz/sounds/517860/, CC0 1.0 licence, 2020.

Aud-A2-2: xserra, "piano-phrase.wav," Website https://freesound.org/people/xserra/sounds/196765/, CC BY 4.0 licence, 2013.