Introduction to Audio Content Analysis

Module 3.4.2: Time-Frequency Representations — Constant Q Transform

alexander lerch

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Section 3.4.2

- **lecture content**
  - constant-Q transform (CQT)
- **learning objectives**
  - discussing advantages and disadvantages of different time-frequency transforms
  - explaining the principles of the CQT and auditory filterbanks

corresponding textbook section

Section 3.4.2

- **lecture content**
  - constant-Q transform (CQT)

- **learning objectives**
  - discussing advantages and disadvantages of different time-frequency transforms
  - explaining the principles of the CQT and auditory filterbanks

non-FT time frequency transforms
introduction

- Fourier transform continues to be much-used tool in audio signal processing and MIR
- but there are disadvantages, e.g.
  - frequency axis does not directly map to (perceptual) pitch axis
  - frequency and time resolution inversely related

$\Rightarrow$ **alternative transforms** can be used

non-FT time frequency transforms
introduction

- Fourier transform continues to be much-used tool in audio signal processing and MIR
- but there are disadvantages, e.g.
  - frequency axis does not directly map to (perceptual) pitch axis
  - frequency and time resolution inversely related

  $\Rightarrow$ **alternative transforms** can be used

constant-Q transform

introduction

- DFT has a *linear* frequency axis:
  - not perceptually meaningful: *logarithmic* is better match
  - low pitch resolution at low frequencies

$\Rightarrow$ compute DFT-like transform at specific frequencies

- space frequencies logarithmically (constant Q)
- resulting abscissa resolution is pitch-related
- parameter c adjusts number of bins per octave

constant-Q transform

introduction

- DFT has a *linear* frequency axis:
  - not perceptually meaningful: *logarithmic* is better match
  - low pitch resolution at low frequencies

$\Rightarrow$ compute DFT-like transform at specific frequencies

- space frequencies logarithmically (constant $\mathcal{Q}$)
- resulting abscissa resolution is pitch-related
- parameter c adjusts number of bins per octave

$$\mathcal{Q} = \frac{f}{\Delta f} = \frac{1}{2^{1/c} - 1}$$

constant Q transform
implementation 1/2

$$X_{\mathrm{CQ}}(k,n) = \frac{1}{\mathcal{K}(k)} \sum_{i=i_{\mathrm{s}}(n)}^{i_{\mathrm{e}}(n)} w_k(i-i_{\mathrm{s}}) \cdot x(i) e^{\mathrm{j}2\pi \frac{\mathcal{Q}\cdot(i-i_{\mathrm{s}})}{\mathcal{K}(k)}}$$

$$\mathcal{K}(k) = \frac{f_{\mathrm{S}}}{f(k)} \mathcal{Q}$$

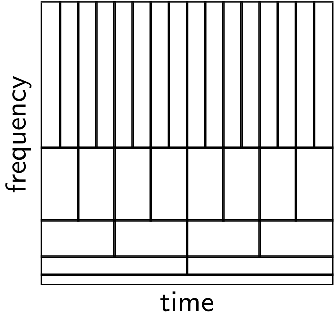
- $f(k)$: frequency of bin index k
- $\mathcal{K}(k)$: blocklength for bin index k
- $\mathcal{Q}$: measure of pitch res.
- w_k: window function
- $i_{\mathrm{s}}, i_{\mathrm{e}}$: start and stop time indices of block
- f_{S}: sample rate

- long window for low frequencies (high freq res, low time res)
- short window for high frequencies (low freq res, high time res)

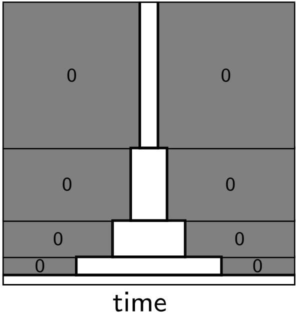
constant Q transform

implementation 2/2

non-overlapping



overlapping



differences

- outputs at multiple vs. one time resolution
- multiple different FFT lengths vs. one FFT length (zero-padded)
- dependent vs. independent definition of block and hop length

constant Q transform
CQT vs. DFT

CQT:

+ perceptually/musically adapted frequency resolution
- time resolution depends on frequency
- not invertible
- no optimized implementation (compare FFT)

constant Q transform
CQT vs. DFT

CQT:

+ perceptually/musically adapted frequency resolution
− time resolution depends on frequency
− not invertible
− no optimized implementation (compare FFT)

CQT:

- $+$ perceptually/musically adapted frequency resolution
- $-$ time resolution depends on frequency
- $-$ not invertible
- $-$ no optimized implementation (compare FFT)

constant Q transform

CQT vs. DFT

CQT:

+ perceptually/musically adapted frequency resolution
– time resolution depends on frequency
– not invertible
– no optimized implementation (compare FFT)

summary
lecture content

- **DFT has disadvantages**
  - low frequency resolution for low pitches
  - non-logarithmic/perceptually relevant pitch resolution
- **CQT**
  - similar to Fourier Transform but logarithmically spaced frequency bins
  - not invertible and inefficient