

Introduction to Audio Content Analysis

module 3.3.2: time-frequency representations — constant Q transform

alexander lerch



introduction overview



corresponding textbook section

section 3.3.2

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



introduction overview



corresponding textbook section

section 3.3.2

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
 - ⇒ 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
 - ⇒ alternative transforms can be used



- DFT has a *linear* frequency axis:
 - not perceptually meaningful: logarithmic is better match
 - low pitch resolution at low frequencies
- - space frequencies logarithmically (constant Q)

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

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

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

$$\mathcal{K}(k) : \text{blocklength for bin index } k$$

$$\mathcal{Q}: \text{ measure of pitch res.}$$

$$\mathcal{W}_k: \text{ window function}$$

$$\mathbf{w}_k: \text{ window function}$$

$$\mathbf{w}_k: \text{ start and stop time indices}$$

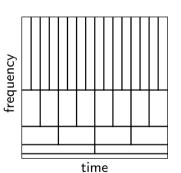
- \bullet i_s , i_e : start and stop time indices of block
- \blacksquare $f_{\rm S}$: sample rate
- long window for low frequencies (high freq res, low time res)
- short window for high frequencies (low freq res, high time res)

CQT oo•o

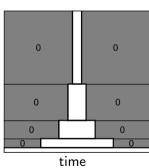
constant Q transform implementation 2/2

Georgia Center for Music Tech Technology

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



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



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



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



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



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

