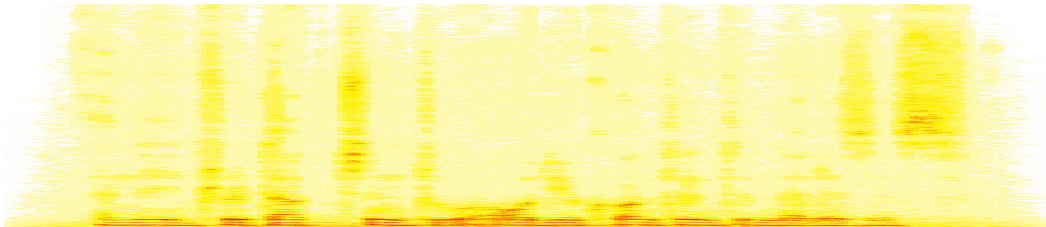


Introduction to Audio Content Analysis

Module 2.6: Fundamentals — Non-Fourier Time-Frequency Transforms

alexander lerch



introduction

overview

corresponding textbook section

Chapter 2 — Fundamentals: pp. 24–26

● lecture content

- constant-Q transform (CQT)
- Gammatone filterbank

● learning objectives

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



introduction

overview

corresponding textbook section

Chapter 2 — Fundamentals: pp. 24–26

- **lecture content**

- constant-Q transform (CQT)
- Gammatone filterbank

- **learning objectives**

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



other 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

other 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

constant-Q transform

introduction

- 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

constant-Q transform

introduction

- 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

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

constant Q transform

implementation 1/2

$$X_{\text{CQ}}(k, n) = \frac{1}{\mathcal{K}(k)} \sum_{i=i_s(n)}^{i_e(n)} w_k(i - i_s) \cdot x(i) e^{j2\pi \frac{\mathcal{Q} \cdot (i - i_s)}{\mathcal{K}(k)}}$$

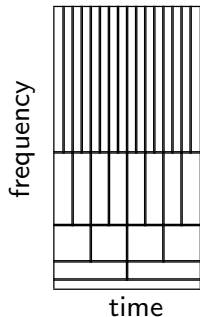
$$\mathcal{K}(k) = \frac{f_s}{f(k)} \mathcal{Q}$$

- $f(k)$: frequency of bin index k
- $\mathcal{K}(k)$: blocklength for bin index k
- \mathcal{Q} : measure of pitch res.
- w_k : window function
- i_s, i_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

- define transformation matrix with maximum window length
 - zeropad higher frequencies (left & right)
- ⇒ 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)

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)

auditory filterbanks

introduction

FT and related transforms bad models of physiological properties of the human ear:

- frequency resolution (critical bands)
- frequency scale (pitch resolution)
- loudness & masking
- event perception & time integration

⇒ **auditory filterbanks**

not as widely used as one might think because

- computationally inefficient
- analysis only: no invertibility (mostly)
- not proven to be superior

auditory filterbanks

introduction

FT and related transforms bad models of physiological properties of the human ear:

- frequency resolution (critical bands)
- frequency scale (pitch resolution)
- loudness & masking
- event perception & time integration

⇒ **auditory filterbanks**

not as widely used as one might think because

- 1 computationally inefficient
- 2 analysis only: no invertibility (mostly)
- 3 not proven to be superior

auditory filterbanks

introduction

FT and related transforms bad models of physiological properties of the human ear:

- frequency resolution (critical bands)
- frequency scale (pitch resolution)
- loudness & masking
- event perception & time integration

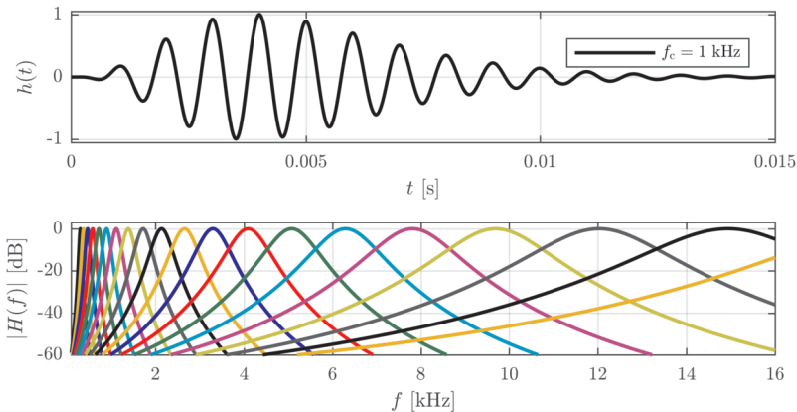
⇒ **auditory filterbanks**

not as widely used as one might think because

- 1 computationally inefficient
- 2 analysis only: no invertibility (mostly)
- 3 not proven to be superior

auditory filterbanks
gammatone filterbank

$$h(i) = \frac{a \cdot (i/f_S)^{O-1} \cdot \cos\left(2\pi \cdot f_c \frac{i}{f_S}\right)}{e^{2\pi i \Delta f / f_S}}$$



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
- **filterbanks**
 - good model of human physiology
 - not invertible and inefficient
 - not proven to be superior

