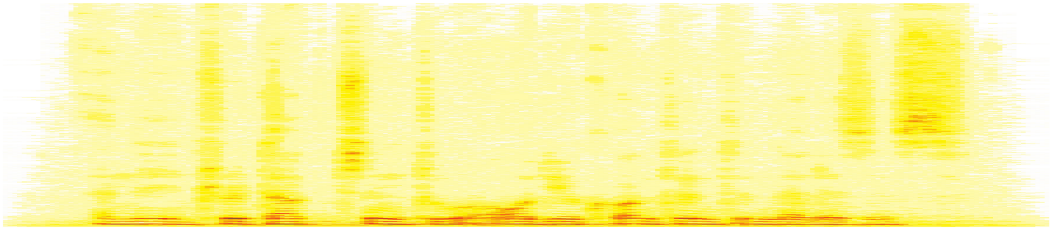
Introduction to Audio Content Analysis

Module 3.2: Feature Extraction — Spectral Shape Features

alexander lerch



Georgia Tech Center for Music Technology
College of Design

introduction

overview

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Chapter 3 — Instantaneous Features: pp. 41–53

- **lecture content**
  - timbre
  - spectral shape instantaneous features
- **learning objectives**
  - describe the general impact of spectral shape on timbre perception
  - summarize features, describe their computation, and discuss their meaning

introduction

overview

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Chapter 3 — Instantaneous Features: pp. 41–53

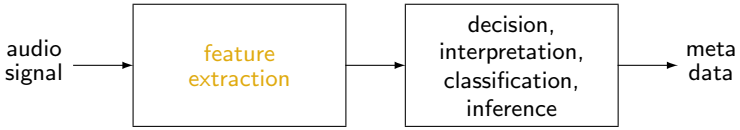
- **lecture content**
  - timbre
  - spectral shape instantaneous features
- **learning objectives**
  - describe the general impact of spectral shape on timbre perception
  - summarize features, describe their computation, and discuss their meaning

introduction

Audio Content Analysis flowchart

Georgia Tech Center for Music Technology
College of Design

remember the flow chart of a general ACA system:



definition (American Standards Association)

...that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar

What is the problem with this definition?

[1] A. S. Bregman, *Auditory Scene Analysis*. MIT Press, 1994.

[2] S. McAdams and A. Bregman, "Hearing Musical Streams," *Computer music journal*, vol. 3, no. 4, pp. 26–60, Dec. 1979, ISSN: 0148-9267.

introduction

timbre 1/2

Georgia Tech Center for Music Technology College of Design

definition (American Standards Association)

...that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar

What is the problem with this definition?

Bregman:[1]

1. implies that timbre *only* exists for sound with pitch!
2. only says that timbre *is not* loudness and pitch

[1] A. S. Bregman, *Auditory Scene Analysis*. MIT Press, 1994.

[2] S. McAdams and A. Bregman, "Hearing Musical Streams," *Computer music journal*, vol. 3, no. 4, pp. 26–60, Dec. 1979, ISSN: 0148-9267.

introduction

timbre 1/2

Georgia Tech Center for Music Technology College of Design

definition (American Standards Association)

...that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar

What is the problem with this definition?

Bregman:[1]

1. implies that timbre *only* exists for sound with pitch!
2. only says that timbre *is not* loudness and pitch

→ [timbre is] "*...the psychoacoustician's multidimensional waste-basket category for everything that cannot be labeled pitch or loudness.*"[2]

[1] A. S. Bregman, *Auditory Scene Analysis*. MIT Press, 1994.

[2] S. McAdams and A. Bregman, "Hearing Musical Streams," *Computer music journal*, vol. 3, no. 4, pp. 26–60, Dec. 1979, ISSN: 0148-9267.

introduction

timbre 2/2

timbre is

- a function of **temporal envelope**
  - attack time characteristics
  - amplitude modulations
  - ...
- a function of **spectral distribution**
  - spectral envelope
  - number of partials
  - energy distribution of partials
  - ...

when dealing with complex mixtures of sound, it is very difficult (maybe impossible?) to extract detailed temporal information for individual tones

$\Rightarrow$ timbre features typically focus on the **spectral shape**

introduction
timbre 2/2

timbre is

- a function of **temporal envelope**
  - attack time characteristics
  - amplitude modulations
  - ...
- a function of **spectral distribution**
  - spectral envelope
  - number of partials
  - energy distribution of partials
  - ...

when dealing with complex mixtures of sound, it is very difficult (maybe impossible?) to extract detailed temporal information for individual tones

$\Rightarrow$ timbre features typically focus on the **spectral shape**

introduction
timbre 2/2

timbre is

- a function of **temporal envelope**
  - attack time characteristics
  - amplitude modulations
  - ...
- a function of **spectral distribution**
  - spectral envelope
  - number of partials
  - energy distribution of partials
  - ...

when dealing with complex mixtures of sound, it is very difficult (maybe impossible?) to extract detailed temporal information for individual tones

$\Rightarrow$ timbre features typically focus on the **spectral shape**

introduction
timbre 2/2

timbre is

- a function of **temporal envelope**
  - attack time characteristics
  - amplitude modulations
  - ...
- a function of **spectral distribution**
  - spectral envelope
  - number of partials
  - energy distribution of partials
  - ...

when dealing with complex mixtures of sound, it is very difficult (maybe impossible?) to extract detailed temporal information for individual tones

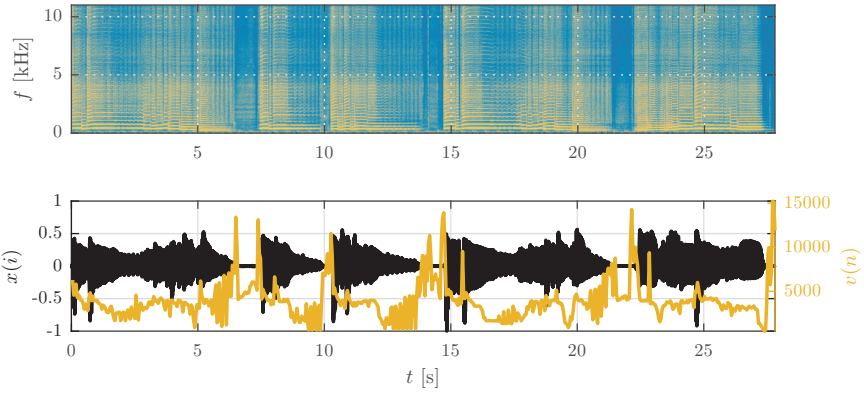
$\Rightarrow$ timbre features typically focus on the **spectral shape**

spectral shape features

spectral rolloff

Georgia Tech Center for Music Technology College of Design

$$v_{\mathrm{SR}}(n) = i \quad \text{at} \sum_{k=0}^{i} |X(k,n)| = \kappa \cdot \sum_{k=0}^{K/2-1} |X(k,n)|$$



matlab source: matlab/displayFeatures.m

spectral shape features

spectral rolloff

Georgia Tech Center for Music Technology College of Design

$$v_{\mathrm{SR}}(n) = i \quad \text{at} \sum_{k=0}^{i} |X(k,n)| = \kappa \cdot \sum_{k=0}^{\mathcal{K}/2-1} |X(k,n)|$$

common variants:

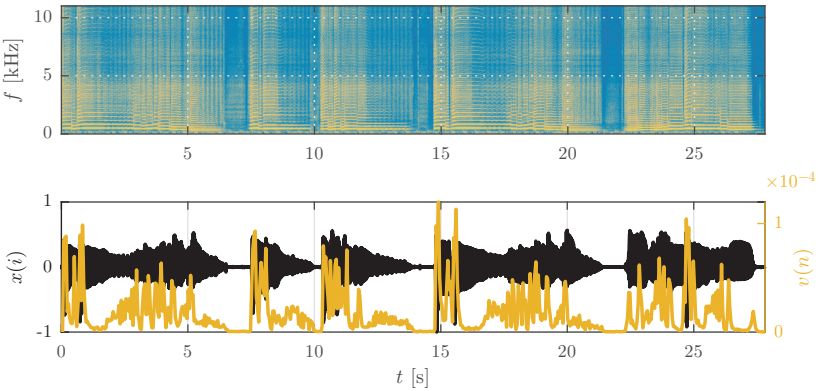
- scaled to frequency
- power spectrum

spectral shape features

spectral flux

Georgia Tech Center for Music Technology College of Design

$$v_{\mathrm{SF}}(n) = \frac{\sqrt{\sum_{k=0}^{\mathcal{K}/2-1} \left(|X(k,n)| - |X(k,n-1)|\right)^2}}{\mathcal{K}/2}$$



matlab source: matlab/displayFeatures.m

spectral shape features

spectral flux

Georgia Tech Center for Music Technology College of Design

$$v_{\mathrm{SF}}(n) = \frac{\sqrt{\sum_{k=0}^{\mathcal{K}/2-1} \left(|X(k,n)| - |X(k,n-1)|\right)^2}}{\mathcal{K}/2}$$

common variants:

$$v_{\mathrm{SF}}(n,\beta) = \frac{\sqrt[\beta]{\sum_{k=0}^{\mathcal{K}/2-1} \left(|X(k,n)| - |X(k,n-1)|\right)^\beta}}{\mathcal{K}/2}$$

$$v_{\mathrm{SF},\sigma}(n) = \sqrt{\frac{2}{\mathcal{K}} \sum_{k=0}^{\mathcal{K}/2-1} \left(\Delta X(k,n) - \mu_{\Delta X}\right)^2}$$

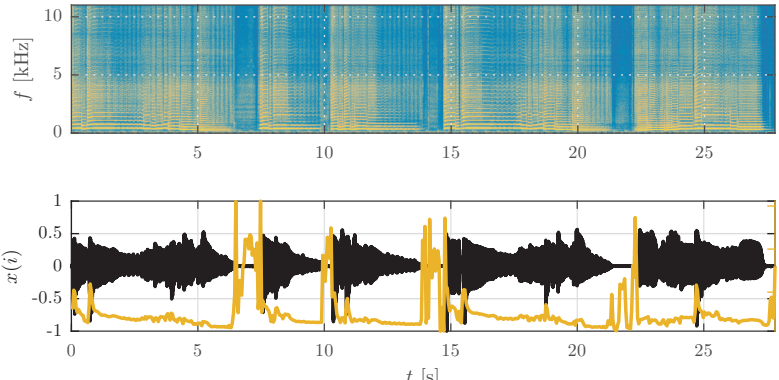
$$v_{\mathrm{SF},\log}(n) = \frac{2}{\mathcal{K}} \sum_{k=0}^{\mathcal{K}/2-1} \log_2\left(\frac{|X(k,n)|}{|X(k,n-1)|}\right)$$

spectral shape features

spectral centroid

Georgia Tech Center for Music Technology
College of Design

$$v_{\mathrm{SC}}(n) = \frac{\sum_{k=0}^{\mathcal{K}/2-1} k \cdot |X(k,n)|}{\sum_{k=0}^{\mathcal{K}/2-1} |X(k,n)|}$$



matlab source: matlab/displayFeatures.m

spectral shape features

spectral centroid

Georgia Tech Center for Music Technology College of Design

$$v_{\mathrm{SC}}(n) = \frac{\sum_{k=0}^{\mathcal{K}/2-1} k \cdot |X(k,n)|}{\sum_{k=0}^{\mathcal{K}/2-1} |X(k,n)|}$$

common variants:

- power spectrum
- logarithmic frequency scale

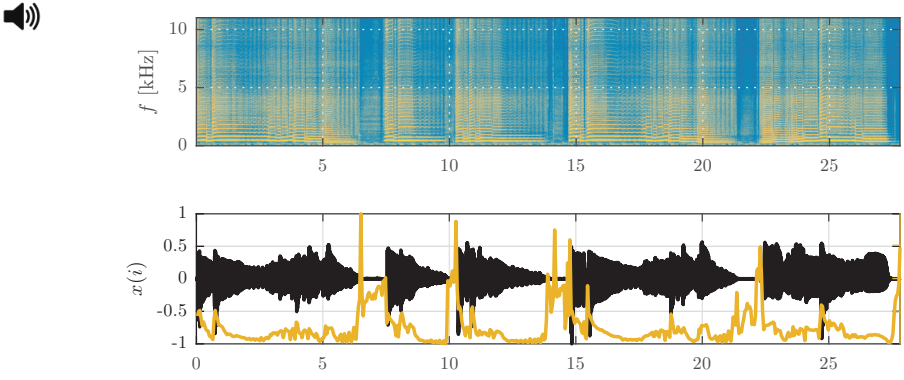
$$v_{\mathrm{SC,log}}(n) = \frac{\sum_{k=k(f_{\min})}^{\mathcal{K}/2-1} \log_2\left(\frac{f(k)}{f_{\mathrm{ref}}}\right) \cdot |X(k,n)|^2}{\sum_{k=k(f_{\min})}^{N/2-1} |X(k,n)|^2}$$

spectral shape features

spectral spread

Georgia Tech Center for Music Technology College of Design

$$v_{\mathrm{SS}}(n) = \sqrt{\frac{\sum_{k=0}^{\mathcal{K}/2-1} (k - v_{\mathrm{SC}}(n))^2 \cdot |X(k,n)|^2}{\sum_{k=0}^{\mathcal{K}/2-1} |X(k,n)|^2}}$$



matlab source: matlab/displayFeatures.m

spectral shape features

spectral spread

Georgia Tech Center for Music Technology College of Design

$$v_{\mathrm{SS}}(n) = \sqrt{\frac{\sum_{k=0}^{\mathcal{K}/2-1} (k - v_{\mathrm{SC}}(n))^2 \cdot |X(k,n)|^2}{\sum_{k=0}^{\mathcal{K}/2-1} |X(k,n)|^2}}$$

common variants:

- same variants as with *Spectral Centroid*, e.g. logarithmic:

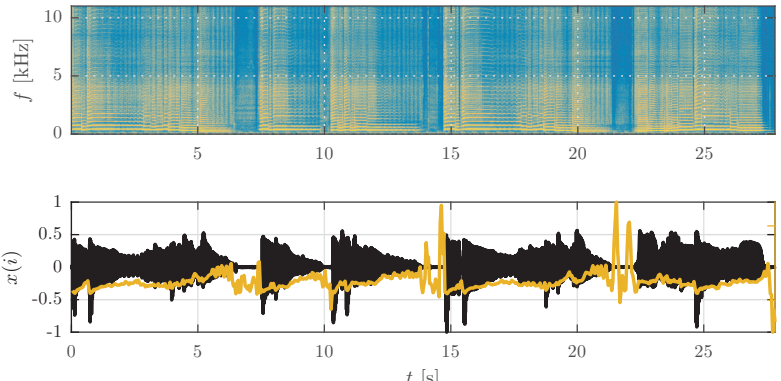
$$v_{\mathrm{SS,log}}(n) = \sqrt{\frac{\sum_{k=k(f_{\min})}^{\mathcal{K}/2-1} \left(\log_2\left(\frac{f(k)}{1000\ \mathrm{Hz}}\right) - v_{\mathrm{SC}}(n)\right)^2 \cdot |X(k,n)|^2}{\sum_{k=k(f_{\min})}^{\mathcal{K}/2-1} |X(k,n)|^2}}$$

spectral shape features

spectral decrease

Georgia Tech Center for Music Technology College of Design

$$v_{\mathrm{SD}}(n) = \frac{\sum_{k=1}^{\mathcal{K}/2-1} \frac{1}{k} \cdot \left(|X(k,n)| - |X(0,n)|\right)}{\sum_{k=1}^{\mathcal{K}/2-1} |X(k,n)|}$$



matlab source: matlab/displayFeatures.m

spectral shape features

spectral decrease

Georgia Tech Center for Music Technology College of Design

$$v_{\mathrm{SD}}(n) = \frac{\sum_{k=1}^{\mathcal{K}/2-1} \frac{1}{k} \cdot \left(|X(k,n)| - |X(0,n)|\right)}{\sum_{k=1}^{\mathcal{K}/2-1} |X(k,n)|}$$

common variants:

- restricted frequency range:

$$v_{\mathrm{SD}}(n) = \frac{\sum_{k=k_{\mathrm{l}}}^{k_{\mathrm{u}}} \frac{1}{k} \cdot \left(|X(k,n)| - |X(k_{\mathrm{l}}-1,n)|\right)}{\sum_{k=k_{\mathrm{l}}}^{k_{\mathrm{u}}} |X(k,n)|}$$

Georgia Tech Center for Music Technology
College of Design

signal model:
convolution of *excitation signal* and *transfer function*

$$x(i) = e(i) * h(i)$$

$$X(\mathrm{j}\omega) = E(\mathrm{j}\omega) \cdot H(\mathrm{j}\omega)$$

$$\begin{aligned} \log\left(X(\mathrm{j}\omega)\right) &= \log\left(E(\mathrm{j}\omega) \cdot H(\mathrm{j}\omega)\right) \\ &= \log\left(E(\mathrm{j}\omega)\right) + \log\left(H(\mathrm{j}\omega)\right) \end{aligned}$$

signal model:
convolution of *excitation signal* and *transfer function*

$$x(i) = e(i) * h(i)$$

$$X(\mathrm{j}\omega) = E(\mathrm{j}\omega) \cdot H(\mathrm{j}\omega)$$

$$\begin{aligned} \log\left(X(\mathrm{j}\omega)\right) &= \log\left(E(\mathrm{j}\omega) \cdot H(\mathrm{j}\omega)\right) \\ &= \log\left(E(\mathrm{j}\omega)\right) + \log\left(H(\mathrm{j}\omega)\right) \end{aligned}$$

signal model:
convolution of *excitation signal* and *transfer function*

$$x(i) = e(i) * h(i)$$

$$X(\mathrm{j}\omega) = E(\mathrm{j}\omega) \cdot H(\mathrm{j}\omega)$$

$$\begin{aligned} \log\big(X(\mathrm{j}\omega)\big) &= \log\big(E(\mathrm{j}\omega) \cdot H(\mathrm{j}\omega)\big) \\ &= \log\big(E(\mathrm{j}\omega)\big) + \log\big(H(\mathrm{j}\omega)\big) \end{aligned}$$

fundamentals

cepstrum 2/3

Georgia Tech Center for Music Technology College of Design

$$
\begin{aligned}
c_x(i) &= \mathfrak{F}^{-1}\left\{\log\left(X(\mathrm{j}\omega)\right)\right\} \\
&= \mathfrak{F}^{-1}\left\{\log\left(E(\mathrm{j}\omega)\right) + \log\left(H(\mathrm{j}\omega)\right)\right\} \\
&= \mathfrak{F}^{-1}\left\{\log\left(E(\mathrm{j}\omega)\right)\right\} + \mathfrak{F}^{-1}\left\{\log\left(H(\mathrm{j}\omega)\right)\right\} \\
\hat{c}_x(i_s(n)\ldots i_e(n)) &= \sum_{k=0}^{\kappa/2-1} \log\left(|X(k,n)|\right) e^{\mathrm{j}ki\Delta\Omega}
\end{aligned}
$$

fundamentals

cepstrum 2/3

Georgia Tech Center for Music Technology
College of Design

$$
\begin{aligned}
c_x(i) &= \mathfrak{F}^{-1}\left\{\log\left(X(\mathrm{j}\omega)\right)\right\} \\
&= \mathfrak{F}^{-1}\left\{\log\left(E(\mathrm{j}\omega)\right) + \log\left(H(\mathrm{j}\omega)\right)\right\} \\
&= \mathfrak{F}^{-1}\left\{\log\left(E(\mathrm{j}\omega)\right)\right\} + \mathfrak{F}^{-1}\left\{\log\left(H(\mathrm{j}\omega)\right)\right\} \\
\hat{c}_x(i_\mathrm{s}(n)\ldots i_\mathrm{e}(n)) &= \sum_{k=0}^{\kappa/2-1} \log\left(|X(k,n)|\right) e^{\mathrm{j}ki\Delta\Omega}
\end{aligned}
$$

fundamentals

cepstrum 2/3

Georgia Tech Center for Music Technology
College of Design

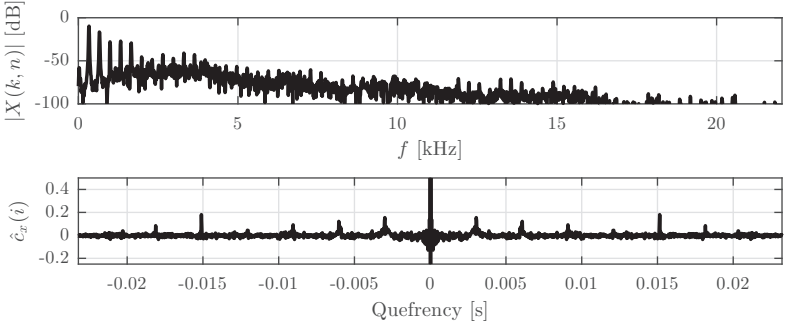
$$
\begin{aligned}
c_x(i) &= \mathfrak{F}^{-1}\left\{\log\left(X(\mathrm{j}\omega)\right)\right\} \\
&= \mathfrak{F}^{-1}\left\{\log\left(E(\mathrm{j}\omega)\right) + \log\left(H(\mathrm{j}\omega)\right)\right\} \\
&= \mathfrak{F}^{-1}\left\{\log\left(E(\mathrm{j}\omega)\right)\right\} + \mathfrak{F}^{-1}\left\{\log\left(H(\mathrm{j}\omega)\right)\right\} \\
\hat{c}_x(i_\mathrm{s}(n)\ldots i_\mathrm{e}(n)) &= \sum_{k=0}^{\mathcal{K}/2-1} \log\left(|X(k,n)|\right) e^{\mathrm{j}ki\Delta\Omega}
\end{aligned}
$$

fundamentals

cepstrum 2/3

Georgia Tech Center for Music Technology College of Design

$$
\begin{aligned}
c_x(i) &= \mathfrak{F}^{-1}\left\{\log\left(X(\mathrm{j}\omega)\right)\right\} \\
&= \mathfrak{F}^{-1}\left\{\log\left(E(\mathrm{j}\omega)\right) + \log\left(H(\mathrm{j}\omega)\right)\right\} \\
&= \mathfrak{F}^{-1}\left\{\log\left(E(\mathrm{j}\omega)\right)\right\} + \mathfrak{F}^{-1}\left\{\log\left(H(\mathrm{j}\omega)\right)\right\} \\
\hat{c}_x(i_\mathrm{s}(n)\ldots i_\mathrm{e}(n)) &= \sum_{k=0}^{\mathcal{K}/2-1} \log\left(|X(k,n)|\right) e^{\mathrm{j}ki\Delta\Omega}
\end{aligned}
$$



matlab source: matlab/displayCepstrum.m

fundamentals

cepstrum 3/3

Georgia Tech Center for Music Technology
College of Design

- **summary**:
  - cepstrum 'replaces' time domain convolution operation with addition
  - result is the *unfiltered* excitation signal *plus* the filter IR (both logarithmic)
  - can be used for, e.g., *spectral envelope extraction* or *pitch detection*

- more naming silliness:
  cepstrum, quefrency, liftering, ...

fundamentals

cepstrum 3/3

- **summary**:
  - cepstrum 'replaces' time domain convolution operation with addition
  - result is the *unfiltered* excitation signal *plus* the filter IR (both logarithmic)
  - can be used for, e.g., *spectral envelope extraction* or *pitch detection*

  - more naming silliness:
    cepstrum, quefrency, liftering, ...

spectral shape features

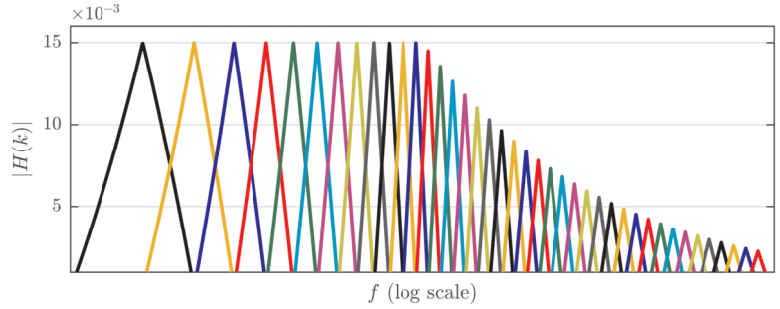
mel frequency cepstral coefficients 1/4

- typical processing steps for the mel frequency cepstral coefficients (MFCCs):
  1. compute magnitude spectrum
  2. convert linear frequency scale to logarithmic
  3. group bins into bands
  4. apply logarithm to all bands
  5. compute (inverse) cosine transform (DCT)

$$v^j_{\mathrm{MFCC}}(n) = \sum_{k'=1}^{\mathcal{K}'} \log\left(|X'(k',n)|\right) \cdot \cos\left(j \cdot \left(k' - \frac{1}{2}\right) \frac{\pi}{\mathcal{K}'}\right)$$

spectral shape features

mel frequency cepstral coefficients 2/4



matlab source: matlab/displayMfccFilterbank.m

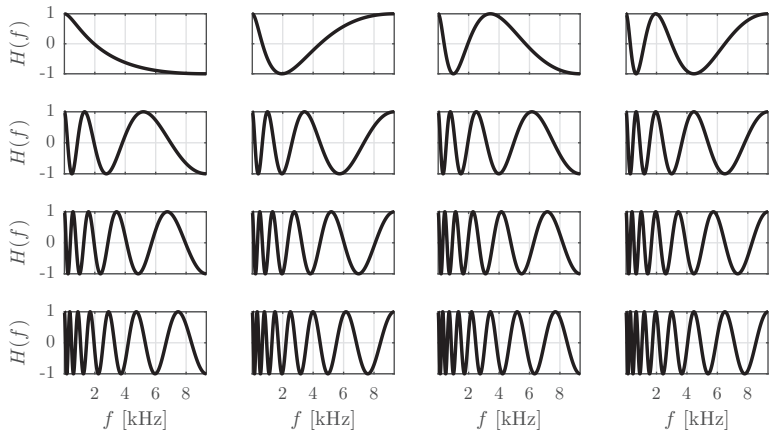
- constant Q filter spacing for higher frequencies (mel scale)
- FFT values are weighted and summed over bins for each band

spectral shape features

mel frequency cepstral coefficients 3/4

Georgia Tech Center for Music Technology College of Design

mel-warped cosine bases for DCT



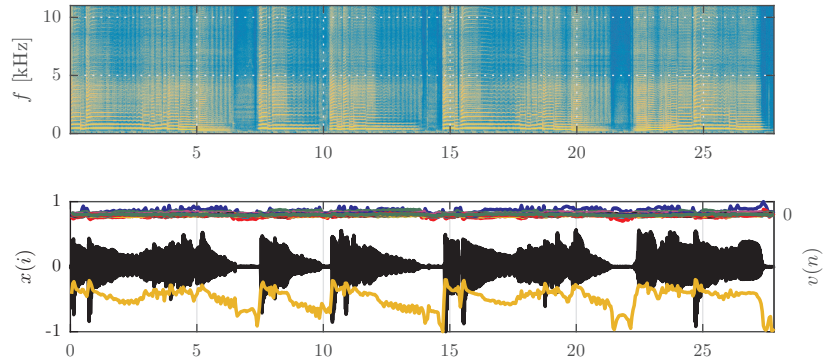
matlab source: matlab/displayMfccMelDct.m

spectral shape features

mel frequency cepstral coefficients 4/4

Georgia Tech Center for Music Technology College of Design

Property	DM	HTK	SAT
Num. filters	20	24	40
Mel scale	lin/log	log	lin/log
Freq. range	[100; 4000]	[100; 4000]	[200; 6400]
Normalization	Equal height	Equal height	Equal area



matlab source: matlab/displayFeatures.m

summary

lecture content

Georgia Tech Center for Music Technology
College of Design

- **timbre**
  - mostly dependent on both spectral shape and time domain envelope characteristics
  - multi-dimensional perceptual property not as clearly defined as pitch or loudness
- **instantaneous spectral shape features**
  - established set of baseline features
  - usually extracted from the magnitude spectrum
  - condensing various properties of the spectral shape into single values
  - there exist multiple variants of "the same" feature
- **cepstrum**
  - approach to "separate" excitation from filter signal
  - makes use of the logarithm turning spectral multiplication into addition