Advanced Course Computer Science

Music Processing

Summer Term 2009

Meinard Müller

Saarland University and MPI Informatik
meinard@mpi-inf.mpg.de

Audio Features

UNIVERSITÄT DES SAARLANDES

max planck institut informatik

Fourier Transform

- Tells **which** notes (frequencies) are played, but does not tell **when** the notes are played
- Frequency information is averaged over the entire time interval
- Time information is hidden in the phase

$\rightarrow$ Windowed Fourier Transform (WFT)
Short Time Fourier Transform (STFT)
(Dennis Gabor, 1946)

Short Time Fourier Transform

Idea:

- To recover time information, only a **small section** of the signal is used for the spectral analysis
- This section is determined by a **window function** $g:\mathbb{R}\to\mathbb{R}$ ($g\in L^2(\mathbb{R})$, $\|g\|=1$)

Definition:

STFT w.r.t. *g* of a signal $f:\mathbb{R}\to\mathbb{R}$

$$\tilde{f}(\omega,t) := \int_{\mathbb{R}} f(u)\bar{g}(u-t)e^{-2\pi i\omega u}du = \langle f|g_{\omega,t}\rangle$$

with $g_{\omega,t}(u) := e^{2\pi i\omega u}g(u-t), \quad u\in\mathbb{R}$

Short Time Fourier Transform

Interpretation:

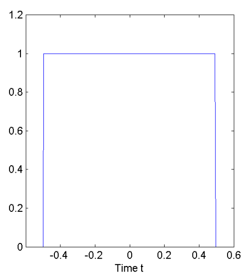
- $g_{\omega,t}$ represents a „musical note“ of frequency ω which oscillates within the translated window given by $u\rightarrow g(u-t)$

- Inner product $\langle f|g_{\omega,t}\rangle$ measures the correlation between the signal *f* and the musical note $g_{\omega,t}$

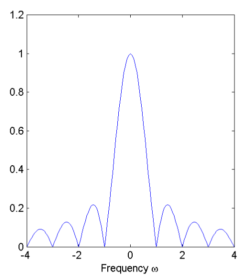
Short Time Fourier Transform

Box window: discontinuities at window boundaries cause artefacts in the frequency domain

g



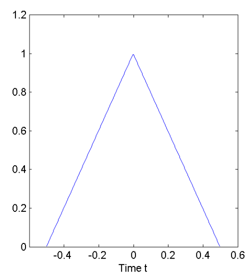
$|\hat{g}|$



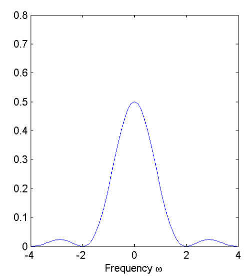
Short Time Fourier Transform

Triangle window

g

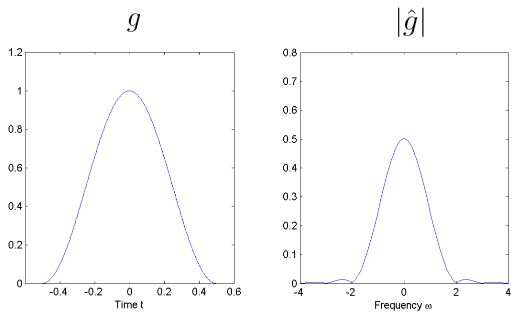


$|\hat{g}|$



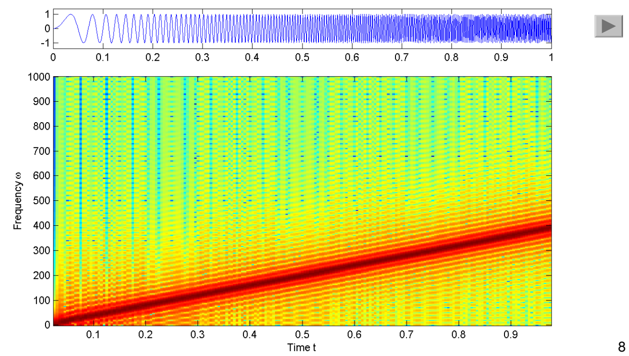
Short Time Fourier Transform

Hann window



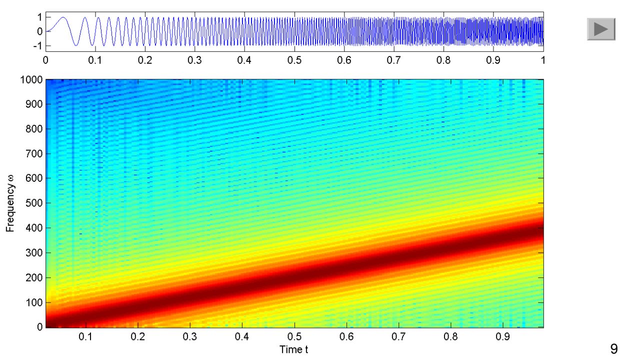
Short Time Fourier Transform

Chirp signal and STFT with **box window** of length 0.05



Short Time Fourier Transform

Chirp signal and STFT with **hann window** of length 0.05



Time-Frequency Localization

- Size of window constitutes a compromise between time resolution and frequency resolution:

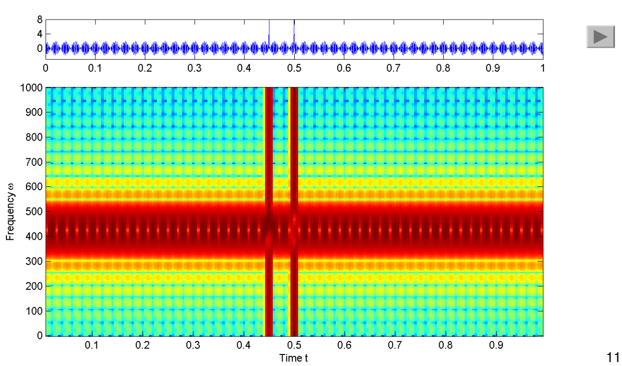
  Large window : poor time resolution
  good frequency resolution

  Small window : good time resolution
  poor frequency resolution

- Heisenberg Uncertainty Principle: there is no window function that localizes in time and frequency with arbitrary position.

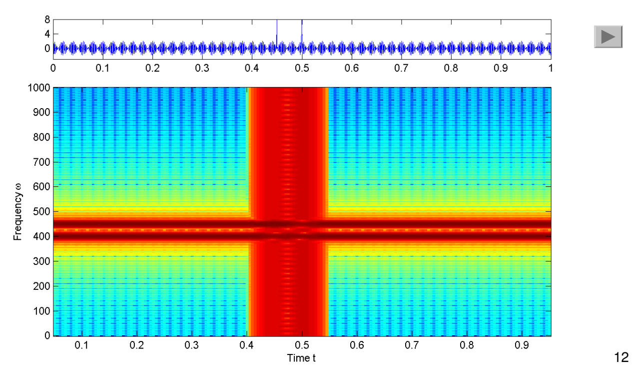
Short Time Fourier Transform

Signal and STFT with hann window of **length 0.02**



Short Time Fourier Transform

Signal and STFT with hann window of **length 0.1**



Heisenberg Uncertainty Principle

Window function $g \in L^2(\mathbb{R})$ with $\|g\| = 1$

Center

$$t_0 = t_0(g) := \int_{-\infty}^{\infty} t|g(t)|^2 dt$$

$$\omega_0 = \omega_0(g) := \int_{-\infty}^{\infty} \omega|\hat{g}(\omega)|^2 d\omega$$

Width

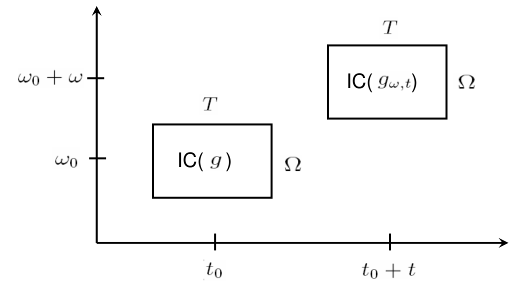
$$T(g) := \left(\int_{-\infty}^{\infty} (t-t_0)^2 |g(t)|^2 dt\right)^{\frac{1}{2}}$$

$$\Omega(g) := \left(\int_{-\infty}^{\infty} (\omega-\omega_0)^2 |\hat{g}(\omega)|^2 d\omega\right)^{\frac{1}{2}}$$

$$\boxed{T(g) \cdot \Omega(g) \geq \frac{1}{4\pi}}$$

Information Cells

$g_{\omega,t}(u) := e^{2\pi i \omega u} g(u-t)$ "musical note"



MATLAB

- MATLAB function SPECTROGRAM
- *N* = window length (in samples)
- *M* = overlap (usually $N/2$)
- Compute DFT_N for every windowed section
- Keep lower $N/2$ Fourier coefficients

→ Sequence of spectral vectors
(for each window a vector of dimension $N/2$)

Example

Let *x* be a DT–Signal $x(n) = f(Tn)$

Sampling rate: $1/T = 22050$ Hz

Window length: $N = 4096$

Overlap: $N/2 = 2048$

Hopsize: window length − overlap

Let

$$\begin{aligned} v_0 &:= (x(0),\ x(1),\ \ldots,\ x(4095)) \\ v_1 &:= (x(2048),\ \ldots,\ x(6143)) \\ v_2 &:= (x(4096),\ \ldots,\ x(8191)) \\ &\vdots \end{aligned}$$

v_m corresponds to window $[m \cdot 2048 : m \cdot 2048 + 4095]$

Example

Time resolution:

$$\frac{\text{hopsize}}{\text{sampling rate}} = \frac{4096 - 2048}{22050} = 0.093 = 93\ ms$$

Frequency resolution:

$$v = v_0\ ,\ \hat{v} := \mathrm{DFT}_N(v)$$

$$\hat{v}(k) \approx \frac{1}{T} \cdot \hat{f}\left(\frac{k}{N} \cdot \frac{1}{T}\right)$$

$$\omega = \frac{k}{N} \cdot \frac{1}{T} = k \cdot \frac{22050}{4096} = k \cdot 5.38\ Hz$$

Pitch Features

Model assumption: Equal – tempered scale

- MIDI pitches: $p \in [1:128]$
- Piano notes: $p = 21\ (\mathrm{A0})$ to $p = 108\ (\mathrm{C8})$
- Concert pitch: $p = 69\ (\mathrm{A4}) = 440$ Hz
- Center frequency: $f_{\mathrm{MIDI}}(p) = 2^{\frac{p-69}{12}} \cdot 440$

→ Logarithmic frequency distribution
Octave: doubling of frequency

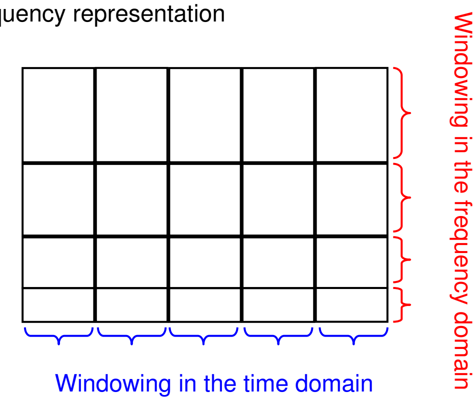
Pitch Features

Idea: Binning of Fourier coefficients

Divide up the fequency axis into logarithmically spaced „pitch regions“ and combine **spectral coefficients** of each region to a single **pitch coefficient**.

Pitch Features

Time-frequency representation



Pitch Features

Note	MIDI pitch	Center [Hz] frequency	Left [Hz] boundary	Right [Hz] boundary	Width [Hz]
A3	57	220.0	213.7	226.4	12.7
A#3	58	233.1	226.4	239.9	13.5
B3	59	246.9	239.9	254.2	14.3
C4	60	261.6	254.2	269.3	15.1
C#4	61	277.2	269.3	285.3	16.0
D4	62	293.7	285.3	302.3	17.0
D#4	63	311.1	302.3	320.2	18.0
E4	64	329.6	320.2	339.3	19.0
F4	65	349.2	339.3	359.5	20.2
F#4	66	370.0	359.5	380.8	21.4
G4	67	392.0	380.8	403.5	22.6
G#4	68	415.3	403.5	427.5	24.0
A4	69	440.0	427.5	452.9	25.4

Pitch Features

Details:

- Let $\hat{v}$ be a spectral vector obtained from a spectrogram w.r.t. a sampling rate $1/T$ and a window length *N*. The spectral coefficient $\hat{v}(k)$ corresponds to the frequency

$$f_{\mathrm{coeff}}(k) := \frac{k}{N} \cdot \frac{1}{T}$$

- Let
$S(p) := \{k : f_{\mathrm{MIDI}}(p - 0.5) \leq f_{\mathrm{coeff}}(k) < f_{\mathrm{MIDI}}(p + 0.5)\}$
be the set of coefficients assigned to a pitch $p \in [1:128]$
Then the pitch coefficient $P(p)$ is defined as

$$P(p) := \sum_{k \in S(p)} |\hat{v}(k)|^2$$

Pitch Features

Example: A4, *p* = 69

- Center frequency: $f(p = 69) = 2^{\frac{0}{12}} \cdot 440 = 440\ Hz$
- Lower bound: $f(p = 68.5) = 2^{\frac{-0.5}{12}} \cdot 440 = 427.5\ Hz$
- Upper bound: $f(p = 69.5) = 2^{\frac{0.5}{12}} \cdot 440 = 452.9\ Hz$
- STFT with $N = 4096$, $1/T = 22050$

$$\begin{array}{rcl} & \vdots & \\ f(k=79) & = & 425.3\ Hz \\ f(k=80) & = & 430.7\ Hz \\ f(k=81) & = & 436.0\ Hz \\ f(k=82) & = & 441.4\ Hz \\ f(k=83) & = & 446.8\ Hz \\ f(k=84) & = & 452.2\ Hz \\ f(k=85) & = & 457.6\ Hz \\ & \vdots & \end{array}$$

Pitch Features

Example: A4, *p* = 69

- Center frequency: $f(p = 69) = 2^{\frac{0}{12}} \cdot 440 = 440\ Hz$
- Lower bound: $f(p = 68.5) = 2^{\frac{-0.5}{12}} \cdot 440 = 427.5\ Hz$
- Upper bound: $f(p = 69.5) = 2^{\frac{0.5}{12}} \cdot 440 = 452.9\ Hz$
- STFT with $N = 4096$, $1/T = 22050$

$$\begin{array}{rcl} & \vdots & \\ f(k=79) & = & 425.3\ Hz \\ f(k=80) & = & 430.7\ Hz \\ f(k=81) & = & 436.0\ Hz \\ f(k=82) & = & 441.4\ Hz \\ f(k=83) & = & 446.8\ Hz \\ f(k=84) & = & 452.2\ Hz \\ f(k=85) & = & 457.6\ Hz \\ & \vdots & \end{array}$$

S(*p* = 69)

$$P(p = 69) = \sum_{k=80}^{84} |\hat{v}(k)|^2$$

Pitch Features

Note:

- $P \in \mathbb{R}^{128}$
- For some pitches, *S(p)* may be empty. This particularly holds for low notes corresponding to narrow frequency bands.

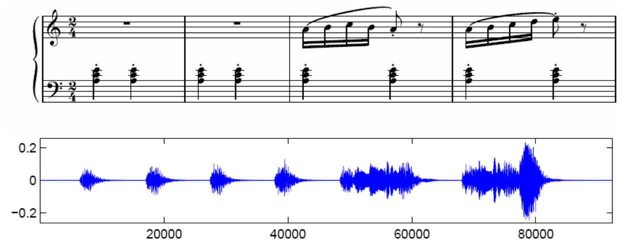
→ Linear frequency sampling is problematic!

Solution:

Multi-resolution spectrograms or multirate filterbanks

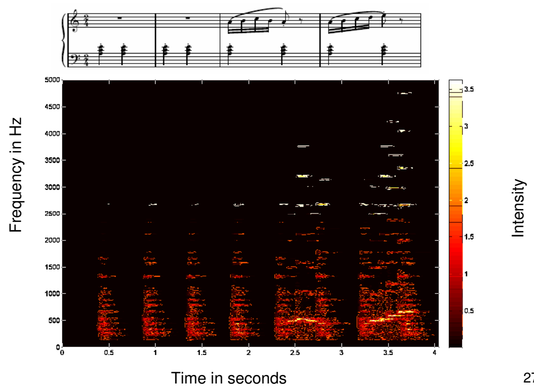
Audio Representation

Example: Op. 100, No. 2 by Friedrich Burgmüller



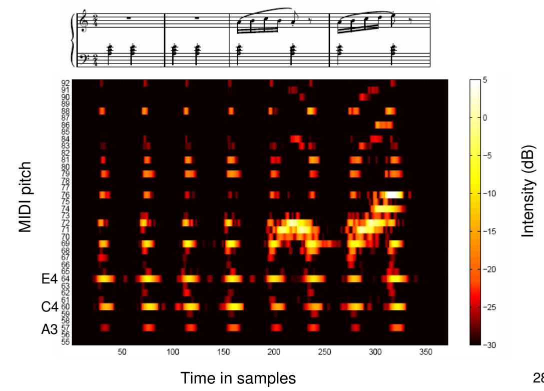
Short Time Fourier Transform

Example: Op. 100, No. 2 by Friedrich Burgmüller



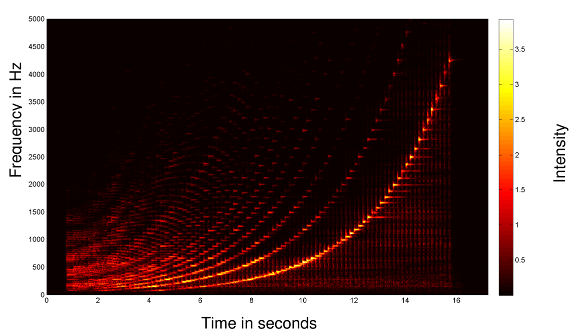
Pitch Features

Example: Op. 100, No. 2 by Friedrich Burgmüller



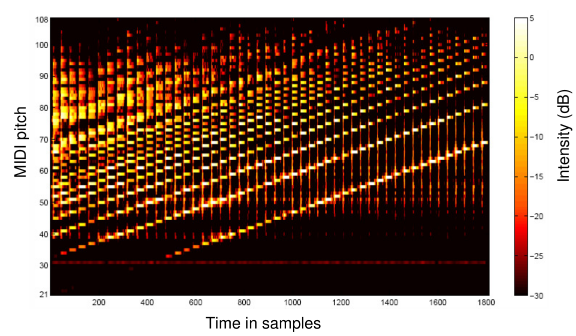
Short Time Fourier Transform

Example: Chromatic Scale



Pitch Features

Example: Chromatic Scale



Chroma Features

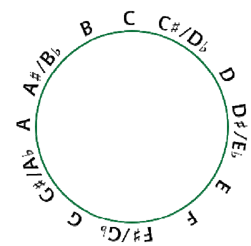
- Human perception of pitch is periodic in the sense that two pitches are perceived as similar in „color“ if they differ by an octave.
- Seperate pitch into two components: tone height (octave number) and chroma.
- Chroma : 12 traditional pitch classes of the equal-tempered scale. For example

  Chroma C $\widehat{=} \{\ldots\ ,\ \mathrm{C0}\ ,\ \mathrm{C1}\ ,\ \mathrm{C2}\ ,\ \mathrm{C3}\ ,\ \ldots\}$
- Computation: pitch features → chroma features

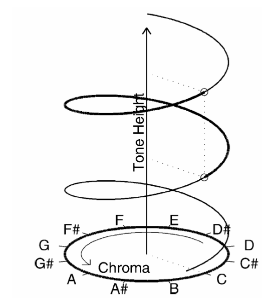
  Add up all pitches belonging to the same class
- Result: 12-dimensional chroma vector.

Chroma Features

Chromatic circle



Shepard‘s helix of pitch perception



http://en.wikipedia.org/wiki/Pitch_class_space

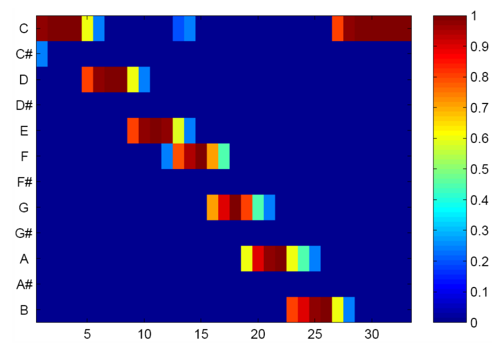
Bartsch/Wakefield, IEEE Trans. Multimedia, 2005

Chroma Features

- Sequence of chroma vectors correlates to the harmonic progression
- Normalization $v \rightarrow \frac{v}{\|v\|}$ makes features invariant to changes in dynamics
- Further quantization and smoothing: CENS features
- Taking logarithm before adding up pitch coefficients accounts for logarithmic sensation of intensity

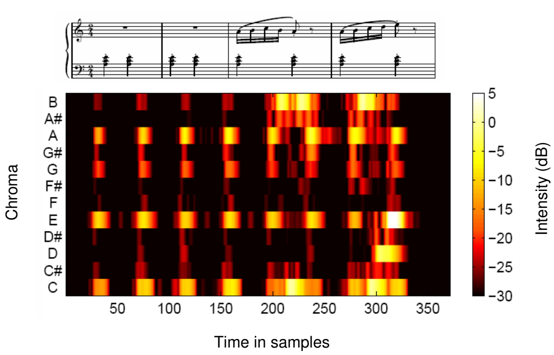
Chroma Features

Example: C-Major Scale



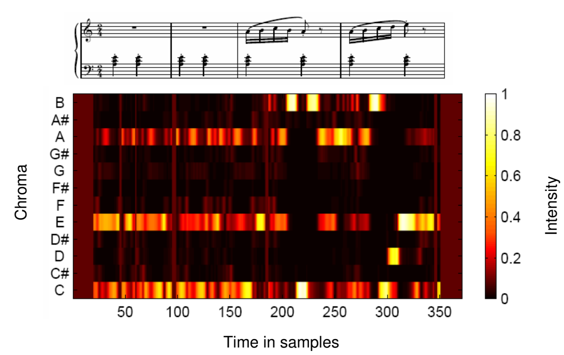
Chroma Features

Example: Burgmüller Op. 100, No. 2



Chroma Features

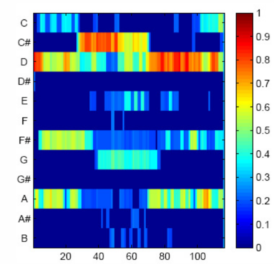
Normalization



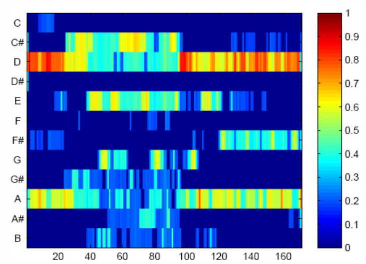
Chroma Features

Example: Bach Toccata

Koopman



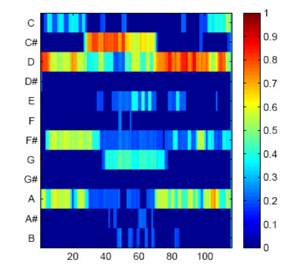
Ruebsam



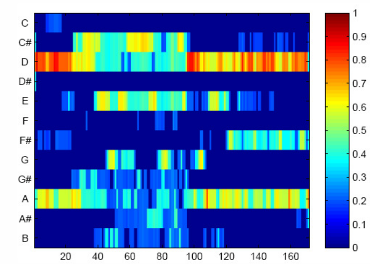
Chroma Features

Example: Bach Toccata

Koopman



Ruebsam

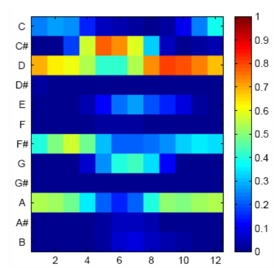


Feature resolution: 10 Hz

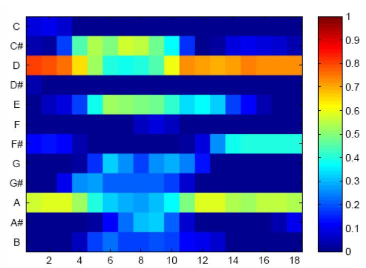
Chroma Features

Example: Bach Toccata

Koopman



Ruebsam

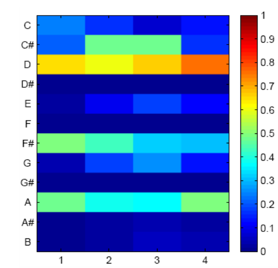


Feature resolution: 1 Hz

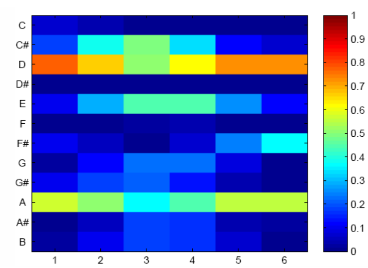
Chroma Features

Example: Bach Toccata

Koopman



Ruebsam



Feature resolution: 0.33 Hz

Chroma Features

	WAV	Chroma (10 Hz)	CENS (1 Hz)
???			
???			
???			

Chroma Features

	WAV	Chroma (10 Hz)	CENS (1 Hz)
Beethoven's Fifth (Bernstein)			
???			
???			

Chroma Features

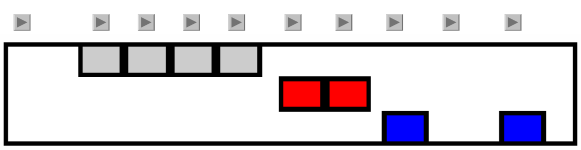
	WAV	Chroma (10 Hz)	CENS (1 Hz)
Beethoven's Fifth (Bernstein)			
Beethoven's Fifth (Piano/Sherbakov)			
???			

Chroma Features

	WAV	Chroma (10 Hz)	CENS (1 Hz)
Beethoven's Fifth (Bernstein)			
Beethoven's Fifth (Piano/Sherbakov)			
Brahms Hungarian Dance No. 5			

Chroma Features

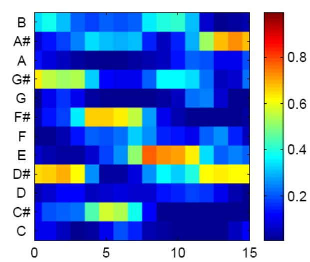
Example: Zager & Evans "In The Year 2525"



How to deal with transpositions?

Chroma Features

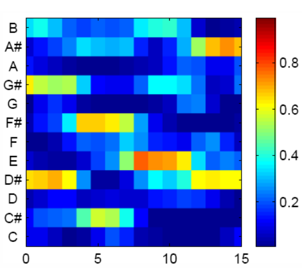
Example: Zager & Evans "In The Year 2525"



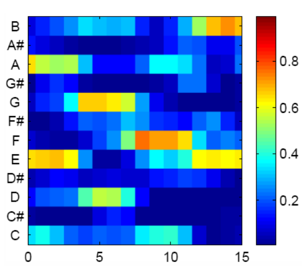
Original: $(v^1, \ldots, v^N)$

Chroma Features

Example: Zager & Evans "In The Year 2525"



Original: $(v^1, \ldots, v^N)$



Shifted: $(\sigma(v^1), \ldots, \sigma(v^N))$