Lecture 11: Chroma and Chords

1. Features for Music Audio
2. Chroma Features
3. Chord Recognition

Dan Ellis
Dept. Electrical Engineering, Columbia University
dpwe@ee.columbia.edu   http://www.ee.columbia.edu/~dpwe/e4896/
I. Features for Music Audio

- **Challenges of large music databases**
  - how to find “what we want”...

- **Euclidean metaphor**
  - music tracks as points in space

- **What are the dimensions?**
  - “sound” - timbre, instruments → MFCC
  - melody, chords → Chroma
  - rhythm, tempo → Rhythmic bases
• The standard feature for **speech recognition**

![Diagram of MFCCs process](image)

- **Sound** → **FFT X[k]** → **Mel scale freq. warp** → **log |X[k]|** → **IFFT** → **Truncate** → **MFCCs**
**MFCC Example**

- **Resynthesize by imposing spectrum on noise**
- MFCCs capture *instruments*, not notes

Let It Be - log-freq specgram (LIB-1)

![Log-freq specgram](image)

MFCCs

![MFCCs](image)

Noise excited MFCC resynthesis (LIB-2)

![Noise excited MFCC resynthesis](image)
MFCC Artist Classification

- 20 Artists x 6 albums each
  - train models on 5 albums, classify tracks from last

- Model as MFCC mean + covariance per artist
  - “single Gaussian” model
  - 20 (mean) + 10 x 19 (covariance) parameters
  - 55% correct (guessing ~5%)

Confusion: MFCCs (acc 55.13%)
2. Chroma Features

- What about modeling **tonal content** (notes)?
  - melody spotting
  - chord recognition
  - cover songs...

- **MFCCs** exclude tonal content

- **Polyphonic transcription** is too hard
  - e.g. sinusoidal tracking: confused by harmonics

- **Chroma features** as solution...
**Chroma Features**

- Idea: Project all energy onto 12 semitones regardless of octave
  - maintains main “musical” distinction
  - invariant to musical equivalence
  - no need to worry about harmonics?

\[
C(b) = \sum_{k=0}^{N_M} B(12 \log_2(k/k_0) - b)W(k)|X[k]| 
\]

- \(W(k)\) is weighting, \(B(b)\) selects every \(\sim \mod 12\)
Better Chroma

- **Problems:**
  - **blurring** of bins close to edges
  - limitation of FFT bin **resolution**

- **Solutions:**
  - peak picking - only keep energy at center of peaks
  - **Instantaneous Frequency** - high-resolution estimates
  - adapt **tuning** center based on histogram of pitches
Chroma Resynthesis

• Chroma describes the notes in an octave
  ○ ... but not the octave

• Can resynthesize by presenting all octaves
  ○ ... with a smooth envelope
  ○ “Shepard tones” - octave is ambiguous

\[ y_b(t) = \sum_{o=1}^{M} W(o + \frac{b}{12}) \cos 2^{o+\frac{b}{12}} w_0 t \]

○ endless sequence illusion
Chroma Example

- Simple Shepard tone resynthesis
  - can also reimpose broad spectrum from MFCCs

Let It Be - log-freq specgram (LIB-1)

Chroma features

Shepard tone resynthesis of chroma (LIB-3)

MFCC-filtered shepard tones (LIB-4)
Beat-Synchronous Chroma

Bartsch & Wakefield 2001

- Drastically reduce data size by recording one chroma frame per beat
3. Chord Recognition

- Beat synchronous chroma look like chords

- Can we transcribe them?

- **Two approaches**
  - Manual templates (prior knowledge)
  - Learned models (from training data)
Chord Recognition System

- Analogous to speech recognition
  - Gaussian models of features for each chord
  - Hidden Markov Models for chord transitions
HMMs

- Hidden Markov Models are good for inferring hidden states
  - underlying Markov “generative model”
  - each state has emission distribution

- observations tell us something about state...

- infer smoothed state sequence
HMM Inference

- HMM defines emission distribution $p(x|q)$ and transition probabilities $p(q_n|q_{n-1})$
- Likelihood of observed given state sequence:
  $$p(\{x_n\}|\{q_n\}) = \prod_{n} p(x_n|q_n)p(q_n|q_{n-1})$$

- By dynamic programming, we can also identify the best state sequence given just the observations
Key Normalization

- Chord transitions depend on **key** of piece
  - dominant, relative minor, etc...

- Chord transition probabilities should be **key-relative**
  - estimate main key of piece
  - rotate all chroma features
  - learn models
Chord Recognition

- Often works:

- But only about 60% of the time

<table>
<thead>
<tr>
<th></th>
<th>12 chroma</th>
<th>+bass</th>
</tr>
</thead>
<tbody>
<tr>
<td>indep. models</td>
<td>0.539</td>
<td>0.552</td>
</tr>
<tr>
<td>pooled models</td>
<td>0.556</td>
<td>0.578</td>
</tr>
</tbody>
</table>
Summary

• **Music Audio Features**
  capture information useful for classification

• **Chroma Features**
  12 bins to robustly summarize notes

• **Chord Recognition**
  Sometimes easy, sometimes subtle
References


