Beat-Synchronous Chroma Representations for Music Analysis

Dan Ellis

Laboratory for Recognition and Organization of Speech and Audio
Dept. Electrical Eng., Columbia Univ., NY USA

dpwe@ee.columbia.edu          http://labrosa.ee.columbia.edu/

1. Chroma Features
2. Beat Tracking
3. Matching Cover Songs
4. Artist Identification
Beyond MFCCs...

- **MFCCs** have been useful in Audio Music IR
  - “timbral similarity”
  - artist ID, segmentation, thumbnailing, singing ...
- **Separate tradition of Symbolic MIR**
  - melody matching, chord detection, meter analysis
- **It’s time to bring them together**
  - ... with robust audio mid-level representations
  - ... that capture **tonal** (melodic-harmonic) content

= beat-synchronous chroma features
1. Chroma Features

- Chroma features map spectral energy into one canonical octave
  - i.e. 12 semitone bins

- Can resynthesize as “Shepard Tones”
  - all octaves at once
Calculating Chroma Features

- **Method 1**: Map every STFT bin
  - Blurs non-tonal energy

- **Method 2**: Map only STFT peaks
  - Still blurry at low frequencies

- **Method 3**: Instantaneous Frequency $\frac{\delta \theta}{\delta t}$
  - Escapes frequency resolution limit
2. Beat Tracking (1)

• Goal: One feature vector per ‘beat’ (tatum)
  ○ for tempo normalization, efficiency

• “Onset Strength Envelope”
  ○ \( \text{sum}\left(\max(0, \text{diff}_t(\log |X(t, f)|))\right) \)

• Autocorr. + window \( \rightarrow \) global tempo estimate

168.5 BPM
Beat Tracking (2)

• Dynamic Programming finds beat times \{t_i\}
  - optimizes \( \sum_i O(t_i) + \alpha \sum_i W((t_{i+1} - t_i - \tau_p)/\beta) \)
  - where \( O(t) \) is onset strength envelope (local score)
    \( W(t) \) is a log-Gaussian window (transition cost)
  - \( \tau_p \) is the default beat period per measured tempo
  - incrementally find best predecessor at every time
  - backtrace from largest final score to get beats

\[
C^*(t) = \gamma O(t) + (1-\gamma)\max_\tau \{ W((\tau - \tau_p)/\beta) C^*(\tau) \}
\]

\[
P(t) = \arg\max_\tau \{ W((\tau - \tau_p)/\beta) C^*(\tau) \}
\]
Beat Tracking Results

• DP will **bridge gaps** (non-causal)
  - there is always a best path ...

• 2nd place in MIREX 2006 Beat Tracking
  - compared to McKinney & Moelants human data
**Beat-Synchronous Chroma Features**

- **Beat + chroma features / 30ms frames**
  - → *average chroma* within each beat
  - compact; sufficient?
3. Cover Song Detection

- “Cover Songs” = reinterpretation of a piece
  - different instrumentation, character
  - no match with “timbral” features

Let It Be - The Beatles

Let It Be - Nick Cave

Need a different representation!
- beat-synchronous chroma features
Matching (1): Little Fragments

- Cover versions may change song structure
  - multiple local matches at different alignments
- Match query and target as many small pieces?
  - how big are the pieces?
  - how do we combine individual scores?
  - do we have all day?
Matching (2): Global Correlation

- Cross-correlate *entire* beat-chroma matrices
  - ... at all possible transpositions
  - implicit combination of match quality and duration

- One good matching fragment is sufficient...?
Filtered Cross-Correlation

- Raw correlation not as important as precise local match
  - looking for large contrast at ±1 beat skew
  - i.e. high-pass filter
• 23 pairs of cover songs from uspop2002 +...
  ○ one correct match per query
Results (2): MIREX 06

- Cover song contest
  - 30 songs x 11 versions of each (!)
  - (data has not been disclosed)
  - # true covers in top 10
  - 8 systems compared
    (4 cover song + 4 similarity)
- Found 761/3300
  = 23% recall
  - next best: 11%
  - guess: 3%
Where are the matches?

- **Look inside** global cross-correlation to find matching fragments...

\[ x_{\text{corr}} = \sum_t \sum_f (C_1(t,f) \cdot C_2(t,f)) \] - view along time
What are the mistakes?

- **False reject** - missed true match
  - cover version is too different, beat tracking wrong ...

- **False alarm** - invalid match
  - “Cocaine” (Clapton) vs. “Satisfaction” (Stones)
4. Artist Identification (AID)

- Baseline system: “Bag of (timbral) frames”
  - MFCC frames, model as Gaussian or GMM
  - Distance by likelihood or KL

- Dataset: [Mandel et al. 2006]
  - 18 artists \times 5 or 6 albums each
  - 18x3 albums for training, 18x2 for test, 10x1 dev
Beat Chroma Features for AID?

• Artists may use tonality in particular ways...
  ○ density, variety
  ○ particular chords
  ○ (influence of instruments on chroma features)

• Try bag-of-frames on beat-chroma rep’n
  ○ use several consecutive beats?
  ○ key-normalization of each piece?
Key Normalization

- Could try matching at all possible rotations..
- .. or just **transpose** every piece initially
  - single Gaussian model of one piece
  - find ML rotation of other pieces
  - model all transposed pieces
  - iterate until convergence

---

Beat-Chroma Representations - Ellis

2007-05-23 - 19/23
• Preliminary Mandel18 Artist ID accuracy:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Model</th>
<th>T win</th>
<th>Acc</th>
<th>Exec. time</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFCC20</td>
<td>FullCov</td>
<td>1</td>
<td>48%</td>
<td>212 s</td>
</tr>
<tr>
<td>MFCC20</td>
<td>64 GMM</td>
<td>1</td>
<td>33%</td>
<td>1952 s</td>
</tr>
<tr>
<td>Chroma</td>
<td>FullCov</td>
<td>1</td>
<td>15%</td>
<td>46 s</td>
</tr>
<tr>
<td>Chroma</td>
<td>FullCov</td>
<td>4</td>
<td>14%</td>
<td>117 s</td>
</tr>
<tr>
<td>Chroma</td>
<td>64GMM</td>
<td>1</td>
<td>24%</td>
<td>850 s</td>
</tr>
<tr>
<td>Chroma</td>
<td>64GMM</td>
<td>4</td>
<td>15%</td>
<td>2242 s</td>
</tr>
<tr>
<td>ChromaKN</td>
<td>FullCov</td>
<td>1</td>
<td>17%</td>
<td>110 s</td>
</tr>
<tr>
<td>ChromaKN</td>
<td>FullCov</td>
<td>4</td>
<td>14%</td>
<td>258 s</td>
</tr>
<tr>
<td>ChromaKN</td>
<td>64GMM</td>
<td>1</td>
<td>25%</td>
<td>2533 s</td>
</tr>
<tr>
<td>ChromaKN</td>
<td>64GMM</td>
<td>4</td>
<td>16%</td>
<td>5803 s</td>
</tr>
<tr>
<td>MFCC + Chroma fusion</td>
<td></td>
<td></td>
<td>52%</td>
<td></td>
</tr>
</tbody>
</table>
**Artist Fragments**

- Idea: Find the most **discriminант** beat-chroma fragments per artist
  - k-means cluster 16 beat fragments within piece
    - Chroma Features - The Beatles, "All You Need is Love"
  - keep fragments largest ratio
    (avg. similarity to **same artist**) / (avg. sim. to **others**)
  - classify test pieces by ID of best-scoring fragment
Artist Fragment Results

- Preliminary, 5 way artist ID, ~32% correct

- need to search more fragments
- way to choose phrase beginnings?
- a basis set for all tonal content?
Conclusions and Future Work

• **Beat-synchronous chroma features** are successful for matching cover songs
  ○ captures **melody-harmony**, not instruments

• **Further uses:**
  Beat-chroma fragments as **musical building blocks**
  ○ e.g. VQ over large body of music
  ○ find recurrent **motifs**
  ○ artist identification?

• **Code available!**
  Google “matlab chroma features”