Cover Song ID with Beat-Synchronous Chroma Features

Dan Ellis • Columbia University • dpwe@ee.columbia.edu

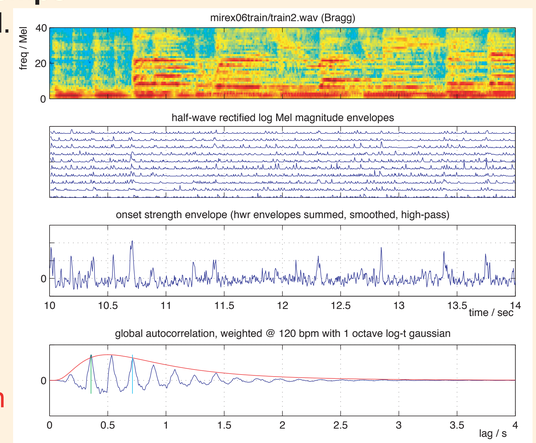
Columbia University

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

Summary: **Beat-synchronous chroma features** capture melodic-harmonic content of music audio, and successfully **detect cover versions** of songs. We describe our system, including **beat tracking by dynamic-programming.**

Tempo Estimation by Onset Autocorrelation

- Cover song system requires **beat tracking** (see bottom right panel)
  - our beat tracker first estimates **tempo**
  - did ~OK in MIREX06 tempo eval.
- Music Audio is converted to an **Onset Strength** envelope by:
  - Mel-freq. log-mag. spectrogram
  - Half-wave rectify (negative values → zero)
  - Sum across all channels
  - Highpass at ~ 0.5 Hz
  - Smooth at ~ 20 Hz
- **Autocorrelate** out to lag of ~ 4 s
- **Weight** autocorrelation by Gaussian on log-time axis (centered on 120 bpm for human tapping tempo, or 240 bpm for chroma features)
- Pick highest peak as main period & 2'dry peak from [0.33, 0.5, 2, 3] × main

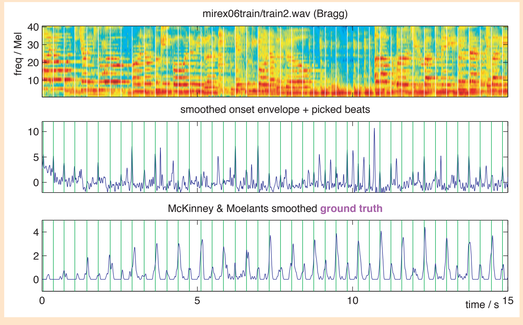


Beat Tracking by Dynamic Programming

- Beat tracker takes **global tempo** period t_B (from above) as input
- **Beat tracking score** is calculated recursively for every time sample as:

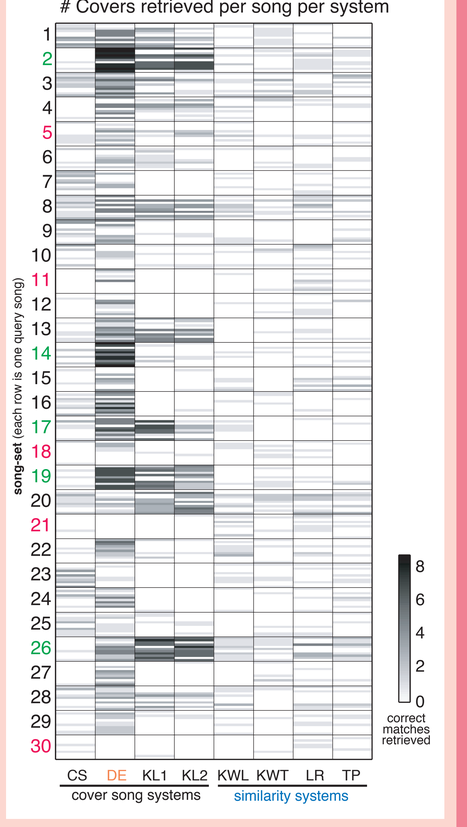
$$S_B(t) = (1-\alpha)O(t) + \alpha \max_{\tau_p} W(t-\tau_p)\, S_B(\tau_p)$$

  - $O(t)$ is onset strength envelope from above (local match)
  - $W(t)$ is a **log-time Gaussian window** between $t_B/2$ and $2t_B$ centered on t_B (width determines tempo tolerance)
  - best **predecessor** τ_p is recorded for each time t
  - largest S_B near end of track is **traced back** for entire beat time sequence
- Beat tracker came **2nd of 5** systems compared to **human ground truth** in MIREX06 beat tracking evaluation



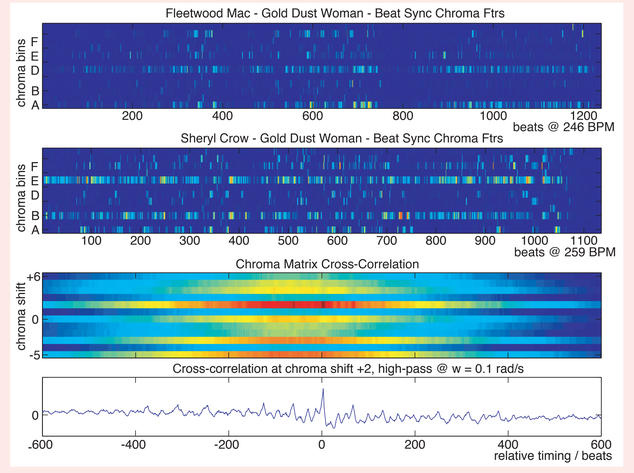
MIREX-06 Cover Song Identification: Evaluation Results & Analysis

- **Task:** Identify Cover Songs (alternate versions of the same musical piece) within a database of music audio recordings
- **Data:** 30 songs with 11 different versions of each (330 songs), from mixed genres ("classical, jazz, gospel, rock, folk-rock, etc.")
- **Metric:** How many of the 10 other versions of each query song are among the 10 most similar found by each algorithm? (3300 max. total) (also used Mean Reciprocal Rank, not reported here)
- **Results:** 4 cover song systems and 4 music similarity systems tested
  - Our system (DE) is **best** with 761 covers found (23% recall)
  - Music similarity systems score a little above random (≈ 100 covers)
  - → Cover song task is very different from current similarity approaches
- **Analysis:** Many song-sets difficult for all systems (e.g. 5, 11, 18, 21, 30)
  - most hits from a few song sets (2, 14, 19 for DE; 2, 17, 26 for KL)
  - different systems have different strengths



Song Matching by Cross-Correlating Beat-Chroma Matrices

- **Cover versions** may differ in tempo, instrumentation, style
  - devise features to **normalize** these variations
- **Chroma** features map whole spectrum to one octave
  - preserve essence of melody/harmony without detail
  - high-resolution **instantaneous frequency** finds peaks
- **Beat-synchronous** features factor out tempo variations
  - ... provided beat tracker finds matching metrical levels
  - average chroma bin energy over whole beat
- **Cross-correlation** of entire beats × semitones matrix has **sharp maxima** when large subsections match
  - repeat at all 12 chroma rotations
  - **high-pass** filter to emhasize peaks



Download Matlab code from: http://labrosa.ee.columbia.edu/projects/coversongs/