The 2007 LabROSA cover song detection system

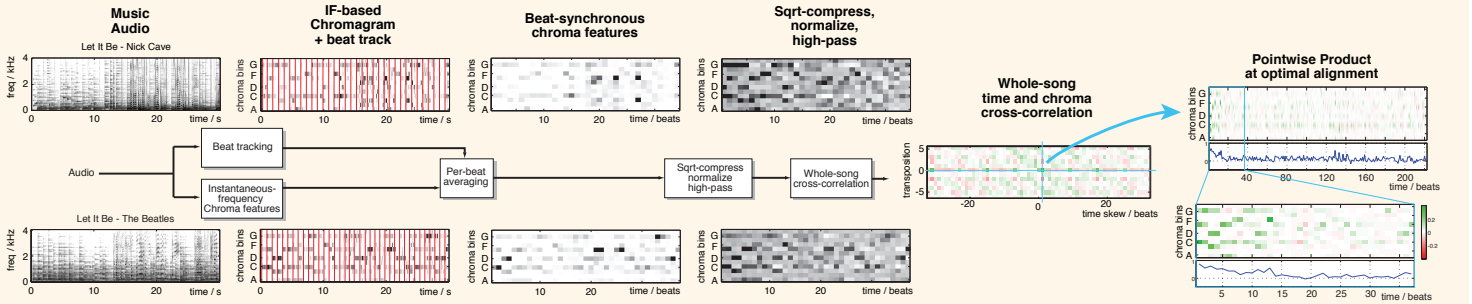
Dan Ellis & Courtenay Cotton • {dpwe,cvcotton}@ee.columbia.edu

Columbia University

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

Summary: A **beat-synchronous chroma** representation enables the matching of cover versions of popular music using **global cross-correlation** across time- and transposition-skew.

System Overview



- **Cover Songs** are alternative performances that frequently change genre, tempo, instrumentation, etc.
- We strive for a representation of music audio largely invariant to such changes by:
  - **beat-tracking** to the dominant tempo
  - storing 12-dimensional **chroma** feature vectors, one per beat
- This beat-chroma representation is compared between songs by **cross-correlating** entire songs (after normalization). All 12 possible transpositions are tested.
- Normalization involves amplitude compression and **high-pass filtering** along time to emphasize changes and downweight sustained pitches.
- **Whole-song correlation** can find only one best relative alignment, but we often find that even a small region of alignment leads to a clear peak value. It is implemented efficiently as a single FFT.
- Examining the pointwise product of the beat-chroma matrices at their optimal alignment reveals which **chroma and time bins** are most responsible for the correlation score (green blobs in bottom left figure).

Changes for 2007

- We modified our 2006 Cover Song system [1] as follows:
- Previously, we normalized each cross-correlation by the length of the shorter song to bound the maximum result. **Removing normalization** improved results.
- **High-pass filtering** to emphasize locally-optimal skews used to be performed after cross-correlation; we now high-pass the beat-chroma before correlation, and optimized the filter.
- Our beat tracker includes a **tempo bias** towards a particular range. Before, we aimed for around 240 BPM to get a denser representation. Reducing to 120 BPM improved performance.
- Sometimes original and cover are beat-tracked at different metrical levels. **Beat tracking twice**, at 120 and 240 BPM, then choosing the best out of the 2x2 comparisons, fixed most of these.

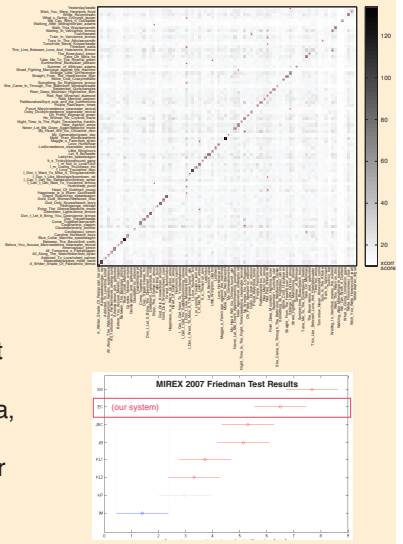
Data - "covers80" dataset & MIREX 2007

- **"covers80"** is our development set of 80 pairs of pop-music cover songs drawn from uspop2002 and other sources.
- MFCCs, Chroma features, etc. are available for **download** at `http://labrosa.ee.columbia.edu/projects/coversongs/covers80/`
- MIREX Cover Song evaluation has 330 covers among 1000 tracks.

Results

"covers80" results	Correct
MIREX 06 baseline	34/80 = 42.5%
Without cross-corr norm	41/80 = 51.3%
Improved high-pass filtering	46/80 = 57.5%
Tempo bias = 120 BPM	48/80 = 60.0%
Dual tempo levels	54/80 = 67.5%

- The modifications led to an overall **59% relative improvement** over the 2006 on both covers80 accuracy and MIREX top-10 count.
- Examination of the full confusion matrix reveals certain 'preferred' **decoy matches** (I Don't Like Mondays/Tori Amos) – to be investigated.
- Our system **ranked 2nd** of 8 in the 2007 MIREX evaluation, with a Mean AP of 0.330, compared to 0.521 for the best system (Serrá & Gómez). Their system also uses chroma, but aligns tracks by Dynamic Time Warping. We believe our correlation approach allows much **faster** comparisons.



[1] D. Ellis and G. Poliner, "Identifying Cover Songs With Chroma Features and Dynamic Programming Beat Tracking, Proc. ICASSP-07 Hawai'i, pp. IV-1429-1432.

MATLAB code to run this system is available at:

http://labrosa.ee.columbia.edu/projects/coversongs/