Clustering Beat-Chroma Patterns in a Large Music Database

Thierry Bertin-Mahieux †, Ron J. Weiss ‡ and Daniel P. W. Ellis †

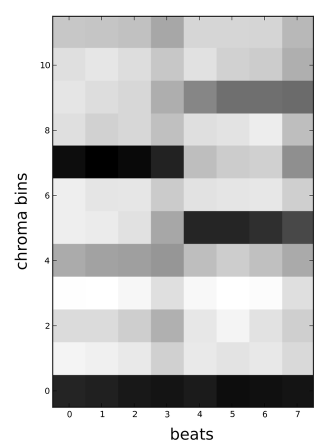
† LabROSA, Dept. of Electrical Engineering, Columbia University

‡ MARL, New York University

{thierry,ronw,dpwe}@ee.columbia.edu

Introduction

- Availability of very large collections of music audio: can we infer anything about the underlying structure and common features of e.g. commercial pop music?
- Our interest: tonal content of the music – i.e. the harmony and melody.
- Beat-synchronous chromagrams: rich enough to generate musically-relevant results, simplified enough to abstract away instrumentation and other stylistic details.



- Goal: identify meaningful information about the musical structure represented in the entire database by examining individual entries in this codebook.
- Method: identify common patterns in beat-synchronous chromagrams by learning codebooks from a large set of examples. Individual codewords consist of short beat-chroma patches of between 1 and 8 beats, optionally aligned to bar boundaries.
- Prior work: "shingles" of [1], beat-synchronous analysis to indeitfy the chorus by [2], and cover recognition by [3].

Audio Features - Echo Nest

Chromagrams from the Echo Nest online API.

- Feature analysis based on Echo Nest analyze API [4].
- For any song, EchoNest provides a chroma vector (length 12) for every music event (called "segment"), and a segmentation of the song into beats and bars. Beats may span or subdivide segments; bars span multiple beats.
- Averaging per-segment chroma over beat times results in a beat-synchronous chroma feature representation.
- Dataset size: $43,000$ songs.

Beat-Chroma Patches.

- We use Echo Nest analysis to break a song into a collection of beat-chroma "patches", typically one or two bars in length.
- 82% of the bars in our data were 4 beats long.
- Other cases: we resample patches to a fixed length of 4 beats.
- We rotate patches so that the first row contains the most energy. Each patch is normalized independently.

Vector Quantization

- Vector Quantization algorithm [5] to cluster beat-chroma patches
- VQ can be seen as online k-means
- VQ initialized with k random patches from the data
- VQ, although not optimal, scales linearly with the number of patches seen

ℓ learning rate
$\{P_n\}$ set of patches
$\{C_k\}$ codebook of K codes
for $nIters$ **do**
  for $p \in \{P_n\}$ **do**
    $c \leftarrow \min_{c \in C_k} \text{dist}(p, c)$
    $c \leftarrow c + (p - c) * \ell$
  end for
end for
return $\{C_k\}$

Intuitively:

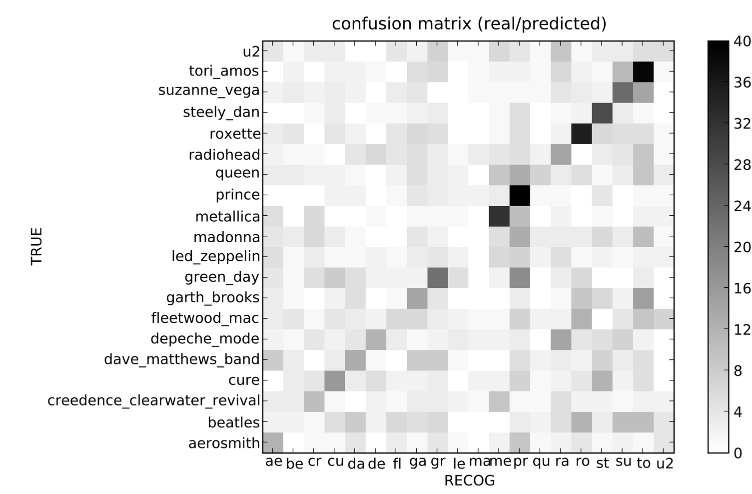
- For each new patch, find the closest codeword.
- Bring that codewords closer to the patch by some learning rate.
- Iterate.

Pattern Analysis

Experiments

We present two applications of the beat-chroma codebooks to illustrate how the "natural" structure identified via unsupervised clustering can provide useful features for subsequent supervised tasks.

Artist recognition task. We use the *artist20* data set: 1402 songs from 20 artists, mostly rock and pop of different subgenres. Previously published results using GMMs on MFCC features achieve an accuracy of 59%, whereas using only chroma as a representation yields an accuracy of 33% [6].
We get an accuracy of **23.4**%, random baseline is around 5%. The confusion matrix is shown here, note that certain artists are recognized at an accuracy far above the average.



Bar alignment task. Since the clustering described is based on the segmentation of the signal in to bars, the codewords should contain information related to bar alignment, such as the presence of a strong beat on the first beat.

Offset	% of times chosen
0	**62.6**
1	16.5
2	9.4
3	11.5

- A bullet point
- A bullet point
- A bullet point
- A bullet point

Conclusion

- A bullet point
- A bullet point

References

[1] M. Casey and M. Slaney, "Fast recognition of remixed music audio," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007.

[2] M. A. Bartsch and G. H. Wakefield, "To catch a chorus: using chroma-based representations for audio thumbnailing," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, New York, October 2001.

[3] D. Ellis and G. Poliner, "Identifying cover songs with chroma features and dynamic programming beat tracking," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007.

[4] The Echo Nest Analyze, API, http://developer.echonest.com.

[5] A. Gersho and R. Gray, *Vector quantization and signal compression.* Kluwer Academic Publishers, 1991.

[6] D. Ellis, "Classifying music audio with timbral and chroma features," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, 2007.

Poster presented at ISMIR 2010 (Utrecht) for the paper: Clustering Beat-Chroma Patterns in a Large Music Database by T. Bertin-Mahieux, R. J. Weiss and D. P. W. Ellis. Contact: tb2332@columbia.edu, code: http://www.columbia.edu/~tb2332/.