The Harmonix Set

Beats, downbeats, and structural annotations for Western pop music

Sevag Hanssian

MUMT 621, Winter 2021

January 26, 2021

SOUND CHECK

The Harmonix Set

Annotations of beats, downbeats, and functional segmentation for over 900 full tracks that covers a wide range of western popular music, to foster research that focuses on multiple retrieval tasks at once.[1],[2]

1. MIREX task "Audio Beat Tracking"[3]
2. MIREX task "Audio Downbeat Estimation"[4]
3. MIREX task "Structural Segmentation"[5]

---

[1] Oriol Nieto et al. "The Harmonix Set: Beats, Downbeats, and Functional Segment Annotations of Western Popular Music". In: *Proc. of the 20th International Society for Music Information Retrieval Conference (ISMIR), Delft, The Netherlands, 2019* (2019).

[2] *urinieto/harmonixset*. URL: https://github.com/urinieto/harmonixset.

[3] *2006:Audio Beat Tracking*. URL: https://www.music-ir.org/mirex/wiki/2006:Audio_Beat_Tracking.

[4] *2014:Audio Downbeat Estimation*. URL: https://www.music-ir.org/mirex/wiki/2014:Audio_Downbeat_Estimation.

[5] *2009:Structual Segmentation*. URL: https://www.music-ir.org/mirex/wiki/2009:Structural_Segmentation.

MIREX beat tracking datasets

Summary of MIREX beat tracking datasets[6]

2006 First appearance of challenge; *the MCK dataset contains 160 30-second audio excerpts created by the MIREX team in 2006. Characterized by stable tempo, wide variety of instrumentations and musical styles. 20% of the files have non-binary meters.*

2009 Second dataset, Chopin Mazurkas; *the MAZ dataset contains piano recordings of 322 Chopin Mazurkas, which include tempo changes.*

2012 Third dataset; *consists of 217 excerpts around 40s each, majority is difficult to track (e.g. changes in meter and tempo, bad sound quality, expressive timing). It includes romantic music, film soundtracks, blues, chanson, and solo guitar*

[6] E. Krebs and S. Böck. *MIREX 2012 AUDIO BEAT TRACKING EVALUATION : NEUROBEAT*. 2012.

MIREX downbeat estimation datasets

2014 Six different datasets from diverse geographic and stylistic sources:
The Beatles[7], **HJDB**[8] (Hardcore, Jungle, Drum and Bass)
Turkish[9], **Ballroom**[10], **Carnatic**[11], **Cretan**[12]

---

[7] M. Mauch et al. "OMRAS2 metadata project 2009". In: *In Late-breaking session at the 10th International Conference on Music Information Retrieval (ISMIR)*. 2009.

[8] Jason A. Hockman, Matthew E. P. Davies, and Ichiro Fujinaga. *One in the Jungle: Downbeat Detection in Hardcore, Jungle, and Drum and Bass*. 2012.

[9] Ajay Srinivasamurthy, André Holzapfel, and Xavier Serra. "In Search of Automatic Rhythm Analysis Methods for Turkish and Indian Art Music". In: *Journal of New Music Research* 43 (Mar. 2014). DOI: 10.1080/09298215.2013.879902.

[10] F. Krebs, S. Böck, and G. Widmer. "Rhythmic Pattern Modeling for Beat and Downbeat Tracking in Musical Audio". In: *ISMIR*. 2013.

[11] Ajay Srinivasamurthy and Xavier Serra. "A supervised approach to hierarchical metrical cycle tracking from audio music recordings". In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. May 2014, pp. 5217–5221. ISBN: 978-1-4799-2893-4. DOI: 10.1109/ICASSP.2014.6854598.

[12] Andre Holzapfel, Florian Krebs, and Ajay Srinivasamurthy. *Tracking the "odd": meter inference in a culturally diverse music corpus*. Oct. 2014.

MIREX structural segmentation datasets

MIREX initial dataset included **The Beatles** and a subset of **RWC**[13]
Traits of structural segmentation:

1. Ambiguous – more than one valid annotation for a given track
2. Subjective – different listeners can perceive different segments

Solutions: SALAMI[14], SPAM[15] include multiple annotations per track by different experts

---

[13] Masataka Goto et al. "RWC Music Database: Popular, Classical, and Jazz Music Databases". In: *ISMIR 2002, 3rd International Conference on Music Information Retrieval, Paris, France, October 13-17, 2002, Proceedings*. Jan. 2002.

[14] Jordan Smith et al. "Design and creation of a large-scale database of structural annotations.". In: *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011, Miami, Florida, USA*. Oct. 2011, pp. 555–560.

[15] Oriol Nieto and Juan Pablo Bello. "Systematic Exploration of Computational Music Structure Research.". In: *Proceedings of the 17th International Society for Music Information Retrieval Conference* (New York City, United States). New York City, United States: ISMIR, Aug. 2016, pp. 547–553. DOI: 10.5281/zenodo.1417661.

Traits of datasets

MIREX datasets:

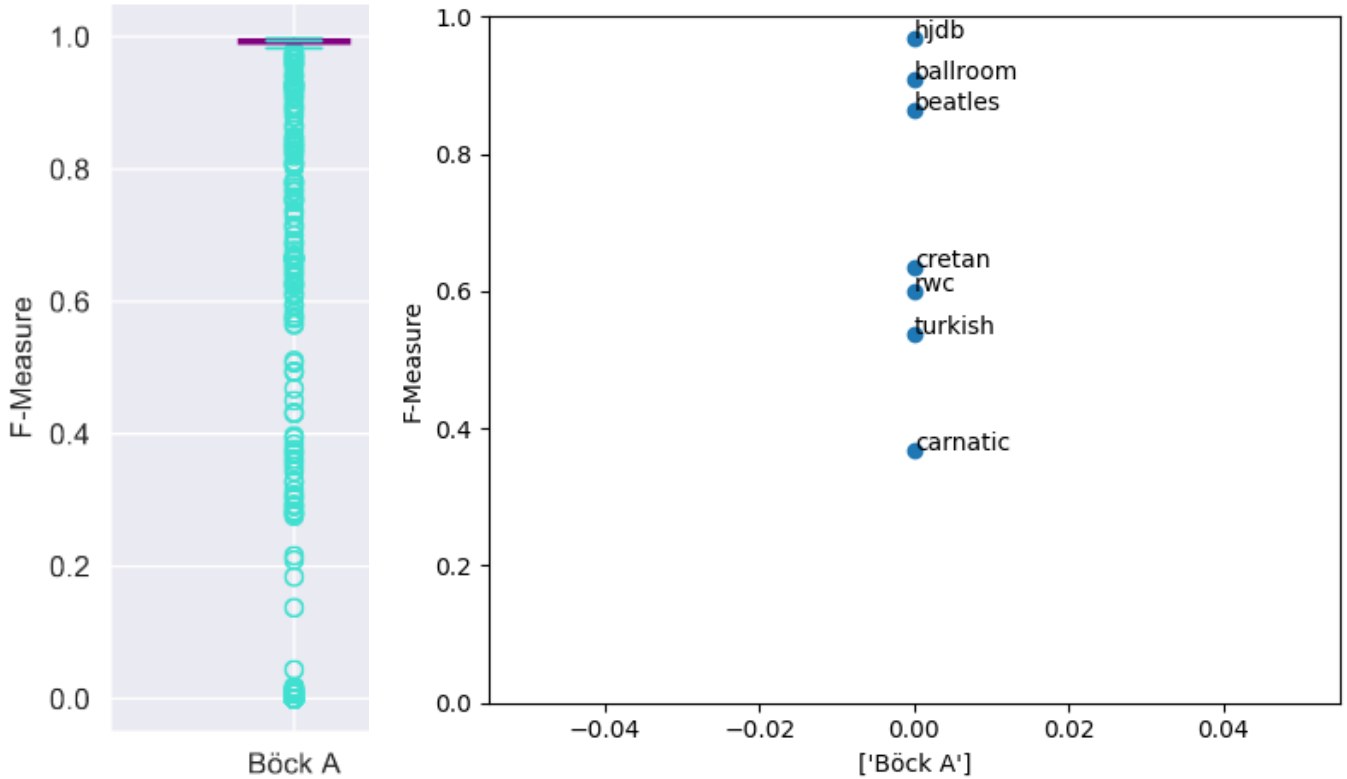
- MIREX datasets can have multiple annotators, a single annotator, or even semi-automated annotations using algorithm outputs
- Post-processing of first-pass, raw annotations involves iterative adjustment until annotators are satisfied
- Selection of songs for style-specificity (either targeting a specific style, or including diverse styles), to adjust western/non-western bias, or based on perceived difficulty (on a musical or signal processing level)

HarmonixSet:

- Songs were annotated by trained professional musicians who regularly work in music production environments (using DAW + MIDI)
- Mix of genres were chosen to be typical of ones used in the rhythm-action games, tendency to pop/EDM for dancing
- Tend to have a very stable tempo and a 4/4 time signature, however some atypical songs (classic rock, country, metal) were included with less stable tempo and which may deviate from a strict 4/4 meter

MIREX vs. HarmonixSet result – single case study

Results from MIREX 2016:
https://www.music-ir.org/mirex/wiki/2016:MIREX2016_Results
F-measure for downbeat estimation
"Böck A" (in HarmonixSet), "BK4" in MIREX[16]



[16] S. Böck, F. Krebs, and G. Widmer. "Joint Beat and Downbeat Tracking with Recurrent Neural Networks". In: *ISMIR*. 2016.

Gangnam style – HarmonixSet annotations

1. `# dataset/beats_and_downbeats/0388_gangnamstyle.txt`

```
0.079    1        1
0.54     2        1
1.017    3        1
1.494    4        1
1.92     1        2
2.374545          2        2
2.82909  3        2
3.283635          4        2
```

2. `# dataset/segments/0388_gangnamstyle.txt`

```
0.079     intro
6.010905  chorus
14.64726  verse
```

3. https://github.com/urinieto/harmonixset/blob/master/dataset/jams/0388_gangnamstyle.jams[17]

[17] Eric Humphrey et al. "JAMS: A JSON Annotated Music Specification for Reproducible MIR Research". In: Oct. 2014.

Gangnam style results – beats

Beat time output[18]

```
beat times: [ 0.09  0.55  1. 1.46 1.91 2.37 2.82 3.28 ... ]
```

vs. HarmonixSet ground truth

```
0.079    1        1
0.54     2        1
1.017    3        1
1.494    4        1
1.92     1        2
2.374545          2        2
2.82909  3        2
3.283635          4        2
```

Clicks: BEAT CLICKS

[18] *madmom.features.beats*. URL: https://madmom.readthedocs.io/en/latest/modules/features/beats.html.

Gangnam style results – downbeats

Downbeat time output[19]

```
downbeat times: [1.0, 2.82, 4.64, 6.46, ...]
```

vs. HarmonixSet ground truth

```
# awk '{ if ($2 == 3) { print }}' \
#     dataset/beats_and_downbeats/0388_gangnamstyle.txt
1.017   3       1
2.82909 3       2
4.64727 3       3
6.46545 3       4
...
```

Clicks: DOWNBEAT CLICKS

[19] *madmom.features.downbeats*. URL: https://madmom.readthedocs.io/en/latest/modules/features/downbeats.html.

Gangnam style results – segmentation

Structural segmentation output[20]

```
segments: [  0.58049887   8.08054422  11.74929705
            14.09451247 ...
```

vs. HarmonixSet ground truth

```
0.079 intro 6.010905 chorus 14.64726 verse ...
```

SEGMENT PAUSES
GROUND TRUTH SEGMENT PAUSES

[20] *Laplacian segmentation*. URL: https://librosa.org/librosa_gallery/auto_examples/plot_segmentation.html.

Audio alignment

YouTube music videos, or different file formats or recordings obtained by researchers, may have temporal differences with the original mp3 files. Alignment data is included to

> *... help align the audio in case researchers obtain audio data with different compression formats that might include certain small temporal offsets.*

Algorithms used for alignment:

1. Dynamic time warping[21],[22]
2. Onsets[23]

---

[21] Meinard Müller. "Dynamic time warping". In: *Information Retrieval for Music and Motion* 2 (Jan. 2007), pp. 69–84. DOI: 10.1007/978-3-540-74048-3_4.

[22] *Audio Alignment for Harmonix Set*. URL: https://github.com/urinieto/harmonixset/blob/master/notebooks/Audio%20Alignment.ipynb.

[23] *librosa.onset.onset_detect*. URL: https://librosa.org/doc/main/generated/librosa.onset.onset_detect.html.

Dataset recreation and copyright

Data provided to allow independent recreation of dataset includes:
Identifiers in shared music databases:

1. MusicBrainz ID[24], open music encyclopedia including unique identifiers for recordings, releases, artists, etc.
2. AcoustID (https://acoustid.org/), open source fingerprinting service to easily match audio content associated with MusicBrainz ids
3. YouTube URLs, including alignment information with the original mp3 files used in the paper

Audio/DSP features:

1. mel spectrograms for the original mp3 files
2. estimated onsets for the first 30 seconds of audio from librosa

[24] A. Swartz. "MusicBrainz: a semantic Web service". In: *IEEE Intelligent Systems* 17.1 (2002), pp. 76–77. DOI: 10.1109/5254.988466.

Conclusion

Future work: meta-study to analyze

1. Results for algorithms across MIREX challenge datasets
2. Same algorithms applied to the HarmonixSet
3. Comparison of results to gauge the characteristics of the HarmonixSet over established datasets, e.g. in the vein of Peter Grosche, Meinard Müller, and Craig Sapp. "What Makes Beat Tracking Difficult? A Case Study on Chopin Mazurkas". In: *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010, Utrecht, Netherlands*. Aug. 2010, pp. 649–654

Source latex and Python code for this presentation: https://gitlab.com/sevagh/MIR-presentations/-/tree/master/01-harmonix-set