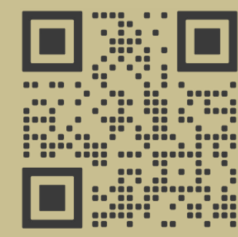


A Meta-Corpus of Melodic and Harmonic Transcriptions

Georgia Tech Center for Music Technology



Computational & Cognitive Musicology Lab



CoCoPops is a dataset of **expert transcriptions** of (currently) 20th century popular music, extending a preexisting dataset, and combining with a second into a common format. CoCoPops **currently** has two main parts:

- (Coming Soon: The Billboard Melodic Music Dataset (BiMMuDa), *forthcoming*)

Harmony and melody are two perceptually salient musical features, yet most datasets consist of either harmonic annotations, or melody for classical music. We **lack expert-quality melodic data for popular music**. Adding (and unifying) melodic as well as perceptual (**emotion**) data to existing corpora facilitates research in **computational musicology and perception**.

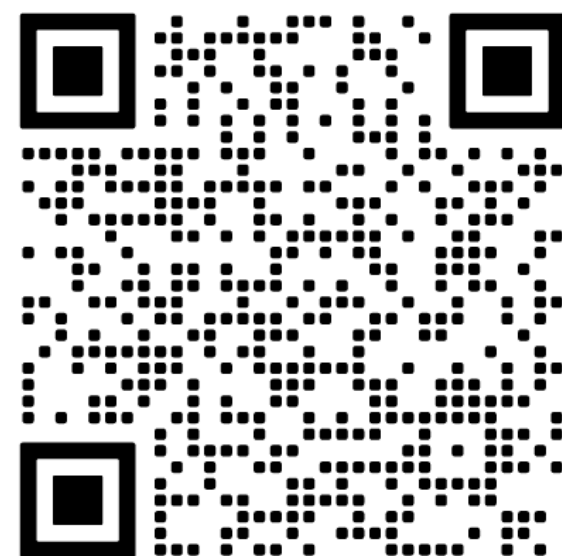
- Multiple encoding formats (e.g., musicXML, kern, MIDI)
- Different representation schemes (e.g., RN, Leadsheet, Harte, etc.)
- Low or unknown quality *or* data scarcity
- Aimed at symbolic *or* audio

The CoCoPops dataset, released under a **CC-BY-4.0 license**, aims to **facilitate cross-corpus study** by **extending** and **integrating** expert data using widely-used representations in a common format (humdrum) **supported** by all modern computational musicology software (e.g., humdrumR, music21) with a tabular format (i.e., csv) that can support audio and symbolic data.

"Rebel Yell" by Billy Idol (1984)

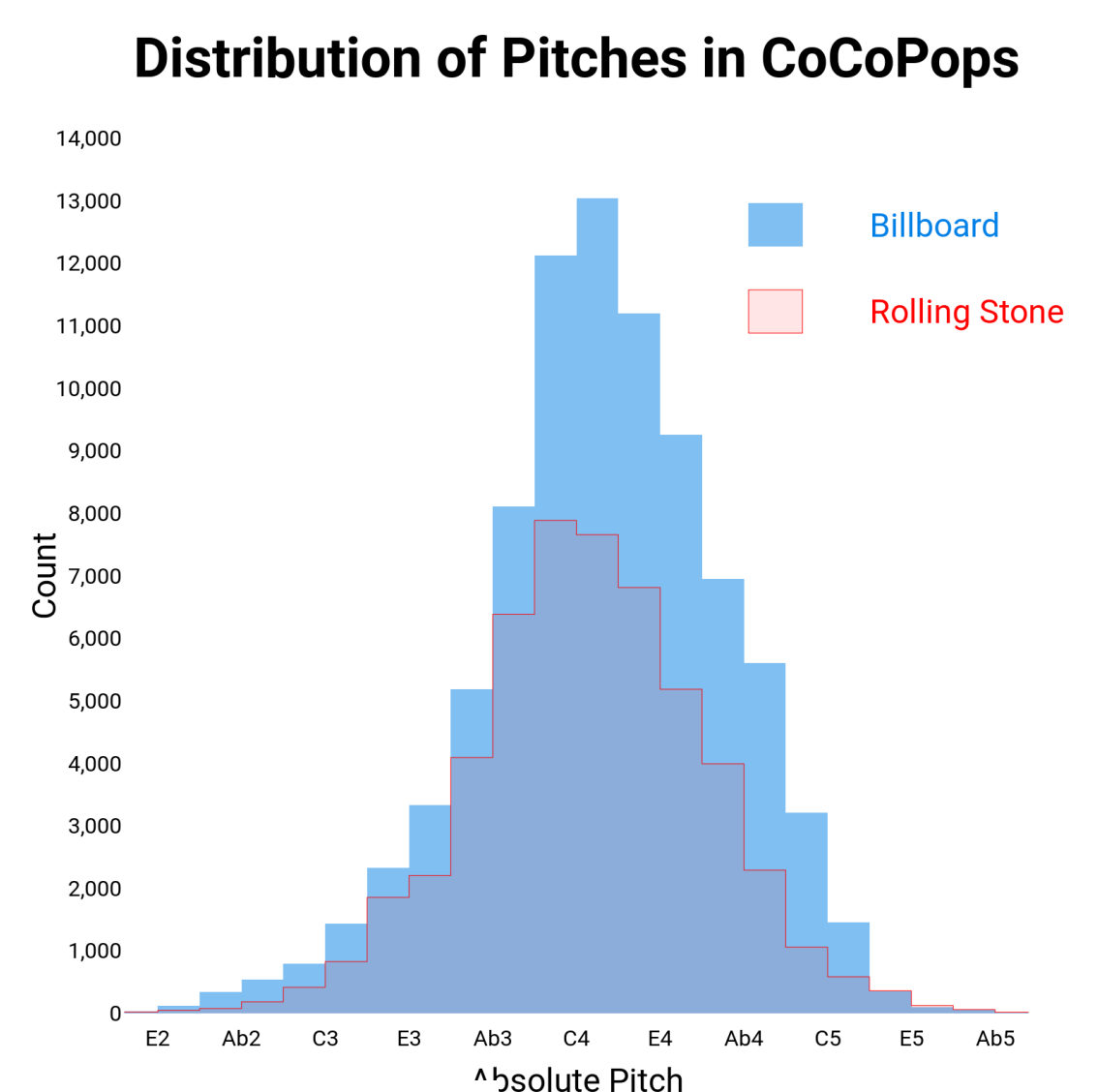
Metadata

Find
CoCoPops
on
GitHub



Search GitHub for #humdrum #digital-scores
to find other humdrum datasets

he distributions of absolute (vocal) pitch in the two datasets are similar.



Scale Degree	Billboard Count	Rolling Stone Count
1	~18,500	~17,000
2	~10,500	~6,500
3	~10,500	~6,500
4	~8,500	~6,500
5	~15,000	~10,000
6	~8,500	~6,500
7	~4,500	~4,500

The ranked frequency of scale degrees in the two datasets are nearly identical.

Nine of the ten most frequent functional harmonies in are the same between the two datasets.

Only the fourth and fifth spots (bVII and I) and the tenth and eleventh spots (iv and iii) differ slightly in relative frequency.

Nine of the ten most frequent functional harmonies in are the same between the two datasets.

Only the fourth and fifth spots (bVII and I) and the tenth and eleventh spots (iv and iii) differ slightly in relative frequency.

<https://github.com/Computational-Cognitive-Musicology-Lab/CoCoPops>