The Coordinated Corpus of Popular Musics (CoCoPops)

A Meta-Corpus of Melodic and Harmonic Transcriptions

Claire Arthur (claire.arthur@gatech.edu) & Nat Condit-Schultz (natcs@gatech.edu)





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:

- The Billboard subset, consisting of new melodic transcriptions added to 214 harmonic transcriptions from the existing McGill Billboard dataset (Burgoyne et al., 2011).
- The Rolling Stone subset, a conversion of Temperley's and de Clercq's (2011) RS200 corpus to a common tabular (humdrum) format.

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

Rationale

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.

Computational research is frequently burdened by datasets across:

- 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.

Sample CoCoPops File

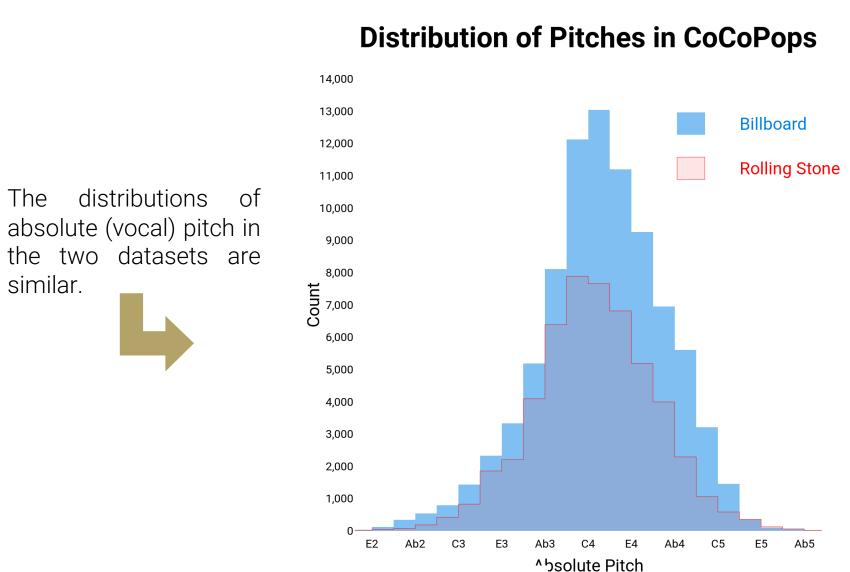
			"Re	ebel Yell	by Billy	y Idol (19	984)			
Melody	Lyrics	Rhyme	Interp.Harm	Harmony	Phrase	Instrument	t Time	Valence	Arousal	RMS
441	•		•			***				
**kern *ICvox	**silbe	**rhyme	**harmony *M4/4	**harte *M4/4	**phrase	**leadinstrument	**timestamp	**valence_ratings	**arousal_ratings	** FMS
*VRlead *Hstimme	*	*	*	*	*	*	*	*	*	*
* r;	*	*	*b: 1r;	*b: N.C.	*b:	*b:	*b:	* r;	* r;	* r;
=1 *>Letter>A	=1 *>Letter>A	=1 *>Letter>A	=1 *>Letter>A	=1 *>Letter>A	=1 *>Letter>A	=1 *>Letter>A	=1 *>Letter>A	=1 *>Letter>A	=1 *>Letter>A	=1 *>Letter>A
*>Label>Intro *M4/4	*>Label>Intro	*>Label>Intro	*>Label>Intro *	*>Label>Intro	*>Label>Intro	*>Label>Intro *	*>Label>Intro *	*>Label>Intro	*>Label>Intro	*>Label>Intro
*clefF4	*	*	* 1Bmm	* B:min7	* newline	* synthesizer	* 1.022	* 102 117 68 62	* 0 95 65 113	* 64.05
=2	=2	=2	=2	=2	=2	=2	=2	=2	=2	=2
1r =3	• =3	=3	1Bmm =3	B:min7 =3	=3	synthesizer =3	2.473 =3	62 117 68 63	59 69 60 68 =3	63.39 =3
1r =4	=4	=4	1Bmm =4	B:min7 =4	=4	synthesizer =4	3.924 =4	62 62 68 63 =4	59 69 60 68 =4	62.24 =4
1r =5	=5	=5	1Bmm =5	B:min7 =5	=5	synthesizer =5	5.376 =5	93 62 68 63 =5	59 69 77 68 =5	62.28 =5
<pre>*>Letter>B *>Label>Interlude</pre>	*>Letter>B *>Label>Interlude	*>Letter>B e *>Label>Interluc	*>Letter>B de *>Label>Interlud	*>Letter>B e *>Label>Interlude	*>Letter>B *>Label>Interlud	*>Letter>B e *>Label>Interlude	*>Letter>B *>Label>Interlude	*>Letter>B *>Label>Interlude	*>Letter>B *>Label>Interlude	*>Letter>B *>Label>Interl
1r =6	• =6	• =6	1GMM =6	G:maj7 =6	newline =6	synthesizer =6	6.827 =6	100 62 68 63 =6	59 75 80 80 =6	76.26 = 6
1r =7	• =7	• =7	1F#mm	F#:min7	• =7	synthesizer	8.269 =7	100 62 68 63	59 79 81 88 =7	75.01 =7
1r			1Em	E:sus2(b7)		synthesizer	9.711	127 62 68 63	59 80 89 92	74.85
=8 1r	=8	=8	=8 1Em	=8 E:sus2(b7)	=8	synthesizer	=8 11.153	127 85 68 63	59 80 92 94	=8 74.5
=9 1r	=9	=9	=9 1GMM	=9 G:maj7	=9 newline	synthesizer	=9 12.594	127 85 68 63	127 81 92 94	72.88
=10 1r	=10	=10	=10 1F#mm	=10 F#:min7	=10	=10 synthesizer	=10 14.04	=10 127 85 68 63	=10 127 81 92 94	=10 71.83
=11 1r	=11	=11	=11 1Em	=11 E:sus2(b7)	=11	=11 synthesizer	=11 15.485	=11 127 85 68 63	=11 127 81 92 94	=11 74.09
=12 1r	=12	=12	=12 2Em	=12 E:sus2(b7)	=12	=12 synthesizer	=12 16.931	=12 127 85 68 63	=12 127 81 92 94	=12 73.29
			4DM 4AM	D:maj A:sus2		synthesizer synthesizer	17.654 18.015			•
=13	=13	=13	=13	=13	=13	=13	=13	=13	=13	=13
*>Letter>A *>Label>Trans	*>Letter>A *>Label>Trans	*>Letter>A *>Label>Trans	<pre>*>Letter>A *>Label>Trans</pre>	<pre>*>Letter>A *>Label>Trans</pre>	*>Letter>A *>Label>Trans	*>Letter>A *>Label>Trans	*>Letter>A *>Label>Trans	*>Letter>A *>Label>Trans	*>Letter>A *>Label>Trans	*>Letter>A *>Label>Trans
1r =14	=14	=14	1Bmm =14	B:min7 =14	newline =14	drums =14	18.376 =14	127 85 68 63 =14	127 84 92 94 =14	71.5 =14
1r =15	• =15	=15	1Bmm = 15	B:min7 =15	• =15	drums =15	19.817 =15	127 85 68 63 =15	127 84 92 94 =15	72.71 = 15
1r =16	• =16	• =16	1Bmm =16	B:min7 =16	• =16	drums =16	21.257 =16	127 85 68 63 =16	127 84 92 94 =zZZ16	71.78 =16
1r =17	=17	=17	1Bmm =17	B:min7 =17	=17	drums =17	22.698 =17	115 85 68 63 =17	127 90 92 101 =17	74.54 =17
*>Letter>C	*>Letter>C	*>Letter>C	*>Letter>C	*>Letter>C	*>Letter>C *>Label>Verse	*>Letter>C	*>Letter>C	*>Letter>C	*>Letter>C	*>Letter>C
*>Label>Verse 4r	*>Label>Verse	*>Label>Verse	*>Label>Verse 1Bm	*>Label>Verse B:min	newline	*>Label>Verse voice	*>Label>Verse 24.138	*>Label>Verse 94 85 68 63	*>Label>Verse 127 97 98 109	*>Label>Verse 72.37
BF# 4E	Last night		:	:	:	:	:	90 85 68 63 90 85 68 63	127 97 100 109 127 97 100 109	72.36 72.45
8D 8E	a lit-	:	:				:	90 85 68 63 90 85 68 63	127 97 100 109 127 97 100 109	71.48 70.97
8F# =18	-tle =18	· =18	• =18	• =18	• =18	• =18	• =18	96 85 68 63 =18	127 97 100 109 =18	70.12 =18
4E 8D	dan- -cer		1Bm	B:min	:	voice	25.582	96 85 68 63 96 85 68 63	127 97 100 109 127 97 100 109	75.41 73.64
8r 4r		:	:	•				96 85 68 63 96 85 68 63	127 97 100 109	68.65 73.13
4E	came	:	:	:	:	:	:	96 85 68 63	127 97 100 109 127 97 100 109	72.87
=19 4F#	=19 danc-	=19	=19 1Bm	=19 B:min	=19	=19 voice	=19 27.026	96 85 68 63	=19 127 97 100 109	=19 73.78
4F# 8E	-in' to	:	:		:			96 85 62 63 96 85 55 63	127 97 100 109 127 97 100 109	72.99 73.35
4E [8BB	my door.	A	:	:	:	:	:	96 85 55 63 96 85 55 63	127 97 100 109 127 97 100 109	72.2 71.95
-		*>Label>Outro	*>Label>Outro	*>Label>Outro	1 more records *>Label>Outro	*>Label>Outro	*>Label>Outro	*>Label>Outro		
4f# 2B	more.		1Bmm	B:min7	newline	synthesizer	278.017	104 92 68 99 104 92 68 99	127 115 115 124 127 115 115 124	75.95 72.83
4r	-194	-104	-194		-104	-104	-104	103 92 68 99	127 115 115 124	71.51
=194 1r	=194	=194	=194 1Bmm	=194 B:min7	=194	synthesizer	=194 279.47	103 92 68 99	=194 127 115 115 124	=194 70.81
=195 1r	=195	=195	=195 1Bmm	=195 B:min7	=195	=195 synthesizer	=195 280.924	=195 103 92 68 99	=195 127 114 107 124	=195 73.12
=196 1r	=196	=196	=196 1Bmm	=196 B:min7	=196	=196 synthesizer	=196 282.377	=196 110 92 68 99	=196 127 110 96 124	=196 71.62
=197 1r	=197	=197	=197 1Bmm	=197 B:min7	=197 newline	=197 synthesizer	=197 283.83	=197 127 92 68 99	=197 127 106 91 120	=197 68.41
r; r;	:	:	1r; 1r;	N.C. N.C.		r r	285.884 290.664	r; r;	r; r;	r; r;
*- !!!BillboardChart	*- Date: 1984/04/07	*_	*-	*-	*-	*-	*-	*-	*-	*_
!!!BillboardPeak:	46									
!!!BillboardWeeks	l)				
	nanual error check	ing and editing.								
!!!OTL: Rebel Yel	l		+ (implied pitch cl	ass set. with bass	pitch in first no	sition)	Metada	ata		
!!!RDF**harmony:	**recip + pitch c	lass of root + has					iviciaud	ıla		
!!!RDF**phrase: r	ewline = linebrea : McGill audio ti	k in original McG	ill transcription	- 2011		(
!!!RDT: 2017/01/1	.2/	mescamp in second								
!!!SampleActualRa	7									
!!!SampleTargetRa	ink: 61									

New Data Includes:

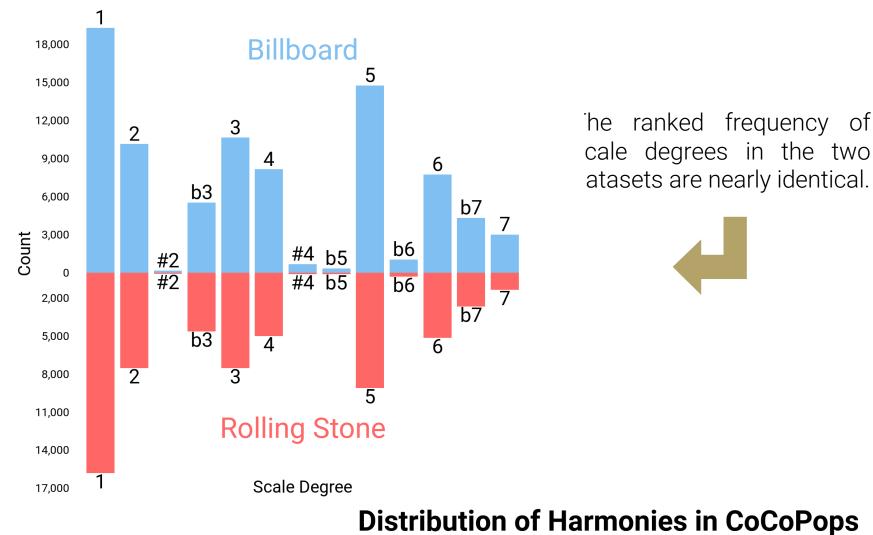
- 214 expert melodic transcriptions
- Continuous valence & arousal ratings (100 song subset)
- Root Mean Square (RMS) (100 song subset)
- Linked MusicBrainz IDs
- All Billboard and RS200 data to common format

Summary Statistics Comparison

Despite differences in their sampling and transcription procedures, basic musical features are quite consistent between the two datasets, which suggests that treating them like a single large dataset is appropriate.



Distribution of Scale Degrees in CoCoPops



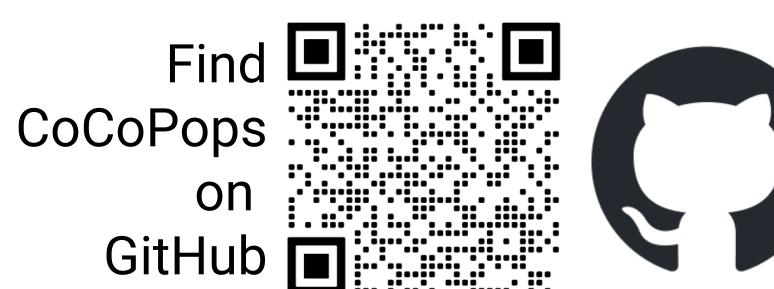
4,000 Billboard IV V 3,000 2,000 bVII Nine of the ten most vi bVI ii bIII iv iii functional frequent 1,000 harmonies in are the Cour same between the two datasets. Only the fourth and fifth

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

1,000
2,000
3,000
4,000

Rolling Stone
Harmony

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



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