THE COORDINATED CORPUS OF POPULAR MUSICS (COCOPOPS): A META-CORPUS OF MELODIC AND HARMONIC TRANSCRIPTIONS

Claire Arthur
Georgia Institute of Technology
School of Music
claire.arthur@gatech.edu

Nathaniel Condit-Schultz
Georgia Institute of Technology
School of Music
natcs@gatech.edu

ABSTRACT

This paper introduces a new corpus, CoCoPops: The Coordinated Corpus of Popular Musics. The corpus can be considered a "meta corpus" in that it both extends and combines two existing corpora—the widely-used McGill Billboard corpus the and RS200 corpus. Both the McGill Billboard corpus and the RS200 contain expert harmonic annotations using different encoding schemes and each represent harmony in fundamentally different ways: Billboard using a root-quality representation and the RS200 using Roman numerals. By combining these corpora into a unified format, using the well-known `**kern` and `**harm` representations, we aim to facilitate research in computational musicology, which is frequently burdened by corpora spread across multiple encoding formats. The format will also facilitate cross-corpus comparison with the large body of existing works in `**kern` format. For a 100-song subset of the CoCoPops-Billboard collection, we also provide participant ratings of continuous valence and arousal ratings, along with the RMS (Root Mean Square) signal level and associated timestamps. In this paper we describe the corpus and the procedures used to create it.

1. INTRODUCTION

In 2011, Burgoyne et al. [1] introduced a dataset that would have a lasting influence in the ISMIR community: the McGill Billboard corpus, a set of expert harmonic analyses of commercial pop songs. This dataset—and the Harte [2] standard for encoding chord symbols that it adopted—has become a standard in the MIR community, for example, being used as training and testing data in the MIREX competition for Audio Chord Estimation since 2008. Around the same time, Trevor de Clercq and David Temperley independently created another rock music dataset—the RS200 corpus—which would ultimately consist of 200 harmonic *and* melodic transcriptions [3, 4]; Though perhaps less well known in the MIR community, their corpus has been the basis for several computational musicology papers [4]. While other datasets of popular-style music harmony have been released (e.g., Isophonics [2]), the Billboard and RS200 datasets stand out for their use of experts to encode the annotations, the rigor of their sampling methodologies, and the detail of their procedural documentation.

The field of computational musicology suffers from perennial data scarcity [5]; What few symbolic corpora exist are largely biased towards Western classical music [6], which is relatively easy to digitize due to its basis in notated scores. Unlike classical music, popular music must generally be transcribed from audio recordings, with melody transcription being a particularly time-consuming task. Although more open-source data can be found (e.g., crowd-sourced arrangements from www.musescore.com) and MIR algorithms for tasks such as source separation and automatic transcription are improving, both procedures are prone to high levels of error that is undesirable for either computational music analysis or training machine learning models [6]. The RS200 is still the only major corpus of expert melodic transcriptions of popular music; the pairing of these melodic transcriptions with harmonic analyses affords sophisticated analysis of tonality in popular music.

In this paper we present a corpus which extends the Billboard corpus to include expert-transcribed melodies for a sizable subset of the original corpus (214 songs presently). By adding melodic transcriptions to an existing corpus of harmonic annotations (the Billboard corpus), we create a dataset fully comparable to the RS200. We also translate both the Billboard and RS corpora into humdrum data formats, creating two comparable datasets which together form a super-corpus we call the Coordinated Corpus of Popular Music (CoCoPops). In addition to melodic and harmonic transcriptions, CoCoPops includes entirely new annotations of rhyme schemes in both subcorpora and continuous valence and arousal ratings in a 100-song subset. Like the When In Rome project [7], CoCoPops aims to facilitate musicological and MIR research by making a large body of data available in a consistent, standard format. In the sections that follow, we describe in detail the original two datasets that CoCoPops is built on, the procedures we used to generate new data, and the content of CoCoPops.

```
# title: Honky Tonk Women
# artist: The Rolling Stones
# metre: 4/4
# tonic: G

0.0	silence
0.424489795	A, intro, | N | N | N | N |
9.2	| G:maj | G:maj | G:maj | G:maj |
17.689795918	B, verse, | G:maj | G:maj | C:maj . . C:sus4 | C:maj |, (voice
26.048979591	| G:maj | A:maj | D:maj . . D:sus4 | D:maj |
34.351020408	| G:maj | G:maj | C:maj . . C:sus4 | C:maj |
42.579591836	| G:maj | D:maj | G:maj | G:maj |
50.669387755	C, chorus, | G:maj | D:maj | G:maj | G:maj |
58.734693877	| G:maj | D:maj | G:maj | G:maj |
66.620408163	B, verse, | G:maj | G:maj | C:maj . . C:sus4 | C:maj |
74.620408163	| G:maj | A:maj | D:maj . . D:sus4 | D:maj |
82.579591836	| G:maj | G:maj | C:maj . . C:sus4 | C:maj |
90.553151927	| G:maj | D:maj | G:maj | G:maj |
98.489795918	C, chorus, | G:maj | D:maj | G:maj | G:maj |
106.375510204	| G:maj | D:maj | G:maj | G:maj |, voice)
114.188639455	B, solo, | G:maj | G:maj | C:maj . . C:sus4 | C:maj |, (guitar
121.991836734	| G:maj | A:maj | D:maj . . D:sus4 | D:maj |
129.75510204	| G:maj | G:maj | C:maj . . C:sus4 | C:maj |
137.534693877	| G:maj | D:maj | G:maj | G:maj |, guitar)
145.255873015	C, chorus, | G:maj | D:maj | G:maj | G:maj |, (voice
153.028571428	| G:maj | D:maj | G:maj | G:maj |
160.718367346	C, chorus, | G:maj | D:maj | G:maj | G:maj |
168.440816326	| G:maj | D:maj | G:maj G:maj/3 G:maj/11 G:maj/#11 | G:maj/5 G:maj |
179.118730158	silence
182.282448979	end
```

Figure 1. Sample annotation file from the original McGill Billboard corpus ("Honky Tonk Woman," The Rolling Stones).

2. BACKGROUND

2.1 The McGill Billboard Corpus

The McGill Billboard [1] corpus contains annotations of 739 [1] unique songs, all sampled from the Billboard Hot 100 charts between 1958 (when Billboard magazine began publishing this chart) and 1991. The authors used a stratified sampling procedure to gather as representative a sample as possible, sampling a (roughly) equal number of songs from each of three "eras" (60s, 70s, 80s) while also accounting for chart position (1–100).

The McGill Billboard transcription process involved a team of more than two dozen people, included "auditions to identify musicians with sufficient skill to transcribe reliably and efficiently," and cost upwards of $20,000 [1]. The process included creating manual annotations of the chords, formal sections (e.g., verse, chorus), phrases (loosely defined), key(s), and meter in each sampled song, conducted by two independent annotators. A third "meta-annotator" compared the two versions for differences and combined them into a single, final transcription.

The McGill Billboard chord annotations are encoded using the representation scheme proposed by Harte in 2005 [8] and later expanded and revised in 2010 [2]. This representation uses a syntax that is common in popular music lead sheets, where chords are represented as a root note with a set of intervals above the root, with the most common chord types given a list of shorthand symbols (e.g., `C:maj`, `A:7`). The McGill annotations are encoded in plain-text files with line breaks representing new phrases, each line tagged with the dominant instrument (or vocals) in that phrase. An example of an original file from the McGill Billboard corpus is shown in Figure 1. [2]

2.2 The RS200 Corpus

The Rolling Stone corpus was first described in a paper published in 2011 [3], initially dubbed the RS5x20 corpus. This original 100-song corpus (RS5x20) contained harmonic annotations of the top 20 songs listed, for each of five decades from the 1950s through the 1990s, on Rolling Stone magazine's list of the "500 Greatest Songs of All Time" (as first published in 2004). The corpus was later expanded to 200 songs (the RS200 corpus), and also added melodic transcriptions for each song [4], making it the first public corpus of expert melodic transcriptions of popular music. Since the remaining 400 songs on Rolling Stone's list were not chronologically balanced, the second set of 100 songs was chosen based on rank position alone. While the Billboard charts are based on commercial sales, the Rolling Stone list was based on votes from experts (specifically, "172 rock stars and leading authorities"). Although one may suspect that these two corpora would substantially overlap, in fact there are only fifteen songs in common.

The RS200 annotations are spread over multiple separate files per song: one with the timestamps, two with the harmonic analyses (one per annotator), another with the melody transcription, and (for an 80-song subset) a fifth with lyrics. Unlike the Billboard corpus, the RS200 chords are annotated using Roman numerals; Similarly, the melody transcriptions are encoded as scale-degree annotations, with direction markers to clarify octave and contour. Rhythmic durations are not encoded at all, only the timing of note onsets: each measure of music is divided into regular steps representing metric positions, with notes placed at steps indicating onsets and dots representing empty steps. The number of steps per measure is dynamic, depending on the meter and the lowest metric position needed to represent onsets in that measure. For instance, a measure that contains only one note that arrives on the second half of the first beat (e.g., the "and of 1") requires division into eighth notes, so that measure will have eight steps with only a note at the fourth step and the rest dots. However, a measure with only a single note that lands on the downbeat can be represented with just one token. Sample files from the RS200 corpus can be seen in Figure 2.

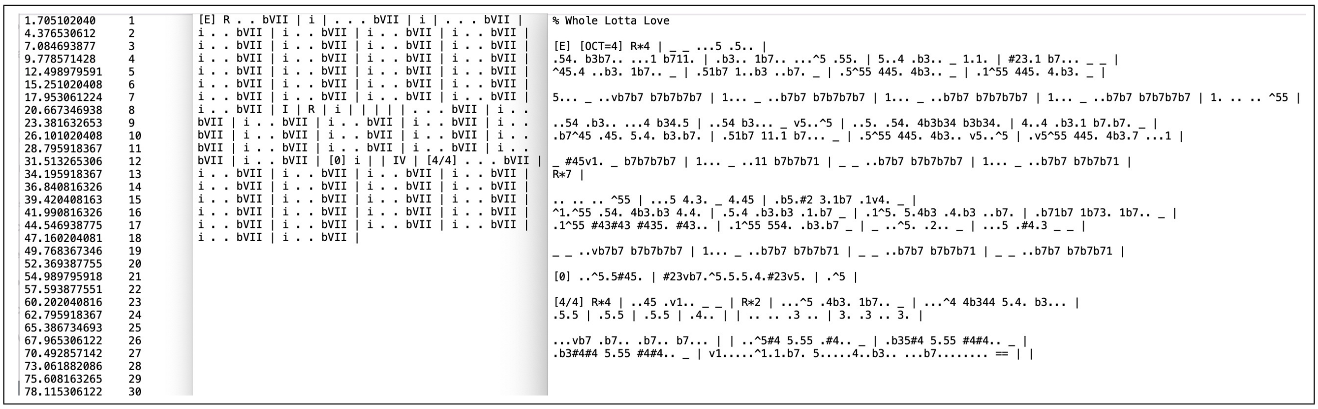
2.3 Related Work

The most closely-related work to ours is another extension to the McGill Billboard corpus by Christopher White et al. [9], which adds timbral and textural annotations to the entire Billboard corpus. Annotators of this corpus listened to the songs and notated "all moments of change" within each track according to three broad categories: the "domain" of change (such as the instrument group, harmony, lyrics, texture, etc.); the "genera" of each change within the relevant domain (such as a change to "solo" within a texture category); and an "event type" which solely denotes one of three options: a change, entry, or exit. We intend to work with the authors for a future release of CoCoPops to incorporate this textural and timbral information as well.

A major drawback of both the Billboard and Rolling Stone samples is their overwhelming bias towards music from before 1991. Two recent projects have sought to right this imbalance by creating corpora of more modern popular music to complement the Billboard sample: White et

[1] Note that a small subset has been withheld from the public to serve as testing data for the MIREX competition.

[2] A separate set of `mirex` text-files includes only the chords, but with a timestamp for every chord.



```
1.705102040	1
4.376530612	2
7.084693877	3
9.778571428	4
12.498979591	5
15.251020408	6
17.953061224	7
20.667346938	8
23.381632653	9
26.101020408	10
28.795918367	11
31.513265306	12
34.195918367	13
36.840816326	14
39.420408163	15
41.990816326	16
44.546938775	17
47.160204081	18
49.768367346	19
52.369387755	20
54.989795918	21
57.593877551	22
60.20204816	23
62.795918367	24
65.386734693	25
67.965306122	26
70.492857142	27
73.061882086	28
75.608163265	29
78.115306122	30
```

```
[E] R . . bVII | i | . . . bVII | i | . . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | I | R | i | | | | | . . . bVII | i . .
bVII | i . . bVII | i . . bVII | i . . bVII | i . .
bVII | i . . bVII | i . . bVII | i . . bVII | i . .
bVII | i . . bVII | i . . bVII | i . . bVII | i . .
bVII | i . . bVII | [0] i | | IV | [4/4] . . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII | i . . bVII | i . . bVII |
i . . bVII | i . . bVII |
```

```
% Whole Lotta Love

[E] [OCT=4] R*4 | _ _ ...5 .5.. |
.54. b3b7.. ...1 b711. | .b3.. 1b7.. ...^5 .55. | 5..4 .b3.. _ 1.1. | #23.1 b7... _ _ |
^45.4 ..b3. 1b7.. _ | .51b7 1..b3 ..b7. _ | .5^55 445. 4b3.. _ | .1^55 445. 4.b3. _ |

5... _ ..vb7b7 b7b7b7b7 | 1... _ ..b7b7 b7b7b7b7 | 1... _ ..b7b7 b7b7b7b7 | 1... _ ..b7b7 b7b7b7b7 | 1. .. .. ^55 |

..54 .b3.. ...4 b34.5 | ..54 b3... _ v5..^5 | ..5. .54. 4b3b34 b3b34. | 4..4 .b3.1 b7.b7. _ |
.b7^45 .45. 5.4. b3.b7. | .51b7 11.1 b7... _ | .5^55 445. 4b3.. v5..^5 | .v5^55 445. 4b3.7 ...1 |

_ #45v1. _ b7b7b7b7 | 1... _ ..11 b7b7b71 | _ _ ..b7b7 b7b7b7b7 | 1... _ ..b7b7 b7b7b71 |
R*7 |

.. .. .. ^55 | ...5 4.3. _ 4.45 | .b5.#2 3.1b7 .1v4. _ |
^1.^55 .54. 4b3.b3 4.4. | .5.4 .b3.b3 .1.b7 _ | .1^5. 5.4b3 .4.b3 ..b7. | .b71b7 1b73. 1b7.. _ |
.1^55 #43#43 #435. #43.. | .1^55 554. .b3.b7 _ | _ ..^5. .2.. _ | ...5 .#4.3 _ _ |

_ _ ..vb7b7 b7b7b7b7 | 1... _ ..b7b7 b7b7b71 | _ _ ..b7b7 b7b7b71 | _ _ ..b7b7 b7b7b71 |

[0] ..^5.5#45. | #23vb7.^5.5.5.4.#23v5. | .^5 |

[4/4] R*4 | ..45 .v1.. _ _ | R*2 | ...^5 .4b3. 1b7.. _ | ...^4 4b344 5.4. b3... |
.5.5 | .5.5 | .5.5 | .4.. | | .. .. .3 .. | 3. .3 .. 3. |

...vb7 .b7.. .b7.. b7... | | ..^5#4 5.55 .#4.. _ | .b35#4 5.55 #4#4.. _ |
.b3#4#4 5.55 #4#4.. _ | v1.....^1.1.b7. 5.....4..b3.. ...b7........ == | |
```

Figure 2. Sample annotation files from the RS200 corpus ("Whole Lotta Love," Led Zeppelin). The image shows three files overlaid on top of each other from left to right: timestamps of each measure, key and chord annotations, and melody transcription.

al. [10] introduce the "Millenial corpus," a dataset of expert melodic transcriptions of twenty five popular songs written between 2015 and 2019. Beach and Arthur [6] created a much larger corpus of popular songs with annotations, although their annotations are derived algorithmically from the audio, and are quite noisy.

Our aim to combine two existing corpora into a single, homogeneous dataset is inspired by Mark Gotham's "when in Rome" project [7], which merges and reformats several existing classical corpora with Roman numeral annotations into a single collection in a common format. Our project to gather valence and arousal data for the Billboard sample was similarly inspired by the DEAM dataset: a dataset containing dynamic annotations of valence and arousal for 1,809 non-copyrighted (Creative Commons license) songs and song excerpts [11]. The majority of these annotations are of short excerpts ($\approx 45_s$) across numerous musical styles (folk, world, jazz, instrumental, pop); however, the dataset also includes ratings for 56 complete songs, which provide the most valuable information, according to the creators [11]. In addition, the quality of the audio recordings (and the musical content) in the DEAM sample is highly variable, as these recordings do not represent professionally published works. Our valence- and arousal-ratings for 100 complete, successful, commercial recordings will serve as a useful complement to the DEAM sample.

3. CORPUS OVERVIEW

The CoCoPops corpus consists of two collections: the Billboard and RS200 subcorpora. Each collection contains one file per song. In the CoCoPops-Billboard collection, all 739 of the original McGill Billboard files have an equivalent humdrum file. The contents of each file, however, vary: All 739 files contain all the originally encoded information (chords, keys, formal section labels, timestamps, phrase information) from the original McGill dataset, but all converted to humdrum format, and with a significant number of corrections (see Section 5). At present, 214 out of the 739 files include new expert melodic and lyric transcriptions, as well as an encoding of the rhyme scheme; 100 of those 214 songs also contain continuous user ratings of valence and arousal, as well as rolling RMS (root mean square) amplitude values of the audio, to approximate the changing sound level of the music—both sampled at a rate of 2_{Hz}. A sample CoCoPops-Billboard file is shown in Figure 3. In the CoCoPops-RS200 collection, each file contains the information originally spread over separate files—e.g., melody, harmony, time stamps, lyrics—in a single humdrum file. Unlike the original Billboard annotations which used Harte's encoding scheme (i.e., root+quality), the RS200 were originally annotated with Roman numerals. To facilitate analysis, we provide both types of harmonic annotations in both collections. In addition, since the original RS200 contained two independent transcriptions of the harmony, each CoCoPops-RS200 file includes two Roman numeral annotations (i.e., two `**harm` spines) side-by-side. Eighty of the files also include lyrics and syllable stress information.

The humdrum syntax is a plain-text format for representing musical information, organized into tab-delineated columns—called "spines"—representing different streams of data [12] (see `www.humdrum.org` for more information). Within the general humdrum syntax, various specific representation schemes can be defined [3]: Two of the most common representation schemes include the widely-known `**kern` representation of pitch information, the `**silbe` representation of lyrics, and the `**harm` representation of harmonic information in Roman-numeral format. Other relevant representations for the present collections include `**harte`—a humdrum representation for root+quality-style harmonic annotations (near-identical to the original annotation scheme used in the McGill Billboard corpus. This scheme is based on the syntax proposed by Chris Harte [2, 8] and the humdrum representation is described in Arthur et al. [13]); and `**rhyme`—a representation for rhyme schemes [14].

In the following sections we describe our procedures for gathering new data (e.g., melodic transcriptions), and

[3] Chapter 18 of the Humdrum User Guide illustrates how to create new humdrum representations.

```
**kern          **silbe         **rhyme         **harm          **harte         **phrase        **timestamp     **leadinstrument
*ICvox          *               *               *M4/4           *M4/4           *               *               *
*G:             *               *               *G:             *G:             *               *               *
*M4/4           *               *               *               *               *               *               *
*clefG2         *               *               *               *               *               *               *
r;              .               .               1r;             r               .               0               r
r;              .               .               1r;             r               .               0.511           r
=1              =1              =1              =1              =1              =1              =1              =1
8r              .               .               1I              G:maj           newline         7.616           voice
8b-             My              .               .               .               .               .               .
8b-             ba-             .               .               .               .               .               .
8b-             -by             .               .               .               .               .               .
8b-             whis-           .               .               .               .               .               .
8a              -pers           .               .               .               .               .               .
8g              in              .               .               .               .               .               .
8d              my              .               .               .               .               .               .
=2              =2              =2              =2              =2              =2              =2              =2
*>Letter>A      *>Letter>A      *>Letter>A      *>Letter>A      *>Letter>A      *>Letter>A      *>Letter>A      *>Letter>A
*>Label>Verse   *>Label>Verse   *>Label>Verse   *>Label>Verse   *>Label>Verse   *>Label>Verse   *>Label>Verse   *>Label>Verse
8e              ear             A               1I              G:maj           newline         10.017          voice
8d              _               .               .               .               .               .               .
4r              .               .               .               .               .               .               .
4r              .               .               .               .               .               .               .
8r              .               .               .               .               .               .               .
[8g             Mmm             .               .               .               .               .               .
=3              =3              =3              =3              =3              =3              =3              =3
2.g]            _               .               1IV             C:maj           .               .               .
8g              Sweet           .               .               .               .               .               .
[8e             no-             .               .               .               .               .               .
=4              =4              =4              =4              =4              =4              =4              =4
8e]             _               .               1I              G:maj           .               .               .
4.d             -thin's         .               .               .               .               .               .
```

Figure 3. Sample file from the CoCoPops corpus. This file ("Sweet Nothins,", Brenda Lee) includes the original McGill Billboard information alongside new melody and lyric information. Files in the valence and arousal subset (see Section 6) include three additional spines.

how we converted the preexisting datasets into humdrum formats.

4. MELODY TRANSCRIPTION

In the early stages of our project, we worked with four collaborators [4] to define transcription guidelines which could be applied consistently. We elected to transcribe only vocal parts, with focus on the "lead" vocal melody in each song—however, we agreed to encode important vocal harmonies or other "backing" vocals as needed. The vocal performances in the sample are often challenging to transcribe, including unpitched or quasi-pitched vocals, "blue" notes, glissandi, loose rhythms, and syncopation. Our goal was to create readable transcriptions using conventional musical syntax (beat positions, durations, notes) rather than mechanical, empirical terms (milliseconds, F0, etc.). This requires significant interpretation and quantization; However, we took care to not over-simplify melodies such that they became melodic reductions. Our transcriptions generally interpret rhythms using a 16th-note grid, but triplets and 32nd-notes are used sparingly at slow tempos; Similarly, pitches are encoded in standard western pitch categories (e.g., C#5, B4), ignoring most glissandi and blue notes. However, many vocal performances simply cannot be faithfully represented in traditional score categories: as such, we included provisions for indicating, as needed, unpitched or approximate pitch, "free" or approximate rhythms, glissandi, and blue notes—the complete details of these encodings are documented directly in the CoCoPops repository.

Ultimately, ten individuals contributed to our 214 melodic transcriptions: 94 transcriptions by the authors; 40 transcriptions by our four early collaborators, all graduate students in music performance or theory; 10 transcriptions by three (paid) undergraduate music students; and 70 transcriptions by one (paid) professional jazz performer, also a graduate student in jazz performance at the time. When transcribers were uncertain of their transcriptions, a second transcriber would collaborate on the final version. We gave our paid transcribers detailed instructions and have personally vetted and edited all transcriptions for consistency. The complete transcription guidelines are provided in the supplementary materials.

The exact audio files used for the original McGill transcriptions are not publicly available; for our transcriptions we accessed targeted songs via YouTube, taking care to confirm that each recording was the correct Billboard Hot 100 single. Unfortunately, some of the original McGill transcriptions do *not* match the targeted single, instead matching an album version, live version, or some other version of the same song; In a few cases, we could not find any recording that clearly matched the transcription. To improve consistency, we elected to modify the harmonic transcriptions for sixteen tracks to match the correct, single version from the Hot 100 chart. In most cases, these versions were very similar but slightly longer or shorter; in a few cases, the alternate version was in a different key or contained other significant differences. For these sixteen altered versions, the original timestamps were discarded and replaced with corrected timestamps in the correct single version, as available on YouTube. The CoCoPops repository includes files with links to each song's reference recording on YouTube, as well as MusicBrainz MBIDs for our 214-song melodic transcription subset.

5. CONVERTING EXISTING DATA

To create the new data, we converted the preexisting Billboard and RS data into humdrum format. During this process, we noted some errors in the Billboard transcriptions, which we corrected in our new data. Our expertise (education/credentials) in music performance and analysis are comparable to the original transcribers'. Most of these er-

[4] Thanks to Hubert Léveillé Gauvin, Gary Yim, Dana DeVlieger, Lissa Reed.

rors are unambiguous—for instance, a measure of music missing or a clear change of key that is not indicated. In only a few cases our "corrections" might by considered debatable. All errors and corrections are documented in our corpus repository. Each file in CoCoPops also includes a wealth of meta-data, including track information—title, original artist, release date, etc.—and sampling information, like the rank on the Rolling Stone 500 list and chart position on the Billboard Hot 100.

5.1 Billboard Data

We created a custom R script to convert the original Billboard corpus files (available at ddmal.music.mcgill.ca/research) into a humdrum representation. The harmonic annotations are encoded in a `**harte` spine with the timestamps in a `**timestamp` spine. Along with the `**harte` representation, we also include a `**harm` spine in each file: the humdrum standard for representing Roman numerals. Whereas the original harmonic transcriptions focus on the literal pitch-content played by rhythm-section instruments (ignoring vocal parts), Roman numerals represent harmony at a higher level of abstraction, incorporating the broader tonal context. This means that, for example, open-fifth "power chords" are interpreted as major or minor triads (numerals) based on the key, context, and vocal melody. For illustration, the original transcription of the track "I'm Going Down," by Bruce Springsteen, consists entirely of two repeated patterns: `A:5-E:5-F#:5-D:5` and `A:maj-E:maj-F#:min-D:maj`. We interpret both of these patterns as `I-V-vi-IV`. To create this `**harm` information, we wrote an R script to parse each file and replace under-specified chords (like `C5`) with the full triad expected given the key-signature and/or explicit triads indicated on the same root in the same song. This process was effective in the vast majority of cases; however, for songs with ambiguous modality we identified the triad manually. The harmonic rhythm is also indicated in the `**harm` spine using standard humdrum rhythmic duration tokens.

The original McGill data includes two, parallel, formal encodings: named sections (e.g., verse, chorus) and abstract letters (e.g., AABA). These parallel encodings are not redundant, as the transcribers used letters to indicate more abstract repetition (mainly of chord progressions)—for example, a guitar solo section which reuses the chord progression from the verse will be labeled "solo", but use the same letter designation (e.g., `A`) as the verse. We encode both formal representations, independently, in hierarchical https://www.humdrum.org/guide/ch20/: Abstract formal labels are encoded in interpretation records of the form `*>Letter>A`; formal names are encoded in separate records of the form `*>Label>Verse`. Phrases in the music (originally represented with line breaks) are indicated by the presence of the token `newline` in a `**phrase` spine, with a parallel `**leadinstrument` spine for lead-instrument annotations.

Transcribers worked in music notation software of their choice (e.g., Musescore, Sibelius) transcribing pitch, rhythm, and lyrics. The transcription was then exported into musicXML format. We wrote a Haskell program to parse musicXML scores into humdrum notation (`**kern` for pitch/rhythm, and `**silbe` for lyrics), and align this information with the already generate humdrum data described in the previous paragraphs. When transcribers included more than one vocal part for a song, each part appears as a separate pair of spines(`**kern` and `**silbe`) in the humdrum file.

5.2 Rolling Stone Data

The RS200 dataset is available at rockcorpus.midside.com, with data for each song encoded in four or five separate files—Figure 1 shows three such files. In addition, David Temperley provided us with files indicating the hierarchical structure built into their original transcriptions, which can be interpreted as formal labels. We created a Haskell program to parse these files and generate a single humdrum-syntax file for each track.[5] In some cases, we had to correct inconsistencies between harmonic and melodic transcriptions—e.g., music notated as 4/4 in the harmonic analysis but 12/8 in the melodic transcription. Each humdrum file created includes two `**harm` spines, representing Temperley and de Clercq's separate harmonic transcriptions, labeled with comment tokens '!D.T.' or '!T.d.C' respectively. The RS200's original step-sequencer-like approach to rhythm transcription is faithfully encoded using humdrum's "timebase" function where `*tb` interpretations indicate the duration of each step. For the 80-song subset with lyrics, a `**silbe` spine indicates the lyric alongside a `**stress` spine to indicate three levels of lexical/prosodic stress.

The original RS200 transcriptions indicate only tonal center (tonic), not mode, which can be ambiguous in popular music [3]. For consistency with the Billboard data, the key in each `**harm` spine is indicated as either major or minor, depending on what would be the most likely interpretation. The RS200 melodic transcriptions *do* include key-signature-like indications of raised/lowered scale degrees. Using these scale indications and the humdrumR package [16], we were able to convert the original scale-degree representation to `**kern` in the final dataset.

6. VALENCE AND AROUSAL SUBSET

In addition to the musical data itself, we gathered continuous-response ratings of perceived valence and arousal, in a 100-song subset of the Billboard data. Valence and arousal are the two core dimensions of Russel's circumplex model of affect [17], and, while perhaps limiting [18, 19], has been used widely in both music perception research [18, 20] and music emotion recognition (MER) [21–23]. We focused on valence and arousal due to their simplicity (i.e., only two variables) and ubiquity

[5] Though the music21 Python library [15] includes a parser for the RS200 harmonic transcriptions, it was easier to assure consistency and alignment between melodic, harmonic, lyrical, and formal information by using a single custom parser.

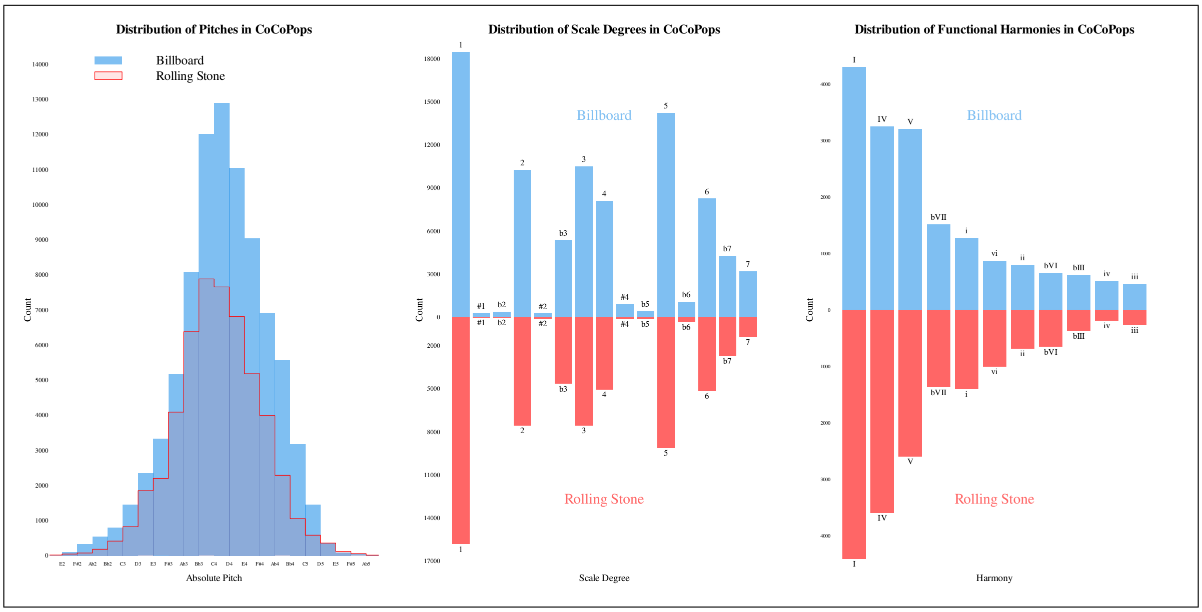


Figure 4. Left: distribution of absolute pitches in each corpus. Middle: distribution of 15 most common scale degrees in each corpus. Right: distribution of ten most common functional harmonies in each corpus (11 in total), sorted by rank in the Billboard data. (Only Temperley's harmonic annotations are counted; Immediate repetitions of a chord are not counted.)

in the literature, though it is acknowledged that there are likely additional, overlooked dimensions such as tension and power [24]. Since arousal is highly correlated with sound level, we also include the rolling RMS values for each track in an `**rms` spine.

6.1 Perceptual Data

Perceptual data was gathered in a human-subject experiment, approved by Georgia Tech's Institutional Review Board (protocol H22086). Eighty participants took part in our experiment, each paid $15 for their time. All participants were students at Georgia Tech, and were mainly non-music majors. Experiments took place in person, in a sound-attenuated booth using professional-quality loudspeakers set at the same fixed sound-level for all participants. Participants used a physical slider (Monogram Creative) to make their continuous ratings, with the slider position sampled every 500_{ms}.

The concepts of valence and arousal were explained to each participant in simple terms: arousal being how calm-energetic they perceived the music to be at any given moment, and valence being the polarity (negative-positive) of the music [25]. Participants were instructed to rate what they perceived the music to express, not necessarily what they themselves felt. Since continuously tracking valence and arousal simultaneously is challenging, we had participants rate each independently—the same approach taken for the DEAM dataset [11]. The authors of the DEAM project also reported an increase in the usability (i.e. variation) [11] of the ratings when they used full songs as opposed to shorter clips; Accordingly, participants in our experiment listened to the full songs. To encourage sustained engagement and attention throughout the experiment, we had each participant rate only ten songs. Participants were randomly assigned to rate valence in five songs and arousal in the other five, with the order of tasks counterbalanced. Ultimately, each of the 100 songs was independently rated for valence and arousal by eight different participants (four for valence and four for arousal). The full experiment took approximately forty minutes.

Files in the 100-song subset include independent `**valence`, `**arousal`, and `**rms` spines. The four independent arousal and valence ratings are encoded in the same spine, in space-separated humdrum sub-tokens.

7. SUMMARY

The CoCoPops corpus includes complete melodic and harmonic data for 398 unique popular songs released between 1949 and 2002. 95% of songs (379) come from the years 1956–1991 with more than half (203) from the years 1965–1980. The corpus includes 145,822 note onsets (86,215 in the Billboard subset), 37,010 chord changes (19,682 in the Billboard subset), and 63,809 words in the lyrics (48,018 in the Billboard subset). Figure 4 shows the distributions, in each subcorpus, of three fundamental pitch parameters—absolute pitch height, scale degree, and the ten most frequent Roman numerals. Though the two subcorpora originate in data generated by different sampling criteria and different measurement/encoding procedures (see Section 5), these distributions are nonetheless broadly similar, highlighting the potential value of treating these two separate subcorpora as a single united corpus.

The CoCoPops dataset is hosted at github.com/Computational-Cognitive-Musicology-Lab/CoCoPops, shared under a CC-BY-4.0 license. Many further methodological and encoding details are included in the repository files, as well as our recommendations about the usage, distribution, and citation of the data.

8. REFERENCES

[1] J. A. Burgoyne, J. Wild, and I. Fujinaga, "An expert ground truth set for audio chord recognition and music analysis," in *Proceedings of the International Society for Music Information Retrieval*, A. Klapuri and C. Leider, Eds., Miami, FL, 2011, pp. 633–638.

[2] C. Harte, "Towards automatic extraction of harmony information from music signals," Doctor of Philosophy, University of London, Department of Electronic Engineering, Queen Mary, University of London, 2010.

[3] T. De Clercq and D. Temperley, "A corpus analysis of rock harmony," *Popular Music*, vol. 30, no. 1, pp. 47–70, 2011, publisher: Cambridge University Press. [Online]. Available: https://www.cambridge.org/core/journals/popular-music/article/abs/corpus-analysis-of-rock-harmony/C5210A8EC985DDF170B53124F4464DA4

[4] D. Temperley and T. d. Clercq, "Statistical analysis of harmony and melody in rock music," *Journal of New Music Research*, vol. 42, no. 3, pp. 187–204, 2013. [Online]. Available: https://doi.org/10.1080/09298215.2013.788039

[5] D. Huron, "On the virtuous and the vexatious in an age of big data," *Music Perception: An Interdisciplinary Journal*, vol. 31, no. 1, pp. 4–9, 2013. [Online]. Available: https://www.jstor.org/stable/10.1525/mp.2013.31.1.4

[6] B. Clark and C. Arthur, "Is melody 'dead'?: A large-scale analysis of pop music melodies from 1960 through 2019," *Empirical Musicology Review*, In Press.

[7] G. Micchi, M. Gotham, and M. Giraud, "Not all roads lead to rome: Pitch representation and model architecture for automatic harmonic Analysis," *Transactions of the International Society for Music Information Retrieval (TISMIR)*, vol. 3, no. 1, p. 42, 2020. [Online]. Available: https://hal.science/hal-02934374

[8] C. Harte, M. B. Sandler, S. A. Abdallah, and E. Gómez, "Symbolic representation of musical chords: A proposed syntax for text annotations." in *Proceedings of the International Society for Music Information Retrieval*, 2005, p. 66–71.

[9] C. W. White, J. Fulmer, B. Cordova, A. Black, C. Danitz, W. Evans, A. Fischer, R. Greene, J. He, E. Kenyon, J. Miller, M. Moylan, A. Ring, E. Schwitzgebel, and Y. Wang, "A new corpus of texture, timbre, and change in 20th-century American popular music," in *Future Directions of Music Cognition*. The Ohio State University Libraries, 2021. [Online]. Available: https://kb.osu.edu/handle/1811/93133

[10] C. W. White, J. Pater, and M. Breen, "A comparative analysis of melodic rhythm in two corpora of American popular music," *Journal of Mathematics and Music*, vol. 16, no. 2, pp. 160–182, 2022. [Online]. Available: https://doi.org/10.1080/17459737.2022.2075946

[11] A. Aljanaki, Y.-H. Yang, and M. Soleymani, "Developing a benchmark for emotional analysis of music," *PLOS ONE*, vol. 12, no. 3, p. e0173392, 2017. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0173392

[12] C. S. Sapp, "Online database of scores in the humdrum file format." in *Proceedings of the International Society for Music Information Retrieval*, 2005, p. 664–665.

[13] C. Arthur, F. Lehman, and J. McNamara, "Presenting the SWTC: A symbolic corpus of themes from John Williams' *Star Wars* episodes I–IX," *Empirical Musicology Review*, In Press.

[14] N. Condit-Schultz, "MCFlow: A digital corpus of rap transcriptions," *Empirical Musicology Review*, vol. 11, no. 2, p. 124–147, 2016.

[15] M. S. Cuthbert and C. Ariza, "Music21: A toolkit for computer-aided musicology and symbolic music data." in *Proceedings of the International Society for Music Information Retrieval*, 2010, p. 9–13.

[16] N. Condit-Schultz and C. Arthur, "humdrumR: a new take on an old approach to computational musicology," in *Proceedings of the International Society for Music Information Retrieval*, 2019, p. 715–722.

[17] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, place: US Publisher: American Psychological Association.

[18] A. Micallef Grimaud and T. Eerola, "An interactive approach to emotional expression through musical cues," *Music & Science*, vol. 5, p. 20592043211061745, 2022, publisher: SAGE Publications Ltd. [Online]. Available: https://doi.org/10.1177/20592043211061745

[19] G. L. Collier, "Beyond valence and activity in the emotional connotations of music," *Psychology of Music*, vol. 35, no. 1, pp. 110–131, 2007.

[20] T. Eerola and J. K. Vuoskoski, "A review of music and emotion studies: Approaches, emotion models, and stimuli," *Music Perception: An Interdisciplinary Journal*, vol. 30, no. 3, pp. 307–340, 2013. [Online]. Available: http://mp.ucpress.edu/cgi/doi/10.1525/mp.2012.30.3.307

[21] Y. Agrawal, R. G. R. Shanker, and V. Alluri, "Transformer-based approach towards music emotion recognition from lyrics," in *Advances in Information Retrieval*, ser. Lecture Notes in Computer Science, D. Hiemstra, M.-F. Moens, J. Mothe, R. Perego,

M. Potthast, and F. Sebastiani, Eds. Cham: Springer International Publishing, 2021, pp. 167–175.

[22] J. de Berardinis, A. Cangelosi, and E. Coutinho, "The multiple voices of musical emotions: source separation for improving music emotion recognition models and their interpretability," in *Proceedings of the 21st International Society for Music Information Retrieval Conference*, Online, 2020, pp. 310–317. [Online]. Available: https://livrepository.liverpool.ac.uk/3105085

[23] D. Bogdanov, X. Lizarraga Seijas, P. Alonso-Jiménez, and X. Serra, "MusAV: A dataset of relative arousal-valence annotations for validation of audio models," in *Proceedings of the International Society for Music Information Retrieval*. International Society for Music Information Retrieval (ISMIR), 2022, pp. 650–658.

[24] J. Cespedes-Guevara and T. Eerola, "Music communicates affects, not basic emotions – A constructionist account of attribution of emotional meanings to music," *Frontiers in Psychology*, vol. 9, 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00215/full

[25] K. N. Olsen, R. T. Dean, and C. J. Stevens, "A Continuous Measure of Musical Engagement Contributes to Prediction of Perceived Arousal and Valence," *Psychomusicology: Music, Mind, and Brain*, vol. 24, no. 2, pp. 147–156, 2014.