

Data conventions

This document is intended to gather all relevant encoding conventions used in the creation of the dataset.

Spreadsheet

Columns:

1. SongID
2. Title
3. Artist
4. Key
5. Genre
6. Tempo
7. Time Signature

MIDI files (Filenames: `songid_title.midi`, for example `075_diva.midi`)

Track 1: Melody

Track 2: Chords (root position)

Track 3: Bass

Track 4: Sections

Track 5: Instrumental (Not available for every song)

Section labels on track 4:

- Verse - E
- Chorus - C
- Bridge - G
- Pre-chorus - D
- Hook / instrumental - B
- Intro - A

Notes should be on the first beat of the bar. The first note might be in bar 1 or bar 2, if there is an upbeat in the melody.

Accidentals are irrelevant in form markings. $C\# = C$, $Ab/A\# = A$, etc.

The Midi files contain only the skeleton parts of songs. If a part repeats twice it will appear only once in the MIDI file because there is no new information to use. The midi files do not stand as full or complete songs when played back, but just a sequence of the significant parts of any song.

Shorter midi files might indicate to songs that had only a couple of musical parts repeating, or a song with shorter parts (short verse, short chorus and so).

Hook/Instrumental - is a combined category for either:

Hook: will either be Instrumental or vocal but with no full phrase lyrics- either a single syllable or word etc.

Instrumental: any instrumental part that is significant to the musical DNA of the song, usually in between two lyric parts (Chorus to Verse or so).

These two are combined as in the general scheme they both bare a similar significance to a song, as a non-lyrical part that usually connects between two lyrical parts, and serve as a compositional addition to a song.