Universiteit Utrecht

chordify

NWO
Netherlands Organisation for Scientific Research

MuSiVa
Modelling musical similarity over time through the variation principle

Information Retrieval · Musicology · Cognitive Science · MUSIVA

# Chordify Annotator Subjectivity Dataset

Hendrik Vincent Koops*, W. Bas de Haas#, John Ashley Burgoyne§, Jeroen Bransen#, Anja Volk*
* Department of Information and Computing Sciences, Utrecht University
§ Music Cognition Group, University of Amsterdam
# Chordify

## Motivation

- Reference annotation datasets containing single harmony annotations are at the core of a wide range of studies in MIR and related fields. However, annotator subjectivity is (usually) not taken into account.

- Currently available chord-label annotation datasets containing more than one reference annotation are limited by size, sampling strategy, or lack of a standardized encoding.

## Dataset

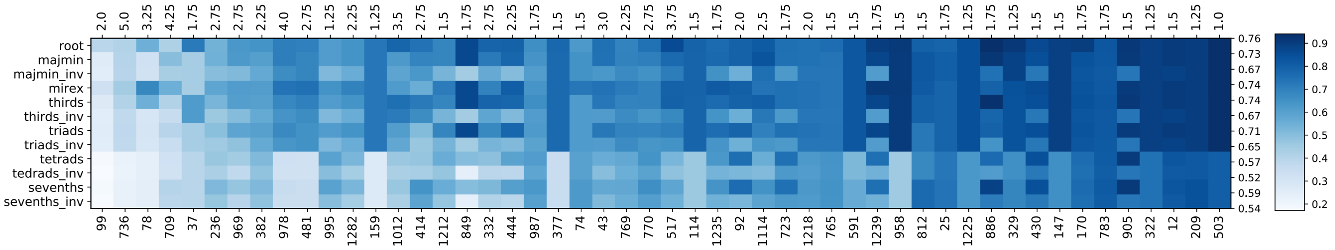
- **Fifty songs** from the *Billboard* dataset
  - Having a stable on-line presence in widely accessible music repositories
- Annotated by **four** *expert* annotators
  - Chord labels encoded in standard *Harte et al.* syntax
  - Annotations encoded in *JAMS* format

www.github.com/chordify/CASD

## Analysis

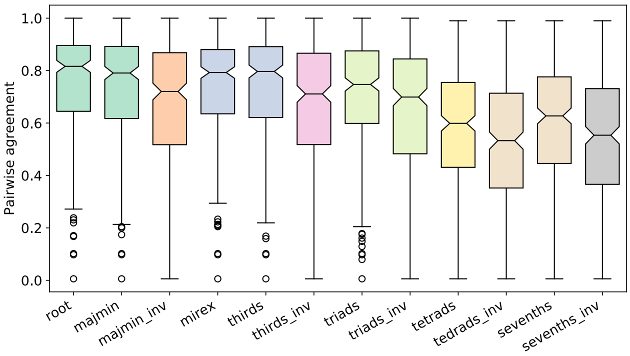
We find **low agreement** between annotators. Only ±80% root note overlap, and only ±54% full chord overlap

Within this dataset, significant **differences** exist between annotators, in **chord labels** as well as in perceived **difficulty** and **annotation times**. These results show that annotator subjectivity is an important factor in harmonic transcriptions, which should be taken into account in **future automatic chord estimation** and related computational harmonic research.

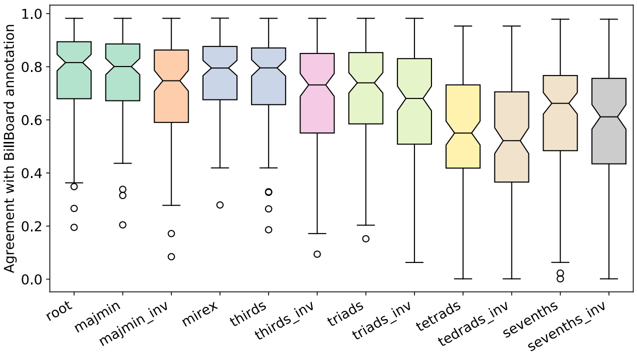


Krippendorff's-α inter-rater agreement of all songs in the dataset. Agreement **decreases** with **increased chord granularity.** The level of **agreement decreases significantly** when **inversions** are taken into account. Average reported difficulties can be found above the columns. The numbers on the right show the average agreement for each chord granularity level.

A factor analysis shows that each annotator is unique. Differences in factor variance, along with the factor structure itself, suggest that the core of annotator subjectivity lies in the relative importance of triads, sevenths, inversions, and other musical factors for each annotator.



Pairwise **agreement among four annotators** for all MIREX chord granularity levels. Agreement is significantly lower with inversions (⋆ vs ⋆_inv) with ($p \ll 0.001$).



**Agreement** of the four annotators **with the *Billboard*** annotations. Agreement is significantly lower with inversions (⋆ vs ⋆_inv) with ($p \ll 0.001$).

These results show that annotators do not significantly agree more with a *Billboard* annotation than with the annotations from the other three annotators. Compared to this dataset, the *Billboard* annotations appear to be another expert opinion.

{h.v.koops}{a.volk}@uu.nl, {bas}{jeroen}@chordify.net, j.a.burgoyne@uva.nl

MuSiVa is supported by NWO-VIDI grant 276-35-001
Contact: h.v.koops@uu.nl