CHORDIFY ANNOTATOR SUBJECTIVITY DATASET

Hendrik Vincent Koops[1] **W. Bas de Haas**[2] **Jeroen Bransen**[2]
John Ashley Burgoyne[3] **Anja Volk**[1]
[1] Department of Information and Computing Sciences, Utrecht University, the Netherlands
[2] Chordify, Utrecht, the Netherlands
[3] Music Cognition Group, University of Amsterdam, the Netherlands
{h.v.koops}{a.volk}@uu.nl, {bas}{jeroen}@chordify.net, j.a.burgoyne@uva.nl

ABSTRACT

Reference annotation datasets containing harmony annotations are at the core of a wide range of studies in music information retrieval and related fields. The majority of these datasets contain single reference annotations describing the harmony of each piece or song. Nevertheless, music theoretical insights on harmonic ambiguity and studies showing differences among annotators in many MIR tasks make the notion of a single "ground-truth" reference annotation a tenuous one. In order to gain a better understanding of differences between harmony annotators, we introduce the *Chordify Annotator Subjectivity Dataset* (CASD) containing chord labels for fifty songs from four annotators.

1. CHORDIFY ANNOTATOR SUBJECTIVITY DATASET

The *Chordify Annotator Subjectivity Dataset* (CASD) [1] contains chord labels for 50 songs from 4 annotators.

2. SONG SELECTION

Currently available chord-label annotation datasets containing more than one reference annotation are limited by size, sampling strategy, or lack of a standardized encoding [2, 5]. To account for these potential problems in our own dataset, we chose to select fifty songs from the *Billboard* dataset [1] that have a stable online presence in widely accessible music repositories. This way, listening to the songs is easy, stimulating future research with the dataset. After searching the YouTube website for the title and artist tags of the *Billboard* dataset, we ranked the results of each query by number of views and selected the top fifty songs by this ranking. At the time they were collected, the least-viewed song in the dataset had 67 thousand views and the most-viewed song over 13 million, and an average of 11.9 unique chords according to the *Billboard* dataset annotations.

3. ANNOTATOR SELECTION

To study annotator subjectivity and account for a potential instrument bias, we recruited four annotators: two guitarists and two pianists. All annotators had either studied composition or music performance at the undergraduate or graduate level. All annotators were also successful professional music performers, with between 15 and 20 years of expertise in playing their primary instrument. Two of the annotators further identified themselves as composers. We reviewed the first ten transcriptions from each annotator to ensure the annotators had sufficient aptitude to continue; all four annotators completed the initial screening successfully and were hired to continue to annotate the remaining forty songs. The annotators were compensated financially for their annotations at a fixed rate per song.

4. TRANSCRIPTION PROCESS

To ensure the annotators were all focused on the same task, we provided them with a guideline for the annotating process. We asked them to listen to the songs as if they wanted to play the song on their instrument in a band, and to transcribe the chords with this purpose in mind. They were instructed to assume that the band would have a rhythm section (drum and bass) and melody (e.g., a singer). Therefore, their goal was to transcribe the complete harmony of the song in a way that, in their view, best matched their instrument.

To provide the annotators with a central, unified transcription method, we provided the annotators with an instance of the Chordify chord editing interface [2] that was specifically adapted for the creation of the CASD . This interface provided the annotators with a grid of beat-aligned elements, which we manually verified for correctness. Chord labels could be chosen for each beat. The standard YouTube web player was used to provide the reference recording of the song. Through the interface, the annotators were free to select any chord of their choice for

[1] https://www.github.com/chordify/CASD

[2] see https://chordify.net

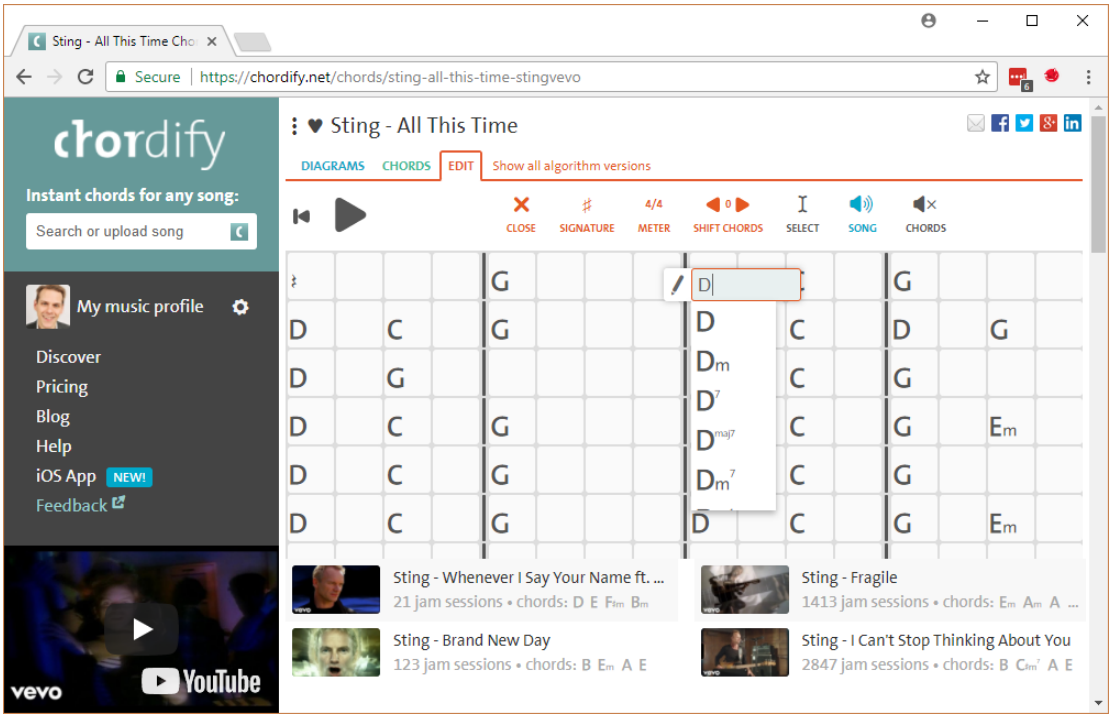

Figure 1. The Chordify chord editing interface

each beat. While transcribing, the annotators were able to watch and listen not only to the YouTube video of the song, but also a synthesized version of their chord transcription.

In addition to providing chords and information about their musical background, we asked the annotators to provide for each song a difficulty rating on a scale of 1 (easy) to 5 (hard), the amount of time it took them to annotate the song in minutes, and any remarks they might have on the transcription process.

5. DATASET TECHNICAL SPECIFICATIONS

To provide the MIR research community with a dataset that is easily accessible and expandable, that encourages reproducibility and stimulates future research into annotator subjectivity, we adopted a number of standard encodings that are commonly used in MIR research.

For each of the fifty songs, the dataset contains chord labels provided by four annotators. These chord labels are encoded using the chord-label syntax introduced by Harte et al. [3]. This syntax provides a simple and intuitive encoding that is highly structured and unambiguous to parse with computational means. To promote and stimulate future research, we include identifiers for music repositories (e.g., YouTube), allowing researchers to listen to the tracks easily. Furthermore, we provide *Billboard* dataset identifiers which make it possible to cross-reference our dataset with data from the *Billboard* dataset, ACE output from the MIREX task, and other datasets that use these identifiers.

The complete dataset is encoded using the JAMS format: a JSON-annotated music specification for reproducible MIR research, which was introduced by [4]. JAMS provides an interface with the standard MIREX evaluation measures, making it very easy to evaluate and compare annotations. To provide easy access, we make the dataset publicly available in a Git repository. By way of Git and JAMS, we encourage the MIR community to exchange, update, and expand the dataset.

6. REFERENCES

[1] J.A. Burgoyne, J. Wild, and I. Fujinaga. An expert ground truth set for audio chord recognition and music analysis. In *Proc. of the 12th International Society for Music Information Retrieval Conference, ISMIR*, volume 11, pages 633–638, 2011.

[2] T. De Clercq and D. Temperley. A corpus analysis of rock harmony. *Popular Music*, 30(01):47–70, 2011.

[3] C. Harte, M.B. Sandler, S.A. Abdallah, and E. Gómez. Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proc. of the 6th International Society for Music Information Retrieval Conference, ISMIR*, volume 5, pages 66–71, 2005.

[4] E.J. Humphrey, J. Salamon, O. Nieto, J. Forsyth, R.M. Bittner, and J.P. Bello. JAMS: A JSON annotated music specification for reproducible MIR research. In *Proc. of the 15th International Society for Music Information Retrieval Conference, ISMIR*, pages 591–596, 2014.

[5] Y. Ni, M. McVicar, R. Santos-Rodriguez, and T. De Bie. Understanding effects of subjectivity in measuring chord estimation accuracy. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(12):2607–2615, 2013.