


FiloBass: A Dataset and Corpus Based Study of Jazz Basslines


Xavier Riley
C4DM
j.x.riley@qmul.ac.uk


Simon Dixon
C4DM
s.e.dixon@qmul.ac.uk




Highlights

 48 scores, audio stems and annotations focusing on jazz double bass.

 50k+ annotated note events based on backing tracks from the FiloSax dataset (Foster 2021).

 In-depth musicological analysis of bass lines, derived from data.

 Code released as Jupyter notebooks

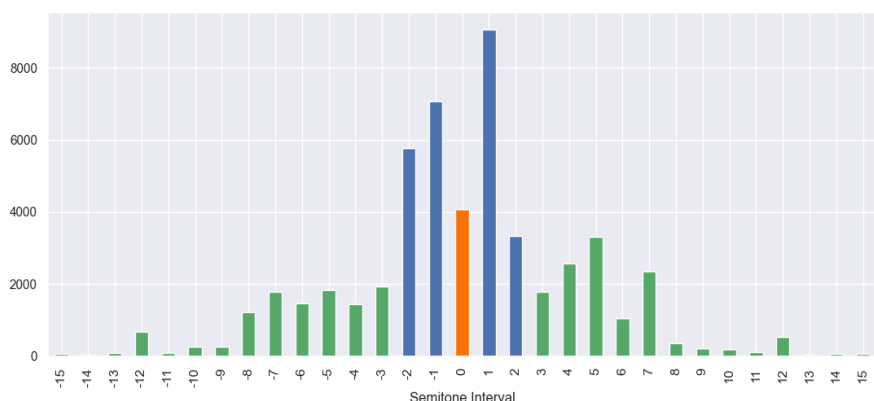


FIG 1. Analysis of Interval Distribution in dataset

Methodology

- Dataset uses professionally produced jazz backing tracks from the Aebersold series. Each piece (a jazz standard) was:
 - Fully transcribed using a manually verified pipeline
 - Fully annotated with downbeats, chords and performance timings
 - Written as a MusicXML score, reviewed for accuracy by a jazz bassist

Results

- Analysis includes chord degrees, rhythmic variety, common n-grams, intervals used to approach chord resolutions and melodic contour analysis
- Baseline results for automatic note transcription (a bassline baseline):

	CREPE Notes	Basic Pitch	Melodyne
Recall _{no}	74.11 \pm 12.09	81.28 \pm 6.26	79.52 \pm 14.77
Precision _{no}	71.81 \pm 13.33	51.40 \pm 6.28	78.48 \pm 15.41
F-measure _{no}	72.89 \pm 12.68	62.73 \pm 4.51	78.95 \pm 15.02



Fig 3. Common patterns

Conclusions

- This dataset allows for a new level of detailed analysis regarding the role of the double bass in jazz.
- Data Driven Rules for basslines: play the roots on new chords; approach them via a semitone steps; 66% bars play straight four quarters; maintain a balance between ascending and descending line contours
- Possible downstream tasks: musicology, automatic transcription, chord recognition, beat tracking, expressive performance studies and more

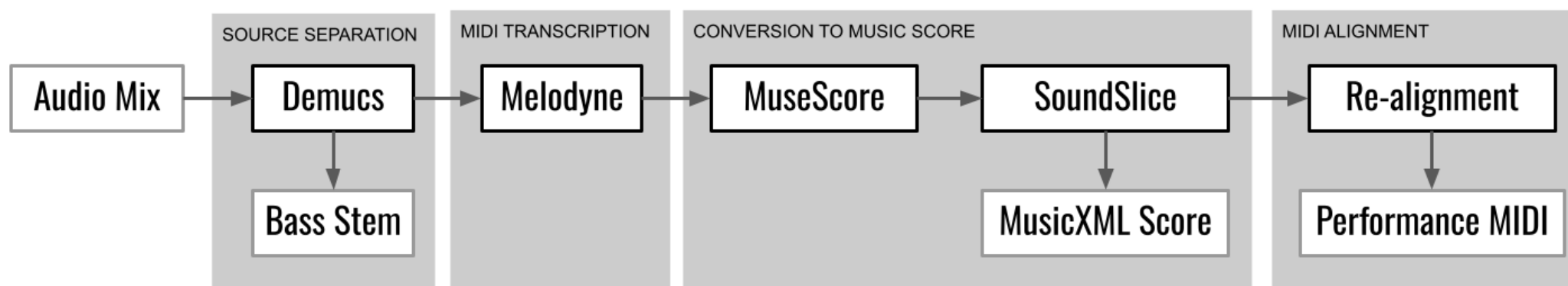


FIG 2. Workflow for transcription process

This work was supported by the UKRI and EPSRC under grant EP/S022694/1

