The Jazz Ontology: Organising Metadata in Dig That Lick

Simon Dixon[1], Polina Proutskova[1], Tillman Weyde[2], Daniel Wolff[2], Martin Pfleiderer[3], Klaus Frieler[3], Frank Höger[3], Hélène-Camille Crayencour[4], Jordan Smith[1,4], Geoffroy Peeters[5], Doğaç Başaran[6], Gabriel Solis[7], Lucas Henry[7], Krin Gabbard[8], Andrew Vogel[8]

(1) Queen Mary University of London; (2) City, University of London; (3) University of Music Weimar; (4) CNRS, IRCAM Lab, Sorbonne Université; (5) Telecom ParisTech; (6) Audible Magic; (7) University of Illinois; (8) Columbia University

Linked Music Workshop, McGill University, 19 November, 2022

The *Jazzomat* Project (2012-2017)

- Background to the *Dig That Lick* project
- Martin Pfleiderer, Klaus Frieler, Jakob Abeßer at the University of Music Franz Liszt Weimar
- Investigation of jazz improvisation
- Produced the **Weimar Jazz Database**:
  - 456 transcriptions of monophonic solos from well-known musicians
  - Aligned to audio, chords, beats
  - Released as an SQL database

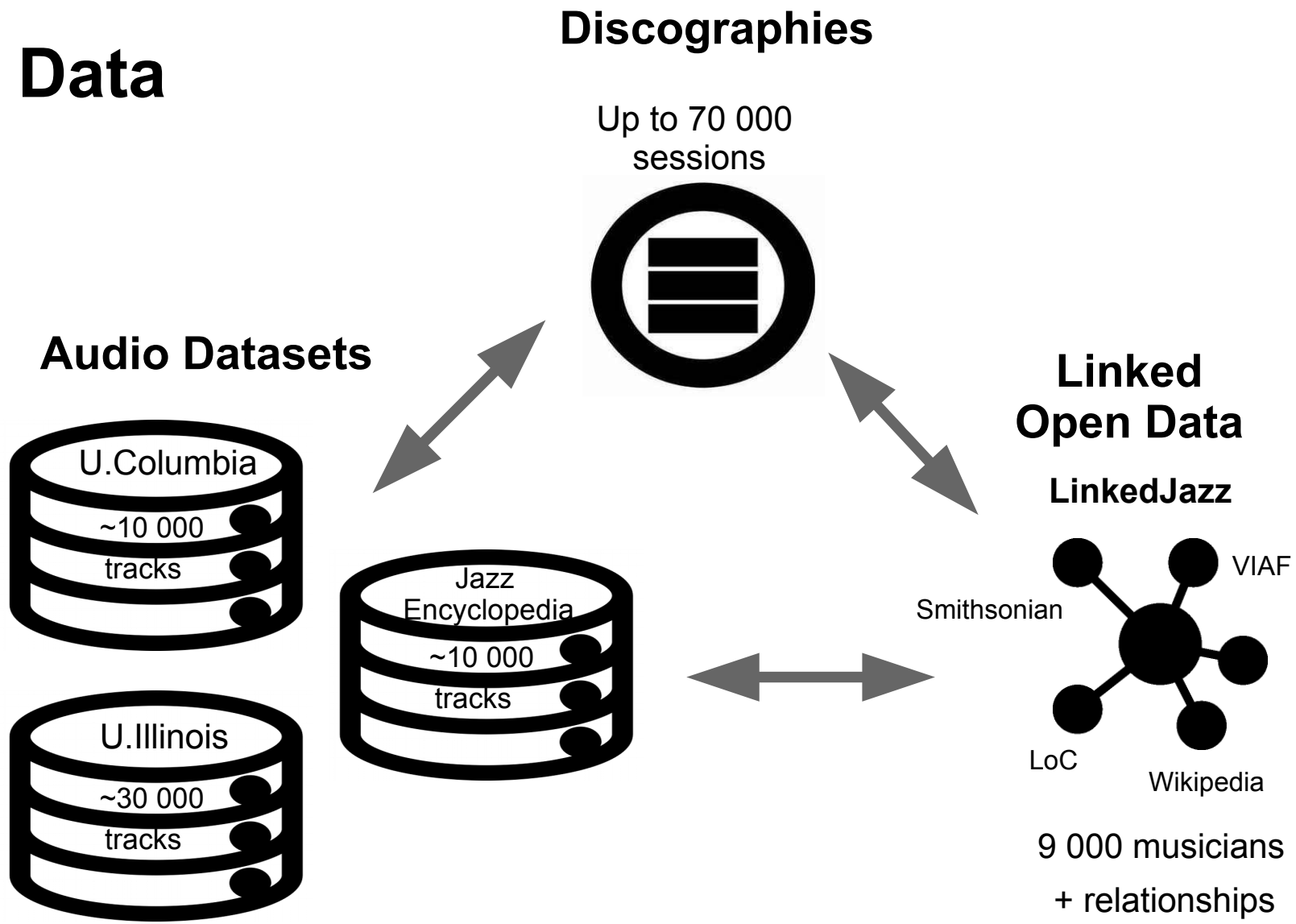
WJD Database Structure

Table name	Description
`beats`	Table for beat annotation of WJD melodies, referenced by `melody(id)`
`composition_info`	Info re composition, refd by `melody(id)`
`db_info`	Info re distributed database file (version information, license, etc.)
`esac_info`	EsAC info (if EsAC), refd by `melody(id)`
`melody`	Main table for all melody events
`melody_type`	Type of melody: WJD solo or EsAC (Essen Folk songs), refd by `melody(id)`
`popsong_info`	Pop song info, refd by `melody(id)`
`record_info`	Info re audio source, refd by `melody(id)`
`sections`	All sections (phrase, chorus, form, chords, etc.), refd by `melody(id)`
`solo_info`	Solo info refd by `melody(id)`
`track_info`	Info specific to a track on a CD
`transcription_info`	Transcription info refd by `melody(id)`

The *Dig that Lick* Project (2017-2019)

- Full title: *Dig that lick: Analysing large-scale data for melodic patterns in jazz performances*
- **Infrastructures** for large-scale semantic audio analysis
- **Interfaces** for selection, analysis, and aggregation
- **Analysis** of melodic patterns in a large jazz corpus
- **Link** to metadata
- **Interpret** evidence of musical influence

Data: Audio and Metadata

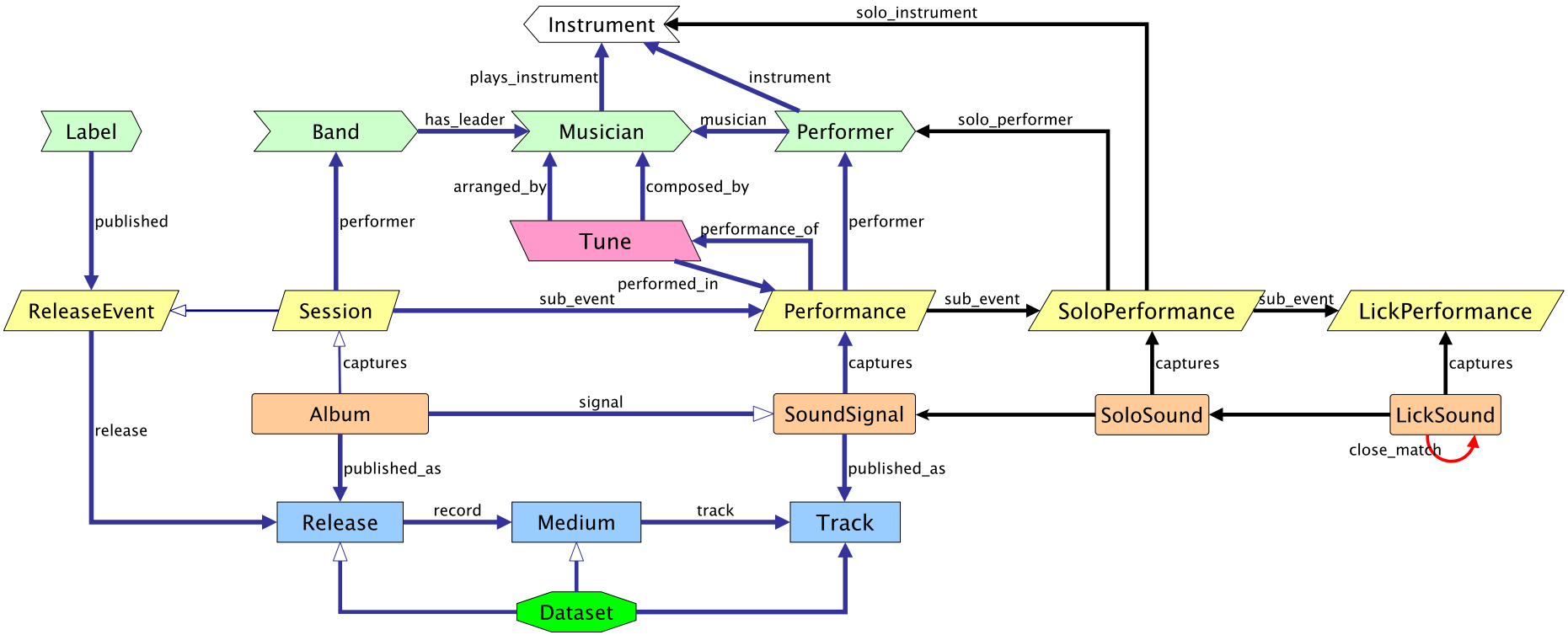


The Metadata Bottleneck

- Who is soloing now?
- Audio obtained from music libraries included only: Label, Catalog number, Artist, CD title, Track title
- Semiautomatic matching to Lord's Jazz Discography, MusicBrainz and Discogs provided much richer data (full lineup, dates, places)
- Jazz Encyclopedia had all this, already digitised in a CSV file (!)
- For high quality, human effort is hard to avoid
  - Named entity resolution
  - Disambiguation
  - Reconciliation

The Jazz Ontology

[Proutskova et al., 2022] The Jazz Ontology: A semantic model and large-scale RDF repositories for jazz. *J. Web Semantics*, 74:100735.



The *DTL1000* Dataset

- 1000 tracks selected randomly from jazz collections (100 per decade from 1920-2019)
- Note tracks automatically extracted from monophonic solos
- 1700 solos, 6M pitch n-gram instances, 5.6M interval n-grams
- Metadata expressed in RDF using a bespoke ontology and accessed via SPARQL requests
- Metadata used to filter searches and shown in results
- Similarity search combines DTL1000 with the Weimar Jazz Database, Charlie Parker Omnibook and Essen Folk Song Collection
- Prereleased on `https://osf.io/buxvr/`

Pattern Search: List Results

Dig That Lick

Switch to Pattern search

New search | All searches | Documentation | Help | About | Print
Login

Pattern Similarity Search

Similarity search

Pattern: -1,-2,-1,3,3,3,-1,-2

Transformation: Semitone Intervals

Pin pattern elements

First Last

Search

Options

Minimum similarity (80%)

65 70 75 80 85 90 95 100

Maximum length difference: 2

Maximum edit distance: 1

Minimum frequency: 2

Keep overlapping instances

Within single phrase

Preserve contour (ascending)

Preserve pitch range (9)

Databases

Dig That Lick — Metadata filter

Weimar Jazz Database — Metadata filter

Charlie Parker Omnibook

EsAC Folksong Database

Show columns

Raw frequency Pitch range Contour Start position Duration

Group by

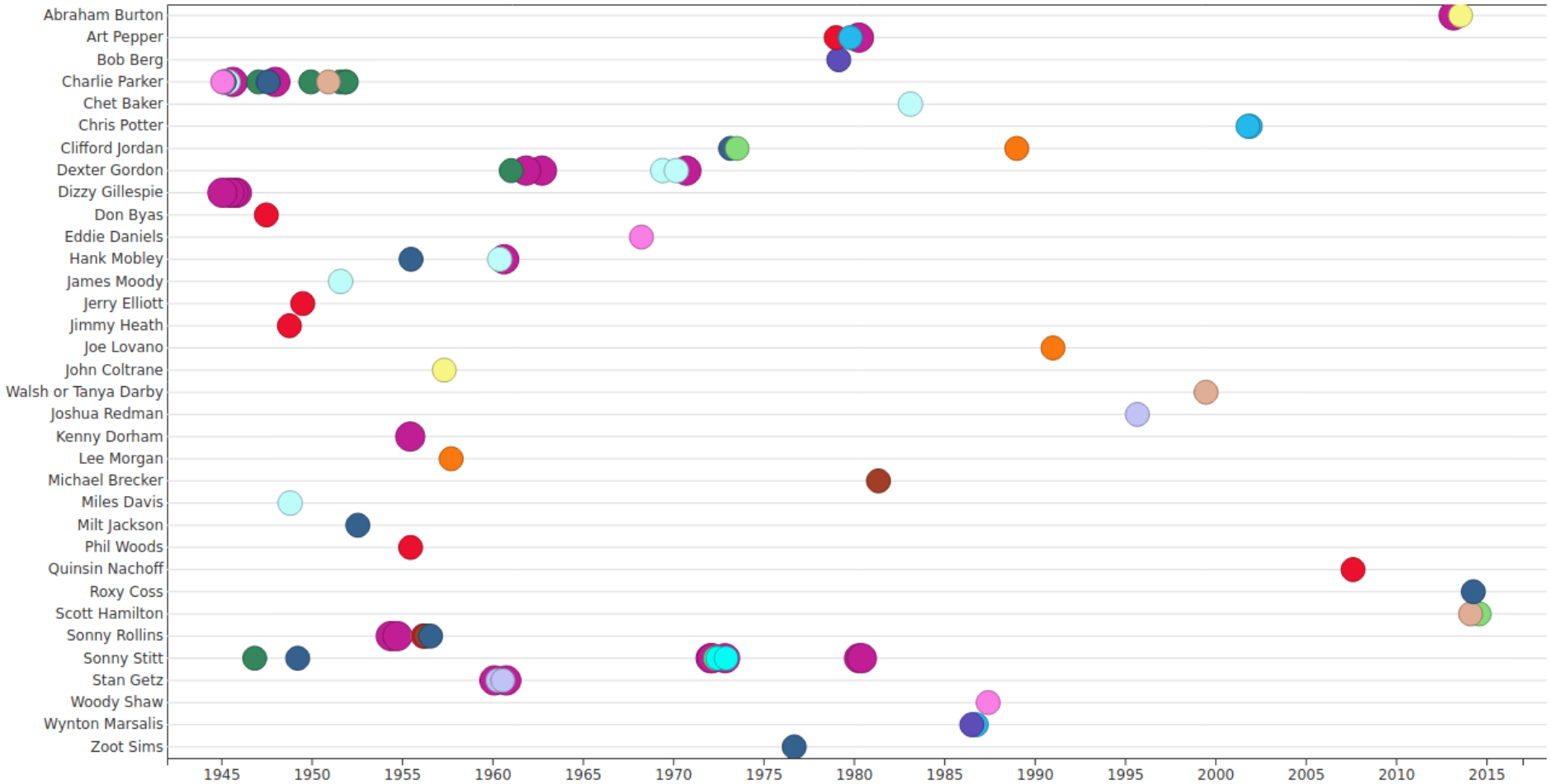
Pattern Performer

Found 82 similar (15 unique) pattern instances:

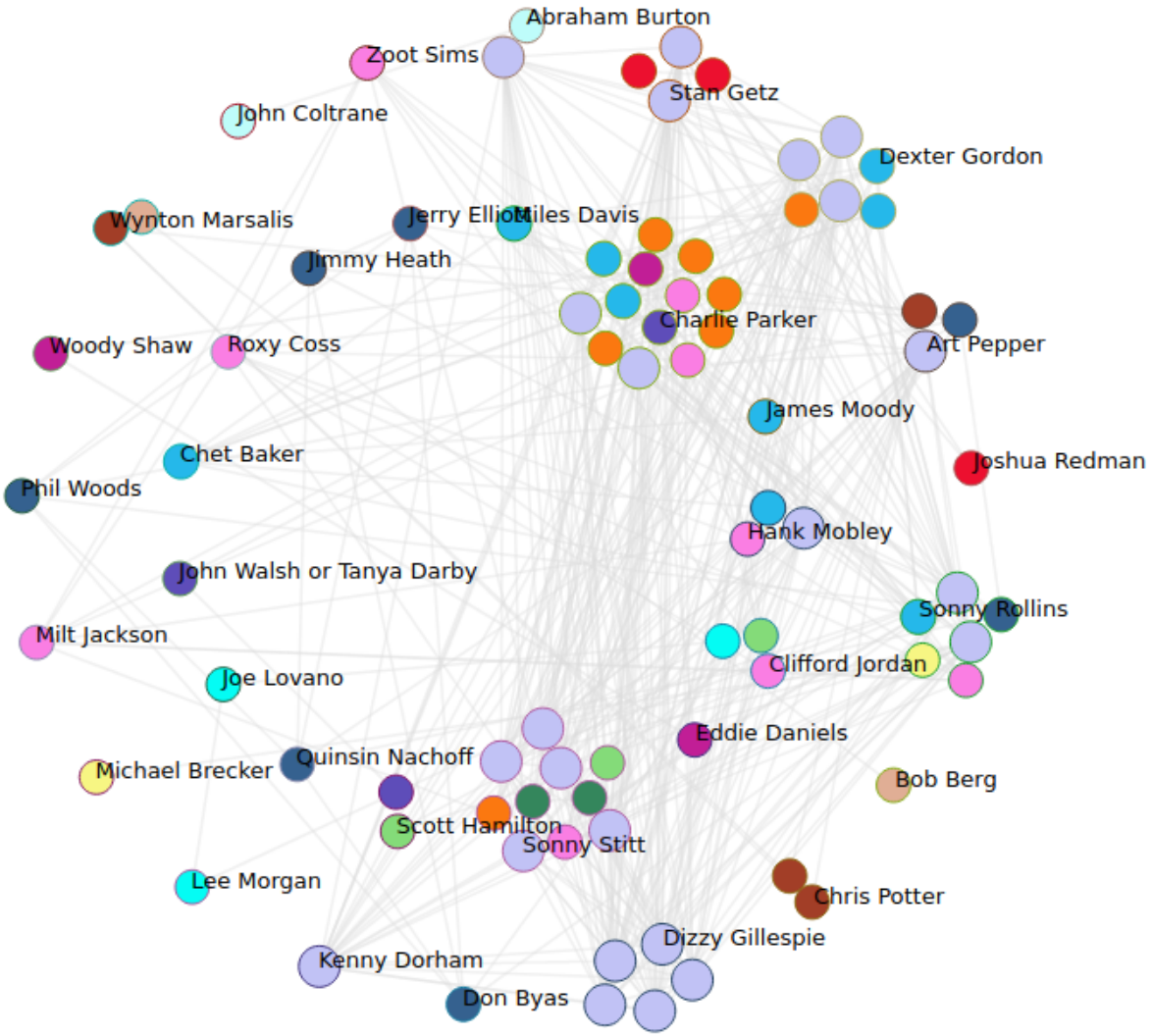
(44) (38)

#	Pattern	Performer	Title	Recording year	Instrument	Style	Similarity	Edit distance
1	-1,-2,-1,3,3,3,-1,-2 (8)							
		Abraham Burton	Without a song	2013	Tenor saxophone	Hardbop	1.00	0
		Art Pepper	How high the moon	1980	Alto saxophone	Cool	1.00	0
		Charlie Parker	Donna Lee	1947	Alto saxophone	Bebop	1.00	0
		Charlie Parker	Ko-Ko	1945	Alto saxophone	Bebop	1.00	0
		Dexter Gordon	Cheese Cake	1962	Tenor saxophone	Hardbop	1.00	0
		Dexter Gordon	Society Red	1961	Tenor saxophone	Hardbop	1.00	0

Pattern Similarity Search: Timeline Results



Pattern Similarity Search: Network Results



Publications and Presentations

Başaran, D., Essid, S., and Peeters, G. (2018). Main melody estimation with source-filter NMF and CRNN. In *19th International Society for Music Information Retrieval Conference*, pages 82–89.

Frieler, K. (2019). Constructing jazz lines: Taxonomy, vocabulary, grammar. In M. Pfleiderer, W.-G. Z., editor, *Jazzforschung heute: Themen, Methoden, Perspektiven*, pages 103–132. Edition EMVAS, Berlin.

Frieler, K., Başaran, D., Höger, F., Crayencour, H.-C., Peeters, G., and Dixon, S. (2019a). Don't hide in the frames: Note- and pattern-based evaluation of automated melody extraction algorithms. In *6th International Conference on Digital Libraries for Musicology*, pages 25–32.

Frieler, K., Höger, F., and Pfleiderer, M. (2019b). Anatomy of a lick: Structure and variants, history and transmission. In *Book of Abstracts of the Digital Humanities Conference*.

Frieler, K., Höger, F., and Pfleiderer, M. (2019c). Towards a history of melodic patterns in jazz performance. In *6th Rhythm Changes Conference*.

Frieler, K., Höger, F., Pfleiderer, M., and Dixon, S. (2018). Two web applications for exploring melodic patterns in jazz solos. In *19th International Society for Music Information Retrieval Conference*, pages 777–783.

Gabbard, K. (2019). What we are digging out of the data? In *6th Rhythm Changes Conference*.

Henry, L., Frieler, K., Solis, G., Pfleiderer, M., Dixon, S., Höger, F., Weyde, T., and Crayencour, H.-C. (2023). Dig that lick: Exploring patterns in jazz with computational methods. *Jazzforschung / Jazz Research*, 50/51:to appear.

Höger, F., Frieler, K., Pfleiderer, M., and Dixon, S. (2019). Dig that lick: Exploring melodic patterns in jazz improvisation. In *20th International Society for Music Information Retrieval Conference: Late Breaking Demo*.

Pfleiderer, M., Frieler, K., Abeßer, J., Zaddach, W.-G., and Burkhart, B., editors (2017). *Inside the Jazzomat: New Perspectives for Jazz Research*. Schott Campus.

Proutskova, P., Wolff, D., Fazekas, G., Frieler, K., Höger, F., Velichkina, O., Solis, G., Weyde, T., Pfleiderer, M., Crayencour, H. C., Peeters, G., and Dixon, S. (2022). The Jazz Ontology: A semantic model and large-scale RDF repositories for jazz. *Journal of Web Semantics*, 74:100735.

Solis, G. and Henry, L. (2019). Chasing the Trane: Quantifying the social journey of a Coltrane solo. In *6th Rhythm Changes Conference*.

Weyde, T., Wolff, D., Dixon, S., Proutskova, P., Crayencour, H.-C., Smith, J., Peeters, G., and Başaran, D. (2019). Dig that lick: A technical primer for big data jazz studies. In *6th Rhythm Changes Conference*.

Acknowledgements

This research was funded under the Trans-Atlantic Program Digging into Data Challenge with the support of the UK Economic and Social Research Council (ES/R004005/1), the French National Research Agency (ANR-16-DATA-0005), the German Research Foundation (PF 669/9-1), and the US National Endowment for the Humanities (NEH-HJ-253587-17).