DATASET

Cambridge Jazz Trio Database: Automated Timing Annotation of Jazz Piano Trio Recordings Processed Using Audio Source Separation

Huw Cheston[†], *Joshua L. Schlichting, Ian Cross, Peter M. C. Harrison*[‡]

[†] *hwc31@cam.ac.uk* [‡] *pmch2@cam.ac.uk*

Abstract

Recent advances in automatic music transcription have facilitated the creation of large databases of improvised music forms (including jazz), where traditional notated scores are typically not available. However, most of these datasets focus only on capturing the improvisations of soloists, omitting the contributions of the accompanying members in an ensemble to a performance. We introduce the Cambridge Jazz Trio Database, a dataset of 12 hours of jazz piano trio recordings with automatically generated timing annotations for every performer (piano soloist, bass and drums accompaniment) in the ensemble. Appropriate recordings are identified by scraping user-based listening and discographic data, source separation models are applied to isolate audio for each performer in the piano trio, and timing annotations are generated by applying beat and onset detection algorithms to the separated audio sources. We conduct several analyses of the dataset, including with relation to swing and inter-performer synchronization. We anticipate the dataset will be useful in a variety of music information retrieval tasks, including performer identification and symbolic music generation. The database, including the source code and related documentation, is available at https://github.com/huwcheston/Cambridge-Jazz-Trio-Database.

Version Date: January 31, 2024
Word Count: 8,000

This is an unpublished preprint that has yet to undergo peer review.

Keywords: dataset, jazz, timing, beat tracking, onset detection

1. Introduction

Given its lack of notated scores and the freedoms afforded to its performers, improvised jazz is a musical genre that resists computational analysis. Recent advances in automatic music transcription systems, however, have enabled the creation of several large-scale databases of annotated jazz recordings. Most of these datasets focus on the improvisations of a single lead soloist within an ensemble (e.g., Pfleiderer et al., 2017). This is surprising, as jazz musicians place great importance on how the interaction between a soloist and their accompaniment contributes towards a successful performance (Monson, 1996).

In this project, our goal was to develop a database that included data extracted from every musician in an improvising jazz ensemble. This, we hoped, would facilitate the analysis of interesting group-level musical features (such as interaction and synchronization) that researchers have previously been unable to study using existing datasets. It would also provide a dataset for jazz comparable to those already in existence for other forms of improvised ensemble music, such as Cuban son and Hindustani classical (Clayton et al., 2020).

To extract data from a mixed ensemble recording, we leveraged recent developments in audio source separation and timing annotation. Using deep learning, it has become possible to separate isolated sources from an audio mixture with massively increased fidelity compared to earlier approaches. The quality of separation still depends on the instrument, however, with vocals, bass, drums, and piano separation having seen the majority of work, while the separation of brass and stringed instruments remains at an earlier stage.

As a consequence, we elected to focus on collating performances by jazz "piano trios", where a piano soloist would have improvised with bass and drums accompaniment. These instruments form the standard

"rhythm section" that has accompanied vocalists and soloists in jazz since the 1940s (Carr et al., 1988) and can be thought of as providing the crucial contexts within which the lead solos that have been the focus of previous analyses are articulated and developed. To this extent, results obtained from our database could readily be generalized beyond the piano trio.

We identified recordings for inclusion in our database by scraping user-based listening and discographic data and pulling audio from YouTube. We then used existing source separation software to extract isolated audio from each instrument, and finally applied beat and onset detection algorithms to automatically extract timing data from every source. We focussed on timing, rather than pitch or harmony, as this ensured parity between instruments – neither pitch nor harmony being relevant in drum performances, for instance.

In this paper, we describe several related datasets, discuss the curation of our database, outline the automated data extraction pipeline we developed, and explore our database by conducting several example analyses. We envisage that our database will be useful to researchers engaged in music information retrieval and other forms of empirical music research; for instance, in the development of performer identification, symbolic music generation, and beat detection algorithms.

2. Related Work

2.1 Weimar Jazz Database

The Weimar Jazz Database (WJD) contains note-for-note transcriptions of 456 improvised jazz solos (Pfleiderer et al., 2017). The note-level (pitches, onsets, offsets, intonation, and dynamics) annotations are aligned to the original audio, and the database also includes annotations of chord sequences, beat, and measure numbers. The majority of the annotations were created manually: only dynamics and intonation were measured automatically. A subsequent project has since extended the WJD to include structure and instrumentation annotations for each piece (Balke et al., 2022). While the WJD is undoubtedly a landmark in the computational analysis of jazz, it only captures the performance of soloists, not their accompaniment.

2.2 Filosax

The Filosax dataset (Foster & Dixon, 2021) is a collection of 24 hours of annotated jazz performances on the tenor saxophone. Five professional saxophonists each recorded themselves playing and improvising over 48 "standard" jazz compositions against a pre-recorded backing of piano, bass, and drums. Note-level annotations (including pitches, onsets, and offsets) were created automatically for each recording using an algorithm and through manual transcription. As with the WJD, Filosax only includes annotations for the soloist's performance, and does not attempt to characterize the playing of the accompanying musicians.

2.3 PiJAMA

PiJAMA consists of 200+ hours of solo jazz piano performances transcribed using a pitch-to-MIDI algorithm (Edwards et al., 2023). The methodology used by the PiJAMA authors has several similarities with our contribution: curated playlists of relevant music were developed by scraping discographic services and audio was downloaded from YouTube. Although comprehensive, PiJAMA only covers solo piano recordings, while the authors acknowledge that it is far more common for jazz pianists to perform with bass and drums accompaniment. Additionally, PiJAMA only includes MIDI annotations (pitch names, onsets, offsets, and velocity), and does not attempt to map these onto structural details like beats, measures, or sections.

2.4 Interpersonal Entrainment in Music Performance

The Interpersonal Entrainment in Music Performance project (Clayton et al., 2020) has led to the creation of numerous databases of timing onsets. Unlike the other datasets described in this section, the IEMP databases include computationally extracted event onset annotations for multiple performers within a single ensemble, facilitating the analysis of inter-ensemble performance features like group synchrony. The datasets created during this project span Cuban son, Hindustani classical, and Tunisian stambeli styles, amongst others. Rather than deriving these performances from commercial sources, they were either recorded during fieldwork sessions or in experiments, typically involving only a small number of professional ensembles.

3. Database Curation

When compiling any database of music recordings, deciding which material to include can be a challenging task. To simplify this process, we opted to identify a body of suitable ensembles first, and only after doing so would we select appropriate recordings from their respective discographies to include in our database.

3.1 Performer Selection

The criteria used to decide whether a piano trio should be included in our database was that they should be both "popular" and "prolific". With relation to "popularity", we wanted to ensure that groups would only be included if they were both representative of general listening habits and highly regarded by experts. With relation to "prolificacy", we wanted to ensure that groups would only be included if they had made a significant contribution to the overall body of piano trio recordings.

3.1.1 Identifying "Popular" Performers

We searched the Last.fm platform to obtain an overview of jazz listening habits. Last.fm is an online recommendation service that allows users to build profiles of their personal taste by tracking the music they listen to across many streaming platforms. We scraped the names of the top 250,000 performers and groups on

Last.fm most frequently tagged by users with the genre "Jazz", using their official API.[1]

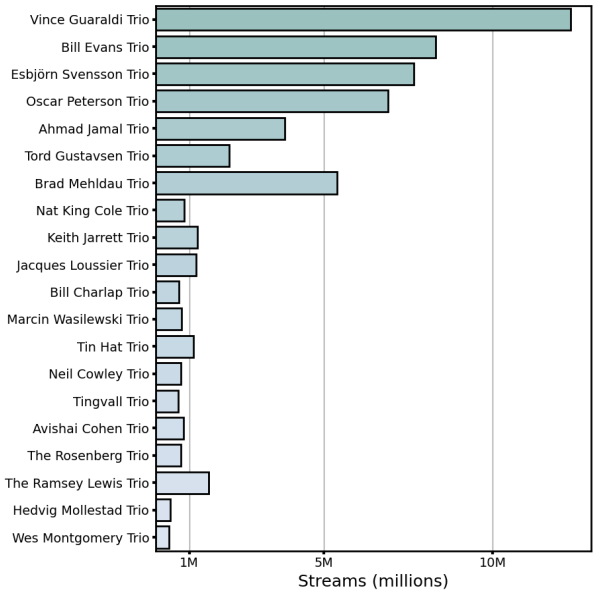

Figure 1: The total Last.fm streams ("scrobbles") of all recordings made by the top 20 "trio" artists most-frequently tagged as "Jazz".

We ordered by tag count, rather than by plays or favorites, as we wanted to find the most quintessentially "jazz" artists, rather than those who fused jazz with other styles. From the resulting list of performers and groups, we selected only those with the word "Trio" in their name. This left 249 unique performers or groups; we show the total play count of all recordings made by the 20 most-tagged performers or groups in Figure 1.

While the names of many of the "usual suspects" featured prominently in this list (Bill Evans, Keith Jarrett), it also included several performers that were mainly known in other genres (e.g., Dee Felice Trio, who accompanied soul singer James Brown when he performed jazz standards) or that mostly composed soundtrack or "stock" music (e.g., Vince Guaraldi Trio, famous for performing the soundtrack to the film "A Charlie Brown Christmas").

These artists were removed by cross-referencing the Last.fm results against two prominent jazz textbooks, keeping only those that received a mention in the discographies within either Ted Gioia's "The History of Jazz" (2011) or Mark Levine's "The Jazz Piano Book" (2011). The intention here was to capture artists who would be highly regarded by expert jazz listeners and performers.

This narrowed the total number of groups to 34, all of which were named after a single musician. This musician would have both led the ensemble and traditionally been expected to compose the majority of the compositions for them to play (e.g., the Dave Brubeck Trio, led by pianist-composer Dave Brubeck). Two bandleaders were bassists (Dave Holland and Ray Brown) and the remainder were pianists; no bandleaders were drummers.

3.1.2 Identifying "Prolific" Performers

Next, we turned to searching the MusicBrainz service to acquire a more detailed summary of each bandleader's recorded discography. MusicBrainz is a community-driven service that provides a comprehensive and open index of discographical metadata (including artist names, recording locations, and release dates) and is commonly used in music information retrieval tasks. We scraped MusicBrainz using the NGS Python bindings to gather metadata relating to every individual recording ever made by each of our 34 bandleaders.[2] This resulted in the identification of 18,504 recordings.

We removed tracks that (1) were duplicated across several releases (for instance, those that also appeared on compilation albums), (2) did not contain a complete trio lineup, or that included multiple musicians performing on one instrument (for example, a piano "four hands" recording), (3) featured a musician doubling on a second instrument (for instance, a pianist who also played synthesizer, or a drummer who played auxiliary percussion), and (4) contained keywords in their title that suggested that they were incomplete performances (e.g., "breakdown", "outtake", "false start"). The cleaned dataset comprised 4,692 unique tracks, with a total runtime of 451 hours.

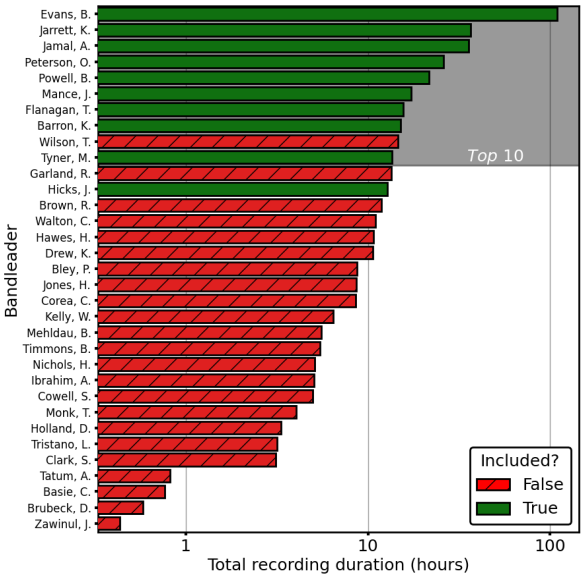

Figure 2: The duration of all recordings produced by the 34 bandleaders, scraped from MusicBrainz; bar color and hatching indicate whether that bandleader was included in the database.

We ordered the 34 bandleaders by the combined duration of all their recordings and selected the top ten most prolific for further study (Figure 2). These ten musicians were all pianists; between them, they had recorded 3,071 unique tracks, featuring the performances of 78 different bassists and 82 drummers. Bassists Ron Carter and George Mraz had the widest performing networks, each having recorded with five of the ten pianists, while drummer Roy Haynes had performed with

four different pianists. Bassist Eddie Gomez and drummer Marty Morell were most frequently heard together as a pair, appearing as a pair on 260 different recordings – all of which had Bill Evans at the piano.

3.2 Recording Selection

We began the process of selecting recordings for inclusion by sorting each of the tracks made by the remaining ten bandleaders chronologically by their date of recording. In cases where a full date could not be obtained, a track was estimated to have been recorded either midway through the month (in cases where both a month and year were given) or year (when only a year was given) that was provided from MusicBrainz. In cases where multiple dates were given for one track (i.e., when an album was recorded over a period of time, without dates being assigned to individual tracks), we took the midpoint of these dates. If no dates were provided, the track was excluded from selection.

Next, we sorted each track into one of 30 equally spaced bins; the left edge of the first bin coincided with the date of a bandleader's earliest recording, and the right edge of the final bin the day of their final (or most recent) recording. Tracks were ordered within each bin by the proximity of their recording date to the midpoint of that bin; if multiple recordings were made on the same day, they were arranged following the order in which they appeared on their original release.

We then identified the first track within each bin that met the inclusion criteria detailed below by listening to it in full. Any tracks that did not meet the inclusion criteria were excluded from selection. In cases where a bin either contained no tracks (i.e., the bandleader did not record during that period) or none that met the inclusion criteria, that bin was excluded. If it proved impossible to obtain one acceptable track from each of the 30 bins, then we obtained additional tracks by choosing the second acceptable track from the first bin, and continuing on as before until 30 tracks were obtained for every bandleader.

3.2.1 Exceptional Inclusions

Not all of the ten bandleaders had recorded enough material that met our inclusion criteria. In the case of the ninth bandleader, Teddy Wilson, we could only identify 27 tracks out of a possible 253 that met the inclusion criteria. This was true also for the 11th Red Garland (29/122 tracks); so, we instead sampled from the 12th bandleader, John Hicks, from whom 30 acceptable tracks could be identified.

3.2.2 Inclusion Criteria

To be included in the database, a recording must have had: (1) an approximate tempo between 100 and 300 quarter-note beats-per-minute (BPM), assessed by tapping along to the opening measures of the performance, (2) a time signature of either three or four quarter note beats per measure, with no changes in meter, (3) an identifiable piano solo, accompanied by bass and drums with no interruptions in the ensemble texture (i.e., "solo breaks"), and (4) an uninterrupted "swing eighths" rhythmic feel.

These inclusion criteria were set to obtain a relatively consistent set of tracks, representative of the traditional "straight ahead" jazz improvisation style, that could be reliably analyzed by our pipeline. We elected only to analyze audio from the piano solo in each track as this was the only section in a performance where every musician in the trio would be expected to improvise. For instance, in the "head" (melodic statements that occur at the beginning and ending of "straight ahead" jazz performances), the material is pre-composed; while, in bass or drum solos, the accompanying musicians may choose to "lay out" and not play at all (Monson, 1996).

Additional criteria were specified for each instrument in the trio. During their solo, pianists must have played on acoustic instruments (rather than, e.g., synthesizer or "Rhodes" piano), and without any external FX like echo or distortion. Bassists must also have played on an unmanipulated upright, rather than electric, instrument, and with their fingers, as opposed to a bow; this was to ensure that every note had a clearly identifiable onset time. Drummers must have played on the traditional jazz drum set configuration (snare and kick drums; hi-hat, ride, and crash cymbals; tom-toms), without auxiliary percussion (shakers, maracas). They must also have used sticks, rather than wire brushes or mallets, to play their instrument, for the same reason as to why we excluded bassists' use of the bow.

Note that these criteria were only enforced during the piano solo. A recording could feasibly be included in the database if a drummer began the performance on wire brushes but switched to drum sticks before the piano solo began, for example, or if a pianist played the opening "head" melodic statement on an electronic instrument but switched to acoustic piano to take their solo. The requirement for drummers to use sticks, rather than wire brushes, resulted in the exclusion of the greatest number of tracks (299 tracks, 10%).

3.2.3 Metadata Curation

For every track that met the inclusion criteria, we compiled metadata from MusicBrainz including its name, when it was recorded, the album it first appeared on, the names of the musicians in the trio, and a URL linking to the performance on YouTube. These URLs were individually checked and replaced with an alternative source from an official YouTube channel (uploaded either by the artist themselves or their recording company) in the case of any false positive matches.

We compiled other metadata manually, including timestamps for the beginning and ending of the piano solo, the position of individual instrument sources across the stereo spectrum, the time signature of the recording, and a timestamp for a single, clearly identifiable "downbeat" (i.e., the first beat of a measure) at the start of the piano solo. The whole database was then converted to a text file. Finally, audio was downloaded from the YouTube URLs, trimmed to the piano solo, and stored in a lossless format on a local machine.

4. Annotation

4.1 Source Separation

4.1.1 Use of the Stereo Spectrum

We exploited a quirk in the historical development of stereophonic audio to provide our source separation models with the cleanest possible signal for each instrumental source. Before the widespread adoption of full-spectrum panning in the 1970s, recording engineers mastering for stereo used a three-way switch to assign an audio signal to either the left or right speaker output, or both (center); positioning tracks elsewhere across the spectrum was otherwise not possible.

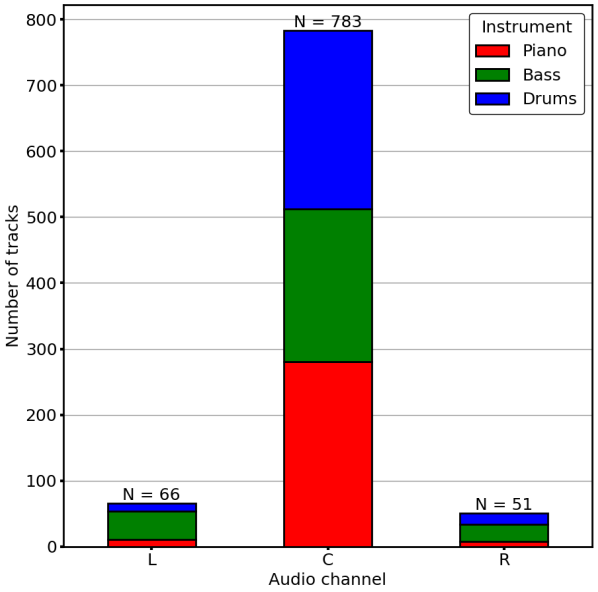

Figure 3: The number of recordings where individual audio sources were panned to either the left or right channel, with color representing the instrument panned.

At least one instrument was panned to either the left or right channel in 28% (84) of the tracks in our database (Figure 3), most commonly the bass (69 tracks, 23%). For these tracks, we processed the left and right monaural signals separately from the center channel.

4.1.2 Model Selection

We then applied one of two models to the audio mixture (or individual channels taken from this) to separate each instrumental source. We used Demucs, a hybrid spectrogram- and waveform-based separation model using transformers (Rouard et al., 2022), to separate double bass and drums. We used Spleeter, a spectrogram-based model using convolutional neural networks (Hennequin et al., 2020), to separate the piano. Both models are released under the MIT license.

Both Demucs and Spleeter have achieved good results in comparison to other available models and have appeared as baselines in several music demixing community challenges. Demucs has performed better than Spleeter on tests of drums and bass separation (Rouard et al., 2022), but the Demucs authors warn that the quality of separation for piano is poor; this led to our decision to use Spleeter for this instrument.

Spleeter was trained on a private, internal dataset, while Demucs was trained on both an internal dataset and the musdb18 dataset (Rafii et al., 2017). Of the 150 tracks in this dataset, only three were tagged as "jazz". For this reason, we designed our inclusion criteria (see above) to maximize the similarity between the database and model training audio – mandating, for instance, the use of sticks (rather than wire brushes) on the drums, as these are more common in the pop/rock style that dominates the recordings that constitute the musdb18 dataset. Neither model received additional fine-tuning for the jazz genre or the piano trio format.

4.1.3 Audio Filtering

After source separation, we filtered the audio for each isolated instrumental source using a second-order Butterworth filter. Our goal was to attenuate frequency bands that multiple instruments might have competed for (such as the double bass and the drummer's kick drum) and that could have "bled" through the source separation model.

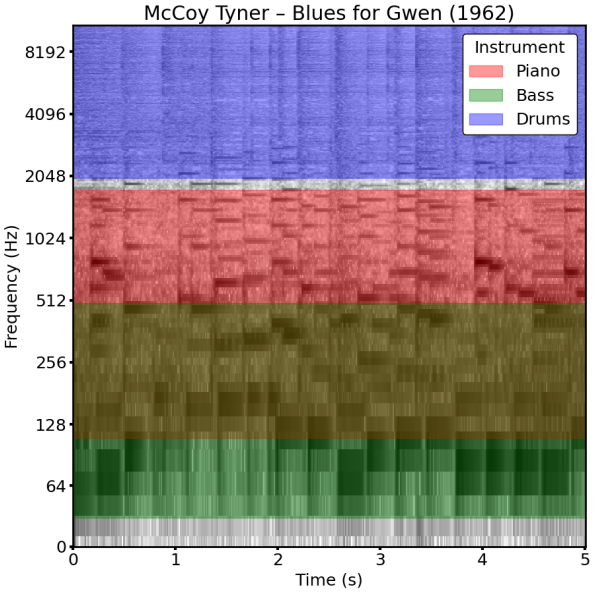

Figure 4: A spectrogram of five seconds of database audio; the colored horizontal spans correspond to the audio frequency range considered for each instrument when detecting onsets.

For the bass, we allowed frequencies between 30–494 Hz to pass (B_0–B_4, a four-octave span from the lowest string on a five-string instrument) and attenuated all others; for the piano, we passed between 110–1760 Hz (A_2–A_6, a four-octave span from the A two octaves below middle C_4); and, for the drums, we passed between 2000–11000 Hz (the approximate frequency range of the ride, hi-hat, and crash cymbal; Figure 4).

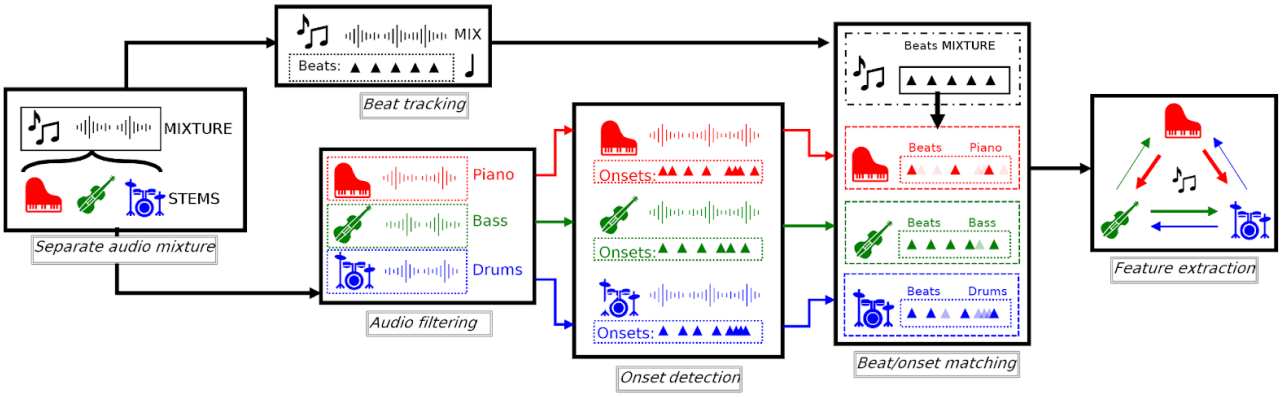

Figure 5. Diagram shows the pipeline used to extract features from the input audio signal, involving processes of source separation, beat tracking, audio filtering, onset detection, and beat/onset matching

4.2 Data Extraction

Our data extraction pipeline (Figure 5) consisted of three components: (1) an onset detection algorithm, used to estimate timestamps for every note onset in the source-separated audio for each instrument, (2) a beat tracking algorithm, used to estimate timestamps for every quarter note beat in the raw audio mixture, and (3) an algorithm to match onsets with their nearest beat, used to estimate the overall metrical structure of a piece.

4.2.1 Onset Detection

We used the algorithm developed by Böck & Widmer (2013) to detect onsets in the isolated audio obtained for each instrument in the trio, after filtering. First, a normalized spectral flux envelope was computed for a track; then, onsets were identified as local maxima within this envelope using a peak-picking algorithm, the parameters of which were set to maximize the concordance between the results and a reference set of annotations created manually (see below).

4.2.2 Beat Detection

We used the algorithm developed by Böck et al. (2016) to detect the position of quarter note beats in the audio mixture. First, a recurrent neural network was applied to an audio spectrogram in order to distinguish between beats, downbeats, and non-beat classes; then, a dynamic Bayesian network inferred meter from the RNN observations. Meter changes cannot be detected by this algorithm, which led us to incorporate this into our inclusion criteria (see above). The range in which the tempo of the detected quarter note beats is allowed to vary can either be inferred from the RNN observations or specified by the user.

We applied this algorithm multiple times to a single track in order to gradually narrow down the range in which the tempo of the detected quarter note beats was allowed to vary. First, the default parameters specified by the authors were used, with the tempo allowed to vary between 100 and 300 beats-per-minute. This resulted in the tempo of the detected beats changing frequently, due to the lack of constraint in the parameter settings. We extracted the inter-beat intervals from successive timestamps and removed outliers according to the 1.5 * IQR rule. Finally, we re-ran the beat tracking algorithm, using the first and third quartile of the cleaned inter-beat interval array as the new minimum and maximum tempo. The total number of iterations this process ran for was optimized during the validation process (see below).

The downbeat classes estimated from the RNN were then combined with the detected beat positions and the time signature for the recording in order to assign beat and bar numbers to every detected quarter note. In addition, we generated a second estimate for the track downbeats and beat numbers by extrapolating forwards and backwards from the timestamp of a single clear downbeat located at the start of an excerpt, identified manually when entering a track into the database (see "Metadata Curation", above). We provide both the "manual" and "automatic" meter annotations for each track, which agreed in the majority of cases (181, 60%). Informally, the manual annotations seemed more accurate in cases where both did not agree, so these were used to create the related figures and analyses in this paper.

4.2.3 Meter Estimation

The final stage of our detection pipeline involved matching every beat with the nearest onset detected in each instrumental source. The size of the window used to match any given onset to the nearest beat varied depending on the tempo of a track. Onsets played up to one thirty-second note before and one sixteenth note after any given quarter-note beat were included within the window; whichever onset had the smallest absolute distance to the beat was understood to mark the pulse. If no onsets were contained within a window, then the musician was considered to not have marked the pulse at that beat.

4.3 Validation

The procedure used to validate our pipeline involved three steps: (1) reference onset and beat annotations were created for a proportion of tracks in the database; (2) the

agreement between these ground truth annotations and those detected by the extraction pipeline was calculated; (3) the parameters used by both detection algorithms were optimized to maximize the overall agreement with the reference annotation set.

4.3.1 Ground Truth Annotations

To evaluate the effectiveness of this detection pipeline, we first created a set of ground truth annotations for a sample of tracks taken from the database. These were identical in format to the annotations created by our pipeline; the only difference was that they were created manually through a process of listening to the audio file and viewing representations of it (waveforms and spectrograms). Two of the authors and one research assistant created the reference annotation set, using the Sonic Visualiser software (Cannam et al., 2010). The annotations of the assistant were later checked for consistency by the lead author. Annotations were created for the earliest, middle, and last recordings made by each bandleader. This meant that 10% (30) of the tracks in the database were annotated, equivalent to approximately 5 hours of audio across all audio sources (instruments + mixture).

4.3.2 Pipeline Performance

The performance of the detection pipeline was determined by considering the proportion of automatically detected annotations that occurred within a small window (50 ms) of a ground truth annotation. The precision and recall of the algorithm was calculated separately for every instrumental source in each track, and an overall *F*-measure was then calculated as their harmonic mean, where $0 \leq F \leq 1$. When $F = 1$, every onset that was detected by the algorithm could be matched with one onset in the reference set, with no onsets left unmatched. In weighting both precision and recall equally, we aimed to develop an algorithm that was neither inaccurate (in terms of missing onsets identified by a human) nor indiscriminate (in terms of identifying every possible auditory event as an onset).

The performance of the pipeline could also be evaluated by looking at the temporal difference between equivalent automatic and manual annotations. The pipeline tended to annotate beats slightly earlier than the human annotators did; the mean difference in beat location time (algorithm – human) was –4.33 ms (*SD* = 13.41). In comparison, the algorithm tended to detect note onsets slightly after the annotators: piano: +8.53 ms (*SD* = 13.56), bass: +4.05 ms (*SD* = 16.24), drums: 2.38 ms (12.45). In context, however, these differences were likely perceptually sub-threshold, and could have resulted from variation between annotators.

4.3.3 Parameter Optimization

We then applied the gradient-free, nonlinear optimization algorithm "subplex" (Rowan, 1990) implemented in the "NLopt" library to set the parameters of the beat and onset detection algorithms separately for each audio source.[3] In each case, the mean value of *F* across the entire reference set was treated as the objective function to maximize.

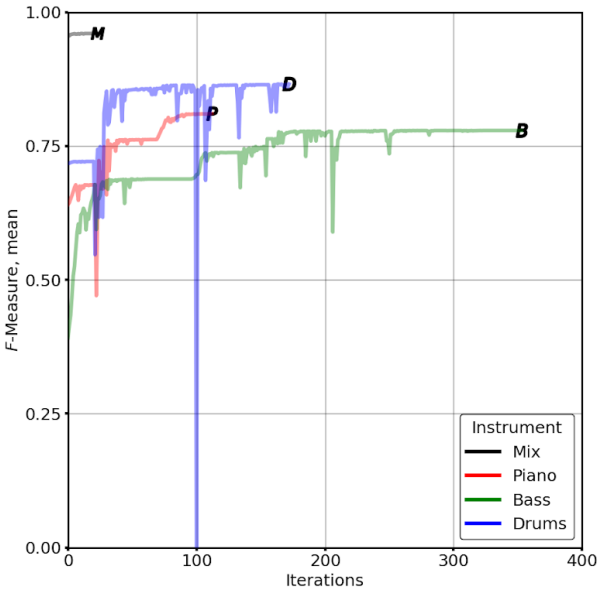

Figure 6: The mean *F*-score across all reference tracks (n = 30) at every iteration of the optimization algorithm. Line color indicates audio source.

For the piano, the algorithm took 113 iterations to converge (Figure 6), with the optimized parameter set yielding mean $F = 0.81$ ($SD = 0.07$) over the reference set. For the bass, the algorithm took 354 iterations, with mean $F = 0.78$ ($SD = 0.10$). For the drums, the algorithm took 173 iterations, with mean $F = 0.87$ ($SD = 0.06$). Finally, when tracking beats in the audio mixture, the algorithm took 24 iterations, yielding mean $F = 0.96$ ($SD = 0.09$) .

This discrepancy in detection performance between the different instruments in the trio was not unexpected. Both drums and piano are percussion instruments, with short attacks that facilitate peak-picking, whereas the double bass has a quantitatively longer attack, making it harder to annotate automatically by picking maxima from the spectral flux envelope.

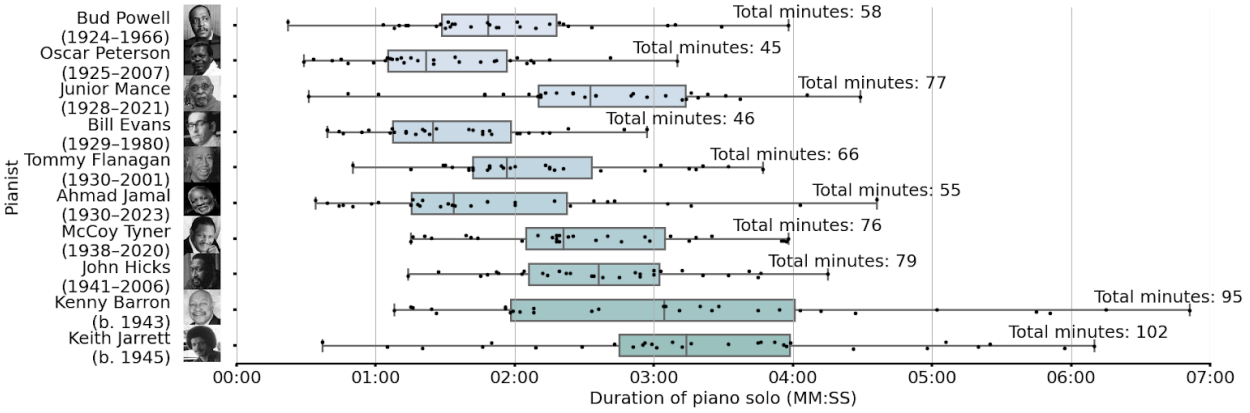

Figure 7. The whiskers of each box plot show the duration of the shortest and longest solos by each pianist in the database. Markers show the duration of individual solos. The total duration of all recordings by a bandleader is given above each box.

4.4 Subjective Evaluation

As a final check on the validity of our pipeline, we obtained evaluations of processing quality from a human participant. A research assistant who had not been involved in designing the pipeline was asked to listen to a sample of 30 tracks drawn randomly from the database (including those with ground truth annotations), and rate for each instrumental source (a) the quality of the audio separation, and (b) the quality of onset detection. Both metrics were evaluated on a scale from 1 to 3, with 3 being the best grade.

The subjective ratings of separation quality followed the distribution of *F*-scores: drums separation was rated the best (mean rating = 2.72, *SD* = 0.45), with piano performing slightly (2.41, *SD* = 0.82) and bass substantially (1.62, *SD* = 0.62) worse. The subjective ratings of onset detection quality also followed this trend; quality of detection was highest for the drums (2.79, *SD* = 0.41), then piano (2.41, *SD* = 0.68), and finally bass (2.21, *SD* = 0.73).

We also found a moderately strong positive correlation between ratings of separation and detection quality, across all instrumental sources: $r(88) = .65$, $p < .001$. This suggested that the quality of timing annotation for an audio source was likely related, at least in part, to how well it could be separated from the audio mixture.

5. Analysis

The database includes excerpts from 300 different jazz trio recordings led by ten different pianists (30 tracks per pianist), recorded between 1947 and 2015 (median = 1978). There were 524,155 total onsets (piano: 210,733, bass: 102,339, drums: 211,083) and 139,161 total beats in the annotation set. On average, each excerpt contained 702 piano onsets, 341 bass onsets, 704 drums onsets, and 464 quarter note beats. The vast majority of tracks (95%, 284 recordings) had a time signature of four quarter note beats per measure. Only 16 tracks (5%) had a time signature of three quarter note beats per measure.

5.1 Solo Duration

The total duration of all piano solo excerpts in the database was 11 hours and 40 minutes. The shortest solo lasted 22 seconds (Bud Powell, "Salt Peanuts", 1956), and the longest 6 minutes 51 seconds (Kenny Barron, "Well You Needn't", 1996). On average, solos lasted for 2 minutes and 20 seconds.

While our database contained the same number of excerpts from each pianist, the duration of these solos varied considerably (Figure 7). Notably, the later a pianist's first recording was made (i.e., the later they started their career), the longer they tended to solo for: the three pianists whose first recordings were made the latest (Kenny Barron, John Hicks, and Keith Jarrett, with recordings made in 1982, 1981, and 1967, respectively) also had the longest average solo duration in the database (Jarrett; 3 minutes 24 seconds; Barron, 3 minutes 10 seconds; Hicks; 2 minutes 38 seconds). Meanwhile, Oscar Peterson, whose earliest recording was also the earliest in the database (1947), had the shortest average solo duration (1 minute 29 seconds) of all the pianists.

One possible explanation for this phenomenon is that solos in chronologically later jazz styles ("free" and "fusion" jazz, for instance) may not have been based on traditional harmonic structures (e.g., the 32-bar "song form"), but on open-ended structures (e.g., "vamps") more suitable for extended improvisation (Gioia, 2011). Alternatively, as the storage capacity of recorded media increased throughout the twentieth century, musicians may have been able to devote more time in a performance to improvised solos (Carr et al., 1988).

5.2 "Standard" Jazz Compositions

The jazz repertoire contains numerous compositions that have been recorded by many different musicians and which are commonly referred to as "standards". Our database contains 252 unique compositions, with 71% of compositions (214 tracks) recorded only once in the database. The most performed composition was "Beautiful Love", with four recordings (three by Bill

Evans, one by Kenny Barron), while eight compositions (including "Autumn Leaves", "Stella by Starlight", and "Whisper Not") had three recordings each.

Compared with the PiJAMA database, where 51% of indexed compositions have only one recording (Edwards et al., 2023), this suggests a substantially lower presence of "standard" compositions in the trio format when compared with solo jazz piano performance. This could suggest a greater tendency for pianists to play original material in an ensemble setting versus performing unaccompanied.

5.3 Tempo

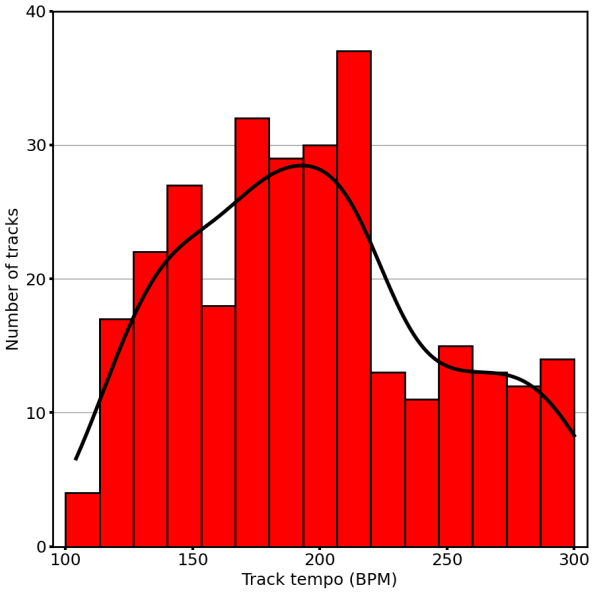

Figure 8: Distribution of tempi (in quarter note beats-per-minute) for tracks in the database.

The average tempo of a recording in our database was 192 BPM (Figure 8). The slowest performance had a tempo of 104 BPM (Junior Mance, "Rainy Mornin' Blues", 1963) and the fastest 310 BPM (Kenny Barron, "Guess What", 2005).

As measures of tempo stability, for each track we obtained (1) the standard deviation of the tempo normalized by the mean tempo (i.e., the percentage fluctuation about the overall tempo, "tempo fluctuation"), and (2) the slope of a linear regression of instantaneous tempo against beat onset time, with positive values implying acceleration and negative values deceleration ("tempo slope"). The average tempo fluctuation was 4.59%, while the average tempo slope was 0.03 BPM/s. This suggests that performances were typically stable, with a very slight tendency towards acceleration (with a predicted change of +1 BPM approximately every 30 seconds).

There was no correlation between mean tempo and tempo fluctuation, $r(298) = .02$, $p = .72$, nor between mean tempo and tempo slope, $r(298) = .04$, $p = 0.52$. This suggested that neither the stability of a track nor whether it changed tempo related to its overall pace.

5.4 Swing

In jazz, swing refers to the subdivision of the musical pulse into alternating long and short intervals. Expressed in Western musical notation, the long interval is typically written as a quarter note triplet, and the short as an eighth note triplet. Empirically, swing can be measured by taking the ratio of these long and short durations (the swing or "beat-upbeat ratio", commonly expressed in binary logarithmic form in the literature). For notated "swung" eighths, the beat-upbeat ratio = 2:1 ($log_2 = 1$); for "straight" eighths (i.e., the equal-equal subdivision of the beat), beat-upbeat ratio = 1:1 ($log_2 = 0$).

We searched our database for all discrete groupings of three onsets where the first and last had marked the quarter note pulse. The total number of such groupings was 88,025. Following Corcoran & Frieler's analysis of the WJD (2021), we classified ratios above 4:1 ($log_2 = 2$) and below 1:4 ($log_2 = –2$) as outliers, which resulted in an exclusion rate of <1% of total identified groupings. The final number of beat-upbeat ratios in the dataset was 86,639 (piano: 30,177, bass: 7,362, drums: 49,100).

5.4.1 Differences in Swing Between Instruments

To evaluate the differences in swing ratio between the instruments in the piano trio, we first smoothed the distribution of beat-upbeat ratios obtained for each instrument through kernel density estimation (with the optimal bandwidth calculated using Scott's rule-of-thumb). Then, we applied the peak-picking algorithm (using the default parameters) from the "SciPy" library to obtain the local maxima of the smoothed curve.[4] Confidence intervals for these peaks were generated by bootstrapping over bandleaders ($n = 10{,}000$ replicates).

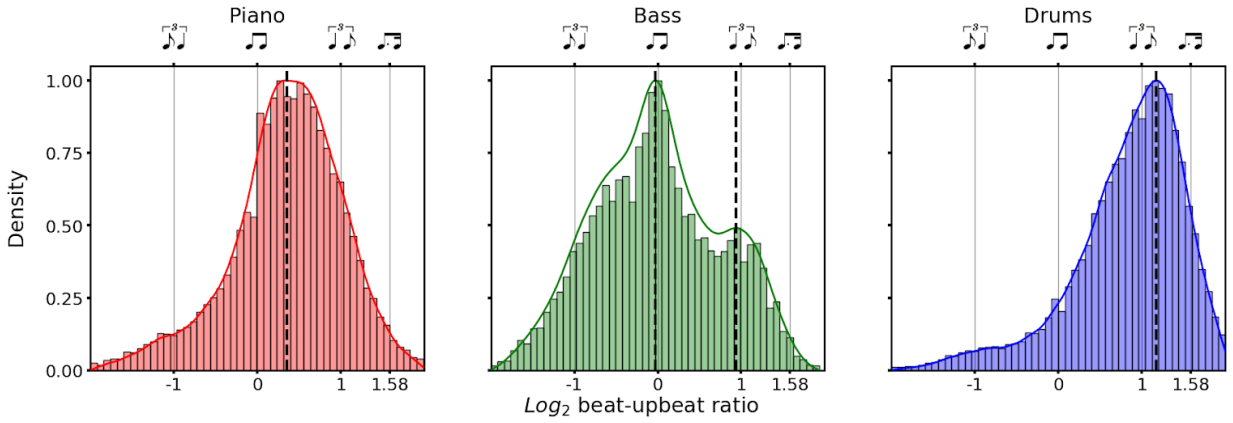

Figure 9. Each panel shows the distribution of log_2 beat-upbeat ratios between instruments across the whole dataset, normalized such that the height of the largest bar in each panel is 1. Dotted vertical lines show peaks of the density estimates; straight lines correspond with the musical notation given along the top of the panel.

For the piano, we found one peak in the density estimate, corresponding to a log_2 beat-upbeat ratio of 0.35 (beat-upbeat ratio: 1.27:1, 95% CI: [0.28, 0.51]). For the bass, we found two separate peaks, corresponding with log_2 beat-upbeat ratio values of –0.03 (0.98:1, [–0.05, –0.01]) and 0.93 (1.91:1, [0.85, 1.01]). For the drummers' cymbals, we found one peak, at log_2 beat-upbeat ratio 1.17 (2.26:1, [1.15, 1.20]) (Figure 9, dotted lines).

We took from this analysis that: (1) pianists targeted long-short subdivisions of the quarter note, with the peak of their density estimate suggesting that these were typically closer to notated "straight" than "swung" eighths, (2) bassists targeted both equal-equal and long-short subdivisions of the beat, with their peaks aligning nearly exactly with notated straight and swung eighths, and (3) drummers primarily targeted the long-short subdivision of the beat, with their peak lying slightly above that implied by notated triplet swing.

We noted that the peak obtained from the pianists in our dataset was similar to the mean log_2 beat-upbeat ratio of 0.38 found in the analysis of the WJD conducted by Corcoran & Frieler (2021). Given that this dataset included solo improvisations made on many different instruments (beyond just the piano), this suggested that swing feel may differ more between solo and accompanying roles, rather than just between different instruments that fulfill the soloist role. In comparison to the WJD, our work is the first to include data from these accompanying roles.

5.4.2 *Effect of Tempo on Swing*

Next, we considered the relationship between swing and the tempo of a performance. We fitted a linear mixed effects model, predicting a performer's mean log_2 beat-upbeat ratio using the mean tempo of the recording (standardized through z-transformation), their instrument, and the interaction between tempo and instrument as fixed effects (piano = reference category). Bandleader was used as a random effect (slopes and intercepts). An individual musician's performance was excluded if 15 ratios could not be obtained, resulting in the exclusion of 136 out of 900 performances, mostly by bassists.

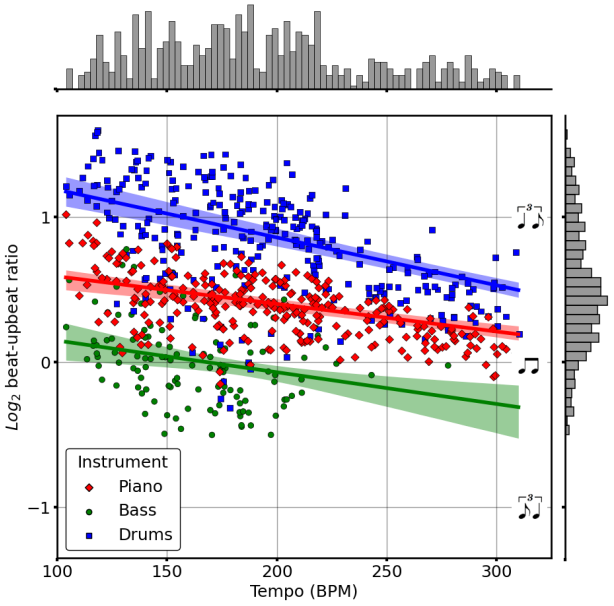

Figure 10: Markers show the mean log_2 beat-upbeat ratio and tempo for a recording, solid lines indicate predictions (without random effects), and shaded areas indicate bootstrapped 95% confidence intervals (n = 10,000 replicates).

An increase in mean tempo was a significant predictor of a decrease in mean beat-upbeat ratio for a recording, with a one standard deviation change in BPM associated with a change of –0.11 log_2 beat-upbeat ratio ($p < .001$, 95% CI: [–0.15, –0.08]). This suggested that it became harder for musicians to articulate long-short subdivisions of the quarter note as its duration decreased. This "straightening" effect has also been observed in

analyses of the WJD (Corcoran & Frieler, 2021) and Filosax datasets (Foster & Dixon, 2021). There was a significant interaction between instrument and tempo for the drummer ($\beta = -0.08$, $p < .05$, 95% CI: [–0.12, –0.05]), suggesting that the effect of tempo on swing was more severe for this instrument (Figure 10).

The standard deviation in mean log_2 beat-upbeat ratio estimated for the random effect of the bandleader was 0.05. The amount of variance in the data explained by both the fixed and random effects of the model was 69%, compared with 68% for the fixed effects only. This suggested only minimal differences in the effect of tempo on swing between groups led by different pianists.

5.5 Synchronization

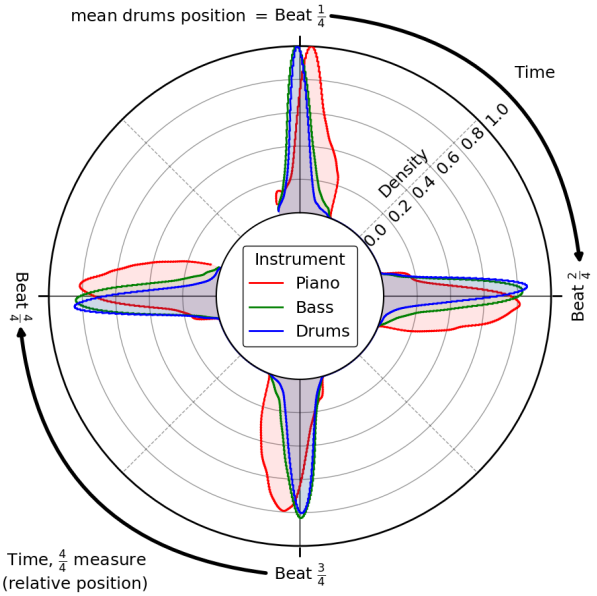

Figure 11: Diagram shows kernel density estimates for the relative position of beats by each instrument, indicated by color. The flow of time unfolds in a clockwise direction, with the lines at 0, 90, 180, and 270 degrees corresponding to the position of each beat in a measure of four quarter notes. Values are shifted so that the mean position of drummers' first beat aligns with 0 degrees. Density estimates are scaled such that the maximum height of the curve for each instrument is 1.

Synchrony can be defined as the temporal difference between onsets played by two musicians that demarcate the same moment in a piece (e.g., the same beat). While synchrony can be expressed in "raw" units (milliseconds, frames), we chose to express it as a percentage of a single quarter note beat at the tempo of the given track, which allowed for comparison across performances made at different tempi. For example, a value of 25% would imply that one musician played a sixteenth note after another. We calculated the synchrony between all pairs of instruments in the trio at every quarter note beat, across all tracks in the database (Figure 11). Confidence intervals were again obtained by bootstrapping over bandleaders (n = 10,000).

5.5.1 *Soloist – Accompaniment Synchrony*

On average, pianists marked the beat 5.67% (95% CI: [4.34, 6.67]) of the duration of a quarter note later than drummers and 4.34% ([3.14, 5.38]) later than bassists. This was equivalent to slightly less of a sixty-fourth note (6.25% of a beat) delay between the soloist and the accompaniment. This phenomenon (the "relaxed" or "laid back" solo feel), has been observed frequently in the literature on jazz improvisation (e.g., Butterfield, 2010).

To investigate whether this effect depended on the tempo of a performance, we fitted a mixed effects model that predicted the average piano asynchrony to the bassist or drummer using the performance tempo (after z-transformation), the accompanying instrument, and the interaction between tempo and instrument as fixed effects (bass = reference category). Bandleader was used as a random effect (slopes and intercepts). As with the model used to predict mean log_2 beat-upbeat ratio (see above), individual soloist-accompaniment pairings were excluded if 15 values could not be obtained for a performance. This resulted in the exclusion of 13 out of 600 total values.

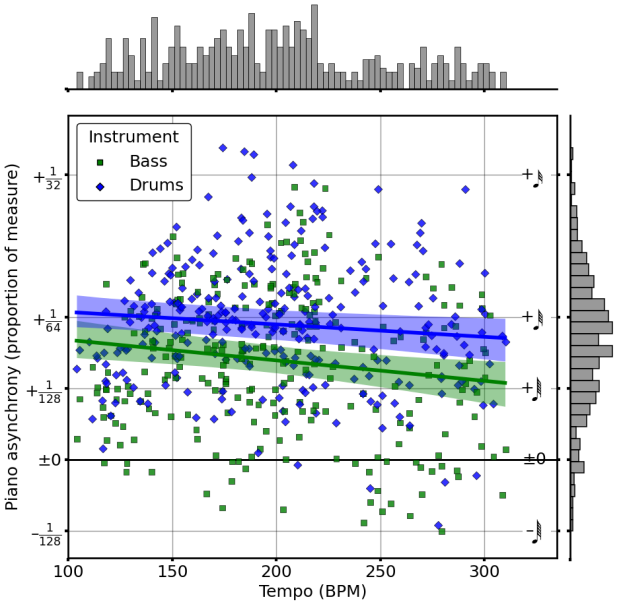

Figure 12: Markers show the mean piano asynchrony and tempo for a recording, solid lines indicate predictions (without random effects), and shaded areas indicate bootstrapped 95% confidence intervals (n = 10,000 replicates).

Increased tempo predicted significantly reduced asynchrony; for every *SD* increase in BPM, the pianist played closer to the accompaniment by a predicted 0.46% of the duration of a quarter note beat ($p < .05$, 95% CI: [–0.87, –0.05]). There was no significant interaction observed between accompanying instrument and tempo ($\beta = 0.16$, $p = .46$, 95% CI: [–0.27, 0.60]). Faster performances thereby had closer soloist-accompaniment

synchronization than slower ones, but this did not differ between either bass or drums (Figure 12).

The size of this effect was relatively small, however, with less than a 256th note difference (1.56% of a quarter note beat) in predicted mean asynchrony between pianist and bassist at both the slowest and fastest tempo in the corpus. The amount of variation in the data explained by the fixed effects was only 10%, compared with 19% for fixed and random effects – suggesting that differences between pianists were an equally likely source of variation in synchrony.

5.5.2 Accompaniment – Accompaniment Synchrony

On average, bassists marked the beat 1.73% (95% CI: [1.04, 2.43]) of a quarter note later than drummers. This is consistent with previous research that has noted the importance of close synchronization between bass and drums to act as an anchor for the soloist's performance in jazz (Butterfield, 2010). We also noted that the asynchrony between bass and drums was slightly higher on the even-numbered beats of a bar (see Figure 11). This would have had the effect of stretching the duration of even-numbered beats, and in so doing would have emphasized the underlying metrical hierarchy. This is because the second and fourth beats of a measure (the 'backbeat') are generally considered to be metrically strong in jazz (Levine, 2011).

6. Conclusion

We have introduced the Cambridge Jazz Trio Database, a dataset of 300 jazz piano trio recordings with automatically generated timing annotations. Appropriate recordings were identified by scraping user-based listening and discographic data, source separation models were applied to isolate audio for piano, bass, and drums, and timing annotations were generated by applying beat and onset detection algorithms to the separated audio. The pipeline achieved a mean *F* score of 0.85 when compared with equivalent ground truth annotations.

Several analyses were conducted using the database that reproduced and extended findings from earlier computational studies of jazz. This has exciting implications for the analysis of this and other forms of improvised music as it removes the need to manually annotate recordings, enabling researchers to massively scale up their work in this area through the use of automated methods.

We encourage the use of our database in the development of performer identification models and in symbolic music generation. We would also welcome extensions to the database beyond the current v.01 that include annotations of additional features, such as harmony or melody.

We can foresee some limitations of our work. Our criteria for including a recording in the database were particularly strict, with the aim of providing the separation models with material as close as possible to the music they had been trained on. This necessitated having to identify acceptable tracks manually. Methods for the automatic tagging of recordings (e.g., distinguishing between a drummer's use of brushes or sticks) would enable a more efficient and scalable data collection process. Expanding to include larger ensemble recordings containing the piano-bass-drums lineup would also increase the number of recordings that could be included, provided that suitable piano solo excerpts could be located within them. Finally, the database shows an imbalance towards male bandleaders and musicians that perhaps represents a gender imbalance in jazz listening habits. Exceptional inclusions could be made in future revisions of the dataset to include prolific female bandleaders who did not appear in the results of the Last.fm search results.

We have released the code to build the database and provide the timing annotations for download.[5] We have also developed a web app, which includes full code documentation and a variety of interactive resources enabling the exploration of the database and related analyses without downloading or building it from source.[6]

Notes

[1] https://www.last.fm/api
[2] https://github.com/alastair/python-musicbrainzngs
[3] https://github.com/stevengj/nlopt
[4] https://github.com/scipy/scipy
[5] https://github.com/HuwCheston/Cambridge-Jazz-Trio-Database/
[6] https://huwcheston.github.io/Cambridge-Jazz-Trio-Database/

Acknowledgements

The authors express their thanks to Tessa Pastor for her help in constructing the dataset.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

HC: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, writing – original draft, writing – review & editing.
JLS: conceptualisation, methodology, investigation, writing – review & editing.
IC: conceptualization, writing – review & editing, supervision.
PMCH: conceptualization, writing – review & editing, supervision.

Author Affiliations

HC: Centre for Music and Science, Faculty of Music, University of Cambridge
JLS: Department for Psychology, Neuroscience & Behaviour, McMaster University
IC: Centre for Music and Science, Faculty of Music, University of Cambridge

PMCH: Centre for Music and Science, Faculty of Music, University of Cambridge

References

Balke, S., Reck, J., Weiß, C., Abeßer, J., & Müller, M. (2022). JSD: A Dataset for Structure Analysis in Jazz Music. *Transactions of the International Society for Music Information Retrieval*, *5*(1), 156–172. https://doi.org/10.5334/tismir.131

Böck, S., Krebs, F., & Widmer, G. (2016). Joint Beat and Downbeat Tracking with Recurrent Neural Networks. *Proceedings of the 17th International Society for Music Information Retrieval Conference*. 17th International Society for Music Information Retrieval Conference, New York.

Böck, S., & Widmer, G. (2013). Maximum Filter Vibrato Suppression for Onset Detection. *Proceedings of the 16th International Conference on Digital Audio Effects*. 16th International Conference on Digital Audio Effects, Maynooth, Ireland.

Butterfield, M. (2010). Participatory Discrepancies and the Perception of Beats in Jazz. *Music Perception*, *27*(3), 157–176. https://doi.org/10.1525/mp.2010.27.3.157

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files. *Proceedings of the ACM Multimedia 2010 International Conference*, 1467–1468.

Carr, I., Fairweather, D., & Priestley, B. (1988). *Jazz: The Essential Companion*. Paladin.

Clayton, M., Jakubowski, K., Eerola, T., Keller, P. E., Camurri, A., Volpe, G., & Alborno, P. (2020). Interpersonal Entrainment in Music Performance. *Music Perception*, *38*(2), 136–194. https://doi.org/10.1525/mp.2020.38.2.136

Corcoran, C., & Frieler, K. (2021). Playing It Straight: Analyzing Jazz Soloists' Swing Eighth-Note Distributions with the Weimar Jazz Database. *Music Perception*, *38*(4), 372–385. https://doi.org/10.1525/mp.2021.38.4.372

Edwards, D., Dixon, S., & Benetos, E. (2023). PiJAMA: Piano Jazz with Automatic MIDI Annotations. *Transactions of the International Society for Music Information Retrieval*, *6*(1), 89–102. https://doi.org/10.5334/tismir.162

Foster, D., & Dixon, S. (2021). Filosax: A Dataset of Annotated Jazz Saxophone Recordings. *Proc. of the 22nd Int. Society for Music Information Retrieval Conf*.

Gioia, T. (2011). *The History of Jazz* (2nd ed.). Oxford University Press.

Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. (2020). Spleeter: A fast and efficient music source separation tool with pre-trained models. *Journal of Open Source Software*, *5*(50), 2154. https://doi.org/10.21105/joss.02154

Levine, M. (2011). *The Jazz Theory Book*. Sebastopol: Sher Music Company.

Monson, I. T. (1996). *Saying something: Jazz improvisation and interaction*. Chicago: University of Chicago Press.

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

Rafii, Z., Liutkus, A., Stöter, F.-R., Mimilakis, S. I., & Bittner, R. (2017). *The MUSDB18 corpus for music separation* [dataset]. Zenodo. https://doi.org/10.5281/zenodo.1117372

Rouard, S., Massa, F., & Défossez, A. (2022). *Hybrid Transformers for Music Source Separation* (arXiv:2211.08553). arXiv. http://arxiv.org/abs/2211.08553