Klaus Frieler

A feature history of jazz improvisation

1. Introduction

This study tries to explore the possibilities statistical and computational methods can offer for historical jazz research. As a corpus study it is not focusing on persons and events but on measurable quantities extracted from historical documents, here: jazz improvisations transcribed from recordings as contained in the Weimar Jazz Database (see Pfleiderer et al 2017), which cover the years from 1925 to 2009. With the ongoing proliferation of digital tools and data in the humanities, methods such as "distant reading" become increasingly common (see Moretti 2013). It is clear that such approaches forego details and precise historical context for a broader view on the subject matter. Using a huge number of "thin" descriptions of cultural objects contrasts sharply with more traditional ways of focusing on few selected items with very rich or "thick" descriptions. But we think that both approaches can complement each other. Brought together, they can result in a more complete view.

This approach is not without antecedents in the broader field of digital humanities, e.g., Franco Moretti's work in literature research (Heuser et al 2016). In the field of musicology, there have recently been two studies about the history of pop music, using automatically extracted features from large sets of audio files (Mauch et al 2015, Serrá et al 2011), as well as one dealing with classical Western music (Weiß et al 2018).

We focus on a transpersonal historical development of the improvised solo – one of the central elements in jazz. This is done in a mainly descriptive manner by tracing certain objectively measureable features of a large set of solos over time. Some attempts are made towards causal explanations. Moreover, we try to identify "points of maximal change" using statistical methods to illustrate the power of this approach to recover or discover stylistic changes.

The present study should be considered as a proof of concept and as a step towards introducing new methods in jazz research which can serve as a starting point of future research.

2. Method

We used a large set of 159 features extracted from the solos in the Weimar Jazz Database. These features are numerical descriptors based on several musical dimensions such as pitch, interval, rhythm, tone formation, and others. For each feature, we calculated the Spearman rank correlation with the recording year of the solo as well as a linear regression for the sake of visualization[1]. Mostly, only features with a correlation greater than .20 were kept. Due to

---

[1] Spearman correlation uses ranks rather than values. This has the advantage of being more robust to outliers in the data. Mostly, however, Spearman rank and standard (Pearson) correlation give very similar results. Linear regression is a method to fit a straight line to a given set of data, such that the error, i.e., the

this paper's limited space and certain redundancies in the features, we will report only some of the most relevant findings using a reduced set of 24 features (cf. Tab. 2). We plan to prepare an accompanying website with more complete information (http://jazzomat.hfm-weimar.de/feature_history). All features were extracted with the help of the MeloSpyGUI software, developed by the Jazzomat Research Project. Software and data are freely available from the project website (http://jazzomat.hfm-weimar.de). All statistical analyses were carried out using the software R (R Development Core Team 2008).

2.1 Data

The Weimar Jazz Database (WJD) contains high-quality transcriptions of 456 monophonic solos by 78 different soloists with over 200,000 tone events. A solo transcription is represented as a list of tone events with the three core parameters *onset* (in seconds), *duration* (sec) and *pitch* (in MIDI numbers, 0-127, C4=60, representing semitones (see for example, Selfridge-Field 1997)) and a vast array of annotations such as metrical position, intensity, and f0-modulations (vibrato, glides, slides, etc.), as well as segmental annotations such as phrase and chord contexts, chorus IDs, and mid-level units (Frieler et al 2016). Transcriptions and annotations were produced manually by expert transcribers and carefully double-checked. Timbre and intensity properties were assembled using state-of-the art audio algorithms (Abeßer et al 2016). Extensive metadata, such as discographic information, is also available for each solo. An overview of the data set can be found in Table 1, a full list of solos is available on our website (http://jazzomat.hfm-weimar.de/dbformat/dbcontent.html).

Table 1: Short Overview on the Weimar Jazz Database

Solos	456
Performers	78
Number of Tones	200,809
Performers with the most solos	Coltrane (20), Davis (19), Parker (17), Rollins (13), Liebman (11), Brecker (10), Shorter (10), S. Coleman (10)
Styles	Traditional (32), swing (66), bebop (56), cool (54), hardbop (76), postbop (147), free (5 = O. Coleman)
Instruments	ts (158), tp (101), as (80), tb (26), ss (23), other (68)
Time range	1925–2009

For the present study, the distribution of solos with respect to jazz history is a crucial factor. The solos in the WJD were not specifically selected for the purpose of a historical analysis, though one of the guiding principles was to cover a broad range of styles from jazz history (Pfleiderer 2017). The focus is on US jazz and the established canon of eminent master improvisers, which are also represented with larger numbers of solos in the database, but lesser known improviser were also included. Great care was taken to receive a rather

squared distance between the line and the corresponding data points is minimal. There is a close connection between linear regression and Pearson correlation, since the error will be smaller the higher the correlation (and vice versa).

representative sample of this inarguably very important, segment of jazz. But there are also limitations, which partly stem from the used data format, which is well-suited to represent jazz based on pitches, beats, and chords, but not for styles – such as avant-garde and free jazz – which abandon one or more of these principles. Hence, the database does not contain free jazz solos (except for a small set of early Ornette Coleman solos). Other points are: (1) solos on one and the same composition (e.g., "Body and Soul", blues, "Rhythm" changes) to enable comparisons; (2) alternate takes for the same solo on the same recording date, for the same reason; (3) several solos taken from the same track by the same or different soloist (or both); (4) several tracks from the same album and/or recording session. For example, the year with the most recordings, 1964, has 23 solo transcriptions by 8 soloists on 14 different compositions. These specifics introduce some bias and imbalance in the dataset, which is, however, alleviated by the large number of solos.

Defining a representative sample is an intricate problem. Representativeness is a relation, i.e., it is always: "representative for what?" In the case of jazz solos this is a specifically pertinent question. First of all, recorded jazz solos might not be in themselves fully representative for the actual practice of improvising in a live context. The situation in a recording studio has its own rules and affordances which certainly influences the improviser and the band. An obvious example is that before the introduction of vinyl albums in the 1950s, recorded jazz solos had to be very restricted in time due to the three-and-a-half minute time limit imposed by 78-RPM records. Moreover, jazz styles went in and out of fashion but they effectively never fully ceased to exist – at least in a niche. Players who grew up and got established in a certain style tend to stick to that style. Another issue is that the selection (of solos) was often based on innovation and post-hoc importance. For instance, in the 1940s, the majority of jazz that could be heard at concerts, dances, and on records was big band swing, traditional New Orleans jazz, and Rhythm and Blues, but none of this entered the WJD, since the most important development at this time was bebop. Hence, in an attempt to cover the jazz canon as defined by jazz scholar, historians, musicians, and writers (and widely accepted by the jazz audience), the WJD incorporates mostly (but not exclusively) solos from the major figures deemed most representative for the developments at a certain point in time. In this sense, the WJD can be regarded as representative for the most important stylistic developments at a given time span, but it might not be representative with respect to other parameters such as popularity.

Looking at the temporal distribution of the solos in the WJD (Fig. 1), there are certain clusters and gaps noticeable. The first gap is located in the 1930s, the era of big band jazz, in which smaller ensembles and longer solos were not in fashion. This gap reflects a factual scarcity of interesting and/or long enough recorded solos in small ensemble context (i.e., without the backing of a big band). The dominating style of the 1970s was probably jazz-rock and fusion with a diminished role of the solo. This and the difficulty to find lead sheets for these recordings explain the second big gap in the WJD in the 1970s.

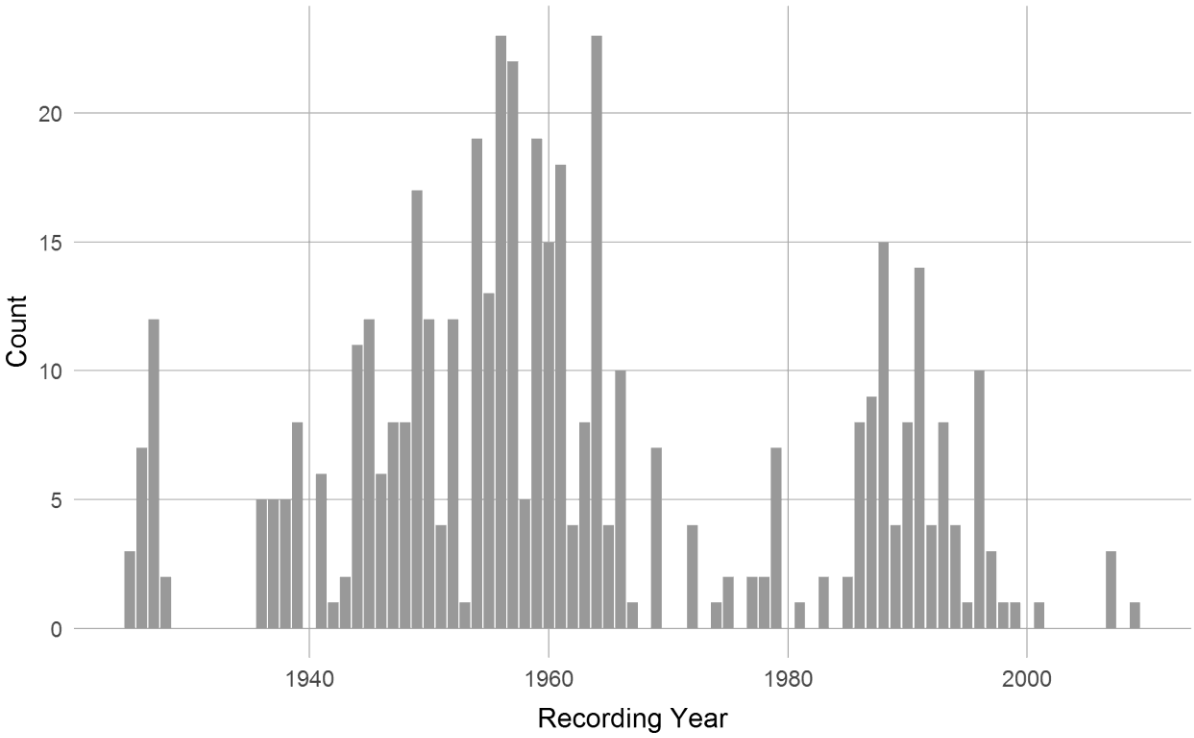

Fig. 1: Distribution of recordings years for all 456 solos in the WJD.

2.2 Features

For the present study, a broad range of features was used, defined as numerical values capturing some measurable aspect of the solo improvisation. The MeloSpyGUI software currently offers about 600 features, but not all of them are suitable for this kind of task and some features are contained more than once or in slight variations. Hence, a subset of 159 features was selected (very similar to those used in Frieler et al 2015), covering a wide variety of musical dimensions, e.g., pitch, interval, rhythm, tone formation and playing ideas, which were further reduced after calculating correlations. Specificities of the features will be discussed along with the results. An overview of the reduced set of 24 features can be found in Table 2.

Table 2: Overview of the reduced set of 24 features reported in the study

Feature name	Description
abs_int_range	Size of largest interval
art_range	Range of articulation
cdpcx_density_1	Rel. frequency of chordal roots
cdpcx_density_10	Rel. frequency of chordal sevenths
cdpx_density_3	Rel. frequency of chordal thirds
cdpcx_density_5	Rel. frequency of chordal fifths
cdcpx_density_6	Rel. frequency of chordal sixths
cdpcx_density_b2	Rel. frequency of flat ninths
cdpcx_density_T	Rel. frequency of sharp elevenths
durclass_abs_entropy	Entropy of absolute duration classes
durclass_abs_hist_very_short	Rel. frequency of *very short* tones (< 16th notes in 120 bpm)

expressive	Rel. frequency of *expressive* MLUs
f0_median_dev	Median deviation from nominal 12TET pitch
fragment	Rel. frequency of *fragment* MLUs
int_bigram_entropy	Entropy of interval combinations
lick	Rel. frequency of *lick* MLUs
line	Rel. frequency of *line* MLUS
loudness_sd	Standard deviation of tone loudness
pitch_entropy	Entropy of pitch distribution
pitch_range	Pitch range
pitch_std	Mean distance from mean pitch.
rhythm	Rel. frequency of *rhythm* MLUS
total_duration	Total duration of solo (in sec)
void	Rel. frequency of *void* MLUS

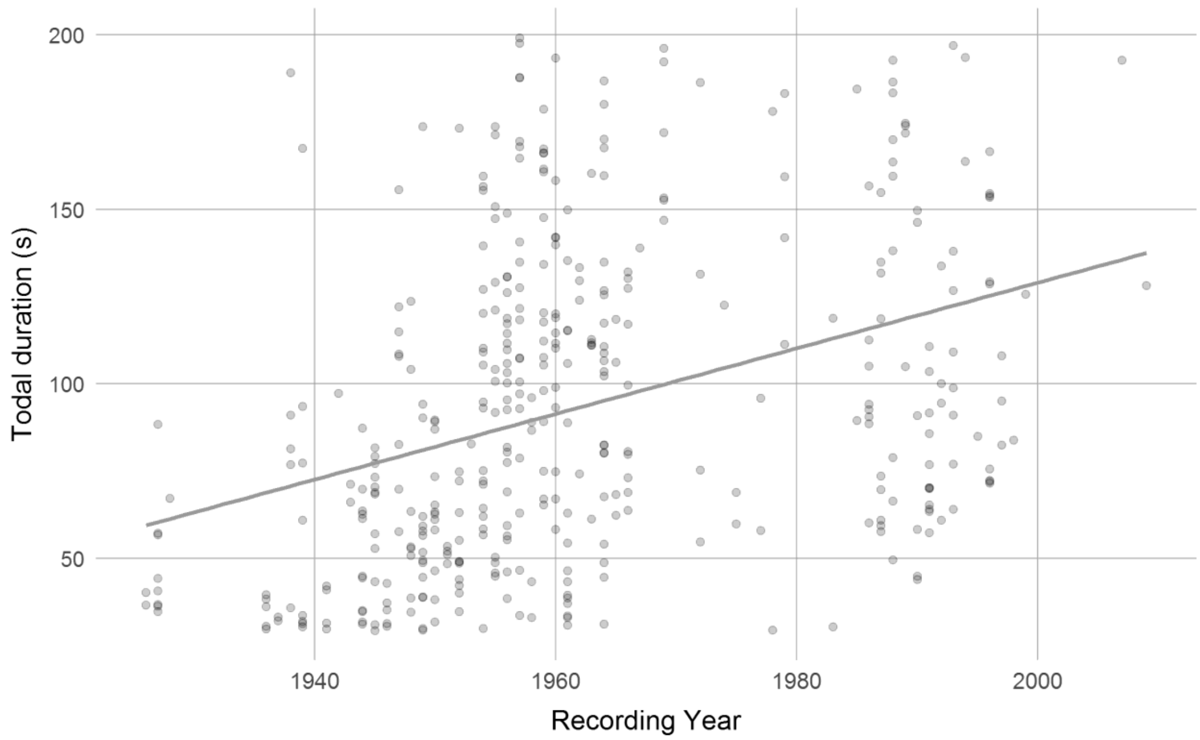

Fig. 2: Total duration (in seconds) over time.

3. Results

3.1. General

The first result is a rather expected one: (recorded) solos were becoming longer (s. Fig 2). The correlation is high with $\rho = .534$ ($p < .001$), but basically vanishes after 1955, though there is a smaller boost at the end of the 1990s (cf. Fig. 12, Panel "total_duration"). This trend is caused by the substitution of the old "three and a half minute" 78-RPM records with longer 12" vinyl records, which had conquered the market by the mid-1950s.

3.2. Rhythm

There are four representations for rhythmical durations in the full set of features, but we present here only the ones based on the actual sounding durations of tone events. Based on an absolute reference time of 500 ms (= 120 bpm), five (absolute) duration classes are defined: *very short*, *short*, *medium*, *long*, and *very long*. The border between these classes is set so that the classes roughly correspond to the sixteenth, the eighth, the quarter, the half, and the whole note level (in tempo 120 bpm). In this manner, all events can be classified and frequency distribution of classes and other numerical features can be derived, e.g., the estimated probability of a given class or the information entropy of the entire distribution. Information entropy is a concept from communication theory (Shannon 1948) which measures, amongst others, the uniformity or predictability of a probability distribution[2].

Amongst the features connected to rhythm, the normalized absolute duration class entropy was clearly trending downwards over time ($\rho = -.304$, $p < .001$). Likewise, the probability to find a *very short* note duration was clearly trending upwards ($\rho = .266$, $p < .001$). Together these findings indicate a trend to playing more shorter notes, since the entropy will decline, if one possibility (here: the *very short* durations) dominates.

3.3. Pitches and intervals

We examined indicators of pitch variability and usage, such as pitch entropy (predictability of pitches), pitch standard deviation (average deviation from the mean value), and pitch range (difference between the highest and lowest pitch). All these values correlate very highly, partly mediated by pitch range, since using a wider scope of pitches makes pitches less predictable and increases also the standard deviation. The results for pitch range can be found in Fig. 3. The upward trend is very pronounced, indeed, one of the strongest trends amongst all features. The Spearman correlation is $\rho = .596$ ($p < .001$). It is most pronounced for tenor saxophones players ($\rho = .721$), but also present for all other instruments ($\rho = .50$).

---

[2] The more predictable the outcome, the lower is the information entropy. In the extreme case, where one event is sure to happen, the entropy is 0. In the other extreme case, where all possible events are equally likely to happen, the entropy is $\log_2 N$, where N is the number of possible outcomes. Using this maximal value, entropy can be normalized to values in the range between 0 and 1.

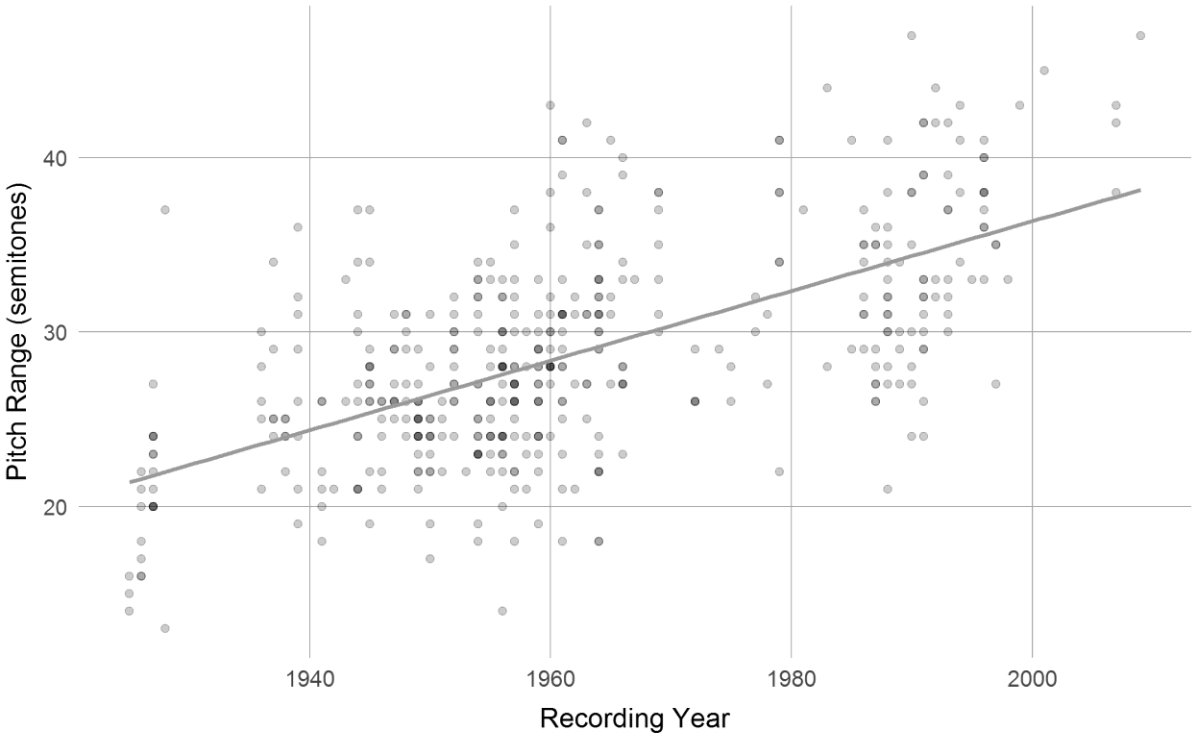

Fig. 3: Development of pitch range/ambitus with recording year.

Next, we had a look at pitch choices with respect to the underlying chord. To this end, we used a representation called "extended chordal diatonic pitch class" (CDPCX), which classifies pitches into 13 possible classes depending on the chordal context. Base is the major scale for chords with major thirds and the Dorian scale for chords with minor thirds. This gives seven possible diatonic pitch classes (labelled 1-7), whereby the seventh scale degree is always determined by the seventh prescribed in the chord. The remaining possibilities are: #11 (augmented fourth/diminished fifth), #9 (minor third over major-triad chord), b9 (minor ninth), b13 (minor sixth), #7 (major seventh over chord with minor seventh), and #10 (major third over minor-triad chord, "blue third"). The relative frequencies of the CDPCX classes were correlated with recording year. After selecting all significant correlations regardless of the magnitude of the correlation, we found seven trends (Fig. 4).

The strongest trend could be observed for the fifth of a chord, which clearly decreases over time ($\rho = -.30$, $p < .001$). This is somewhat complemented by an increasing trend of the #11 ($\rho = .29$). Likewise, the usage of roots declined with time ($\rho = -.27$), as well as that of thirds ($\rho = -.171$) and sixths ($\rho = -.23$). On the other hand, the #10 ($\rho = -.26$) and the b9 ($\rho = .23$) showed small but significant upward slopes. All in all, this indicates an increasing trend towards chromaticism and a de-emphasis of the primary triad tones.

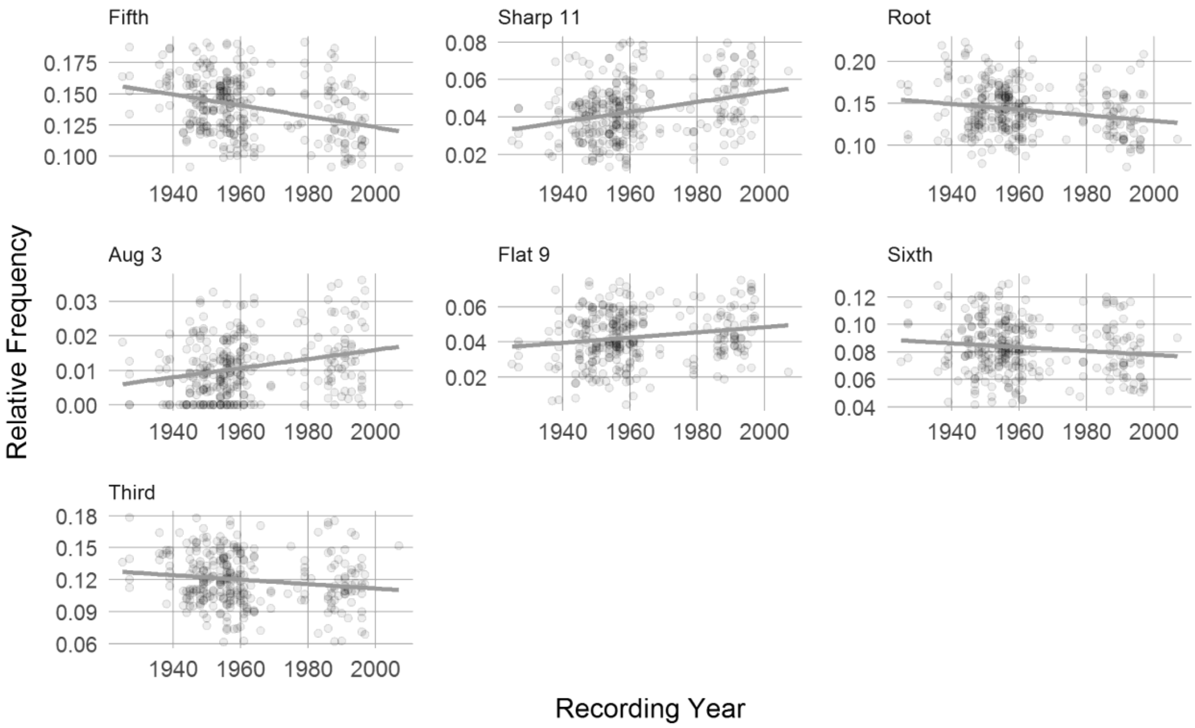

Fig. 4: Significant trends for usage of single pitches classes with respect to the underlying chord.

We also looked at semitone interval distributions and associated features. Two developments stand out, which are related to each other. Firstly, over the course of jazz history, soloists increasingly use larger intervals ($\rho = .347$, $p < .001$, Fig. 5). Before the 1930s, interval sizes beyond 20 semitones were hardly ever played – this also includes intervals between phrases –, whereas since the mid-1980s maximum intervals over three octaves are rather common. This trend is mostly driven by the tenor saxophone players ($\rho = .604$), whereas for all other instruments it is hardly significant ($\rho = .123$, $p = .03$). However, the trend only concerns the extreme intervals in a solo, the overall mean interval size did not change with time ($\rho = .03$, $p = .5$). With a mean interval size of about 2.7 semitones, this indicated that soloists of all times mostly proceed in scalar fashion using mostly semi- and whole-tones.

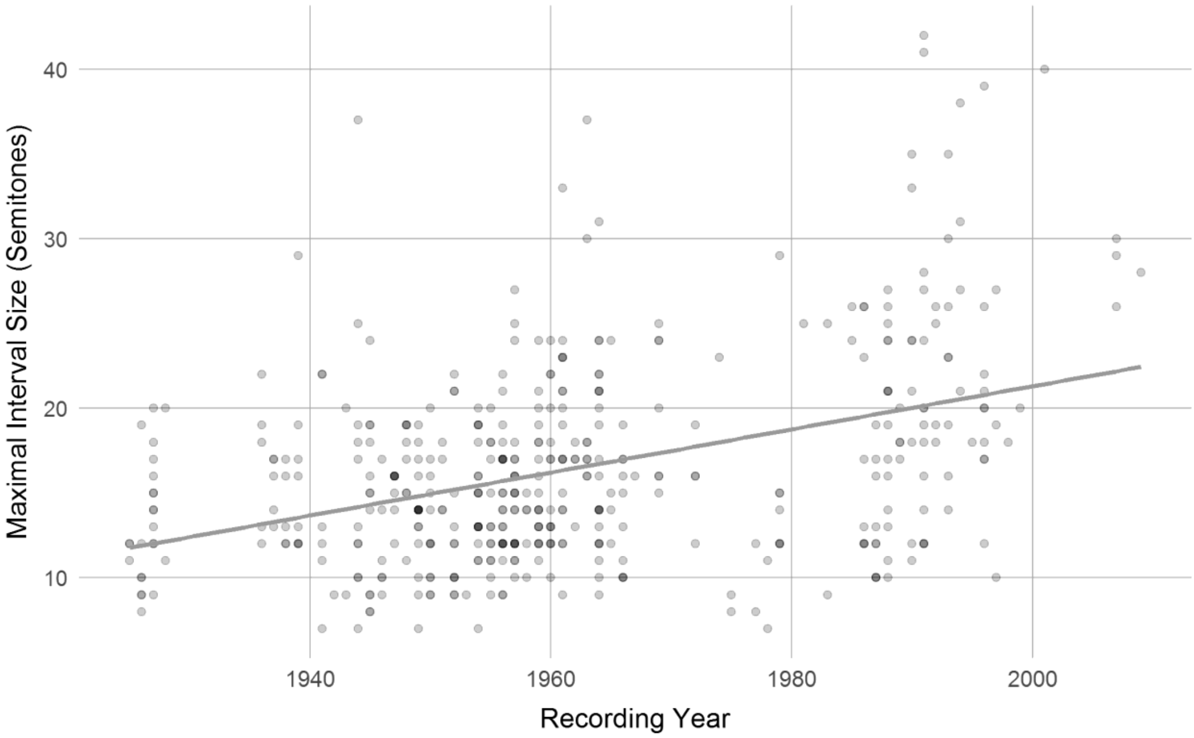

Fig. 5: Maximal interval size over time.

Secondly, the interval variability increased even more as measured by entropy of interval bigrams (ρ = .43, p < .001, Fig. 6), which is the predictability of consecutive interval combinations. The interval (unigram) entropy, i.e., the predictability of single intervals, went only slightly up (ρ = .16, p < .001). This can be explained by the fact that the pre-bebop players rather frequently played thirds, i.e., arpeggios, a technique that fell slightly out of fashion during the era between 1945 and 1965, when a more scalar approach took hold. With the advent of modern postbop in the mid-1980s, the mean interval size and predictability went up again, however not because of the return of the arpeggio, but because of a higher fluidity of all intervals, as indicated by the higher interval bigram entropy.

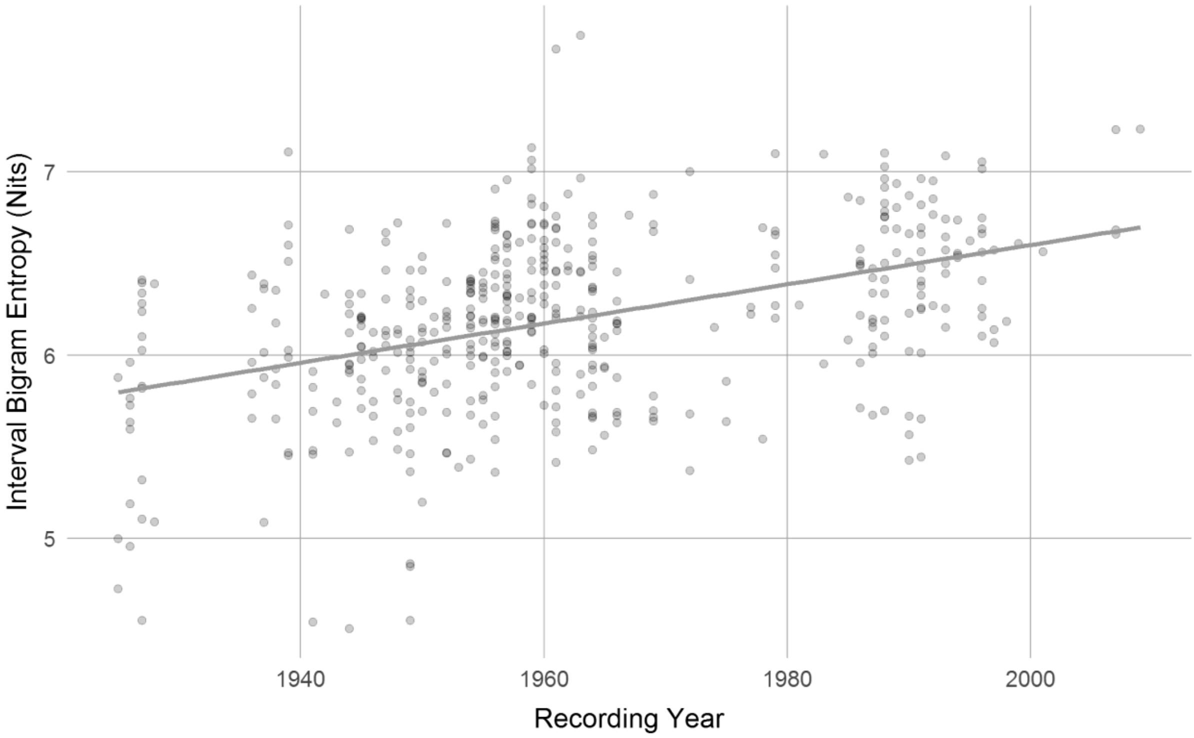

Fig. 6: Interval bigram entropy over time. The entropy is measured in 'nits', i.e., the entropy was measured based on the natural logarithm instead the binary logarithm (bits).

3.4. Tone formation

We also investigated tone formation parameters, which include loudness, intonation and f0-modulation. These parameters were made accessible by an audio algorithm which used onsets and durations of all tones in the WJD to (1) separate the solo voice from the background and (2) to measure these variables from the isolated solos (Abeßer et al 2017). To assess intonation, the overall reference tuning of the audio recording was estimated first, since this sometimes differs from 440 Hz either because of a mistuned studio piano, speed differences of tape machines, or audio post-processing. Twelve-tone equal temperament pitches based on the tuning frequency were used as a reference to assess intonation. We also estimated articulation by using the ratio of tone duration to inter-onset interval, which yields values between 0 (staccato) and 1 (legato).

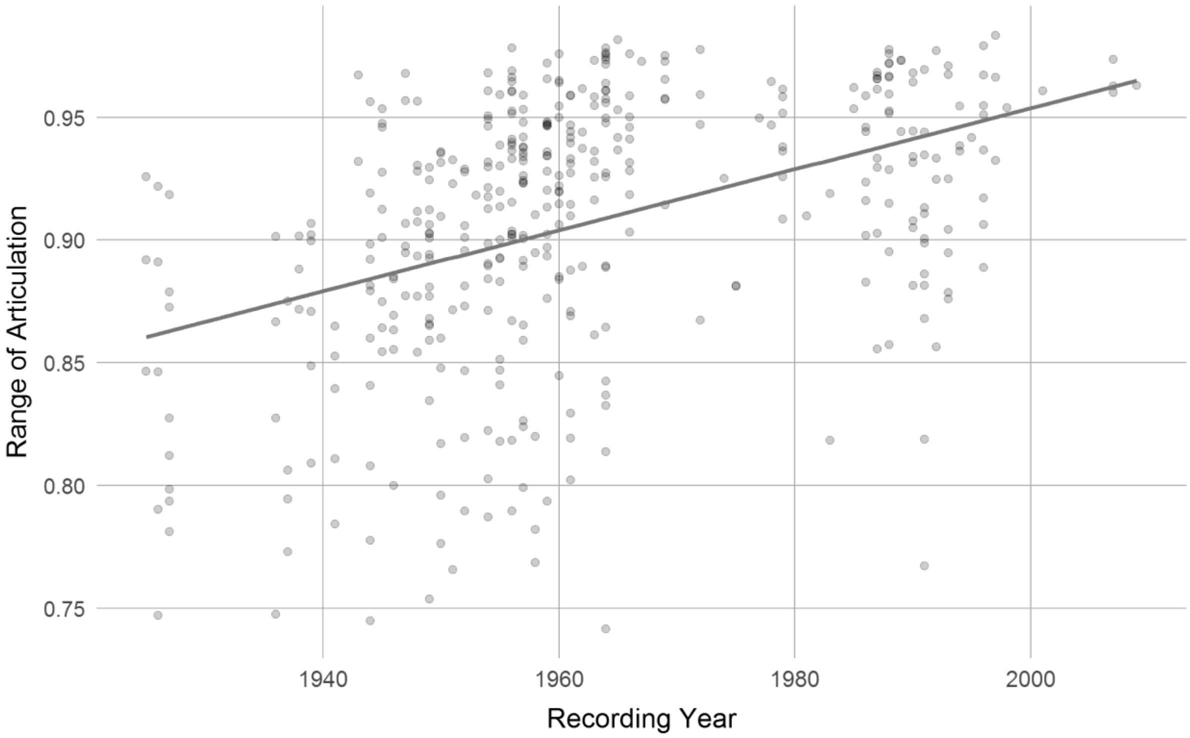

Fig. 7: Articulation range over time. Articulation is defined here as the ratio of durations to inter-onset intervals with no unit.

The strongest trend could be found for articulation, specifically the variability of articulation as measured by the range (difference between maximum and minimum articulation, $\rho = .51$, $p < .001$, Fig. 7) and the standard deviation ($\rho = .29$, $p < .001$). This is a rather unexpected result, since the mean values did not change over time ($\rho = .07$, $p = .11$). This result, though, might be due to an increased need for expressivity resulting in a more variable articulation.

The stability of tone intensities was measured by the standard deviation of intensity across the duration of a tone ($\rho = -.276$, $p < .001$, Fig. 8), which is decreasing with time, hence, the stability is increasing. This might partly be due to the increase of very short tones, which naturally are more stable. Another observation is the increase of intonation precision, as measured by the deviation of median of the pitch frequency (f0) measured over the course of a tone from the nominal frequency of the tone with respect to equal temperament ($\rho = -.159$, $p = .001$, Fig. 9).

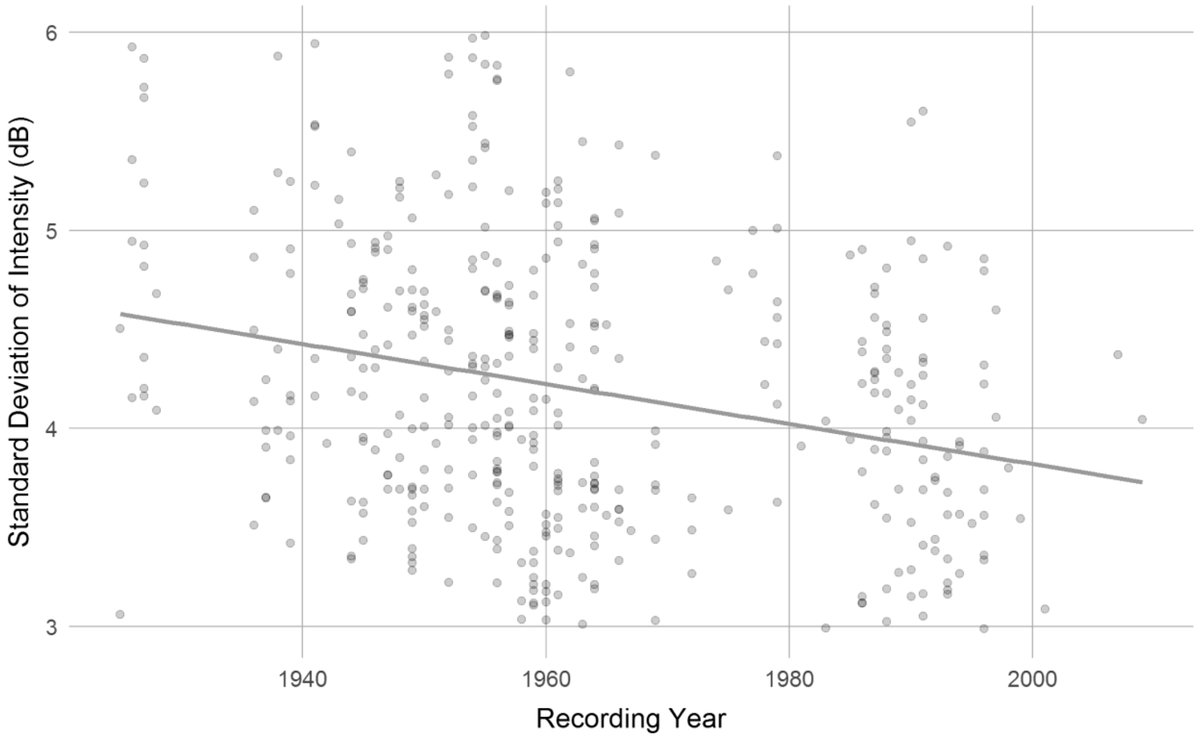

Fig. 8. Intensity stability over time as measured by a the average tone-wise standard deviation of intensity (dB).

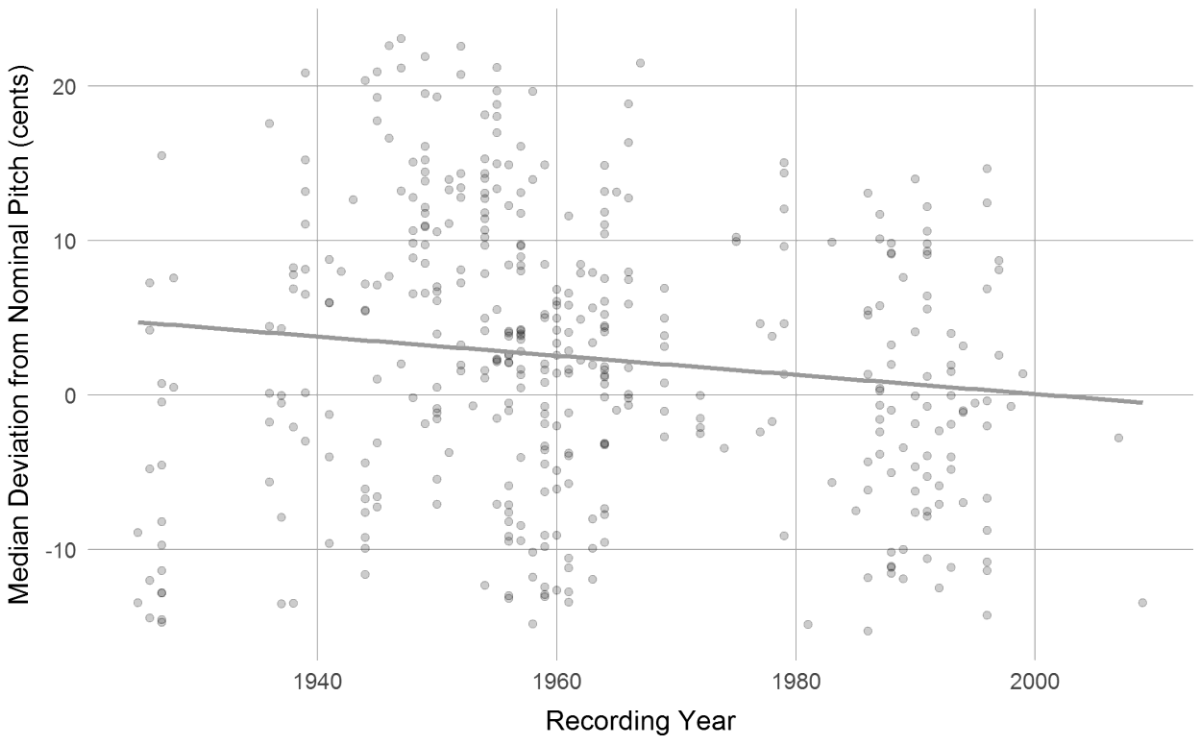

Fig. 9: Median deviation from nominal pitch (in cents, 100 cents = one semitone). For sake of display, 95% quantiles are shown to remove extreme outliers in intonation measurement.

A closer look into the data reveals that before 1940 there is a tendency to play flat, in the 1940s and 1950 a tendency to play sharp, whereas after 1950 no clear tendency can be found (Table 2). One reason for this result might be that intonation in the early days was more flexible as a remnant of blues intonation (or simply a lack of technical skills).

Tab. 2: Median intonation (median deviation from nominal f0) by decade.

Decade	Median intonation (cents)
1920s	-10.5
1930s	4.3
1940s	10.9
1950s	3.7
1960s	1.9
1970s	2.6
1980s	-2.4
1990s	-0.9
2000s	-2.8

3.5. Midlevel units

Midlevel analysis was developed as part of the Jazzomat Research Project to achieve a content-based categorization of the musical surface (Frieler et al 2016). Guiding principle were (hypothetical) underlying playing ideas on a middle time scale of about 2-5 s resulting in various midlevel units (MLUs). After analyzing a larger set of solos, a classification system with nine main MLU types, 18 sub-types and 39 sub-subtypes was devised and codified. Subsequently, all solos in the WJD were annotated with midlevel units by expert annotators and cross-checked. The nine main types are *line*, *lick*, *melody*, *rhythm*, *expressive*, *theme*, *quote*, *fragment*, and *void. Line* refers to melodic lines of mostly uniform rhythm; *lick*s are short concise melodic units (not necessarily stock phrases), which are similar to *melody*, but the latter are conceived to be more singable. In *rhythm* MLUs the rhythmical aspect is foregrounded, typically realized using a single pitch or only a few pitches. For *expressive* MLUs, the timbral aspect is dominating, e.g., howls and honks. Reference to external or internal melodic material is captured with the *theme* and *quote* MLUs. The first means a reference to the theme of the song, and the latter a quotation from any other source. Sometimes, *fragments* are encountered, which are one or two isolated pitches, without any proper melodic characteristics. Finally, deliberate pauses are codified with the *void* MLU.

The relative frequency of the nine main types per solo was examined. For six of the nine main types, significant trends could be observed (Fig. 9). First of all, the amount of *expressive* MLUs was increasing over time ($\rho = .23$, $p < .001$) as well as *rhythm* ($\rho = .21$, $p < .001$), *void* ($\rho = .21$, $p < .001$), and *fragment* MLUs ($\rho = .18$, $p < .001$). Moreover, the usage of *line*s was also trending positively ($\rho = .16$, $p < .001$) whereas *lick*s showed the opposite trend ($\rho = -.19$, $p < .001$). The other three main types (*theme*, *quote*, and *melody*) showed no significant trends. All in all, this indicates a tendency to more expressivity and more diversity. In the early decades, playing short *lick*s was the prevailing mode, with up to 80% of all MLUs. This changed in the 1940s, where *line*s became more prominent. An overview of changes of MLUs by decade with aggregated mean values can be seen in Fig. 11. In this representation, the decrease of *theme* MLUs becomes also significant.

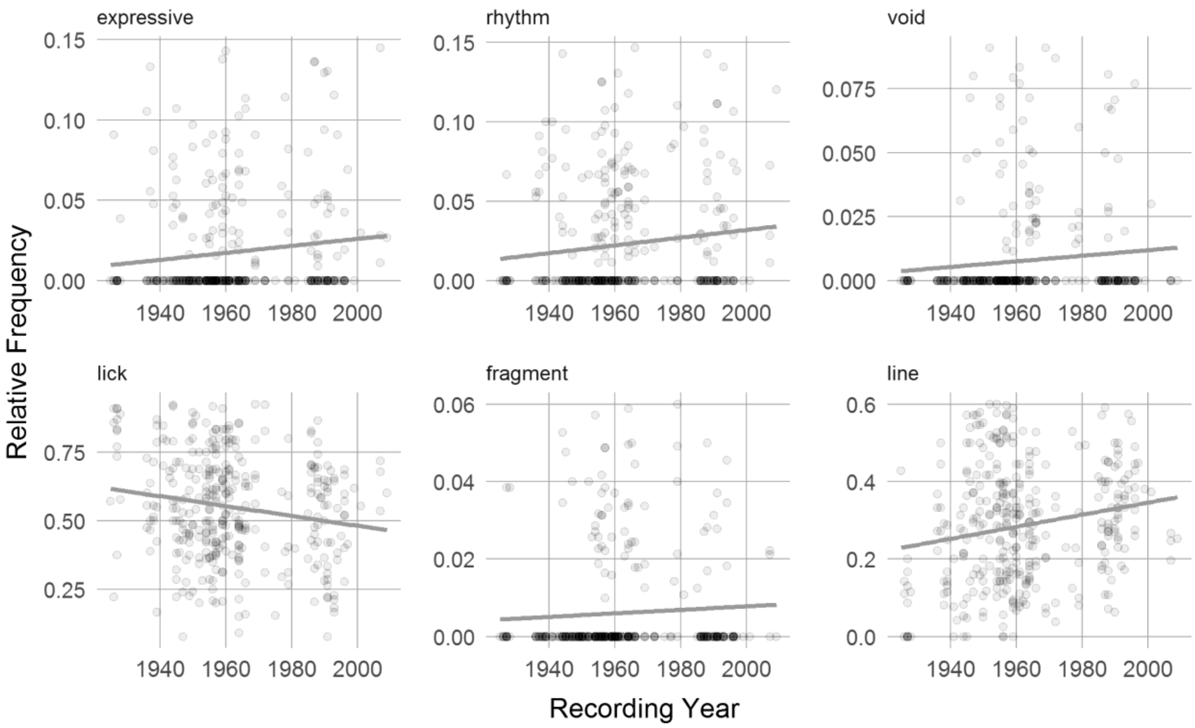

Fig. 10: Significant trends for relative frequency of midlevel units over time.

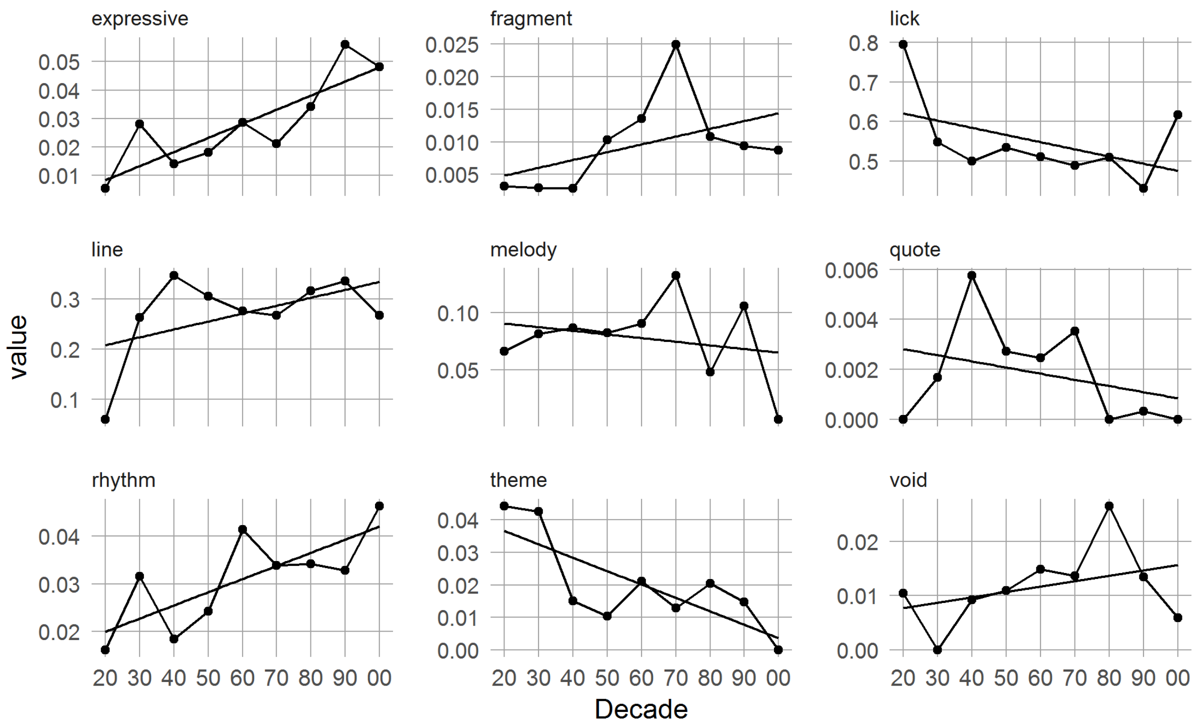

Fig. 11: Average percentage of the nine main types of midlevel units per decade. Note the different scales on the y-axes. Because of averaging, the percentages of different MLU types do not have to add to 1. (The peak of void *MLUs in the 1980s is mostly due to several solos by David Liebman, whereas the peak for* quote MLUs *in the 1940s is due to one solo by Charlie Parker (on Donna Lee). The peak for* fragment *MLUs in the 1970s is coming from several solos by Art Pepper*.)

3.6. Points of maximal change

Our last investigation is dedicated to finding "points of maximal change" within our corpus in regard to recording date, as these might indicate disruptive developments in the approach to improvisation. One would readily expect such a change in the middle of the 1940s with the advent of bebop, which is regarded by many (DeVeaux 1991) as a revolution, or even the moment when (modern) jazz finally realized itself. To this end, we used decision trees (Breiman 1998) to iteratively partition the timeline for a given feature into segments of maximal between-contrast and minimal within-variation. The algorithm automatically finds the best partition for an optimal number of splits by using statistical criteria to balance specificity and prediction error. For each feature in the reduced feature set, we collected the start and end years of each segment found by the decision tree as well as the mean values of the segments. The results can be found in Fig. 12 (median deviation from nominal pitch omitted since it produced only trivial solution by the decision tree).

The effect of this procedure is an approximation of the data with step functions of constant values over a certain time span. In this representation, fast changes and relative stable phases can be identified rather easily. For example, it can be seen how the pitch range quickly expanded from 1925 to 1928, then stayed relatively stable for a few decades before it underwent a sudden acceleration at the end of the 1980s. Of interest is also the development of *line* MLUs, which tripled from 1928 to the end of the 1930s, to drop shortly at the beginning of the 1940s and then to settle on a constant high level at the end of the 1940s. However, the peak at the end of the 1930s is caused by a sampling artefact, because the years 1936 and 1937 only contain solos by Benny Goodman and Lionel Hampton.

To find points of maximal changes, we plotted a histogram of start years of segments (excluding the year 1925) for all features (Fig. 13). The results are quite striking. Five to six clusters can be identified: in the years 1927-28, at the end of 1930s, in the mid of the 1940s, at the end of the 1950s, and between 1985 and 1995. The first cluster reflects the change from traditional jazz to swing. More concretely, the use of chordal thirds decreased in the years from 1927 and 1928, whereas pitch range, pitch entropy and interval bigram entropy increased. Loudness stability increased also, but this might be partly a result of recording quality. The next cluster reflects the transition from swing to bebop. Pitch entropy and pitch standard deviations increase, as well as total duration. The usage of chordal sixth decreased, whereas the use of the #11 increased. There is also a short peak in *line* MLUs in 1936 and 1937, which is, as mentioned earlier, solely due to seven solos by Benny Goodman and three solos by Lionel Hampton, which might indicate a trendsetting role of these performers, but more likely seems to be an artefact of our sample.

The next cluster is located in the mid-1940s, at the advent of bebop. Notable changes here are an increase of b9 and sevenths, pitch entropy and articulation range. Dramatic changes can be seen for *line* (up by 50%) and *lick* MLUs (down by 39%), which did not change since then. On the other hand, chordal roots and fifths strongly declined as well as absolute duration class entropy. All in all, these results seem to neatly reflect the rise of bebop with its extended harmonic thinking and the growing penchant of long lines.

The cluster in the mid to the end of the 1950s is dominated by a huge increase in solo durations, due to the success of long playing vinyl records, and a jump for *rhythm* and *fragment* MLUs to a level which lasts until the end of the timeline. Tonally, the usage of chordal sevenths increased whereas it decreased for thirds and sixths. Also, pitch range and variability and interval size were expanding.

The 1960s saw only a further decline of the chordal fifth, whereas the peaks in the 1970s are both results of an increase in pitch range and pitch entropy.

The mid to end 1980s brought a large increase in expressive MLUs (up by 60%) lasting till the end of the timeline. The usage of chordal roots declined further while the usage of #11 increased.

The cluster at the end of the 1990s is mainly caused by another boost for the total duration of solos (maybe due to the success of the CD with its even longer playing time), whereas duration class entropy sharply declined and the frequency of *very short* durations rose twice by about 10% in 1992 and 1998. In regard to pitch, the b9 further increased as well as pitch range, pitch standard deviation, and absolute interval size.

The large number of changes in this time span is somewhat unexpected and reflects the modern postbop style of Michael Brecker, Bob Berg, the Marsalis brothers, Steve Coleman and others, who introduced a new level of virtuosity and expressivity. It should also be noted that the sample is slightly biased at this time span to (tenor) saxophone players which might explain some of the results with regard to pitch range and absolute interval size, but it still indicates overall improved virtuosity and musical expressivity. This finding suggests that at the end of the 1980s / beginning of the 1990s a significant new style emerged, which seems to be neglected in the jazz literature so far.

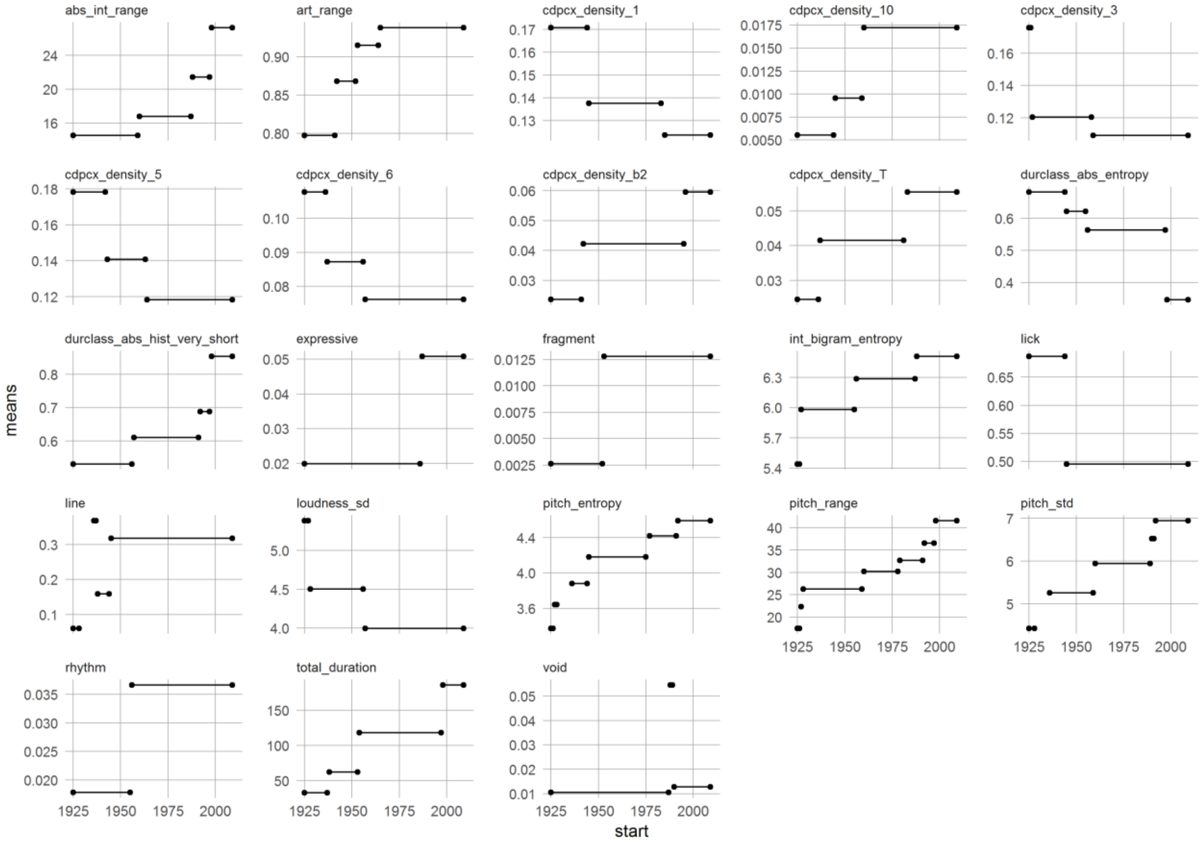

Fig. 12: Segments and mean values for the 23 features of a reduced feature set as discovered statistically by the decisions trees. A segment is indicated by a horizontal line with dots as end points. The height of the line corresponds to the mean value in this segment. See Table 2 and the text for an explanation of the features.

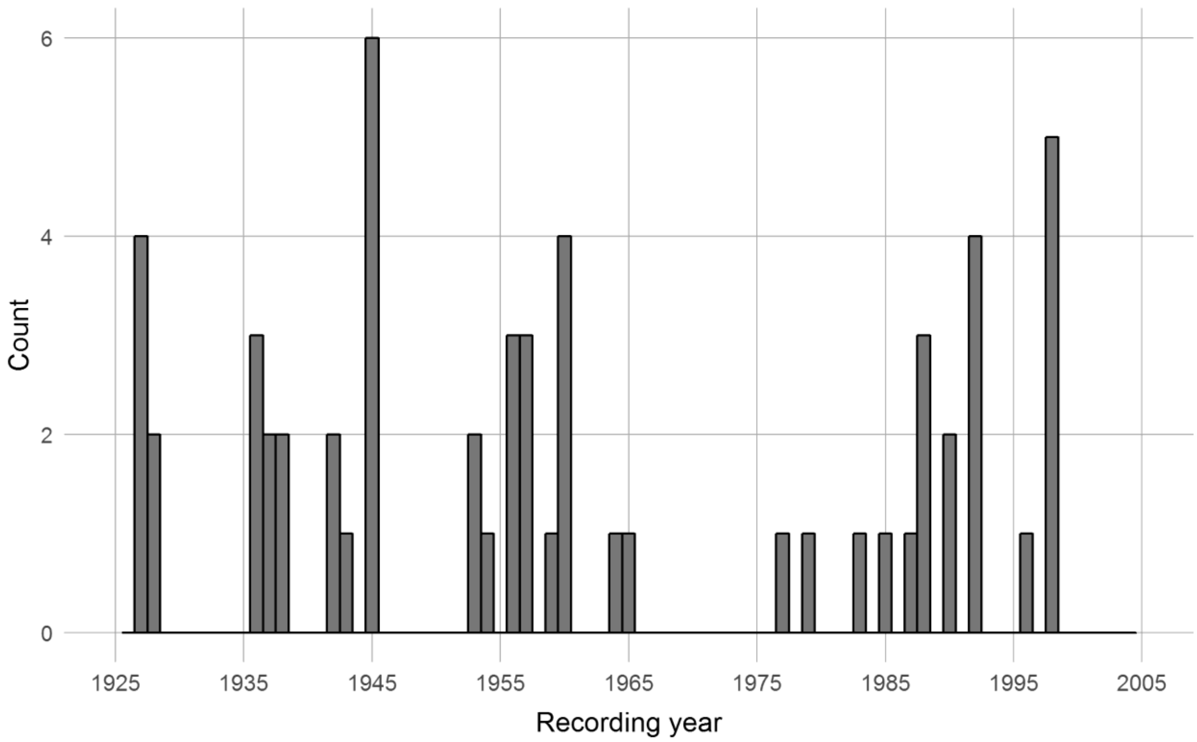

Fig. 13: Distribution of segment start years resulting form decision tree based partition the timeline with respect to the reduced set of features (cf. Fig. 12).

4. Discussion

The main result of our study is that jazz solo improvisation strived to increased complexity over time. Nearly all of the considered features can be interpreted in this regard. To name a few: On the tonal side, an increase in chromaticism can be seen, as measured by an increase in pitch entropy, more frequent usage of non-diatonic tones (e.g., the flatted fifth, the flat ninth) along with a decrease of primary chord tones (root, fifth, and third). Likewise, a tendency to use more interval combinations as well as larger intervals and to use the full instrument range can be observed. Short licks (*lick* MLU) decreased to the advantage of faster and longer lines (*line* MLU), which are cognitively and technically more demanding to play. Generally, all these increases in complexity necessitated ever increasing technical and cognitive skills, which can be interpreted as an increase of virtuosity in jazz. Additionally, these trends go only in one direction and rather unequivocally so.

The question is thus, why did this happen and how can it be explained? Here, we can only speculate. One possibility is that higher virtuosity and complexity are ends in themselves by providing intrinsic reward for mastership. Additionally, this could also serve as a means of distinction from older generations of players; spawned by a need to imitate and impress but also to surpass the previous masters. On the other hand, it could be driven or necessitated by striving for heightened expressivity, as an overall trend for larger expressivity could also be observed indicated by the rise of *expressive, void, fragment*, and *rhythm* MLUs along with a more diversified use of MLUs. The increased tonal complexity can also be interpreted in this way.

The presented results provide a perspective on jazz history which is complementary and, in regard to methodology, independent to common narratives of US jazz history[3]. However, since we only examined improvisation, other important musical parameters (e.g., compositions and accompaniment) that change in conjunction with improvisational style are out of the scope of the present study.

The identification of "points of maximal change" allows some new or more detailed insights, in particular, with regard to the critical time window in the 1930s and 1940s. A rather quick succession of changes in improvisatory techniques could be shown, with a peak in the middle of the 1940s, where bebop started to be documented on recordings. The succession of clusters somewhat neatly follows the standard succession of styles, from traditional to swing at the end of the 1920s, from swing to bebop and later to hard bop and cool jazz during the 1940s, and then to postbop at the end of the 1950s. A rather unexpected result in this context is the number of changes occurring at the end of the 1980s, where a distinctive boost further in direction of higher complexity and virtuosity could be observed. To our knowledge, this change in style was not clearly identified so far. This could be a starting point for further research.

---

[3] However, we have to point out that the evidence is not fully independent since the selection of solos in the WJD was already informed by these narratives.

5. Conclusion and outlook

We are inclined to view our results as a successful demonstration of the potential of computational and corpus-based techniques both for a statistical inspection and specification of historical narratives and for the discovery of musical developments in jazz which might have been garnered little attention so far.

However, all generalizations from our results to a "real" jazz history have to be taken with some caution since they are tied to the specifics of our corpus. The number and selection of solos in the Weimar Jazz Database is – though rather impressive with regard to the effort to create it– still rather small. Text corpora as, for instance, used in literary research, are usually much larger (e.g., a corpus of about 5000 English novels from between 1700 and 1900 used in a study of Franco Moretti and colleagues, see Heuser et al 2016). This is no surprise, since digitizing of text is much cheaper and easier than transcribing music.

We think that the problem of representativeness is hardly solvable, since there might always be some parameter of interest for which a given data set is not fully representative or stratified. But this problem diminishes with increasing size of the data. Therefore, the hopes for computational jazz research certainly lie on automated methods (Müller 2015). Today, transcription algorithms are not yet mature enough to provide jazz solo transcriptions of sufficiently high quality. However, the progress made in the last two decades is promising. We already pursue this line of research in the ongoing research project "Dig That Lick: Analysing Large-scale Data for Melodic Patterns in Jazz Performances" which is an interdisciplinary effort within the transatlantic "Digging-Into-Data Challenge" involving music information retrieval, computational musicology, and jazz research[4] .

Supplementary Material

Please visit http://jazzomat.hfm-weimar.de/feature_history_jazz/ for an interactive website ´to explore and download our data set. There is also the full set of feature available.

[4] https://diggingintodata.org/awards/2016/project/dig-lick-analysing-large-scale-data-melodic-patterns-jazz-performances

Abeßer et al 2017: Jakob Abeßer & Klaus Frieler: Score-Informed Audio Analysis of Jazz Improvisation, in: Martin Pfleiderer, Klaus Frieler, Jakob Abeßer, Wolf-Georg Zaddach & Benjamin Burkhart (eds.): *Inside the Jazzomat. New Perspectives for Jazz Research*, Mainz 2017 (Schott-Campus): 97-131; <http://schott-campus.com/jazzomat/>

Abeßer et al 2016: Jakob Abeßer, Klaus Frieler, Estefania Cano, Martin Pfleiderer & Wolf-Georg Zaddach: Score-Informed Analysis of Tuning, Intonation, Pitch Modulation, and Dynamics in Jazz Solos, in: *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2016: 1-1; <https://doi.org/10.1109/TASLP.2016.2627186>

Breiman 1998: Leo Breiman (ed.): *Classification and Regression Trees*, Boca Raton 1998 (Repr.: Chapman & Hall [u.a.])

DeVeaux 1991: DeVeaux, Scott: Constructing the Jazz Tradition: Jazz Historiography, in: *Black American Literature Forum*, 25 (1991): 525-60

Frieler et al 2015: Klaus Frieler, Kelly Jakubowski & Daniel Müllensiefen: Is It the Song and Not the Singer? Hit Song Prediction Using Structural Features of Melodies, in: *Jahrbuch Musikpsychologie*, 25 (edited by Wolfgang Auhagen, Claudia Bullerjahn, and Richard von Georgi), Göttingen 2015 (Hogrefe-Verlag): 41-54

Frieler et al 2016: Klaus Frieler, Martin Pfleiderer, Jakob Abeßer & Wolf-Georg Zaddach: Midlevel Analysis of Monophonic Jazz Solos. A New Approach to the Study of Improvisation, in: *Musicae Scientiae*, 20/2 (2016): 143-62

Heuser et al 2016: Ryan Heuser, Franco Moretti & Erik Steiner: The Emotions of London, in: *Pamphlets of the Stanford Literary Lab*, Oct.2016; <https://litlab.stanford.edu/LiteraryLabPamphlet13.pdf>

Mauch et al 2015: Matthias Mauch, Robert M. MacCallum, Mark Levy & Armand M. Leroi: The Evolution of Popular Music: USA 1960-2010, in: *Royal Society Open Science*, 2/5 (6 May 2015): 150081-150081 <https://doi.org/10.1098/rsos.150081>

Moretti 2013: Franco Moretti: *Distant Reading*, London / New York 2013 (Verso)

Müller 2015: Meinard Müller: *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*, Berlin / Heidelberg 2015 (Springer)

Pfleiderer 2017: Martin Pfleiderer: The Weimar Jazz Database, in: Martin Pfleiderer, Klaus Frieler, Jakob Abeßer, Wolf-Georg Zaddach & Benjamin Burkhart (eds.): *Inside the Jazzomat. New Perspectives for Jazz Research*, Mainz 2017 (Schott-Campus): 19-41; <http://schott-campus.com/jazzomat/>

Pfleiderer et al 2017: Martin Pfleiderer, Klaus Frieler, Jakob Abeßer, Wolf-Georg Zaddach & Benjamin Burkhart (eds.): *Inside the Jazzomat. New Perspectives for Jazz Research*, Mainz 2017 (Schott-Campus)

R Development Core Team 2008: R Development Core Team: *R: A Language and Environment for Statistical Computing*, Vienna 2008

Selfridge-Field 1997: Eleanor Selfridge-Field (ed.): *Beyond MIDI. The Handbook of Musical Codes*, Cambridge, MA 1007 (The MIT Press)

Serra et al 2011: Joan Serrá, Gopala K. Koduri, Marius Miron & Xavier Serra: Assessing the Tuning of Sung Indian Classical Music, in: *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami 2011: 157-162

Shannon 1948: Claude E. Shannon: A Mathematical Theory of Communication, in: *Bell System Technical Journal*, 27/3 (1948): 379-423 and 623-656

Weiß et al 2018: Christoph Weiß, Matthias Mauch, Simon Dixon & Meinard Müller: Investigating style evolution of Western classical music: A computational approach. Musicae Scientiae (Online First). URL: <http://journals.sagepub.com/doi/abs/10.1177/1029864918757595>

Abstract

Typically, jazz history is written in an integrated fashion, mixing biographical accounts of eminent figures, descriptions of sociological and cultural context and genuine musical characterizations. Purely musical accounts of jazz history are rare – and much less so for improvisational styles and techniques. Individual analyses of specific players are more common but they often focus on traits that make a player outstanding, revolutionary or influential. In this study, we attempt to trace the evolution of certain features of (monophonic) jazz improvisations over the course of about 90 years, from the 1920s until the present time using a large sample of 456 high-quality solo transcriptions from Weimar Jazz Database and the analytical software MeloSpyGUI, both developed within the Jazzomat Research Project. To this end, for each solo a comprehensive set of scalar features will be extracted that capture aspects of solo improvisations such as tonal and rhythmic complexity, pitch range, and interval content including midlevel playing ideas and sound aspects. These features are correlated with recording year and trend lines were fitted. A selection of most significant trends pertaining to a subset of 24 features are presented and discussed. Overall, trends to higher complexity and virtuosity could be observed in all dimensions. Finally, we identified points of maximal changes in the features and were able to reconstruct the well-known succession of jazz styles, but also found evidence that in the late 1980s a new style emerged. Finally, implication and limitations of the study are discussed.

About the author

Klaus Frieler graduated in theoretical physics (diploma) and received a PhD in systematic musicology in 2008 from the University of Hamburg. He worked as a freelance software developer for several years, before taking up a post as a lecturer in systematic musicology at the University of Hamburg in 2008. In 2012, he spent a brief period at the Centre for Digital Music, Queen Mary University of London. From 2012 to 2017, he has been working as a post-doctoral researcher with the Jazzomat Research Project at the University of Music "Franz Liszt" Weimar. After a short intermezzo at the Max Planck Institute for Empirical Aesthetics in Frankfurt/Main, he is now a post-doctoral research with the transatlantic Dig That Lick Project. His main research interests are computational and statistical music psychology with a focus on jazz research, creativity, melody perception, and singing intonation..