Proceedings of the 9th Conference on Interdisciplinary Musicology – CIM14. Berlin, Germany 2014

Dynamics in jazz improvisation - score-informed estimation and contextual analysis of tone intensities in trumpet and saxophone solos

Jakob Abeßer[1,2], Estefanía Cano[1], Klaus Frieler[2], Martin Pfleiderer[2]

[1] *Semantic Music Technologies, Fraunhofer IDMT, Ilmenau, Germany*

[2] *Jazzomat Research Project, Liszt School of Music, Weimar, Germany*

Correspondence should be addressed to: jakob.abesser@idmt.fraunhofer.de

Abstract: **In this paper, we aim at analyzing the use of dynamics in jazz improvisation by applying score-informed source separation and automatic estimation of note intensities. A set of 120 jazz solos taken from the Weimar Jazz Database covering many different jazz styles was manually transcribed and annotated by musicology and jazz students within the Jazzomat Research Project. In order to enrich these symbolic parameters with note-wise intensity annotations, the solo instrument tracks are extracted from the original audio files by applying a pitch-informed separation algorithm that uses the manual transcriptions as prior information. Subsequently, the magnitude envelope and spectral energy are analyzed in order to extract intensity measures for all note events in the solo. Next, we investigate how dynamics are used as a stylistic tool in jazz improvisation. To this end, we analyze how the note intensity values correlate with contextual information encoded in the note's pitch, duration, position within a musical phrase, perceptual accents, and structural markers. Additionally, we compare the use of dynamics among different instruments (alto and tenor saxophone, trumpet, and trombone). The results of this interdisciplinary study have implications for jazz research, jazz education, performance research, as well as for Music Information Retrieval fields such as automatic music transcription and source separation.**

1. Introduction

1.1. Motivation

Dynamics are a crucial dimension for any music performance (*e.g.*, [1, 2]). Musicians give liveliness to music by playing different phrases with differing degrees of intensity or by accentuating single tones by playing them louder, *i.e.*, "local stresses" or "phenomenal accents" according to [3]. Additionally, longer tones could be played with subtle changes of dynamics.

Presumably, dynamics are shaped following various intentions and according to several implicit syntactical and expressive rules. On the one hand, musicians could strengthen various metrical or structural aspects of a certain piece by stressing metrically or structurally salient tones with additional intensity (see [3, 76]). On the other hand, if certain tones of a melodic line are played louder than others, the stressed tones could form an additional overlaid rhythmical component (see [4]). This is a common strategy in African music, jazz, or rock and pop music. For example, many jazz musicians such as seminal jazz saxophonist Charlie Parker or clarinet and soprano saxophone player Sidney Bechet are claimed to deliberately accentuate off-beats (every second eighth note) or use cross-rhythmic superposition (*e.g.*, by stressing every third eighth note) in their improvisations (cf. [5]).

However, dynamics are often neglected in jazz research, because it is a hard task to reliably discern and annotate dynamic differences within a melodic line of a single musician from recordings of ensemble performances. And while it is quite easy to detect the overall dynamics of a recording automatically, it is very hard to detect dynamics of one musician from an ensemble recording except when single tracks of a multi-track recording are available.

1.2. Research Goals

In this paper, we introduce a new method for the detection of dynamics within melodic lines from commercial jazz recordings. The analysis is based on transcriptions of monophonic jazz improvisations from the *Weimar Jazz Database* (cf. section 3), which are created aurally/manually within the Jazzomat Research Project. A score-based source separation algorithm is applied to the original ensemble recordings in order to isolate audio tracks with only the soloist playing (cf. section 2.1). Based on the isolated audio track, the note intensity values are estimated as will be shown in section 2.2. Since algorithms for automatic source separation can produce audible artifacts, the robustness of the note intensity estimation is evaluated in section 2.3 by comparing intensity values extracted from score-based separated tracks with intensity values extracted from perfectly isolated multi-track recording tracks.

In the second part of this paper, the isolated melodic lines of 120 solos by 44 jazz musicians are explored statistically with regards to their structure of dynamics (section 4). In particular, we are looking for overall tendencies and regularities of dynamics with regards to pitch, duration, onset, and position within a phrase (sections 4.2, 4.4, and 4.5), as well as for correlations between the phenomenal dynamic structure and metrical accents according to various accent rules (section 4.6), and for the stress of off-beats within lines of eights through intensity as sometimes asserted for Charlie Parker and jazz phrasing in general (section 4.7). Finally, some conclusions with regards to future music performance research are drawn in section 5.

1.3. Related Work

As discussed in [6], most algorithms for automatic music transcription do not include a loudness estimation stage. The main reason for that is the lack of reliable ground truth annotations for evaluation. Electric keyboards that allow to record MIDI velocity values are a potential solution, since this parameter is directly related to the note intensity. However, this approach is not transferable to other instruments. In the field of expressive performance analysis, several authors tried to estimate intensity from isolated instrument recordings. For instance, Ren et al. extract the note-wise perceptual loudness values as part of a real-time analysis framework [7]. Ewert and Müller propose to estimate note intensity values from spectrogram representations of polyphonic piano recordings after aligning given score information to audio performances using Dynamic Time Warping [8].

2. Proposed approach

2.1. Score-informed Source Separation

The algorithm for pitch-informed solo and accompaniment separation presented in [9] was used to perform separation of the solo instrument. As initially proposed, the algorithm automatically extracts pitch sequences of the solo instrument and uses them as prior information in the separation scheme. In order to obtain more accurate spectral estimates of the solo instrument, the algorithm creates tone objects from the pitch sequences, and performs separation on a tone-by-tone basis. Tone segmentation allows more accurate modeling of the temporal evolution of the spectral parameters of the solo instrument. The algorithm performs an iterative search in the magnitude spectrogram in order to find the exact frequency locations of the different partials of the tone.

A smoothness constraint is enforced on the temporal envelopes of

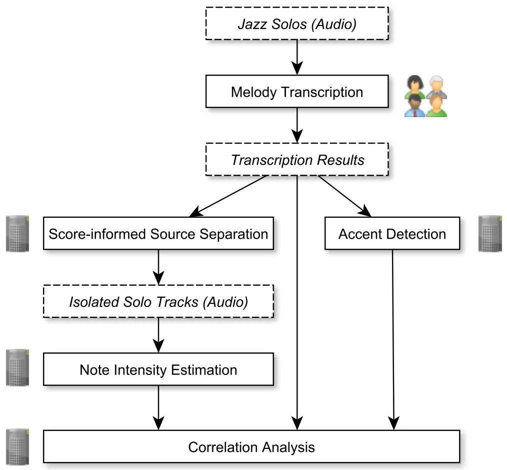

Figure 1: Flowchart of the proposed approach. The individual processing steps are detailed in section 2. While the melody transcription is performed manually, all other processing steps are performed automatically.

each partial. In order to reduce interference from other sources caused by overlapping of spectral components in the time-frequency representation, a common amplitude modulation is required for the temporal envelopes of the partials. Additionally, a post-processing stage based on median filtering is used to reduce the interference from percussive instruments in the solo estimation.

As detailed in section 3, pitch information is taken from manual solo melody transcriptions. Hence, the automatic pitch extraction stage in the separation algorithm is bypassed and the tone objects taken from the manual transcriptions are used as prior-information in the separation scheme.

2.2. Estimation of Note Intensity Values

As a result of the score-informed source separation, we obtain an audio track with the solo instrument being isolated from the other instruments, which will be referred in the following as *solo track*. This section explains how we obtain note intensity values for all notes in the solo. We follow the approach proposed in [10].

The (monaural) solo track is processed with overlapping time frames with a hopsize of 480 samples and a blocksize of 512 samples. The sampling rate is $f_s = 44.1$ kHz. The Short-time Fourier Transform $X(k,n)$ is computed with n denoting the time frame and k denoting the frequency bin.

We compute the intensity $L(i)$ of the i-th note as follows. From the power spectrogram $|X(k,n)|^2$, we first compute *band-wise intensity* values $I_b(n)$ for each of the $N_b = 24$ critical bands (with the indices $b \in [1:N_b]$) as

$$I_b(n) = 10\log_{10} \sum_{k\in[k_{\min,b}:k_{\max,b}]} |X(k,n)|^2 . \tag{1}$$

$k_{\min,b}$ and $k_{\max,b}$ denote the frequency bins that correspond to the lower and upper boundaries of the b-th critical band.

In the next step, the band-wise intensity values $I_b(n)$ are accumulated over all bands as

$$I_{\mathrm{acc}}(n) = \sum_{b\in[1:N_b]} 10^{\frac{I_b(n)}{10}} . \tag{2}$$

Finally, the *frame-wise intensity* value in the n-th frame is computed as

$$L(n) = 90.302 + 10\log_{10} I_{\mathrm{acc}}(n). \tag{3}$$

In order to compute *note-wise intensity* values $L(i)$, we take the highest frame-wise intensity value over the duration of the i-th note as

$$L(i) = \max_{n\in[n_{\mathrm{on},i}:n_{\mathrm{off},i}]} L(n). \tag{4}$$

$n_{\mathrm{on},i}$ $n_{\mathrm{off},i}$ denote the time frames that correspond to the onset and offset time of the i-th note. We assume a strictly monophonic melody without any note overlap.

2.3. Robustness of the Note Intensity Estimation within a Source Separation Context

Source separation algorithms can lead to audible artifacts in the separated audio tracks [9]. Hence, we wanted to investigate to what extend these artifacts affect the computation of note intensity values on the isolated solo instrument track.

In this experiment, we analyzed audio tracks from a multi-track recording session of the Jazz standard “You And The Night And The Music” (performed in September 2014 at the Liszt School of Music). We could access both the isolated tracks (without any cross-talk between the instruments) as well as a professional mix performed by a sound engineer. In particular, we analyzed two solos from the electric guitar and the trumpet. The solos were first manually transcribed by music experts in the same manner as described in section 3. Based on the given solo transcriptions, we applied the source separation procedure as explained in section 2.1 in order to separate the solo parts from the mix for both instruments. Also, we obtained the corresponding solo parts from the original multi-track session tracks for each instrument.

Then, we computed the note-wise intensity values $L(i)$ as described in section 2.2 for each solo over the (automatically) isolated solo track and the (perfectly isolated) multi-track session track of the corresponding instrument. Table 1 summarizes the results obtained for the guitar and trumpet solo. The correlation coefficient r and root mean square error (RMSE) between the intensity curves computed from the automatically isolated solo track and the corresponding multi-track session instrument track are given.

Figure 2 illustrates an example excerpt taken from the analyzed guitar solo. It can be observed that despite of some local variations, the general intensity trend is barely affected by the source separation artifacts. Hence, we assume that the proposed method for automatic estimation of note intensities based on automatically separated solo tracks is a solid basis for the statistical evaluations presented in the following sections.

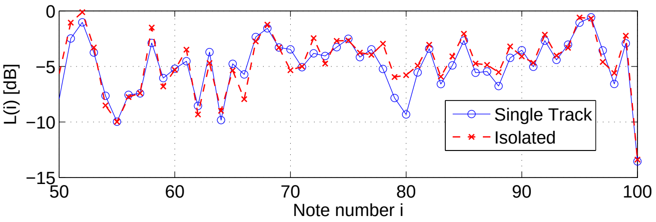

Figure 2: Note-wise intensity values $L(i)$ obtained from an excerpt of the isolated guitar solo track (“Isolated”) and from multi-track session solo track (“Single Track”).

Instrument	r ($p < .0001$)	RMSE	Average Note Pitch
Guitar	0.817	1.944	65.1
Trumpet	0.792	2.106	69.4

Table 1: Correlation coefficient r and root mean square error (RMSE) between the intensity curves computed from the isolated solo track and the original instrument track from the multi-track session. In addition, the average note pitch is given for each solo.

3. Dataset

120 jazz solos by 44 performers taken from the *Weimar Jazz Database* (WJazzD) were investigated. The Weimar Jazz Database

is publicly available at `http://jazzomat.hfm-weimar.de/`. For our explorative approach it seemed fit to restrict ourselves to the most important jazz wind instruments, *i.e.*, alto sax, tenor sax, and trumpet. We added trombone to this set in order to include a low register brass instrument. The distribution of instruments in our dataset can be found in Table 2 and the list of artists and solo counts in Table 3.

The database contains high-quality transcription of jazz solos from various artists and styles. The transcriptions were manually performed and cross-checked by musicology and jazz students at the Liszt School of Music. Each transcription contains basic note annotations (pitch, onset, duration) as well as contextual annotations (metrical structure, beat times, chords, phrases, form parts, choruses). Due to copyright restrictions, only the transcriptions are published. However, the corresponding audio recordings can be identified with the given MusicBrainz-ID's and the given solo start and end times.

Table 2: Distribution of instruments in the dataset.

Tenor Sax	Trumpet	Alto Sax	Trombone	Total
51	40	23	6	120

Table 3: Overview of performers, number of solos, and instruments in the dataset.

Performer	#	Inst.	Performer	#	Inst.
Art Pepper	2	as	Ben Webster	3	ts
Benny Carter	2	as	Bob Berg	4	ts
Buck Clayton	2	tp	Cannonball Adderley	4	as
Charlie Parker	2	as	Chet Baker	6	tp
Chu Berry	1	ts	Clifford Brown	4	tp
Coleman Hawkins	3	ts	Curtis Fuller	2	tb
David Murray	3	ts	Dexter Gordon	4	ts
Dizzy Gillespie	3	tp	Don Byas	3	ts
Don Ellis	2	tp	Eric Dolphy	1	as
Freddie Hubbard	5	tp	Hank Mobley	1	ts
Harry Edison	1	tp	J. J. Johnson	2	tb
Joe Henderson	4	ts	John Coltrane	3	ts
Joshua Redman	4	ts	Kenny Dorham	3	tp
Kenny Garrett	2	as	Lee Konitz	2	as
Lee Morgan	1	tp	Lester Young	2	ts
Michael Brecker	1	ts	Miles Davis	6	tp
Nat Adderley	1	tp	Paul Desmond	6	as
Roy Eldridge	2	tp	Sonny Rollins	6	ts
Sonny Stitt	1	as	Steve Coleman	2	ts
Steve Turre	2	tb	Warne Marsh	2	ts
Wayne Shorter	4	ts	Woody Shaw	3	tp
Wynton Marsalis	1	tp	Zoot Sims	2	ts

4. Results

4.1. Data Analysis

Due to the absence of an independently evaluated gauge for the extracted intensities, we decided to work in a solo-based manner, *i.e.*, we avoided to pool intensity data across solos, if not justified by single tests. This meant that a large number of statistical tests (mostly Wilcoxon rank tests and Kendall rank correlation) had to be carried out. We addressed the problem of multiple testing by using second-order statistics, *i.e.*, statistics of p-values from single tests. Furthermore, large differences between solos and performer with respect to intensity shaping can be already expected from the outset due to personal, instrumental, stylistic, and other reasons. Thus, solo-wise comparison seem to be an adequate approach to examine these differences. However, in most cases the results from the multiple tests were in quite good agreement with the results from global tests, indicating that the extracted intensity values might be sufficiently consistent across different solos. If this was the case, we also resort to global tests and plots to facilitate our way of presentation.

In order to ease comparison and discard outliers, we normalized the intensity data solo-wise by mapping the 5%-95% percentile to the interval of [0,1], which resulted in a distribution of the medians of relative intensity between 0.44 and 0.81 (SD=0.076, IQ=0.11), with a median of medians of 0.594, which seems to be sufficiently close to the midpoint of the normalized scale.

We make frequently use of logarithmic Bayes factors (log-BF) in the context of multiple testing. Logarithmic bayes factors are defined here as $\log_{10}\mathrm{BF}_\alpha = \log_{10} N_{obs;\alpha}/N_{exp;\alpha}$ for a certain significance level α, where $N_{obs;\alpha}$ is the number of observed significant tests, and $N_{exp;\alpha}$ is the number of expected significant tests by chance alone. Also, to compact information further, we use average log-BFs where averaging is done over a set of tests on a fixed range of significance levels from 0.05 down to 10^{-6}.

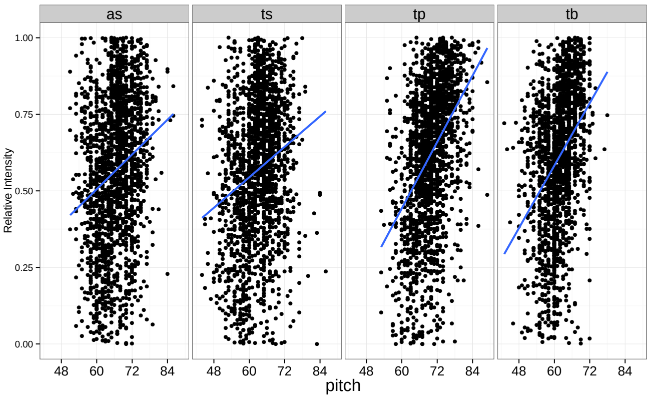

Figure 3: Scatterplots of pitch (MIDI units) vs. relative intensity (trimmed to the range of [0,1] and thinned out for displaying purposes) by instruments (as=alto sax, ts=tenor sax, tp=trumpet, tb=trombone). Linear fits are shown in blue. The positive correlation of pitch and intensity is more pronounced for brass instruments (tp, tb) than for reed instruments (as, ts).

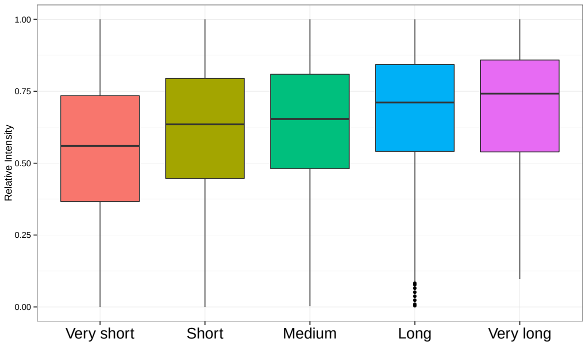

Figure 4: Boxplots of relative duration classes vs. relative intensity (trimmed to [0,1] for displaying purposes).

4.2. Correlation with Pitch

First, correlations between relative intensity and pitch height are explored. To this end, we carried out 120 Kendall rank correlation tests, from which 107 became significant at the 5%-level, 100 at the 1%-level, and 94 at the 0.1%-level, with a mean log-BF of 3.44. Hence, a highly significant but moderate correlation of pitch height and intensity can be found across all players ($\tau = 0.188, p < 0.000$). This can partly be explained with instrument specificities. The correlations are about twice as strong for trumpets ($\tau = 0.352, p < 0.000$) and trombones ($\tau = 0.324, p < 0.000$) than for alto ($\tau = 0.162, p < 0.000$) and tenor saxophones ($\tau = 0.139, p < 0.000$), cf. Figure 3. However, there were some exceptions, where pitch was even anti-correlated with relative intensity on the 0.001-level with a mean correlation coefficient of about $\overline{\tau} = -0.18$, *e.g.*, Bob Berg (2 out 4 solos), Coleman Hawkins (1/3), Joshua Redman (2/4), Steve

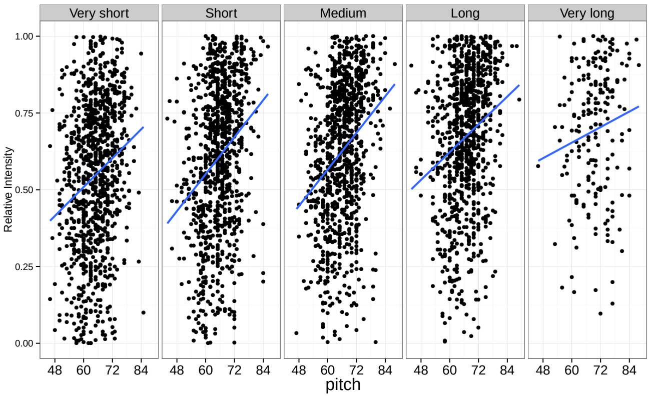

Figure 5: Scatterplot of relative duration classes vs. relative intensity (trimmed to [0,1] and thinned out for displaying purposes) by relative duation classes. Linear fits are shown in blue. Median correlations coefficients were $\overline{\tau}_{\text{very short}} = 0.299$, $\overline{\tau}_{\text{short}} = 0.338$, $\overline{\tau}_{\text{medium}} = 0.427$, $\overline{\tau}_{\text{long}} = 0.551$, $\overline{\tau}_{\text{very long}} = -0.617$ for the 90, 87, 60, 19 and 1 solo(s) resp. with significant correlations at the 5% level.

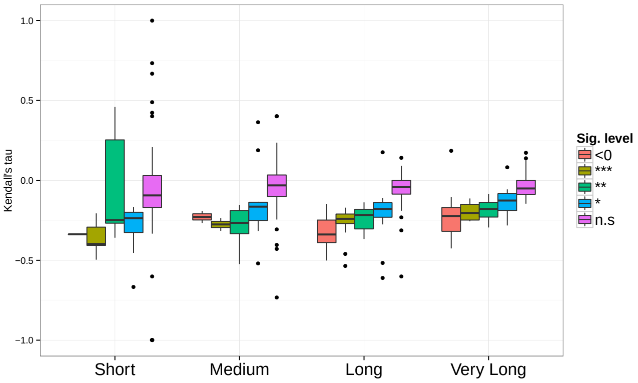

Figure 6: Boxplot of correlation coefficients (Kendall's τ) for phrase duration classes for 5 different significance levels.

Coleman (1/2), Miles Davis (1/6), Paul Desmond (1/6), Wynton Marsalis (1/1).

4.3. Correlation with Duration and Duration Classes

Second, relative intensity and duration classes were correlated. Here, duration classes have five gradations ("Very Short", "Short", "Medium", "Long", and "Very Long") as compared to a reference time T_r, which was either the (local) beat duration (`durclass_rel`) or 500 ms (=120 bpm) (`durclass_abs`). Classes (starting with $n = -2$ for "very short") are the intervals $[2^{\frac{n-1}{2}} T_r, 2^{\frac{n}{2}} T_r]$ with extension to ±∞ for the end classes. We correlated with the class index n and received 80 solos being significant at the 5%-level, 68 at the 1%-level and 51 at the 0.1%-level for absolute duration classes, and 91, 80, and 66 solos, respectively, for relative duration classes. Mean correlation coefficients were $\tau_{abs} = 0.192$ and $\tau_{rel} = 0.219$, with mean log-BFs of 3.25 and 3.12. Overall correlations were consequently a bit weaker $\tau_{abs} = 0.111$ and $\tau_{rel} = 0.147$. There were no exceptions with respect to the direction of the correlation. In Figure 4 one clearly sees, how relative intensity raises with duration class. Furthermore, we checked the interaction between pitch, duration class and relative intensity. There was clear trend of higher correlation of intensity with higher pitches and longer durations, except for very long durations, for which barely any correlation became significant (cf. Figure 5). Also, raw, unclassified duration correlated with relative intensity with a mean correlation of $\overline{\tau} = 0.175$ and 106, 93 and 81 correlations at the 5%, 1%, and 0.1% level respectively, yielding a mean log-BF of 3.36.

4.4. Correlation with Relative Position in Phrase

Third, relative position in a phrase and relative intensity are correlated. One might expect correlations here for two reasons: general musical expressiveness and shortness of breath. In the second case, the correlations should be negative. Furthermore, they should become stronger and more frequent with phrase duration. We classified the phrase durations in four classes, "Short", "Medium, "Long" and "Very Long" according to the 1st to 3rd quartiles (1.004, 1.895, 3.288 seconds) of the overall phrase length distribution, and normalized the tone position in a phrase to values in the interval from 0 to 1. As expected, there were several significant correlations, increasing with phrase duration. On the 5% level, we found 20, 25, 53, and 81 significant correlations, whereas on the 0.1%-level, the corresponding sequence was 6, 4, 25, and 40. Overall log-BFs were 2.00, 1.86, 2.66 and 3.05. The pattern is in good concordance with the shortness of breath hypothesis. Generally, we found that the more significant the correlation, the larger is the absolute correlation coefficient. However, the correlation coefficient is located exclusively in the small to medium negative range (mean correlation across all classes and significance levels: $\overline{\tau} = -0.12$). Hence, relative intensity of phrases tends slightly to go down—if there is a trend (cf. Figure 6).

The natural follow-up question is then: Are there any performer or instrument specific patterns? To investigate this, we defined a group of 18 "short breathers", who showed significant correlations on the 5% level for "Long" and "Very Long" phrases in more than two-thirds of their solos, see Table 4. Indeed, the mean correlations coefficient for this group is $\overline{\tau} = -0.21$ for "Long" and "Very Long" phrases, whereas for the remaining soloists, it is only $\overline{\tau} = -0.10$, whereas for "Short" and "Medium" phrases the correlations did not significantly differ. Hence, "short breathers" might not be a misnomer. However, the effect is mostly quite small, except for Don Byas, which might be related to the fact that his three solos are all in slow or medium slow tempo, and falling in intensity during a phrase could as well be an expressive tool in context of a ballad. Furthermore, for different instruments no significant correlation patterns could be found, corroborating the fact that this is a performer-related effect.

Performer	$\overline{\tau}_L$	$\overline{\tau}$	Number of Solos
Don Byas	-0.42	-0.37	3
Hank Mobley	-0.26	-0.12	1
Coleman Hawkins	-0.24	-0.10	3
Freddie Hubbard	-0.24	-0.13	5
Clifford Brown	-0.24	-0.17	4
David Murray	-0.23	0.01	3
Buck Clayton	-0.23	-0.18	2
Benny Carter	-0.21	-0.10	2
Sonny Stitt	-0.21	-0.12	1
Dizzy Gillespie	-0.20	-0.11	3
Art Pepper	-0.18	-0.10	2
Don Ellis	-0.17	-0.09	2
Chu Berry	-0.16	-0.14	1
Steve Turre	-0.15	-0.17	2
Cannonball Adderley	-0.15	-0.16	4
Joe Henderson	-0.15	-0.15	4
Steve Coleman	-0.14	-0.07	2
Eric Dolphy	-0.12	-0.13	1

Table 4: Mean correlation for the group of short breathers with at least two-third significant correlations of all possible correlations. Second column shows the average rank correlations for "Long" and "Very Long" phrases, and third column for all phrase duration classes.

4.5. Correlation with Onsets

Finally, we correlated relative intensity with the onsets of tones to reveal global trends in intensity change, complementing the phrase-based analysis. We found a large amount of significant correlations, 62, 47, and 25 for the 5%, 1%, and 0.1% level and an overall mean log-Bayes factor of 2.87. As indicated by the log-Bayes factor, this is actually a rather large effect—there are still 13 solos with significant correlations to be found at the 10^{-6} level. However, the directions of correlations are very diverse with a median of $\tau = -0.08$, $SD_\tau = 0.139$ and ranging bimodally from -0.275 up to 0.23. Inspection of differences between performer, style, tempo class, rhythm feel, and tonality type using Kruskal-Wallis tests did not reveal any systematic connection. Likewise, no correlation with the total duration of the solo was found. Only a slight trend for high tempo to rise and for slow tempo to drop in intensity could be observed, but became not significant in a Kruskal-Wallis test. Hence, it seems to be a strong but very solo-specific effect, possibly a result of spontaneous interaction with the rhythm group or of unobserved performer-related variables.

4.6. Accent Rules and Structural Markers

In [11] a large set of accent rules taken from the literature was compared with experimental data of perceived accents for pop melodies. These (melodic) accent rules are mostly formulated in a way, that make them equivalent to binary structural markers, which evaluate to "true" only at certain locations in a melody as defined by the rule, and "false" at all other locations. Examples for such special locations are the downbeat of a bar or the pitch peak in a line.

Accents rules can be classified in 6 classes (cf. [11]): duration, pitch jump, contour, meter, harmony, and phrase accents. We extracted a selection of 31 accents across all categories for all solos using the MeloSpySuite[1]. These accent rules are not all independent, some are subsuming others, *e.g.*, `beat13` = accent on the primary and secondary downbeat of a bar is the logical disjunction of `beat1` (accent on primary downbeat) and `beat3` (accent on secondary downbeat). Some rules are orthogonal by construction, *e.g.*, the syncopation rules and the metrical downbeat rules, or jump accents on the tone before or after the pitch jump. Moreover, it is very well possible that the structural markers are correlated via music syntactical rules or by the creativity and expressivity of the performer. For instance, the `phrasend` rule (accent on the last tone in a phrase) coincides very often with durational accents (accents of longer tones than the previous tone(s)), because long gaps are strong hints for phrase endings. Finally, we used thresholded version of two optimized accent rules from [11] which itself are combinations (additive or tree-like) of primitive accent rules and hence not independent from its constituent rules.

Due to space restrictions, we will limit ourselves to a simple differential study of intensity with respect to structural positions—taking internal correlations of accents rules only sometimes into account. For each of the 31 binary accent rules, we conducted a Wilcoxon rank test in order to find significant differences in relative intensity between marked and unmarked locations. In the following, we will report only those tests that became significant for the largest share of performers and solos. An overview of the results can be found in Table 5, where accent rules with a mean log-BF of higher than 2 (decisive effect) are listed. We also calculated *p*-values for the corresponding global Wilcoxon tests across all solos. Strong effects with high mean log-BF did not always result in globally significant tests, since the effects were sometime in different directions and effectively canceled each other out (another reason to resort on tests for single solos). To estimate the effect size, Cohen's *d*'s were calculated per solo and then averaged. Likewise, the direction of the effect is of interest. To assess this, we define the *q*-factor as the difference between the number of positive and negative Cohen's *d* values divided by the number of solos. The range for *q* is $[-1, 1]$ with 1 meaning only positive, -1 only negative, and 0 an equal number of positive and negative effect sizes.

The highest ranked difference (according to mean log-BF) was found for durational accent `longmod_abs`, which marks tones that have an duration class higher than the mode of all duration classes in the solo, whereby duration classes are defined with respect to an absolute reference value of 500 ms. This accent condition is true for nearly exactly one-third of all tones. The direction of effect is nearly always positive, hence, longer tones (in this sense) are played louder ($\overline{d} = 0.29$), which is a medium effect.

Accent	$\overline{\log_{10} BF}$	$N_{\alpha=0.001}$	p_{glob}	$\overline{d}$	q
`longmod_abs`	3.07	46	0.00	0.29	0.75
`phrasbeg`	2.96	39	0.00	0.49	0.78
`longmod`	2.95	30	0.00	0.26	0.65
`sync1234`	2.73	26	0.00	0.25	0.70
`thom_thr`	2.52	17	0.00	0.26	0.68
`longpr_rel`	2.43	10	0.00	0.14	0.52
`long2pr`	2.29	5	0.00	0.13	0.47
`pextrem`	2.18	5	0.00	-0.09	-0.40
`sync13`	2.16	9	0.00	0.22	0.63
`jumpaft5`	2.15	6	0.00	0.12	0.32
`pextrmf`	2.11	4	0.00	-0.07	-0.23
`long2mod_win5`	2.06	4	0.04	0.02	0.02
`phrasend`	2.02	4	0.39	-0.04	0.00

Table 5: Table of accent rules with a mean log-Bayes factor higher than 2 (decisive effect). Bayes factors are defined as $\text{BF} = N_{\text{obs}}/N_{\text{exp}}$. Third column shows the number of significant Wilcoxon test at the $\alpha = 0.001$ level. The p-values of correponding global Wilcoxon test across all solos can be found int he fourth column. Estimates of the mean of Cohen's *d*'s per solo makes up the fifth column. The last column contains the *q*-factor, which is defined as the difference between the number of positive and negative Cohen's *d* value divided by the number of solos.

Next in the list is the `phrasbeg` accent, which marks the first note in a phrase, which applies to about 6% of all tones. Phrase starts are nearly always played louder ($q = 0.78$) with a rather large effect size of $d = 0.49$. The two highest ranking accents are moderately correlated. Only about one-third of phrase beginnings are also `longmod_abs` accents.

The next accent rule is `longmod`, which is true for tones that have longer inter-onset intervals (IOI) than the mean value of all IOIs in the solo. Nearly all `longmod` accents are also `longmod_abs` accents (but not vice versa), so this result is no surprise.

The fourth ranking rule is `sync1234` with captures syncopation occurring right before the beat positions in the bar. The direction is mostly positive ($q = 0.70$), hence, syncopations tend to be played louder than unsyncopated tones. But note that this holds only for about one-fourth of all solos on the 0.001-level. Even for the 5%-level this tendency is only observed in about half (55) of the solos. The effect size is nevertheless small to medium.

The following accent rule `thom_thr` is a thresholded version of Thomassen's accent, which is (a rather complicated) pitch contour accent working with three-note groups and was derived from results of labor experiments. The original Thomassen's accent gives a tone-wise probability for the perception of an accent by a listener. The thresholded version used here is true for probabilities larger than 75%. For a thorough discussion see [12].

Following up are again two durational accents `longpr_rel` and `long2pr` with mostly positive direction and small effect size. The first one marks tones that have a IOI class higher than the previous tone; classes are build using the beat duration of the solo as reference. The second one is defined for tones which have an IOI which as least two times larger than the preceding IOI. Since all duration accents are more or less correlated, this is no new result here.

The next one is the contour accent `pextrem`, which marks every extremal pitch value (local maxima and minima in pitch space). Interestingly, the effect is mostly negative in direction with effect sizes which almost cancel each other out. This means, some performer tend to play extremal pitches louder, while others tend to play them lower–if at all. Even at the 5%-level, only 23 solos became significant. However, there were some solos in which pitch extrema are strongly de-emphasized. On the 0.001 level these are two solos by Coleman Hawkins, two solos by Miles Davis and one solo by Bob Berg, with an overall large mean effect size of $\overline{d} = -0.45$.

The next rule in line is `sync13`, a subset of `sync1234`, which means that anticipated primary and secondary downbeats get emphasis

[1] `http://jazzomat.hfm-weimar.de/`.

($q = 0.63$). Then comes `pextrmf`, a subset of `pextrem` where cambiatas are excluded. Consequently, directions and effect sizes of both accent rules are similar as for the corresponding supersets.

The following rule is the pitch jump accent `jumpaft5` with positive direction ($q = 0.32$) and small effect size ($\bar{d} = 0.12$). It marks tones that follow a pitch jump of at least 5 semitones, *i.e.*, of at least a fourth up or down.

Next in row is another duration accent (`long2mod_win5`), followed by the phrase end marker `phrasend`. For this no clear direction can be found, not even for single performers, which sometime accentuate phrase ends in one solo, and de-emphasize them in another.

To sum up, the largest intensity differences can be found for durational and syncopation accents as well as for phrase beginnings. Some pitch related accents occur also in the Top 13, but these are a minority. No harmonic accent turned up, on the contrary—no systematic difference can be found between the intensity of chord and non-chord tones. Interestingly, no downbeat metrical accents made it into the Top 13, likewise, the optimised accents from [11] did not succeed. However, one must bear in mind, that the original accent rules were devised to model accent perception of pop melodies, whereas we investigate actually performed accents in jazz solos. It would be an interesting follow-up study to measure also perceived accents for our sample of solos.

4.7. First and Second Eighths

Last but not least, we investigated the intensity differences between the first and second eights in binary divided beats. In only 13 cases, solos showed significant differences on the 5%-level. However, there were some clear cases, see Table 6 for an overview. Notably, Chet Baker shows up with four of his six solos in the list, with positive d, hence, he seems to be a strong off-beat accentuator. The overall q-factor for all 120 solos is with $q = 0.15$ only slightly positive, showing a tendency for off-beat emphasis across the board, but in general there seems to be no agreement among players how to shape the eights with respect to intensity.

Performer	Title	p	Cohen's d
Chet Baker	You'd Be So Nice ...	0.000	0.783
Steve Turre	Steve's Blues	0.000	0.792
Chet Baker	Just Friends	0.000	0.552
Paul Desmond	Alone Together	0.001	0.364
John Coltrane	Blue Train	0.001	0.655
Chet Baker	Long Ago And Far Away	0.005	0.506
Joe Henderson	In 'n Out (1)	0.009	-0.253
Chet Baker	Two's Blues	0.019	0.421
Zoot Sims	Dancing In The Dark (2)	0.020	0.607
Kenny Garrett	Brother Hubbard (2)	0.028	-0.407
Miles Davis	Airegin	0.034	0.372
Joe Henderson	Serenity	0.037	0.468
Wayne Shorter	Footprints	0.041	-0.424

Table 6: Table of all solos with significant differences between first and second eights of binary beats. Positive d means that the second eights are played louder, *i.e.*, off-beats are emphasized.

5. Conclusion & Outlook

We presented a novel approach to measure tone intensities of monophonic jazz improvisations from audio files by using score-informed source-separation in order to explore dynamics in music performance. Evaluation on a multi-track recording revealed sufficient precision to justify investigating a set of 120 solos for correlations of intensity and several structural parameters, which revealed results specific to instruments or performers as well as related to syntactical and expressivity. A general rule of thumb is: the higher and the longer a tone, the louder it is played. Furthermore, structural accents such as phrase beginnings, long and syncopated notes as well as pitch peaks and pitch jump targets tend to be emphasized by performers, however, some interactions with instrumental techniques might be at play here. Furthermore, the hypothesis that the second eights in a binary divided beat are played louder could only be ascertained in some solos, notably by Chet Baker, whereas other solos tended even to use the opposite emphasis. Particularly for the two solos of Charlie Parker included in our set, no significant differences in intensity for the two eighths could be found. All in all, the two eighths of a binary divided beat are normally played equally loud with only a slight tendency to stress the second one. In general, all effects are typically of small to medium size with a hugh variety across single solos and performer, so they should understood as tendencies.

These promising first results have implications for the explorations of dynamics in jazz studies as well as for jazz education and performance reserach in general. Our findings could be easily extended to a wider range of instruments and performers. Taking more metadata, *e.g.*, style, rhythmic feel or tempo, into account might reveal significant insights in expressive techniques. Using loudness instead of intensity values (*i.e.*, using the sone scale), and in this way taking perceptual aspects into account, maybe could lead to new results. Furthermore, a more sophisticated and fine grained analysis of temporal features of intensity curves in interaction with structural accents and microtiming aspects is a highly desirable and promising approach to gain further understanding of that "magic" swing feeling that is a trademark of all jazz music.

6. Acknowledgments

The Jazzomat Research Project is supported by a of the German Research Foundation ("Melodisch-rhythmische Gestaltung von Jazzimprovisationen. Rechnerbasierte Musikanalyse einstimmiger Jazzsoli", DFG-PF 669/7-1). The authors would like to thank all jazz and musicology students participating in the transcription and annotation process.

References

[1] A. Gabrielsson: *Psychology of Music*, chapter Music Performance, pages 501–602. Academic Press, San Diego, second edition. edition, 1999.

[2] J. Langner and W. Goebel: *Visualizing Expressive Performance in Tempo–Loudness Space*. In *Computer Music Journal*, volume 27(4):69–83, 2003.

[3] F. Lerdahl and R. Jackendoff: *A generative theory of tonal music*. The MIT press, Cambridge. MA, 1983.

[4] M. Pfleiderer: *Rhythmus. Psychologische, theoretische und stilanalytische Aspekte populärer Musik*. transcript, Bielefeld, Germany, 2006.

[5] T. Owens: *Bebop. The Music and Its Players*. Oxford University Press, New York, US, 1995.

[6] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri: *Automatic music transcription: challenges and future directions*. In *Journal of Intelligent Information Systems*, pages 1–28, 2013.

[7] G. Ren, G. Bocko, J. Lundberg, S. Roessner, D. Headlam, and M. Bocko: *A Real-Time Signal Processing Framework of Musical Expressive Feature Extraction Using Matlab*. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, pages 115–120. 2011.

[8] S. Ewert and M. Müller: *Estimating Note Intensities in Music Recordings*. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 385–388. 2011.

[9] E. Cano, G. Schuller, and C. Dittmar: *Pitch-informed solo and accompaniment separation towards its use in music education applications*. In *EURASIP Journal on Advances in Signal Processing*, volume 2014(1):23, 2014.

[10] T. Painter and A. Spanias: *Perceptual Coding of Digital Audio*. In *Proceedings of the IEEE*, volume 88(4):451–515, 2000.

[11] D. Müllensiefen, M. Pfleiderer, and K. Frieler: *The Perception of Accents in Pop Music Melodies*. In *Journal of New Music Research*, volume 1:19–44, 2009.

[12] J. Thomassen: *Melodic accent: Experiments and a tentative model*. In *Journal of the Acoustical Society of America*, volume 71:1596–1605, 1982.