Score-Informed Analysis of Tuning, Intonation, Pitch Modulation, and Dynamics in Jazz Solos

Jakob Abeßer, Klaus Frieler, Estefanía Cano, Martin Pfleiderer, Wolf-Georg Zaddach

***Abstract*—Both the collection and analysis of large music repertoires constitute major challenges within musicological disciplines such as jazz research. Automatic methods of music analysis based on audio signal processing have the potential to assist researchers and to accelerate the transcription and analysis of music recordings significantly. In this paper, we propose a framework for analyzing improvised monophonic solos in multi-instrumental jazz recordings with special focus on reed and brass instruments. The analysis algorithms rely on prior score-information, which is taken from high quality manual solo transcriptions. Following an initial solo and accompaniment source separation, we propose algorithms for tone-wise extraction of fundamental frequency and intensity contours. Based on this fine-grained representation of recorded jazz solos, we perform several exploratory experiments motivated by questions relating to jazz research in order to analyze the use of expressive stylistic devices such as intonation, pitch modulation, and dynamics in jazz solos. The results show that a score-informed audio analysis of jazz recordings can provide valuable insights into the individual stylistic characteristics of jazz musicians.**

***Index Terms*—Jazz research, music analysis, score-informed audio analysis, fundamental frequency contours, dynamics, source separation**

I. INTRODUCTION

Improvised solo parts, which are shaped both by the individual approach of the performing musician and by overarching stylistic conventions, are key elements of jazz music. Improvisations can be described and analyzed with regard to syntactical and structural properties on the one hand, and expressive and timbral aspects on the other. The syntactical dimensions of jazz solos include for instance pitch and interval choices, specifically in relation to underlying harmonies, as well as rhythmical and metrical structures. Symbolic music representations such as Western staff notation or MIDI are best suited to analyzing these properties as they directly encode sounding tone events as basic syntactical note elements [1].

In contrast, non-syntactical or expressive properties relate to the actual performance of the improvised melody on a specific instrument. Some important aspects are timbre and special sound characteristics such as roughness and breathiness, micro-timing (deviations from the underlying metric grid), dynamics (intensity changes), intonation (pitch accuracy with respect to the tuning of the accompanying musicians and to a given tone system), articulation (e.g., legato and staccato), as well as pitch modulation (variation of the fundamental frequency within a tone). These properties require a complementary audio-based analysis of the recording.

The Jazzomat Research Project aims to analyze all aforementioned dimensions of jazz improvisation using a large database of jazz recordings in order to examine, among other aspects, the personal styles of jazz musicians. In this paper, we focus on the audio-based analysis of expressive properties such as intonation, pitch modulation, and dynamics of monophonic solos extracted from polyphonic jazz recordings. We combine several previously proposed audio analysis methods to a unified framework and analyze a novel large dataset to gain insights on jazz solo performances.

The paper is structured as follows: Section II discusses related work on the analysis of tuning, pitch modulation and intonation, as well as dynamics in music recordings. Section III presents the applied method for expressive analysis of improvised jazz solos in audio recordings in detail. The evaluation dataset and several experiments in analyzing different aspects of jazz musicians' personal styles are discussed in Section IV. Finally, Section V concludes our work and highlights further perspectives.

II. BACKGROUND

A. Tuning

Tuning describes the adjustment of pitch frequencies of musical instruments to a given reference frequency. This allows musicians to coordinate the intonation of several instruments when playing together. Nowadays, tuning is most often standardized by the concert pitch A4 with the frequency $f = 440\,\mathrm{Hz}$. However, there are various reasons why music recordings may deviate from a standardized tuning frequency. Firstly, before 1955, there was no authoritative international standard but only informal recommendations using varying reference values. The tuning frequency of $f = 440\,\mathrm{Hz}$ was included in the ISO 16 standard in 1955 and reaffirmed in 1975 [2]. Secondly, pianos, as the main tuning reference for other instruments in jazz, go out of tune over time [3]. Thirdly, jazz recordings up to the 1940s often exhibit tuning deviations due to speed variations of grammophones or tape recorders caused by technical imperfections [3].

Tuning estimation constitutes an important pre-processing step for several MIR tasks such as key or chord estimation [4] as well as music transcription [5]. Tuning estimation is often performed based on the estimated frequency position of spectral peaks that coincide with harmonic frequencies of the performing instruments. By focusing on stable pitch contours of a certain minimum length, the robustness of the tuning estimation can be improved [6], [7], [8]. Serrà *et al.* use high-resolution interval histograms based on peak frequencies and compare these with interval histograms that represent just intonation and equal temperament [7]. Several authors propose testing different tuning hypotheses by using adjustable semitone filterbanks [9] or by minimizing the distance of spectral peaks to semitone grids in order to select the most likely tuning frequency [10]. Dressler and Streich map the tuning deviation to the complex plane and apply circular statistics of tuning deviations on selected stable melody pitches to obtain a tuning frequency estimate [6]. Dixon *et al.* propose a system that matches peak frequencies to frequency templates of 15 different temperaments from solo harpsichord recordings [8]. Müller and Ewert create triangular filterbanks for different tuning frequency candidates by aligning their center frequencies to the chromatic scale within the full piano pitch range [11]. A final estimate of the tuning frequency is derived by filtering the averaged magnitude spectrogram and by maximizing the filterbank output energy. Mauch proposes computing a harmonic saliency function based on a logarithmic frequency axis with a resolution of three bins per semitone [4]. The tuning deviation from $440\,\mathrm{Hz}$ is obtained from the phase spectogram.

B. Pitch Modulation & Intonation

Notes of a given musical composition are rarely transformed into played or sung tones with constant pitches. For the analysis of pitch modulation, the fundamental frequency (f_0) contour must be estimated for each tone in a recorded melody. One suitable approach involves the use of an automatic melody transcription algorithms that include a pitch estimation step, in which f_0 contours are formed [12], [13]. A general challenge in the analysis of polyphonic music recordings is the interference between different instrument signals [5]. This can be circumvented either by analyzing isolated tracks from multi-track recordings, which are usually not available for commercial recordings, or by isolating the melody signal using source separation techniques [14].

Vibrato is one of the most important modulation techniques and describes a periodic pitch change around a target pitch. It is often used by vocalists [15] as well as by string and wind instrument players [16]. We investigated intonation in our previous work [17] as well as the use of pitch modulation techniques in reed and brass jazz solos. Other publications focus on the deviation of f_0 contours from the target pitch, i. e., intonation [18], as well as on modulation techniques such as vibrato and pitch glides [19] or pitch bends [20].

C. Dynamics

Dynamics are an important part of any musical performance [21], [22]. Musicians commonly play musical phrases with differing degrees of intensity and accentuate certain notes by playing them more loudly. Lerdahl and Jackendoff refer to this as "local stresses" or "phenomenal accents" [23]. Intensity can also change within the duration of a longer tone. The acoustic properties of the musical instruments are another important source of dynamic variations. Presumably, dynamics are shaped for various reasons and according to several implicit syntactical rules and expressive intentions. Metrically or structurally salient tones can be emphasized using higher intensity to convey the underlying syntactical structure [23]. On the other hand, single tones in a melodic line that are played more loudly can form an additionally overlaid rhythmical stratum [24], e.g., giving rise to syncopations. This is common practice in African music, jazz, rock, and pop music. As an example, many jazz musicians such as jazz saxophonist Charlie Parker or clarinet and soprano saxophone player Sidney Bechet are said 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 [25]. However, dynamics are often neglected in jazz research since they are difficult to measure using conventional methods such as manual transcription.

Similar to the estimation of the f_0 contours discussed in the previous section, a precise estimation of single tone intensity values in ensemble recordings is complicated by the inherent interference between instrument signals. Therefore, most of the work in the field of expressive performance analysis has focused on isolated instrument recordings [26], [27] but also on ensemble recordings [28]. For instance, Arcos *et al.* use the Spectral Modeling Synthesis technique to extract expressive performance parameters such as dynamics from tenor saxophone performances [29]. Score-informed methods for extracting expressive parameters such as note onset times and intensity values from polyphonic piano recordings are proposed by Scheirer [30] as well as Ewert and Müller [31]. Ramirez et al. use a genetic algorithm to automatically learn performance rules concerning timing and dynamics from recorded jazz saxophone recordings [32]. Here, we use given score-information of the solo melody to isolate the solo instrument from the mixture signal using a source separation algorithm as detailed in Section III-A. As will be shown in Section III-D, we apply a perceptually-motivated method to estimate tone-wise intensity values [33] and to cope with the timbre variety of the different solo instruments we investigate.

III. Method

In this paper, we conduct score-informed audio analysis based on high-quality melody transcriptions of monophonic jazz improvisations, which were notated manually by music experts. The transcription process is described in detail in Section IV-A. The proposed analysis framework is shown in Figure 1. It combines the approaches from our previous work on analysis of pitch modulations [17] and dynamics in [33] in a unified framework including several parameter optimizations with respect to the sampling frequency and frequency resolution. Using the score-information, we first apply a source separation algorithm (Section III-A) to split the solo part taken from the original audio recording into a solo instrument track featuring the improvising melody instrument and a backing track with the accompaniment, i. e., the rhythm section (drums, double bass, and piano). Based on the backing track, we estimate the tuning frequency of the rhythm section (Section III-B). Given both the tuning frequency and the score information, we track the fundamental frequency (f_0) contour of each tone in the solo instrument track (Section III-C). Finally, we estimate tone-wise intensity values (Section III-D).

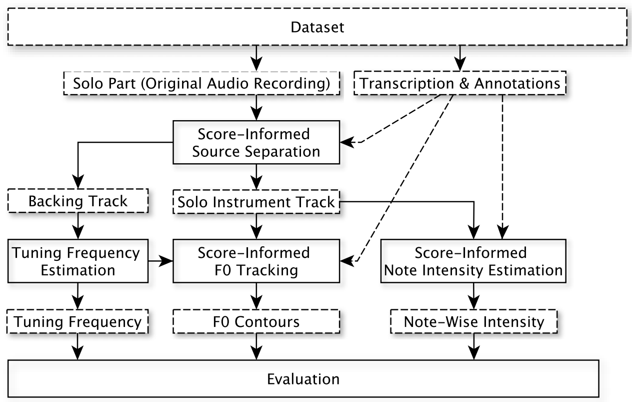

Fig. 1. Overview of the proposed method for score-informed expressive analysis of jazz solos. Processing steps are shown as solid line boxes, intermediate results are shown as dotted line boxes.

A. Score-Informed Solo and Accompaniment Separation

An initial sound source separation stage allows us to isolate independent signals for the solo instrument and the accompanying instruments from the audio mix. This facilitates a more detailed and focused analyses than would be possible with the original mix. In this study, the backing track is used to extract tuning information from the rhythm section only. With the separated solo signal and the extracted tuning information, we are able to study the intonation, dynamics, and pitch modulations of the soloist's performance. The separation algorithm can be guided by using the transcriptions as high-quality prior information. As opposed to most score-informed separation tasks, we have the great benefit of having melody transcriptions already aligned to the audio track. This greatly reduces the complexity of the process and results in better quality of separation.

To separate the solo instrument from the backing track, the method for pitch-informed solo and accompaniment separation proposed in [14] was used. The automatic pitch detection stage in the separation algorithm was bypassed and the melody transcriptions were used as prior information instead. The separation method is based on a spectral model of the solo instrument that takes into account characteristics of musical instruments such as common amplitude modulation, inharmonicity, magnitude and frequency smoothness, among others. The separation method has proven to be robust in the extraction of a great variety of solo instruments, and particularly efficient, with computation times that permit real-time processing.

B. Tuning Frequency Estimation

1) Method: In this paper, we use the two methods proposed by Müller and Ewert [11], and Mauch [4], as implemented in the publicly available Chroma Toolbox[1] and NNLS Chroma Vamp Plugin[2], respectively. Our goal is to evaluate the intonation of the improvising soloist with respect to the accompanying reference section. Therefore, we estimate the tuning frequency from the backing track (compare Section III-A). For each solo, we obtain two estimates $f_{\text{ref}}^{\text{CTB}}$ and $f_{\text{ref}}^{\text{NNLS}}$ from the Chroma Toolbox for Matlab and the NNLS Vamp plugin, respectively. For the Chroma Toolbox approach, we modified the originally proposed search range for $\hat{f}_{\text{ref}}$ to $440\text{Hz} \pm 0.5$ semitone (corresponding MIDI pitch range: 69 ± 0.5) and the stepsize to 0.1 cents. We were able to show in [17] that the artifacts of the source separation are negligible with regard to the precision of the automatic tuning estimation.

2) Evaluation: As an initial experiment, we compare the tuning frequency estimates $f_{\text{ref}}^{\text{CTB}}$ and $f_{\text{ref}}^{\text{NNLS}}$. Figure 2 shows pairs of tuning estimates for all solos in the dataset. The agreement is very high for most of the solos (sample correlation of $r = 0.96$ with $p < .001$, root mean squared error (RMSE) of 0.13 cent), apart from seven outliers visible in the lower right of the scatter plot.[3] Due to the faster computational speed of the NNLS Vamp Plugin, we use $f_{\text{ref}} = f_{\text{ref}}^{\text{NNLS}}$ as the tuning frequency estimate in the remaining experiments.

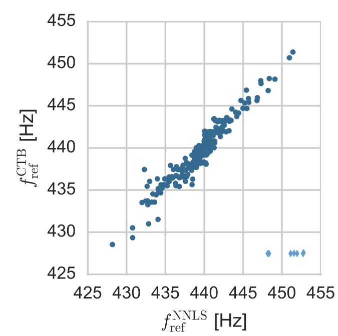

Fig. 2. Tuning Frequency estimates $f_{\text{ref}}^{\text{CTB}}$ and $f_{\text{ref}}^{\text{NNLS}}$ in Hz extracted with the *Chroma Toolbox for Matlab* and the *NNLS Vamp plugin*, respectively.

C. Score-Informed Fundamental Frequency Tracking

1) Spectral Estimation: For the f_0 tracking of a given tone, we compute a *reassigned magnitude spectrogram* between its temporal boundaries defined by its onset and offset time and a safety margin around its pitch [34]. The principal idea is to reassign magnitude values to different frequency positions such that the resulting time-frequency representation exhibits sharper peaks at harmonic frequencies compared to the STFT magnitude spectrogram. Therefore, the reassigned spectrogram is better suited for tracking time-varying f_0 contours.

We first compute the magnitude spectrogram $M \in \mathbb{R}_+^{K \times N}$ from the Short-Time Fourier Spectrogram (STFT) of the solo part segment that corresponds to the given tone. K is the number of frequency bins and N is the number of time frames. We use a zero-padding factor of 8, a blocksize of $b = 2048$, a hopsize of $h = 128$, and a sampling frequency of $f_s = 44.1\,\text{kHz}$. We estimate the instantaneous frequency

[1] http://resources.mpi-inf.mpg.de/MIR/chromatoolbox/ (last visited: 12.10.2016)

[2] http://isophonics.net/nnls-chroma (last visited: 12.10.2016)

[3] We did not observe any common acoustic features among the outliers. Their tuning estimates $f_{\text{ref}}^{\text{CTB}}$ and $f_{\text{ref}}^{\text{NNLS}}$ are close to the lower and upper tuning frequency range, respectively. Therefore, the high difference between both methods presumably comes from wrapping to the fixed pitch range of $\pm\,0.5$ semitones around $440\,\text{Hz}$.

$\hat{f} \in \mathbb{R}^{K \times N}$ for each time-frequency bin using the method proposed by Abe in [35], which exploits the time derivative of the phase for frequency correction.

For computing the reassigned spectrogram $M_{\text{IF}} \in \mathbb{R}^{K_{\log} \times N}$, we then define a logarithmically-spaced frequency axis as $f(k_{\log}) = f_{\text{ref}} \cdot 2^{\frac{P-69-2+k_{\log}/25}{12}}$ with $k_{\log} \in [0, K_{\log}]$ and $K_{\log} = 100$ within a margin of $\pm$ 2 semitones around the tone's pitch P and a frequency resolution of 25 bins per semitone. The frequency axis is aligned to the estimated tuning frequency f_{ref} (see Section III-B).[4] Finally, each STFT magnitude value $M(k, n)$ is reassigned to the target frequency bin $M_{\text{IF}}(k_{\log}, n)$ of the reassigned spectrogram M_{IF} whose frequency $f(k_{\log})$ is closest to the corresponding instantaneous frequency value $\hat{f}(k, n)$. Therefore, in each time frame n, the original STFT magnitude values $M(k, n)$ are accumulated in M_{IF} as

$$M_{\text{IF}}(k_{\log}, n) = \sum_{k \in [1,K]} \delta(k, n) \cdot M(k, n) \text{ with} \tag{1}$$

$$\delta(k, n) = \begin{cases} 1, & \text{for } k_{\log} \equiv \text{argmin}_{\tilde{k}_{\log}} \left| f(\tilde{k}_{\log}) - \hat{f}(k, n) \right| \\ 0, & \text{otherwise.} \end{cases}$$

2) Starting Location: As a first step of the f_0 tracking of a given tone, we first identify a suitable time-frequency position in the tone's reassigned spectrogram M_{IF} as starting location: In each frame n within the tone's duration, we retrieve the frequency bin position $k_{\log}^{\text{peak}}(n)$ of the magnitude peak as

$$k_{\log}^{\text{peak}}(n) = \arg\max_{k_{\log}} M_{\text{IF}}(k_{\log}, n). \tag{2}$$

Then, we select the frame n^0 with a magnitude peak as close as possible to the annotated fundamental frequency and the corresponding frequency bin $k_{\log}^0 = k_{\log}^{\text{peak}}(n^0)$ as starting coordinates for the subsequent f_0 tracking.

3) Contour Tracking: Based on the starting location $(k_{\log}^0, n^0)$, the f_0-contour is tracked on a frame-wise basis forwards and backwards in time. Since we assume harmonic instruments to have continuous f_0-contours, we allow a maximum absolute frequency deviation of only $5\,\text{cent}$ between the f_0 values in adjacent frames. In each frame, we choose the f_0 frequency bin based on the peak position in the search range around the previous f_0 estimate. Finally, we obtain a fundamental frequency contour $f_0(n)$ for each tone, which we use for intonation and pitch modulation analysis (Section IV-C).

4) Evaluation: For evaluating the f_0 tracking, we selected eight solos[5] from our dataset (see Section IV-A) as test-set. Three human annotators annotated f_0 contours using a graphical user interface described in [36]. For comparison, we extracted contours using the pYIN algorithm [37].

[4]Since pitch bends, slides, and other frequency modulation techniques can lead to a significant pitch deviation in the beginning or end of the tone, these parameters were chosen to allow to capture the f_0 contour over the full duration of each tone.

[5]Clifford Brown: Sandu, Jordu, Joy Spring; Curtis Fuller: Blue Train; John Coltrane: Blue Train; Sidney Bechet: Summertime; Stan Getz: The Girl From Ipanema; Wayne Shorter: Footprints

We compared for each solo both of the two automatically estimated f_0 contours with each of the three manually annotated contours in a pair-wise fashion. For each comparison, we compute the standard MIREX evaluation measures Voicing Detection Rate VDR (proportion of correctly detected ground-truth melody frames), Voicing False Alarm Rate VFAR (proportion of ground-truth non-melody frames mistakenly detected as melody frames), and Raw Pitch Accuracy RPA (proportion of detected melody frames with an absolute pitch error of less than a quartertone) as detailed in [38]. In addition, we calculate the median over the absolute frame-wise pitch errors in cent (MAPE).

Table I shows the evaluation metrics over all solos & annotations (mean and standard deviation values are given). Firstly, the mean VFAR value of 0.25 for the proposed method indicates the amount of disagreement in the annotation of note onset and offset times between the annotators who contributed to the Weimar Jazz Database (see Section IV-A) and those who transcribed the f_0 contours (see also [36]). Since we obtained f_0 for each time frame from the pYIN algorithm (using the "smoothedpitchtrack" parameter), we did not compute the two voicing-related measures VDR and VFAR for this algorithm. Secondly, in terms of pitch detection, both algorithm show comparable results. While the pYIN gets slightly better scores for RPA and MAPE, the proposed algorithm shows lower standard deviation values for the 7 test pieces. These results indicate that the proposed method performs comparable to the state-of-the-art pYIN algorithm and is suitable for the analysis of pitch modulation and intonation.

TABLE I
RESULTS OF f_0 TRACKING. BOTH MEAN VALUES (AVERAGED OVER ALL THREE ANNOTATIONS AND SEVEN TEST PIECES) AND STANDARD DEVIATION VALUES (IN BRACKETS) ARE GIVEN.

Algorithm	VDR	VFAR	RPA	MAPE
Proposed	0.92 (0.02)	0.25 (0.06)	0.84 (0.04)	11.76 (0.38)
pYIN	-	-	0.87 (0.08)	11.10 (1.28)

D. Score-Informed Intensity Estimation

We compute intensity values for all tones in the solo following the approach described by Painter in [39]. We assume a strictly monophonic melody without any overlap. First, we compute an STFT spectrogram $X \in \mathbb{R}_+^{K \times N}$ using a blocksize of $b = 512$, a hopsize of $h = 480$, a sampling frequency of $f_s = 44.1\,\text{kHz}$, and a Hann window.

Based on 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

$$I_b(n) = \frac{1}{b} \sum_{k \in [k_{\min,b}, k_{\max,b}]} |X(k, n)|^2 . \tag{3}$$

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

Finally, we compute the *frame-wise intensity* value in the n-th frame as

$$I(n) = 90.302 + 10 \log_{10} \sum_{b=1}^{24} I_b(n). \quad (4)$$

using the sound pressure level (SPL) normalization discussed in [39]. In order to compute the *intensity* value I_i of the i-th *tone*, we take the highest frame-wise intensity value between its onset frame $n_{\mathrm{on},i}$ and offset frame $n_{\mathrm{off},i}$ as

$$I_i = \max_{n \in [n_{\mathrm{on},i}, n_{\mathrm{off},i}]} I(n). \quad (5)$$

We follow the assumption here that the maximum value affects the overall perceived intensity [31]. In order to facilitate the comparison between different solos, we normalize the tone intensity values by mapping the 5%–95% percentiles to the interval $[0, 1]$ for each solo. All tones played with intensity values outside these percentile ranges are discarded as outliers in our experiments.

IV. Experiments & Results

A. Dataset

All experiments presented in this paper are based on a subset of 264 solos taken from the publicly available Weimar Jazz Database[6], which currently contains 299 high-quality transcriptions of instrumental solos from various jazz styles and performers. The transcriptions were created manually using the Sonic Visualiser software [40], and cross-checked by jazz and musicology students. Each transcription includes note annotations (pitch, onset, and duration) as well as several contextual annotations: metrical structure, beat times, musical phrases, salient pitch modulations such as vibrato or slides, as well as form sections, chorus numbering, and chords corresponding to the lead sheet of the original composition.[7] In this paper, we focus on instrumental solos performed on reed and brass instruments. Also, we only consider notes of a minimum duration of 50 ms. Table II gives an overview of all performers included, their musical instruments[8], as well as their respective number of solos and total number of notes. In total, our dataset includes 104,964 annotated note events in 264 jazz solos played by 47 performers.

In the following sections, we describe several exploratory experiments motivated by musicological research questions concerning the stylistic peculiarities of improvising jazz musicians: the musicians' characteristic pitch modulation and intonation (Section IV-C), as well as their use of dynamics (Section IV-D). Additionally, the tuning of jazz recordings will also be examined (Section IV-B).

[6]http://jazzomat.hfm-weimar.de/dbformat/dboverview.html

[7]Copyright restrictions allow us to only publish the transcriptions. However, we include both the MusicBrainz-IDs (https://musicbrainz.org, last visited: 12.10.2016) as well as the precise solo start times, which permit the identification of the corresponding audio segments in the original recordings. In addition, raw analysis results as well as Python code for the analysis steps are available on https://github.com/jazzomat/article_2016.

[8]The abbreviations of the instruments are: as - alto saxophone, bs - bass saxophone, cl - clarinet, cor - cornet, tb - trombone, tp - trumpet, ts - tenor saxophone, ts-c - C melody saxophone, ss - soprano saxophone.

TABLE II
Overview of performers, instrument(s), number of solos, and number of notes.

Performer	Instrument	# Solos	# Notes
Art Pepper	as, cl	6	3482
Ben Webster	ts	5	852
Benny Carter	as	5	1750
Benny Goodman	cl	7	1966
Bix Beiderbecke	cor	4	518
Bob Berg	ts	6	4000
Buck Clayton	tp	3	561
Cannonball Adderley	as	5	2475
Charlie Parker	as	6	1606
Chet Baker	tp	6	1079
Clifford Brown	tp	7	2890
Coleman Hawkins	ts	6	2465
David Liebman	ss, ts	5	3210
David Murray	ts	6	2810
Dexter Gordon	ts	5	3702
Dickie Wells	tb	3	387
Dizzy Gillespie	tp	5	1384
Don Byas	ts	7	1928
Eric Dolphy	as	3	1500
Fats Navarro	tp	4	876
Freddie Hubbard	tp	6	2016
Gerry Mulligan	bs	3	1049
Hank Mobley	ts	3	1462
J.J. Johnson	tb	5	2215
Joe Henderson	ts	6	3534
Joe Lovano	ss, ts, ts-c	6	4116
John Coltrane	ss, ts	11	8042
Joshua Redman	ts	5	2344
Kenny Dorham	tp	6	1922
Kid Ory	tb	3	174
Lee Konitz	as	5	2202
Lester Young	ts	6	1452
Louis Armstrong	cor, tp	6	782
Michael Brecker	ts	6	4076
Miles Davis	tp	8	2377
Ornette Coleman	as	5	2718
Paul Desmond	as	8	2119
Roy Eldridge	tp	6	1643
Sidney Bechet	ss	3	695
Sonny Rollins	ts	12	4797
Sonny Stitt	as, ts	4	1239
Stan Getz	ts	6	3129
Steve Coleman	as	7	2776
Steve Lacy	ss	5	1661
Steve Turre	tb	3	1038
Wayne Shorter	ts	10	3510
Woody Shaw	cor, tp	6	2435
Total		**264**	**104964**

B. Tuning

1) Tuning Deviations by Recording Year: As outlined in Section II, the tuning frequency differs slightly between music recordings, particularly in the first half of the 20th century. Since the jazz recordings in the Weimar Jazz Database cover large parts of the 20th century, it allows us to investigate whether this phenomenon can be discovered in our dataset. We compute the deviation of the estimated tuning frequency from the ideal concert pitch of $440\,\mathrm{Hz}$ in cent as $\Delta f_{\mathrm{ref}}^{440} = 1200 \log_2 \frac{f_{\mathrm{ref}}}{440}$. As shown in Figure 3, we observe a negative correlation between $\Delta f_{\mathrm{ref}}^{440}$ and the recording year of $r = -0.33$ ($p < 0.001$). Hence, the absolute deviation has decreased over the course of the 20th century.

Notably, before the year 1960, deviations up a to maximum of 50 cent can be found, while after around 1960, the highest tuning deviations drop below 25 cent. One possible reason might be a (slightly delayed) adoption of the standardized $440\,\mathrm{Hz}$ tuning frequency issued by the International Standards Organization in 1955.

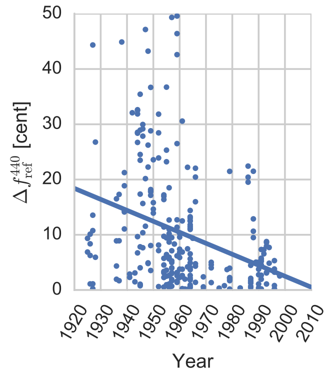

Fig. 3. Tuning deviation $\Delta f_{\text{ref}}^{440}$ from 440 Hz in cent vs. recording year for all solos in the dataset. A negative correlation of $r = -0.33$ ($p < 0.001$) is observed. The linear regression line is shown.

C. Pitch Modulation and Intonation

This section presents experiments that focus on the characteristics of the fundamental frequency contours estimated for each tone that corresponds to a note in the Weimar Jazz Database. The variation of the fundamental frequency within each tone's duration constitutes an important means of expression in various music genres, including jazz.

1) Artist-Specific Pitch Modulation & Intonation: References to the personal "sound" of a jazz performer are abundant in jazz literature [41]. Personal sound is regarded as one of the main defining features for an artistic personality. Two key ingredients of the personal sound of jazz players are pitch modulation techniques (e. g. vibrato) and, presumably, intonation characteristics.

The *modulation range* is a measure of pitch stability. It can be defined as the amount of f_0 variation within a single tone. We measure it for each tone as the average interquartile range (IQR), i.e., the difference between the 75th and the 25th percentiles, over frame-wise f_0 deviations from the annotated note pitch. Table III lists the ten lowest-ranked and the ten highest-ranked artists with respect to their average modulation range. Strikingly, tenor saxophone players—along with trombone players Dickie Wells, Steve Turre, and Kid Ory—have the highest IQR values (see also Figure 4), presumably due to specificities of that particular instrument [42] or, in case of the swing players Byas, Young, Hawkins and Webster, to their pronounced vibrato. On the other hand, soloists playing soprano saxophone, clarinet, and trumpet show lower IQR values, which is probably caused by a lesser use of vibrato.

The *pitch intonation deviation* is measured by the median f_0 deviation from the ground-truth pitch frequencies, taking into account the estimated tuning frequency. We chose the median as it is more robust to possible outliers caused by f_0 tracking errors. In order to facilitate the interpretation of the results, we classify notes with intonation deviations of below -25 cents as "flat" (-), notes above 25 cents deviations as "sharp" (+), and all others as "normal" (o). We define the *intonation tendency* as $T = (N_{\text{sharp}} - N_{\text{flat}})/(N_{\text{sharp}} + N_{\text{flat}})$ using the number of sharp notes N_{sharp} and flat notes N_{flat}. We designate the number of notes per solo with an absolute intonation deviation of less than 25 cent as N_{on} and the total number of notes as N. Table IV shows a list of performers with the highest and lowest values N_{on}/N and the corresponding intonation tendency T.

Results suggest that there is no general tendency regarding intonation with respect to instrument or jazz style. Instead, intonation seems to be a personal characteristic of a particular jazz musician—or even a peculiarity of a certain soloist playing with a certain group during a certain recording session.

TABLE III
PERFORMERS WITH LOWEST AND HIGHEST AVERAGE FUNDAMENTAL FREQUENCY MODULATION RANGE OVER ALL CORRESPONDING SOLOS.

#	Performer	Instrument	Average IQR [cent]
1	Steve Lacy	ss	17.4
2	Benny Goodman	cl	17.5
3	Woody Shaw	cor, tp	19.8
4	Dizzy Gillespie	tp	21.2
5	Freddie Hubbard	tp	21.5
6	Art Pepper	as, cl	22.1
7	David Liebman	ss, ts	22.1
8	Clifford Brown	tp	22.4
9	Miles Davis	tp	22.6
10	Buck Clayton	tp	22.6
38	Gerry Mulligan	bs	30.2
39	Kid Ory	tb	30.3
40	Steve Turre	tb	30.3
41	Sonny Rollins	ts	30.4
42	Michael Brecker	ts	30.5
43	Ben Webster	ts	30.8
44	Dickie Wells	tb	30.9
45	Coleman Hawkins	ts	31.4
46	Lester Young	ts	31.7
47	Don Byas	ts	33.7

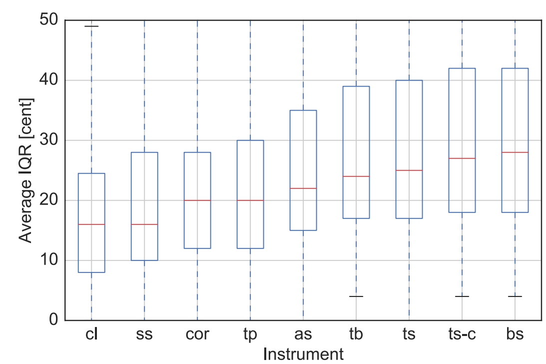

Fig. 4. Boxplot over fundamental frequency modulation range (interquartile range in cent) for different instruments sorted by median values. The vertical axis is truncated at 50 cent for better readability.

D. Dynamics

1) Note Intensity within the Musical Context: In this experiment, we investigate to what extent the correlations between tone intensity and contextual parameters such as pitch and duration are characteristic features for a musician. For each solo in the dataset, we compute the Pearson correlation coefficient r between a tone's intensity and pitch, its duration, and its relative position within the corresponding phrase ("RelPosInPhrase", normalized to $[0, 1]$). For each musician, we average the correlation coefficients over all solos that show significant correlations (with $p < 0.05$). Table V gives an

TABLE IV
PERFORMERS WITH HIGHEST AND LOWEST PERCENTAGE N_{ON}/N OF NOTES WITHIN AN ABSOLUTE INTONATION DEVIATION OF LESS THAN 25 CENT. THE GENERAL INTONATION TENDENCY T AS WELL AS A CLASSIFICATION INTO "FLAT" ("-" FOR $T < -0.15$), "NO TENDENCY" ("O" FOR $T \in [-0.15, 0.15]$), AND "SHARP" ("+" FOR $T > 0.15$) ARE SHOWN.

#	Performer	Instrument	N_{on}/N	Intonation Tendency T	Class
1	Benny Goodman	cl	0.81	0.24	+
2	Chet Baker	tp	0.81	0.26	+
3	Bix Beiderbecke	cor	0.80	0.24	+
4	Louis Armstrong	cor, tp	0.79	-0.01	o
5	Freddie Hubbard	tp	0.79	-0.26	-
6	Woody Shaw	cor, tp	0.79	-0.07	o
7	Paul Desmond	as	0.79	0.39	+
8	Buck Clayton	tp	0.76	0.07	o
9	Sidney Bechet	ss	0.76	-0.15	-
10	Benny Carter	as	0.75	-0.15	o
38	David Liebman	ss, ts	0.69	-0.13	o
39	Dizzy Gillespie	tp	0.69	0.37	+
40	Coleman Hawkins	ts	0.68	0.22	+
41	Ornette Coleman	as	0.68	-0.18	-
42	Eric Dolphy	as	0.68	0.01	o
43	Fats Navarro	tp	0.67	0.18	+
44	Roy Eldridge	tp	0.67	0.56	+
45	Michael Brecker	ts	0.65	0.05	o
46	Don Byas	ts	0.63	0.43	+
47	Charlie Parker	as	0.59	0.53	+

TABLE V
HIGHEST-RANKED AND LOWEST-RANKED PERFORMERS SORTED IN DESCENDING ORDER OF PEARSON CORRELATION COEFFICIENT VALUES BETWEEN INTENSITY AND PITCH, NOTE DURATION, AND RELATIVE NOTE POSITION IN A PHRASE AVERAGED OVER THEIR SOLOS. BOTH MEAN AND STANDARD DEVIATION VALUES ARE GIVEN.

#	Pitch	Duration	RelPosInPhrase
1	Sonny Stitt (0.67 ± 0.16)	Benny Carter (0.33 ± 0.00)	Roy Eldridge (0.12 ± 0.00)
2	Bix Beiderbecke (0.59 ± 0.05)	Buck Clayton (0.27 ± 0.00)	Art Pepper (-0.12 ± 0.00)
3	Woody Shaw (0.57 ± 0.09)	Dizzy Gillespie (0.26 ± 0.08)	Benny Carter (-0.12 ± 0.00)
4	Ornette Coleman (0.57 ± 0.08)	Fats Navarro (0.26 ± 0.07)	John Coltrane (-0.12 ± 0.01)
5	Clifford Brown (0.57 ± 0.06)	Benny Goodman (0.26 ± 0.08)	Stan Getz (-0.13 ± 0.06)
6	Chet Baker (0.56 ± 0.13)	Charlie Parker (0.26 ± 0.07)	Benny Goodman (-0.13 ± 0.00)
7	J.J. Johnson (0.56 ± 0.08)	Clifford Brown (0.21 ± 0.02)	Bix Beiderbecke (-0.13 ± 0.00)
41	Roy Eldridge (0.19 ± 0.29)	Joe Henderson (0.12 ± 0.03)	Sonny Stitt (-0.27 ± 0.00)
42	David Liebman (0.19 ± 0.08)	Bob Berg (0.10 ± 0.00)	Kenny Dorham (-0.29 ± 0.06)
43	Paul Desmond (0.17 ± 0.35)	Steve Lacy (0.09 ± 0.00)	Dizzy Gillespie (-0.29 ± 0.01)
44	Michael Brecker (0.10 ± 0.00)	Coleman Hawkins (0.09 ± 0.14)	Lester Young (-0.33 ± 0.09)
45	Steve Turre (0.08 ± 0.44)	Ben Webster (0.09 ± 0.17)	Freddie Hubbard (-0.34 ± 0.12)
46	Joshua Redman (-0.09 ± 0.19)	Michael Brecker (0.07 ± 0.00)	Miles Davis (-0.35 ± 0.00)
47	Bob Berg (-0.19 ± 0.03)	Stan Getz (-0.02 ± 0.11)	Coleman Hawkins (-0.39 ± 0.08)

overview of the highest-ranked and lowest-ranked musicians for each contextual parameter sorted in descending order of correlation effect size. Both the mean and standard deviation over all solos of a particular musician are depicted.

Throughout, we can observe a strong correlation between intensity and pitch, which is in agreement with the common assumption that higher notes are played more loudly. Significant exceptions are the three postbop musicians Joshua Redman, Michael Brecker, and Bob Berg who appear to play with a fairly constant intensity over the full range of the tenor saxophone and even tend to play low tones more loudly ("honking"). It seems that these younger musicians are very homogeneous players who withstand the supposedly natural tendency to play higher tones with more breath and pressure. The same seems to be true for postbop trombone player Steve Turre. Similarly to our previous work [33], we observe on average a stronger correlation for brass instruments (cor, tp, tb) as compared to reed instruments (as, ts, bs, ss, cl). We also observe a positive correlation between intensity and note duration with, however, a smaller effect size. With very few exceptions, all musicians tend to play longer tones more loudly.

In the *Weimar Jazz Database*, solos are segmented by the transcribers into musical phrases, i. e. groups of consecutive notes which form perceptually closed melodic units. With the exception of Roy Eldridge, the results show that all soloists play tones at the beginning of musical phrases more loudly than at the end, which is presumably due to running out of breath, but could also be a deliberate artistic decision. Presumably, uniformity of intensity in regard to both dimensions, pitch and relative position in a phrase, is a characteristic trait of the individual playing aesthetics of certain musicians.

2) Phrase-Wise Intensity and Pitch Contours: The overall movement of a phrase within the pitch space as well as time-varying intensity curves (dynamics) contribute significantly both to the emotional and semantic impression of a melody. Furthermore, it is well-established in music psychology that pitch contours are important for melodic perception and memory [43], and are useful for melodic classification and similarity judgments [44]. Hence, analyzing pitch and intensity contours is of fundamental interest with regard to several musicological questions.

In order to classify the contour types of individual phrases, we modify the approach taken by Huron [45] for folk song contours to make it suitable for the very long phrases often encountered in jazz. We first compute the median intensity values v_1, v_2, and v_3 within the first 25 %, the central 50 %, and the final 25 % of the phrase duration over note-wise pitch and intensity values. Then, we compare the segment intensity difference Δv_i via

$$\Delta v_i = \begin{cases} \text{sign}(v_{i+1} - v_i), & \text{if } |v_{i+1} - v_i| \geq \Delta v_{\text{min}} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

using the threshold $\Delta v_{\text{min}} = 0.1\,(\max_i v_i - \min_i v_i)$. Finally, we classify the contour type based on the pairs of segment intensity differences $(\Delta v_1, \Delta v_2)$ using the heuristic indicated in Table VI. Our definition can only be applied to phrases with a minimum length of four notes. Hence, shorter phrases are treated as a special category.

Table VII shows the percentage of different phrase contour types in regard to both the intensity and pitch contours in our dataset. It can be observed that horizontal and concave contour types appear more often in intensity contours while convex and

TABLE VI
HEURISTIC FOR PHRASE CONTOUR SHAPE CLASSIFICATION. EACH PHRASE IS SEGMENTED INTO THREE PARTS (FIRST 25 %, THE CENTRAL 50 %, AND THE FINAL 25 % OF THE PHRASE DURATION). THE PHRASE CONTOUR TYPE IS CLASSIFIED BASED ON THE INTENSITY DIFFERENCES Δv_1 AND Δv_2 BETWEEN ADJACENT SEGMENT PAIRS.

Contour Type	$(\Delta v_1, \Delta v_2)$
Horizontal	$(0,0)$
Ascending	$(0,1)$, $(1,0)$, $(1,1)$
Descending	$(0,-1)$, $(-1,0)$, $(-1,-1)$
Concave	$(-1,1)$
Convex	$(1,-1)$

ascending types occur more often in pitch contours. Around 35% of all intensity and pitch contours are descending.

TABLE VII
PERCENTAGE OF INTENSITY AND PITCH CONTOUR TYPES IN JAZZ SOLO PHRASES.

Contour Type	Intensity	Pitch
$<$ 4 notes	17.1	17.1
Horizontal	6.0	2.9
Convex	12.9	17.0
Concave	14.3	10.2
Ascending	14.1	17.6
Descending	35.7	35.2

Table VIII shows the percentage of co-occurance of intensity and pitch contour types over phrases. All pitch contour types coincide most often with a descending intensity contour (40.7% on average) or with the intensity contour of the same shape (32.4% on average), which reflects the positive correlation between pitch and intensity (see Section IV-D1).

TABLE VIII
PERCENTAGE OF CO-OCCURANCE OF INTENSITY AND PITCH CONTOUR IN JAZZ SOLO PHRASES (MINIMUM LENGTH IS 4 NOTES).

Pitch Contour Type	Intensity Contour Type				
	Horiz.	Conv.	Conc.	Ascend.	Descend.
Horizontal	14.1	7.3	15.7	14.1	48.7
Convex	6.2	26.8	13.5	17.2	36.4
Concave	6.9	9.3	36.6	13.2	34.0
Ascending	7.8	17.4	17.0	29.0	28.8
Descending	7.0	11.7	13.6	12.1	55.6

3) Alternating Eighth-Note Accentuations: In a last experiment, we analyzed intensity differences between first and second eighth notes. Using the metrical annotation included in the Weimar Jazz Database, we identified pairs of successive eighth notes in solos. We performed a paired t-test between the intensity values on the second and first eighths and computed Cohen's d [46] as a measure of effect size. We found 33 solos that showed significant intensity differences on the 5%-level for at least ten successive eight-note pairs. 18 of the 33 solos showed higher intensity values for the first eighth notes. Therefore, in contrast to our findings for a smaller dataset (13 solos with significant intensity differences) in [33], we observe a stronger tendency towards on-beat accentuation. Table IX shows the ten solos with the highest effect size for positive intensity differences (first eighths are louder than second eighths) as well as the ten solos with the highest effect size for negative intensity differences, respectively. It seems that there is a tendency for musicians from traditional styles, swing, and cool jazz to accent the on-beat eighths by playing them more loudly. However, cool jazz trumpeter Chet Baker as well as soprano player Steve Lacy both tend to play the second eighths more loudly, along with several hardbop and bebop players. Therefore, intensity differences of eighth pairs could also be an individual stylistic trait of a particular musician[9] or even depend on other aspects such as the overall tempo.

TABLE IX
TEN SOLOS WITH THE HIGHEST EFFECT SIZE FOR BOTH POSITIVE INTENSITY DIFFERENCES (FIRST EIGHTHS ARE LOUDER THAN SECOND EIGHTHS) AND NEGATIVE INTENSITY DIFFERENCES. THE SIGNIFICANCE LEVEL IS PROVIDED IN THE LAST COLUMN (*** $p < 0.001$, ** $p < 0.05$, * $p < 0.01$).

Performer	Title	Cohen's d	Significance Level
Benny Goodman	Tiger Rag	1.1	***
Coleman Hawkins	Body And Soul	0.9	*
Sidney Bechet	Limehouse Blues	0.9	***
Kenny Dorham	Blues In Be-Bop	0.7	*
Coleman Hawkins	Perdido	0.7	***
Benny Goodman	Whispering	0.7	**
Freddie Hubbard	Speak No Evil	0.6	*
Lee Konitz	Wow	0.6	*
David Liebman	No Greater Love	0.5	*
Woody Shaw	Rosewood	0.5	**
John Coltrane	Mr. P.C.	-0.2	**
Stan Getz	Blues In The Closet	-0.2	**
Clifford Brown	Daahoud	-0.3	*
Steve Lacy	Easy To Love	-0.3	*
Miles Davis	Airegin	-0.3	*
Miles Davis	Blues By Five	-0.3	*
John Coltrane	Blue Train	-0.4	*
Chet Baker	Long Ago And Far Away	-0.4	*
Kenny Dorham	Punjab	-0.4	*
Steve Turre	Steve's Blues	-0.5	*

V. CONCLUSIONS

In this paper, we proposed a score-informed framework for the analysis of monophonic jazz solos, which can be extended to other music styles. Based on high-quality solo transcriptions by music experts, we separated the audio recording into the improvising solo instrument and the accompanying rhythm section. After estimating the tuning frequency from the backing track, fundamental frequency and intensity contours were estimated for each note in the solo using the score-parameters as prior information. From this data, several features, such as modulation range and intonation tendencies were derived and analyzed in conjunction with data taken from existing symbolic representations and manual annotations as well as external metadata.

We demonstrated that this method opens up several new and interesting ways of investigating musicological research questions that are difficult to tackle using conventional methods (such as manual transcriptions). Moreover, the results of our score-informed analysis framework might contribute to the

[9] Note that the two solos "Blues in Be-Bop" and "Punjab" by Kenny Dorham that are both within the highest and lowest effect size ranges are recorded in two different stylistic contexts: the first, recorded in 1946, within bebop style, the second, recorded in 1964, within postbop style.

further development of audio analysis and MIR techniques, such as main melody extraction and audio source separation, in providing more detailed a priori knowledge. However, the exploratory experiments presented in this paper show that, at least in the case of jazz solos, the actual results vary strongly with different context parameters and that there are few overarching trends to be observed. Firstly, the personal style of a performer seems to be the most crucial factor, but overall style and instrument specifics also play a role. Some universal tendencies in melody design, such as a prevalence of convex and descending melodic contours, could be found. This study focused on global distributions of those features while the deliberate and expressive deployment of vibrato, intonation deviations, and dynamics on certain tones and within certain phrases has not been explored. This could be the topic of future detailed analyses of single solos by certain artists based on the extracted intensity and intonation data already included in the Weimar Jazz Database. All in all, our studies demonstrate the very complexity and variety of jazz performances in particular, and of music performances in general, which should be taken into account by further research.

In the future, we would like to extend the set of features extracted from the separated melody track with additional timbre-related features to investigate the "sound" of a performer in more detail. Finally, we plan to compare the performance of our sound-related features with that of symbolic features in the context of artist and style classification tasks.

ACKNOWLEDGMENT

The Jazzomat Research Project is supported by a grant provided by the German Research Foundation (Melodisch-rhythmische Gestaltung von Jazzimprovisationen. Rechner-basierte Musikanalyse einstimmiger Jazzsoli, DFG-PF 669/7-2). The authors would like to thank Meinard Müller, Sebastian Ewert, as well as Matthias Mauch for making the Chroma Toolbox for Matlab and the NNLS Chroma and pYIN VAMP plugins available. Also, thanks to Stefan Balke and Jonathan Driedger for their help in annotating the f_0 contours and to all the jazz and musicology students who transcribed and annotated the jazz solos.

REFERENCES

[1] M. Müller, *Fundamentals of Music Processing*. Springer Verlag, 2015.

[2] "ISO 16:1975 Acoustics - Standard tuning frequency (Standard musical pitch)," International Organization for Standardization, Standard, 1975.

[3] G. Ballou, *Handbook for sound engineers*, 4th ed. Taylor & Francis Ltd., 2008.

[4] M. Mauch, "Automatic chord transcription from audio using computational models of musical context," Ph.D. dissertation, Queen Mary University of London, London, 2010.

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

[6] K. Dressler and S. Streich, "Tuning frequency estimation using circular statistics," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, 2007, pp. 357–360.

[7] J. Serrá, G. K. Koduri, M. Miron, and X. Serra, "Assessing the tuning of sung indian classical music," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami, USA, 2011, pp. 157–162.

[8] S. Dixon, D. Tidhar, and E. Benetos, "The temperament police: the truth, the ground truth, and nothing but the truth," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami, USA, 2011, pp. 281–286.

[9] A. Lerch, "On the requirement of automatic tuning frequency estimation," in *Proceedings of the 7th Internations Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada, 2006, pp. 212–215.

[10] V. Gnann, M. Kitza, J. Becker, and M. Spiertz, "Least-squares local tuning frequency estimation for choir music," in *Proceedings of the 131st Audio Engineering Society (AES) Convention*, New York, NY, USA, 2011, pp. 1–7.

[11] M. Müller and S. Ewert, "Chroma toolbox: Matlab implementations for extracting variants of chroma-based audio features," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami, USA, 2011, pp. 215–220.

[12] J. Salamon, G. Peeters, and A. Röbel, "Statistical characterisation of melodic pitch contours and its application for melody extraction," in *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, Porto, Portugal, 2012, pp. 187–192.

[13] K. Dressler, "Pitch estimation by the pair-wise evaluation of spectral peaks," in *Proceedings of the 42nd AES International Conference on Semantic Audio*, Ilmenau, Germany, 2011, pp. 1–10.

[14] E. Cano, G. Schuller, and C. Dittmar, "Pitch-informed solo and accompaniment separation: towards its use in music education applications," *EURASIP Journal on Advances in Signal Processing*, pp. 1–19, 2014.

[15] J. Sundberg, "Acoustic and psychoacoustic aspects of vocal vibrato," *Speech, Music and Hearing Quarterly Progress and Status Report (STL-QPSR)*, vol. 35, no. 2-3, pp. 45–68, 1994.

[16] T. Liley, "Vibrato: Enhancing the sound," *Clarinet & Saxophone Magazine*, pp. 1–4, 2007.

[17] J. Abeßer, E. Cano, K. Frieler, M. Pfleiderer, and W.-G. Zaddach, "Score-informed analysis of intonation and pitch modulation in jazz solos," in *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Málaga, Spain, 2015, pp. 823–829.

[18] D. Gerhard, "Pitch track target deviation in natural singing," in *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, London, UK, 2005, pp. 514–519.

[19] S. S. Miryala, K. Bali, R. Bhagwan, and M. Choudhury, "Automatically identifying vocal expressions for music transcription," in *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, Brazil, 2013, pp. 239–244.

[20] C. Kehling, J. Abeßer, C. Dittmar, and G. Schuller, "Automatic tablature transcription of electric guitar recordings by estimation of score- and instrument-related parameters," in *Proceedings of the 17th International Conference on Digital Audio Effects (DAFx-14)*, Erlangen, Germany, 2014, pp. 1–8.

[21] A. Gabrielsson, *Psychology of Music*, second edition ed. San Diego: Academic Press, 1999, ch. Music Performance, pp. 501–602.

[22] J. Langner and W. Goebel, "Visualizing expressive performance in tempo–loudness space," *Computer Music Journal*, vol. 27, no. 4, pp. 69–83, 2003.

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

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

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

[26] G. Widmer and W. Goebl, "Computational models of expressive music performance: The state of the art," *Journal of New Music Research*, vol. 33, no. 3, pp. 203–216, 2004.

[27] A. Friberg, E. Schoonderwaldt, and P. N. Juslin, "CUEX : An algorithm for automatic extraction of expressive tone parameters in music performance from acoustic signals," *Acta Acoustica united with Acustica*, vol. 93, no. 3, pp. 411–420, 2007.

[28] A. Lerch, "Software-based extraction of objective parameters from music performances," Ph.D. dissertation, Technische Universität Berlin, 2008.

[29] J. L. Arcos, R. L. de Mántaras, and X. Serra, "Saxex a case-based reasoning system for generating expressive musical performances," *Journal of New Music Research*, vol. 27, pp. 194–210, 1998.

[30] E. D. Scheirer, "Extracting expressive performance information from recorded music," Master's thesis, Massachusetts Institute of Technology, 1995.

[31] 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), Praha, Czech Republic, 2011, pp. 385–388.

[32] R. Ramirez, A. Hazan, E. Maestre, and X. Serra, “A genetic rule-based expressive performance model for jazz saxophone,” *Computer Music Journal*, vol. 32, pp. 38–50, 2008.

[33] J. Abeßer, E. Cano, K. Frieler, and M. Pfleiderer, “Dynamics in jazz improvisation - score-informed estimation and contextual analysis of tone intensities in trumpet and saxophone solos,” in *Proceedings of the 9th Conference on Interdisciplinary Musicology (CIM)*, Berlin, Germany, 2014, pp. 156–161.

[34] F. Auger and P. Flandrin, “Improving the readability of time-frequency and time-scale representations by the reassignment method,” *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[35] T. Abe, T. Kobayashi, and S. Imai, “The IF spectrogram: a new spectral representation,” in *Proceedings of the International Symposium on Simulation, Visualization and Auralization for Acoustic Research and Education (ASVA)*, Tokyo, Japan, 1997, pp. 423–430.

[36] S. Balke, J. Driedger, J. Abeßer, C. Dittmar, and M. Müller, “Towards evaluating multiple predominant melody annotations in jazz recordings,” in *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, 2016, pp. 246–252.

[37] M. Mauch and S. Dixon, “pYIN: A fundamental frequency estimator using probabilistic threshold distributions,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, Florence, Italy, 2014, pp. 659–663.

[38] G. E. Poliner, D. P. W. Ellis, A. F. Ehmann, E. Gómez, S. Streich, and B. Ong, “Melody Transcription From Music Audio - Approaches and Evaluation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1247–1256, 2007.

[39] T. Painter and A. Spanias, “Perceptual Coding of Digital Audio,” *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–515, 2000.

[40] C. Cannam, M. O. Jewell, C. Rhodes, M. Sandler, and M. d’Inverno, “Linked data and you: Bringing music research software into the semantic web,” *Journal of New Music Research*, vol. 39, no. 4, pp. 313–325, 2010.

[41] P. F. Berliner, *Thinking in Jazz: The Infinite Art of Improvisation*, ser. Chicago Studies in Ethnomusicology. University of Chicago Press, 1994.

[42] L. Teal, *The Art of Saxophone Playing*. Princeton, N.J.: Summy-Birchard Music, 1963.

[43] W. J. Dowling and D. Fujitani, “Contour, interval, and pitch recognition in memory for melodies,” *The Journal of the Acoustical Society of America*, vol. 49, no. 2, pp. 524–531, 1971.

[44] D. Müllensiefen and G. Wiggins, “Polynomial functions as a representation of melodic phrase contour,” in *Systematic Musicology: Empirical and Theoretcial Studies*, A. Schneider and A. von Ruschkowski, Eds. Frankfurt a.M.: Peter Lang, 2011, pp. 63–88.

[45] D. Huron, “The melodic arch in Western folksongs,” *Computing in Musicology*, vol. 10, pp. 3–23, 1996.

[46] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. New York University, 1988.

Klaus Frieler Klaus Frieler graduated in theoretical physics (diploma) and received a PhD in systematic musicology in 2008. 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 C4DM, Queen Mary University of London, working on singing intonation with Matthias Mauch and Simon Dixon. Since late 2012, he has been working as a post-doctoral researcher with the Jazzomat Research Project at the University of Music Franz Liszt Weimar. His main research interests are computational and statistical music psychology with a focus on creativity, melody perception, singing intonation, and jazz research. Since 2006, he has also been working as an independent music expert specializing in copyright cases.

Estefanía Cano Estefanía Cano holds degrees in Electronic Engineering (B.Sc), Music Saxophone Performance (B.A), Music Engineering (M.Sc), and Media Technology (Dr.-Ing.). In 2009, Estefanía joined the Fraunhofer Institute for Digital Media Technology IDMT, where she currently works as a post-doctoral researcher on various music information retrieval topics such as sound source separation, audio matching, automatic music transcription, and music education.

Martin Pfleiderer Martin Pfleiderer studied musicology, philosophy, and sociology at Giessen University (1988-93) where he received a doctorate in 1998. He has worked as assistant professor for systematic musicology at Hamburg University (1999-2005), attained a postdoctoral lecture qualification (2006), as well as a temporary professorship at Paderborn University (2007/08) and several guest lectures. Since 2009, he holds the professorship for the history of jazz and popular music at the Department for Musicology, University of Music Franz Liszt Weimar. His research is dedicated to rhythm, singing, and improvisation in jazz and popular music, to the sociology of music, and to the history of jazz and popular music in Germany, Europe, and the USA.

Jakob Abeßer Jakob Abeßer holds a degree in computer engineering (Dipl.-Ing.) from Ilmenau University of Technology. He is a postdoctoral researcher in the Semantic Music Technologies group at Fraunhofer IDMT and obtained a PhD degree (Dr.-Ing.) in Media Technology from Ilmenau University of Technology in 2014. During his PhD, he was a visiting researcher at the Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Finland in 2010. As a research scientist at Fraunhofer, he has experience with algorithm development in the fields of automatic music transcription, symbolic music analysis, machine learning, and music instrument recognition. In addition, he works as a postdoctoral researcher at the Liszt School of Music in Weimar at the Jazzomat Research Project, focussing on the analysis of Jazz solo recordings using MIR technologies.

Wolf-Georg Zaddach Wolf-Georg Zaddach studied musicology, arts administration, and history in Weimar and Jena, and music management and jazz guitar in Prague. After completing his Magister Artium with a thesis on jazz in Czechoslovakia in the 50s and 60s, he worked as assistant professor at the department of musicology in Weimar. Since 2012 he has been working at the Jazzomat Research Project. Since 2014, he has held a scholarship by the German National Academic Foundation (Studienstiftung des deutschen Volkes) for his Ph.D. on heavy and extreme metal in the 1980s GDR/East Germany. He frequently performs live and on recordings as a guitarist.