SCORE-INFORMED ANALYSIS OF INTONATION AND PITCH MODULATION IN JAZZ SOLOS

Jakob Abeßer[1,2] **Estefanía Cano**[2] **Klaus Frieler**[1]
Martin Pfleiderer[1] **Wolf-Georg Zaddach**[1]
[1] Jazzomat Research Project, University of Music Franz Liszt, Weimar, Germany
[2] Semantic Music Technologies Group, Fraunhofer IDMT, Ilmenau, Germany
jakob.abesser@idmt.fraunhofer.de

ABSTRACT

The paper presents new approaches for analyzing the characteristics of intonation and pitch modulation of woodwind and brass solos in jazz recordings. To this end, we use score-informed analysis techniques for source separation and fundamental frequency tracking. After splitting the audio into a solo and a backing track, a reference tuning frequency is estimated from the backing track. Next, we compute the fundamental frequency contour for each tone in the solo and a set of features describing its temporal shape. Based on this data, we first investigate, whether the tuning frequencies of jazz recordings changed over the decades of the last century. Second, we analyze whether the intonation is artist-specific. Finally, we examine how the modulation frequency of vibrato tones depends on contextual parameters such as pitch, duration, and tempo as well as the performing artist.

1. INTRODUCTION

The personal styles of improvising jazz musicians can be described from various musical perspectives. There are several structural or syntactical aspects of the improvised melodic lines, which could be idiosyncratic for a certain musician, e.g., preferred pitches, intervals, scales, melodic contours, rhythms or typical patterns, licks, and formulas. These dimensions can be best explored using a symbolic representation, e.g., Western staff notation or MIDI. However, there are other important aspects, which define personal style and make it recognizable: *timbre* (sound characteristics such as roughness or breathiness), *microtiming* (systematic deviations from the underlying metric structure), *dynamics* (the changes in intensity of tones or phrases), *intonation* (the pitch accuracy with respect to a given tone system), *articulation* (e.g., legato or staccato playing) and *pitch modulation* (the variation of the fundamental frequency within the duration of a tone). Symbolic representation does not reveal information about timbre, intonation, and pitch modulation. Therefore, audio-level analysis of recorded improvisations is necessary to characterize those non-syntactical, expressive dimensions in order to get a comprehensive and exhaustive description of a personal style.

2. GOALS

Polyphonic music recordings exhibit strong spectral and temporal overlaps between harmonic components of different instrument sources. Hence, the transcription and analysis of the individual sound sources remain one of the most challenging tasks in Music Information Retrieval (MIR). We approach this task by using high-quality melody transcriptions provided by music experts as foundation for a score-informed audio analysis. In particular, we use score information for the source separation of the solo instrument from the audio mixture and for the frame-wise tracking of the fundamental frequency of each tone. Our main goal is to investigate, which intonation and modulation strategies are applied by woodwind and brass instrument players in jazz solos.

3. RELATED WORK

Various MIR publications investigate the *intonation* and *tuning* of music recordings, ranging from historic solo harpsichord recordings [5], over classical music recordings [11], to Non-Western music styles such as Carnatic and Hindustani music [17]. The tuning frequency of audio recordings is commonly estimated based on pitch frequencies [6], high-resolution interval histograms [17], or adjustable filterbanks [11]. Just intonation and equal temperament are generally used as reference tunings for the analysis of music performances. Lerch [11] points out that observed tuning deviations can have different reasons ranging from deviation of harmonic frequencies from the equal tempered scale to deviations due to non-equal temperament.

Most automatic music transcription algorithm aim at a symbolic representation of tone events, which are described by distinct onset times, durations, and constant pitches [15]. Some automatic melody extraction algorithms such as proposed in [16] and [6] include an estimation of the tone-wise contours of the fundamental frequency (f_0)

as well, which is an essential pre-processing step for analyzing the applied *frequency modulation techniques*. There are many studies on *vibrato* detection in audio recordings [14], particularly for singing voice [8,9,12]. Other publications deal with analyzing the deviation of f_0 contours from the target pitch [8] as well as with segmenting f_0 contours based on modulations such as vibrato and *pitch glides* [12] or *bendings* [10]. To the best knowledge of the authors, no publication so far analyzes intonation and modulation techniques in recorded jazz solos.

4. METHOD

Figure 1 gives an overview over our analysis approach, all processing steps are detailed in the following sections. Section 4.1 describes the dataset of jazz solo audio excerpts and transcriptions. Two separate score-informed analysis techniques are involved. At first, a *source separation* algorithm is performed (see Section 4.2), which separates the original audio recording into a solo track containing the improvising solo instrument and a backing track containing the accompanying band, i.e., the rhythm section (most often piano, double bass, and drums). The backing track is used to estimate the *reference tuning frequency* (see Section 4.4). The second step is the *tracking of frame-wise f_0 contours* for each note played by the solo instrument (see Section 4.3). Based on the extracted f_0 contours, we compute several *contour features* (see Section 4.5) to describe their temporal shape. In the experiments reported in Section 5, we analyze how these features depend on contextual parameters such as tone duration and pitch and whether these might be specific for the personal style.

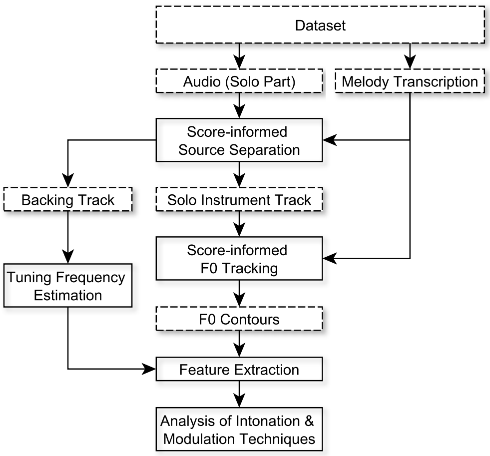

Figure 1: Proposed algorithm for score-informed analysis of tuning and modulation in improvised jazz solos.

4.1 Dataset & Melody Annotations

The dataset used in this publication is a subset of 207 jazz solos taken from the *Weimar Jazz Database* [1]. Table 1 lists all musicians in the dataset with their instrument, the number of solos N_{S}, and the total number of tones and f_0 contours N_{N}, respectively. The solos were manually annotated by musicology and jazz students based on excerpts from commercial audio recordings. The annotations include score-level melody transcription (MIDI pitch, tone onset, and duration) as well as additional annotation layers with respect to melody phrases, metric structure, chords, and modulation techniques. So far, the tone-wise annotations of modulation techniques are incomplete and only represent the most clear examples within the solos. Figure 2 gives an overview over the number of annotated tones per artist. In total, 87643 tones and f_0 contours are included in the dataset.

[1] `http://jazzomat.hfm-weimar.de` (last accessed Juli 10, 2015)

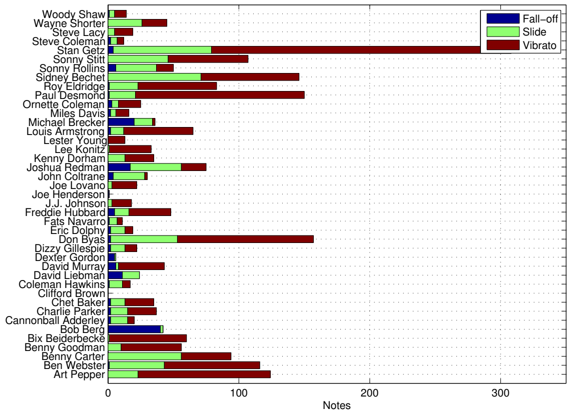

Figure 2: Number of tones of each artist which are annotated with fall-off, slide, and vibrato.

4.2 Score-informed Source Separation

To separate the solo/lead instrument from the backing track, the method for pitch-informed solo and accompaniment separation proposed in [4] was used. For this study, the automatic pitch detection stage in the separation algorithm was bypassed, and the manual melody transcriptions were used as prior information. The separation method is based on an iterative modeling of the solo instrument in the spectral domain. The model is constructed taking into account characteristics of musical instruments such as common amplitude modulation, inharmonicity, and enforcing magnitude and frequency smoothness constraints in the estimation. The separation method has proven to be robust in the extraction of a great variety of solo instruments, as well as being particularly efficient, with computation times that allow real-time processing. The complete dataset was processed and independent signals for the solo instruments and the backing tracks were extracted.

4.3 Score-informed f_0 tracking

The original audio recordings are processed at a sampling rate of 22.05 kHz. In order to track the f_0 contour of each tone, the signal is analyzed between the annotated note onset and offset time, for which a *reassigned magnitude spectrogram* $M \in \mathbb{R}_+^{K \times N}$ with K frequency bins and N frames

Performer	Inst.	N_S	N_N	Performer	Inst.	N_S	N_N	Performer	Inst.	N_S	N_N	Performer	Inst.	N_S	N_N
Art Pepper	cl/as	4	2134	David Liebman	ss/ts	4	3286	John Coltrane	ts/ss	11	8969	Sidney Bechet	ss	2	489
Ben Webster	ts	4	1497	David Murray	ts	4	2295	Joshua Redman	ts	5	2429	Sonny Rollins	ts	10	4639
Benny Carter	as	3	1153	Dexter Gordon	ts	4	3056	Kenny Dorham	tp	6	2149	Sonny Stitt	ts	4	1284
Benny Goodman	cl	7	1154	Dizzy Gillespie	tp	4	967	Lee Konitz	as	4	1839	Stan Getz	ts	6	3253
Bix Beiderbecke	tp	4	519	Don Byas	ts	7	2022	Lester Young	ts	4	887	Steve Coleman	as	3	1353
Bob Berg	ts	5	3275	Eric Dolphy	as	2	1109	Louis Armstrong	tp	4	634	Steve Lacy	ss	4	1437
Cannonball Adderley	as	5	2623	Fats Navarro	tp	4	937	Michael Brecker	ts	4	2605	Wayne Shorter	ts	9	3013
Charlie Parker	as	6	1688	Freddie Hubbard	tp	6	2266	Miles Davis	tp	7	2080	Woody Shaw	tp	5	1822
Chet Baker	tp	6	1100	J.J. Johnson	tb	2	754	Ornette Coleman	as	3	1782				
Clifford Brown	tp	4	1676	Joe Henderson	ts	6	3830	Paul Desmond	as	8	2142				
Coleman Hawkins	ts	6	2613	Joe Lovano	ts/ts-c	2	1787	Roy Eldridge	tp	6	1744				

Table 1: Overview over all artists in the dataset. For each artist, the number of solos N_S, the total number of notes N_N, as well as the instrument is given (ts: tenor saxophone, ss: soprano saxophone, as: alto saxophone, cl: clarinet, tp: trumpet, cor: cornet, tb: trombone, ts-c: C melody tenor saxophone).

is computed. We use a logarithmic frequency axis with a high resolution of 50 bins/semitone and a frequency range of $\pm$ 2 semitones around the annotated pitch. Based on an initial short-time Fourier transform (STFT) with a blocksize of 1024, a hopsize of 64, and a zero-padding factor of 16, the magnitude values are mapped (reassigned) towards the frequency bins that correspond to their instantaneous frequency values at the original frequency bins computed using the method proposed by Abe in [1]. Two steps are performed for each tone to estimate its f_0 contour. First, we estimate a suitable *starting frame* within the tone's duration with a prominent peak close to the annotated pitch. Second, we perform a *contour tracking* both forwards and backwards in time. Further details are provided in [3].

4.4 Tuning Frequency Estimation

The oldest recordings in our dataset date back to the year 1924, two years before the American music industry recommended 440 Hz for A4 as standard tuning, and 12 years before the American Standards Association officially adopted it. Hence, we can not rely on the assumption of a constant and fixed overall tuning. Moreover, the technical level of recording studios were rather low at this time, which might result in tuning deviations by speed fluctuations of recording machines as well as from instruments tuned to another reference pitch such as studio or live venue pianos. Hence, we estimate a *reference tuning frequency* f_ref prior to the intonation analysis of the solo instrument from the backing track of the rhythm section, which we obtain from the source separation process explained in Section 4.2. The reference tuning frequency corresponds to the fundamental frequency of the pitch A4 in the backing track.

In the Chroma Toolbox [13], a triangular filterbank is generated based on a given tuning frequency in such way that its center frequencies are aligned to the chromatic scale within the full piano pitch range. For a given audio signal, the magnitude spectrogram is averaged over the full signal duration and processed using the filterbank. By maximizing the filterbank output energy over different tuning frequency hypotheses, a final tuning frequency estimate $\hat{f}_\mathrm{ref}$ is derived. We modified the originally proposed search range for $\hat{f}_\mathrm{ref}$ to 440 Hz± 0.5 semitone (corresponding MIDI pitch range: 69 ± 0.5) and the stepsize to 0.1 cents. As will be shown in Section 5.1, the influence of source separation artifacts on the estimation accuracy of the reference tuning frequency can be neglected.

4.5 Feature Extraction

Based on the estimated contour $f_0(n)$ of each tone, we first perform a smoothing using a two-element moving average filter in order to compensate for local irregularities and possible estimation errors. The extracted audio features describe the *gradient* of the f_0 contour as well as its temporal *modulation*. Table 2 lists all computed audio features and their dimensionality.

Category	Feature Label	Dim.
Gradient	Linear slope	1
Gradient	Median gradients (first half, second half, overall)	3
Gradient	Ratio of ascending frames	1
Gradient	Ratio of ascending / descending / constant segments	3
Gradient	Median gradient of longest segments	1
Gradient	Relative duration of longest segments	1
Gradient	Pitch progression	1
Modulation	Modulation frequency [Hz]	1
Modulation	Modulation dominance	1
Modulation	Modulation range [cent]	1
Modulation	Number of modulation periods	1
Modulation	Average relative / absolute f_0 deviation	2
Modulation	f_0 deviation inter-quartile-range	1

Table 2: Summary of audio features to descript the f_0 contours.

4.5.1 *Gradient features*

Based on the gradient $\Delta f_0(n) = f_0(n{+}1) - f_0(n)$, we first determine frames and segments of adjacent frames with ascending ($\Delta f_0(n) > 0$), descending ($\Delta f_0(n) < 0$), and constant frequency. We use the relative duration (with respect to the note duration) of each segment class as features. Also, we compute median gradients in the first and second halves, over the whole note, as well as over the longest segment. Overall pitch progression is measured by the difference of average f_0 values in the end and beginning of each tone. Furthermore, we use linear regression to estimate the linear slope of the f_0 contour.

4.5.2 *Modulation features*

We analyze the modulation of the f_0 contour by computing the autocorrelation over $f_0(n)$. Fletcher [7] reported for woodwind instruments that a vibrato frequency range between 5 and 8 Hz is comfortable for listeners and common for players. We add a safety margin of 2 Hz and search for the lag position $\tau_{\max}$ of the highest local maximum within the lag range that corresponds to fundamental frequency values of $f_{\text{mod}} \in [3, 10]$ Hz and estimate the modulation frequency as $\hat{f}_{\text{mod}} = 1/\tau_{\max}$. The difference between the maximum and median magnitude within this frequency band is used as dominance measure for the modulation. Other applied features are the number of modulation periods and the frequency modulation range in cent.

4.6 Analysis of Intonation and Modulation Techniques

We distinguish three modulation techniques *fall-off*, *slide*, and *vibrato*. Table 3 provides a description of the characteristic f_0 contour shape for each technique. The number of tones in our dataset annotated with each technique is given in Table 3.

Technique	Description	Notes
Fall-off	Drop of the f_0 contour in the end of the tone after a stationary part.	146
Slide	Rise or drop of the f_0 in the beginning of the tone towards a stationary part.	708
Vibrato	Periodic modulation of the f_0 contour during the stationary part of the tone.	1380
None	No discernible modulation of the f_0 contour / No modulation technique annotated.	83587

Table 3: Frequency modulation techniques considered and number of annotated notes in the dataset for each technique.

5. RESULTS

5.1 Influence of Source Separation on the Reference Tuning Estimation

After the application of source separation algorithms, parts of the isolated solo instrument often remain audible in the backing track due to artifacts or interference. We first investigated the influence of the source separation step described in Section 4.2 on the reference tuning estimation. A subset of 13 solos was randomly selected from the dataset, covering various recording decades and solo instruments. For each solo, we took a 20s segment from the original recording, where only the rhythm section and no solo instrument is playing. We used the tuning estimation method described in Section 4.4 on both this 20s segment as well as on the backing track obtained from the source separation of the solo part (compare Section 4.2) to get two estimates $f_{\text{ref}}^{\text{NoSolo}}$ and $f_{\text{ref}}^{\text{Backing}}$ of the reference tuning frequency.

The results show a very high sample correlation of $r = 0.97$ ($p < 0.001$) and a small root mean squared error of $\text{RMSE} = 1.05$ Hz between both estimates. These results indicate that the influence of source separation artifacts is negligible for the tuning estimation process. Therefore, we will use $\hat{f}_{\text{ref}} = f_{\text{ref}}^{\text{Backing}}$ as an estimate of the reference tuning frequency throughout the paper.

5.2 Relationship between the Reference Tuning and the Recording Year / Decade

How did the tuning frequency f_{ref} of commercial jazz recordings change during the 20th century? Figure 3 shows the distribution of solos in the dataset over the from the 1920s to the 2000s. Moreover, the inserted boxplots illustrate the deviation $\Delta f = 1200 \log_2 \frac{f_{\text{ref}}}{440}$ between the tuning frequency f_{ref} and 440 Hz in cent.

Absolute tuning deviation $|\Delta f|$ and recording year of each solo are weakly negatively correlated ($r = -0.33$, $p < 0.001$). Hence, the absolute deviation from the tuning frequency from 440 Hz decreased over the the course of the 20th century, reflecting the spread of the 440 Hz standard (1955 adopted by the International Standards Organization), as well as the progress of studio technology.

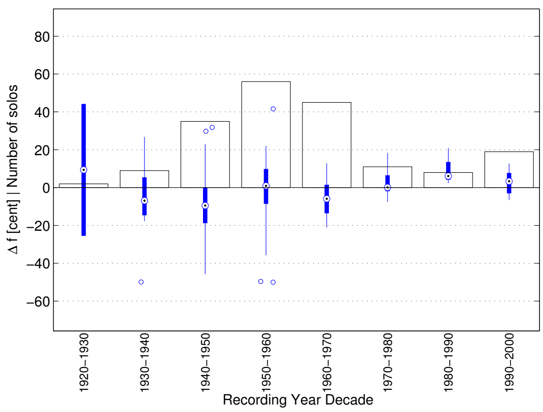

Figure 3: The box plot shows the reference tuning deviation from 440 Hz in cent for different recording year decades. The bars show the number of solos in the dataset for each decade.

5.3 Dependency of Intonation from Artist and Instrument

The distribution of the absolute deviation of tone-wise fundamental frequency values from the estimated tuning frequency as well as modulation range are shown for all musicians in Figure 4, and for all instruments in Figure 5.

According to Figure 4, the overall pitch intonation of jazz musicians is astonishingly accurate. Some woodwind and brass players tend to play a bit sharp, few a bit flat—but throughout in a range of less than 25 cent. There are few exceptions: Sidney Bechet, a traditional soprano saxophonist, has very high values; however, presumably this is caused not by a sharp intonation but by the high percentage of pitch slides played by him (almost 15 % of the tones, cf. Figure 2).

For most players, the range of frequency modulation, i.e., the size of vibrato, is around 25 cent. There are some bigger modulation ranges from 35 to 50 cent, predominantly used by tenor saxophone players associated with swing style (Ben Webster, Coleman Hawkins, Don Byas, and Lester Young), but also by postbop tenor saxophonist Joe Lovano, and, again, by Sidney Bechet, showing the largest variance of modulation ranges. Therefore, there are some slight personal and stylistic peculiarities in the use of vibrato size. However, there are no obvious trends of intonation according to different instruments (cf. Figure 5), since for each instrument there seem to be players who play a bit sharp as well as players who play a bit low; note that for trombone and c-melody sax there is only one musician (J.J. Johnson resp. Joe Lovano) in our sample. Likewise, there is no evidence for general trends of modulation ranges with respect to instrument.

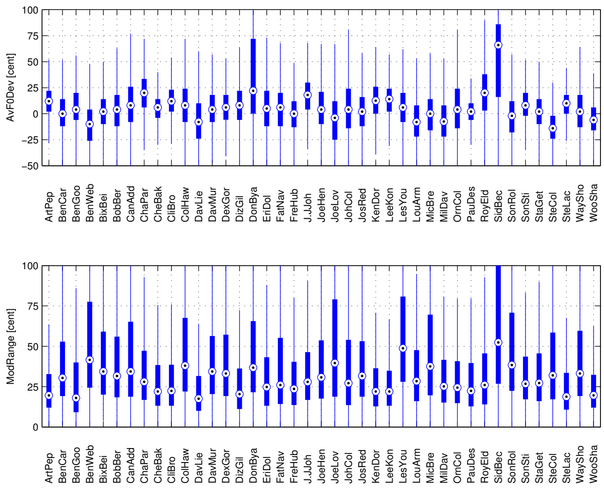

Figure 4: Absolute deviation of tone-wise fundamental frequency values from the estimated tuning frequency in cent and modulation range in cent for all musicians (for their full names see Table 1).

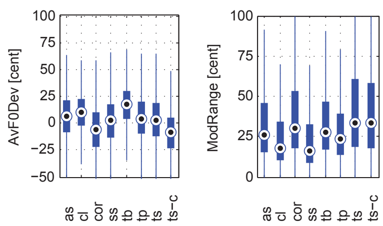

Figure 5: Absolute deviation of note-wise fundamental frequency values from the estimated tuning frequency in cent and modulation range in cent for all instruments.

5.4 Context-dependency of the Modulation Frequency of Vibrato

Does the modulation frequency of vibrato depend on pitch, or duration of the vibrato tones, or on the tempo of the piece? For the 1380 tones with vibrato notes (cf. Table 3), we found no significant correlations between modulation frequency and pitch ($r = 0.02$, $p = 0.42$), duration ($r = 0.02$, $p = 0.5$), nor tempo ($r = 0.0$, $p = 0.83$). The small effect size of the correlation indicates that despite the high variety of tempo values in the dataset (mean tempo 154.52 bpm, standard deviation 68.16 bpm), the modulation frequency only slightly increases with increasing tempo.

Furthermore, we investigated, whether and how the modulation frequency of vibrato is connected to the underlying metrical structure of a solo. We computed the ratio $r = T_{\text{mod}}/T_{\text{solo}}$ between the modulation tempo and the average tempo of the solo. The modulation tempo is computed as $T_{\text{mod}} = 60 f_{\text{mod}}$. Figure 6 shows the ratio r against the average tempo of the solo. There is no evidence in our data for a strategy to adapt the modulation frequency of vibrato to integer multiples of the tempo of the piece, e.g., to use a vibrato speed according to simple subdivision of the beat (e.g. eighth notes or eighth triplets). As Figure 6 shows, for medium and fast tempos (100 to 350 bpm) the vibrato frequency varies only between the beat and the 16th note level. For slower tempos, the vibrato tempo could be up to six or seven times as fast as the beat—but rarely much faster.

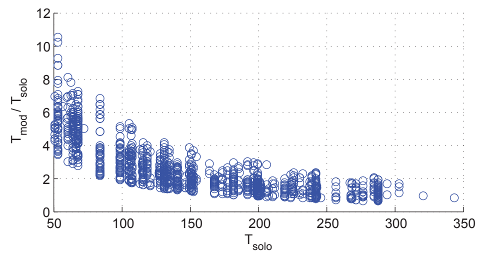

Figure 6: Ratio between the modulation frequency of vibrato tones and the average tempo of a piece vs. the average tempo.

5.5 Artist-dependency of the Modulation Frequency of Vibrato

Although there is no obvious correlation between modulation frequency and pitch, duration, or tempo, there are some peculiarities of musicians according to vibrato modulation speed. In Figure 7 only those musicians are included for which more than twenty annotated vibrato tones could be found in our data set. All in all, there seems to be no clear correlation between vibrato speed and jazz style or instrument, which indicates that modulation technique is mostly an idiosyncratic part of personal styles. Strikingly, several trumpet players can be found there (Louis Armstrong, Kenny Dorham, Roy Eldridge) using vibrato to an considerable amount and size. This is in sharp contrast to

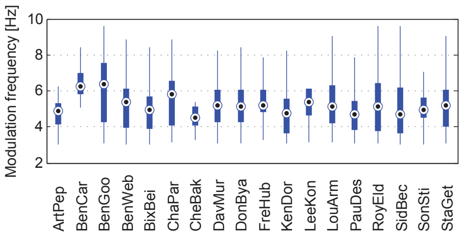

Figure 7: Modulation frequency in Hz in vibrato notes for different performers. Only performers with more than 20 vibrato notes are shown.

playing standards for brass instruments in classical music, where it is custom to play without any vibrato [7].

5.6 Automatic Classification of Frequency Modulation Techniques

Using the set of features discussed in Section 4.5, we extracted an 18-dimensional feature vector for each tone, which was used to automatically classify tones with respect to their modulation class. To this end, we only considered tones annotated with fall-off, slide, and vibrato since all remaining tones were not explicitly annotated. We used a Support Vector Machine (SVM) classifier with a linear kernel function as classification algorithm and perform a 10-fold cross-validation. Due to the imbalanced class sizes (cf. Table 3), we repeatedly re-sampled from the existing class items such that all classes have the same number of items as the largest class from the original dataset.

The confusion matrix is shown in Table 4. The highest accuracy of 92.25 % was achieved for vibrato tones. The classes fall-off and slide show lower accuracy values of 48.04 % and 67.32 %, respectively. One might assume, that the similar f_0 contour shapes of fall-offs and the slide-downs causes part of the confusions between both classes.

Correct	Classified		
	Fall-off	**Slide**	**Vibrato**
Fall-off	**48.04**	37.46	14.49
Slide	23.55	**67.32**	9.13
Vibrato	4.06	3.7	**92.25**

Table 4: Confusion matrix for the automatic classification of frequency modulation techniques. All values are given in percent.

6. CONCLUSIONS

In this exploratory study, we proposed a score-informed algorithm for the extraction of non-syntactical features in jazz solos played with wind and brass instruments. This method allows for an analysis of performative and expressive aspects of jazz improvisation which are not captured by the traditional approaches of jazz research such as transcriptions (even though some rudimentary notation for f_0-modulations are used sometimes). Combining transcriptions with state-of-art MIR algorithms significantly enhances the methodical and analytical tool box of jazz research (as well as other subfields of musicology and performance studies). In turn, this kind of fine-structured analysis might be useful in guiding automatic transcription algorithms by providing relevant background information on tone characteristics. Moreover, in this study we demonstrated exemplarily that our method can be readily applied for a range of different research questions, from historical analysis of reference tuning in 20th century jazz recordings to more general questions such as intonation accuracy or differences in f_0 modulations with respect to tempo, instrument class, stylistic trends, or personal style.

As a case study, we investigated whether some these expressive aspects, i.e., intonation, slides, vibrato speed and vibrato range, are correlated with structural features of the solos (absolute pitch, tone duration, overall tempo, meter) and whether those aspects are characteristic for an instrument, a jazz style or the personal style of a musician. While there is little evidence for a general correlation between intonation and pitch modulation (slide, vibrato) on the one hand, and structural features on the other hand, the issue of how intonation and pitch modulation contributes to the formation of a jazz style and personal style needs further examination with more data and including listening tests for style discrimination.

For the future, we plan to complete and refine the f_0-modulation annotations for the dataset, with the overall goal of the design of an automated f_0-modulation annotation algorithm. Finally, we aim at a complete description of personal timbre characteristics, the so-called "sound" of a player, which is an important dimension of jazz music, and not yet fully addressed. Dynamics [2], intonation, articulation, and f_0-modulation are part of this "sound", but other aspects such as breathiness, roughness and general spectral characteristics (and their classification) are still to be explored.

7. ACKNOWLEDGEMENTS

The Jazzomat research project is supported by a grant DFG-PF 669/7-1 ("Melodisch-rhythmische Gestaltung von Jazzimprovisationen. Rechnerbasierte Musikanalyse einstimmiger Jazzsoli") by the Deutsche Forschungsgemeinschaft (DFG). The authors would like to thank all jazz and musicology students participating in the transcription and annotation process.

8. REFERENCES

[1] Toshihiko Abe, Takao Kobayashi, and Satoshi Imai. Harmonics tracking and pitch extraction based on instantaneous frequency. In *Proceedings of the 1995 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 756–759, Detroit, USA, 1995.

[2] Jakob Abeßer, Estefanía Cano, Klaus Frieler, and Martin 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.

[3] Jakob Abeßer, Martin Pfleiderer, Klaus Frieler, and Wolf-Georg Zaddach. Score-informed tracking and contextual analysis of fundamental frequency contours in trumpet and saxophone jazz solos. In *Proceedings of the 17th International Conference on Digital Audio Effects (DAFx-14)*, Erlangen, Germany, 2014.

[4] Estefanía Cano, Gerald Schuller, and Christian Dittmar. Pitch-informed solo and accompaniment separation: towards its use in music education applications. *EURASIP Journal on Advances in Signal Processing*, 23:1–19, 2014.

[5] Simon Dixon, Dan Tidhar, and Emmanouil 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)*, pages 281–286, Miami, USA, 2011.

[6] Karin Dressler. Pitch estimation by the pair-wise evaluation of spectral peaks. In *Proceedings of the 42nd AES International Conference on Semantic Audio*, pages 1–10, Ilmenau, Germany, 2011.

[7] N. H. Fletcher. Vibrato in music – physics and psychophysics. In *Proceedings of the International Symposium on Music Acoustics*, pages 1–4, 2010.

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

[9] Chao-Ling Hsu and Jyh-Shing Roger Jang. Singing pitch extraction by voice vibrato/tremolo estimation and instrument partial deletion. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, pages 525–530, Utrecht, Netherlands, 2010.

[10] Christian Kehling, Jakob Abeßer, Christian Dittmar, and Gerald 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.

[11] Alexander Lerch. On the requirement of automatic tuning frequency estimation. In *Proceedings of the 7th Internations Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada, 2006.

[12] Sai Sumanth Miryala, Kalika Bali, Ranjita Bhagwan, and Monojit Choudhury. Automatically identifying vocal expressions for music transcription. In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, Brazil, 2013.

[13] Meinard Müller and Sebastian 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)*, pages 215–220, Miami, USA, 2011.

[14] T. Özaslan, Serra X., and Arcos J. L. Characterization of embellishments in ney performances of makam music in turkey. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, pages 13–18, 2012.

[15] Matti P. Ryynänen and Anssi Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 32:72–86, 2008.

[16] Justin Salamon, Geoffroy Peeters, and Axel Röbel. Statistical characterization 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.

[17] Joan Serrá, Gopala K. Koduri, Marius Miron, and Xavier Serra. Assessing the tuning of sung indian classical music. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, pages 157–162, Miami, USA, 2011.