COMPUTATIONAL CORPUS ANALYSIS: A CASE STUDY ON JAZZ SOLOS

Christof Weiß[1] **Stefan Balke**[1] **Jakob Abeßer**[2] **Meinard Müller**[1]
[1] International Audio Laboratories Erlangen, Germany
[2] Semantic Music Technologies Group, Fraunhofer IDMT, Ilmenau, Germany
christof.weiss@audiolabs-erlangen.de

ABSTRACT

For musicological studies on large corpora, the compilation of suitable data constitutes a time-consuming step. In particular, this is true for high-quality symbolic representations that are generated manually in a tedious process. A recent study on Western classical music has shown that musical phenomena such as the evolution of tonal complexity over history can also be analyzed on the basis of audio recordings. As our first contribution, we transfer this corpus analysis method to jazz music using the Weimar Jazz Database, which contains high-level symbolic transcriptions of jazz solos along with the audio recordings. Second, we investigate the influence of the input representation type on the corpus-level observations. In our experiments, all representation types led to qualitatively similar results. We conclude that audio recordings can build a reasonable basis for conducting such type of corpus analysis.

1. INTRODUCTION

Characterized by keywords such as *systematic musicology* or *computational music analysis*, quantitative and data-driven methods have recently gained importance within musicology. As one central benefit, computational methods enable corpus-based studies on a large scale. Several studies have been conducted recently for different music genres including pop music [13], jazz [1, 6, 9], and Western classical music [2, 17, 21, 24], and also in the field of ethnomusicology [14, 16, 19]. For conducting such corpus studies, a number of different aspects are important. Besides methodological questions such as the musical characteristics under investigation (e. g., melodic, harmonic, or rhythmic aspects), also the way these characteristics are measured, evaluated, and presented matters. Moreover, the corpus itself plays a crucial role. Beyond its size and composition, the representation of the music data constitutes an important aspect. For example, the data can be given as a symbolic transcription [9, 16], as a graphical score [17], or as an audio recording [6, 13, 18].

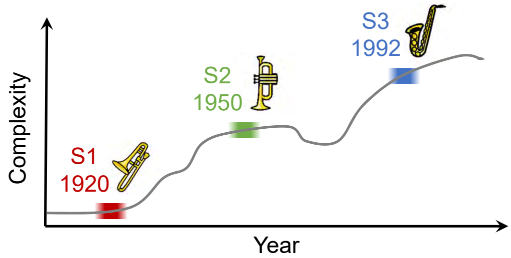

Figure 1. Procedure for mapping feature values from individual solos onto the timeline using the recording years.

In this paper, we investigate the influence of the music representation type on the corpus analysis results. For this purpose, we present a case study for jazz music using the solos contained in the *Weimar Jazz Database* [15]. As an example for a corpus analysis, we investigate the tonal complexity of the jazz solos using a measure introduced in [22]. Inspired by recent work on pop [13] and classical music [21], we apply a visualization technique where quantitative descriptors for individual pieces are mapped onto a timeline as shown in Figure 1. The resulting *evolution curves* [21] allow for studying the evolution of musical phenomena (here: tonal complexity) over history.

As input data for this study, we compare different representations of the jazz solos including a high-quality symbolic transcription of the solo melody as well as the full mix audio recording of the solo section. Furthermore, we investigate intermediate representations, which rely on signal processing techniques [3,4,7,8] for enhancing the presence of the solo instrument and for suppressing accompanying instruments and audio-specific artifacts. Specifically, we consider the approaches proposed in [4, 7]. Though the music representations—as well as the derived features—exhibit a different behavior on the *piece level*, our experiments show that on the *corpus level*, results are qualitatively similar for audio-based procedures and for analyses based on high-quality symbolic transcriptions. Our findings encourage to perform corpus studies on the basis of audio recordings. This opens up new ways for musicological research since audio recordings are available easily without an extensive transcription or annotation process that often needs to be done manually.

The remainder of this paper is structured as follows. First, we describe our music scenario and sketch some musicological hypotheses (Section 2). Second, we detail on

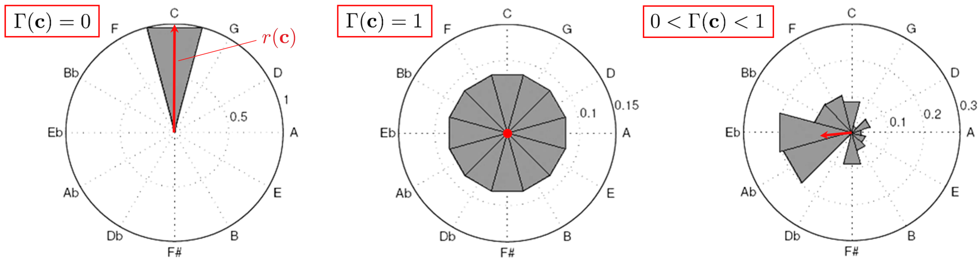

Figure 2. Complexity measure Γ based on the circle of fifths. Values for a sparse chroma vector (left), a flat chroma vector (middle), and a more realistic chroma vector (right) are shown. The red arrows denote the resultant vectors.

our tonal complexity measure and explain its musical implications (Section 3). We then describe the different representations and signal processing techniques we use in this study (Section 4). Next, we describe our corpus analysis strategy and present the experimental results (Section 5). Finally, we discuss the implications of our findings.

2. JAZZ SCENARIO

Within the scope of jazz music, the *Weimar Jazz Database* (WJD) with its 456 manually generated transcriptions of famous jazz solos constitutes a unique dataset [15]. A major benefit of the WJD lies in its clean annotations of the solo melody (fundamental frequency, F0), which create a controlled environment for systematic experiments. The data served as basis for a number of musicological studies, which mainly focus on performance analysis [1, 6, 9].

Besides the rhythmical aspects of the solos [6] and the melodic phrasing [9], also the played pitch material (scales) can be of musicological interest. In our experiments, we consider this dimension by measuring the tonal complexity of the pitches played by the soloist. We expect to find a lower tonal complexity for solos from the *Chicago Jazz* era (1920s), compared to, for instance, *Bebop* solos from the 1950s. However, there might be some outliers in each period. For example, Chet Baker's intimate solos will probably obtain lower complexity values than Clifford Brown's solos—although both perform in the same period.

3. MEASURING TONAL COMPLEXITY

The analysis of music complexity has been an important task within MIR research in the past years. Streich [20] tackled multiple dimensions of this notion denoted as acoustic, timbral, rhythmic, and tonal complexity. Concerning tonality, many studies [5, 12, 20] focus on sequential complexity aspects such as the complexity of chord progressions [5]. As opposed to this, chroma-based complexity measures were introduced in [22], which locally describe the pitch class distribution without explicitly capturing transitional characteristics. Despite their simplicity, these features have shown a high correspondence to an intuitive understanding of music complexity over the course of an individual piece [22]. Beyond that, they have turned out to be useful for classifying music recordings according to style categories [23]. Averaging such complexity features over many pieces provides meaningful and stable results, which has been shown in a large-scale study of musical evolution in classical music [21]. As one contribution, we transfer this concept to jazz music and show that complexity features also yield meaningful results for this scenario. In contrast to [21], the WJD scenario provides data in different representations (see Section 4), whose influence we want to investigate. Moreover, we have detailed metadata such as the recording year of each solo.

The complexity measures introduced in [22, 23] describe statistical properties of an underlying normalized chroma distribution. Flat distributions result in high complexity values while sharp distributions result in low ones. In [23], several different measures are introduced for this purpose such as entropy-, sparsity-, and flatness-based quantities. Here, we restrict ourselves to one feature that additionally accounts for the tonal relationship of the prominent pitch classes. Following [23], we now summarize the definition of this measure $\Gamma : \mathbb{R}^{12} \rightarrow [0, 1]$. Let $\mathbf{c} = (c_0, c_1, \ldots, c_{11})^{\mathrm{T}} \in \mathbb{R}^{12}$ denote a chroma vector with positive entries ($c_n \geq 0$) normalized with respect to the ℓ^1-norm $\left(\sum_{n=0}^{11} c_n = 1\right)$. The entries c_n with $n \in [0:11]$ indicate the salience of the twelve pitch classes C, C$^\sharp$, ..., B, respectively. Because of octave invariance, the features show a cyclic behavior so that a transposition in pitch leads to a circular shift.

For computing the complexity $\Gamma(\mathbf{c}) \in [0, 1]$ of a chroma vector $\mathbf{c} \in \mathbb{R}^{12}$, we first re-sort the chroma values to an ordering of perfect fifth intervals (7 semitones) resulting in the vector $\mathbf{c}^{\mathrm{fifth}}$ defined by:

$$c_n^{\mathrm{fifth}} = c_{(n \cdot 7) \bmod 12}. \tag{1}$$

Based on the reordered vector $\mathbf{c}^{\mathrm{fifth}}$, we define the resultant vector $\mathbf{r}(\mathbf{c})$ with a length of

$$r(\mathbf{c}) = \left| \frac{1}{N} \sum_{n=0}^{N-1} c_n^{\mathrm{fifth}} \exp\left(\frac{2\pi i n}{12}\right) \right|. \tag{2}$$

Then, the complexity $\Gamma(\mathbf{c})$ is defined as:

$$\Gamma(\mathbf{c}) = \sqrt{1 - r(\mathbf{c})}. \tag{3}$$

This measure corresponds to the angular deviation and describes the spread of the pitch classes around the circle of fifths. Figure 2 shows the complexity feature and the resultant vector $\mathbf{r}(\mathbf{c})$ (in red) for three input chroma vectors $\mathbf{c}$. For a sparse vector (left), the complexity is minimal ($\Gamma(\mathbf{c}) = 0$). For a flat vector (middle), we obtain maximal complexity ($\Gamma(\mathbf{c}) = 1$).

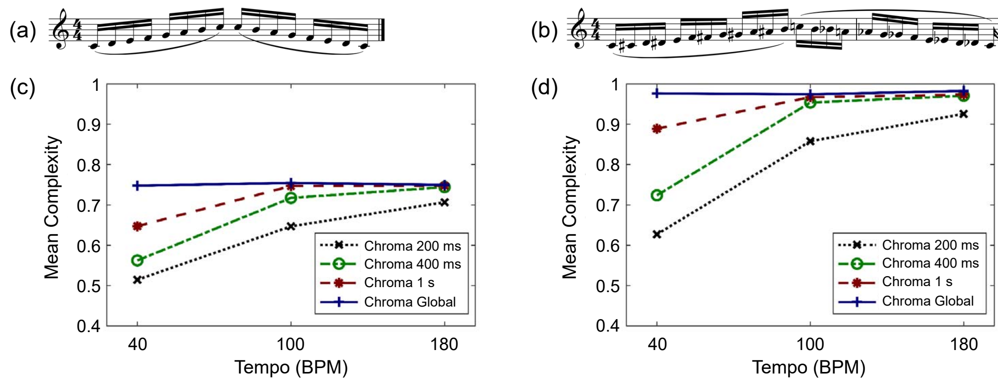

Figure 3. Complexity values for musical scales in several tempi, computed with different window lengths. (a) Diatonic scale. (b) Chromatic scale. (c) Complexity values for the diatonic scale. (d) Complexity values for the chromatic scale.

In this paper, we compute complexity features for jazz solos. Relying on chroma features of the full audio recordings, the features describe the complexity of the overall tonal content—comprising the sounding pitches of the solo instrument as well as the accompanying instruments (e. g., piano, double bass, drums). Since noise-like sounds such as drum hits contribute in an approximately equal fashion to each of the twelve chroma values, this results in an overall increase of complexity. As opposed to the full mix recording, a symbolic transcription of the solo only captures the pitches played by the solo instrument. Since we deal with monophonic solo instruments (mainly saxophone, trumpet, trombone), there is only one non-zero pitch class at a time. Using a small window length (fine-grained resolution) for the chroma features, this results in low complexity values. As soon as we use a larger window length—e. g., by smoothing over several chroma frames—the complexity features are computed from local pitch class histograms and, thus, show mostly non-zero values in case that different pitch classes are played within the analysis interval. Hereby, the feature values depend on the number of pitch classes played but also, on their tonal relationship. Playing many fifth-related pitch classes—such as a diatonic scale—yields a distribution pointing towards a specific direction in the circle of fifths and, thus, results in a rather low complexity value (see Figure 3a and c). For a chromatic scale, in contrast, pitch classes all over the circle of fifths contribute equally resulting in a high complexity value (Figure 3b and d).

Beyond the pitch classes and their relationship, the duration of the notes has a crucial effect on the complexity features. To illustrate this effect, we show in Figure 3 complexity values for scales played in different tempi. For this experiment, we synthesized a diatonic scale and a chromatic scale from music notation software using a saxophone sound. From the generated audio, we computed chroma features in different temporal resolutions. On the basis of these chroma features, we calculated complexity values and averaged these over the full segment. Figures 3c and d show the resulting complexity features for different resolutions and playing tempi. In a higher tempo, more pitch classes are sounding within a window leading to higher complexity. The absolute complexity values also depend on the analysis window length. The four curves in Figures 3c and d refer to different chroma window lengths of 200 ms, 400 ms, 1 s, and a global chroma histogram, respectively. With larger smoothing windows, we obtain higher complexity values. Using global chroma statistics, the complexity is practically independent of the tempo since it always relies on the same pitch class distribution. For a monophonic input signal, our feature captures the tonal complexity of the melody pitches rather than describing a "melodic complexity," which usually accounts for further properties such as direction, jumps, melodic intervals. etc. Despite these simplifications, our complexity feature mostly behaves in a musically meaningful way.

4. INPUT DATA AND PRE-PROCESSING

The complexity feature $\Gamma(\mathbf{c})$ can be computed from different pitch class representations. This enables us to compare the feature values for different representation types. Besides symbolic representations with explicit pitch information, we can also use audio-based chromagrams. [1] In our experiments (Section 5), we investigate how the choice of the input representation influences the complexity features (see Figure 4).

Beyond the symbolic transcription (Figure 4a) created in the Jazzomat project (manual F0 annotation of the solo melody), we consider the full mix audio signal (d), as well as two modified audio versions (b, c). For this, we use signal processing methods to suppress components that might affect our harmony analysis. One such method is harmonic–percussive–residual separation (HPRS) [7], which is an extension of the technique presented by Fitzgerald [8]. HPRS aims to decompose a given audio recording into a harmonic component, a percussive component, and a residual component. The residual component captures portions of the audio recording which are neither of harmonic, nor percussive nature, e. g., noise-like signals such as applause or the breathy component of the saxophone sound. For enhancing the tonal parts of the jazz

[1] In contrast to our complexity measure, high-level measures as presented in [5, 20] often require pre-processing steps that involve challenging tasks such as automatic transcription.

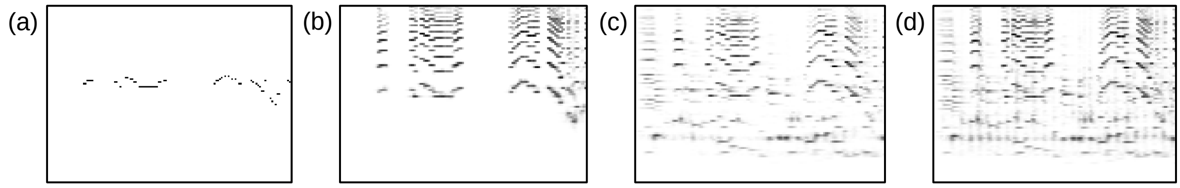

Figure 4. Log-frequency representations of Dexter Gordon's solo from "Society Red" (excerpt of 14 seconds). (a) Symbolic transcription. (b) Source-separated melody (score-informed). (c) Harmonic–Percussive–Residual separation, harmonic part. (d) Full audio mix.

recordings, we use HPRS and throw away both the residual and the percussive components (see Figure 4c).

Beyond this straight-forward separation, we also use a more sophisticated decomposition. Hereby, we try to extract the solo signal from the full mix via source separation. Similar to previous approaches [10, 11], we make use of score information (F0 trajectories) for the separation into solo instrument and backing track [4]. The fundamental frequency trajectory of the solo instrument is used to construct time-variant masks that follow in principle a comb filter structure covering a certain number of the instrument's partials. Several post-processing steps ensure that the bandwidth of the single comb spikes covers the range of the individual partials and that interference from transient sound events is attenuated. Due to the score information, the resulting solo track is almost free of background instruments (see Figure 4b). Only signals that overlap the solo instrument's partials (such as broad-band percussive components) are sometimes perceivable.

From the four representations, we compute pitch class features by summing up energies from different octaves. A comparison of the representation types is interesting since they fundamentally differ from each other in several respects. First, the representations capture different musical parts. Symbolic transcription (a) and source-separated signal (b) only contain the solo instrument, whereas in the other representations, accompaniment is also present. Second, the transcription (a) only contains the fundamental frequency while all other representations also capture overtones. Third, transcription (a) and HPRS-enhancement (c) only capture harmonic information while the separated solo (b) and the full mix (d) also contain residual and percussive components. We will now study how these properties influence a large-scale analysis on the corpus level.

5. CORPUS ANALYSIS

Based on the different types of music representations discussed above, we conduct studies on the tonal complexity of the WJD solos. Inspired by [21], we compute *evolution curves* mapping solo-wise complexity features onto a historical timeline. For this purpose, we use the annotated recording year of each solo. To smooth the curve, we use a soft mapping employing a Gaussian window of size 11 years. Thus, a solo contributes not only to its concrete recording year but also, to a smaller degree, to each 5 years before and after. [2] With this technique, the jazz solos distribute over the timeline as shown in Figure 5a. At about 1955, more than 15 solos contribute on average. Around 1932 (beginning of our timeline) and 2002 (end), there are hardly any solos. This means that a solo contributing to these years has a higher influence on the evolution curve.

To investigate the complexity of the jazz solos, we first analyze each solo individually by computing complexity features in one resolution using the global chroma histogram. In Figure 5b and c, we show these complexity values of individual solos as gray crosses. Figure 5b relies on the symbolic transcription and Figure 5c on the HPRS-enhanced audio (harmonic part). We find a broad range of values for most years. Except for the first 15 years, which do not show very high complexity values, there are solos of diverse complexity at all times. Thus, it is hard to find general structures and trends for individual solos. The overall distribution, however, is similar in both figures.

To analyze this in more detail, we now compute evolution curves. We project the feature value of every piece onto the timeline using the procedure described above. The complexity curves are normalized regarding the number of solos contributing to each year. [3] Figure 5 shows the resulting curves as blue lines. As an additional cue, we compute for the most frequent soloists the complexity value averaged over all their solos, respectively. For each soloist, we plot the average value as horizontal bar from the first to the last solo's recording year. Overall, we observe a slight increase of complexity over the years. The first major increase develops towards the year 1948, where soloists such as Don Byas and Charlie Parker start to contribute. Around the 1960s, we find soloists such as Chet Baker with lower complexity as well as Clifford Brown or Joe Henderson with higher complexity. During the 1970s, there is a major drop, before the complexity again increases towards the early 2000s (David Liebman or Michael Brecker).

Comparing the two curves in Figures 5b and c, we observe that their shape is similar—only the overall scale of the complexity values differs slightly. Most of the prominent changes in complexity can be observed on the basis of both representations—such as the increase around 1945, the drop in the 1970s, and even smaller changes such as the local minimum around 1950. The peak and drop after

[2] The window is normalized so that the total weight of a solo summed up over all 11 years is one.

[3] We sum up the weighted complexity values for all pieces and divide by the number of solos per year as shown in Figure 5a.

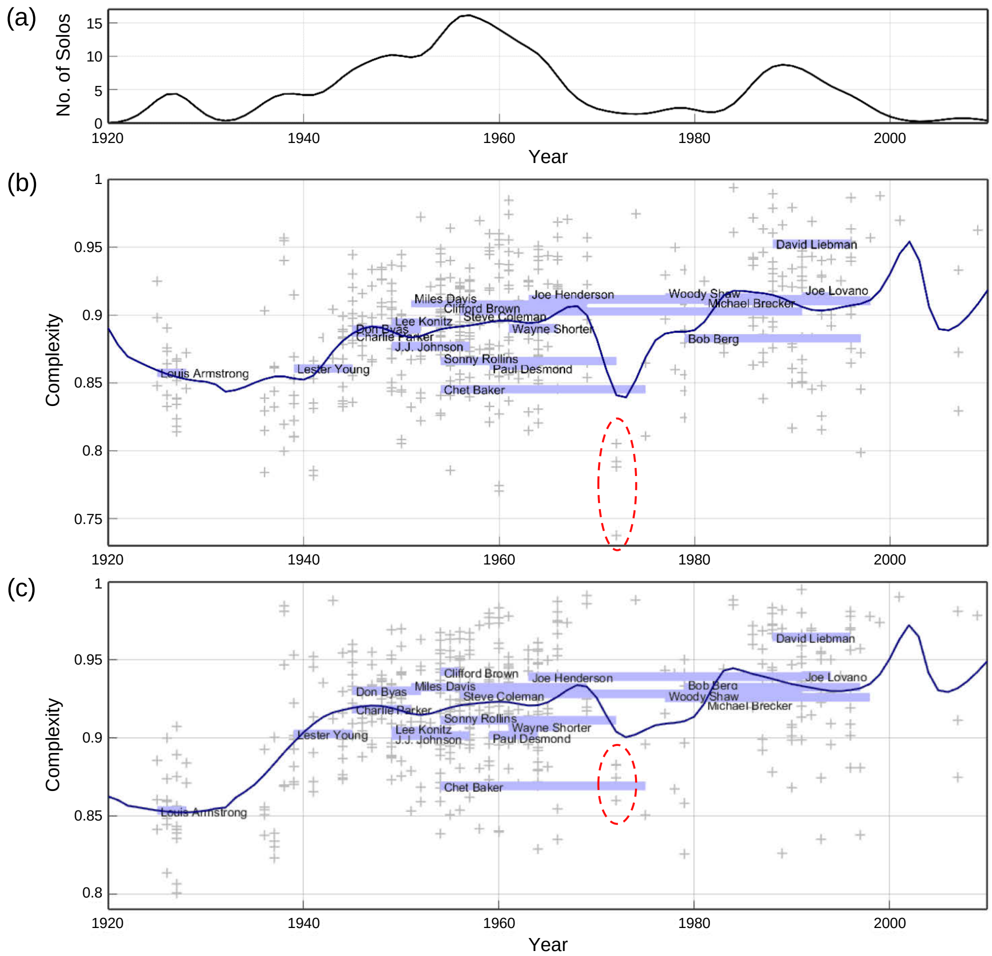

Figure 5. (a) Average number of solos per year contained in the dataset. Evolution curve and artist means based on (b) symbolic transcriptions and (c) harmonic component of audio recordings.

2000 behave very similar. However, we have to take these results with care since only a few solos contribute here. We also find differences between the two plots. For the first years, there are higher values in the symbolic-based plot (b). Here, we could identify several solos with longer silence between the phrases such as Kid Ory's solo in "Gut Bucket Blues." In the symbolic representation, these silent frames are all zero which results in a flat chroma vector (high complexity). This leads to a higher overall complexity of these solos.[4] In the audio-based chromagrams, there are accompanying instruments playing between the phrases, which leads to a lower complexity here. At the year 1972, the drop in Figure 5b is more extreme than in Figure 5c. Looking at the individual solos, we can identify four points of low complexity here. These are solos by Sonny Rollins, two of them played within the piece "Playin' in the Yard" and two within "The Everywhere Calypso" (red ellipses in Figure 5). Indeed, these solos are constructed of only a few pitch classes with clear tonal relationships. For "Playin' in the Yard", Rollins only uses a pentatonic scale for both solos whereas the solos in "The Everywhere Calypso" mainly consist of major scales and broken major triads (arpeggios). In the symbolic representation, these structures lead to a low complexity since there is no accompaniment. In the audio, the background instruments dampen this drop. Overall, we can observe several interesting structures that might be relevant for jazz history. These phenomena could be observed in a similar way on the basis of both symbolic and audio representations.

To test these observations in more detail, we now consider four different feature resolutions (see Section 4). Beyond the influence of the representation type, we want to test how signal processing technologies for suppressing background instruments affect the evolution curves. Figure 6 summarizes this experiment's results. In addition to the global complexity, we use chroma window lengths of 20 s, 1 s, and 400 ms. Looking at the vertical axes, we observe different absolute ranges. For the symbolic transcription (Figure 6a), the values of Γ for the global complexity (blue curve) lie in the interval $[0.84, 0.95]$. In contrast, the audio-based complexity curve (d) lies in the range $[0.93, 0.98]$. The enhanced audio versions are located between these extremes. HPRS-enhancement (c) leads to a curve with values in $[0.85, 0.97]$. Score-informed source separation (b) produces a global complexity curve ranging

[4] Removing silent frames before computing features suppresses this effect to some degree but, at the same time, produces artificial pitch combinations within local windows (phrases squeezed together).

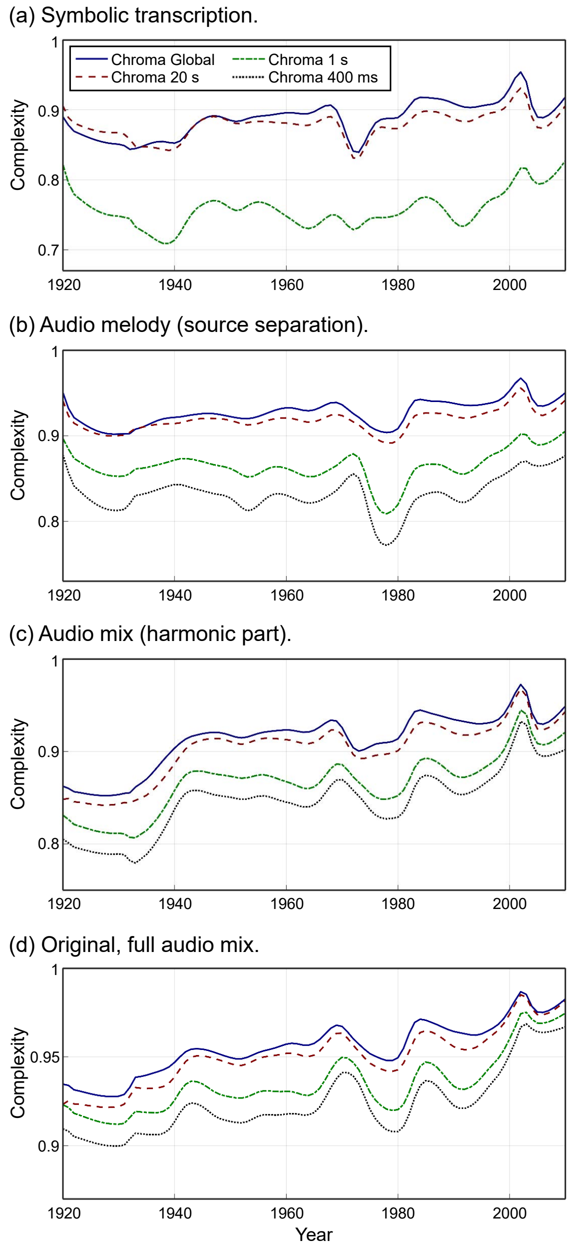

Figure 6. Evolution curve based on (a) symbolic transcription, (b) source-separated melody (score-informed), (c) harmonic part of audio (HPRS), (d) full audio mix.

in $[0.9, 0.96]$. Interestingly, these values are higher than in the HPRS-enhanced case (c). It seems that the percussive components or other artifacts remaining in the separated signal affect the complexity more than the harmonic parts of the background instruments do. For other window lengths, the behavior is similar. Only for the symbolic transcription (a), the smaller window lengths of 1 s and 400 ms (outside the plotting range) behave differently. Since the transcription of a monophonic solo exhibits only one nonzero pitch class at a time, this is no surprise—our complexity feature drops to zero then. With larger window lengths, we capture several pitch classes simultaneously leading to higher complexity.

Apart from the different ranges, we find only minor differences between the curves. As in Figure 5, the first years show higher complexity for the symbolic transcription (a) but also for the source-separated audio (b). As mentioned above, this is due to the long silence gaps between solo phrases. Considering the background instruments leads to a lower complexity and thus, stabilizes the analysis in some way. We also discover a special behavior at the year 1972. The symbolic-based curve (a) shows a sharp drop here stemming from Rollins' solos discussed above. This drop is weakened when using source separation (b) or the full mix (d) but it can still be observed in the HPRS-enhanced analysis (c). We conclude that not the background instrument but the percussive and residual components of the melody instrument (and possible overlap signals) eliminate this drop.

Beyond these rather subtle differences, the overall behavior is similar for all curves. In all settings, we observe a major increase around 1940 followed by a slightly increasing plateau between 1945 and 1967. Then, all curves drop, again reach a peak around 1983, and finally rise towards the 2000s. Even detailed structures are preserved throughout all representations such as the small drops around 1950 and 1965, or the curvature during the 1990s. Even for years with a low number of contributing solos where we have to take the results with care, the behavior is stable across representations. These observations show that corpus-level characteristics of the WJD appear in a widely coherent way over all of our experimental settings.

6. DISCUSSION

From our experiments, we conclude that meaningful corpus analyses can be performed on the basis of different music representations. Though our evolution curves for the WJD vary in their absolute range, general trends can be observed for all representations. Some audio-related artifacts in the analysis could be suppressed with standard signal processing tools such as harmonic–percussive separation. In contrast, using a high-quality score-informed technology for melody separation did not necessarily improve the results regarding audio-specific artifacts. It seems that timbral characteristics have a greater effect on the curves than the presence of background instruments. Quite the contrary, the presence of background instruments could even stabilize the analysis since it helps to suppress extreme complexity values when the solo instrument is silent. The high similarity between symbolic- and audio-based analyses lets us conclude that in a typical jazz scenario, the solo instrument is prominent enough in the full mix for analyzing some interesting solo characteristics directly from audio. This is an encouraging finding since audio-based studies can be scaled up to a large number of solos easily—in contrast to the time-consuming procedure needed for creating the WJD melody annotations. Since the deviations between our curves occurred in regions with low solo coverage, we suppose that in a large-scale corpus study, individual outliers are suppressed even better leading to more reliable results.

Acknowledgments: This work has been supported by the German Research Foundation (MU 2686/10-1, MU 2686/11-1, AB 675/2-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS. The authors want to thank the members of the Jazzomat research project led by Martin Pfleiderer for creating the WJD.

7. REFERENCES

[1] Jakob Abeßer, Klaus Frieler, Estefanía Cano, Martin Pfleiderer, and Wolf-Georg Zaddach. Score-informed analysis of tuning, intonation, pitch modulation, and dynamics in jazz solos. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(1):168–177, 2017.

[2] Héctor G. Bellmann. *Categorization of Tonal Music Style: A Quantitative Investigation*. PhD thesis, Griffith University, Brisbane, Australia, 2012.

[3] Juan J. Bosch, Rachel M. Bittner, Justin Salamon, and Emilia Gómez. A comparison of melody extraction methods based on source-filter modelling. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 571–577, New York City, USA, 2016.

[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*, 2014(23), 2014.

[5] Bruno Di Giorgi, Simon Dixon, Massimiliano Zanoni, and Augusto Sarti. A data-driven model of tonal chord sequence complexity. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(11):2237–2250, 2017.

[6] Christian Dittmar, Martin Pfleiderer, Stefan Balke, and Meinard Müller. A swingogram representation for tracking micro-rhythmic variation in jazz performances. *Journal of New Music Research*, 47(2):97–113, 2018.

[7] Jonathan Driedger, Meinard Müller, and Sascha Disch. Extending harmonic–percussive separation of audio signals. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 611–616, Taipei, Taiwan, 2014.

[8] Derry FitzGerald. Harmonic/percussive separation using median filtering. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, pages 246–253, Graz, Austria, 2010.

[9] Klaus Frieler, Martin Pfleiderer, Wolf-Georg Zaddach, and Jakob Abeßer. Midlevel analysis of monophonic jazz solos: A new approach to the study of improvisation. *Musicae Scientiae*, 20(2):143–162, 2016.

[10] Romain Hennequin, Bertrand David, and Roland Badeau. Score informed audio source separation using a parametric model of non-negative spectrogram. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 45–48, Prague, Czech Republic, 2011.

[11] Antoine Liutkus, Jean-Louis Durrieu, Laurent Daudet, and Gaël Richard. An overview of informed audio source separation. In *Proceedings of the International Workshop on Image and Audio Analysis for Multimedia Interactive Services (WIAMIS)*, pages 93–96, Paris, France, 2013.

[12] Matthias Mauch and Mark Levy. Structural change on multiple time scales as a correlate of musical complexity. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 489–494, Miami, Florida, USA, 2011.

[13] Matthias Mauch, Robert M. MacCallum, Mark Levy, and Armand M. Leroi. The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, 2(5), 2015.

[14] Maria Panteli, Emmanouil Benetos, and Simon Dixon. A review of manual and computational approaches for the study of world music corpora. *Journal of New Music Research*, 47(2):176–189, 2018.

[15] Martin Pfleiderer, Klaus Frieler, Jakob Abeßer, Wolf-Georg Zaddach, and Benjamin Burkhart. *Inside the Jazzomat*. Schott Campus, Mainz, Germany, 2017.

[16] Marcelo Rodríguez-López and Anja Volk. Symbolic segmentation: A corpus-based analysis of melodic phrases. In *Sound, Music, and Motion*, pages 548–557, Cham, 2014. Springer.

[17] Pablo H. Rodriguez Zivic, Favio Shifres, and Guillermo A. Cecchi. Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences*, 110(24):10034–10038, 2013.

[18] Xavier Serra. Creating research corpora for the computational study of music: The case of the CompMusic project. In *Proceedings of the AES International Conference on Semantic Audio*, London, UK, 2014.

[19] Ajay Srinivasamurthy, Andre Holzapfel, Kaustuv Kanti Ganguli, and Xavier Serra. Aspects of tempo and rhythmic elaboration in hindustani music: A corpus study. *Frontiers in Digital Humanities*, 4(20), 2017.

[20] Sebastian Streich. *Music Complexity a Multi-Faceted Description of Audio Content*. PhD thesis, University Pompeu Fabra, Barcelona, Spain, 2007.

[21] Christof Weiß, Matthias Mauch, Simon Dixon, and Meinard Müller. Investigating style evolution of Western classical music: A computational approach. *Musicae Scientiae*, 2018.

[22] Christof Weiß and Meinard Müller. Quantifying and visualizing tonal complexity. In *Proceedings of the Conference on Interdisciplinary Musicology (CIM)*, pages 184–187, Berlin, Germany, 2014.

[23] Christof Weiß and Meinard Müller. Tonal complexity features for style classification of classical music. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 688–692, Brisbane, Australia, 2015.

[24] Christopher Wm. White. *Some Statistical Properties of Tonality, 1650-1900*. PhD thesis, Yale University, New Haven, Connecticut, USA, 2013.