

AUTOMATED ESTIMATION OF RIDE CYMBAL SWING RATIOS IN JAZZ RECORDINGS

Christian Dittmar
International Audio
Laboratories Erlangen
christian.dittmar
@audiolabs-erlangen.de

Martin Pfeleiderer
The Liszt University of Music Weimar
martin.pfleiderer
@hfm-weimar.de

Meinard Müller
International Audio
Laboratories Erlangen
meinard.mueller
@audiolabs-erlangen.de

ABSTRACT

In this paper, we propose a new method suitable for the automatic analysis of microtiming played by drummers in jazz recordings. Specifically, we aim to estimate the drummers' swing ratio in excerpts of jazz recordings taken from the Weimar Jazz Database. A first approach is based on automatic detection of ride cymbal (RC) onsets and evaluation of relative time intervals between them. However, small errors in the onset detection propagate considerably into the swing ratio estimates. As our main technical contribution, we propose to use the log-lag autocorrelation function (LLACF) as a mid-level representation for estimating swing ratios, circumventing the error-prone detection of RC onsets. In our experiments, the LLACF-based swing ratio estimates prove to be more reliable than the ones based on RC onset detection. Therefore, the LLACF seems to be the method of choice to process large amounts of jazz recordings. Finally, we indicate some implications of our method for microtiming studies in jazz research.

1 Introduction

Jazz drummers usually keep time by using the ride cymbal (RC) and hi-hat (HH), especially in styles with so-called "swing feel" [2]. They commonly emphasize the "backbeat," i.e., the metric-harmonically unaccented beat, on the HH while playing typical patterns on the RC. According to [21, p. 248], this supports the "light" character of jazz rhythm. Instead of playing the beat in a steady manner, variations and additional "offbeat" strokes are usually added on the RC as well as on other drum parts. These variations differ from drummer to drummer and from performance to performance [2, pp. 617-629].

The most common time-keeping pattern played on the RC is shown in Figure 1. In addition to conventional drum notation in the top row, we show a corresponding time-domain signal at 240 BPM with overlaid amplitude envelope (bold black curve) and the so-called novelty curve

(thin black curve). We color-code the relevant beats and subdivisions thereof as follows. The sequence starts with the so-called "downbeat" quarter note (light blue), followed by the backbeat eighth note (light green), and the offbeat eighth note (light red) before starting over again with the downbeat. We will refer to this prototype sequence of onsets as RC pattern.

The so-called swing ratio expresses the beat subdivision and relates to the phrasing of the eighth notes in the RC pattern. Swinging eighth notes are typically played in different ratios, ranging continuously from straight eighths (1 : 1), over triplet eighths (2 : 1), to dotted eighths (3 : 1), or more extreme ratios. The swing ratio is reported to be tempo dependent [4, 9, 15], cf. Section 2.1. In Figure 1, the color-coded tone durations show how the backbeat duration grows with increasing swing factor, while the complementing offbeat duration shrinks. In Figure 1(a), backbeat and offbeat have equal duration, corresponding to straight eighths as given in the drum notation. In Figure 1(b), the RC pattern is notated as tied-triplets. In Figure 1(c), the backbeat duration equals a dotted eighth. Consequently, the offbeat duration equals that of a sixteenth note as shown in the drum notation.

There are several case studies concerning the swing ratio in jazz (cf. [21, pp. 262-273], and Section 2.1). While most of the studies examine swing ratios of soloists, it is widely acknowledged that the swing ratio of the RC pattern crucially contributes to the "swinging" character of the music. Most of the studies are based on manual transcription of onsets, often by visual inspection of the amplitude envelope of jazz excerpts. Few studies specifically examine the RC pattern [15] and its interaction with the soloist's timing [9]. This inspired us to develop and to evaluate methods for automated swing ratio estimation from RC patterns in jazz recordings. For sure, an automated generation of large amounts of reliable swing ratio data is essential for meaningful and more differentiated research on microtiming in jazz. Besides onset-based swing ratio estimation, our main approach is a log-lag variant of a local autocorrelation function (ACF) applied to onset-related novelty functions (see Sections 3.3). We refer to this representation as log-lag ACF (LLACF) and show its applicability to swing ratio estimation in Section 3.4.



© Christian Dittmar, Martin Pfeleiderer, Meinard Müller.
Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Christian Dittmar, Martin Pfeleiderer, Meinard Müller. "Automated Estimation of Ride Cymbal Swing Ratios in Jazz Recordings", 16th International Society for Music Information Retrieval Conference, 2015.

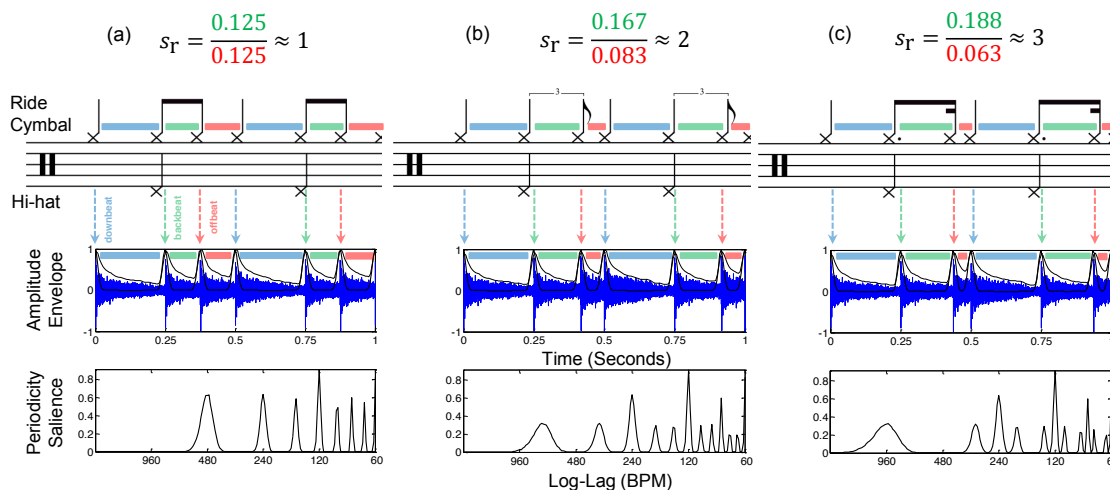


Figure 1. Illustration of prototypical RC patterns as drum notation (top), time-domain signal (mid), and LLACF (bottom). **(a):** Swing factor of $s_r = 1$ corresponding to straight eighth notes. **(b):** Swing factor of $s_r = 2$ corresponding to the idealized “tied-triplet feel”. **(c):** Swing factor $s_r = 3$, where the duration of the backbeat equals a dotted eighth note.

2 Related Work

A number of papers are concerned with systematic studies on swing ratio in jazz music. Since most of the studies use comparably small data sets and manual annotation, we think that swing ratio estimation is a suitable task to apply automatic methods from Music Information Retrieval (MIR) research in order to enable analysis of larger music data sets.

2.1 Jazz Microtiming Analysis

An early attempt to analyze swing ratios in jazz recordings is described in [17]. The author relies on visual inspection of spectrograms but does not report quantitative results. In [22], the swing ratios in the analyzed jazz recordings are reported to range from 1.48 to 1.82. Rose [23] reports an average swing ratio of 2.38 measured from amplitude envelopes. In [7], an average swing ratio of 1.75 is measured using a MIDI wind controller played by saxophonists. In [19], the analysis focuses on the RC and swing factors between 1.0 and 3.3 are reported without detailing the measurement method. In [6], an average swing ratio of 1.6 is measured using amplitude envelopes. Friberg and Sundström [9] annotated RC onsets in spectrograms of jazz excerpts. They report trends indicating a high negative correlation between the tempo and the swing ratio which seems to be valid across different drummers. In [3], an average swing ratio of 2.45 is measured in the performances of pianists playing a MIDI piano. In [1], comparably low swing ratios in the range between 0.9 to 1.7 are measured from amplitude envelopes. Honing and de Haas [15] conducted experiments with professional jazz drummers performing on a MIDI drum kit. Besides further evidence for the tempo dependency of swing ratios, the results show that jazz drummers have enormous control over their timing.

2.2 Rhythmic Mid-Level Features

Motivated by the need to design specialized mid-level features for music similarity estimation, several authors proposed conceptually similar, tempo-independent representations of rhythmic patterns. The basic observation is, that rhythmic patterns that are perceived as similar by human listeners may not be judged as similar by automatic methods. One of the main reasons is that the patterns are typically played in different tempi, which makes them unsuited for direct comparison. Therefore, Peeters [20] used tempo normalized spectral rhythm patterns to automatically classify ballroom dance styles. Holzapfel and Stylianou [13, 14] proposed to apply the scale transform to periodicity spectra to enable the use of conventional distance measures between rhythmic patterns despite tempo differences. Around the same time, the LLACF was proposed in [12] as well as the tempo-insensitive representation used for classification of ballroom dances in [16]. The LLACF was reported to be favorable over the scale transform for classification of Latin American rhythm patterns in [24]. The tempogram as described in [11] is based on similar ideas and additionally features a cyclic post-processing to remedy the problem of octave ambiguity. Marchand and Peeters [18] revisited the scale transform and applied it to modulation spectra as tempo-independent feature, again for classification of ballroom dances. Eppler et al. [8] used peak ratios in the LLACF as features for detecting the swing feel but did not explicitly try to estimate swing ratios.

3 Method

In this section, we describe our approaches to automatic swing ratio estimation from excerpts of jazz recordings with swing feel. The first variant relies on peak-picking in an onset-related novelty curve (Section 3.1). The sec-

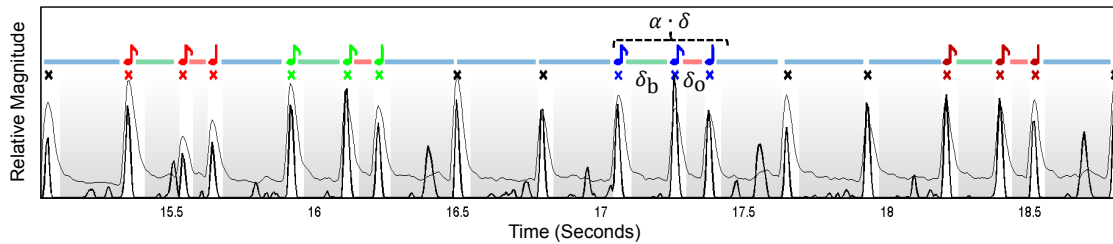


Figure 2. A four seconds excerpt from the 1979 recording of “Anthropology”, performed by Art Pepper playing solo clarinet, with Charlie Haden on bass and Billy Higgins on drums. The bold black curve depicts the novelty function Δ , the thin black curve shows the RC related threshold H . Automatically detected RC onsets are marked by the bold black crosses, colored crosses represent the four onset triples accepted for swing ratio estimation. The note durations are color-coded in the same way as in Figure 1.

ond approach relies on computation of the LLACF from the novelty curve (Section 3.3) and comparison to prototype LLACFs. As will be explained in Section 4.1, we have a rough tempo estimate $\tau_e \in \mathbb{R}_{>0}$ available for each jazz excerpt. Let $\delta_b, \delta_o \in \mathbb{R}_{>0}$ be the tone duration of the backbeat and the offbeat in an RC pattern as shown in Figure 1. They relate to the tempo by $\tau_e \approx (\delta_b + \delta_o)^{-1} \approx \delta^{-1}$, with the beat (quarter note) duration $\delta \in \mathbb{R}_{>0}$. The targeted swing ratio is given by:

$$s_r = \frac{\delta_b}{\delta_o} \quad (1)$$

Consequently, $\delta_b = \delta \cdot s_r \cdot (1 + s_r)^{-1}$ yields the tone duration of the backbeat and $\delta_o = \delta \cdot (1 + s_r)^{-1}$ yields the tone duration of the offbeat.

3.1 Ride Cymbal Onset Detection

With regard to Eqn (1), we aim to measure δ_b and δ_o from the jazz excerpts under analysis. One possibility is to search for RC onsets and use the time differences between consecutive onsets as estimate for note durations. To this end, we compute a time-frequency (TF) representation of an excerpt using the short-time Fourier transform (STFT) with blocksize w and hopsize r given in seconds. Let $\mathcal{X}(m, k)$ with $m \in [1 : M]$, $k \in [0 : K]$ be a complex-valued STFT coefficient at the m^{th} time frame and k^{th} spectral bin. Here, the interval $[1 : M]$ represents the time axis and K corresponds to the Nyquist frequency. Following the approaches in [10, 11], we compute a novelty curve $\Delta : [1 : M] \rightarrow \mathbb{R}$ as follows. First, we derive the logarithmically compressed magnitude spectrogram $\mathcal{Y}(m, k) := \log(1 + \gamma \cdot |\mathcal{X}(m, k)|)$ for a suitable constant $\gamma \geq 1$. Then, the novelty function is given as

$$\Delta(m) := \sum_{k=0}^K |\mathcal{Y}(m+1, k) - \mathcal{Y}(m, k)|_{\geq 0}, \quad (2)$$

where $|\cdot|_{\geq 0}$ denotes half-wave rectification. The resulting Δ exhibits salient peaks at frames corresponding to tone onsets. Inevitably, spurious peaks may occur in Δ that could be mistaken for RC onsets. Thus, we derive an RC

related threshold function as

$$H(m) := \sum_{k=k_0}^K |\mathcal{X}(m, k)|, \quad (3)$$

where the bin k_0 corresponds to the lower cutoff frequency. Figure 2 shows an example of Δ as bold black curve and the corresponding H as thin black curve. For the sake of visibility, both curves are normalized to unit maximum in the plot. We take the average value of H as threshold criterion and only accept peaks from Δ in frames where H exceeds this value (indicated by the white background). The $N = 18$ local maxima accepted as RC onsets are marked by bold crosses. Multiplication of the corresponding frame indices with the hopsize r yields a set of strictly monotonically increasing onset times $B = \{b_1, b_2, \dots, b_N\}$ for onset-based swing ratio estimation.

3.2 Onset-Based Swing Ratio Estimation

Once we obtained a sequence B of RC onsets, we estimate s_r in a tempo-informed manner. Assuming a roughly constant tempo τ_e throughout the excerpt, the time interval $\delta = \tau_e^{-1}$ between two consecutive beats should be close to $\delta_b + \delta_o$. To account for small deviations from the ideal beat period δ , we introduce a tolerance $\alpha \geq 1$. Now, we go through every previously detected RC onset and test the hypothesis that it could be the first in a series of three consecutive onsets (backbeat, offbeat, downbeat). We denote this sub-sequence as $B_n = \{b_n, b_{n+1}, b_{n+2}\}$, $B_n \subset B$ and refer to it as onset triple. From all possible triples $B_n, n \in [1 : N - 2]$ we accept the ones that fulfill the criterion

$$(b_{n+2} - b_n) < \alpha \cdot \delta \quad (4)$$

as instances of triples embedded in an RC pattern. The swing ratio is estimated from a valid onset triple by setting $\delta_b = b_{n+1} - b_n$ and $\delta_o = b_{n+2} - b_{n+1}$ in Eqn (1). In Figure 2, we illustrate this procedure. All RC onset candidates are marked by black crosses but only the triples that fulfill the constraint in Eqn (4) are marked with different colors. Above the third triple (blue note symbols) we depict the extent of the search range $\alpha \cdot \delta$ that covers both δ_b and δ_o . As indicated in the plot, we try to find multiple

occurrences of the RC pattern triples per excerpt, so we can obtain a more robust estimate for the swing ratio by averaging over the individual s_r -values computed for each triple. For that reason, we also accept variations of the RC pattern where the offbeat impulse occurs in succession to the downbeat instead of the backbeat. As will be explained in Section 4.4, there are situations where estimation of s_r from RC onsets may deliver erroneous results. To obtain more robust estimates, we introduce LLACF-based swing ratio estimation in the next two sections.

3.3 LLACF Mid-Level Representation

We propose to employ the LLACF as a tempo-normalized mid-level representation capturing the swing ratio that is implicitly encoded in the peaks of Δ . Using the LLACF, we can circumvent the selection of onset candidates and instead transform the complete Δ into a phase-invariant, tempo-normalized representation. Swing ratio estimation then boils down to matching this representation to LLACFs with known swing ratios (see Section 3.4). To this end, we first compute a normalized ACF from the novelty function Δ as:

$$R_{\Delta\Delta}(\ell) = \frac{\sum_{m=1}^{M-\ell} \Delta(m)\Delta(m-\ell)}{\sum_{m=1}^M \Delta(m)^2}, \quad (5)$$

where we only consider the positive lags $\ell \in [0 : M - 1]$. Note that $R_{\Delta\Delta}(\ell) = R_{\Delta\Delta}(-\ell)$ due to symmetry. Moreover, $R_{\Delta\Delta}(0) = 1$ and $R_{\Delta\Delta}(\ell) < 1$ for $\ell \in [1 : M - 1]$. Each lag can be expressed as tempo value by the relation $\tau = \frac{60}{r \cdot \ell}$. We now define a logarithmically spaced tempo (log-tempo) axis, that has equal distance q between tempo octaves and has the reference tempo τ_r at a defined position. After correction for the ratio between the excerpt's tempo estimate τ_e and the reference tempo τ_r , we use linear interpolation to warp $R_{\Delta\Delta}$ onto this axis, yielding our tempo-normalized LLACF \mathcal{A} . Despite using a log-tempo axis, we stick to the term log-lag ACF since the inverse relation $\ell = \frac{60}{r \cdot \tau}$ retains the logarithmic spacing, just in opposite direction.

In the bottom row of Figure 1, we show the LLACFs corresponding to the prototypical RC patterns. Variation of s_r gives an intuition how the salience of different periodicities in the RC pattern is represented by the LLACF. Since τ_r is constant, all three LLACFs have clear peaks at the beat periodicity (240 BPM) and its integer subdivisions. For $s_r = 1$ in Figure 1(a), there is a strong peak at 480 BPM (corresponding to the straight eighth notes). With increasing swing ratio, this peak diverges into two lobes that move to other periodicities. In Figure 1(c), the first peak resides at 960 BPM (offbeat equals a sixteenth note) and the second peak is at 320 BPM (backbeat equals a dotted eighth note).

3.4 LLACF-Based Swing Ratio Estimation

In order to estimate a swing ratio from the shape of \mathcal{A} , we construct a set $\mathcal{A}_{s_r}, s_r \in \mathbb{R}$ with $1 \leq s_r \leq 4$ of prototype

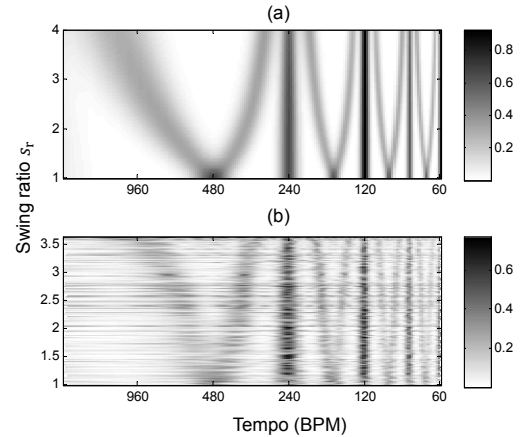


Figure 3. Evolution of the LLACF computed from RC patterns with increasing swing ratio. **(a):** LLACFs derived from novelty functions of idealized prototype RC patterns at a reference tempo τ_r of 240 BPM. **(b):** LLACFs extracted from our test corpus that have been warped to match τ_r .

LLACFs. They are extracted from novelty functions of idealized RC patterns with fixed reference tempo τ_r and varying swing ratio s_r (cf. the time-domain plots in Figure 1). In Figure 3(a) we show the complete set of prototype LLACFs with the log-tempo axis in BPM and the swing ratio increasing from bottom to top. Darker shade of gray corresponds to higher periodicity salience. One can clearly see how the offbeat-related peaks change their periodicity with the swing ratio while the peaks related to the beat (and subdivisions thereof) reside at the same periodicity.

Now, our approach to swing ratio estimation is to compare the extracted \mathcal{A} to each of these prototype LLACFs and to select the swing ratio corresponding to the best match. For the comparison, we employ Pearson's correlation coefficient. We have to take into account that the tempo estimate τ_e used for warping the LLACF to the reference log-tempo axis underlying \mathcal{A}_{s_r} may be slightly inaccurate. As a consequence, the resulting \mathcal{A} might exhibit a constant offset with respect to the prototype \mathcal{A}_{s_r} . Thus, we shift the \mathcal{A} against the log-tempo axis of each \mathcal{A}_{s_r} in a restricted interval $[-q \cdot \log_2(\alpha) : +q \cdot \log_2(\alpha)]$ to find the best alignment. Finally, the s_r corresponding the maximum correlation coefficient over all entries in \mathcal{A}_{s_r} is selected.

4 Evaluation

In this section, we describe the setup, metrics, and results of the experiments we conducted in order to compare manual, onset-based, and LLACF-based swing ratio estimation. In addition, some trends visible in the data are discussed.

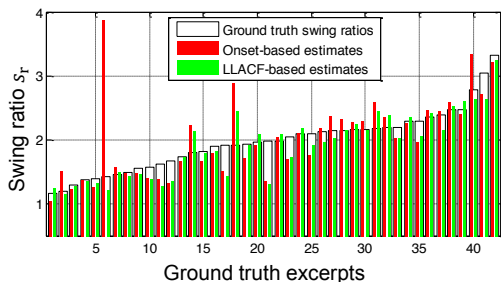


Figure 4. Comparison of the swing ratios estimated from ground truth RC onsets, automatically detected RC onsets and LLACF analysis.

4.1 The Weimar Jazz Database

The Weimar Jazz Database¹ consists of 299 (as in July 2015) transcriptions of instrumental solos in jazz recordings performed by a wide range of renowned jazz musicians. The solos have been manually annotated by musicology and jazz students at Liszt School of Music Weimar as part of the Jazzomat Research Project.² Several music properties are annotated, most notably the pitch, onset and offset of all tones played by the soloists, as well as a manually tapped beat grid, chords, form parts, phrase boundaries, and articulation. For our work, we only use the beat grid. From the complete Weimar Jazz Database, we automatically selected a subset of 921 excerpts that had been labeled with swing feel. Because we will compare the swing ratios of drummers and soloists in our future work, the excerpts had to contain at least 5 consecutive eighth notes played by the soloists. The total playtime of the selected excerpts amounts to roughly 50 minutes (out of 8 hours), their average duration is 3.3 seconds.

4.2 Evaluation Setting

A subset of 42 excerpts have been manually annotated for RC onsets in order to create a ground truth for swing ratio estimation. The reference onsets were transcribed by two experienced student assistants of the Jazzomat Research Project using the software Sonic Visualiser [5]. The ground truth subset was split in two, approximately equal parts and each part was given to one of the annotators. In total, 834 RC onsets were manually annotated. In our evaluation (cf. Sections 4.3, 4.4, and 4.5), we used the well-known metrics recall, precision and F-measure for quantitative evaluation. In order to count an onset candidate as true positive, we allowed a maximum deviation of ± 30 ms to the ground truth onset time. Furthermore, we used Pearson’s correlation coefficient as a means to quantify the agreement between reference swing ratios and automatically estimated swing ratios. We fixed the following extraction parameters for the automatic estimation of swing

ratios: The STFT blocksize w was appr. 46 ms and the hop-size r was appr. 5.8 ms. The compression-constant γ was 1000, the lower cutoff k_0 was set to equal appr. 12.9 kHz, the reference tempo τ_r was 240 BPM, the LLACF octave-resolution q was 36. The tolerance α for tempo deviations was 1.2.

4.3 Cross-Validation

At first, we are interested in the agreement between our human annotators, since we suspect that there may be ambiguous cases where it is not clear where an RC onset is exactly located in time or if there is an onset at all. Thus, we selected a small subset of 11 excerpts for which the annotators created a cross-validation transcription. Running these against the larger set, we receive an F-measure of appr. 0.96. The average absolute time difference between matched onsets in the reference and the cross-validation set amounts to 7.8 ms.

4.4 Onset-Based Evaluation

Next, we used the previously validated ground truth annotations as reference to assess the performance of our automated RC onset detection described in Section 3.2. In this scenario, we received an F-measure of appr. 0.93 and an average onset deviation of 2.5 ms. Since these results seem surprisingly good, we wanted to quantify how much potential onset detection errors would propagate into the swing ratio estimation. Using the procedure described in Section 3.2, we determined ground truth swing ratios for all manually annotated excerpts. When we compared these to the swing ratios estimated from automatically detected RC onsets, we yielded a correlation coefficient of appr. 0.66 (see Figure 4). With regard to the comparably high F-measure obtained for the onset detection, this unsatisfactory result may seem surprising at first, but can be explained using the example in Figure 2. There, we see that only 12 out of 18 RC onsets are considered for swing ratio estimation. Intuitively, small deviations in the detected onset times can lead to under- or overestimation of the swing ratio, especially for fast tempi, where subtle timing differences may get lost due to the coarse sampling of the analysis frames. Even worse errors may be caused by spurious onsets that fulfill the threshold criterion but are actually not RC patterns. This is the case for the sixth excerpt in Figure 4, where some sort of RC swell is mistaken for an onset triple, leading to an overestimation of s_T .

4.5 LLACF-Based Evaluation

Since we found the correlation between ground truth swing ratios and onset-based swing ratios to be unsatisfactory, we repeated the comparison with respect to swing ratios estimated from the LLACF as described in Section 3.3. This time, we received a correlation coefficient of appr. 0.9. In Figure 4, one can see that both methods behave similar

¹ <http://jazzomat.hfm-weimar.de/dbformat/dboverview.html>

² <http://jazzomat.hfm-weimar.de/>

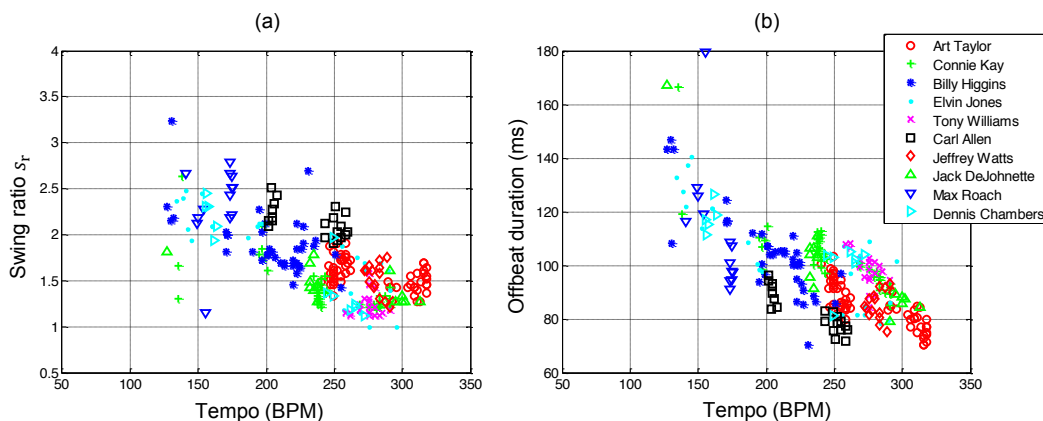


Figure 5. Scatter plots showing the relationship of tempo vs. **(a):** swing ratio and **(b):** offbeat duration. Each marker corresponds to one jazz excerpt. We only show the 10 most frequently represented drummers.

but the onset-based swing ratios exhibit some pronounced outliers. Moreover, Figure 3 shows that the prototypical LLACFs in \mathcal{A}_{s_r} correspond quite well to the LLACFs extracted from our test corpus. Both plots depict the LLACFs ordered by the corresponding swing ratio. The typical structure of periodicity peaks is clearly visible, although the LLACFs extracted from the jazz excerpts are much more noisy than the idealized LLACFs. This leads us to the conclusion that the LLACF-based swing ratio estimation is a reliable method that should be preferred over the onset-based swing ratio estimation.

4.6 Comparison to Friberg and Sundström

In Section 1, we already indicated our aim to re-examine the findings of Friberg and Sundström [9] on a larger scale. As can be seen in Figure 5(a), our automatically estimated swing ratios show similar trends as the manually annotated data used in the original paper. However, while Friberg and Sundström only had around 40 excerpts from various pieces of four drummers, we are able to study several hundreds of RC patterns played by a wide range of drummers due to our automated method (three among them—Tony Williams, Jack DeJohnette, and Jeffrey Watts—were examined by Friberg and Sundström, too).

In Figure 5, we show the results obtained for the 10 drummers represented with the most excerpts. Each point in the scatterplots is placed according to (a) s_r vs. τ_e and (b) δ_o vs. τ_e . In general, the negative correlation of swing ratio and tempo is clearly discernable—for the whole data set as well as for certain drummers like Elvin Jones or Billy Higgins, who vary their swing ratio from appr. 2.5 around 150 BPM to appr. 1.5 at 250 BPM, and in the case of Jones even to around 1.0 at 300 BPM. However, there are also drummers who seems to keep almost the same swing ratio at different tempi, e.g., Art Taylor or Carl Allen.

Additionally, Friberg and Sundström report the duration between the offbeat impulse and the next beat to be roughly constant at 100 ms for all tempi faster than 150 BPM (cf. [9, p. 337]). In general, this finding is supported by

our data (see Figure 5(b)), but the offbeat durations have a wider range from 110 ms to 80 ms and even 70 ms.

5 Conclusions and Future Work

In this paper, we presented a microtiming study conducted on a subset of the publicly available Weimar Jazz Database. Future work will be directed towards extending our method to more drummers and other recordings as well as to the comparison between RC patterns and soloists. Exact onset times of all tones of the soloists, and thus their microtiming and swing ratio, are at hand within the Weimar Jazz Database. A comparison between drummers' and soloists' microtiming will allow for a larger scale re-examination of one of the central findings in [9]: The swing ratio of soloists is in general lower than the swing ratio of the accompanying drummer since soloists deliberately play behind the beat while synchronizing the offbeat with the drummer. They do so, because, as Friberg and Sundström claim, “delayed downbeats and synchronized offbeats may create both the impression of the laid-back soloist, which is often strived for in jazz, and at the same time an impression of good synchronization” [9, p. 345]. Therefore, using microtiming data from the Weimar Jazz Database as well as automatically estimated swing ratios of RC patterns may lead to new insights in the interactive art of improvising together in a professional jazz ensemble.

6 Acknowledgments

The Jazzomat Research Project is supported by the German Research Foundation (Melodisch-rhythmische Gestaltung von Jazzimprovisationen. Rechnerbasierte Musikanalyse einstimmiger Jazzsoli, DFG-PF 669/7-1). The authors would like to thank all student assistants participating in the transcription and annotation process. The International Audio Laboratories Erlangen (AudioLabs) is a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer IIS.

7 References

- [1] Fernando Benadon. Slicing the beat: Jazz eighth-notes as expressive microrhythm. *Ethnomusicology*, pages 73–98, 2006.
- [2] Paul F. Berliner. *Thinking in Jazz. The Infinite Art of Improvisation*. University of Chicago Press, 1994.
- [3] Walter Gerard Busse. Toward objective measurement and evaluation of jazz piano performance via midi-based groove quantize templates. *Music Perception*, 19(3):443–461, 2002.
- [4] Matthew W. Butterfield. Why do jazz musicians swing their eighth notes? *Music Theory Spectrum*, 33(1):3–26, 2011.
- [5] Chris Cannam, Christian Landone, and Mark B. Sandler. Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proc. of the International Conference on Multimedia*, pages 1467–1468, Florence, Italy, 2010.
- [6] Geoffrey L. Collier and James Lincoln Collier. A study of timing in two louis armstrong solos. *Music Perception*, 19(3):463–483, 2002.
- [7] Mark C. Ellis. An analysis of ‘swing’ subdivision and asynchronization in three jazz saxophonists. *Perceptual and Motor Skills*, 73(3):707–713, 1991.
- [8] Arndt Eppler, Andreas Männchen, Jakob Abeßer, Christof Weiß, and Klaus Frieler. Automatic style classification of jazz records with respect to rhythm, tempo, and tonality. In *Proc. of the Conference on Interdisciplinary Musicology (CIM)*, December 2014.
- [9] Anders Friberg and Andreas Sundström. Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, 19(3):333–349, 2002.
- [10] Peter Grosche and Meinard Müller. Extracting predominant local pulse information from music recordings. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(6):1688–1701, 2011.
- [11] Peter Grosche, Meinard Müller, and Frank Kurth. Cyclic tempogram – a mid-level tempo representation for music signals. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5522–5525, Dallas, Texas, USA, March 2010.
- [12] Matthias Grühne and Christian Dittmar. Improving Rhythmic Pattern Features Based on Logarithmic Pre-processing. In *Proc. of the Audio Engineering Society Convention (AES)*, Munich, Germany, May 2009. Preprint 7817.
- [13] André Holzapfel and Yannis Stylianou. A scale transform based method for rhythmic similarity of music. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 317–320, April 2009.
- [14] André Holzapfel and Yannis Stylianou. Scale transform in rhythmic similarity of music. *IEEE Transactions on Audio, Speech & Language Processing*, 19(1):176–185, 2011.
- [15] Henkjan Honing and W. Bas de Haas. Swing once more: Relating timing and tempo in expert jazz drumming. *Music Perception: An Interdisciplinary Journal*, 25(5):471–476, 2008.
- [16] Jesper Højvang Jensen, Mads Græsbøll Christensen, and Søren Holdt Jensen. A tempo-insensitive representation of rhythmic patterns. In *Proc. of the European Signal Processing Conference (EUSIPCO)*, pages 1509–1512, Glasgow, Scotland, August 2009.
- [17] Franz Kerschbaumer. *Miles Davis: Stilkritische Untersuchungen zur musikalischen Entwicklung seines Personalstils*. Studies in jazz research. Akademische Druck und Verlagsanstalt, 1978.
- [18] Ugo Marchand and Geoffroy Peeters. The modulation scale spectrum and its application to rhythm-content description. In *Proc. of the International Conference on Digital Audio Effects (DAFx)*, pages 167–172, Erlangen, Germany, September 2014.
- [19] Will Parsons and Ernest Cholakis. It dont mean a thing if it aint dang, dang-a dang! *Downbeat*, 52(8):61, 1995.
- [20] Geoffroy Peeters. Rhythm classification using spectral rhythm patterns. In *Proc. of the International Conference on Music Information Retrieval (ISMIR)*, pages 644–647, London, UK, September 2005.
- [21] Martin Pfeleiderer. *Rhythmus: Psychologische, theoretische und stilanalytische Aspekte populärer Musik*. Transcript, 2006.
- [22] Peter Reinholdsson. Approaching jazz performances empirically. some reflections on methods and problems. *Action and perception in rhythm and music*, 55:105–125, 1987.
- [23] Richard Franklin Rose. *An Analysis of Timing in Jazz Rhythm Section Performances*. PhD thesis, University of Texas, 1989.
- [24] Thomas Völkel, Jakob Abeßer, Christian Dittmar, and Holger Großmann. Automatic genre classification on latin music using characteristic rhythmic patterns. In *Proc. of the Audio Mostly: A Conference on Interaction with Sound*, pages 16:1–16:7, Piteå, Sweden, September 2010.