# EXPLORING PHRASE FORM STRUCTURES.
# PART II: MONOPHONIC JAZZ SOLOS.

**Klaus Frieler, Wolf-Georg Zaddach**
Liszt School of Music Weimar
{klaus.frieler, wolf-georg.zaddach@hfm-weimar.de

**Jakob Abeßer**
Fraunhofer IDMT Ilmenau
{jakob.abesser@idmt.fraunhofer.de

## ABSTRACT

In this explorative study, we investigate the phrasal structure of a set of 100 monophonic jazz solo taken from the WEIMAR JAZZ DATABASE. The main purpose was to see whether phrase form structure might lead to useful features for computational jazz solo analysis. To this end, we extracted basic statistical descriptors for phrases such as the number of notes, event density, total duration etc. Furthermore, we analysed the self-similarity of phrase sequences with regard to semitone intervals and duration classes and in combination. Phrase form structure can be characterised by coherence values and runlengths. As expected, form coherence values are generally very low with duration-based form coherence being higher than interval-based or combined form structure. John Coltrane was found to be an exceptional case with very high duration-based coherences. Furthermore, a global tendency to increase in event density (i. e., intensity) at the beginning of solos was observed.

## 1. INTRODUCTION

In the second part of our explorations into phrase form structure (PFS), we examine monophonic jazz solos. Compared to folk songs, these solos represent quite a different set of data: solos are improvised not composed (or fixed by tradition), they are much longer, do not contain lyrics, and employ much more advanced rhythmic and tonal devices. Nevertheless, they are also structured in phrases, on a generative as well as on a perceptional level. Moreover, jazz musicians and their solos are also rooted in Western music traditions, which might be reflected in the PFS. From the outset, we do not except very large inner coherence in the PFS, since jazz solos are often thought to be structured like ad-lib speeches (Johnson-Laird, 2002). But motivic improvisation might lead to phrase coherence. Moreover, one could suppose that form coherence is used as a dramaturgical device. For example, a series of phrases with high similarity in the rhythmic or tonal domain can be used by the soloist to increase intensity or just to convey coherence. Likewise, phrase lengths and event densities can be employed to increase or lower the "power" level. The present studies sets out to explore such relationships, to check some of the stated hypotheses and to search for useful features.

To our knowledge, this is the first study which explores phrase form structure in jazz solos using a large set of data. Jazz researchers examined phrase structures in the past (e. g., Downs, 2001; Love, 2012), but rarely with regard to phrase similarities, and usually not using statistical methods but focusing on selected examples.

Form in music is generally a hierarchical and multi-layered phenomenon. Phrase form structure as considered here is not to be confused with the form of the underlying composition. Those structural levels might coincide, but often do not (Love, 2012). In the following, "form" is nearly always to be understood as "phrase form structure".

## 2. DATA

The analysis was carried with help of the WEIMAR JAZZ DATABASE included in the MeloSpySuite software toolkit [1] (Frieler et al., 2013), which at the time of the study (February 2014) contained 106 annotated monophonic jazz solos covering a wide range of soloist and styles. For technical reasons, only a subset of 100 solos from 38 soloist with a total of 368 choruses, 2643 phrases and 42015 tones was included in the study. Phrase annotation for jazz solos suffers from the same issues as folk songs, even though in the case of wind instruments phrase boundaries often coincide with breathing rests. The phrase annotation for this study was done by the transcribers of the solos which are musicology and jazz students. The data are in a pre-final status, which means they were cross-checked by an independent transcriber for basic correctness, which, however, did not include a revisal of phrase boundaries.

## 3. METHOD

To extract the form information, similarity values between each phrase of a song were calculated using edit distance (Levenshtein, 1965) on a interval-based or duration-class-based representation of the melody. Form strings (such as AAAA, ABCD etc.) were extracted from the resulting self-similarity matrices using fixed numerical thresholds. Two phrases $p_i$, $p_j$ were deemed similar if $\sigma_I(p_i, p_j) \geq 0.6$ for intervals and $\sigma_D(p_i, p_j) \geq 0.7$ for duration classes, where $\sigma(p_i, p_j)$ is the normed edit similarity taking values $\in [0, 1]$ (Müllensiefen & Frieler, 2004). For a more in-depth discussion on the choice of these values see the accompanying paper Frieler (2014), this volume. Similar phrases were denoted using the same symbol, i. e., no distinction between identical and similar phrases was made for sake of simplictity. The analysis was carried out on the level of whole solos and of single choruses. Additionally, number of notes, total duration (in seconds and in bar units), and event densities (per seconds or per bar)

---

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

were calculated for each phrase. Duration of a phrase is defined here as the time-interval between the onset of the first and the offset of the last tone. Bar units are based on a fractional representation of metrical positions, where each measure equals one numerical unit.

This analysis was carried out using the `melfeature` commandline tool from the MeloSpySuite (Frieler et al., 2013). The resulting data were imported into R (R Core Team, 2013) for further analysis.

## 4. RESULTS

A majority of 75% of the solos comprise 4 or less choruses, the median is 2 and the mean is 3.5 choruses. 30 solos consist of only 1 chorus and 28 of 2 choruses. A small peak can be found at 8 choruses (6 instances), possibly due to a previous arrangement of the musicians. Descriptive statistics of all used variables can be found in Tab. 1. Phrases have an average duration of about 3 sec, which coincides with estimates for the subjective presence time (Fraisse, 1982). This is also in good agreement with previous studies on ideational flow in jazz piano solos (Schütz, 2011; Lothwesen & Frieler, 2013). The mean length in bar units is about 2 bars.

Curiously, the mode of the distribution lies at 19 phrases with 8 instances, most likely just a chance result. For phrase lengths less or equal 16 phrases, there is a preference for an even number of phrases, which is the case for 77% solos. This is in agreement with our observations for folk songs (Frieler, 2014), and might be a reflection of the form of the underlying compositions, which are rooted in Western music traditions and thus prefer even-numbered structures. Beyond a length of 16 phrases, the situation is rather opposite—about 60% have odd number of phrases. Interestingly, the number of phrases is strongly decreasing with chorus position. A Spearman rank test for phrase count and chorus position became highly significant ($p < 0.0001, \rho = -0.42$, cf. Tab. 2). While the first choruses have a mean of 10.6 phrases, this number drops down to 4.0 for the sixth chorus. This correlation is rather stable; disregarding very short (less than 3 choruses) and very long solos (more than 11 choruses it is still highly significant ($p < 0.0001, \rho = -0.46$).

In Tab. 2, Spearman rank correlations between chorus and phrase position and various descriptors can be found. Most of these correlation are either non-significant or they indicate uncorrelatedness. Only event density (tones per second) is increasing with chorus and phrase position. However, this correlation is only occasionally present on the level of single solos (9 solos had significant correlation of event density with chorus position, and 13 with phrase position, but not always in the same direction).

In total, 72 different interval phrase form strings (IF) and 97 duration-based phase form strings (DF) were found, which gave rise to 58 distinct combined phrase form classes (CF). Combining was done by enumerating each unique pair of IF and DF symbols of a song with a new form symbol. As expected, many form strings were basically sequences of different form parts with only occasional repetitions. The most frequent form was `ABCDEFGH` with 5 instances for IF and 3 instances for DF.

One can define the *coherence* of a form as the amount of contained repetition, i. e. the number of unique elements divided by the total length of the form (subtracted from 1 for better interpretation). A coherence of 0 means that no form part is repeated, whereas a coherence of 1 can only be reached in the limit of a single, infinitely repeated part. We found a median coherence for IFs and CFs of 0.0, and for DFs of 0.25 (c.f. Tab. 1). As for folk songs, DF coherence is much higher than IF coherence (Frieler, 2014). Generally, the very low IF coherence might indicate that motivic improvisation is not very common or, alternatively, that it is not captured by our (rather simplistic) method of similarity calculation. Moreover, motivic improvisations should result in runs of similar phrases. To check this, we calculated run lengths for all form strings of all types. The median IF runlength is 0 (AM=0.42), the median DF runlength is 1 (AM=1.52), and the median CF runlenght is 0 (AM=0.18). Three-quarter of all solos have not a single IF run, whereas 38 have no DF run, and 88 solo no CF run. The longest IF run with 5 repetitions, which was also the longest CF run, occured in Freddie Hubbards solo on "Society Red". The longest DF with 11 elements was performed by John Coltrane in his legendary solo on "Giant Steps". A analysis of variance on DF run lengths revealed that John Coltrane is the only soloist in the database with a tendency to use DF runs ($F(37, 62) = 1.704, p = 0.031, R_{adj}^2 = 0.21$). With regard to IF and CF runs, no difference between performers could be found, the same holds for IF and CF coherences. Again, John Coltrane was the only soloist which differed in DF coherence ($F(37, 62) = 2.011, p = 0.007, R_{adj}^2 = 0.27$). Coltrane's mean DF coherence is 0.22 [2], about twice as high as the overall mean of 0.13. Furthermore, a analysis of variances for mean run lengths of the three different form types was carried out. For IF, Curtis Fuller and Kenny Garrett showed a tendency to longer runlengths. For DF, this was the case for Coleman Hawkins, Bob Berg, and Clifford Brown. However, this analysis has to be taken with care, because in general runs occur rarely, and for the majority of soloist there are only 1 or 2 solos in the database.

## 5. CONCLUSION

In this explorative study, we reported on basic statistics of jazz solo phrases in a relatively large set of 100 solos. We found that event density correlates high with chorus and phrase position, which might result in an increase in intensity during the course of a solo. The fact, that this correlation is only found at the corpus level and only occasionally for individual solos, can be interpreted such, that solos in general show a tendency to increase in intensity at the beginning. The same effect is observed for phrase positions as well.

---

[2] The DF coherence values for all 7 solos of John Coltrane are "Central Park West": 0.00, "26-2": 0.13, "Mr. P.C.": 0.16, "Blue Train": 0.17, "Countdown": 0.26, "Giant Steps": 0.31, "So What": 0.54.

| Variable | Median | AM | SD | Range |
|---|---|---|---|---|
| Choruses/solo | 2.00 | 3.54 | 3.57 | $(1, 19)$ |
| Phrases/solo | 23.00 | 25.94 | 15.15 | $(3, 66)$ |
| Phrases/chorus | 18.00 | 20.97 | 13.26 | $(1, 66)$ |
| Tones/solo | 370.50 | 411.90 | 261.87 | $(50, 1172)$ |
| Tones/chorus | 92.00 | 113.90 | 80.71 | $(10, 616)$ |
| Tones/phrase | 12.00 | 15.9 | 13.5 | $(1, 129)$ |
| Event density/phrase (sec) | 5.38 | 5.76 | 2.43 | $(0.26, 19.47)$ |
| Event density/phrase (bars) | 7.62 | 8.65 | 4.36 | $(0.74, 38.79)$ |
| Duration/phrase (sec) | 2.37 | 2.95 | 2.28 | $(0.03, 20.64)$ |
| Duration/phrase (bars) | 1.53 | 2.04 | 1.71 | $(0.23, 12.62)$ |
| IF coherence | 0.00 | 0.04 | 0.06 | $(0.00, 0.25)$ |
| DF coherence | 0.26 | 0.28 | 0.20 | $(0.00, 0.83)$ |
| CF coherence | 0.00 | 0.02 | 0.04 | $(0.00, 0.22)$ |

**Table 1**: Descriptive statistics of various indices. Range is indicated in the format (min, max). AM=arithmetic mean, SD=standard devisation. X/Y means "(Number of) X per Y". Densities and duration are measured absolutely in seconds or relatively in bar units. Very small event-densities are due to one-note phrases. IF, CF, and DF coherences measure the amount of repetition in a phrase form structure. For more details see text.

| Position | Variable | $\rho$ | p |
|---|---|---|---|
| Chorus position vs. | Number of phrases | $-0.51$ | $0.000^{***}$ |
| | Number of notes | $+0.07$ | $0.002^{***}$ |
| | Event density (sec) | $+0.23$ | $0.000^{***}$ |
| | Event density (bars) | $-0.03$ | $0.129$ |
| | Total duration (sec) | $-0.04$ | $0.024^{*}$ |
| | Total duration (bars) | $+0.09$ | $0.000^{***}$ |
| | IF coherence | $-0.06$ | $0.204$ |
| | DF coherence | $-0.08$ | $0.112$ |
| | CF coherence | $-0.03$ | $0.566$ |
| | Coherence difference | $-0.06$ | $0.253$ |
| Phrase position vs. | Number of notes | $-0.02$ | $0.248$ |
| | Event density (sec) | $+0.17$ | $0.000^{***}$ |
| | Event density (bars) | $+0.04$ | $0.029$ |
| | Total duration (sec) | $-0.11$ | $0.000^{***}$ |
| | Total duration (bars) | $-0.03$ | $0.083$ |

**Table 2**: Spearman rank correlations $\rho$ of phrase statistics vs. chorus and phrase position.

For the significant decrease of phrase count with chorus position, we have currently no explanation. Regarding coherence, a few individual differences could be found, especially for John Coltrane, who has a much higher DF coherence. As expected, all coherences are generally rather low, but DF coherence is always higher than IF or CF coherence, just as for folk songs (Frieler, 2014). This could hint to an universal phenomenon in music. However, it might primarily be the result of a much smaller space of possible durations classes (here: five) compared to a larger event space for intervals, even though both event spaces are constrained by further tonal or metrical conditions. This need further examination.

All in all, phrase form structure as such seems to be only weakly exploitable for useful features, though, as the solos of John Coltrane show, it might be possible to capture exceptional cases. Other phrase characteristics such as event densities are more likely to be useful. However, our findings nicely corroborate the above cited analogy of jazz solos to ad-lib speeches (Johnson-Laird, 2002). In contrast, folk songs might more comparable to poems, which classically show more inner coherence and regularities than speeches or other narrative prose.

# 6. REFERENCES

Downs, C. (2000-2001). Metric displacement in the improvisation of Charlie Christian. *Annual Review of Jazz Studies*, *11*, 39–68.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 149–180). New York: Academic Press.

Frieler, K. (2014). Exploring phrase form structure. Part I: European Folk Songs. In Holzapfel, A., Cemgil, T., Mungan,

E., & Bozkurt, B. (Eds.), *Proceedings of the Fourth International Workshop on Folk Music Analysis (FMA2014), Istanbul*. Bogazici University.

Frieler, K., Abeßer, J., Zaddach, W.-G., & Pfleiderer, M. (2013). Introducing the Jazzomat Project and the Melo(S)py Library. In van Kranenburg, P., C. Anagnostopoulou, C., & Volk, A. (Eds.), *Proceedings of the Third International Workshop on Folk Music Analysis, Meertens Institute and Utrecht University Department of Information and Computing Sciences*, (pp. 76–78).

Johnson-Laird, P. N. (2002). How jazz musicians improvise. *Music Perception*, *10*, 415–442.

Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, *163*(4), 845–848. Englische Übersetzung in: Soviet Physics Doklady, 10(8) S. 707-710, 1966.

Lothwesen, K. & Frieler, K. (2013). Einflussfaktoren und Gestaltungsprinzipien formelbasierter Jazzimprovisation. In Jeßulat, A., Lehmann, A., & Wünsch, C. (Eds.), *Kreativität - Struktur und Emotion*, (pp. 256–265)., Würzburg. Königshausen & Neumann.

Love, S. (2012). An approach to phrase rhythm in jazz. *Journal of Jazz Studies*, *8*(1), 4–32.

Müllensiefen, D. & Frieler, K. (2004). Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments. *Computing in Musicology*, *13*, 147–176.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Schütz, M. (2011). Improvisation im Jazz: Ideenflussanalyse von Jazzpiano-Improvisationen. Master, University of Hamburg.