TOWARDS A FORMALIZATION OF MUSICAL RHYTHM

Martin Rohrmeier
Digital and Cognitive Musicology Lab, École Polytechnique Fédérale de Lausanne
martin.rohrmeier@epfl.ch

ABSTRACT

Temporality lies at the very heart of music, and the play with rhythmic and metrical structures constitutes a major device across musical styles and genres. Rhythmic and metrical structure are closely intertwined, particularly in the tonal idiom. While there have been many approaches for modeling musical tempo, beat and meter and their inference, musical rhythm and its complexity have been comparably less explored and formally modeled. The model formulates a generative grammar of symbolic rhythmic musical structure and its internal recursive substructure. The approach characterizes rhythmic groups in alignment with meter in terms of the recursive subdivision of temporal units, as well as dependencies established by recursive operations such as preparation and different kinds of shifting (such as anticipation and delay). The model is formulated in terms of an abstract context-free grammar and applies for monophonic rhythms and harmonic rhythm.

1. INTRODUCTION

Temporality lies at the very heart of music, and the play with rhythmic and metrical structures constitutes a major device across musical styles and genres. However, there is comparably less research on rhythm itself than on other temporal structures of music. For instance, there is a lot of work on modeling beat and beat inference [1–3], tempo estimation [4] as well metrical structure [5] and its inference [6–10]. There has been major theoretical work differentiating between grouping and meter [11], and between rhythm and meter [5, 12]. In comparison, there is less formalization work on musical rhythm [13–15], and some major studies such as the GTTM [11] or [5] avoid a formal characterization of musical rhythm. The purpose of this theoretical paper is to address this gap and to provide a generative model of musical rhythm in terms of an abstract context-free grammar that generates rhythmic structure in alignment with metrical structure.

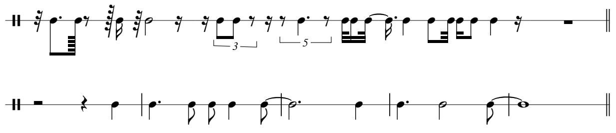


Figure 1: Two different series of onsets and durations.

2. MOTIVATION

Rhythm is commonly thought of as a series of onsets and durations of musical events. While duration patterns are essential for musical rhythm, the core idea of the model is to capture that rhythmic structures, especially those in tonal music, are more than (fully) freely placed onsets over time and that the concept of rhythm involves different kinds of dependencies that are constituted between its musical events. To illustrate this point, Figure 1 displays two different series of onsets and durations. Only the lower one looks like a plausible candidate for a rhythm from a tonal piece of music. There are several points underpinning this distinction. In essence, it is argued that rhythm is understood involving an *interpretation* in terms of hierarchical dependencies of temporal events and their assignment to the metrical grid, which result in a surface projection of patterns of onsets and durations.

One central point is that the rhythmic Gestalt is fundamentally defined by its relation to metrical structure; rhythm cannot be separated from a metrical interpretation. This point is illustrated by Figure 2. Both rhythms have the identical sequence of onsets and durations, but different metrical structures associated with them, and this results in both rhythms sounding very different. In particular, the ways in which events are linked to weak and strong metrical beats and also their underlying meter have a major impact on the interpretation of a given pattern of durations.

A second major point lies in the fact that we characterize rhythmic structures in terms of an *interpretation* by event dependencies and transformations. For instance, certain events lead to other events; we understand a certain event or group of events as an upbeat to (or preparation of) another event; and we understand certain events as subdivisions of longer units (such as triplets). Furthermore, when we speak about syncopation, anticipation, or delay, it means that certain events occur earlier or later, implying that there is an (underlying) position where these events would have been expected normally before the transformation (shift) [13, 16]. Rhythmic events are also recognized as grouped.

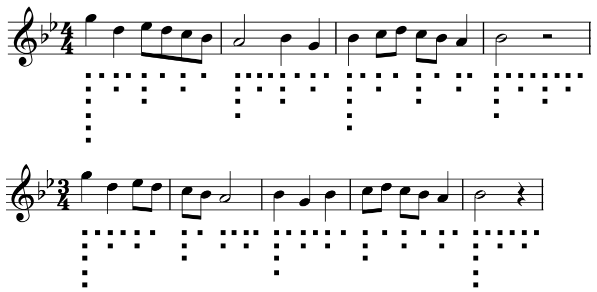


Figure 2: The alignment with metrical structure may yield two very different rhythmic interpretations for the same pattern of onsets and durations.

Generally, it is useful to distinguish different levels of musical time [17–20]. Rhythmic structure, in the aspect that is modeled here, lives in *idealized* time. The metrical grid presumes an underlying isochronic beat, and rhythmic patterns are related by simple integer ratios in relation to the grid and the beat. The rhythmic structure at this symbolic level in relation to the symbolic beat is different from the level of tempo variation, expressive timing, swing, groove, performance errors and other subsymbolic variations of timing.

In sum, we generally conceive of rhythms as structured both in terms of an associated metrical structure as well as in terms of event dependencies such as the ones mentioned. In contrast, events occurring with purely random onsets and durations (like the first example in Figure 1) sound erratic—which in turn means that they have no *interpretation* in terms of the dependencies outlined. It is the purpose of the proposed model to express the various rhythmic dependencies at the deep structure that give rise to the patterns of event onsets and durations observed at the surface.

One common observation in rhythmic structure is that events may reach into the timespan of other events. This is particularly common with preparations before an event, anticipations or syncopations that may enter during the timespan of the directly preceding event. This may cause the preceding event to be shortened, which we refer to as *time stealing* when it is discussed below.

2.1 Related literature

Numerous approaches have addressed rhythmic and metrical structure in music [5, 21]. While there are several research directions in terms of rhythmic corpus studies [22–27] and mathematical analyses [28–30], there is less research proposing formal theoretical frameworks generating rhythmic structure. Differentiating metrical structure from grouping, the GTTM [11] laid a foundation for the understanding of meter that is still in place today. Several endeavors have been devoted to implementing the GTTM in a computational way [31, 32].

Several computational approaches to rhythm have proposed sequential models such as Markov models, HMMs or other graphical models [33, 34]. More recently various hierarchical approaches and probabilistic grammars have been used for rhythmic inference and transcription problems [15,35–37]. These approaches are essentially built on recursive subdivision (split). [14] proposed an algorithmic model of rhythm using transformations of syncopation, figural, and density (split) based on the transformation vector proposed by [13]. From the perspective of mathematical music theory, rhythmic structure has also been modeled in terms of subdivision of a graph [38].

The present model extends previous approaches [14,15, 39] by characterizing an overarching abstract context-free grammar of recursive rhythmic dependencies. It is based on five abstract operations of splits, preparations, and shifts using a tripartite representation of rhythmic categories, and models rhythmic conflicts using the concept of *time stealing*. As a grammar-based generalized model of rhythm, it can be naturally integrated with syntactic models of harmony [39–48] for modeling harmonic rhythm.

3. THE FORMALISM

3.1 Metrical structure

Metrical structure has been famously modeled by [11] with a recursive grid of metrical weights and a notation adopted from metrical phonology in linguistics [49–53]. Examples of the metrical grid are shown in Figures 2 and 3. In the grid, each level m is characterized by a multiple (2 or 3) of the period of the subordinate level $m-1$ and an offset. Generally, the subdivision for regular meters is binary or ternary [5]; for irregular meters the formalization would need to be extended to combinations of twos and threes (e.g. $\frac{7}{8} = \frac{3}{8} + \frac{2}{8} + \frac{2}{8}$).

The metrical grid can be characterized as follows for regular meters: the beat level is marked with m_0 and the beginning of the segment or piece is indicated by the index 0, and locations are indicated in reference to m_0. Each higher metrical level i is characterized by the tuple (π_i, o_i), the regularity $\pi_i \in \{2,3\} \cdot \pi_{i-1}$, and the offset $o_i := o_{i-1} + a\pi_{i-1}, with a \in \mathbb{N}_0$, $o_0 = 0$ and $\pi_0 = 1$. This ensures that higher levels can only subselect beats established in all lower metrical levels. Metrical levels below the beat ($i < 0$) are characterized in the same way with $\pi_i \in \{\frac{1}{2}, \frac{1}{3}\} \cdot \pi_{i+1}$, and the offset $o_i = 0$. Accordingly, a metrical grid M is fully defined by the list (or series) of all tuples $M := (\pi_i, o_i)$. The metrical grid is potentially infinite in duration and has an arbitrary number of metrical levels i. Most commonly, almost all values of π are 2, except for the levels 1 and 2 in ternary meters.

The metrical weight at position t is characterized as:

$$W_{M:=(\pi_i, o_i)}(t) = \sum_i \big(1 - sign(\mid t - o_i \mid \bmod \pi_i)\big) \quad (1)$$

where the $1 - sign(\cdot)$ function is used to compute 1 for a position falling on the metrical grid and 0 otherwise. Furthermore, a subsegment of a metrical grid M is characterized by $M_{[a,b]}$, where a and b denote the beginning and end locactions of the open or closed subsegment interval.

3.2 Generative rhythm

The formalism is modeled employing abstract context-free grammars [47]. The grammar consists of four parts: $\mathcal{G} = (C, \Sigma, P, C_0)$, non-terminal rhythmic categories C, terminal symbols Σ, production rules P, and, in this case, a set of start symbols $C_0 \subseteq C$. All sets C, Σ, C_0 are infinite.

A rhythmic category consists of a tuple (t, m): a timespan t and the metrical grid m associated with the timespan. The timespan represents a formalization of the rhythmic time interval (which is going to be generatively subdivided). It is itself a triple $[a : b : c]$ that combines three parts: a downbeat part of a duration b, also called the *body* of t, and an initial offset (upbeat) of duration a, and final offset (coda) of duration c. Durations are defined in beats ($\in \mathbb{Q}$) including fractions of beats, such as $\frac{1}{8}$. The offset parts can be positive or negative. If positive, a time segment is added to the core length of b; if negative, b is shortened by that amount. The total duration of t is $a + b + c$. While in practice, a and c are each mostly no longer than $\frac{b}{2}$, it is avoided to postulate such a restriction theoretically rather than empirically. However, $-b < a < b$, $-b < c < b$, and $a + c <= b$ are required.

The set of surface symbols Σ consists of musical events that are characterized by pairs (l, m) of event durations $l \in \mathbb{Q}, l \leq b$ in beats with associated metrical weigths $m \in \mathbb{Q}$. The separate modeling of the event length in addition to the timespan is important because a realized event may in turn occupy a shorter duration than its timespan (e.g. a quarternote in a half-note timespan, or a staccato note). This makes it possible to model the type of rests that can occur in this case. Since the focus in this paper is about the rhythmic grammar, other features of the musical event (like pitch) will be left out of account.

The set of start symbols $C_0 \subseteq C$ consists of rhythmic categories that do not have a coda offset:

$$C_0 = \{k \mid k = (t = [a : b : c], m) \in C \land c = 0\} \quad (2)$$

Because the symbol space is infinite, rules do not constitute rewrite operations over symbols (as in classical context-free grammars), but as rewrite functions. The set of generative production rules P is characterized as

$$P \subseteq \{r \mid r : C \to (C \uplus \Sigma)^*\}, \quad (3)$$

where the rules are functions that map sets of categories onto sets of categories or surface symbols, establishing an abstraction over different paramaterized instantiations of analogous rules.

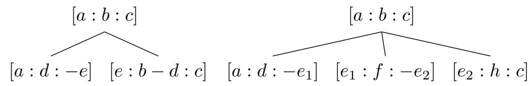
The set of core rules constitutes the heart of the generative formalism. For the generation of rhythmic structure, five main rules are assumed: *split*, *prepare*, *(c-split)*, *anticipate*, and *delay*.

Split rule. The main rule of the formalism is *split*, which subdivides the body of the rhythmic category, while the outer upbeat and coda parts remain identical. The split can induce *timestealing* such that a timespan protrudes into an adjacent one by an upbeat or coda of length e. Note that it does not result in an update of the length of the timespan body b, but of its upbeat or coda part. This maintains the core durations at the deep structure. Split can subdivide a timespan into two or three parts; other subdivisions, such as four or five, etc., require multiple split operations.

$$\begin{aligned}&([a : b : c], m_0) \longrightarrow ([a : d : -e], m_1)\ ([e : b - d : c], m_2)\\ &\quad \mid 0 < d < b, m_1 = m_{0[0,a+d-e[}, m_2 = m_{0[a+d-e,a+b+c]}\end{aligned} \quad (4)$$

$$\begin{aligned}&([a : b : c], m_0) \longrightarrow\\ &([a : d : -e_1], m_1)\ ([e_1 : f : -e_2], m_2)\ ([e_2 : h : c], m_3)\\ &\quad \mid d + f + h = b, \quad m_1 = m_{0[0,a+d-e_1[},\\ &m_2 = m_{0[a+d-e_1,a+d+f-e_2[}, m_3 = m_{0[a+d+f-e_2,a+b+c]}\end{aligned} \quad (5)$$

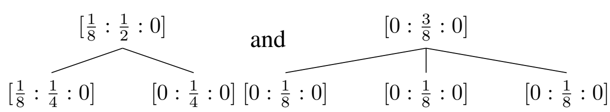
The working of these rules may be visualized by the subtrees they produce (ignoring the metrical assignment):



It is further assumed that the majority of these split operations divide the body of the timespan equally ($d = f = \frac{b}{2}$, for even b) or into simple integer ratios. The actual instantiation of these splits in practice is, however, not a matter that should be decided a-priori at the level of the formalism. Note, for instance, that an unequal subdivision like 3:1 in the context of a long note and an upbeat to the next bar, is not required since such cases are rather expressed with the upbeat (u-split) rule that is explained next.

Also note that for categories where the metrical accent lies at the beginning of the body, it is sufficient to metrically characterize the split segments by the metrical weight at the downbeat (onset) of the body. The notation $m = u \oplus v$ is used to denote the weight of the downbeat in terms of the metrical level u generated by the split operation plus the metrical weight v inherited from the parent node in the tree. Figure 3 and 4 illustrate this notation.

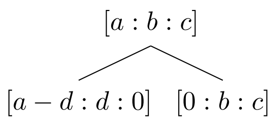
As an example, a halfnote split into two quarter-notes or dotted quarter-note split into three eighth-notes would be expressed like this:



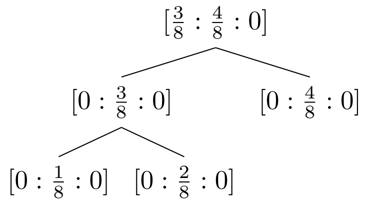
Prepare (U-Split) rule. The second core rule models upbeat structures. It takes the upbeat part of the timespan of a rhythmic category and generates an own rhythmic category from it:

$$\begin{aligned}&([a : b : c], m_0) \longrightarrow ([a - d : d : 0], m_1)([0 : b : c], m_2)\\ &\quad \mid 0 < d \leq a, m_1 = m_{0[0,a[}, m_2 = m_{0[a,b+c]}\end{aligned} \quad (6)$$

The corresponding tree fragment looks like this:



For example, a half-note that is prepared by a combined upbeat of an eighth and a quarternote would be expressed as follows:



One may postulate a corresponding counterpart to the upbeat split in the *prepare* rule; this rule (*c-split*) would instantiate a rhythmic category from the coda-part of a given rhythmic category. While it is unclear if such a rule would indeed be required for tonal rhythm (i.e. a phenomenon of "post-paration" as opposite to "pre-paration"), still the rule is listed even though it may be dropped from the formalism (or be found to not occur empirically):

$$\begin{aligned}([a:b:c],m_0) &\longrightarrow ([a:b:0],m_1)\,([c-d:d:0],m_2)\\ &\mid 0<d\leq c, m_1=m_{0[0,a+b[}, m_2=m_{0[a+b,a+b+c]}\end{aligned} \tag{7}$$

Shift rules. Two further rhythmic phenomena do not relate to subdivision but to the shift of events, such as in the context of syncopations. In these cases, a rhythmic category c may be shifted to occur early or late. The corresponding rules are *anticipate (e-shift)* and *delay (l-shift)*. These rules are unary rules that transform a rhythmic category rather than creating a new one.

$$\text{e-shift: } ([a:b:c],m_0) \longrightarrow ([0:b:a+c],m_0) \tag{8}$$

$$\text{l-shift:} ([a:b:c],m_0) \longrightarrow ([a+c:b:0],m_0) \tag{9}$$

Surface rules. Finally, from the set of recursive generated rhythmic subdivisions and transformations a rhythmic surface will be generated. If events are shortened by *timestealing*, the lengths of the core are updated (by the first two rules). In order to ensure that all upbeat parts and codas have been instantiated either with events or shifts, surface symbols can only be generated for categories that have an empty upbeat and coda part. Once generated, surface symbols cannot reenter the generative process.

$$([a:b:c],m) \longrightarrow ([0:b+a:c],m) \text{ for } a<0 \tag{10}$$

$$([a:b:c],m) \longrightarrow ([a:b+c:0],m) \text{ for } c<0 \tag{11}$$

`surface:`

$$([0:b:0],m) \longrightarrow (b, W_m(0)) \tag{12}$$

$$\begin{aligned}([0:b:0],m) &\longrightarrow (l, W_{m_{[0,l[}}(0))(\epsilon, b-l, W_{m_{[l,b]}}(0))\\ &\mid 0<l\leq b\end{aligned} \tag{13}$$

In other words, the surface rule yields the rhythmic surface duration $l=b$ or $l\leq b$ as well as its metrical weight m. If the event is shorter than its timespan b, the surface rule also creates a rest event ϵ that fills up the remaining space, so that the subsequent events are not affected by the shortening.

The surface rule is designed in such a way that all lengths of all surface events add up to the full length of the entire musical segment from the start symbol:

$$\sum l_i = a+b \quad \text{for } c_0=[a:b:0]\in C_0 \tag{14}$$

An illustrative example of this sum can be reconstructed from the surface-note durations in Figure 3 and 4.

4. EXAMPLES

4.1 Melodic rhythm

A first detailed analysis is carried out on the first two bars of the jazz standard "Blue Bossa". The tree analysis based on the generative model is displayed in Figure 3 together with a corresponding analysis that visualizes the recursive rhythmic subdivisions and shifts of the same generation using musical score lines. All durations are encoded following common music notation, i.e. $\frac{1}{2}$ refers to a half-note, 1 to a whole note, or $\frac{3}{8}$ to a dotted quarter-note; $\frac{1}{4}$ refers to the quarter-note beat level. The figure characterizes the metrical (sub)grids of each category employing the $m=u\oplus v$ notation.

Several observations can be made based on the figure. All of the applications of *split* illustrate that timespans may be subdivided with equal subdivision of the core, yet resulting in unequal timespan durations based on time-stealing effects encoded in the upbeat and coda parts of the timespan category. The derivation of the sixth note provides an example where an event at the metrical whole-note level is syncopated by an eighth note and at the surface instantiated shorter than its timespan resulting in the surface generation of an additional rest.

Further, the first, third, and seventh note may be understood as upbeats to the subsequent events. The second halves of measures 1 and 3 have syncopations in which the shifted events reach into the timespan of the previous events, causing the notes on beats 3 and 1, respectively, to be shortened by one eighth note.

4.2 Harmonic rhythm

The formalism proposed has a different application in the modeling of harmonic rhythm. For instance, this concerns modeling the harmonic-rhythmic structure as it is contained in leadsheets. A major difference between the previous case of melodic rhythm is that harmonic rhythm in leadsheets may well employ (harmonic) upbeats, yet no rhythmic shifts in the sense of syncopation, anticipation or delay; also time overlaps that involve the timespan coda have not been observed. A computational version of this (sub)model has been proposed in [39].

Figure 4 shows an example analysis of the first 8 bars of the harmonic phrase of "Blue Bossa". The tree analysis displays the harmonic syntactic dependencies following [47, 54, 55] in conjunction with the harmonic rhythm

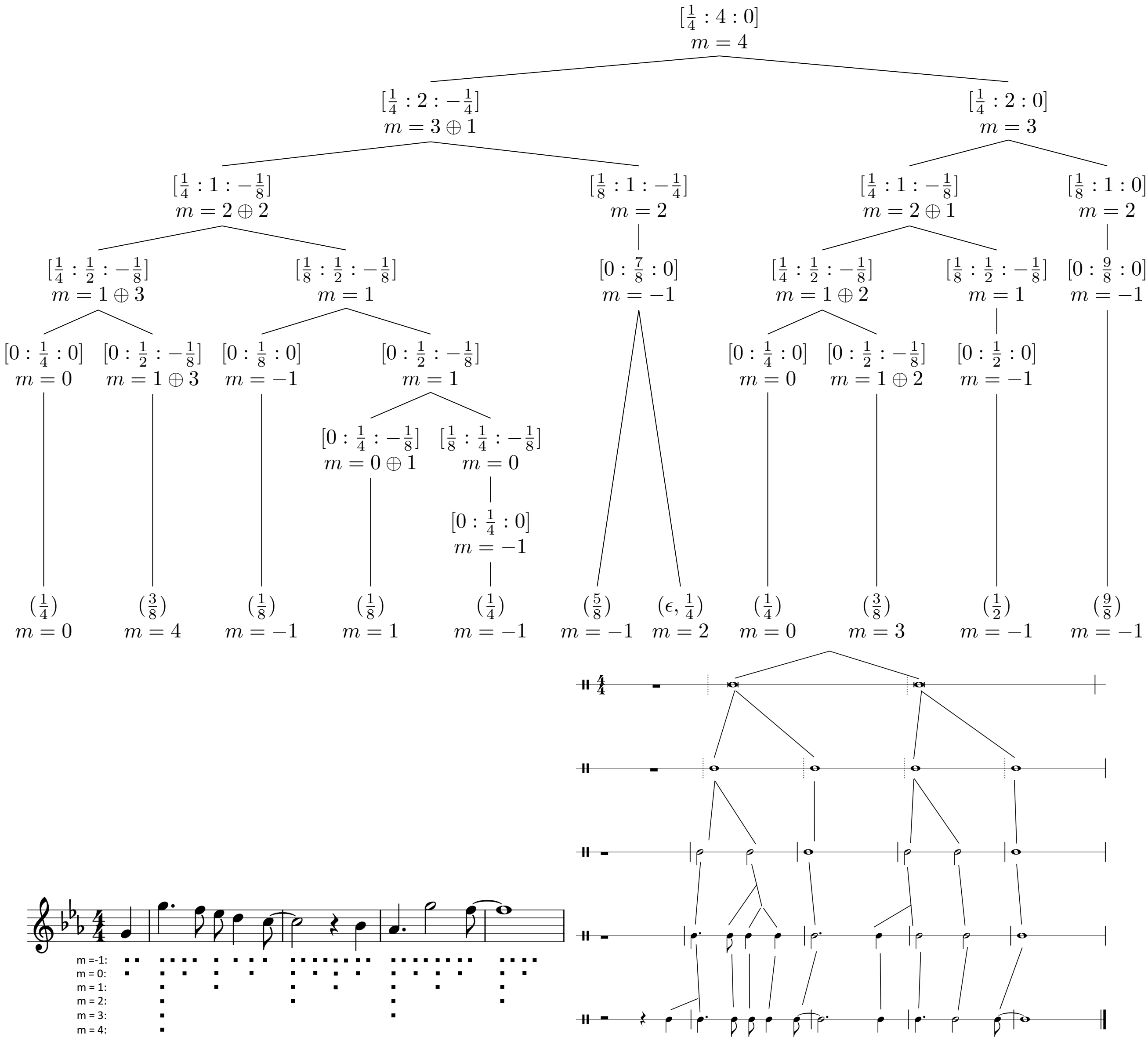


Figure 3: A rhythmic analysis of the first four bars of the melody of the Jazz standard "Blue Bossa".

and the metrical levels. This requires a product grammar as defined by [39] for the formulation of the coordination between harmonic and rhythmic structures—which is an application of the formalism as proposed here. The tonic i at the highest level is equally split into two tonic timespans at level 3. When the preparing dominant V^7 is inserted, it causes to take up half of the space of the timespan of i and causes i to appear later. When iv is introduced preparing V^7 it is introduced at the same metrical level (level 2) and reaches into the time domain of the initial i (timestealing). By analogy, the introduction of $ii^{\varnothing}$ takes up the half of the V^7 time domain. Accordingly, the analysis reveals that the metrical domains of the chords in the hierarchical analysis are not identical with the position where the chords occur on the surface. Figure 4 (a) displays the stepwise joint derivation of harmonic syntactic dependencies and harmonic rhythm.

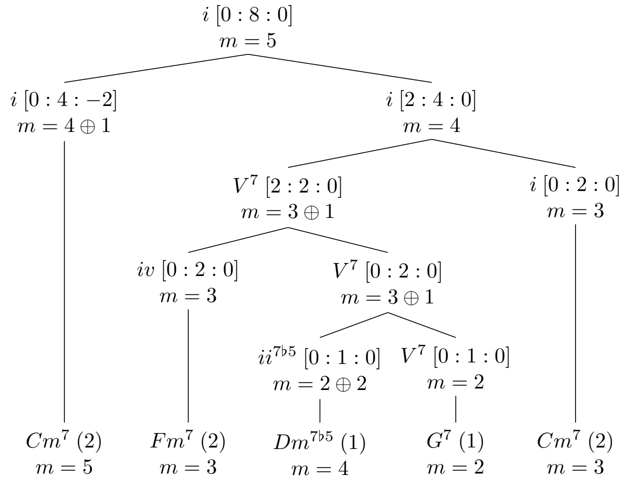
5. DISCUSSION

The contribution of this paper is to characterize the recursive internal structure of musical rhythms using a formal grammar. This goes beyond the GTTM, which does not propose a model of rhythm, and further argues that the inference of the hierarchical rhythmic deep structure is central to music cognition. Because of the joint representation of rhythmic and metrical structure in the model, a parser of the proposed abstract grammar of musical rhythm instantiates rhythmic interpretation and metrical inference at the same time.

In this formalism, the concept of *timestealing* is proposed. It is modeled at the highest metrical level it affects, and the split operation already sets up the timespans for subsequent preparation or shift operations in the upbeat or coda parts of the timespan category. This modeling ensures that all operations remain context-free and could be implemented and parsed efficiently with a parser of abstract

4:	i							
3:	i				i			
2a:	i				V^7		i	
2b:	i		iv		V^7		i	
1:	i		iv		$ii^{\varnothing}$	V^7	i	
s:	Cm^7		Fm^7		$Dm^{\varnothing}$	G^7	Cm^7	

(a) Schematic stepwise top-down generation of the chord sequence including harmonic rhythm. The generation occurs top-down from the single tonic i at metrical level 4. If a bar is empty, it is in the domain of the preceding chord. The row numbers indicate the metrical level. There are two operations at the same metrical level. The last row contains the chord sequence as it appears in the leadsheet.



(b) Harmonic syntax tree co-generated with rhythmic-metrical structure. At the surface, the numbers in parenthesis encode duration and m its metrical weight (see also part (a) of the figure).

Figure 4: Syntactic analysis of the harmony and its rhythm in the first phrase of the Jazz standard "Blue Bossa" in C minor.

context-free grammars. This would not be guaranteed if the upbeat feature would only be instantiated at the upbeat or shift operation, for instance. If such an upbeat would reach across a border at a different metrical level the information to adapt the coda part of the adjacent subtree would have to perculate through the tree in a context-sensitive fashion, thus resulting in a model of much higher computational complexity. Such an instance could, for example, be observed in the syncopation of the sixth note in Figure 3. The eighth note syncopation of a note at the whole-note level results in a shift of the corresponding right neighbor at the quarternote level, accordingly the information would have to traverse one node up and four nodes down the tree to reach the right node. Further, a generation of upbeats and shifts without timespan reservation may result in the generation of impossible structures if both sides expand in a unrestricted context-free fashion.

Because of the hierarchical modeling of shifts of timespans, it is not necessary to include "hacks" such as binding-over of events as in musical notation (as in the long notes C and F in Figure 3) since the logic of syncopation can be modeled directly. With the formalism and the upbeat feature it is further possible to model the rhythmic displacement of an entire group of events, such as the syncopation of four quarternotes by an eighth note.

The tripartite representation of a timespan with upbeat, core, and coda parts makes it possible to maintain the simple split ratios at the deep structure. It also models the normalized locations where syncopations originated, as well as the overarching timespan that a deeper event dominates even though it may only occur at a different surface position (such as the tonic or dominant symbols in Figure 4). Maintaining simpler deep structure relations aggregates of similar rules, which establishes theoretical parsimony and facilitates probabilistic modeling and inference.

In the presented model rhythmic structure is generated in alignment with meter. It is possible to devise a variant of the model such that metrical structure is *co-generated* jointly with rhythm rather than having it defined with the start category. The additive $m = u \oplus v$ characterization as used in Figure 3 and 4 defines metrical weight with a current metrical level and a part inherited from the parent. Split, prepare and shift rules can be redefined in such a way that they recursively generate each successive metrical level. While such an approach has advantages for complex and irregular rhythms, it would require additional constraints to ensure metrical consistency across independent context-free subtrees.

The proposed formalism models the generation of a single rhythmic sequence. For musical structures with multiple streams or voices, additional parallel trees can be instantiated which need to fulfill the constraint that their derived metrical structures are aligned. Moreover, complex (non-Western) rhythms and meters can be modeled with an extended model of meter that allows for non-isochronous or additive subdivisions [5]. Application of a computational implementation of the model would be measures of rhythmic complexity based on the derivation tree as well as rhythmic similarity based on largest embeddable common subtrees as for instance employed in [43].

The fact that rhythmic relations instantiate upbeats and splits is closely related to the core syntactic principles of preparation and prolongation [54]. This is corroborated by the fact that harmonic syntax and rhythm are found to be highly correlated in computational modeling [39].

Finally, there is also a close relation between such recursive rhythm and grouping; in fact, the higher order rhythmic categories reflect or *constitute* the grouping structure, refining the concept from the GTTM [11]. Crucially, it has not been argued that this model holds for all rhythms found in musical practice. Rather, the formalism models rhythmic interpretability based on the deep structural rhythmic dependencies outlined; while music is highly flexible, certain complex rhythms (Figure 1) may not be interpretable.

6. ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No 760081 – PMSB.

7. REFERENCES

[1] A. Holzapfel, F. Krebs, and A. Srinivasamurthy, "Tracking the 'odd': Meter inference in a culturally diverse music corpus," in *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR)*. ISMIR, 2014, pp. 425–430.

[2] F. Krebs, S. Böck, and G. Widmer, "Rhythmic pattern modeling for beat and downbeat tracking in musical audio." in *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, 2013, pp. 227–232.

[3] H. Merchant, J. Grahn, L. Trainor, M. Rohrmeier, and T. Fitch, "Finding the beat: a neural perspective across humans and non-human primates," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 370, no. 1664, p. 20140093, 2015.

[4] S. Dixon, "Automatic extraction of tempo and beat from expressive performances," *Journal of New Music Research*, vol. 30, no. 1, pp. 39–58, 2001.

[5] J. London, *Hearing in time*. Oxford: Oxford University Press, 2004.

[6] C. Lee, "The rhythmic interpretation of simple musical sequences: Towards a perceptual model," in *Musical structure and cognition*, P. Howell, I. Cross, and R. West, Eds. London: Academic Press., 1985, pp. 53–69.

[7] ——, "The perception of metrical structure: Experimental evidence and a model," in *Representing Musical Structure*, P. Howell, R. West, and I. Cross, Eds. London: Academic Press, 1991, pp. 59–127.

[8] D. Temperley and D. Sleator, "Modeling meter and harmony: A preference-rule approach," *Computer Music Journal*, vol. 23, no. 1, pp. 10–27, 1999.

[9] A. P. Klapuri, A. J. Eronen, and J. T. Astola, "Analysis of the meter of acoustic musical signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 342–355, 2005.

[10] B. van der Weij, M. T. Pearce, and H. Honing, "A probabilistic model of meter perception: Simulating enculturation," *Frontiers in Psychology*, vol. 8, p. 824, 2017.

[11] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press, 1983.

[12] C. Hasty, *Meter as Rhythm*. Oxford: Oxford University Press, 1997.

[13] G. Sioros and C. Guedes, "Syncopation as transformation," in *Sound, Music, and Motion*, M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, Eds. Springer International Publishing, 2014, vol. Vol. 8905, pp. 635–658.

[14] G. Sioros, M. E. P. Davies, and C. Guedes, "A generative model for the characterization of musical rhythms," *Journal of New Music Research*, vol. 47, no. 2, pp. 114–128, 2018.

[15] F. Foscarin, F. Jacquemard, and P. Rigaux, "Modeling and learning rhythm structure," in *Proceedings of the Sound and Music Computing Conference (SMC)*, 2019.

[16] D. Temperley, "Syncopation in rock: a perceptual perspective," *Popular Music*, vol. 18, no. 1, pp. 19–40, 1999.

[17] E. Clarke, "Levels of structure in the organization of musical time," *Contemporary Music Review*, vol. 2, no. 1, pp. 211–238, 1987.

[18] ——, "Rhythm and timing in music," in *The Psychology of Music*, 2nd ed., D. Deutsch, Ed. San Diego: Academic Press, 1999, pp. 473–500.

[19] H. Honing, "Structure and interpretation of rhythm and timing," *Tijdschrift voor Muziektheorie*, vol. 7, no. 3, pp. 227–232, 2002.

[20] ——, "Structure and interpretation of rhythm in music," in *The Psychology of Music*, 3rd ed., D. Deutsch, Ed. Academic Press, 2013, pp. 369–404.

[21] D. Temperley, *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press, 2001.

[22] M. Mauch, "A Corpus-based study of rhythm patterns," in *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, 2012, pp. 163–168.

[23] A. Holzapfel, "Relation between surface rhythm and rhythmic modes in turkish makam music," *Journal of New Music Research*, vol. 44, no. 1, pp. 25–38, 2015.

[24] H. V. Koops, A. Volk, and W. B. De Haas, "Corpus-based rhythmic pattern analysis of ragtime syncopation," in *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*. ISMIR press, 2015, pp. 483–489.

[25] K. Salley and D. T. Shanahan, "Phrase rhythm in standard jazz repertoire: A taxonomy and corpus study," *Journal of Jazz Studies*, vol. 11, no. 1, pp. 1–39, 2016.

[26] J. London, R. Polak, and N. Jacoby, "Rhythm histograms and musical meter: A corpus study of malian percussion music," *Psychonomic bulletin & review*, vol. 24, no. 2, pp. 474–480, 2017.

[27] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, Massachusetts: MIT Press, 2006.

[28] E. Amiot, *Music through Fourier space*. Berlin: Springer, 2016.

[29] G. T. Toussaint, "A mathematical analysis of african, brazilian, and cuban clave rhythms," in *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science*, 2002, pp. 157–168.

[30] E. D. Demaine, F. Gomez-Martin, H. Meijer, D. Rappaport, P. Taslakian, G. T. Toussaint, T. Winograd, and D. R. Wood, "The distance geometry of music," *Computational Geometry: Theory and Applications*, vol. 42, no. 5, pp. 429–454, 2009.

[31] M. Hamanaka, K. Hirata, and S. Tojo, "Automatic generation of grouping structure based on the GTTM," in *Proceedings of the International Computer Music Conference (ICMC)*, 2004, pp. 141–144.

[32] ——, "Implementing "A Generative Theory of Tonal Music"," *Journal of New Music Research*, vol. 35, no. 4, pp. 249–277, 2006.

[33] M. Pearce, "The construction and evaluation of statistical models of melodic structure in music perception and composition," Ph.D. dissertation, City University, London, 2005.

[34] C. Raphael, "Automated rhythm transcription," in *Proceedings of the 2nd International Society for Music Information Retrieval Conference (ISMIR)*, 2001, pp. 99–107.

[35] F. Foscarin, F. Jacquemard, P. Rigaux, and M. Sakai, "A parse-based framework for coupled rhythm quantization and score structuring," in *Mathematics and Computation in Music*, 2019.

[36] A. McLeod and M. Steedman, "Meter detection in symbolic music using a lexicalized pcfg," in *Proceedings of the 14th Sound and Music Computing Conference (SMC)*, 2017, pp. 373–379.

[37] ——, "Meter detection and alignment of midi performance," in *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, 2018, pp. 113–119.

[38] J. Yust, *Organized Time: Rhythm, Tonality, and Form*. Oxford: Oxford University Press, 2018.

[39] D. Harasim, T. J. O'Donnell, and M. A. Rohrmeier, "Harmonic syntax in time: Rhythm improves grammatical models of harmony," in *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR)*. ISMIR, 2019, pp. 335–342.

[40] M. Steedman, "The Blues and the abstract truth: Music and mental models," in *Mental Models in Cognitive Science: Essays in Honour of Phil Johnson-Laird*, J. Oakhill and A. Garnham, Eds. Psychology Press, 1996, ch. 15, pp. 305–318.

[41] S. Tojo, Y. Oka, and M. Nishida, "Analysis of chord progression by HPSG," in *Proceedings of the 24th IASTED International Conference on Artificial Intelligence and Applications*, Innsbruck, Austria, 2006, pp. 305–310.

[42] M. Rohrmeier, "Towards a generative syntax of tonal harmony," *Journal of Mathematics and Music*, vol. 5, no. 1, pp. 35–53, 2011.

[43] B. De Haas, M. Rohrmeier, R. Veltkamp, and F. Wiering, "Modeling Harmonic Similarity Using a Generative Grammar of Tonal Harmony," in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, G. et al., Ed., 2009.

[44] W. B. De Haas, "Music information retrieval based on tonal harmony," Ph.D. dissertation, Utrecht University, 2012.

[45] M. Granroth-Wilding and M. Steedman, "Statistical Parsing for Harmonic Analysis of Jazz Chord Sequences," *Proceedings of the International Computer Music Conference (ICMC)*, vol. 2012, pp. 478–485, 2012.

[46] M. Granroth-wilding and M. Steedman, "A Robust Parser-Interpreter for Jazz Chord Sequences A Robust Parser-Interpreter for Jazz Chord Sequences," *Journal of New Music Research*, vol. 43, no. 4, pp. 355–374, 2014.

[47] D. Harasim, M. A. Rohrmeier, and T. J. O'Donnell, "A generalized parsing framework for generative models of harmonic syntax," in *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, 2018, pp. 152–159.

[48] D. Quick and P. Hudak, "Grammar-based automated music composition in Haskell," *Proceedings of the first ACM SIGPLAN workshop on Functional Art, Music, Modeling & Design - FARM '13*, pp. 59–70, 2013.

[49] N. Chomsky and M. Halle, *The sound pattern of English*. New York: Harper & Row, 1968.

[50] M. Liberman and A. Prince, "On stress and linguistic rhythm," *Linguistic inquiry*, vol. 8, no. 2, pp. 249–336, 1977.

[51] B. Hayes, *Metrical stress theory: Principles and case studies*. University of Chicago Press, 1995.

[52] J. Katz, "Hip-hop rhymes mirror phonological typology," *Lingua*, vol. 160, pp. 54–73, 2015.

[53] U. Reich and M. Rohrmeier, "Batidas latinas – rhythm and meter in spanish and portuguese and other forms of music," in *Phonological Typology of Syllable and Word Languages in Theory and Practice*, ser. (linguae & licerae), J. C. Reina and R. Szczepaniak, Eds. Berlin / New York: Walter de Gruyter, 2014, pp. 391–420.

[54] M. Rohrmeier, “The syntax of jazz harmony: Diatonic tonality, phrase structure, and form,” *Music Theory and Analysis (MTA)*, vol. 7, no. 1, pp. 1–63, 2020.

[55] M. Rohrmeier and M. Neuwirth, “Towards a syntax of the classical cadence,” in *What is a Cadence*, M. Neuwirth and P. Bergé, Eds. Leuven University Press, 2015, pp. 287–338.