A PROSODIC LAYER FOR SUPRASEGMENTAL FEATURES SPANNING RHYTHM TREES, MOTIVATED BY CARNATIC RHYTHMICAL STRUCTURES

Daniel Miller
Independent Researcher
danielmiller.js@gmail.com

ABSTRACT

Rhythm trees model durational structure as recursive equal subdivision, but musical rhythm often involves suprasegmental patterns that cross-cut the subdivision hierarchy—long-distance dependencies that context-free grammars cannot express. Carnatic rhythmic techniques provide clear test cases: *rhythmic sangati* preserves motives across changes in subdivision density and accent placement, while *jathi bhedam* treats irregular accent spans as compositional units.

Drawing on work in prosodic phonology which models intonation as a tree-based hierarchy distinct from segmental structure, we introduce an *accent tree* that overlays a base rhythm tree. The accent tree acts on a depth projection—a sequence of positions at a chosen metrical level—and partitions those positions into spans that may cross subdivision boundaries. Each span either inherits the underlying rhythm or hosts an attached rhythm tree time-scaled to fill it. Accent trees compile to labeled duration sequences that map back to rhythm trees, so the formalism remains compatible with existing rhythm-tree grammars while capturing long-distance dependencies.

1. INTRODUCTION

Rhythm trees provide a compact, context-free account of durational structure as recursive equal-subdivision of time. More precisely, they are *duration* trees: they encode how time is split, not how the resulting units are grouped, stressed, or otherwise made salient. Yeston [1] observed that rhythm, as perceived, involves criteria beyond duration: timbre, dynamics, density, and pattern recurrence help pick out groupings not derivable from subdivision alone. For Yeston, however, these groupings result from interacting periodic strata, not as independent structures in their own right. The Carnatic techniques developed in Section 3 suggest a need for more flexible accentual structures that may form their own hierarchies, unfolding over extended spans of musical time and crossing durational boundaries. [1]

Just as linguistic prosody organizes words into phrases that may cross-cut syntactic boundaries, we argue that an analogous separation is needed for formal models of rhythm: a *duration–accent interface* that treats these as distinct but related structural layers.

Carnatic techniques like jathi bhedam and rhythmic sangati encode musical organization through accentual patterns that have no clear representation as rhythm trees. Recent work has extended context-free rhythm grammars to handle syncopation and anticipation through local mechanisms such as timestealing [2], but the Carnatic cases involve accent constraints that span entire cycles and cannot be localized to adjacent timespans. This is the kind of non-locality for which prosodic phonology introduced a separate structural layer in linguistics, and we argue that the same separation is needed for rhythm.

In prosodic phonology and speech technology, tree based models of intonation treat prosodic structure as a separate hierarchical representation related to, but not reducible to, syntactic structure or segmental strings [3, 4]. Prosodic trees in this tradition organize words or syllables into constituents at several levels, and every node in the tree is interpretively active for prediction and classification tasks. The prosodic hierarchy is anchored in syntax and segmental timing, but it has its own topology and supports systematic patterns of alignment or misalignment between levels.

We propose an analogous architecture for musical rhythm: a two layer model in which a rhythm tree provides an equal subdivision grid and a separate accent tree captures the hierarchical grouping of metrical positions into accented spans, independent of their internal durational content. The accent tree operates over a *depth projection* of the rhythm tree whose groupings may cross-cut the subdivision hierarchy from which its constituent positions are drawn. In line with tree based models of prosodic structure in speech, we treat every node in the accent tree as interpretively active: each node defines a potential accent domain, and the accentual status of any surface event is determined by the lowest accent node that dominates it. This provides a simple formal account of the accentual clash between gati and jathi. Gati fixes a durational grid and suggests one accentual pattern at the subdivision level; jathi induces a potentially conflicting pattern that groups those same positions into spans of different cardinality. Both are represented in the accent tree, and where they conflict, the jathi grouping takes precedence.

[1] Reflecting the term's prominence in the Carnatic theory literature, in what follows we use the term "accent" as a cover term for any rhythmic grouping induced by non-durational features such as timbre, dynamics, or orchestration, imposed on a durational grid that partitions the same temporal span as the subdivision hierarchy but is not derived from it.

Although the formal motivation comes from Carnatic rhythmic theory, the structures it identifies are not confined to that tradition. Figures 2 and 3 reconstruct passages from Elliott Carter's *Eight Pieces for Four Timpani* using the accent-tree formalism, illustrating how Carnatic-like structural elements—cross-cutting accent groupings, recursive subdivision of accent spans, and metric modulation as a consequence of hierarchical repartitioning—arise independently in twentieth-century concert music.

The remainder of the paper proceeds as follows. Section 2 reviews rhythm trees and their rewrite theory. Section 3 summarizes Carnatic rhythmic organization and distills the formal challenges these techniques pose for rhythm-tree representations. Section 4 defines accent trees over depth projections, introduces the attachment model with recursive and iterative modes of application, and shows how rhythmic sangati arises naturally as a compatibility condition between accent structures sharing equal capacity. We conclude with a discussion of directions for future work, including generative grammars for accent trees and computable measures of rhythmic tension.

2. BACKGROUND: RHYTHM TREES

A rhythm tree represents a single span of time that is recursively subdivided into equal parts. Each internal node specifies that its duration is split into p equal segments, where $2 \leq p \leq P$ for some fixed maximal arity P.

Let $\Sigma = \{n, \_\}$ be the alphabet of terminal symbols, where n marks the onset of a new event and $\_$ marks a tie that prolongs the most recent onset. A rhythm tree over Σ is defined recursively as either a leaf $\ell \in \Sigma$ or an internal node of the form

$$t = p(t_1, \ldots, t_p), \tag{1}$$

where $2 \leq p \leq P$ and each t_i is itself a rhythm tree. Following Jacquemard et al. [5], we label each internal node with its arity. The *depth* of a node is its distance from the root, with the root at depth 0.

We assume the root of every rhythm tree spans a duration of 1. Each internal node of arity p splits its incoming duration equally among its children: if a node has duration D and arity p, each child receives duration D/p.

Let $(\ell_0, \ldots, \ell_{L-1})$ be the leaves of t in left-to-right order, with corresponding durations $(d_0, \ldots, d_{L-1})$. The *labeled duration sequence* of t is

$$\mathrm{lds}(t) = ((\ell_0, d_0), \ldots, (\ell_{L-1}, d_{L-1})). \tag{2}$$

Two rhythm trees are *equivalent* when they have the same labeled duration sequence.

Jacquemard et al. [6] give a rewrite system for rhythm trees that is sound and complete: equivalent trees are connected by rewrite paths, and every rewrite preserves rhythmic value. Crucially, each rewrite step is local to the subdivision hierarchy: rules can normalize or repartition existing constituents, but cannot impose groupings that cross-cut the durational tree. No sequence of rewrites can produce a partition of the frontier that does not correspond to some subtree. Any formalism that operates entirely within the subdivision hierarchy will face the same constraint. When musical practice demands groupings that cut across subdivision boundaries, as the Carnatic techniques in Section 3 systematically do, a second structural layer is required.

Rhythm trees, as developed in the computer-aided composition literature [7, 6], provide a structural representation of rhythmic notation without commitment to models of rhythmic cognition. The extensions proposed here maintain this agnosticism: the accent tree is a generative device for imposing groupings that cross-cut the subdivision hierarchy, not a model of metrical perception or structural hearing. In particular, we make no claim about the relationship between accent-tree structure and felt or heard metrical layers. Whether and how these constructions might support analytical or cognitive accounts of layered metric hierarchy remains an open question to which we return in the conclusion.

3. CARNATIC RHYTHMIC ORGANIZATION

We draw our test cases from Carnatic rhythm, focusing on Reina's systematic treatment [8, 9]. Carnatic music provides a particularly stringent test case for rhythmic formalisms because its theoretical vocabulary already distinguishes durational and accentual organization as independent parameters and because its compositional techniques explicitly manipulate the tension between them.

Rhythm in Carnatic practice is organized around the *tala*, a cyclical metric framework comprising a fixed number of beats. [2] The first beat of the cycle, called *sam*, serves as the primary point of resolution; phrases are structured to align with sam. Each beat in the tala is subdivided into equal pulse units according to a *gati*: *tisra* (3), *chatusra* (4), *khanda* (5), *misra* (7), or *sankīrna* (9). These subdivisions define the *matra* level, the finest regular grid on which rhythmic patterns are realized. A separate parameter, the jathi, governs how matras are grouped by accent. While gati determines subdivision *within* beats, jathi determines how those subdivisions are *grouped*. The two parameters are independent: the same 20-matra grid can receive accent groupings of $(4, 4, 4, 4, 4)$ (aligned with beats), $(5, 5, 5, 5)$ (cross-cutting beats), or $(1, 7, 3, 4, 3, 2)$ (irregular).

This independence is the crux of the formal problem. A rhythm tree encodes the cycle, beat, and matra hierarchy naturally: the root represents the cycle, depth-1 nodes represent beats, and depth-2 nodes represent matras. But an accent pattern like $(1, 7, 3, 4, 3, 2)$ partitions the 20 matras into spans that correspond to no subtree. The technique of jathi bhedam exploits exactly this: irregular accent sequences are superimposed on a uniform gati grid, creating the impression of continuous meter change while the underlying pulse remains fixed. Each accent span is performed with "downbeat feeling" [8, 9], yet the spans cross-cut beat boundaries. The rhythm tree stays fixed; only the grouping varies.

[2] This account is necessarily simplified. In practice, tala also encodes structural constraints that govern how and when surface-level rhythmic detail may develop. Since these issues are not central to the present discussion, we set them aside here.

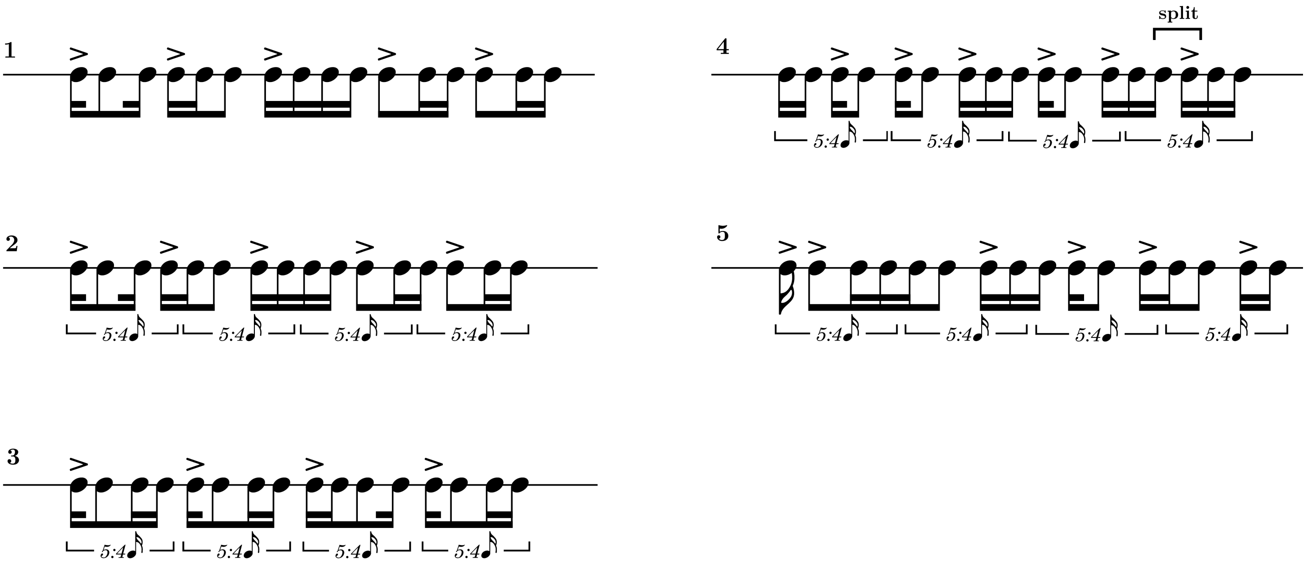


Figure 1. Development of a rhythmic motif through rhythmic sangati. (The beaming follows the convention used by Reina for Carnatic music in Western notation [8, 9]. (1) The original motif in chatusra (four subdivisions per beat), with accents every four matras. (2) The motif rephrased in khanda (five subdivisions per beat) while preserving the accent every four matras; the original gati becomes the jathi, creating the effect of a local metric modulation. (3) The motif takes on the accent pattern of the new gati, with accents now falling every five matras, leading to motivic development. (4) Both gati and jathi are changed simultaneously. When the original phrase does not fit evenly within the new grouping, notes may be split (as marked) to align with the new jathi accent pattern, with any remainder phrased as an anacrusis to ensure the sangati coincides with tala sam. (5) The motif is rephrased as jathi bhedam, applying an irregular accent pattern that does not coincide with the khanda gati.

What makes Carnatic rhythm particularly revealing for our purposes is that these accent spans are not merely perceptual effects; they are first-class compositional objects. Spans can be reordered, fragmented, and treated as units of subdividable time, and a single rhythmic phrase can be re-phrased across multiple gati environments. This is the technique of rhythmic sangati: a phrase constructed in one gati is adapted to another, with the relationship between original and target accent structures varying according to compositional intent. In Figure 1, a phrase in chatusra (four subdivisions per beat) becomes a phrase in khanda (five subdivisions per beat). The phrase spans different numbers of beats in each gati, but it is recognized as the "same rhythm"—similar in feeling to metric modulation in Western music.

Reina distinguishes four approaches to such transformations: 1, the original gati becomes the jathi in the new context, preserving the accent pattern exactly but at a different tempo relationship; 2, the note-value sequence is preserved but rebeamed into the new gati's grouping, imposing a fresh accent structure; 3, both gati and jathi change, requiring reconciliation when the new grouping does not evenly divide the phrase length; 4, the source phrase is itself constructed using jathi bhedam, and its irregular accent structure is preserved through the gati transformation, yielding jathi bhedam phrasing in the target gati. Each approach implies a different formal relationship between source and target accent structures, from identity, to replacement, to constrained compatibility, to the preservation of already-irregular groupings across metric contexts.

Rhythmic sangati poses a precise combinatorial question: given two accent structures over the same number of pulses, which onset patterns are compatible with both? In Section 4 we introduce *accent trees*, which formalize this interplay between rhythmic content and accentual grouping. The model addresses sangati and related techniques in two complementary ways: First, an *attachment model* allows rhythmic content to be time-scaled into accent spans of varying sizes, capturing the tempo-fluctuation effects that arise when the same phrase is stretched or compressed across different gati environments. This also provides a natural account of gati bhedam: a tala cycle can be progressively elaborated by attaching rhythm trees into accent spans, increasing rhythmic depth while the accent skeleton remains fixed. Second, we observe that accent compatibility reduces to a simple combinatorial condition: a rhythm is compatible with both accent structures exactly when its onset positions contain all required accent heads. We encode this as a bitmask and enumerate the compatible rhythms by iterating over subsets of the free positions. Together, these mechanisms yield not a single sangati but a structured compositional space: the complete set of rhythms satisfying the sangati constraint, each available for attachment into the accent spans of either context.

4. ACCENT TREES

We work with two distinct structural layers over rhythmic time: a durational layer, represented by rhythm trees, and an accent layer, represented by accent trees. The durational

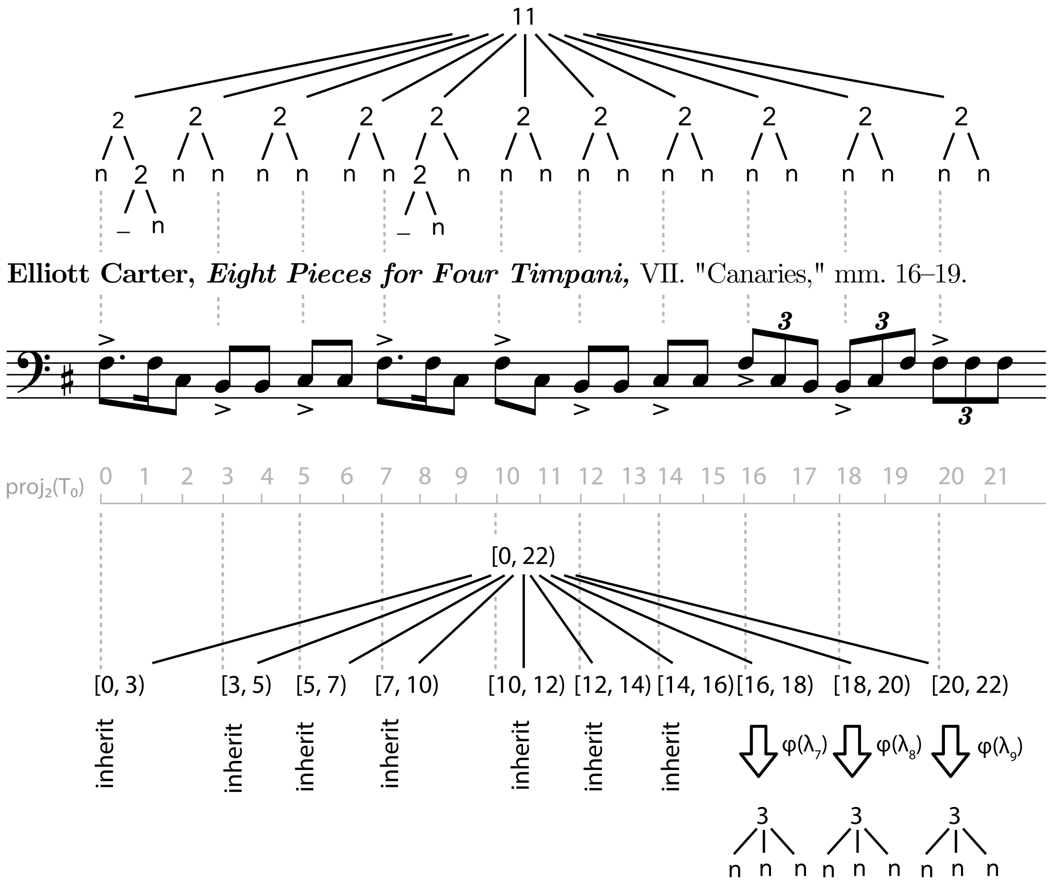


Figure 2. Elliott Carter, *Eight Pieces for Four Timpani* [10], VII. "Canaries," mm. 16–19, reconstructed as a single-level accent product $A \triangleright T_0$ at depth 2. Top: the base rhythm tree $T_0 = 11\big(2(n,n),\ldots,2(n,n)\big)$, with dotted figures at depth 3 in beats 1 and 4. Middle: the score, with positions $\nu_0,\ldots,\nu_{21}$ of $\mathrm{proj}_2(T_0)$ indexed below the staff. Bottom: the accent tree, whose root $[0, 22)$ partitions the projection into spans $(3, 2, 2, 3, 2, 2, 2, 2, 2, 2)$. Where spans cross-cut beat boundaries at depth 1, no span corresponds to any subtree of T_0. The projection presents each position as an undifferentiated index; rhythmic detail below depth 2 is invisible to the accent tree and passes through inheritance unchanged via (22). Spans λ_7–λ_9 replace inherited content with $\phi(\lambda) = 3(n, n, n)$, producing triplets within the unchanged base grid. Figure 3 extends this construction to a later passage in the same movement, where recursive application of the accent tree captures a metric modulation from tisra to chatusra jathi 5.

Elliott Carter, ***Eight Pieces for Four Timpani,*** VII. "Canaries," mm. 43, 45–48.

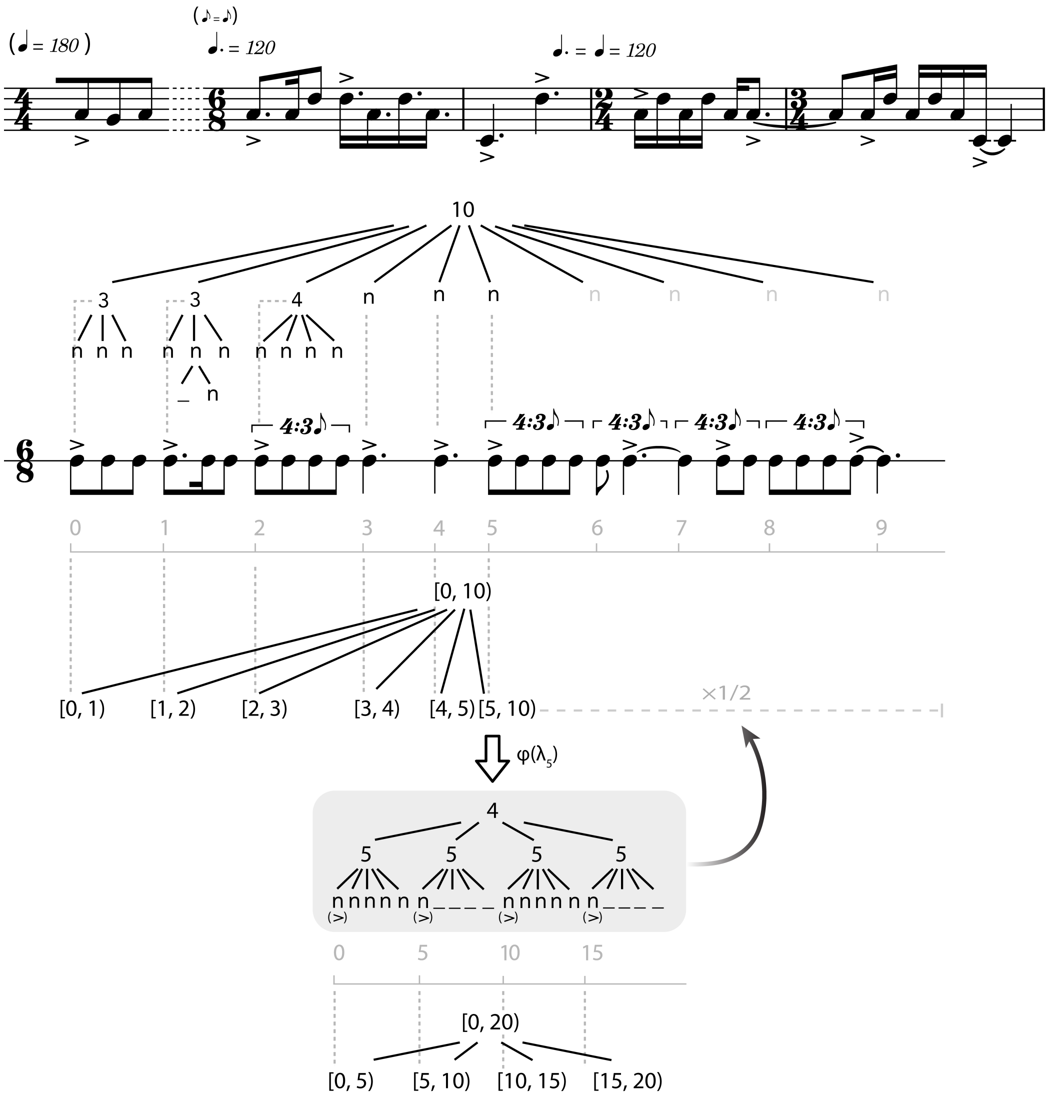


Figure 3. Recursive accent-tree reconstruction of Elliott Carter, *Eight Pieces for Four Timpani* [10], VII. "Canaries," mm. 43, 45–48. Read through a Carnatic lens, the passage begins in tisra gati at second speed. A single beat of chatusra subdivision appears at the second beat following the 6/8 bar line, after which two unsubdivided beats introduce metric ambiguity. The metric modulation ♩. = ♩ resolves this ambiguity into *chatusra jathi 5*, a regrouping of the pulse into spans of five within a fourfold subdivision. The reconstruction (middle) normalizes the passage into a single 6/8 span of ten beats. The accent tree $[0, 10)$ imposes an asymmetric partition whose span $[5, 10)$ receives an attached rhythm tree $\varphi(\lambda_5)$ comprising four groups of five matras, time-scaled by $\times\frac{1}{2}$. A second accent tree $[0, 20)$ partitions the frontier of $\varphi(\lambda_5)$ into four equal spans (bottom). This example illustrates a broader point about the data structure: because the accent tree with its attachment function can represent asymmetric partitioning, recursive elaboration, and time-scaling within a single hierarchical object, metric modulation need not be treated as an external event that interrupts one metric regime and installs another. Instead, it arises internally as a consequence of how accent spans at successive levels relate to one another.

layer specifies how a span of time is recursively subdivided into notated units (notes and ties). The accent layer specifies how those units are grouped into spans that carry downbeat feeling or other forms of rhythmic salience. This mirrors the syntax/prosody split in speech: a syntactic tree organizes words, while a prosodic tree groups them into higher-level units such as prosodic words and phrases. Here the rhythm tree plays the role of syntax, and the accent tree plays the role of prosody.

An attachment model interprets an accent tree over a base rhythm tree by operating on labeled duration sequences (2). Figure 2 illustrates a single-level accent product in which an accent tree partitions a base rhythm tree's depth-2 projection into spans that cross-cut beat boundaries, with some spans inheriting the base rhythm and others receiving attached content. The output is a new labeled duration sequence, which can then be mapped back to a notational rhythm tree by a reconstruction procedure in the style of Ycart et al. [11].

4.1 Depth Projection and Repleteness

Analytical studies of West African drum ensemble music commonly assume a fastest pulse—variously called the "density referent" or "equal-pulse base"—that serves as an orientation grid from which rhythms are composed, independent of surface articulation [12]. Carnatic rhythmic theory makes a similar commitment: accent groupings are defined over matras, the pulse positions of the subdivision grid, not over note onsets.

For this reason, the accent layer operates over a projection of the rhythm tree at a chosen depth, rather than directly over frontier positions. This design reflects a key feature of Carnatic rhythmic theory: accent groupings are defined over matras, the pulse positions of the subdivision grid, not over note onsets. A jathi pattern $(3, 4, 5, 3, 5)$ partitions twenty matras into five spans. These spans are counted in matras regardless of how notes articulate within them. Of course, for an accent to be audible, the first matra of each span must carry an onset. This is why our attachment model requires attached rhythms to begin with an onset (Section 4.2), and why the sangati derivation in Section 4.5 constrains onset positions to include all accent heads. The grouping structure itself, however, is defined at the matra level, independent of surface articulation.

This is why we require the base rhythm tree to be *replete* at the projection depth. Repleteness ensures that every matra exists as an addressable node in the tree, so that accent spans can be measured correctly even when the musical surface contains long notes spanning multiple pulses. Without repleteness, a half-note occupying four matras would appear as a single leaf, and the accent tree would have no way to count those four positions toward span length.

Recall that the depth of a node is its distance from the root, with the root at depth 0. A rhythm tree T is *replete at depth* d if every leaf has depth at least d. Any rhythm tree can be made replete by applying the rewrite rules of Jacquemard et al. [6]. A leaf at insufficient depth is replaced by a subtree of tied notes that preserves the original duration while exposing each matra as a separate node (Figure 4). Since the rewrite system is sound, this transformation preserves rhythmic equivalence. The sounding result is unchanged, but the tree now supports matra-level accent assignment.

Let T be a rhythm tree replete at depth d. The *d-projection* of T is the sequence

$$\mathrm{proj}_d(T) = (\nu_0, \ldots, \nu_{k-1}) \tag{3}$$

of nodes at depth exactly d, collected in left-to-right order. Each node ν_i roots a subtree $T[\nu_i]$ covering a contiguous block of the frontier. The blocks for $i = 0, \ldots, k-1$ partition $\mathrm{fr}(T)$.

We identify the projection with its *index set*

$$I_d(T) = \{0, 1, \ldots, k-1\}, \tag{4}$$

where index i refers to the subtree $T[\nu_i]$.

Different choices of d pick out different levels of metrical structure: a full tala cycle ($d = 0$), beats within a tala ($d = 1$), subdivision pulses ($d = 2$), and so on.

We associate the depth-d projection (3) with a *projected duration sequence*, which records the durational content carried by each projected node. Specifically, the d-projected duration sequence of T is

$$\mathrm{pds}_d(T) = \big((T[\nu_0], D_0), \ldots, (T[\nu_{k-1}], D_{k-1})\big), \tag{5}$$

where $T[\nu_i]$ is the subtree rooted at ν_i and D_i is the total duration of its frontier.

Each pair $(T[\nu_i], D_i)$ is a metrically meaningful unit at the chosen level. Together the projected subtrees cover the entire cycle.

4.2 The Attachment Model

Accent trees partition the index set $I_d(T_0)$ (4) of projection positions into grouped spans. Each node of the accent tree represents one span, and $\mathrm{slot}(\alpha)$ records its length (in number of projection positions).

We write $A \triangleright T_0$ to denote the accent action of A on T_0 and call the result an *accent product*. The action is mediated by the depth projection: A partitions $I_d(T_0)$, and the attachment function ϕ specifies the content of each span.

Formally, an accent tree at depth d for a rhythm tree T_0 is a rooted, ordered tree A with a function

$$\mathrm{slot} : \mathrm{nodes}(A) \to \{1, 2, \ldots\}. \tag{6}$$

We require:

1. **Root span.** The root covers the entire projection:

$$\mathrm{slot}(\mathrm{root}(A)) = k. \tag{7}$$

2. **Tiling.** If an internal node α has ordered children $\beta_1, \ldots, \beta_q$, then the children tile (partition) α's span:

$$\sum_{i=1}^{q} \mathrm{slot}(\beta_i) = \mathrm{slot}(\alpha). \tag{8}$$

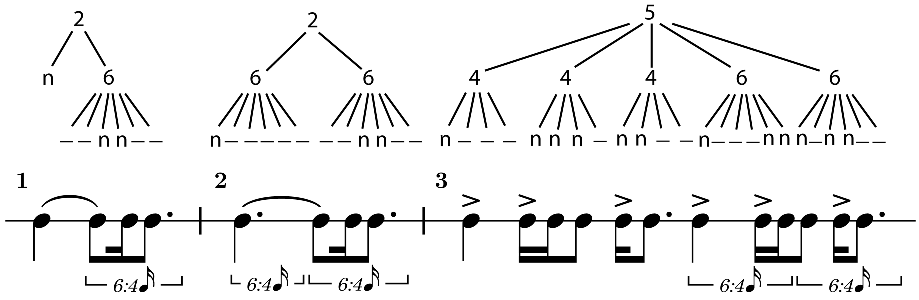


Figure 4. Repleteness at depth 2. (1) The tree $2(n, 6(\_, \_, n, n, \_, \_))$ is non-replete at depth 2. The first child has a terminal at depth 1, so the matras it spans at depth 2 are not individually addressable. (2) The same rhythm rewritten as a tree replete at depth 2: $2(6(n, \_, \_, \_, \_, \_), 6(n, \_, n, n, \_, \_))$. The first note is now represented as equal subdivisions at depth 2, exposing each matra as a separate node. Both trees produce the same labeled duration sequence; only the internal structure differs. (3) A rhythmic sangati: the accent cycle $(4, 4, 4)$ partitions twelve matras while internal articulation varies across beats. Repleteness makes this partition well-defined. Groupings are counted in matras, even though audible accents require onsets.

Condition (2) is the *tiling property*: children's slots partition exactly the parent's slot, in left-to-right order. Thus the leaves of A induce a partition of $I_d(T_0)$ into contiguous blocks.

The only constraint on this partition is that the slot values sum to k; individual spans may be of any length from 1 to k. Since an attached tree filling a span is itself a rhythm tree, it admits its own depth projection and can serve as the base for a further accent product (Section 4.4). Arbitrary span length is thus the mechanism that enables recursive elaboration: a large span can host a complex attached tree, which can in turn be repartitioned by its own accent tree, to any depth.

Since the base tree is replete at depth d, each projected position ν_i has duration D_i. When all projected positions have equal duration (for example, at the beat level, or at the matra level under uniform gati), this simplifies to

$$D_{\text{proj}} = \frac{1}{k}, \tag{9}$$

and the span duration of any accent node α is

$$D_\alpha = \text{slot}(\alpha)\, D_{\text{proj}} = \frac{\text{slot}(\alpha)}{k}. \tag{10}$$

When gati varies across beats, matra durations differ and the general form

$$D_\lambda = \sum_{i=a}^{b-1} D_i \tag{11}$$

must be used for a leaf λ with global interval $[a, b)$.

Let α be an internal node with children $\beta_1, \ldots, \beta_q$. The local interval of β_i relative to α is

$$\text{loc}(\beta_i \mid \alpha) = [s_i,\ s_i + \text{slot}(\beta_i)), \tag{12}$$

where $s_1 = 0$ and

$$s_{i+1} = s_i + \text{slot}(\beta_i). \tag{13}$$

Global intervals are defined compositionally: the root has interval $[0, k)$. If α has global interval $[a, b)$ and β_i has local interval $[s_i,\ s_i + \text{slot}(\beta_i))$, then

$$\text{glob}(\beta_i) = [\, a + s_i,\ a + s_i + \text{slot}(\beta_i)\,). \tag{14}$$

The global interval of a leaf λ specifies exactly which indices of $I_d(T_0)$—hence which projected subtrees—fall under that accent span.

The accent-tree formalism defines grouping at a chosen projection depth, typically the beat level, while the rhythmic content that fills each span is measured at a finer level, typically the matra. This separation requires reasoning simultaneously about two quantities: the *slot*, which counts how many projection-level units a span covers, and the *capacity*, which counts how many matras that span contains, given the gati of each covered beat.

When the accent tree operates at depth d, the number of matras within a span depends on the arity of each covered beat. For a leaf λ with global interval $[a, b)$, we define the capacity as

$$\text{cap}(\lambda) = \sum_{i=a}^{b-1} \text{arity}(\nu_i), \tag{15}$$

where ν_i is the ith node in the depth-d projection and $\text{arity}(\nu_i)$ is the subdivision count of the corresponding beat.

This two-level bookkeeping is a direct consequence of the attachment model. An attached rhythm tree is time-scaled to fill a span whose duration is determined by slot. Two spans with the same slot duration but different capacities—for example, five tisra beats versus five khanda beats—have the same absolute duration but different numbers of matras, so the same attached rhythm will align differently with the underlying pulse grid.

To connect these abstract accent positions to concrete surface events, we define an anchoring map from each leaf-span to the realized leaf event that carries its accent head. Fix an accent attachment model (T_0, d, A, ϕ). For each leaf $\lambda \in \text{leaves}(A)$ with global interval

$$\text{glob}(\lambda) = [a, b), \tag{16}$$

define its anchored accent head $\text{head}(\lambda)$ as the leftmost realized leaf event contributed by λ. If $\phi(\lambda)$ is undefined (inheritance), then $\text{head}(\lambda)$ is the first leaf of $\text{lds}(T_0[\nu_a])$. If $\phi(\lambda) = t$ (attachment), then t is time-scaled to fill the span duration D_λ: each leaf of t retains its proportional share of duration, but the absolute durations are multiplied by D_λ.

An accent tree specifies *where* accent spans fall; an *attachment function* specifies *what* fills them. Formally, let

$$\mathcal{R} = \{\, t \in \mathrm{RhythmTree} : \ell_0(t) = n \,\} \tag{17}$$

be the set of rhythm trees whose leftmost leaf is an onset. An *attachment function* for an accent tree A is a partial map

$$\phi : \mathrm{leaves}(A) \rightharpoonup \mathcal{R}. \tag{18}$$

Requiring attached rhythms to begin with an onset ensures that each accent span carries an audible attack at its head, which is the defining property of an accented position.

An *accent attachment model* is a tuple (T_0, d, A, ϕ) consisting of a base rhythm tree T_0 replete at depth d, an accent tree A at that depth, and an attachment function ϕ. For each leaf $\lambda \in \mathrm{leaves}(A)$, let

$$[a, b) = \mathrm{glob}(\lambda) \tag{19}$$

be its global interval, so that λ claims projection positions

$$a, a+1, \ldots, b-1. \tag{20}$$

The span duration of λ is

$$D_\lambda = \sum_{i=a}^{b-1} D_i, \tag{21}$$

where D_i is the duration of the ith projected subtree.

For a leaf λ with global interval $[a, b)$, we define its realized sequence σ_λ by distinguishing two cases, depending on whether λ inherits material from the base tree or receives an attached rhythm.

If $\phi(\lambda)$ is undefined (inheritance), then σ_λ is the concatenation of the labeled duration sequences (2) of the projected subtrees covered by λ:

$$\sigma_\lambda = \mathrm{lds}(T_0[\nu_a]) \cdot \mathrm{lds}(T_0[\nu_{a+1}]) \cdots \mathrm{lds}(T_0[\nu_{b-1}]). \tag{22}$$

If $\phi(\lambda) = t$ (attachment), then t is time-scaled to fill the span of duration D_λ.

Inheritance preserves the durations already present in the base tree, while attachment rescales the attached rhythm to exactly fill the accent span.

These two cases—inheritance and attachment—correspond to distinct compositional operations. Inheritance regroups existing material under a new accent structure; attachment fills accent spans with new content, time-scaled to fit. Carnatic practice provides particularly clear examples of both.

Inheritance is the default case: the accent tree partitions existing rhythmic material without replacing it. The durations in the base tree remain unchanged; only the grouping varies. A phrase can be partitioned into accent spans of any length—aligned with the subdivision grid or cross-cutting it—without altering its durational content. The accent tree imposes a grouping structure; inheritance copies the underlying rhythm into each span.

Attachment handles cases where the accent span itself becomes a frame for subdivision—where new rhythmic content is inserted and time-scaled to fill the span. Attachment is not mere concatenation. Sequential rhythm trees are independent durational structures joined end-to-end; an attached rhythm is scaled to fill a span whose boundaries are defined by a separate accent hierarchy. In a single-layer model, rhythmic structure is purely durational: grouping is a byproduct of subdivision. The two-layer model treats accent structure as a parallel hierarchy with its own compositional logic. Accent spans can be subdivided, reordered, nested, and transformed independently of the rhythms that fill them. But the two layers share an interface: accent heads must be articulable, so they must coincide with onsets. This coupling means that transformations in either layer propagate constraints to the other. Elaborating an accent structure may force rhythmic adaptation; imposing a new rhythm may render certain accent patterns unsupportable.

Carnatic practice foregrounds this interaction. In the technique of *mixed jathi nadai bhedam* [8, 9], two accent groupings alternate to tile a cycle exactly—say, spans of 3 and 7 matras repeating twice to fill a span of 20 matras. Each span can then be subdivided independently: the 7-span might be split into three, while the 3-span is split into four, yielding the impression of two alternating tempos (Figure 5). The accent spans are first-class compositional units, subject to motivic development through subdivision, augmentation, and transformation. The rhythms that fill them must articulate each span head, but are otherwise free to vary. This is the long-distance dependency that context-free grammars cannot express: the accent structure imposes constraints across the entire cycle, and any surface rhythm must satisfy all of them simultaneously. The attachment model represents this directly—each span hosts its own attached rhythm tree, time-scaled to fit, while the accent skeleton remains fixed.

The accent tree separates *where* accent spans fall from *what* fills them. Inheritance and attachment are the two ways content can relate to that scaffold: either the existing rhythmic surface is regrouped (jathi bhedam), or new content is inserted and scaled to fit. Both operations preserve the matra grid; both produce labeled duration sequences that compile back to rhythm trees.

When subdivision varies across spans, as in mixed jathi nadai bhedam, the accent tree becomes the structural bridge between subdivision contexts that share a temporal framework but not an internal pulse. Carnatic practice pushes this independence to its perceptual limit: adjacent spans in mixed *jathi nadai bhedam* are performed at different effective tempos, yet the tala cycle and its beat grid remain fixed. The effect approaches polymeter, but the structural commitment is to a shared grid with accentual repartitioning on top. The two-layer model captures this: the attachment function imposes no constraint that attached trees share a common arity, but the projection that anchors the accent tree remains a well-defined level of the base tree. Whether other repertoires require the stronger reading, namely the accent skeleton as the sole structural link between incommensurable subdivisions, is a question beyond the scope of this paper.

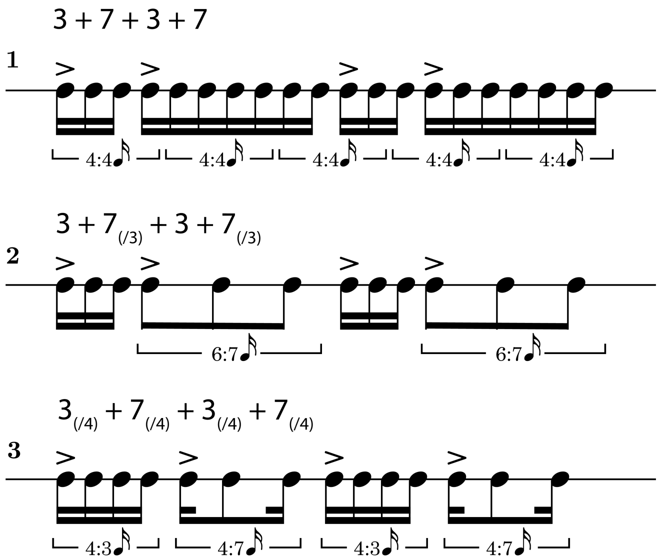


Figure 5. *Mixed jathi nadai bhedam* is a Carnatic technique in which two accent groupings alternate to tile a cycle exactly, with each span subject to independent subdivision. The technique illustrates how accent spans function as first-class compositional units under the attachment model. Here, spans of 3 and 7 matras alternate: (1) the basic alternation; (2) the 7-span subdivided into three; (3) both spans subdivided into four. The result is something perceptually like metric modulation, yet the accent skeleton remains fixed throughout—the accent tree defines the span boundaries, and each span hosts an independently elaborated rhythm, time-scaled to fit.

4.3 Compilation and Reconstruction

The accent attachment model associates a realized sequence σ_λ to each leaf of the accent tree. Let $A \triangleright T_0$ be an accent product, and let $\lambda_1, \ldots, \lambda_m$ be the leaves of A in left-to-right order. The compiled output is the concatenation

$$\sigma_{\text{out}} = \sigma_{\lambda_1} \cdot \sigma_{\lambda_2} \cdots \sigma_{\lambda_m}. \tag{23}$$

By the tiling property (8), the leaves' global intervals partition $[0, k)$, so

$$\text{dur}(\sigma_{\text{out}}) = \text{dur}(\text{lds}(T_0)), \tag{24}$$

that is, attachments redistribute events within each span but preserve total duration. The accent layer thus specifies *what* to play and *when*; it remains to find a rhythm tree, a notational object, that realizes this sequence.

The compiled output σ_{out} is a labeled duration sequence; to obtain conventional notation, one must find a rhythm tree t with $\text{lds}(t) = \sigma_{\text{out}}$. We adopt the k-best enumeration algorithm of Ycart et al. [11], which uses dynamic programming over a subdivision schema to produce candidate trees in order of increasing notational complexity. Since σ_{out} already specifies exact onset times, the algorithm reduces to selecting the simplest tree consistent with the compiled sequence.

4.4 Recursive Accent Products

Since an attached rhythm tree is itself a rhythm tree, it can serve as the base for its own accent product. An attachment $\varphi^0(\lambda) = T_1$ may host a further action $A^1 \triangleright T_1$, elaborating the content within λ's span. This nesting requires no additional machinery: the same attachment model applies at each level, compiling bottom-up. Complex rhythmic structures emerge as a byproduct of recursive attachment rather than as special cases requiring dedicated treatment. Figure 3 illustrates this with mm. 43, 45–48 of Carter's "Canaries" [10], where a passage conventionally described as metric modulation—or, in Carnatic terms, a shift from tisra gati to chatusra jathi 5—arises naturally from nested accent trees at successive levels of attachment.

A second mode, *iterative recursion*, compiles $A^0 \triangleright T_0$ to a rhythm tree T_0' and applies a new action $A^1 \triangleright T_0'$ to the result. Since compilation flattens span boundaries into a labeled duration sequence, A^1 may impose groupings that cross-cut the spans induced by A^0.

4.5 Derivation of Rhythmic Sangati from Figure 6

The attachment model provides one account of sangati: a phrase constructed in one gati can be attached into accent spans defined by another, with time-scaling handling the tempo relationship. But sangatis also pose a complementary question: given two accent structures over the same span, which onset patterns are compatible with both? This is simply a matter of finding rhythms whose onsets include all required accent heads from both structures. The formalism reduces this to a simple combinatorial problem, and the solution defines a structured compositional space rather than a single rhythm.

Figure 6 shows a single tala cycle partitioned into two gati regions: five beats in tisra (three pulses per beat) and three beats in khanda (five pulses per beat). Through attachment, each span takes on a capacity of fifteen matras ($5 \times 3 = 3 \times 5 = 15$), but organizes them differently. One way to construct a rhythmic sangati is to find an onset pattern that works in both metric environments, aligning with beat

$\mathbf{T_0}$ replete at $\boldsymbol{d}$=1

$$T_0 = 8\big(\underbrace{3(\ldots),3(\ldots),3(\ldots),3(\ldots),3(\ldots),}_{\lambda_1 \text{tisra}}\ \underbrace{5(\ldots),5(\ldots),5(\ldots)}_{\lambda_2 \text{khanda}}\big)$$

$$\mathrm{slot}(\lambda_1) = 5, \quad \mathrm{glob}(\lambda_1) = [0, 5)$$

$$\mathrm{slot}(\lambda_2) = 3, \quad \mathrm{glob}(\lambda_2) = [5, 8)$$

Attachment

$$\phi(\lambda_1) = T_1 = 5\big(3(n,n,n),\ 3(n,n,n),\ 3(n,n,n),\ 3(n,n,n),\ 3(n,n,n)\big)$$

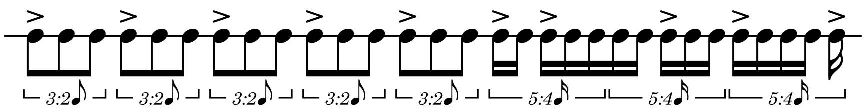
$$\phi(\lambda_2) = T_2 = 3\big(5(n,n,n,n,n)\ \ 5(n,n,n,n,n)\ \ 5(n,n,n,n,n)\big)$$

$$\mathrm{cap}(\lambda_1) = 5 \times 3 = 15\ =\ 3 \times 5 = \mathrm{cap}(\lambda_2)$$

Recursive accent trees

$$A^1 \text{ on } \lambda_1\!:\ \ 3{+}3{+}3{+}3{+}3 \quad \Longrightarrow \quad \mathrm{ons}(A^1) = \{0, 3, 6, 9, 12\}$$

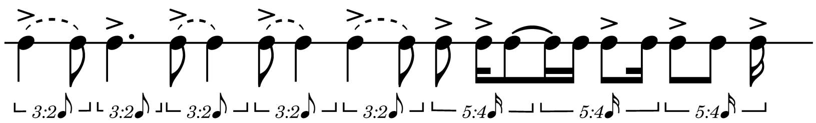
$$A^2 \text{ on } \lambda_2\!:\ \ 2{+}5{+}3{+}4{+}1 \quad \Longrightarrow \quad \mathrm{ons}(A^2) = \{0, 2, 7, 10, 14\}$$



Union of accent masks

$$U = \mathrm{ons}(A^{\text{tisra}}) \cup \mathrm{ons}(A^{\text{khanda}}) = \{0, 2, 3, 6, 7, 9, 10, 12, 14\}$$

Examples of compatible onset patterns



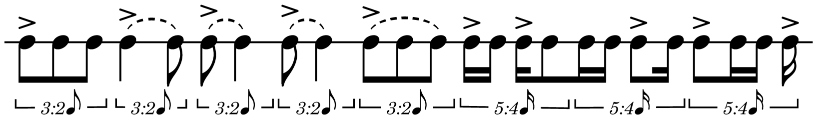


Figure 6. Derivation of a sangati-compatible rhythm space. An accent tree at depth 1 partitions eight beats into spans of five (tisra, slot $= 5$) and three (khanda, slot $= 3$). Through attachment, λ_1 receives five triplets and λ_2 three quintuplets, giving both spans equal capacity: $5 \times 3 = 3 \times 5 = 15$ matras. Recursive accent trees then act on the attached trees: regular gati accents every three matras on T_1, yielding onset set $\{0, 3, 6, 9, 12\}$, and an irregular jathi $(2, 5, 3, 4, 1)$ on T_2, yielding $\{0, 2, 7, 10, 14\}$. Their union determines nine required accent positions; the six free positions yield $2^6 = 64$ compatible rhythms. Two examples are shown.

Added onsets	$(\ell_0, \ell_1, \ldots, \ell_{14})$	total
$\emptyset$	$(n, \_, n, n, \_, \_, n, n, \_, n, n, \_, n, \_, n)$	9
$\{1, 4\}$	$(n, n, n, n, n, \_, n, n, \_, n, n, \_, n, \_, n)$	11
$\{4, 8, 11, 13\}$	$(n, \_, n, n, n, \_, n, n, n, n, n, n, n, n, n)$	13
F	$(n, n, n, n, n, n, n, n, n, n, n, n, n, n, n)$	15

Table 1. Four rhythms from the 64-pattern sangati space. All share the nine required onsets; they differ only in which free positions receive additional onsets.

accents in the first half and with the irregular jathi pattern $(2, 5, 3, 4, 1)$ in the second.

We represent onset patterns as binary words of length n, where bit i is 1 if position i carries an onset and 0 if it carries a tie. The accent trees for the two halves induce onset sets:

$$\begin{aligned} \mathrm{ons}(A_{\mathrm{tisra}}) &= \{0, 3, 6, 9, 12\}, \\ \mathrm{ons}(A_{\mathrm{khanda}}) &= \{0, 2, 7, 10, 14\}. \end{aligned} \tag{25}$$

A rhythm compatible with both must have onsets at every position in their union:

$$U = \{0, 2, 3, 6, 7, 9, 10, 12, 14\}. \tag{26}$$

We encode U as a bitmask M by setting bit i to 1 for each $i \in U$. Bits are indexed left-to-right by pulse position: bit i corresponds to position i, with bit 0 the leftmost bit.

$$M = (101100110110101)_2. \tag{27}$$

The positions where M has bit 0 are *free*, meaning each may independently carry an onset or a tie:

$$F = \{1, 4, 5, 8, 11, 13\}. \tag{28}$$

This yields $2^{|F|} = 64$ compatible rhythms that fit both accent spaces.

Enumerating this space is straightforward. We iterate a $|F|$-bit counter s from 0 to $2^{|F|} - 1$. For each s, we construct a rhythm by starting with the required mask M and inserting additional onsets at free positions according to the bits of s. The key operation is a bitwise OR: we shift a 1 into each free position selected by s and combine the result with M.

Table 1 shows four rhythms from this space. The sparsest (adding no free positions, $\emptyset$) places onsets only at required positions, while the densest (adding all free positions, F) fills every pulse. The 62 intermediate patterns offer a graduated range of rhythmic density, providing composers with a principled vocabulary of sangati-compatible phrases.

The enumeration runs in $O(n \cdot 2^{|F|})$ time, which is output-optimal. In Carnatic contexts this is tractable: projection lengths rarely exceed 20 to 30 matras, each accent structure places 5 to 10 heads, and their union typically covers half or more of the available positions, leaving $|F| \leq 10$ to 15 free positions. The musically interesting cases, where accent patterns create tension by overlapping only partially, are precisely those that shrink the solution space further.

When $|F|$ is larger, or when only a subset of rhythms is desired, the enumeration admits pruning. The iteration over subsets of F forms a binary decision tree; constraints on tie density or maximum run length can prune branches early. Alternatively, a best-first search ordered by a scoring function, such as proximity to a target density or resemblance to a seed pattern, could yield preferred rhythms without exhaustive enumeration.

5. CONCLUSION

A useful test of a formalism is the range of structural phenomena it can make explicit. Carnatic rhythmic theory supplies a concentrated set of such cases, but the formal problem is not Carnatic-specific. Reina [8] demonstrates a broader analytical reach for Carnatic rhythmic structures, applying it systematically across Western repertoire: jathi bhedam in Messiaen, Stravinsky, Ligeti, and Maxwell Davies; independent gati layering in Carter and Donatoni; nadai bhedam and anuloma-pratiloma structures in Xenakis, Castiglione, and Ferneyhough. In each case, the rhythmic interest lies not in the durational structure or the accent structure alone, but in the tension between them—a tension that no single-tree formalism over the subdivision hierarchy can structurally represent. The duration–accent interface proposed here provides a compact representation for these phenomena while remaining compatible with existing rhythm-tree grammars and their associated rewrite and enumeration algorithms.

Future work falls into two main directions. First, accent trees themselves can be placed under an explicit generative grammar, for example by adapting Tree-Adjoining Grammar to treat accent cycles as elementary trees and to model recursive elaboration, nesting, and span-level transformations across metrical contexts. Second, the present model suggests quantitative alignment constraints that measure systematic match or mismatch between accent boundaries and durational constituents at multiple depths, yielding a computable notion of rhythmic tension analogous to prosodic markedness. Wells [13, 14] develops a complementary intervallic approach to Carnatic rhythm, using Lewin's generalized interval systems to model the trajectory of melodic phrases through the tala cycle during trikala performance. While his Met intervals track dynamic displacement—how phrase boundaries migrate across beat positions under speed transformation—accent trees capture the hierarchical grouping structure that listeners recognize as preserved across such transformations. Integrating these perspectives could yield a richer account of sangati, combining Wells's precise tracking of metric tension with the structural characterization proposed here.

6. REFERENCES

[1] M. Yeston, *The Stratification of Musical Rhythm*. New Haven: Yale University Press, 1976.

[2] M. Rohrmeier, "Towards a formalization of musical rhythm," in *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR)*, Montréal, Canada, 2020.

[3] E. O. Selkirk, "The syntax-phonology interface," in *The Handbook of Phonological Theory*, 2nd ed., J. Goldsmith, J. Riggle, and A. C. L. Yu, Eds. Oxford: Blackwell, 2011.

[4] J. Tepperman and S. Narayanan, "Tree grammars as models of prosodic structure," in *Proceedings of Interspeech*, 2008.

[5] F. Jacquemard, A. Ycart, and M. Sakai, "Generating equivalent rhythmic notations based on rhythm tree languages," in *Proceedings of the Third International Conference on Technologies for Music Notation and Representation (TENOR'17)*, A Coruña, Spain, 2017, pp. 145–153.

[6] F. Jacquemard, P. Donat-Bouillud, and J. Bresson, "A structural theory of rhythm notation based on tree representations and term rewriting," in *Mathematics and Computation in Music: 5th International Conference, MCM 2015, London, UK, June 22–25, 2015, Proceedings*. Springer, 2015.

[7] C. Agon, K. Haddad, and G. Assayag, "Representation and rendering of rhythmic structures," in *Proceedings of the Second International Conference on Web Delivering of Music (WEDELMUSIC)*. Darmstadt, Germany: IEEE Computer Press, 2002, pp. 109–113.

[8] R. Reina, "Karnatic rhythmical structures as a source for new thinking in western music," PhD thesis, Middlesex University, 2016.

[9] ——, *Applying Karnatic Rhythmical Techniques to Western Music*. New York: Routledge, 2016.

[10] E. Carter, *Eight Pieces for Four Timpani*. New York: Associated Music Publishers, 1968.

[11] A. Ycart, F. Jacquemard, J. Bresson, and S. Staworko, "A supervised approach for rhythm transcription based on tree series enumeration," in *Proc. Sound and Music Computing Conference (SMC)*, Hamburg, Germany, 2016.

[12] R. Polak, "Rhythmic feel as meter: Non-isochronous beat subdivision in jembe music from Mali," *Music Theory Online*, vol. 16, no. 4, 2010. [Online]. Available: https://www.mtosmt.org/issues/mto.10.16.4/mto.10.16.4.polak.html

[13] R. L. Wells, "A generalized intervallic approach to metric conflict in Liszt," *Music Theory Online*, vol. 23, no. 4, 2017. [Online]. Available: https://mtosmt.org/issues/mto.17.23.4/mto.17.23.4.wells.html

[14] ——, "Theorizing trikāla: A generalized intervallic approach to pulse transformation in south indian carnatic music," *Analytical Approaches to World Music*, vol. 8, no. 1, pp. 48–86, 2020.