Title

Hierarchical Temporal Structure in Carnatic Improvisation: Interaction and Rhythmic Constraint

Permalink

Journal

None

Author

Nanjangud, Siddharth

Publication Date

2026-08-06

Peer reviewed

UNIVERSITY OF CALIFORNIA, MERCED

Hierarchical Temporal Structure in Carnatic Improvisation: Interaction and Rhythmic Constraint

A Capstone Project submitted in partial satisfaction of the requirements
for the degree of Master of Science
in
Cognitive and Information Sciences

by
Siddharth Nanjangud

Committee in charge:
Professor Christopher Kello, Chair
Professor Ramesh Balasubramaniam
Professor Tyler Marghetis

2026

The Capstone Project of Siddharth Nanjangud is approved, and it is acceptable in quality and form for publication electronically:

Tyler Marghetis

Ramesh Balasubramaniam

Christopher Kello

Chair

University of California, Merced

2026

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to everyone who supported me throughout this capstone project. First, I would like to thank my advisor and committee chair, Christopher Kello, for his guidance, encouragement, and thoughtful feedback throughout every stage of this project. His mentorship shaped the theoretical framing, analyses, and interpretation of this work, and I am deeply grateful for his support. I would also like to thank Ramesh Balasubramaniam and Tyler Marghetis for serving on my committee and for their careful feedback. Their questions and recommendations helped strengthen the statistical analyses, clarify the central argument, and improve the final capstone project. I am grateful to Butovens Médé for providing the independently assembled Indian classical music dataset used as convergent evidence in this study. I also thank the faculty, students, and staff of the Cognitive and Information Sciences program at the University of California, Merced, for creating an intellectually supportive community in which this work could develop. Finally, I would like to thank my family and friends for their patience, encouragement, and support throughout my graduate education and the completion of this capstone project.

ABSTRACT OF THE CAPSTONE PROJECT

Hierarchical Temporal Structure in Carnatic Improvisation: Interaction and Rhythmic Constraint

by

Siddharth Nanjangud

Master of Science in Cognitive and Information Sciences

University of California, Merced, 2026

Professor Christopher Kello, Chair

Carnatic improvisation provides a useful contrast between musical interaction and rhythmic constraint. Raga Alapana develops through flexible melodic coordination, whereas Tani Avartanam combines reciprocal improvisation with a persistent rhythmic cycle. This study examined whether those two interactive performance forms would exhibit similar hierarchical temporal structure or whether the rhythmic organization of Tani would produce a distinct pattern. Ten vocal Raga recordings and ten Jugalbandhi Tani recordings were converted into binary acoustic-event sequences and analyzed using Allan Factor analysis across 11 timescales. Their AF functions were compared with conversation, jazz improvisation, and beat-based music. An independent dataset of Carnatic and Hindustani vocal and instrumental performances was also analyzed as convergent evidence. Raga increased steadily across timescales and closely resembled conversation and jazz, whereas Tani showed stronger short-timescale clustering, a flatter intermediate region, and a pattern that closely resembles beat-based music. Statistical and classification analyses supported this distinction, and the independent Carnatic vocal recordings closely matched the present Raga dataset. These findings suggest that rhythmic organization can strongly shape temporal event structure, even when the underlying performance remains highly interactive.

Introduction

Human behavior unfolds across multiple interacting timescales, from milliseconds to minutes. Instead of occurring as isolated events, behaviors such as speech, music, and interaction reveal structured temporal organization across nested timescales. This organization is called hierarchical temporal structure (HTS), and it reflects how events cluster over time. It provides insight into the processes of complex and dynamic behavior. One method for quantifying HTS is Allan Factor (AF) analysis, which measures the clustering of events from small to larger temporal windows. When AF functions are plotted logarithmically, the scaling behaviors indicate long-range temporal organization (Kello et al., 2017). Kello et al. (2017) demonstrated that conversational speech, jazz improvisation, and rhythmically constrained music exhibit distinct hierarchical temporal structures despite containing similar scaling properties. Specifically, jazz improvisation aligned closely with conversational speech, meaning that systems characterized by flexible interaction and coordination organize behavior similarly across multiple timescales. Kello et al. proposed that hierarchical temporal structure emerges from interaction-dominant dynamics, where behavior is shaped by coordination among multiple interacting processes rather than only isolated events. This framework allows HTS to reflect the behavior and social processes underlying complex vocalizations and music.

South Indian classical music, known as Carnatic music, provides a compelling framework for examining these ideas because improvisation is fundamental to performance while allowing different forms of temporal organization. Two major improvisational forms within Carnatic music are Raga Alapana (Raga) and Tani Avartanam (Tani). Raga is primarily melodic and allows relatively flexible timing and phrase development (Kassebaum, 1987), while Tani is percussion-based and organizes around recurring rhythmic cycles known as Tala. Raga typically involves a vocalist or an instrumentalist who improvises while the accompanying musicians follow and respond to the evolving melody. Tani, although engaging in continuous interaction during a performance, is constrained within a strong rhythmic framework. In this study, only Jugalbandhi performances were analyzed, as this interaction-based Tani subtype features reciprocal exchanges between two or more primary percussionists while conforming to the rhythmic constraints characteristic of Tani. This provides the closest comparison to the interactive nature of Raga while preserving the fundamental distinction of beat-based organization.

Both Raga and Tani provide highly interactive forms of improvisation, but they have a clear distinction. Raga allows relatively unconstrained temporal development, but Tani requires performers to coordinate their improvisation within a constant rhythmic cycle. If interaction is an important organizing factor underlying hierarchical temporal structure, then both forms are expected to exhibit similar scaling and temporal organization because both rely on continuous coordination. However, the strong rhythmic organization that characterizes Tani introduces an alternate possibility. Rather than just contributing to the interaction, adhering to a constant beat may fundamentally alter the temporal organization. Will both improvisational forms exhibit interaction-driven structure observed in conversation and jazz, or will the strong rhythmic organization of Tani produce a pattern resembling beat-based music?

The present study addresses this question by applying Allan Factor analysis to Raga and Tani recordings and comparing their temporal organization with previously analyzed recordings of conversation, jazz, and beat-based music from Kello et al. (2017). To evaluate whether the observed Raga patterns generalized beyond the present dataset, an independently assembled collection of Carnatic and Hindustani vocal and instrumental recordings provided by Butovens Mede was analyzed as convergent evidence. Rather than serving as a direct replication, this secondary analysis was used to discern whether similar AF trajectories emerged across an independently collected corpus of Indian classical music. Based on the interaction-dominant framework proposed by Kello et al. (2017), Raga was expected to exhibit a similar hierarchical temporal structure to conversation and jazz because all three categories involve flexible and coordinated interactions. Tani involves continuous interaction among performers, so it is also expected to reveal a similar hierarchical temporal structure. However, the presence of a persistent rhythmic cycle was expected to modify this organization and potentially produce an AF that closely resembles beat-based music more than Raga. By comparing these alternate possibilities, this study evaluates how interaction and rhythmic organization combine and contribute to hierarchical temporal structure in musical improvisation.

Methods

2.1 Recordings

The present study analyzed twenty recordings of South Indian Carnatic improvisation consisting of ten Raga Alapana (Raga) performances and ten Jugalbandhi Tani Avartanam (Tani) performances. All recordings were obtained from publicly available YouTube videos and selected to represent complete and uninterrupted improvisational performances. Recordings were required to contain at least four minutes of analyzable audio following preprocessing. Musical introductions, announcements, applause, and other miscellaneous non-performance-related segments were removed prior to analysis to ensure that only the improvisational portions of each recording were analyzed. Raga recordings consisted exclusively of melodic improvisations from vocal Raga Alapana performances accompanied by supporting musicians. Vocal performances were selected to maintain consistency across the primary dataset and to minimize variability associated with differences in instrumentation. Tani recordings were limited to the Jugalbandhi style, in which two or more primary percussionists engage in reciprocal improvisation. This subtype was selected to maximize performer interaction while maintaining the rhythmic framework characteristic of Tani Avartanam. Restricting the dataset to vocal Raga and Jugalbandhi Tani allows reduced variability arising from differences in performance format while providing a direct comparison between two highly interactive forms of Carnatic improvisation. To place the present recordings within a broader behavioral context, previously analyzed datasets from Kello et al. (2017) were included for comparison. These consisted of conversational, interview-style speech from the Buckeye Corpus, Jazz trio-ensemble improvisation, and a beat-based musical category comprising rock, electronic dance music (EDM), rap, and metal recordings (Kello et al., 2017; Pitt et al., 2007). The four beat-based genres were combined into a single comparison category because they share a common emphasis on persistent rhythmic organization. This comparison provided a reference for evaluating whether the temporal organization of Tani more closely resembled interaction-dominant behaviors such as conversation and jazz versus music adhering to strong rhythmic constraints.

To assess whether the temporal organization observed in the present Raga recordings generalized beyond the primary dataset, an independently assembled collection of Indian classical recordings provided by Butovens Mede was analyzed as a source of convergent evidence. This dataset consisted of four categories: Carnatic vocal, Carnatic instrumental, Hindustani vocal, and Hindustani instrumental. As the present dataset consisted of exclusively vocal Raga performances, the Carnatic vocal recordings provided the most appropriate comparison to evaluate whether a similar hierarchical temporal structure emerged across independently assembled collections of Carnatic improvisation. The Butovens recordings were analyzed using the same preprocessing and Allan Factor analysis pipeline as the present dataset but are excluded from the primary five-category comparisons.

2.2 Audio Preprocessing and Event Extraction

To establish a common representation across all recordings, each audio file was processed using the AFaudio processing pipeline developed by Kello et al. (2017). Recordings were converted to a single audio channel and downsampled by a factor of four prior to the event extraction. The amplitude envelope of each recording was computed using the Hilbert transform, and local maxima were identified using a minimum peak separation of 2.5 ms. The candidate peaks were ranked according to amplitude, and the largest 0.5% of peaks were kept as events. This threshold reduced the background fluctuations while keeping the most salient acoustic events. The retained peaks were converted into a binary event raster where each sample was set to 1 for an event and 0 for a non-event. All recordings were analyzed using four-minute segments following preprocessing. Each recording being represented as a binary sequence of acoustic events allowed for a common point process representation across the recordings and allowed vocal, instrumental, percussion, speech, and beat-based audio to be analyzed using the same Allan Factor framework.

2.3 Allan Factor Analysis

The binary event sequences generated from each recording were analyzed using Allan Factor (AF) analysis to quantify hierarchical temporal structure across multiple timescales. For a given window size (T), the event sequence was divided into consecutive, non-overlapping windows, and the number of events within each window was calculated. AF measures the variability in event counts between adjacent windows relative to the mean number of events:

$$A(T) = \langle [N_{i+1}(T) - N_i(T)]^2 \rangle \;/\; 2\langle N_i(T) \rangle$$

where Ni(T) represents the number of events occurring in the i^{th} counting window of duration (T). Differences between adjacent windows were calculated and averaged over all values of (i) as indicated by the angle brackets. Larger AF values indicate greater differences in event counts between adjacent windows, providing greater clustering of events at that timescale (Kello et al., 2017).

Following Kello et al. (2017), Allan Factor values were calculated at 11 counting-window durations that corresponded to exponents 2^4 through 2^{14}, ranging from approximately 0.014 to 15.01 seconds. These calculations produced an 11-point AF function for each four-minute segment. When a recording contained multiple segments, the AF functions were averaged to produce one recording AF function. Category-level grand averages and standard errors were calculated across recordings at each window duration. AF functions were displayed on logarithmic window sizes and Allan Factor axes. To quantify the scaling, a linear fit was applied to the log-transformed AF values between 0.05 and 5.00 seconds. The slope of this fit was retained as the scaling exponent, α. To compare the overall shapes of the AF functions, a third-order fit was applied to each recording’s function in log-log axes:

$$\log_{10}A(T) = c_3[\log_{10}T]^3 + c_2[\log_{10}T]^2 + c_1 \log_{10}T + c_0$$

The quadratic coefficient c_2 and linear coefficient c_1 were used to represent each recording in the AF feature-space analysis. The quadratic coefficient was plotted on the X-axis and the linear coefficient on the Y-axis (Kello et al., 2017).

2.4 Statistical Analysis

Each recording was treated independently, and the recording-level scaling exponents were compared across conversation, jazz, beat-based music, Raga, and Tani using a one-way ANOVA. Tukey post hoc tests were used to evaluate pairwise differences between categories, and effect sizes were calculated for the overall model. Due to differences in sample sizes, a Brown-Forsythe (BF) test and Welch one-way ANOVA were also conducted to evaluate the robustness of the results. The Welch ANOVA compared all five categories, and the planned Welch independent-samples t-test compared Raga and Tani. The independent dataset procured by Butovens was analyzed separately using a two-by-two factorial linear model with tradition (Carnatic or Hindustani) and modality (vocal or instrumental) as predictors of the recording-level scaling exponent. The model tested the main effects of tradition and modality along with their interaction. A linear support vector machine (LSVM) classifier was used to determine whether the five primary categories could be distinguished from their AF functions. The 11 AF values from each recording were provided as predictors, and classification was evaluated using a repeated fivefold cross-validation. The performances were summarized using overall and balanced accuracy, category classification, and a confusion matrix. A final permutation test was used to determine whether the balanced accuracy exceeded the expected performance from random category labels.

Results

3.1 Primary Raga and Tani AF Functions

Figure 1 represents the individual AF functions and grand average functions for Raga and Tani. AF values increased as window sizes increased for both categories, indicating increased clustering across timescales. At the shortest window sizes, the Tani grand average was slightly higher than the Raga grand average. Both functions converged at approximately 0.1 seconds and diverged at the intermediate and longer window sizes. The Raga grand average function remained above the Tani function from approximately 0.25 seconds onward. The individual functions showed variability within both categories, but the trends overall from both categories remained consistent. Raga recordings generally exhibited steady and stronger increases in AF values across intermediate and longer windows, whereas Tani displayed flatter regions within the middle time windows before rising at the larger windows.

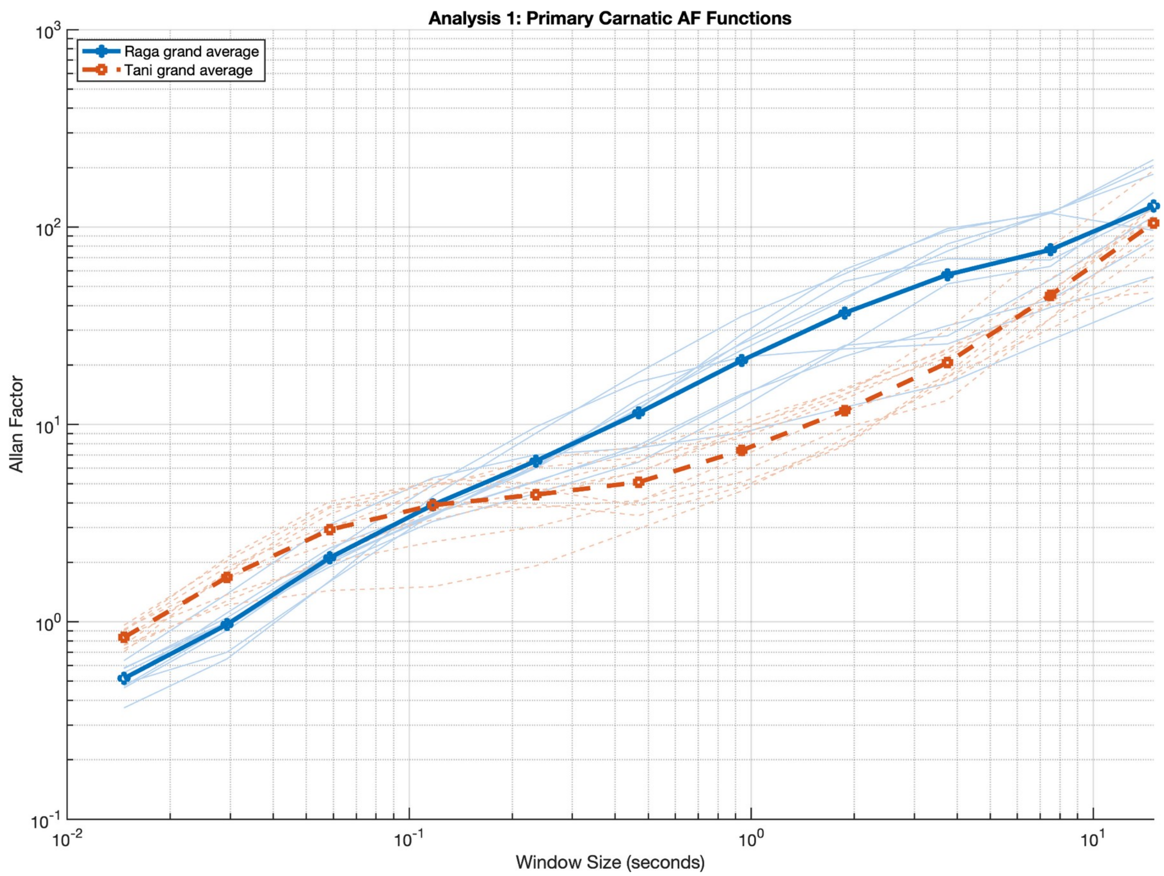

Figure 1. Recording-level and grand-average Allan Factor functions for vocal Raga Alapana and Jugalbandhi Tani Avartanam recordings. Thin lines represent individual recordings and thick lines represent the grand averages of both categories. Window duration and Allan Factor are represented on logarithmic axes.

3.2 Primary Comparison

Figure 2 compares the grand-average AF functions for conversation, jazz, beat-based music, Raga, and Tani. The AF values increased with the counting window duration across all five categories, but differed in shape and increase rate across timescales. Conversation, jazz, and Raga exhibited similar trajectories across the middle and long windows. The AF values remained adjacent through most of the observed function, with Raga showing the greatest increase in the longer window sizes. Tani and beat-based music exhibited lower AF values than conversation, jazz, and Raga across most of the middle time windows. Tani initialized at an observably higher AF value at the shortest window, then plateaued in the middle window, and finally increased sharply at the longer window sizes. At the longest window size, all five category functions stabilized at a similar level, although Raga remained the highest and beat-based the lowest.

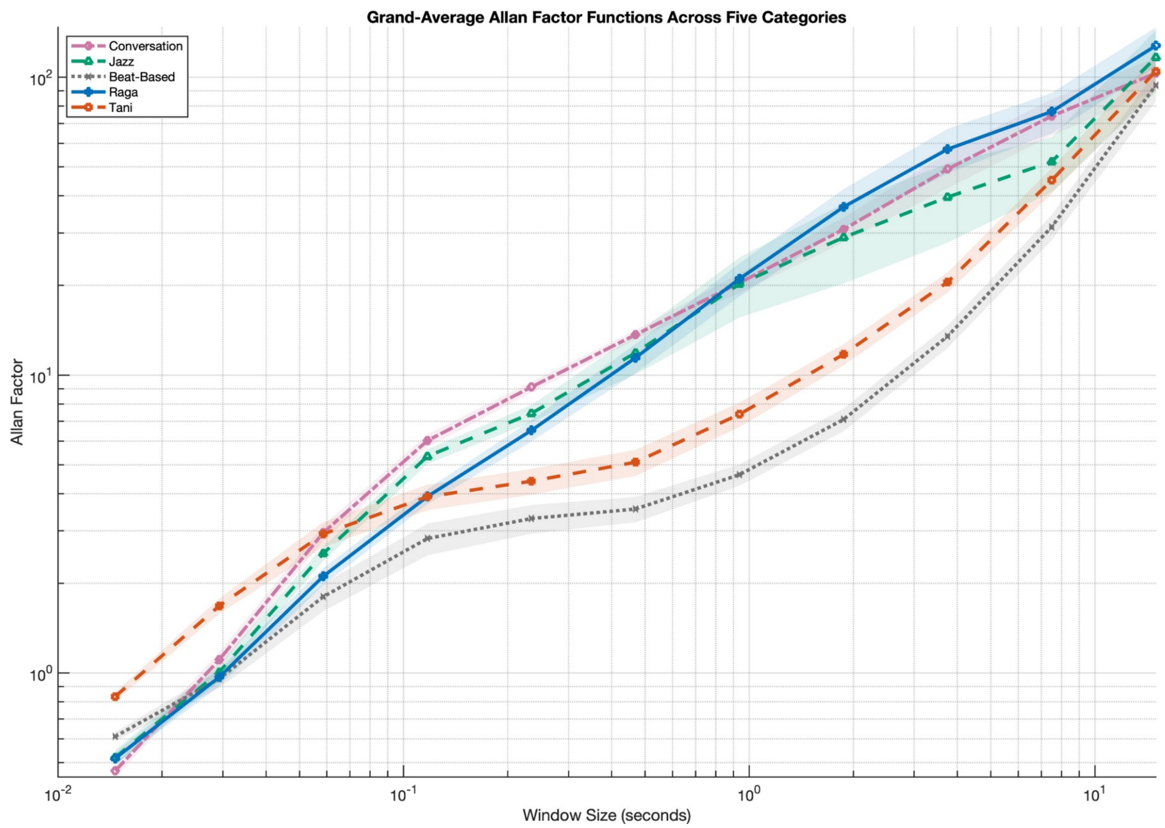

Figure 2. Grand-average Allan Factor functions for conversation, jazz, beat-based music, vocal Raga Alapana, and Jugalbandhi Tani Avartanam. Lines represent categorical averages, and shaded areas represent the standard error of the mean across the recordings. Windows and Allan Factor are plotted on logarithmic axes. Conversation, jazz, and beat-based comparison data were obtained from Kello et al. (2017).

3.3 Feature Space Comparison

Figure 3 shows the AF functions of the recordings using quadratic and linear coefficients obtained from the third-order polynomial fits. The centroids differed in location within the feature space. Raga produced both the most negative mean quadratic coefficient and the highest average linear fit. Conversation concurrently produced a negative average quadratic coefficient, but with a lower mean linear coefficient. Jazz was placed near zero on the quadratic axis and had a similar average linear coefficient to conversation. Tani and beat-based music were positioned towards the right of the plot with positive means for the quadratic coefficients. Their centroids were nearly adjacent on the linear axis, although Tani produced a slightly lower mean quadratic coefficient than beat-based music. Individual recordings reveal the variability within each category. Despite the variability, the categories showed a clear separation between Raga, conversation, and Jazz on the left versus Tani and beat-based music on the right.

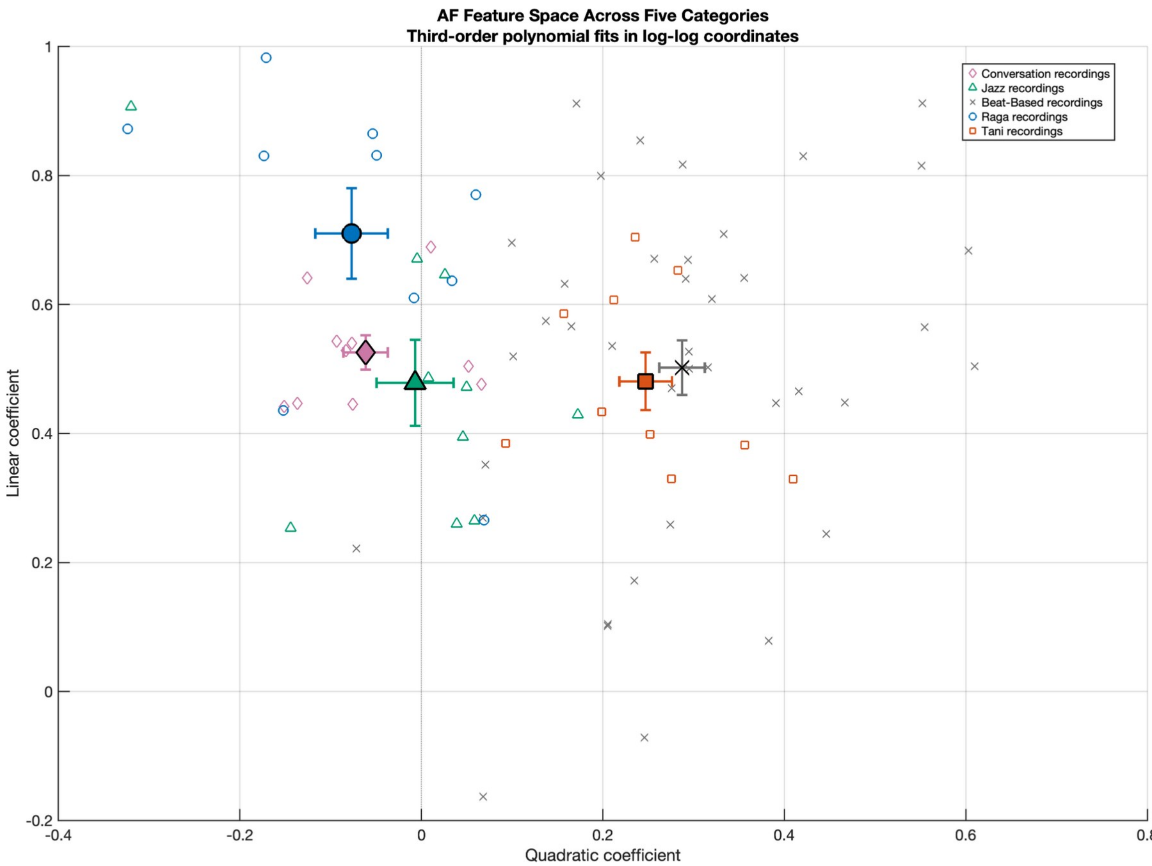

Figure 3. Allan Factor feature space for conversation, jazz, beat-based music, vocal Raga Alapana, and Jugalbandhi Tani Avartanam. Each small hollow symbol represents a single recording. Large filled symbols represent category means, and error bars indicate the standard error of the mean. Conversation, jazz, and beat-based comparison data were obtained from Kello et al. (2017).

3.4 Butovens Dataset Comparison

Figure 4 displays the grand-average AF functions for Carnatic vocal, Carnatic instrumental, Hindustani vocal, and Hindustani instrumental recordings from the independently collected Butovens dataset. All four categories showed an increase in AF values across all windows and followed similar trajectories. Differences among the categories became distinguishable at the longer windows. Hindustani vocal recordings showed higher AF values across many intermediate and long window sizes, while Carnatic vocal and Hindustani instrumental remained close across most of the function, converging near the longest windows. Despite minor differences, the four categories substantially overlapped in their standard error regions.

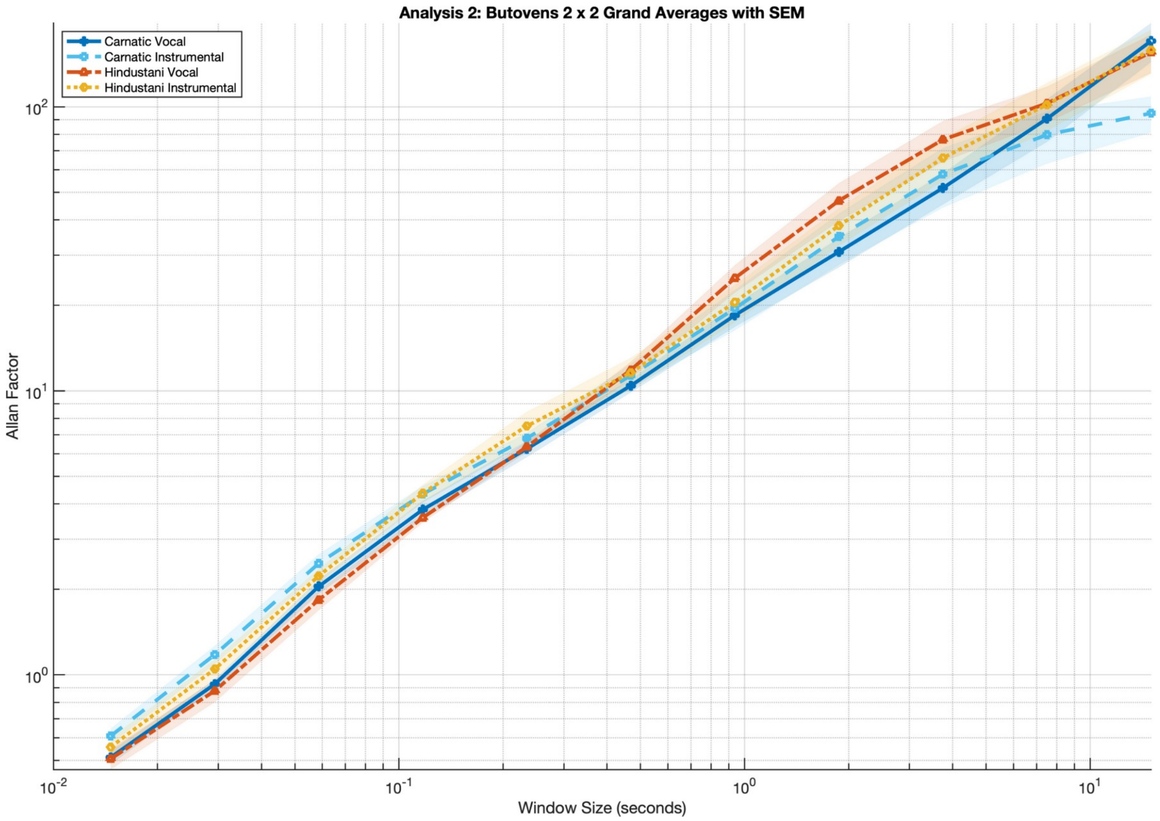

Figure 4. Grand-average Allan Factor functions for Carnatic vocal, Carnatic instrumental, Hindustani vocal, and Hindustani instrumental recordings from the independent Butovens dataset. Lines represent category means and shaded regions represent the standard error of the mean across the recordings. Windows and Allan Factor are plotted on logarithmic axes.

3.5 Raga and Carnatic Vocal Comparison

Figure 5 specifically compares the grand average AF function of the present study's vocal Raga dataset with the Carnatic vocal recordings from the independent Butovens dataset. The two functions followed similar trajectories across most windows. The AF values were almost identical at the small and medium windows, with great overlap in the standard error regions. Minor differences were apparent at the longer timescales, with the present Raga showing higher AF values between 1 and 5 seconds. The Butovens Carnatic vocal recordings increased sharply at the longest windows, slightly exceeding the present Raga dataset. Both AF functions closely aligned across the timescales overall.

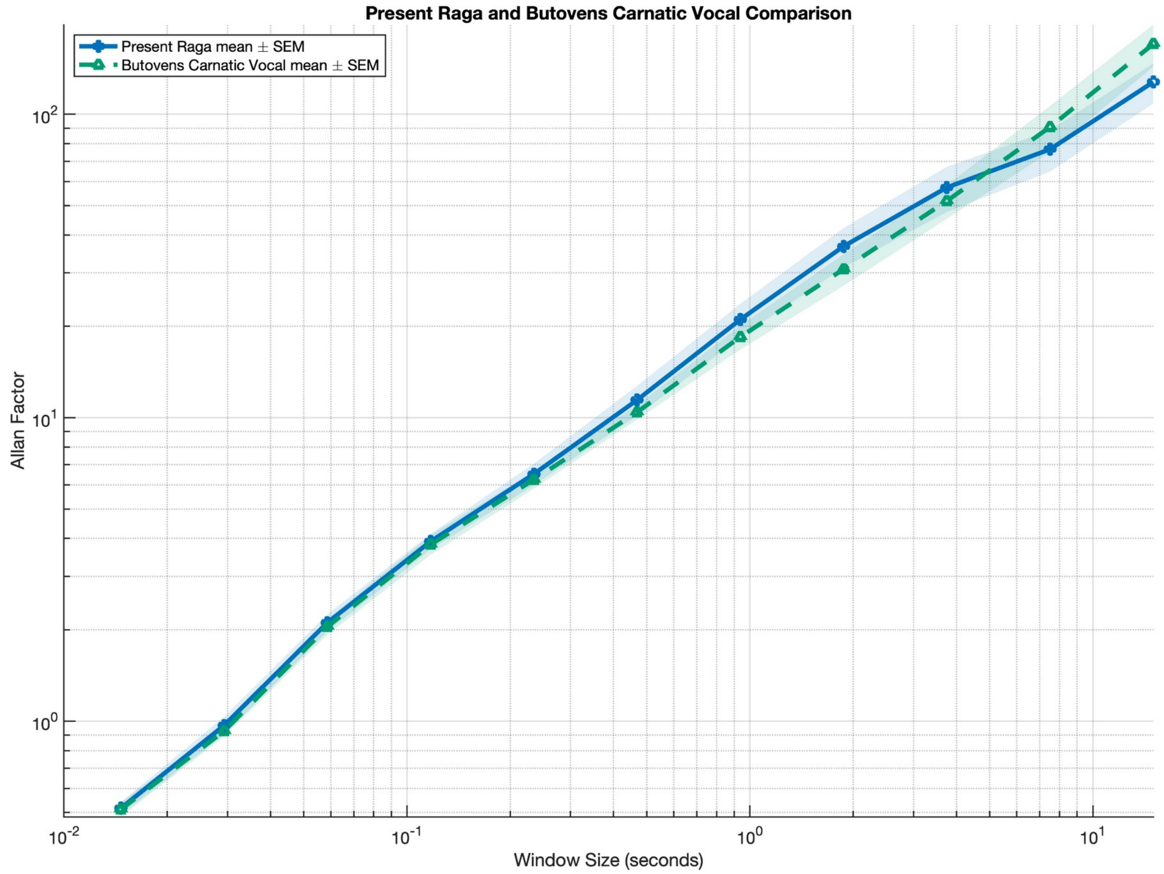

Figure 5. Comparison of grand-average Allan Factor functions for the present vocal Raga Alapana and independent Butovens Carnatic vocal recordings. Lines represent the category averages, with shaded regions representing the standard error of the mean across recordings. Windows and Allan Factor are plotted on logarithmic axes.

3.6 Inferential Statistics

Figure 6 presents the recording-level scaling exponents for the five main categories. Raga had the highest mean scaling exponent (M = 0.766, SE = 0.065), followed by conversation (M = 0.630, SE = 0.023) and jazz (M = 0.575, SE = 0.067). Tani (M = 0.448, SE = 0.036) and beat-based music (M = 0.446, SE = 0.029) had the lowest scaling exponents. A one-way ANOVA showed that scaling exponents differed distinctly across the 5 categories, $F(4,75) = 8.67$, $p < 0.001$, $\eta 2 = 0.316$, and $\omega 2 = 0.277$. The BF was not significant ($p = 0.161$), and a Welch ANOVA confirmed the overall difference ($F(4,24.73) = 10.27$, $p < 0.001$). Tukey comparisons showed that Raga had significantly higher scaling than beat-based music ($p < 0.001$), Raga was greater than Tani ($p = 0.001$), and Conversation was greater than beat-based ($p = 0.027$). The intended comparison between Raga and Tani showed a significant difference, Welch's $t(14.20) = 4.30$, $p < 0.001$, $d = 1.92$. The mean difference was 0.318, with a 95% confidence interval from 0.160 to 0.477. The separate factorial model for the independent Indian classical music dataset was not significant, $F(3,36) = 1.50$, $p = .232$, $R2 = 0.111$. The coefficients linked to tradition ($p = 0.414$), modality ($p = 0.578$), and the interaction between tradition and modality ($p = 0.643$) were not significant.

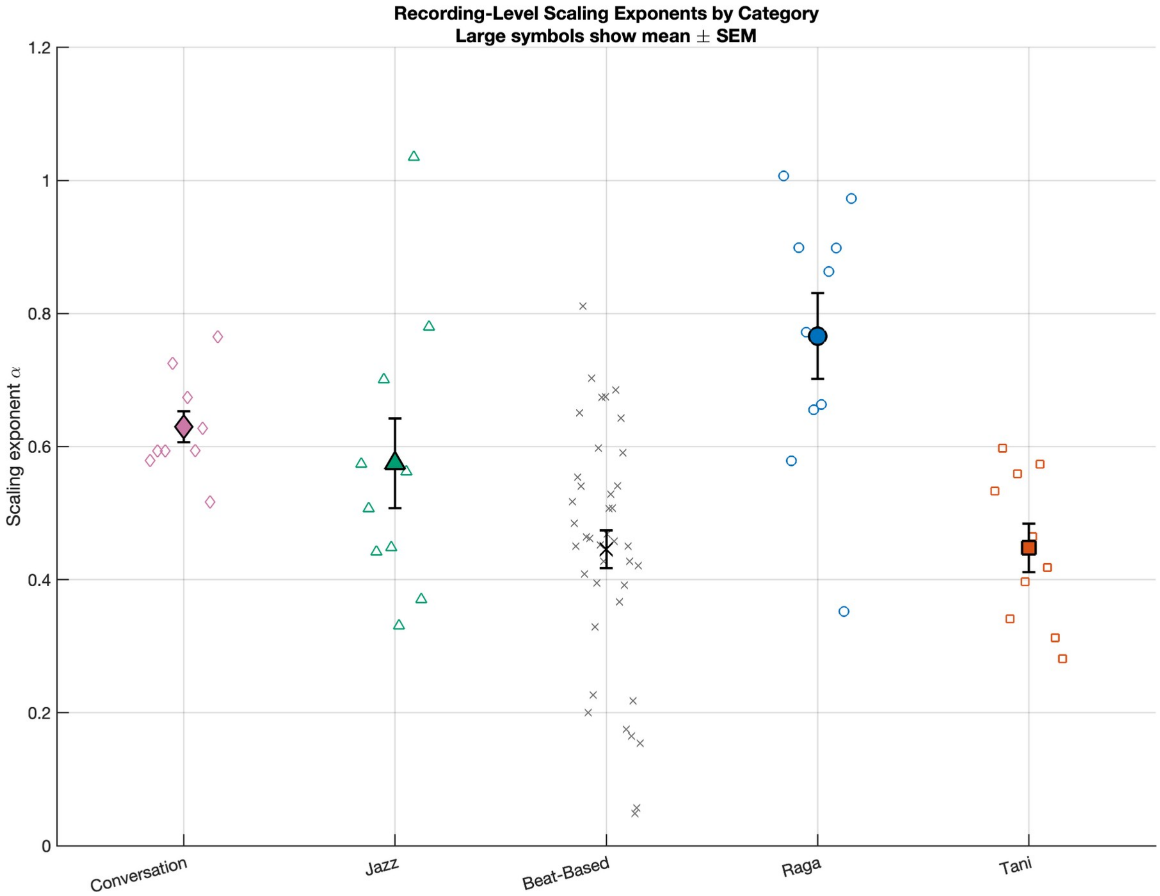

Figure 6. Recording-level scaling exponents for conversation, jazz, beat-based music, vocal Raga Alapana, and Jugalbandhi Tani Avartanam. Small symbols represent individual recordings. Large symbols represent category means, and error bars show the standard error of the mean. Conversation, jazz, and beat-based comparison data were obtained from Kello et al. (2017).

3.7 Classification Analysis

A linear support vector machine was used to classify recordings as conversation, jazz, beat-based music, Raga, or Tani using the complete 11-point AF functions. Across repeated fivefold cross-validation, the classifier had an overall accuracy of 74.1% and a balanced accuracy of 71.7%. The confusion matrix aggregated predictions across repeated trials, showing counts exceeding the number of unique recordings within each category. Category recall was higher for beat-based music (78.0%), followed by conversation (77.4%), Tani (75.8%), Raga (71.2%), and jazz (56.2%). The confusion matrix showed that Raga was most misclassified as conversation, whereas Tani was most misclassified as beat-based music. This pattern was consistent with the similarities observed in the grand-average AF functions and the feature space analysis. Precision ranged from 57.3% for Jazz to 90.2% for beat-based music. A permutation test showed the balanced accuracy was significantly greater than the accuracies after random reassignment, $p = 0.005$.

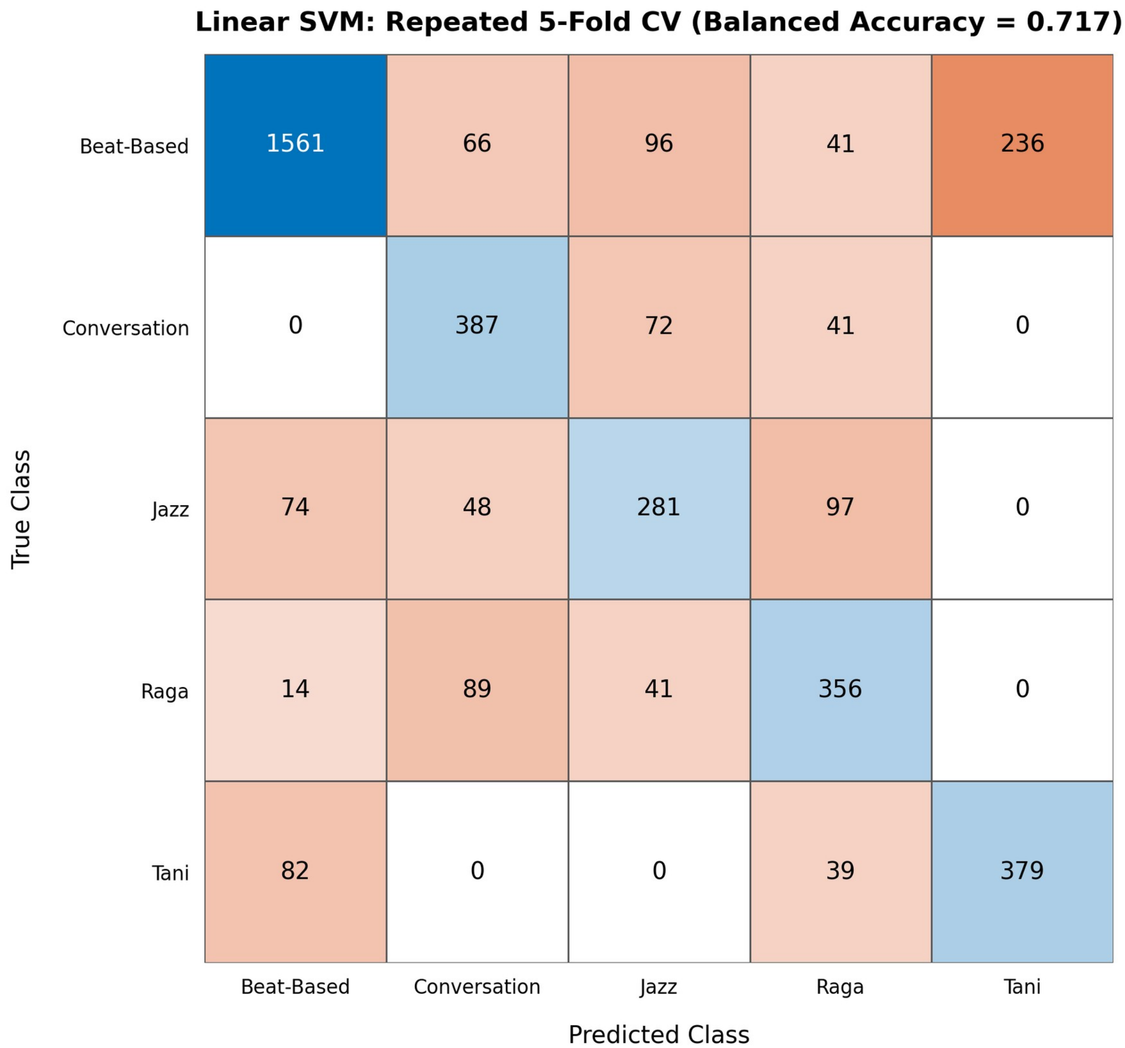

Figure 7. Aggregated confusion matrix for the linear support vector machine classifier using repeated fivefold cross-validation. Rows represent true categories and columns represent predicted categories. Values show the total number of classifications across repetitions. Conversation, jazz, and beat-based comparison data were obtained from Kello et al. (2017).

Discussion

4.1 Summary of Main Findings

The present study examined whether two highly interactive forms of Carnatic improvisation would reveal similar hierarchical temporal structure, or whether persistent rhythmic constraints would alter the temporal organization. The findings supported the second possibility. Although Raga and Tani showed increasing AF values across timescales, their functions followed differing trajectories. Raga started with a lower AF value and steadily increased, whereas Tani began with greater clustering, flattened at the middle levels, and increased at longer levels. The findings should not be interpreted as Raga having structure while Tani does not, but that both exhibited organization in different ways. Raga followed characteristics similar to conversation and jazz, whereas Tani closely followed beat-based music across the AF functions. Raga had a significantly higher mean scaling exponent than Tani and beat-based music, while Tani and beat-based music were almost identical. The feature space analysis created a visualization that showed Raga, conversation, and jazz grouped closer together and Tani positioned with beat-based music in a different region of the feature space. These findings extend the work of Kello et al. (2017), which showed that different acoustic behaviors can have similar temporal organization when involving related forms of coordination.

The classification analysis provided further evidence that the AF functions conveyed information that distinguished the five categories. The linear support vector classified the recordings well above chance, via a balanced accuracy of 71.7%. The permutation test supported these results by indicating that this performance was unlikely to result from a random assignment. At the same time, the classification was not perfect. It showed some shared temporal structures rather than completely separate organizational motifs. This observation is consistent with the 11-point AF functions rather than a single summary coefficient to evaluate both similarities and differences amongst the categories.

The independent Indian classical music dataset provided convergent evidence for the temporal organization observed in the present Raga recordings. The Carnatic and Hindustani vocal and instrumental categories followed similar AF trajectories, and the factorial analysis did not identify significant effects of tradition, modality, or their interaction. For the present study, the independently collected Carnatic vocal recordings closely followed the AF function of the present vocal Raga dataset across most timescales. Both datasets contain vocal Carnatic improvisation but are assembled independently. Their close correspondence suggests that the Raga pattern was not unique to the particular recordings selected for this study’s dataset. However, the independent corpus should be treated as supporting evidence rather than an exact replication because the datasets differed in their sources, recording conditions, and selection. Combined, the findings broadly supported the study’s hypothesis. Raga exhibited the interaction-related organization observed in conversation and jazz, while Tani exhibited a pattern more closely related to music organized around a persistent beat (Kello et al., 2017). The key result is not that melodic and percussion improvisation differed, but that the interactive structure clearly present in Jugalbandhi Tani performance was not reflected in the AF functions in the same manner as the interaction present in Raga, conversation, and jazz.

This contrast raises the possibility that strong rhythmic organization can dominate the acoustic event structure measured by Allan Factor analysis, even when the underlying performance remains highly interactive. This does not establish that interaction is absent from Tani, or that rhythm directly caused the observed difference, but asks whether interaction-related structures can be covered by a persistent rhythmic constraint.

4.2 Interaction and Rhythmic Organization

The clear difference between Raga and Tani suggests that hierarchical temporal structure does not depend only on interaction, but also on the temporal constraints within which the interaction happens. Raga Alapana develops through continuous elaboration and connection of melodic material. Locally produced vocal events contribute to phrases developing over longer portions of the performance (Kassebaum, 1987). The steady increase in the Raga AF function remains consistent with the form of development because the differences in event clustering continuously appear as the window sizes increase. However, AF analysis does not distinguish or isolate notes, phrases, and melodic devices. The findings do not establish that a specific musical technique yielded the Raga pattern. They showed that the timing of salient acoustic events was organized across a wide set of timescales. Tani displayed a different kind of cross-scale organization. The greater AF values at the shortest window sizes were followed by a flattening at the mid-range, before increasing at the longer window sizes. This plateau may result from timescale ranges that highlight cyclic rhythms that limit changes in event clustering. Percussionists introduce variations and develop their performance, but the underlying structure is ultimately coordinated into a strong beat cycle known as tala. The later increase in AF functions measures patterns of acoustic events rather than the literal musical content.

Tani and beat-based music behaving similarly is an important observation, as both domains differ in culture, synthesis, and acoustics. Their similarity is not simply shared musical tradition, but both contain underlying persistent rhythmic organization that determines when the salient events occur. Based on this view, the beat does not take away from the interaction with Tani. Instead, the periodic structure associated with tala may just dominate the event sequence analyzed by AF and make the effects of interaction less visible in the resulting function (Kello et al., 2017). This interpretation is consistent with accounts where complex behavior emerges through coordination among processes that operate across nested timescales (Kello et al., 2017; Kelty-Stephen & Dixon, 2014). Both Raga and Tani exhibited hierarchical organization, but the relationship of events between short and long timescales differed. Raga showed continuous growth across the function, while Tani showed strong short-timescale clustering and a plateau in the middle. These differences show why HTS cannot be reduced to a single estimate of event rate, variability, or tempo.

4.3 Performer Roles and Interaction Structure

The findings also ask about whether AF is sensitive to different forms of interaction within the performers. Raga, jazz, conversation, and Jugalbandhi are all interactive, but the interaction structure differs across these domains. In the vocal Raga recordings, the primary vocalist leads the melodic development while the accompanying musicians follow and respond. Jazz improvisation can involve flexible changes in leadership, where different musicians take the lead at different points in the performance (Kello et al., 2017). Conversation may follow this trend, shifting between more structured speaker roles and reciprocal exchanges. Previous work has shown that speech is organized across multiple temporal levels, including local vocal events, pauses, phrases, and broader conversational structure (Port, 2003).

Despite the differences in performer roles, Raga, jazz, and conversation showed similar AF trajectories. This suggests that identical social roles are not necessary for similar hierarchical temporal organization. Even a relatively stable leader-follower structure in Raga and oscillating leader roles within jazz may still generate comparable event patterns when both involve flexible adaptations over time. The current results suggest that AF may be responsive to the broader temporal consequences of interaction rather than to one specific interaction format. Related research shows that hierarchical event clustering changes across communicative contexts support the possibility that HTS reflects how interactions unfold rather than only the acoustic form of the signal (Falk & Kello, 2017).

Jugalbandhi provides an important contrast because two or more primary percussionists engage in reciprocal improvisation. This format was selected specifically to maximize interaction and avoid reducing Tani to a simple leader accompanied by a timekeeper. Nevertheless, the Tani AF function remained closer to beat-based music than to Raga, jazz, or conversation. This can be interpreted as the persistent tala constrained the acoustic timing of the interaction strongly enough that the reciprocal structure was not clearly represented in the AF function. The present study cannot determine whether different forms of Tani interaction would produce different results. A useful future comparison would examine Jugalbandhi performances separately from leader-follower Tani performances in which one primary percussionist is supported by an auxiliary player or rhythmic timekeeper. If AF is sensitive to the organization of interaction, more reciprocal and more hierarchical performance formats may produce distinguishable AF functions. If both formats are similar to beat-based music, then it would provide stronger evidence that rhythmic constraint has a greater influence on the measured acoustic event structure than the distribution of performer roles.

4.4 Independent Indian Classical Music Comparison

The independent Indian classical music dataset provided supporting evidence for the temporal organization observed in the present Raga corpus. Carnatic vocal, Carnatic instrumental, Hindustani vocal, and Hindustani instrumental recordings all showed increasing AF functions with substantial overlap across timescales. The factorial model did not identify significant effects of tradition, modality, or their interaction. These null effects should not be interpreted as proof that the four categories are equivalent, but that the analysis did not find reliable differences in recording-level scaling exponents within this sample. The most direct comparison involved the present vocal Raga recordings and the independent Carnatic vocal recordings. Their grand-average AF functions closely overlapped through most of the analyzed timescales. Minor differences appeared at the largest windows, but the general trajectories were highly similar. Since the datasets were assembled independently, this suggests that the pattern observed for Raga was not limited to the specific recordings selected in the primary dataset.

At the same time, the comparison was not an exact replication. The datasets differed in recording sources, recording conditions, performer selection procedures, and potentially the forms and contexts of improvisation represented. The close overlap should be treated as convergent evidence rather than proof of pure reproducibility. The purpose of the comparison was to determine whether a separate collection of vocal Carnatic performances showed a broadly similar temporal pattern, and the results supported the expectation. The similarity among the four independent Indian classical categories also suggests that the continuous AF trajectory associated with the present Raga recordings may extend across a much broader range of melodic Indian classical performance. However, tradition and modality were not experimentally manipulated, so the current findings cannot establish whether Carnatic and Hindustani systems, vocal or instrumental, have identical temporal organization. A larger and more systematically sampled dataset is required to test the difference with better precision.

4.5 Limitations and Future Directions

Several limitations should be considered when interpreting the findings. First, the primary dataset contained ten Raga and ten Tani recordings. Although the statistical analyses identified reliable category differences, these samples cannot represent the full variety of Carnatic improvisations. Performance can differ in vocalist, percussionist, ensemble size, tala, tempo, concert setting, and performance style. The Jugalbandhi recordings also differed in the number and types of primary percussion instruments present. Future datasets should include a larger number of recordings and more complete metadata so that the variability can be directly identified and isolated. All primary recordings were obtained from publicly available performances, allowing unadulterated music to be studied, but recording quality and other general quality control features remained uncontrolled. The preprocessing procedure reduces some of these differences by converting recordings into binary event sequences, but it does not remove all effects of the original audio conditions. Carefully recorded performances under more consistent conditions would provide a stronger test of whether the observed differences arose from musical organization rather than the technical recording properties.

The present Raga dataset was limited to vocal performances, while Tani was limited to Jugalbandhi. These restrictions improved consistency but limited the range of Carnatic performance represented. The findings should not automatically be generalized to instrumental Raga, composed Carnatic sections, solo Tani performances, or leader-follower Tani formats. Future work should examine these categories separately rather than combining them into broader groups. The AF pipeline reduced each recording to the timing of the largest acoustic peaks, allowing common representation to compare different sound domains, but removed information regarding pitch, timbre, intensity, performer identity, and musical function. All acoustic properties were represented in the same binary framework, which is useful for studying temporal clustering but does not indicate which performer produced an event or whether the event was a beat, response, or transition. This limitation is significant for Tani, as the current analysis showed that its AF function resembled beat-based music but did not reveal whether interaction-based organization remained present underneath the consistent rhythmic component. Future work may attempt to separate regularly occurring beats from less periodic improvisational occurrences before computing AF. One possibility is to estimate the dominant structure and examine the remaining event sequence after that part is reduced. This kind of filtering needs to be performed carefully, as removing beats could remove meaningful parts of improvisations. The goal is not to eliminate Tani's musical structure, but to test whether a secondary interaction pattern becomes visible when the strongest periodic component is reduced.

Pitch-based percussion provides another possible direction to be considered, as it joins rhythmic interaction and adds melodic variation that may produce an AF function that differs from monotone-pitched or non-pitched beat-based percussion. Comparing pitched and non-pitched percussion could help determine whether the Tani pattern is primarily associated with periodic onset timing, percussion timbre, or the absence of melodic improvisation. The present recordings did not particularly test this distinction, so any apparent difference amongst the individual recordings is exploratory. Future analyses could also represent separate performers as distinct event sources. Source separation could be used to identify when each performer leads, follows, responds, or overlaps with another performer. This would permit the temporal structure of individual contributions and interactions between performers to be measured separately. This kind of approach may reveal patterns that are obscured when the entire ensemble is reduced to a binary combined audio signal. Lastly, the classification results demonstrate that the AF functions contained category-relevant information, but they should not be interpreted as performance on a completely independent external dataset. Repeated cross-validation evaluates how consistently the categories can be identified within the available sample. A stronger test of generalization would train the classifier on one collection and evaluate it on newly collected recordings from different performers and recording sources.

4.6 Conclusion

The present study applied Allan Factor analysis to two highly interactive forms of Carnatic improvisation and examined how their acoustic events were organized across multiple timescales. Raga and Tani both exhibited HTS, but their AF functions followed differing trajectories. Vocal Raga showed a continuously increasing function and was positioned near conversation and jazz across the target comparisons. Jugalbandhi showed stronger short-window clustering, intermediate plateaus, and a similar relationship to beat-based music. These differences were supported by recording-level scaling exponents, feature-space analysis, and an above-chance classification of the AF functions. The findings suggest that interaction alone does not determine how hierarchical temporal structure appears in an acoustic event sequence. Tani remained highly interactive at the performance level, but its persistent rhythmic organization was strongly reflected in the AF results. The beat may dominate the temporal effects of interaction without fully eliminating the interaction itself. Determining whether that interaction-related structure can be recovered through beat filtering, performer-specific analyses, or comparisons with other percussion techniques would be an illuminating direction for future work. The independent Carnatic vocal dataset showed a temporal trajectory closely aligned with the present Raga recordings by providing convergent evidence that the Raga result was not unique to one selected corpus. Broadly, these findings extended the hierarchical temporal framework of Kello et al. (2017) to South Indian classical improvisation. Allan Factor analysis provided a common method for comparing speech, Western music, and Carnatic performance despite major acoustic and cultural differences. The results support the view that complex human behavior is organized across nested timescales while showing that the organizing form depends on the interaction between coordination and rhythmic constraint (Kello et al., 2017; Kelty-Stephen & Dixon, 2014).

References

Falk, S., & Kello, C. T. (2017). Hierarchical organization in the temporal structure of infant-directed speech and song. *Cognition, 163*, 80–86. https://doi.org/10.1016/j.cognition.2017.02.017

Kassebaum, G. R. (1987). Improvisation in Alapana performance: A comparative view of Raga Shankarabharana. *Yearbook for Traditional Music, 19*, 45–64. https://doi.org/10.2307/767877

Kello, C. T., Dalla Bella, S., Médé, B., & Balasubramaniam, R. (2017). Hierarchical temporal structure in music, speech and animal vocalizations: Jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of the Royal Society Interface, 14*(135), 20170231. https://doi.org/10.1098/rsif.2017.0231

Kelty-Stephen, D. G., & Dixon, J. A. (2014). Interwoven fluctuations during intermodal perception: Fractality in head sway supports the use of visual feedback in haptic perceptual judgments by manual wielding. *Journal of Experimental Psychology: Human Perception and Performance, 40*(6), 2289–2309. https://doi.org/10.1037/a0038159

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech* (2nd release) [Data set]. Department of Psychology, The Ohio State University. https://buckeyecorpus.osu.edu

Port, R. F. (2003). Meter and speech. *Journal of Phonetics, 31*(3–4), 599–611. https://doi.org/10.1016/j.wocn.2003.08.001