Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Algorithmic Composition Using Narrative Structure and Tension

Francisco Braga[1], **Gilberto Bernardes**[2], **Roger B. Dannenberg**[3], **Nuno Correia**[1]
[1]NOVA LINCS, NOVA School of Science and Technology
[2]INESC-TEC, Faculty of Engineering, University of Porto
[3]Department of Computer Science, Carnegie Mellon University
fma.braga@campus.fct.unl.pt

Abstract

This paper describes an approach to algorithmic music composition that takes narrative structures as input, allowing composers to create music directly from narrative elements. Creating narrative development in music remains a challenging task in algorithmic composition. Our system addresses this by combining leitmotifs to represent characters, generative grammars for harmonic coherence, and evolutionary algorithms to align musical tension with narrative progression. The system operates at different scales, from overall plot structure to individual motifs, enabling both autonomous composition and co-creation with varying degrees of user control. Evaluation with compositions based on tales demonstrated the system's ability to compose music that supports narrative listening and aligns with its source narratives, while being perceived as familiar and enjoyable.

1 Introduction

Music and narrative have been studied through different disciplines, from musicology, where the question of whether music can be considered a narrative has been discussed [Almén, 2003; Maus, 1988], to music perception, where research has examined the emergence of narrative interpretations during music listening [Margulis *et al.*, 2019; Margulis *et al.*, 2022]. Several algorithmic composition[1] systems have been developed across different communities - artificial intelligence [Davis and Mohammad, 2014; Jeong *et al.*, 2017], computational creativity [Prechtl *et al.*, 2014; Guo *et al.*, 2022], computer music and music information retrieval [Ruiz-Marcos, 2021; Lopez and Alvaro, 2024] - that either compose music for existing narratives, or create musical analogies of existing narratives. However, the ability to create a narrative sense of development in music remains an open challenge in these systems [Carnovalini and Rodà, 2020]. Most systems lack mechanisms for narrative adaptability and struggle to control expressive features in alignment with narrative cues [Carnovalini and Rodà, 2020]. Long-term structural coherence also remains challenging, along with the need for systems that can align music with higher-level narrative concepts and adapt to interactive contexts [Herremans *et al.*, 2017]. These challenges reveal an opportunity for systems capable of integrating narrative structure, expressive control, and thematic coherence within algorithmic composition.

In this paper, we present an approach to algorithmic composition based on abstract narrative concepts. Our system is guided by a narrative representation [Gervás, 2019] that is adaptable to various storytelling domains. It combines generative grammars to generate chord progressions and evolutionary algorithms to optimize musical elements. The system incorporates narrative cues such as character themes (*leitmotifs*), repeated patterns and musical tension to create narrative structure within music [Herremans *et al.*, 2017]. The evolutionary algorithm combines and evolves these elements while optimizing for target musical tension, tonal coherence and recognition of the original *leitmotifs*.

The novelty in our work lies in (i) the creation of a system that composes music based on narrative structures by integrating computational representation of narrative, combining narrative cues and using a hierarchical architecture; and (ii) a hybrid implementation that allows for autonomous composition and co-creation with various degrees of controllability.

The remainder of the paper is structured as follows: Section 2 presents related work on narrative-based algorithmic composition and musical tension. Section 3 provides the foundations for our approach, Section 4 explains our system's architecture and implementation, Section 5 discusses evaluation and results, Section 6 addresses limitations, and Section 7 presents conclusions and future work.

2 Related Work

Our work focuses on algorithmic composition based on narrative structure and tension while serving as a tool for co-creation. As such, we consider related work in two main areas: algorithmic composition, especially approaches using musical tension or narrative analogy, and systems for composition co-creation.

[1]Throughout this paper, we use the term "algorithmic composition" [Fernández and Vico, 2013], although other terms such as "symbolic music generation" or "music generation" are also common in the literature [Carnovalini and Rodà, 2020].

2.1 Algorithmic Composition

Works on algorithmic composition related to ours can be divided into two types: approaches that compose music based on different narrative domains and approaches that utilize tension as a guiding principle.

In the narrative-driven category, several works use different approaches to map narrative elements to musical features. MAgentA [Casella and Paiva, 2001] composes "film-like music" using a rule and agent-based approach, selecting composition algorithms based on emotional states. Brown developed the Mezzo system [Brown, 2012] for real-time composition of game soundtracks. It combines musical form and harmonic tension to reflect the narrative states of the game, while recurring to *leitmotifs* for characters, props, and other important elements in the game. Prechtl *et al.* [2014] implemented a first-order Markov model to define chord transitions for real-time composition for games. The transition matrix varies according to the *danger* level, a game narrative metric defined by the authors. Transpose [Davis and Mohammad, 2014] composes music for novels by analyzing emotional content through Emotion Word Densities and mapping it to both global musical parameters (key, tempo) and local features (pitch, duration).

In the tension-guided category, the approaches vary with respect to the tension models and composition methods used. Ruiz-Marcos [2021] uses Lerdahl's Model of Tonal Tension [Lerdahl and Krumhansl, 2007], with four components combining rules, generative grammars, and statistical methods. MorpheuS [Herremans and Chew, 2017] implements the Tension Ribbons model [Herremans *et al.*, 2016] to compose music that matches specified tension profiles (either from a template piece or from user input). The authors use an optimization approach, implementing a variable neighborhood search algorithm to optimize the music to the tension patterns. Guo *et al.* [2020] use a variational autoencoder (VAE) to compose music controlled by tonal tension. They use two tension measures from the Tension Ribbons model and were able to identify four latent feature vectors for tension manipulation. The same authors developed a system based on the Transformer network for infilling applications with multi-level control [Guo *et al.*, 2022]. They utilize control tokens at the track level for note density, polyphony, and occupation rate, and at the bar level for two of the tension measures from the Tension Ribbons model. MoodLoopGP [Cui *et al.*, 2024] composes loop tablature music using valence, arousal, and mode control tokens, along with the three tension measures from Tension Ribbons. Navarro-Cáceres *et al.* [2019] created an assistive system to compose chord progressions using an artificial immune system. The system uses perceptual properties of chords as objective functions, which were later extended as a tonal tension model [Navarro-Cáceres *et al.*, 2020]. Jeong *et al.* [2017] developed a multi-objective evolutionary approach using NSGA-II [Deb *et al.*, 2002], optimizing for both stability and tension measures. A similar approach is used in another system [Jeong *et al.*, 2022], but instead of the NSGA-II the authors use a Deep Network-Based Estimation of Distribution Algorithm, where a VAE is used for new solutions (trained on each generation of the evolutionary algorithm).

2.2 Co-creative Music Systems

Co-creative music systems employ diverse computational approaches and allow varying degrees of human interaction in the composition process. Farbood *et al.* created Hyperscore [Farbood *et al.*, 2004], a graphical computer-assisted composition system designed for novice composers, particularly children. It maps graphical elements controlled by users (sketches) to both high and low level musical features, such as harmonic tension, melodic contour, pitch, and dynamics. Cococo [Louie *et al.*, 2020] is a web-based music editor that enables novices to co-create music with AI. It does this through targeted generation controls for specific voices (soprano, alto, tenor, bass) and time measures, along with semantic sliders (such as happier/sadder, or more conventional/more surprising) for high-level musical direction. Déguernel *et al.* [2022] proposed a system for co-creative music composition using a machine learning and rule-based approach. Co-creation is achieved through data sharing (a user can provide its own musical data) and knowledge sharing (a user can provide rules and guidelines). The MMM-Cubase [Tchemeube *et al.*, 2023] is the application of the Multi-Track Music Machine [Ens and Pasquier, 2020] as a Digital Audio Workstation (DAW) plugin. It enables multi-track music co-creation with three types of user action: track and bar in-filling, and attribute controls. CHAMELEON [Zacharakis *et al.*, 2021] is an melodic harmonization assistant that offers co-creation through conceptual blending of different harmonic idioms. Micchi *et al.* [2021] described a co-creative approach to songwriting, where AI is used as a suggestion tool for multiple musical layers, including melody, chord sequences, lyrics, and global structure.

3 Background

This section presents the foundations necessary to our narrative approach to algorithmic composition. We start by explaining some of the work on narrative structure representation. Then, we go through some of the existing models of tonal tension with special focus on the Tension Ribbons [Herremans *et al.*, 2016], the model used in our system. Following this, we present Rohrmeier's [2011] generative grammar of tonal harmony, which provides a framework to regulate the tonal functions' hierarchy of our system. Finally, we briefly introduce evolutionary algorithms in the context of algorithmic composition, especially the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [Deb *et al.*, 2002] implemented in our system. These components work together to provide the foundation for our system: narrative structure provides the high-level organization, musical tension enables an analogy for the narrative tension, harmonic grammar ensures musical coherence, and evolutionary algorithms enable flexible solution generation.

3.1 Narrative Structure Representation

Considering computational representations of narrative, there are several approaches in the field of narrative generation. These vary mainly in terms of the granularity considered for the knowledge units and on the control methods used to combine each unit into plots [Gervás, 2017].

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

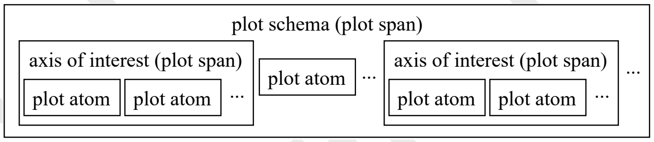

Figure 1: Narrative Representation Overview

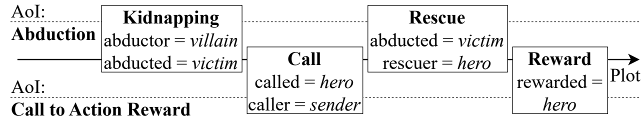

Figure 2: Narrative structure example [Gervás and Méndez, 2024]. The plot combines two Axes of Interest (AoI): "Abduction" (containing plot atoms "Kidnapping" and "Rescue") and "Call to Action Reward" (containing "Call" and "Reward"). Each plot atom assigns specific roles to characters (*e.g.*, in "Kidnapping", the *villain* abducts the *victim*). The final plot (plot schema) is constructed through the non-linear arrangement of these plot atoms from both axes.

The most popular representation baseline for computational representations of narratives is the concept of *character functions* introduced by Propp [1968] in his morphology of the Russian folk tale. Propp structurally analyzed a corpus of Russian folk tales and identified a common set of regularities in terms of the actions performed by characters. He named these character functions, defined according to their importance to the plot's course of action. Examples of these functions include performing a villainous act, starting or winning a fight, departing on a journey, or rewarding someone [Propp, 1968]. Another important aspect identified by Propp is the set of roles for characters in the narrative, which he organized into spheres associated with *dramatis personae* such as the hero, villain, and victim. These two concepts, character function and role, serve to structure the morphology as an outline of the elementary structure of tales.

There are several computational representations based on Propp's morphology, where each approach focuses on different sets of features necessary in a story [Gervás, 2017]. For our model, we adopted the narrative representation developed by Gervás [2019], as it combines a broad set of features necessary for a story while remaining adaptable and simple. The representation is based on Propp's work and introduces a few concepts to be able to generate more complex structures.

The basic unit of this representation is the *plot atom*, which is similar to Propp's character function, adding information on how the roles of the unit (*e.g.*, kidnapper, kidnapped) are filled by roles relevant to the plot (*e.g.*, villain, victim) [Gervás, 2018a]. The unit above this is called *plot span*, and can be constituted by a sequence of plot atoms or smaller spans. If a plot span constitutes the entire plot, it is called *plot schema*, if it is an intermediate unit in the plot structure it is called *axis of interest* [Gervás, 2018b]. Each plot span can have a protagonist defined (role in the axis of interest, character in the plot schema). An overview of the narrative representation can be seen in Figure 1. The axes of interest could be combined with other axes of interest and plot atoms to define a plot schema. The plot atoms of an axis of interest do not need to occur sequentially, but may be intertwined with other axes of interest and plot atoms, with plot links connecting sequential units. This allows the story to maintain both local coherence and broader narrative structure. An example can be seen in Figure 2. The only change to the representation is the addition of a tension curve for each plot atom. When building the complete narrative, these tension curves are concatenated to create the story's overall tension curve.

3.2 Musical Tension

Musical tension is an important tool for evoking emotion in the listener [Herremans *et al.*, 2017], as such it has the potential to connect narrative and music. There are several tension models from the fields of musicology, computer music, and music information retrieval. While some notable approaches include Lerdahls' Model of Tonal Tension [Lerdahl and Krumhansl, 2007] based on GTTM [Lerdahl and Jackendoff, 1996] and TPS [Lerdahl, 2001] theories, and the Computational Model of Tonal Tension Profile [Navarro-Cáceres *et al.*, 2020] using Tonal Interval Space [Bernardes *et al.*, 2016], this work focuses on Tension Ribbons [Herremans *et al.*, 2016] due to its empirical validation and growing adoption in algorithmic composition systems.

Tension Ribbons [Herremans *et al.*, 2016], the model employed in our system, is based on the spiral array representation [Chew, 2002], which provides a geometric model of tonal space where pitch classes, chords, and keys are represented as points in a three-dimensional helix. This model calculates tension through three distinct measures that capture different aspects of musical tension:

1. Cloud Diameter: Measures the largest distance between any two notes at a given time.
2. Cloud Momentum: Calculates the distance between two consecutive centers of effect, where the center of effect represents the weighted center of a set of pitch classes in the spiral array space.
3. Tensile Strain: Evaluates the distance between the center of effect of a cloud of notes and the center of effect of the key.

In addition to the empirical validation [Herremans *et al.*, 2016] and the growing adoption in recent algorithmic composition systems [Herremans and Chew, 2017; Guo *et al.*, 2020; Guo *et al.*, 2022; Cui *et al.*, 2024] mentioned above, this model also has the advantage of being simple with three complementary measures, making it easier to interpret. Musical tension provides a framework for translating narrative tension into musical form. However, maintaining harmonic coherence is important in the creation of musically coherent compositions, which led us to consider formal representations of tonal harmony.

3.3 Generative Grammar of Tonal Harmony

Rohrmeier [2011] proposed a formalization of tonal harmony specifically designed for computational purposes. It is based on functional theories of harmony, and considers two underlying principles: the dependency principle, which states that "each chord in a sequence is structurally connected to its preceding chord or chord group in a dependency relationship" and the functional heads principle, which states that "chords are organized into functional categories which describe their tonal function which may be instantiated or modified by different chords" [Rohrmeier, 2011]. This formalization captures the hierarchical nature of tonal progressions through a recursive structure that allows for both local and long-term dependencies.

The syntax is defined as a set of context-free production rules organized in four different levels: phrase, functional, scale degree, and surface levels. It utilizes the three main tonal function, *tonic*, *dominant* and *subdominant* to characterize both terminal and non-terminal symbols. The non-terminal symbols represent regions (*TR* - tonic region, *SR* - subdominant region and *DR* - dominant region), while terminal symbols represent specific functional terms (*t - tonic, s - subdominant, d - dominant, tp - tonic parallel, sp - subdominant parallel, dp - dominant parallel, tcp - tonic counter parallel*). This approach provides a clear framework for valid harmonic progressions, considering both local relationships and larger-scale tonal structures.

Please refer to the original Rohrmeier paper [Rohrmeier, 2011] for the complete set of production rules and detailed formalization of the syntax.

3.4 Evolutionary Algorithm

Evolutionary algorithms have been used in algorithmic composition for several compositional tasks, such as melody, harmony, rhythm, jazz improvisation, and polyphony [Fernández and Vico, 2013; Herremans *et al.*, 2017]. In our case, the algorithm is used to combine preexisting melody and harmony to find interesting solutions according to a set of objective functions. In addition, the system is intended to be flexible, allowing for varying degrees of control, from fully automatic to manually guided composition.

Evolutionary algorithms are particularly suitable for this task due to their effectiveness in handling large and complex search spaces. Also, since evolutionary algorithms maintain a population of candidate solutions, they allow for the emergence of multiple viable alternatives, which is especially valuable in creative domains such as musical composition. Given these requirements, we opted for a Multi-Objective Evolutionary Algorithm (MOEA). There are three main MOEA paradigms: Pareto-based, Indicator-based and Decomposition-based [Emmerich and Deutz, 2018]. Pareto-based MOEAs rank solutions first by Pareto dominance and then by their contribution to population diversity. Indicator-based MOEAs use quality metrics to evaluate and guide the selection of solution sets. Decomposition-based MOEAs divide the original problem into multiple single-objective subproblems using different weight vectors. We selected a Pareto-based MOEA for its ability to generate diverse solutions, which enables both automated solution selection through preference weights and manual selection from a set of alternatives. Specifically, we implement the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [Deb *et al.*, 2002], which is the most widely adopted Pareto-based MOEA. This algorithm generates solutions along the Pareto front that exhibit significant diversity. The system can therefore function autonomously by selecting solutions based on preference weights or interactively by presenting multiple diverse solutions for user selection.

NSGA-II [Deb *et al.*, 2002] operates through an iterative process that uses typical elements of genetic algorithms with specialized mechanisms to handle multiple objectives. The algorithm maintains a population of candidate solutions and employs three key mechanisms: non-dominated sorting, which ranks solutions based on Pareto dominance; crowding distance calculation, which promotes diversity among solutions of the same rank; and elitism, which preserves the best solutions found. In each generation, parent solutions are selected based on these rankings through binary tournament selection. New offspring are produced through genetic operators, and the combined parent-offspring population is filtered to maintain a fixed population size while preserving the best solutions.

The combination of NSGA-II's multi-objective optimization capabilities with our narrative-based representation, tension modeling, and harmonic grammar constraints enables the generation of musically coherent compositions that effectively translate narrative structure and tension while maintaining harmonic coherence. More details will be presented in the following section.

4 Algorithmic Composition Using Narrative Structure and Tension

In this section, we present our approach to algorithmic composition using narrative structure and tension. The system was designed to be flexible, enabling both autonomous composition and music co-creation. The system composes music consisting of melody and chord progression; it does not produce arrangement or performance interpretations (discussed further in Section 6). The system combines multiple techniques: it utilizes *leitmotifs* as a basis for phrase construction while optimizing for target musical tension curves and maintaining structure coherence through context-free grammar harmony. The *leitmotifs* serve as musical representations of narrative elements - specifically characters in our implementation - establishing narrative associations within the composition. The system utilizes musical tension to align the composition with the narrative tension while using the generative model of tonal harmony to ensure coherent musical structure. This approach prevents ill-defined structure and tension patterns that could result from the use of musical tension alone. To combine all of these elements - *leitmotifs*, tension curves, and harmonic progression - we implemented the multi-objective evolutionary algorithm NSGA-II. The code and examples are available at https://github.com/braga1376/algorithmic-composition-narrative-tension .

4.1 System Architecture

The system architecture has three main hierarchical components: the Plot Span Music Composer, the Plot Atom Music Composer, and the Leitmotif Composer. Figure 3 presents an overview of the architecture of the system.

The Plot Span Music Composer serves as the entry point of the system, receiving the narrative structure (plot schema). It identifies the unique characters and initializes a Leitmotif

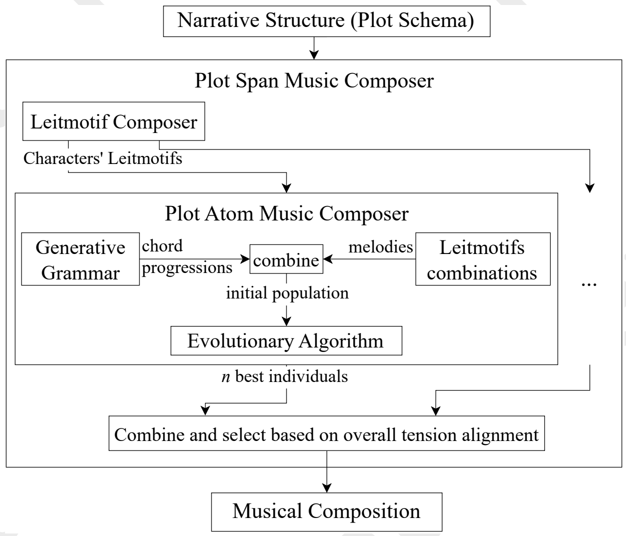

Figure 3: System Architecture

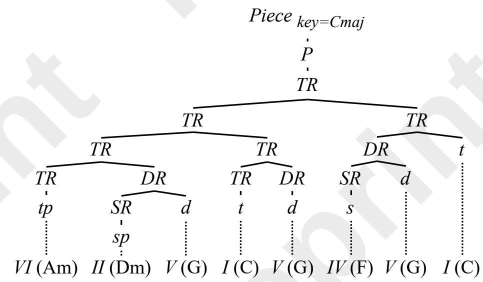

Figure 4: Harmonic Tree Example - non-terminal nodes: *TR*, *DR* and *SR*; terminal nodes: *t*, *d*, *s*, *tp* and *sp*; dashed line: application of rules at the scale degree level (*e.g.*, *tp*→*VI*) and surface level (*e.g.*, *VI*→Am)

Composer for each. Then, it initializes the necessary components for each of the plot schema's components (see Figure 1): Plot Atom Music Composers for each plot atom and Plot Span Music Composers for each axis of interest. Any new Plot Span Music Composers are recursively processed in the same way.

The core composition process occurs in the Plot Atom Music Composer, which processes each plot atom independently. This component receives the *leitmotifs* corresponding to its characters and creates possible melodies by combining these motifs in different permutations. It computes the narrative tension curve from the plot atom and applies the context-free grammar to produce coherent harmonic progressions of appropriate length. These elements are combined to create the initial population. This population goes through an evolutionary composition process using the NSGA-II which balances three objectives: aligning musical tension with narrative tension, maintaining tonal coherence, and preserving character leitmotif recognition. This component produces a set of individuals on the Pareto front. These can be selected automatically using a weighted selection matrix to identify the *n* best individuals according to predefined preferences, or manually chosen. In the automatic scenario, the Plot Span Music Composer combines the selections from each Plot Atom Music Composer and selects the combination with a better correlation with the overall narrative tension.

4.2 Implementation

The input narrative is structured as a plot schema, following Gervás's [2019] representation. As detailed in Section 3.1, this schema represents the overall plot through axis of interest or plot atoms (Figure 1). For musical representation, we use a MIDI-like representation, where notes are defined by their pitch, start time and duration. The harmonic progressions are represented as harmonic trees. The time signature is $\frac{4}{4}$, however, this can be changed in the user scenario.

The *leitmotif* composition process can be provided by the user or automatically composed through a Large Language Model (LLM). For automated composition, we used Anthropic's Claude model through its API[2]. The LLM receives a prompt containing character information, roles, and narrative arc, and outputs melodic motifs in a structured format of pitch-duration tuples [(note, duration), ...] with explanations of musical decisions. In our implementation, each *leitmotif* spans two bars with a predefined duration sum. The *leitmotifs* are combined to compose melodies for the initial population of the evolutionary algorithm, with each plot atom corresponding to eight bars of music.

The harmonic structure is represented through trees built using the context-free grammar of tonal harmony [Rohrmeier, 2011]. To produce harmonic trees, we apply the rules at the phrase level (to start the generative process) and functional level (to characterize harmonic relationships on an abstract level with respect to functions and keys). The key is defined by the protagonist's *leitmotif*. At the functional level, only functional expansion rules are applied initially, while substitution and modulation rules are implemented through mutation operators in the evolutionary process. Regarding scale degree level and surface level, these are only applied to obtain the final harmonic progression. We use the grammar to obtain coherent harmonic progressions with the desired number of terminal nodes. In our implementation we set one terminal node per bar. Figure 4 shows an example of a harmonic tree, where terminal nodes correspond to individual bars.

Regarding the NSGA-II implementation, each individual in the population corresponds to a harmonic tree with melody. The evolutionary process uses three objective functions:

1. Tension Alignment: Evaluates the correlation between musical and narrative tension using the Tension Ribbons model. The musical tension curve is computed as the average of the three tension measurements. For each candidate solution, the musical tension is compared with the target narrative tension using Pearson correlation.

2. Tonal Coherence: Assesses the melodic-harmonic relationship through a weighted beat-based evaluation system. Notes are scored according to their metric position (weights: beat 1>beat 3>beat 2>beat 4>off-beats) and relationship to the current harmony. The scoring hierarchy prioritizes chord tones, followed by scale tones

[2] console.anthropic.com, model: claude-3-5-sonnet-20241022

without minor second conflicts with chord tones, then scale tones with conflicts, and finally non-scale tones.

3. Character Motif Recognition: Measures the preservation of character *leitmotifs* in the evolved melody using a sliding window approach. The evaluation considers both pitch-based features (intervals and contour) and rhythmic features (duration ratios), computing correlations between the original motifs and melodic segments. The four highest correlation values for each of the characters *leitmotif* are averaged for the final score.

Concerning genetic operators, crossover has a probability of $p_c = 0.9$ equally divided into three types:

- Melodic crossover: Single-point crossover on the melodic sequence;
- Harmonic crossover: Exchange of compatible harmonic subtrees (maintaining grammar validity);
- Combined crossover: Application of both melodic and harmonic crossover.

Mutation has a probability of $p_m = 0.2$ equally divided in two types:

- Melodic mutations (at the bar level): transposition ($p_{m-t} = 0.1$), augmentation ($p_{m-a} = 0.1$), diminution ($p_{m-d} = 0.1$), inversion ($p_{m-i} = 0.1$), retrograde ($p_{m-r} = 0.1$), intervallic alterations ($p_{m-ia} = 0.3$), and ornamentation ($p_{m-o} = 0.2$);
- Harmonic mutations: parallel transformations, where the parallel rules of the generative grammar are applied ($p_{m-pt} = 0.5$), and modulations to non-tonic regions, where non-tonic subtrees are modulated ($p_{m-m} = 0.5$).

The evolution process employs standard NSGA-II components including fast non-dominated sorting, crowding distance calculation, and tournament selection. The population size is 200 individuals and evolves for 200 generations. In the end of the evolutionary process, the final composition may be user-selected from Plot Atom Music Composer's Pareto front. In our case, 5 solutions are selected using a preference matrix [0.1, 0.6, 0.3] weighting tonal coherence, character motif recognition, and tension alignment, respectively. These selections are then combined following the narrative order, with the final combination chosen based on optimal alignment with the overall narrative tension curve using Pearson correlation.

5 Evaluation

The evaluation of the system has the purpose of understanding the potential for algorithmic composition from a narrative structure. Three different narrative structures were created based on folk tales archetypes identified by Propp [Propp, 1968]. The music for each was composed autonomously, and the authors chose from the final combinations. The compositions were played using a Virtual Instrument (piano) playing one chord per bar (using a spread voicing) and the melody.

Since we intend to evaluate the narrative potential of the system, we used Margulis *et al.* [2019] procedure to evaluate narrative listening of music. They introduce a Narrative Engagement scale to verify if narrative listening of the compositions is possible. Although the authors' study concluded that listening to music narratively can happen, it is important to note that this is a very complex and subjective task. The study's items to address music familiarity and enjoyment were also used. In addition, we added four items to evaluate the Narrative Alignment with the narrative structure used as input. The scales used were the same as in the study. The items used in our survey are:

- Story Response Question (yes/no): Did you imagine a story?
- Narrative Engagement (1-strongly disagree to 6-strongly agree):
  - "It was easy to imagine a story"
  - "I imagined a vivid story"
  - "I imagined a story with clear setting, characters, events"
  - "I imagined a story while listening, not after"
- Familiarity and Enjoyability (1-strongly disagree to 6-strongly agree):
  - "The music sounded familiar"
  - "I found the music enjoyable"
- Narrative Alignment (1-strongly disagree to 6-strongly agree):
  - "The music matches the story's emotional tone"
  - "The music reflects the story's pacing"
  - "Story events are represented musically"
  - "The overall musical structure follows the narrative arc"

The data and analysis is available at https://github.com/braga1376/algorithmic-composition-narrative-tension.

The surveys were conducted online, and each person responded to a survey for one of the compositions. We used the Allocate Monster [Fergusson, 2016] platform to randomize the survey allocation per person. There were a total of 73 participants (64.4% male, 34.2% female, 1.4% other) distributed in the three surveys ($n_1 = 25$, $n_2 = 18$, $n_3 = 30$). Participants were aged between 21-55 years ($\bar{X} = 29.4$, $S = 7.0$). The sample represented diverse musical backgrounds, from no formal training (31.5%) to professional (21.9%) . Most participants came from Engineering or Computer Science (49.3%) and Visual Arts, Music, or Design (31.0%) fields. We analyzed participants' responses across the three dimensions: Narrative Engagement, Familiarity and Enjoyability, and Narrative Alignment. For each dimension we used Cronbach's alpha to confirm internal consistency: Narrative Engagement, $\alpha = 0.871$; Narrative Alignment, $\alpha = 0.894$; Familiarity and Enjoyability, $\alpha = 0.515$. The high Cronbach's *alpha* values for Narrative Engagement and Alignment validated these as composite measures, whereas for Familiarity and Enjoyability, they did not, as such they were analyzed separately. No statistically significant relationship was found between these dimensions and age, musical background or field of work.

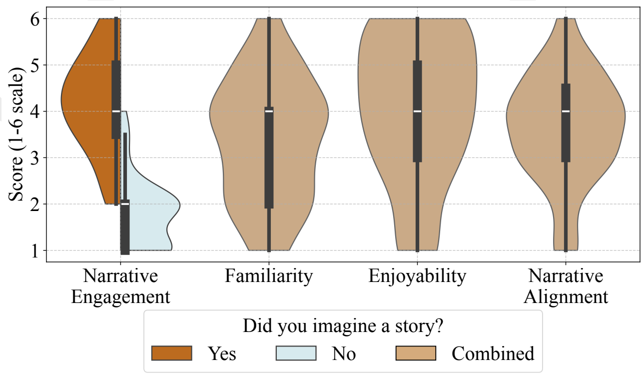

Figure 5: Scores for the dimensions Narrative Engagement, Familiarity, Enjoyability and Narrative Alignment

A Mann-Whitney U test revealed a significant relationship between the Story Response Question and the Narrative Engagement scores ($p < 0.001$, effect size $= 0.939$). Of the participants who reported imagining a story, 94% showed high narrative engagement (above the combined median value), while 87.5% of those who did not imagine a story showed low engagement (below the combined median value). Neither Narrative Alignment nor the individual measures of Familiarity and Enjoyability showed significant relationships with the Story Response Question. Figure 5 illustrates these distributions, with the split violin plot showing the distinct patterns of Narrative Engagement between participants who did and did not imagine stories, while other dimensions show more uniform distributions regardless of story response.

Considering the significant relationship between the Story Response Question and the Narrative Engagement, we can distinguish between participants who listened to the music narratively and those who did not. For those who imagined a story, the narrative engagement scores ($median = 4$, $IQR = 1.5$) suggest that the music successfully supported their narrative listening. It is also noteworthy that narrative listening did not affect familiarity, enjoyability and narrative alignment. Even when participants did not imagine a story, the scores indicate that they found the music familiar ($median = 4$, $IQR = 2$), enjoyable ($median = 4$, $IQR = 2$), and aligned with the original narrative ($median = 4$, $IQR = 1.5$).

6 Limitations

Our approach has several limitations that should be acknowledged. First, the system's output is composed of chord progression and melody. It lacks compositional decisions such as the chords' voicings or other more complex arrangement, instrumentation, and even dynamics. While this may be common in these systems [Herremans *et al.*, 2017; Carnovalini and Rodà, 2020], the absence of arrangement decisions affects aspects such as the narrative engagement and alignment, and even the overall perceived musical tension.

Second, the model of tonal tension focuses exclusively on harmonic aspects. Although the simplicity of the model is useful in our scenario, there are many factors that determine perceived tension besides the harmonic content, such as dynamics, timbre, repetition [Madsen and Fredrickson, 1993], phrase structure, note density [Krumhansl, 1996], tempo, and pitch height [Farbood, 2012].

Third, in evolutionary process, although the different operators are able to produce diverse compositions, the set of possible transformations is bounded by the defined set. This, of course, constraints the search space of the algorithm and might pose a challenge in the co-creation scenario since the user is limited to the predefined operators. These limitations present opportunities for future research to expand the system's capabilities.

7 Conclusions and Future Work

This paper has presented an algorithmic composition system that uses narrative structure and tension that integrates multiple computational approaches. By utilizing a hierarchical architecture that processes narrative elements at different scales - from plot schemas to individual characters - the system demonstrates the feasibility of translating narrative elements into musical form. The system achieves this through three key mechanisms: character representation through *leitmotifs*, maintenance of harmonic coherence through generative grammars, and optimization of musical tension alignment through evolutionary algorithms. This hybrid approach enables both autonomous composition and various degrees of user control in the co-creative process. The integration of these elements - narrative representation, musical tension, and harmonic structure - provides a framework for narrative-based musical composition. Although the current implementation focuses on the first compositional aspects, it establishes a novel approach on musical narratives.

Regarding future work, we have two main focuses: improving co-creation capabilities of the system and refining the current approach. For co-creation capabilities, the system can be extended as a tool by creating an interface for easy manipulation of elements and connecting it to either a Digital Audio Workstation or Musical Notation software as a plugin. This would allow composers to create their own narrative and compose music for it, with further arranging and editing capabilities. Since the system uses narrative representation as input, it could be adapted to compose music for different narrative domains, such as stories and films, by converting original narratives to the narrative representation used.

Regarding the current approach, several improvements are possible. Since many aspects of arrangements influence perceived tension, adding arrangement capabilities could enhance the results. The composition process, currently based on *leitmotifs* and their alterations to obtain the desired tension, could become repetitive when narrative segments are long. To address this, we could add non-motif phrases as infilling. The evaluation approach seems promising, and for future work we intend to use this evaluation method to compare the system to a baseline without guidance from the narrative structure, and with a random narrative structure. Finally, user control could be expanded by allowing selection of specific musical operators in the evolutionary algorithm.

Acknowledgments

This work was supported by FCT - Fundação para a Ciência e Tecnologia, I.P. by project reference and DOI identifier https://doi.org/10.54499/PRT/BD/154690/2023 and project UID/04516/NOVA Laboratory for Computer Science and Informatics (NOVA LINCS).

References

[Almén, 2003] Byron Almén. Narrative archetypes: A critique, theory, and method of narrative analysis. *Journal of Music Theory*, 47(1):1–39, 2003.

[Bernardes *et al.*, 2016] Gilberto Bernardes, Diogo Cocharro, Marcelo Caetano, Carlos Guedes, and Matthew EP Davies. A multi-level tonal interval space for modelling pitch relatedness and musical consonance. *Journal of New Music Research*, 45(4):281–294, 2016.

[Brown, 2012] Daniel Brown. Mezzo: An adaptive, real-time composition program for game soundtracks. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 8(4):68–72, 2012.

[Carnovalini and Rodà, 2020] Filippo Carnovalini and Antonio Rodà. Computational creativity and music generation systems: An introduction to the state of the art. *Frontiers in Artificial Intelligence*, 3:14, 2020.

[Casella and Paiva, 2001] Pietro Casella and Ana Paiva. Magenta: An architecture for real time automatic composition of background music. In *International Workshop on Intelligent Virtual Agents*, pages 224–232. Springer, 2001.

[Chew, 2002] Elaine Chew. The spiral array: An algorithm for determining key boundaries. In *International Conference on Music and Artificial Intelligence*, pages 18–31. Springer, 2002.

[Cui *et al.*, 2024] Wenqian Cui, Pedro Sarmento, and Mathieu Barthet. MoodLoopGP: Generating emotion-conditioned loop tablature music with multi-granular features. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*, pages 97–113. Springer, 2024.

[Davis and Mohammad, 2014] Hannah Davis and Saif Mohammad. Generating music from literature. In Association for Computational Linguistics, editor, *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL) @ EACL*, pages 1–10, April 2014.

[Deb *et al.*, 2002] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

[Déguernel *et al.*, 2022] Ken Déguernel, Mathieu Giraud, Richard Groult, and Sebastien Gulluni. Personalizing AI for co-creative music composition from melody to structure. In *Sound and Music Computing (SMC 2022)*, pages 314–321, 2022.

[Emmerich and Deutz, 2018] Michael TM Emmerich and André H Deutz. A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural computing*, 17:585–609, 2018.

[Ens and Pasquier, 2020] Jeff Ens and Philippe Pasquier. Flexible generation with the multi-track music machine. In *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR)*, pages 1–2, 2020.

[Farbood *et al.*, 2004] Morwaread M Farbood, Egon Pasztor, and Kevin Jennings. Hyperscore: a graphical sketchpad for novice composers. *IEEE Computer Graphics and Applications*, 24(1):50–54, 2004.

[Farbood, 2012] Morwaread M Farbood. A parametric, temporal model of musical tension. *Music Perception*, 29(4):387–428, 2012.

[Fergusson, 2016] A Fergusson. Designing online experiments using Google forms+ random redirect tool, 2016. Retrieved from https://teaching.statistics-is-awesome.org/designing-online-experiments-using-google-forms-random-redirect-tool.

[Fernández and Vico, 2013] Jose D Fernández and Francisco Vico. AI methods in algorithmic composition: A comprehensive survey. *Journal of Artificial Intelligence Research*, 48:513–582, 2013.

[Gervás and Méndez, 2024] Pablo Gervás and Gonzalo Méndez. Evolutionary plot arcs for a series of neurally-generated episodes. In *15th International Conference on Computational Creativity (ICCC 2024)*, 2024.

[Gervás, 2017] Pablo Gervás. Comparative evaluation of elementary plot generation procedures. In *Proceedings of the 6th International Workshop on Computational Creativity, Concept Invention, and General Intelligence*, 2017.

[Gervás, 2018a] Pablo Gervás. Storifying observed events: Could i dress this up as a story. In *5th AISB Symposium on Computational Creativity. AISB, AISB, University of Liverpool, UK*, 2018.

[Gervás, 2018b] Pablo Gervás. Targeted storyfying: Creating stories about particular events. In *9th International Conference on Computational Creativity (ICCC 2018)*, pages 232–239, 2018.

[Gervás, 2019] Pablo Gervás. Generating a search space of acceptable narrative plots. In *10th International Conference on Computational Creativity (ICCC 2019). UNC Charlotte, North Carolina, USA*, 2019.

[Guo *et al.*, 2020] Rui Guo, Ivor Simpson, Thor Magnusson, Chris Kiefer, and Dorien Herremans. A variational autoencoder for music generation controlled by tonal tension. *arXiv preprint arXiv:2010.06230*, 2020.

[Guo *et al.*, 2022] Rui Guo, Ivor Simpson, Chris Kiefer, Thor Magnusson, and Dorien Herremans. MusIAC: An extensible generative framework for music infilling applications with multi-level control. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*, pages 341–356. Springer, 2022.

[Herremans and Chew, 2017] Dorien Herremans and Elaine Chew. MorpheuS: generating structured music with constrained patterns and tension. *IEEE Transactions on Affective Computing*, 10(4):510–523, 2017.

[Herremans *et al.*, 2016] Dorien Herremans, Elaine Chew, et al. Tension ribbons: Quantifying and visualising tonal tension. In *Proceedings of the Second International Conference on Technologies for Music Notation and Representation (TENOR)*, 2016.

[Herremans *et al.*, 2017] Dorien Herremans, Ching-Hua Chuan, and Elaine Chew. A functional taxonomy of music generation systems. *ACM Computing Surveys (CSUR)*, 50(5):1–30, 2017.

[Jeong *et al.*, 2017] Jaehun Jeong, Yusung Kim, and Chang Wook Ahn. A multi-objective evolutionary approach to automatic melody generation. *Expert Systems with Applications*, 90:50–61, 2017.

[Jeong *et al.*, 2022] Jae-Hun Jeong, Eunbin Lee, Jong-Hyun Lee, and Chang Wook Ahn. Multi-objective deep network-based estimation of distribution algorithm for music composition. *IEEE Access*, 10:71973–71985, 2022.

[Krumhansl, 1996] Carol L Krumhansl. A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music perception*, 13(3):401–432, 1996.

[Lerdahl and Jackendoff, 1996] Fred Lerdahl and Ray S Jackendoff. *A Generative Theory of Tonal Music, reissue, with a new preface*. MIT Press, 1996.

[Lerdahl and Krumhansl, 2007] Fred Lerdahl and Carol L Krumhansl. Modeling tonal tension. *Music perception*, 24(4):329–366, 2007.

[Lerdahl, 2001] Fred Lerdahl. *Tonal pitch space*. Oxford University Press, 2001.

[Lopez and Alvaro, 2024] Duarte Lopez and Eduardo Alvaro. A progressive-adaptive music generator (PAMG): An approach to interactive procedural music for videogames. In *Proceedings of the 12th ACM SIGPLAN International Workshop on Functional Art, Music, Modelling, and Design*, pages 65–72, 2024.

[Louie *et al.*, 2020] Ryan Louie, Any Cohen, Cheng-Zhi Anna Huang, Michael Terry, and Carrie J Cai. Cococo: AI-steering tools for music novices co-creating with generative models. In *Joint Proceedings of the Workshops on Human-AI Co-Creation with Generative Models and User-Aware Conversational Agents co-located with 25th International Conference on Intelligent User Interfaces (IUI)*, 2020.

[Madsen and Fredrickson, 1993] Clifford K Madsen and William E Fredrickson. The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface. *Journal of Music Therapy*, 30(1):46–63, 1993.

[Margulis *et al.*, 2019] Elizabeth Hellmuth Margulis, Patrick CM Wong, Rhimmon Simchy-Gross, and J Devin McAuley. What the music said: Narrative listening across cultures. *Palgrave Communications*, 5(146), 2019.

[Margulis *et al.*, 2022] Elizabeth H Margulis, Patrick CM Wong, Cara Turnbull, Benjamin M Kubit, and J Devin McAuley. Narratives imagined in response to instrumental music reveal culture-bounded intersubjectivity. *Proceedings of the National Academy of Sciences*, 119(4):e2110406119, 2022.

[Maus, 1988] Fred Everett Maus. Music as drama. *Music theory spectrum*, 10(1):56–73, 1988.

[Micchi *et al.*, 2021] Gianluca Micchi, Louis Bigo, Mathieu Giraud, Richard Groult, and Florence Levé. I keep counting: An experiment in human/AI co-creative songwriting. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 4(1):263–275, 2021.

[Navarro-Cáceres *et al.*, 2019] María Navarro-Cáceres, Marcelo Caetano, Gilberto Bernardes, and Leandro Nunes de Castro. ChordAIS: An assistive system for the generation of chord progressions with an artificial immune system. *Swarm and Evolutionary Computation*, 50:100543, 2019.

[Navarro-Cáceres *et al.*, 2020] María Navarro-Cáceres, Marcelo Caetano, Gilberto Bernardes, Mercedes Sánchez-Barba, and Javier Merchán Sánchez-Jara. A computational model of tonal tension profile of chord progressions in the tonal interval space. *Entropy*, 22(11):1291, 2020.

[Prechtl *et al.*, 2014] Anthony Prechtl, Robin Laney, Alistair Willis, and Robert Samuels. Algorithmic music as intelligent game music. In *Proceedings of the 50th Annual Convention of the AISB*, London, UK, April 2014.

[Propp, 1968] Vladimir Propp. *Morphology of the Folktale*. University of Texas Press, 1968.

[Rohrmeier, 2011] Martin Rohrmeier. Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1):35–53, 2011.

[Ruiz-Marcos, 2021] Germán Ruiz-Marcos. *Tension-driven automatic music generation*. PhD thesis, Open University (United Kingdom), 2021.

[Tchemeube *et al.*, 2023] Renaud Bougueng Tchemeube, Jeffrey Ens, Cale Plut, Philippe Pasquier, Maryam Safi, Yvan Grabit, and Jean-Baptiste Rolland. Evaluating human-AI interaction via usability, user experience and acceptance measures for MMM-C: A creative AI system for music composition. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI '23, pages 5769–5778, 2023.

[Zacharakis *et al.*, 2021] Asterios Zacharakis, Maximos Kaliakatsos-Papakostas, Stamatia Kalaitzidou, and Emilios Cambouropoulos. Evaluating human-computer co-creative processes in music: A case study on the CHAMELEON melodic harmonizer. *Frontiers in Psychology*, 12:603752, 2021.