

Towards an Online Unsupervised Framework for Accessible Biosignal-Based Co-Creative Musical Performance (Draft)

First Author

Second Author

Received: date / Accepted: December 11, 2025

Abstract Artistic co-creation offers a mode of expression grounded in shared presence, yet many people lack access to such participation due to technological, neurological, or cognitive constraints. Existing assistive systems support symbolic communication but rarely provide entry into the expressive, non-symbolic domain where aesthetic presence emerges through generative interaction. This gap is particularly significant for individuals with disabilities, for whom artistic co-creation can uniquely provide access to lived presence. Motivated by this experiential and ontological challenge, we develop an online, unsupervised optimization framework that transforms biosignals such as EEG or eye movements into meaningful musical contributions.

Our framework treats biosignals as expressive traces of presence and supports three modes of participation: (1) **structured co-creation**, in which participants modulate expressive parameters within predefined musical scores; (2) **semi-structured co-creation**, in which participants interact with a provided non-vocal musical background to generate higher-level musical structure; and (3) **unstructured co-creation**, in which musical form emerges entirely from collaborative biosignal dynamics without any predefined score or background.

To achieve online adaptability, interpretability, and real-time robustness, we combine a physics-informed musical dynamical system with a control-theoretic rendering process and an unsupervised online optimization procedure incorporating a presentness-informed, interpretable noise objective.

We contribute a philosophical and methodological foundation for accessible co-creation integrating relational ontology, embodied expression, and the generative emergence of musical structures. As an initial and ethically responsible step, all empirical evaluation will be conducted with healthy adult participants, with studies involving disabled participants planned as future work once methodological stability and safeguards are secured.

Keywords biosignal-driven creation, co-creative performance, relational ontology, accessible artistic expression, musical structure emergence

1 Introduction

Human existence is ontologically relational, a premise supported by post-structuralist philosophy and contemporary metaphysical analysis grounded in natural science. Under this relational ontological foundation, the human desire for spiritual satisfaction, as defined in need hierarchy theory [4], emerges as a phenomenon arising from the relational existence of human ontology and its mediation through symbolic structures. Individuals identify their presence through generative creation processes. Generative co-creation thus serves as an important approach for humans to engage with their spiritual and existential dimensions, allowing subjectivity to emerge and presence to be identified. In short, the generative co-creation of artistic works is intimately tied to questions of existence, being, and presence, which is also a mode of spiritual and experiential satisfaction, enabling individuals to participate in a shared field of expression that extends beyond symbolic action.

Accessible generative artistic co-creation holds value across populations. For healthy persons, it enables (1) participation in artistic practice without requiring years of technical training, (2) access to non-symbolic expressive domains beyond conventional symbolic structures, and (3) deeper relational engagement where one's physiological and emotional presence actively shapes shared aesthetic experience. For persons with disabilities, the significance runs deeper: they face fundamental difficulties in naturally entering both symbolic relational structures and *the Real* [3, 2] dimension of relational ontology, which manifests as challenges in perceiving and recognizing their own subjectivity, existence, and presence. Accessible generative co-creation provides a pathway through which their existence and presence can become

recognized by themselves, enabling entry into relational structures where spiritual satisfaction emerges. Beyond this ontological function, it also offers opportunities for aesthetic contribution and participation in shared creative spaces. Given its philosophical significance and practical value for existential recognition, inclusiveness, and cultural participation, this work aims to build a framework for biosignal-driven co-creative musical performance that treats biological signals as expressive traces of lived presence rather than control commands.

One promising approach to accessible generative artistic co-creation is the use of biosignals such as EEG, eye movements, or MEG. This work addresses two essential questions regarding biosignal-based artistic participation: (1) **Structured artistic co-creation**: How can participants engage in structured singing activities, such as choral performances of designated songs? and (2) **Unstructured co-creation**: How can participants engage in the creation and performance of entirely new musical works, where structure itself emerges from the collaborative process?

However, to function flexibly and effectively for both healthy and disabled participants, a viable solution must satisfy several requirements: (1) real-time responsiveness, (2) online adaptability, (3) robustness to noise and variability, (4) interpretability of the generative process, (5) unsupervised or self-supervised learning capability, and (6) no pre-creation training burden on participants.

Facing these challenges, this work proposes an online, unsupervised optimization-based framework that transforms biosignals into meaningful contributions to musical performance. The framework integrates a physics-informed musical dynamical system with a control-theoretic rendering process and an interpretable noise-integrated objective function that reflects participants' embodied presence. Rather than decoding intentions or relying on extensive calibration, the method enables generative co-creation through continuous adaptation to biosignal dynamics.

To ensure ethical and methodological soundness, empirical evaluation will first be conducted with healthy adult participants, establishing stability, robustness, and interpretability before extending the system to studies involving disabled participants in future work.

This framework has the potential to provide an accessible foundation for generative artistic co-creation, enhancing contemporary understandings of existence, subjectivity, and presence in a philosophical sense, while also enabling disabled individuals to satisfy their spiritual desires and to have more opportunities to participate fully in the relational structures of society in both practical and social-functional terms.

2 Background and Related Work

2.1 Philosophical Studies

Philosophical investigations of co-creation emphasize the relational and existential dimensions of human interaction. Phenomenology and relational ontology (e.g., Buber, Merleau-Ponty) frame human experience as fundamentally situated within shared presence, while existentialist and post-structuralist thinkers (Heidegger, Sartre, Lacan, Deleuze) highlight contingency as the locus through which existence and subjectivity become perceptible. Contemporary aesthetics also situate generative and participatory artistic practices as modalities of embodied expression beyond symbolic representation. These studies motivate our treatment of biosignals as traces of lived presence and support a framework in which generative musical structures emerge through relational and contingent interaction rather than symbolic intention.

2.2 Neuroscience Basis for Music

Neuroscientific studies [1] show that music perception relies on basic acoustic features: pitch, loudness, and timbre, which are processed in the auditory cortex and integrated across reward, motor, and cognitive networks. Contemporary predictive-processing accounts propose that the brain continuously learns statistical regularities in musical sequences and generates expectations about upcoming events. Violations of these expectations produce prediction-error responses observable in EEG and MEG, linking musical syntax directly to probabilistic neural dynamics. This view motivates our use of EEG-derived structures as interpretable generators of musical form.

2.3 EEG to MIDI Algorithms

Existing EEG-to-MIDI and brain-computer musical interfaces focus primarily on mapping neural activity to musical parameters or discrete control states. Early systems employed spectral features or threshold-based event detection to trigger musical notes or control synthesis parameters. More recent work explores continuous mapping from EEG-derived affective states to musical dynamics, or classification-based control of musical patterns. While these systems provide mechanisms for translating brain activity into musical output, they generally treat EEG as a control source rather than as a generator of musical structure. This distinction motivates our approach, which incorporates biosignals not merely as controllers but as constitutive elements of syntactic and expressive musical emergence.

2.4 Syntax Structures in EEG

EEG microstates and related neural state models provide interpretable symbolic representations of brain activity, segmenting neural dynamics into quasi-stable topographies with well-defined transition properties. Microstate sequences exhibit grammar-like characteristics, including transition matrices, dwell-time distributions, and recurrent motifs. Prior research has linked microstate dynamics to cognitive and perceptual processing, suggesting their role as functional building blocks of thought. Online microstate estimation methods allow real-time tracking of neural state transitions, making them suitable for interactive systems. These properties make microstates a promising basis for syntax generation in unstructured co-creation, enabling EEG-derived symbolic units to drive musical form while maintaining interpretability.

Previous works [1] indicates that the EEG present hierachical structures when utilizing a frequency-dividing based method, which is analogously to the musical syntax structure. And the music appear hierachical and non-adjoin dependence. The hierachical and non-adjoin dependence is simialr to the syntax structure of a structure produced by a language system.

2.5 Probabilistic Context-Free Grammars

Probabilistic context-free grammars (PCFGs) provide a principled framework for describing hierarchical structure in symbolic sequences and have been extensively applied in computational music theory. A PCFG is defined as $G = (V, \Sigma, R, S, P)$, where V denotes non-terminal symbols corresponding to abstract musical categories such as phrases, chords, or melodic contours, Σ denotes terminal symbols representing concrete musical events, R is a set of production rules, S is a designated start symbol, and P assigns probabilities to rule applications. PCFGs capture stylistic regularities and long-range dependencies in musical corpora, enabling the generation of musical sequences that preserve both local coherence and global form. Prior work has used PCFGs to model harmonic progression, phrase structure, motivic development, and rhythmic organization, demonstrating their effectiveness as generative priors in computational composition and music analysis.

In the context of this work, the PCFG serves as a learned musical grammar that provides structural guidance for mapping EEG-derived neural representations into the musical note space. By using the grammar as a probabilistic prior, the system can generate musically plausible structures while allowing real-time neural dynamics to shape the unfolding syntax.

2.6 MIDI Rendering

Real-time MIDI rendering research encompasses expressive performance modeling, timbre control, and parameter mapping frameworks. Traditional models such as rule-based performance systems and expressive timing algorithms replicate features of human interpretation. More recent neural or hybrid systems modulate articulation, dynamics, and timbre through continuous control signals. In accessibility and interactive music contexts, physiological signals have been used to influence emotional character, intensity, or filter parameters in synthesis engines. These approaches inform our semantic rendering layer, where biosignals modulate expressive dimensions of a predefined or emergent score without requiring explicit motor action or musical training.

Table 1: Research Question Taxonomy

Category	RQ ID	Research Question
Structured Co-creation	RQ-S1	How can biosignals modulate expressive musical parameters while preserving a predefined MIDI structure?
Structured Co-creation	RQ-S2	How can interpretable contingency be introduced into syntactic structure to manifest human presence while maintaining musical intelligibility?
Semi-Structured Co-creation	RQ-SS1	How can biosignals generate melodic content that harmonically and rhythmically aligns with existing background instrumental tracks?
Semi-Structured Co-creation	RQ-SS2	How can event-related potentials inform structural decisions in melody generation over background music?
Unstructured Co-creation	RQ-U1	How can biosignal dynamics generate coherent musical structures without predefined templates in real time?
Unstructured Co-creation	RQ-U2	How can multiple participants' biosignals interact to produce emergent collective musical forms?
Biosignal-to-Music Mapping	RQ-M1	Which biosignal features are most effective as interpretable controllers for musical syntax and semantics?
Biosignal-to-Music Mapping	RQ-M2	How should biosignal-to-music mappings be designed to ensure interpretability, stability, and real-time responsiveness?
Evaluation	RQ-V1	How can we evaluate the structural coherence, expressive quality, and relational presence in biosignal-driven co-creation?
Evaluation	RQ-V2	How do different mapping strategies affect user experience, perceived agency, and aesthetic value?
Philosophical Foundations	RQ-P1	What role does contingency play in connecting human existence with symbolic musical structure?
Philosophical Foundations	RQ-P2	How does biosignal-driven co-creation balance symbolic order with existential presence?

2.7 Discussion

Across existing lines of research, two limitations emerge with respect to the requirements of accessible, presence-centered artistic co-creation. First, most biosignal-to-music systems rely on predefined mappings or classifier-based control, treating EEG as a source of discrete commands or continuous modulation. Such systems do not extract or generate musical syntax, and therefore cannot support unstructured co-creation in which musical form must emerge from the participant's physiological presence. Second, contemporary expressive rendering and music generation frameworks depend heavily on pretrained deep models whose latent structures are neither interpretable nor directly coupled to neural dynamics. These models impose externally learned musical grammars, making them unsuitable for participants who cannot reliably produce intentional control signals or who require transparent, traceable mappings between internal states and artistic output. Previous work on Neuroscience basis for music indicates that musical structure emergence and processing need to be analyzed from brain network and dynamics of particular part of the brain.

To address gaps and integrated neuroscience basis for music, our framework adopts a different approach: we extract an interpretable, naturally occurring syntax from EEG microstate dynamics and use this neural syntax as the generative basis for musical structure. Microstates provide symbolic units, dwell-time patterns, and transition probabilities that function analogously to a grammar, enabling the construction of musical-like structures without the need for external training data or intention-driven control. This choice ensures that musical form emerges from participants' lived presence rather than from prelearned models, satisfying the requirements of interpretability, accessibility, and ontological grounding that guide our design philosophy.

3 Research Questions

Research questions are listed in Table 1.

4 Framework Overview

4.1 Structured Co-creation

4.1.1 Design Philosophy and Objectives

Structured co-creation addresses a fundamental question: how can individuals participate meaningfully in the performance of an existing musical work when the structural framework is already established? In traditional musical performance, both the syntactic structure and the semantic structure are determined by the composer and performer. The syntactic structure refers to the temporal organization of notes, rhythms, and phrases. The semantic structure refers to the expressive interpretation, timbre, and emotional coloring. For accessible co-creation, we must preserve the recognizability of the work while opening space for participants to contribute their lived presence.

Our design philosophy separates musical structure into syntactic and semantic dimensions. In structured co-creation, the syntactic structure is already presented through a predefined MIDI score, which ensures musical coherence and recognizability. However, the semantic structure must be provided by humans in being. The semantic structure comprises the expressive qualities that bring the score to life. This division allows participants to shape how the music is experienced without requiring them to generate its underlying structure.

The goal of structured co-creation is therefore twofold. First, to maintain the identity and intelligibility of the original musical work. Second, to create channels through which participants' physiological and existential presence can manifest audibly within that work. Our approach operates along these two parallel dimensions: we introduce controlled contingency into the syntactic layer while mapping biosignals to expressive parameters in the semantic layer. This dual-layer design allows biosignals to shape musical expression while preserving the recognizability and coherence of the predefined musical work.

4.1.2 Syntactic Layer: Introducing Contingency into Musical Structure

The syntactic layer addresses the temporal and structural organization of musical events. Given a predefined MIDI score that establishes the fundamental musical structure of melody, harmony, and rhythm, we introduce controlled syntactic contingency through biosignal-driven perturbations.

This design choice is grounded in existentialist philosophy: contingency allows human existence to become perceptible. It is precisely through contingent events, those small, unpredictable fluctuations that slip through the rigidity of structure, that we become aware of our own existence as human beings. This is also why certain existentialist thinkers emphasize perception, openness, and gratitude toward the contingencies of everyday life. Contingency provides a way to temporarily step outside structure and return to the Real, thereby witnessing one's own existence.

By injecting unpredictable yet bounded variations into timing, note onset, duration, and structural transitions, we create space for the participant's presence to emerge within the musical fabric. Specifically, biosignals modulate temporal micro-variations such as small deviations in note timing and duration, rhythmic elasticity such as dynamic tempo fluctuations within musically acceptable ranges, and structural transitions such as probabilistic choices at phrase boundaries or repeat structures. These perturbations remain constrained to preserve musical intelligibility while allowing the music to breathe with the participant's physiological state.

4.1.3 Semantic Layer: Biosignal-Driven Expressive Control

The semantic layer addresses the expressive and affective qualities of musical rendering. While the syntactic structure is given, the semantic interpretation emerges from human presence and interpretation. The semantic interpretation determines how the music feels, its emotional character, and its timbral quality. In traditional performance, a musician's emotional state, physical gesture, and artistic intention shape these expressive dimensions. In our framework, we enable participants to contribute to this semantic layer through biosignal-driven control.

Here, biosignals function as continuous controllers for timbre, emotional coloring, intensity, and dynamics. Rather than treating biosignals as discrete commands, we map their continuous variation to expressive parameters in the MIDI rendering engine. Timbre modulation allows biosignal features to control synthesis parameters, filter characteristics, or sample selection. Emotional trajectory is shaped by

arousal and valence dimensions derived from biosignals, which influence the emotional contour of the performance. Dynamic intensity responds to physiological variation through adjustments to loudness, attack velocity, and articulation. Harmonic coloring involves subtle modifications to harmonicity, brightness, or spectral envelope.

This semantic mapping ensures that the participant’s internal state becomes audible, transforming invisible physiological presence into shared aesthetic experience. The participant does not need to consciously control these parameters; rather, their mere presence and physiological variation naturally shape the music’s expressive qualities.

4.1.4 Integration and Real-time Rendering

The syntactic and semantic layers operate concurrently in real time. A template MIDI score provides the structural backbone, while biosignal-driven processes continuously modulate both its syntactic realization and semantic expression. The result is a performance that maintains recognizable musical identity while bearing the contingent traces of human presence.

4.2 Unstructured Co-creation

4.2.1 Design Philosophy and Objectives

Unstructured co-creation addresses a more ambitious question: how can musical structure itself emerge from collaborative biosignal dynamics when no predefined template exists? Unlike structured co-creation, where a MIDI score provides the syntactic backbone and participants shape its expressive realization, unstructured co-creation requires both syntactic structure and semantic expression to emerge from biosignals. This represents a fundamentally generative mode of artistic participation where the musical work comes into being through the relational presence of participants.

The objective of unstructured co-creation is to enable participants to collectively generate novel musical works without requiring compositional expertise or predetermined frameworks. This mode of participation aligns closely with the philosophical premise of relational ontology: the musical structure itself emerges from the coexistence of participants, making their shared presence the generative source of the artwork rather than merely its expressive medium.

However, this ambition introduces a significant technical challenge at an earlier stage than in structured co-creation. The primary challenge lies in syntactic structure generation: transforming continuous biosignals into coherent, structured MIDI sequences in real time. Once syntactic structure is generated, the semantic rendering problem remains essentially the same as in structured co-creation. If we successfully generate a coherent MIDI sequence from biosignals, we can directly apply the semantic layer mapping developed for structured co-creation to shape its expressive realization. Therefore, the major distinction in unstructured co-creation concerns biosignal-to-MIDI generation rather than MIDI rendering.

4.2.2 Syntactic Layer: Biosignal-to-MIDI Generation

The syntactic layer in unstructured co-creation must generate discrete, structured musical sequences from continuous, high-dimensional biosignals in real time. Existing research on EEG-to-MIDI generation has developed various algorithms for this purpose, including machine learning approaches that learn mappings from neural activity to musical patterns. However, most existing methods require extensive prior training data and operate as black-box models, making them unsuitable for our application context.

Our design requires three critical properties. First, interpretability: the mapping from biosignal features to musical events must be transparent and understandable, allowing us to comprehend how participants’ physiological states shape musical structure. Second, real-time operation: the system must generate musical output with minimal latency, allowing participants to perceive their influence on the emerging structure and enabling genuine interactive co-creation. Third, unsupervised adaptation: the algorithm must function without participant-specific training data, instead adapting online to the characteristics of incoming biosignals.

We therefore develop an interpretable, real-time biosignal-to-MIDI generation algorithm that operates without prior training. Our approach draws on principles from existing EEG-to-MIDI research while adapting them to satisfy our constraints. The generation process operates in several stages. First, we extract interpretable features from raw biosignals, focusing on characteristics such as spectral power in specific frequency bands, entropy measures, hemispheric asymmetry, and temporal dynamics. Second, we

map these features to musical parameters including pitch, duration, onset timing, and velocity through deterministic yet adaptive mapping functions. Third, we apply musical constraints to ensure the generated sequences satisfy basic principles of tonal coherence, rhythmic regularity, and phrase structure.

The adaptation mechanism allows the system to adjust its mapping parameters online based on the statistical properties of incoming biosignals, ensuring that the generated music responds meaningfully to each participant’s unique physiological patterns without requiring prior calibration.

4.2.3 Semantic Layer: Expressive Rendering of Generated Structure

Once the syntactic layer generates a coherent MIDI sequence from biosignals, the semantic layer addresses its expressive realization. Here, we apply the same biosignal-driven expressive control framework developed for structured co-creation. Biosignals continue to function as continuous controllers for timbre, emotional coloring, intensity, and dynamics, mapping their variation to expressive parameters in the MIDI rendering engine.

This semantic mapping operates identically to structured co-creation: timbre modulation allows biosignal features to control synthesis parameters, emotional trajectory is shaped by arousal and valence dimensions, dynamic intensity responds to physiological variation, and harmonic coloring involves subtle modifications to spectral characteristics. The key difference is that in unstructured co-creation, both the notes being played and how they are expressed emerge from biosignals, creating a fully generative musical experience.

4.2.4 Integration and Emergent Musical Structure

The syntactic and semantic layers operate concurrently in real time. Unlike structured co-creation where a template MIDI score provides stability, unstructured co-creation produces entirely emergent musical structures. The biosignal-to-MIDI generation continuously produces new musical events based on participants’ physiological states, while the semantic rendering simultaneously shapes their expressive qualities. The result is a collaborative sonic environment that has never existed before and cannot be precisely replicated, embodying the contingency and relational presence that define artistic co-creation in its fullest sense.

4.3 Semi-Structured Co-creation

4.3.1 Design Philosophy and Objectives

Semi-structured co-creation occupies an intermediate position between structured and unstructured modes. It addresses the question: how can participants contribute meaningful vocal or melodic content when a background instrumental track already exists? Unlike structured co-creation where both accompaniment and melody are predetermined, or unstructured co-creation where all musical elements emerge from biosignals, semi-structured co-creation involves a complete instrumental arrangement that provides harmonic, rhythmic, and timbral context, while the primary melodic or vocal content emerges from biosignal dynamics.

The objective of semi-structured co-creation is to enable participants to “sing along” or generate melodic content over existing musical backing tracks without requiring vocal training or compositional expertise. The background track establishes key, tempo, harmonic progression, and stylistic character, constraining the generative process to produce musically coherent results while allowing participants’ presence to shape the melodic content. This mode is particularly valuable for karaoke-like scenarios, collaborative performances where some participants provide instrumental backing while others contribute biosignal-driven vocals, or therapeutic settings where familiar musical contexts support creative participation.

The technical challenge lies in generating melodic content that harmonically and rhythmically aligns with the background track while reflecting participants’ biosignal dynamics. Additionally, event-related potentials can provide moment-to-moment insights into participants’ cognitive and attentional states, informing decisions about phrase boundaries, melodic contour changes, or moments of expressive emphasis.

4.3.2 Syntactic Layer: Background-Aware Biosignal-to-MIDI Generation

The syntactic layer in semi-structured co-creation must generate melodic sequences that align with the harmonic and rhythmic structure of the background track. This requires extracting structural information from the background audio and using it to constrain the biosignal-to-MIDI generation process.

Our approach extends the generation framework developed for unstructured co-creation by incorporating real-time analysis of the background track. The generation process operates as follows. First, we analyze the background track to extract key, tempo, chord progressions, metric structure, and phrase boundaries through audio feature extraction and music information retrieval techniques. Second, we extract interpretable features from biosignals including spectral power, entropy measures, hemispheric asymmetry, and event-related potentials that signal cognitive events such as attention shifts or surprise responses. Third, we map biosignal features to melodic parameters such as pitch, duration, and onset timing through adaptive functions that respect the harmonic and rhythmic constraints derived from the background track.

Event-related potentials provide particularly valuable information for melodic structure decisions. ERP components responses can indicate moments of heightened attention, suggesting appropriate locations for melodic peaks, phrase boundaries, or expressive ornaments. By incorporating ERP tracking alongside continuous biosignal features, we create melodic content that responds not only to participants' general physiological state but also to their momentary cognitive engagement with the musical context.

The constraint mechanism ensures that generated pitches belong to appropriate scales or chord tones, that rhythmic placement aligns with the metric grid, and that phrase structure respects the formal organization of the background track.

4.3.3 Semantic Layer: Expressive Rendering of Generated Melody

Once the syntactic layer generates a melodic sequence aligned with the background track, the semantic layer addresses its expressive realization. Here, we apply the biosignal-driven expressive control framework developed for structured co-creation. Biosignals continue to function as continuous controllers for timbre, vocal quality, intensity, and dynamics.

For semi-structured co-creation, semantic control is particularly important for vocal synthesis or melodic instrument selection. Biosignals shape parameters such as vocal timbre, breathiness, vibrato depth and rate, articulation style, and dynamic envelope. These expressive qualities allow the generated melody to carry traces of the participant's emotional and physiological state, creating a sense of authentic vocal presence even when the melodic content emerges from biosignal-to-MIDI generation rather than actual vocalization.

4.3.4 Integration with Background Track

The syntactic and semantic layers operate concurrently in real time, synchronized with the background instrumental track. The biosignal-to-MIDI generation continuously produces melodic content constrained by the harmonic and rhythmic structure of the backing music, while the semantic rendering shapes its expressive qualities. The result is a complete musical performance where the instrumental foundation provides stability and stylistic coherence while the biosignal-driven melody contributes contingency, presence, and relational expressivity.

4.4 EEG to Lyric

We propose a three-stage EEG-to-Lyric pipeline: (1) biosignals are mapped to a semantic trajectory via microstates, oscillations, and ERP responses; (2) a language model generates raw free-form text guided by this trajectory; and (3) a lyric-oriented LLM restructures the text into musically meaningful lyrics aligned with either a provided background track (semi-structured) or EEG dynamics alone (unstructured). This enables participants to co-create lyrics through presence rather than intentional symbolic control.

5 Method

5.1 EEG Phase Space Reconstruction

We first construct an online representation of EEG dynamics that can serve as a basis for musical syntax generation. Instead of relying on fully offline analysis, both microstate estimation and spectral characterization are performed in an online fashion. For each incoming time window, we apply an online k -means procedure to the scalp potential maps, updating cluster centroids iteratively and assigning each time point to a microstate label. This yields a discrete sequence of microstate indices that captures quasi-stable spatial configurations of brain activity in real time.

In parallel, we compute band-limited power for each electrode over a set of pre-defined frequency bands (e.g., θ , α , β , γ) within the same sliding window. The resulting spectral features are discretized into a small number of levels, producing a compact code that summarizes the dominant frequency pattern at each electrode and time t . By combining the online microstate index with the discretized frequency pattern across electrodes, we obtain a symbolic representation in a *pre-note space*, that is, a space whose elements are structured neural symbols intended to be mapped onto musical notes. This pre-note representation evolves in real time and forms the input to the subsequent musical syntax generation stage.

5.2 Musical Syntax Structure Generation

We adopt a generative-grammar-based mapping and real-time dynamic alignment approach for musical syntax structure generation. Specifically, given the online neural representation in pre-note space, we generate musical syntax by mapping these neural symbols into the real note space $\mathcal{N}_o^{88}$, corresponding to the 88-key MIDI pitch domain. For unstructured co-creation, no predefined score is available, and musical structure must be constructed directly from the evolving neural codes. To achieve this, we employ an online optimization procedure that aligns the sequence of pre-note symbols with candidate musical structures generated by a pretrained probabilistic context-free grammar (PCFG).

The PCFG is trained offline on a corpus of target musical pieces, learning probabilistic production rules that encode idiomatic melodic, harmonic, and rhythmic patterns. During interaction, the grammar proposes short candidate continuations in the space of symbolic musical structures, while the online optimizer selects and adapts these candidates so that their note sequences in $\mathcal{N}_o^{88}$ remain consistent with the current neural pre-note representation. In effect, the PCFG provides a learned musical prior, and the online optimization forces its realizations to track the dynamics of the EEG-derived symbols. This yields a stream of musical syntax that is both stylistically grounded by the pretrained grammar and continuously shaped by the participant’s neural activity in real time. For structured co-creation, where a score already specifies the syntactic backbone, this generative step is skipped and the neural representation is instead used only to modulate and perturb the existing structure.

5.3 Syntactic Contingency Injection

To introduce expressiveness while preserving musical coherence, we inject controlled stochasticity into the generated MIDI representation $\mathcal{N}_o^{88}$, which denotes the standard 88-key MIDI pitch space. The final MIDI output is produced by modifying the syntactic MIDI sequence $\mathbf{m}_{\text{syntax}}$ with a contingency term shaped by EEG dynamics,

$$\mathbf{m}_{\text{final}} = \mathbf{m}_{\text{syntax}} + \epsilon \cdot \mathcal{C}(\Phi(t)),$$

where ϵ is a small scaling factor and $\mathcal{C}$ is a mapping from EEG features to interpretable perturbations. These adjustments include micro-variations in timing that mimic humanization, fluctuations in velocity influenced by band-power changes, small pitch ornamentations driven by fast neural components, and subtle modifications of harmonic tension associated with curvature in the reconstructed phase space. Perturbations are restricted to musically meaningful transformations, ensuring that contingency contributes to expressiveness without compromising structural intelligibility.

5.4 MIDI Real-time Rendering with EEG as Controller

Once a musical structure is obtained—either from a predefined MIDI score or from grammar-based generation—it is rendered into audio using an existing pre-trained neural MIDI-to-audio synthesis model.

In this work, the renderer follows a WaveNet- or diffusion-based architecture that models the conditional distribution of the audio waveform given MIDI events and a set of high-level control variables. Formally, the waveform distribution is expressed as

$$p(\mathbf{y} \mid \text{MIDI}, \mathbf{z}) = \prod_{t=1}^T p(y_t \mid y_{<t}, \text{MIDI}_t, \mathbf{z}_t),$$

where $\mathbf{z}_t$ denotes latent synthesis controls such as spectral brightness, dynamic intensity, timbral shaping, and spatialization. Because the renderer is pre-trained and stable, the central methodological question reduces to determining how EEG activity should be mapped into this control space in real time.

To achieve this, we define an online mapping function

$$\mathbf{z}_t = f_{\theta(t)}(\Phi(t)),$$

where $\Phi(t)$ is the real-time EEG phase space representation and $\theta(t)$ are parameters adapted continuously during performance. The adaptation of $\theta(t)$ is governed by an online optimization objective that minimizes the mismatch between the rendered acoustic properties and the target expressive trajectories implied by EEG dynamics. This is formalized as

$$\theta(t+1) = \arg \min_{\theta} \mathcal{L}_{\text{ctrl}}(S(\mathbf{y}_{1:t}), g(\Phi_{1:t})),$$

in which $S(\cdot)$ extracts acoustic descriptors from the synthesized audio and $g(\cdot)$ computes target descriptors derived from EEG features such as oscillatory power, hemispheric asymmetry, microstate transition rates, and phase space curvature. Through this formulation, spectral brightness tends to follow fluctuations in alpha and beta band power, dynamic range becomes sensitive to gamma activity or microstate stability, spatialization evolves with hemispheric imbalance, and temporal articulation responds to geometric properties of the reconstructed EEG trajectory.

Because only the mapping function is optimized online while the renderer itself remains fixed, the system maintains stable and efficient operation with end-to-end latency well below perceptual thresholds, typically under 100 ms. This design produces a direct neural-to-acoustic coupling in which the participant’s evolving physiological state is continuously and audibly reflected, resulting in a coherent, expressive, and deeply embodied musical performance.

5.5 Lyric Generation

For vocal or choral settings, we extend the framework to produce lyrics driven by EEG semantics. The process begins with an encoder that maps EEG phase space features $\Phi(t)$ to a semantic representation $\mathbf{h}_{\text{sem}}$,

$$\mathbf{h}_{\text{sem}} = f_{\text{enc}}(\Phi; \theta_{\text{enc}}).$$

The encoded features reflect neural correlates of emotional valence, cognitive load, and attentional state, and are then used to condition a large language model for lyric generation. The final lyric output is obtained by

$$\text{Lyrics}^* = \text{LLM}(\mathbf{h}_{\text{sem}}, \text{constraints}),$$

where the constraints enforce alignment between syllable structure and melodic rhythm, adherence to rhyme or prosodic patterns, coherence with the musical mood, and compatibility with phonetic flow. The LLM refines the text iteratively to ensure both semantic alignment with EEG-derived states and structural compatibility with the generated music.

5.6 Quantitative Evaluation Metrics

5.6.1 Technical Coherence Metrics

Technical evaluation focuses on whether the system produces musically coherent outputs across structured, semi-structured, and unstructured co-creation modes. In structured co-creation, coherence is assessed by measuring the degree of adherence to the original MIDI template. This includes the percentage of notes that match the predefined score, the mean absolute deviation between generated and template note onsets, and the proportion of pitches that remain harmonically consistent with the original key.

Recognizability is also evaluated through blind identification tests in which listeners determine whether the performance corresponds to the template piece.

In semi-structured settings, coherence is examined through consonance between generated melodies and background harmonies, rhythmic synchronization measured via cross-correlation between melodic onsets and the underlying beat structure, and the alignment of phrase boundaries with structural markers in the background accompaniment. These metrics quantify how effectively a participant’s neural activity guides musical variation while remaining embedded within an existing musical framework.

For unstructured co-creation, where no template exists, coherence is evaluated through emergent structural properties. Tonal stability is estimated using key-profile strength derived from the Krumhansl–Schmuckler method, rhythmic regularity is quantified through auto-correlation of onset interval patterns, and transition smoothness is measured by examining pitch and timing intervals between successive notes. These metrics assess whether purely biosignal-driven musical output develops recognizable musical syntax.

5.6.2 Information-Theoretic Presence Metrics

To confirm that biosignals genuinely shape musical output rather than producing incidental correlations, information-theoretic analyses are employed. Mutual information between biosignal features and musical parameters is compared against the mutual information between musical output and random surrogate signals. A significant gap between these values demonstrates that neural activity contributes non-trivial structure to the music. Directed influence is further assessed through transfer entropy from biosignals to musical output, providing evidence that physiological states causally drive generative decisions rather than merely co-varying with them. These measures collectively establish the presence of an authentic neural-to-musical coupling.

5.6.3 Listener Perception Metrics

Perceptual evaluation examines whether biosignal-driven performances convey qualities associated with human presence. Naive listeners compare biosignal-driven pieces with algorithmic control pieces matched for structural complexity but devoid of physiological modulation. Participants provide ratings on perceived liveness, expressiveness, and the extent to which the music feels “inhabited” or shaped by a human presence. A discrimination test further evaluates whether listeners can distinguish biosignal-driven performances from algorithmic ones with accuracy above chance. High perceptual discriminability would indicate that biosignal influence produces distinctive aesthetic signatures that listeners interpret as traces of presence.

5.7 Participant Experience Questionnaires

Subjective experience is assessed through post-session questionnaires addressing comfort, perceived agency, and the sense of co-presence with the generated music. Participants report whether they felt that their internal state influenced the musical outcome, whether the system supported meaningful engagement, and whether the experience facilitated feelings of expression or embodiment. These subjective measures complement quantitative metrics by capturing dimensions of presence and agency that are phenomenological rather than structural.

Appendix

A Implementation Plan

B Collaboration Roles

The collaboration structure is organized around the major parties involved in the project, their functional roles, and the responsibilities associated with each role. Table 2 summarizes the expected collaboration plan.

Party	Role	Responsibilities
Core Research Team	Technical Lead	Algorithm development; biosignal processing; real-time system design; implementation of structured and unstructured co-creation modules.
Neuroscience / Physiology Partner	Signal Science Advisor	EEG/eye-tracking protocol design; data collection; neural interpretation; ensuring compliance with safety and ethical standards.
Music / Artistic Partner	Artistic Director	Design of musical mappings; shaping performance aesthetics; curating artistic demonstrations; evaluating expressive quality.
Accessibility / Clinical Partner	Inclusion and Safety Lead	Providing expertise on disabled participants' needs; guiding accessible interface design; ensuring safe testing conditions.
Evaluation and UX Team	Evaluation Lead	Design and execution of user studies; quantitative and qualitative assessment; usability and accessibility evaluation.
Philosophy / Theory Contributor	Conceptual Framework Lead	Developing and refining the relational ontology model; articulating presence and contingency; guiding theoretical interpretation.
Project Coordination Office	Project Manager	Timeline planning; communication among partners; documentation; dissemination of publications, demos, and social outcomes.

Table 2: Collaboration roles and responsibilities across the project.

C Ethical Considerations

This project engages directly with human participants and therefore requires careful attention to ethical, psychological, and social dimensions of participation. Our framework is designed to support accessibility and inclusion, yet biosignal-based artistic co-creation involves risks that must be explicitly addressed.

First, biosignals such as EEG and eye movements can reveal sensitive information about cognitive or emotional states. All data collection will follow strict privacy protocols, with secure storage, anonymization, and explicitly limited data use. Participants will receive clear explanations of what biosignals can and cannot reveal, and no diagnostic or evaluative interpretations will be performed.

Second, to ensure safety and ethical responsibility, all empirical evaluation in the initial phase will be conducted solely with healthy adult participants. These studies allow us to establish the stability, interpretability, and robustness of the system before involving disabled participants. Only after methodological validation and ethical board approval will we extend the study to populations with disabilities, ensuring their well-being and preventing psychological or emotional distress.

Third, the co-creative process must respect participants' autonomy and expressive agency. Because the system transforms biosignals into audible artistic output, participants may perceive the music as reflecting their inner state. We therefore implement mechanisms for participants to pause, withdraw, or revise the system's mapping at any time. We also avoid any framing that implies medical assessment or psychological judgment of participants.

Finally, we emphasize relational ethics: co-creation is not simply a technical interaction but a form of shared presence. The research team will adopt sensitive communicative practices, recognize power asymmetries between researchers and participants, and ensure that the co-creative environment affirms participants' subjectivity and dignity. These commitments support both scientific integrity and the broader human values underlying accessible artistic co-creation.

D Expected Outcomes

The outcomes of this project can be organized into three complementary domains: formal technical outputs, philosophical contributions, and social-practical impacts. Each category is summarized below.

D.1 Formal Outcomes

The formal outcomes concern the methodological and technical artifacts generated through the development of the proposed framework. These outputs include algorithms, software prototypes, datasets, and documentation that support reproducibility and future research. A summary of these technical contributions is provided in Table 3.

D.2 Philosophical Outcomes

Beyond technical contributions, this project aims to develop a conceptual foundation for biosignal-driven co-creation grounded in relational ontology, contingency, and embodied presence. These contributions clarify how generative artistic systems can serve as modes of existential expression and human-technology relation. The key philosophical outcomes are summarized in Table 4.

Outcome Type	Description
Algorithms	Online microstate-based syntax generation; interpretable noise-integrated optimization; biosignal-to-expression mapping.
Software Prototypes	Real-time co-creation system supporting structured and unstructured modes.
Datasets	Ethically collected EEG and eye-tracking datasets suitable for reproducible research.
Technical Framework	Unified architecture for accessible, real-time generative music performance.
Documentation	Implementation protocols, design notes, and reproducible pipelines.

Table 3: Summary of formal (technical) outcomes produced by the project.

Contribution Area	Description
Relational Ontology	Operationalizing existence, subjectivity, and presence through biosignal-mediated co-creation.
Contingency and the Real	Showing how controlled contingency allows human presence to manifest within generative systems.
Embodiment Theory	Demonstrating how biosignals mediate embodied engagement beyond symbolic or motor action.
Aesthetics of Co-creation	Developing a presence-centered aesthetic paradigm grounded in generative interaction.
Conceptual Framework	A philosophical model linking neural microstates, structure emergence, and human presence.

Table 4: Philosophical outcomes highlighting conceptual and theoretical contributions.

Outcome Domain	Description
Accessibility	Enabling disabled individuals to participate in musical creation without motor skill or training requirements.
Expressive Agency	Providing new channels for emotional expression, spiritual satisfaction, and recognition of presence.
Relational Integration	Facilitating fuller engagement in social, cultural, and relational structures.
Community Applications	Potential use in inclusive arts programs, rehabilitation settings, and collaborative performances.
Cultural Impact	Supporting inclusive creative ecosystems and expanding participation in artistic expression.

Table 5: Social and practical outcomes demonstrating broader impact and accessibility.

D.3 Social and Practical Outcomes

The project also aims to create measurable social benefits by enabling inclusive artistic participation and supporting expressive agency, particularly for individuals with disabilities. These outcomes strengthen cultural and relational integration while expanding opportunities for community engagement. The main social and practical outcomes are summarized in Table 5.

These outcomes collectively provide technological innovation, conceptual clarity, and meaningful social impact, establishing a foundation for further research, artistic exploration, and inclusive technology development.

References

1. Daikoku, T.: Music and brain. *The Brain Neural Networks* **29**(3), 135–147 (2022). DOI 10.3902/jnns.29.135. URL <http://dx.doi.org/10.3902/jnns.29.135>
2. Lacan, J.: *The Four Fundamental Concepts of Psychoanalysis*. W. W. Norton & Company, New York (1977). Original seminars delivered 1963–1964; published in French 1973
3. Lacan, J.: *Écrits: A Selection*. W. W. Norton & Company, New York (1977). Original work published 1966
4. Maslow, A.H.: A theory of human motivation. *Psychological Review* **50**(4), 370–396 (1943)