Towards a Generative Textural Approach to Computer-Assisted Composition

Dr. Micael Antunes, MSc, PhD
Aix-Marseille Univ., CNRS, PRISM, LIS, InCIAM, Marseille, France.
e-mail: antunes@prism.cnrs.fr

Prof. Dr. Ricard Marxer, MSc, PhD
Toulon Univ., CNRS, LIS, InCIAM, Toulon, France.

Prof. Dr. Vincent Tiffon, MSc, PhD
Aix-Marseille Univ., CNRS, PRISM, InCIAM, Marseille, France.

Abstract

This article presents *TEMA-S* project, which integrates a generative model for symbolic music into a computer-assisted composition (CAC) workflow and evaluates it through a Practice-as-Research (PaR) framework that links analytical assessment to compositional use. The study probes a pretrained chord–texture disentanglement VAE for symbolic music via a focused compositional task: a composition that traces a pathway between Ernesto Nazareth's *Escorregando* and Paganini's *Andantino in C Major* compositions. The workflow couples score-level inspection and symbolic representations with latent-space operations followed by compositional integration of selected outputs. Evaluation combines notated excerpts with targeted listening/reading tasks. Results show rhythmic and textural coherence and stable accompaniment patterns, alongside known limitations in melodic reconstruction and the need for explicit harmonic control, especially at cadences. The contribution is twofold: (i) a compact PaR-based protocol for assessing interpretable generative models within CAC workflows; and (ii) evidence—specific to the chord–texture disentanglement VAE—of practical compositional affordances (ostinatos, transitional passages, hybrid textures) alongside clearly delimited constraints that motivate future development.

1. Introduction

Recent deep-learning architectures (e.g., VAEs, diffusion, LLM models) have expanded the palette of tools for symbolic music generation. Within this landscape, the variational autoencoder (VAE) is a probabilistic dimensionality-reduction model whose continuous latent space supports generation and the transformation of musical data. When coupled with inductive bias (following Wang [1]) via representation, conditioning, or architectural choices, the VAE can be aligned with musically meaningful factors, enabling high-level control over distinct musical dimensions—harmony (pitch-class relations and vertical organisation) and texture (density, diversity, and inter-strata relations).

Despite this promise, it remains underexplored how such high-level controls can be appropriated in compositional practice. Standard probes—UMAP/t-SNE visualisations, reconstruction metrics, and KL balance—diagnose latent structure but do not establish whether the model affords steerable, reliable textural and harmonic control in creative workflows. Quantitative structure does not automatically translate into creative utility.

This article reports on *TEMA-S* (Texture-based Environment for Assisted Symbolic Musical Composition), an interdisciplinary effort to develop CAC [2] strategies grounded in VAEs. This exploratory study employs PaR methods [3] anchored in compositional work. The creative process and its artefacts are treated as evidence for assessing VAE behaviour, guiding subsequent design choices (e.g., adaptations and fine-tuning) and informing the derivation of composition heuristics from the model. The broader project will ultimately deliver a Python-based CAC assistant prototype; the present article focuses on the first methodological framework and its exploratory evaluation.

This article is organised as follows. Section 2 situates the study within related work and the conceptual background, outlining CAC context (2.1), the notion of musical texture (2.2), and interpretable generative models for symbolic music (2.3). Section 3 details the PaR methodology. Section 4 reports the composition and evaluation, first describing the compositional process (4.1) and then presenting analytical outputs (4.2) across Sampling (4.2.1), Projection (4.2.2), Morphing (4.2.3), and Compositional Integration (4.2.4). Section 5 discusses findings and sketches future steps. Section 6 synthesises contributions, limitations, and implications for CAC.

2. Related Work and Conceptual Background

2.1 Computer-Assisted Composition

Computer-assisted composition (CAC) can be understood, following Bresson [4], as the study and design of programming tools, languages, and formalisms that let composers represent and compute musical structures (scores, harmonic frameworks, rhythms, and sounds). Historically, the use of computers to support musical composition can be traced to Lejaren Hiller's algorithmic experiments on the Illiac computer [5] and, soon after, to Iannis Xenakis's stochastic procedures [6]. The field was then consolidated at IRCAM—first with the CRIME group and later within the Music Representations team—through interactive, symbolic environments such as PatchWork and OpenMusic [4]. Across these trajectories, symbolic representation and modelling remained the working substrate. Today, CAC spans a continuum from rule-based systems

(OpenMusic [2], Bach [7]) to transformer-driven generators (ComposerX [8], MIDI-GPT [9]) and production-oriented plugins (e.g., Captain Plugins [10]) that operationalise melodic, harmonic, and rhythmic procedures.

Bresson [4] highlights that CAC can be understood through several polarities that shape both its technical implementation and its creative use. The first lies between *extensionnel* and *intensionnel* approaches: in the former, musical information is specified explicitly as a list of discrete events, while in the latter it is produced through a sequence of procedures or algorithms that generate those events. The second polarity contrasts *signal-based* and *symbolic domains*, distinguishing between interaction with audio signals and the manipulation of score-level or MIDI data. Finally, CAC systems may operate in *real-time* or *off-line* modes, reflecting the degree of computational responsiveness and the nature of interaction afforded to the composer.

This article examines a texture-focused use of a VAE as a CAC tool. The model provides a learned, symbolic latent space that supports procedural navigation while preserving score-level analysis and control. The approach complements rule-based CAC by translating procedural logics into latent structures without sacrificing editability or musical transparency.

2.2 Musical Texture

In symbolic music, texture can be conceived as the emergent relation among constituent parts (voices, registers, rhythmic strata). Beyond any one representation, prior work has treated texture through statistical characterisation—e.g., density (simultaneous note events and registral spread), diversity (variety of pitch classes, rhythms, timbres/instrumental parts), and inter-relation (synchrony, imitation, accompaniment coupling) [11]. Yet texture is not reducible to summary statistics: it also evokes categorical regimes widely used in musical practice and analysis, from monody to melody-and-accompaniment, polyphony/imitative counterpoint, heterophony, and sound-mass/cluster writing.

Datasets and analytical studies reflecting this view of texture span multiple settings. For chamber music repertoire, curated corpora often emphasise melody-and-accompaniment layouts, enabling voice-aware representations and evaluations that separate lead lines from supportive strata [12], [13]. Orchestral studies examine textural behaviours across sections [14], while polyphonic corpora foreground multi-voice coordination and contrapuntal dependencies [15].

This work aligns with the literature but centres on the melody–accompaniment regime, enabling controlled comparisons between VAE-generated scores and recompositions under a shared texture: a salient melody with one or more accompaniment layers whose density, diversity, and interrelation can be measured and qualitatively assessed. Drawing on prior insights into textural features and voice-aware representations, it evaluates the VAE's affordances for texture control in CAC.

2.3 Interpretable Generative Models for Symbolic Music

The generative model employed in this research is a pre-trained Variational Autoencoder (VAE) [1]. It consists of an *encoder* that maps musical data into a latent space and a *decoder* that reconstructs the data from it. This formulation allows the latent space to act as a navigable landscape of stylistic possibilities, where musical ideas can be smoothly transformed and hybridised.

Building on this principle, Wang et al. [1] proposed a VAE designed for interpretable and controllable polyphonic music generation. The model disentangles two complementary musical representations: (1) a chord embedding (z_{chd}) representing the harmonic progression, and (2) a texture embedding (z_{txt}) describing surface organisation—rhythm, register, and note density. This inductive bias allows the independent manipulation of harmony and texture, enabling re-harmonisation, style transfer, among other strategies for musical generation.

Architecturally, the system contains **two encoders** and a **shared decoder**. The *chord encoder* uses a rule-based harmonic analysis to extract symbolic progressions from a chroma vector representation, processed by a bidirectional GRU to produce the latent variable z_{chd}. The *texture encoder* interprets a quasi piano-roll, processed by convolutional layers with chord-invariant kernels followed by another GRU to produce z_{txt}. The concatenated latent vector [z_{chd}, z_{txt}] is passed to a **decoder** based on the PianoTree VAE [16], trained jointly with the encoders using a standard variational objective that balances reconstruction fidelity and regularisation[1]. The model was trained using the POP909 dataset [17].

The PianoTree decoder mirrors the hierarchical syntax of polyphonic music. From the global latent vector, a GRU first generates temporal embeddings (one per sixteenth-note). Each embedding expands into simultaneous-note events, and each note is decoded into pitch and duration attributes.

In the context of CAC, this framework positions the VAE as an exploratory interface, operationalised through three compositional transformations, following [18]: **sampling**, generating textures from an encoded distribution; **projection**, mixing characteristics of the input samples; and **morphing**, creating intermediates between two samples. In this setting, the latent space becomes a navigable field where harmonic and textural qualities can be interpolated and blended, turning VAE representation into tangible creative control.

3. Methodology: Practice-as-Research

This research follows a PaR framework, in which the creative practice itself constitutes a form of inquiry. In line with Nelson's definition — that PaR "involves a research project in which practice is a key method of inquiry and where a practice is submitted as substantial

[1] It is beyond the scope of this paper to detail the full VAE architecture; comprehensive descriptions are available in [1], [16].

evidence of a research inquiry" [3] — the present study uses compositional experimentation with a Variational Autoencoder (VAE) as both method and site of knowledge production. The approach combines creative generation, documentation of process, and critical reflection, forming an investigation in which theory is imbricated within practice.

The research inquiry guiding this work is exploratory rather than hypothesis-driven: *in what ways might the latent space of a VAE afford new forms of textural representation and compositional control within CAC?*

The methodological design integrates three complementary dimensions. **(1) Product:** symbolic scores generated by the VAE and the scores resulting from their recompositional transformation constitute the primary evidence of the inquiry. **(2) Documentation of process:** each stage of musical generation, and compositional decision-making is recorded through code logs, annotated sketches, and textual notes, capturing moments of discovery and critical reflection. **(3) Complementary writing:** analytical commentary situates the praxis within broader debates in CAC and relating emergent insights to conceptual and musicological frameworks. Accordingly, the analytical process is organised into **pre-compositional** and **post-compositional** stages.

In the **pre-compositional** stage, the model is assessed as follow: **(1) Sampling**: the generated excerpt is compared to its original input in terms of (a) melodic fidelity, (b) preservation of harmonic characteristics, (c) rhythmic stability, and (d) maintenance of the melody-and-accompaniment relationship. **(2) Projection**: the model's ability to combine and transfer features across the harmonic and textural latent dimensions is examined. **(3) Interpolation**: interpolated excerpts are analysed to verify whether they retain textural attributes of both source materials, indicating coherent blending.

In the **post-compositional** stage, these outputs are compared vis-à-vis the resulting human-composed scores. The analysis narrates how the generated material was creatively appropriated within the composition process, revealing both the model's potential and limitations as a partner in musical decision-making.

4. Composition and Evaluation

We propose a composition exercise in which a brief miniature recomposes through quotation to trace a continuous path between two pre-existing works. The sources are Ernesto Nazareth (1863–1934), *Escorregando* (1923), Figure 1, and Niccolò Paganini (1782–1840), *Andantino in C Major*, MS. 97 (c. 1800–1810), Figure 2.

The formal plan of the compositional exercise is A–B–C–D–A, with **A** quoting Nazareth and **C** quoting Paganini; sections **B** and **D** are **newly composed transitions** whose role is to achieve a smooth rapprochement between these two musical idioms.

Figure 1 – Opening four bars of Ernesto Nazareth (1863–1934), Escorregando (1923).

Figure 2 - Opening four bars of Niccolò Paganini (1782–1840), Andantino in C Major.

The exercise emphasises textural manipulation so that the quotations remain recognisable while the joins are perceptually continuous. The outcome is a compositional miniature in which the transitions demonstrate how texture can mediate stylistic distance without erasing identity.

4.1 The compositional process

The composition workflow proceeds as follows. Selected musical passages are first engraved in a score editor and exported as MIDI. These files are then converted into the model's inputs—chroma for harmonic content and a quasi-piano-roll for textural information—which are encoded by the VAE to enable latent-space operations such as **sampling**, **projection**, and **morphing** [18]. The transformed latent are decoded back to symbolic form, re-exported as MIDI, and rendered as notation. A dedicated Python routine performs quantisation to ensure a clean symbolic output; however, targeted engraving edits are still applied during composition to secure rhythmic legibility, harmonic consistency, and idiomatic notation.

Each encoded work yields two latent representations—z_{chd} for harmony and z_{txt} for texture—which structure the subsequent compositional operations. In **sampling**, the original pairs (z_{chd}, z_{txt}) are decoded to obtain reconstructions of the source excerpts. In **projection**, cross-pairs are formed (e.g., $z_{chd}^{(A)}$ with $z_{txt}^{(B)}$) to recombine harmonic and textural attributes and probe mixed characteristics. **Morphing** is implemented as linear interpolation in texture-latent space. Given two texture embeddings $z_{txt}^{(A)}$ and $z_{txt}^{(B)}$ in $\mathbb{R}^d$, we form a path $z_{txt}(\alpha) = (1-\alpha)\cdot z_{txt}^{(A)} + \alpha\cdot z_{txt}^{(B)}$ with $\alpha \in [0,1]$, discretised into 10 evenly spaced steps (including both endpoints). Harmony is held fixed (z_{chd} constant), so only textural attributes evolve along this path. Because the decoder is non-linear, a linear path in latent space does not guarantee linear changes in the music information; nonetheless, with an approximately Gaussian, isotropic VAE prior it typically produces smooth, perceptually coherent transitions.

4.2 Analytical outputs

Given the space available, the analysis centres on the reconstruction of a four-bar excerpt by Ernesto Nazareth and on selected projection and morphing derived from it. The discussion details the criteria used to choose morphing trajectories and shows how the resultant materials informed the design of Transition B within the composition[2].

4.2.1 Sampling

[2] The complete composition, additional examples, and source code associated with this project can be accessed at: https://gitlab.com/micael_antunes/GenArts2025Examples.

Figure 3 - Reconstruction from Escorregando sample.

By looking at the sampling output *Figure 3*), we note the following aspects: **Harmony.** Functional harmony is preserved: the chordal progression of the original excerpt is maintained. **Melody.** This is the least coherent aspect: the model fails to capture the original descending melodic pattern of the first 2 bars. **Rhythm.** Largely consistent with the source; despite small variants, onsets align closely enough to sustain the original rhythmic feel. **Texture.** The melody–accompaniment layering is retained, chiefly because the accompaniment ostinati and the melodic rhythmic profile are preserved.

4.2.2 Projection

The projection sample evidences the model's ability to place *Andantino's* harmony within the texture of *Escorregando*, yielding coherent harmony across the first two bars in A minor. In the last two bars, a harmonic ambiguity emerges: the accompaniment outlines an E chord with a major third (G♯) while the principal melody implies a minor third (G♮). As in the sampling results, weaknesses in melodic reconstruction persist, whereas rhythmic representation and the preservation of the melody–accompaniment relationship remain stable.

Figure 4 - Projection output: score combining texture from the *Escorregando* sample with harmony from the *Andantino* sample.

4.2.3 Morphing

Figure 5 – Morphing, sample 3.

Figure 6 – Morphing, sample 7.

The morphing retains salient traits of the source excerpts while exhibiting transitional features. In morph 3 (*Figure 5*), the rhythmic profile of *Escorregando* is preserved, yet the original ostinato (dotted line) are interleaved with new figures and rests; chordal density also decreases, moving from three-note sonorities to dyads or single tones. Morph 7 (*Figure 6*) combines variations of the *Escorregando* ostinato rhythm (indicated by the arrows' trajectories) while, at the close, introducing the low register ostinato of *Andantino* (circled).

4.2.4 Compositional Integration

Figure 7 - Section B of the composition—the transition between Nazareth's Escorregando and Paganini's Andantino.

In Transition B (*Figure 7*), the melodic material from morphings 3 and 7 is appropriated almost intact, as highlighted by the dashed boxes in the score. The black triangles indicate the adapted notes introduced to ensure harmonic coherence and the cadence that marks the end of the section. From the accompaniment perspective, the passage draws on samples from morph 03 (first four bars) and morph 07 (last four bars). This configuration enables a smooth transition between Nazareth's *Escorregando* and Paganini's *Andantino* through two mechanisms: (i) a gradual reduction in note and rhythmic density, and (ii) a shift in accompaniment patterning—from syncopated figures associated with *Escorregando* to a steadier crotchet pulse characteristic of the Andantino.

5. Discussion and future steps

The chose VAE model exhibits limitations in reconstruction fidelity, most notably in capturing melodic complexity: descending contours, longer-span dependencies, and ornamental detail are frequently simplified or misrepresented. By contrast, rhythmic and textural dimensions are rendered with greater coherence; onset placement, surface continuity, and the melody-and-accompaniment layering are consistently preserved, indicating that the latent structure better supports time–surface regularities than pitch-contour nuance. In the composition exercise, this textural strength proved practically valuable: morphing in texture space yielded intermediate materials that retained salient traits of both Nazareth and Paganini while remaining malleable. These hybrids could be appropriated to design new transitional structures without erasing stylistic identity. Taken together, the results suggest a division of labour in which textural transformations serve as reliable scaffolds for form and continuity, while melodic detail may require targeted post-editing or model adaptation to achieve comparable specificity.

The Practice-as-Research design adopted here can be extended into longer, longitudinal studies that stage targeted creative exercises. These exercises should probe specific affordances of the VAE—texture continuity, rhythmic coherence, and cross-source blending—under varied materials and constraints, yielding reproducible protocols and artefacts.

Building on that formative base, consolidation requires user-based evaluation with an interactive interface that operationalises those heuristics as explicit controls. This phase should test perceived agency, melodic fidelity, and textural coherence across diverse workflows, converting exploratory insights into dependable technique.

On the computational side, model adaptation is required to strengthen melodic control. Priorities include retraining or fine-tuning to improve melodic reconstruction and editability, enabling finer adjustments of contour and cadence. This would expand the space for melodic transformation while maintaining harmonic coherence.

The operations explored—sampling, projection, and morphing—should be formalised as compositional heuristics. Each operation needs parameter ranges, constraints, and strategies that translate latent moves into predictable musical outcomes. This formalisation will clarify when and how texture-led navigation supports specific compositional goals.

6. Conclusion

This study presented an exploratory use of a VAE for CAC anchored in musical texture. Under a PaR methodology, composition functioned as the evaluative lens, clarifying both affordances and limits. Results indicated stable rhythmic and textural coherence and productive latent-space operations for creative reworking (notably morphing), yielding material that can be shaped into new structures. Weaknesses in melodic reconstruction and the need for active oversight to secure harmonic cadence remained evident.

These outcomes motivate two near-term directions: heuristics for texture-led navigation and curation in latent space, and a concrete pathway for tool redesign and model adaptation informed by practice. More broadly, the findings argue that texture-oriented representation encoded by the proposed VAE model can deliver controllable generation while preserving compositional agency. In this frame, texture is not merely a descriptor but a handle for action; the model stays an instrument, and VAE representational space remains the site that affords musical creativity.

Acknowledgment

Micael Antunes acknowledge *InCIAM* institute, through which he has received funding from the Excellence Initiative of Aix-Marseille University – A*Midex, a French “Investissements d’Avenir programme” AMX-19-IET-005.

References

[1] Z. Wang, D. Wang, Y. Zhang, and G. Xia, “Learning Interpretable Representation for Controllable Polyphonic Music Generation,” Aug. 17, 2020, *arXiv*: arXiv:2008.07122. doi: 10.48550/arXiv.2008.07122.

[2] G. Assayag, C. Rueda, M. Laurson, C. Agon, and O. Delerue, “Computer-assisted composition at IRCAM: From PatchWork to OpenMusic,” *Computer music journal*, vol. 23, no. 3, pp. 59–72, 1999.

[3] R. Nelson, *Practice As Research in the Arts*. London, England: PALGRAVE MACMILLAN, 2013.

[4] J. Bresson, “Composition assistée par ordinateur: techniques et outils de programmation visuelle pour la création musicale,” PhD Thesis, Université Pierre et Marie Curie, Paris, France, 2017.

[5] L. A. Hiller and L. M. Isaacson, *Experimental Music; Composition with an electronic computer*. Greenwood Publishing Group Inc., 1979. Accessed: Sep. 29, 2025. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/578548

[6] I. Xenakis, *Formalized music: thought and mathematics in*

composition. Pendragon Press, 1992.
[7] D. Ghisi and A. Agostini, “Extending bach: A Family of Libraries for Real-time Computer-assisted Composition in Max,” *Journal of New Music Research*, vol. 46, no. 1, pp. 34–53, Jan. 2017, doi: 10.1080/09298215.2016.1236823.
[8] Q. Deng *et al.*, “ComposerX: Multi-Agent Symbolic Music Composition with LLMs,” Apr. 30, 2024, *arXiv*: arXiv:2404.18081. doi: 10.48550/arXiv.2404.18081.
[9] P. Pasquier *et al.*, “MIDI-GPT: A Controllable Generative Model for Computer-Assisted Multitrack Music Composition,” Feb. 04, 2025, *arXiv*: arXiv:2501.17011. doi: 10.48550/arXiv.2501.17011.
[10] “Captain Plugins from Mixed In Key - Music Composition Software / VST.” Accessed: Sep. 29, 2025. [Online]. Available: https://mixedinkey.com/captain-plugins/
[11] W. Berry, *Structural functions in music*. in Dover books on music, music history. New York: Dover, 1987.
[12] L. Couturier, L. Bigo, and F. Levé, “Annotating Symbolic Texture in Piano Music: a Formal Syntax”.
[13] L. Soum-Fontez, M. Giraud, N. Guiomard-Kagan, and F. Levé, “Symbolic Textural Features and Melody/Accompaniment Detection in String Quartets,” in *Proceedings of the International Symposium on Computer Music Multidisciplinary Research 2021*, Tokyo, 2021, pp. 175–184.
[14] D.-V.-T. Le, M. Giraud, F. Levé, and F. Maccarini, “A Corpus Describing Orchestral Texture in First Movements of Classical and Early-Romantic Symphonies,” in *9th International Conference on Digital Libraries for Musicology*, Prague Czech Republic: ACM, Jul. 2022, pp. 27–35. doi: 10.1145/3543882.3543884.
[15] M. Giraud, R. Groult, E. Leguy, and F. Levé, “Computational fugue analysis,” *Computer Music Journal*, vol. 39, no. 2, pp. 77–96, 2015.
[16] Z. Wang *et al.*, “PIANOTREE VAE: Structured Representation Learning for Polyphonic Music,” Aug. 17, 2020, *arXiv*: arXiv:2008.07118. doi: 10.48550/arXiv.2008.07118.
[17] Z. Wang *et al.*, “POP909: A Pop-song Dataset for Music Arrangement Generation,” Aug. 17, 2020, *arXiv*: arXiv:2008.07142. Accessed: Nov. 20, 2024. [Online]. Available: http://arxiv.org/abs/2008.07142
[18] I. Simon, A. Roberts, C. Raffel, J. Engel, C. Hawthorne, and D. Eck, “Learning a Latent Space of Multitrack Measures,” Jun. 01, 2018, *arXiv*: arXiv:1806.00195. doi: 10.48550/arXiv.1806.00195.