LATENT MORPHOLOGIES: A POST-DIGITAL INSTRUMENT FOR SELF-SUPERVISED AUDIO RECOMPOSITION

LATENT MORPHOLOGIES: UNO STRUMENTO POST-DIGITALE PER LA RICOMPOSIZIONE AUDIO AUTO-SUPERVISIONATA

SHAI COHEN

Abstract (IT): I sistemi contemporanei di musica con IA oscillano tra due paradigmi limitanti: modelli generativi opachi che nascondono i loro processi decisionali musicali e strumenti interattivi in tempo reale che privilegiano l'immediatezza rispetto alla profondità strutturale. Questo articolo propone Latent Morphologies come modello alternativo di liuteria post-digitale: un quadro trasparente e auto-supervisionato per la ricomposizione audio, nel quale la forma musicale emerge dalle relazioni morfologiche intrinseche del materiale sonoro. Combinando l'analisi-risintesi all'interno di Praat AudioTools con processi Python leggeri per l'estrazione di caratteristiche, il clustering, l'apprendimento di rappresentazioni latenti e la generazione di percorsi, il sistema tratta le rappresentazioni analitiche come materia compositiva modificabile, e non come sottoprodotti computazionali nascosti. Gli eventi sonori vengono segmentati, incorporati, raggruppati, navigati e ricomposti attraverso tre operazioni interrelate: scoperta, navigazione e mutazione. Il quadro non utilizza corpora esterni né checkpoint opachi; le trasformazioni restano tracciabili rispetto ai materiali analizzati e ai parametri definiti dal compositore. L'articolo sostiene che Latent Morphologies mette in discussione paradigmi musicali ereditati ridefinendo lo strumento come ambiente analitico-computazionale distribuito, riformulando l'IA come attante verificabile anziché generatore autonomo, e spostando la forma musicale da una sintassi imposta dall'esterno verso una morfologia sonora latente. Situando questa pratica nei discorsi sulla liuteria digitale e sul suono post-digitale, l'articolo presenta un modello di composizione assistita dall'IA fondato su trasparenza, modificabilità e tempo compositivo profondo.
Parole chiave: morfologia sonora; liuteria post-digitale; attante IA; ricomposizione audio; tempo compositivo profondo.

Abstract (EN): Contemporary AI music systems oscillate between two limiting paradigms: opaque generative models that conceal their musical decision-making, and real-time interactive tools that privilege immediacy over structural depth. This article proposes Latent Morphologies as an alternative model of post-digital lutherie: a transparent, self-supervised framework for audio recomposition in which musical form emerges from the intrinsic morphological relations of sound material itself. Combining analysis-resynthesis within Praat AudioTools with lightweight Python processes for feature extraction, clustering, latent representation learning, and path generation, the system treats analytical representations as editable compositional matter rather than hidden computational by-products. Events are segmented, embedded, clustered, navigated, and recomposed through three interrelated operations: discovery, navigation, and mutation. The framework uses no external corpus and no opaque checkpoint; its transformations remain traceable to the analyzed sound materials and composer-defined parameters. The article argues that Latent Morphologies challenges inherited musical paradigms in three ways: it redefines the instrument as a distributed analytical-computational environment rather than a fixed performance interface; it reframes AI as an auditable actant rather than an autonomous generator; and it shifts musical form away from externally imposed syntax toward latent sonic morphology. By situating this practice within discourses of digital lutherie and post-digital sound, the article presents a model of AI-assisted composition grounded in transparency, editability, and compositional deep time.
Keywords: sonic morphology; post-digital lutherie; AI actant; audio recomposition; compositional deep time.

LATENT MORPHOLOGIES: A POST-DIGITAL INSTRUMENT FOR SELF-SUPERVISED AUDIO RECOMPOSITION

SHAI COHEN

Introduction: Against Black-Box Musical AI

Contemporary AI music systems are often presented as instruments of expansion: they promise new sounds, new workflows, and new relations between human intention and machine generation. Yet much of this expansion remains caught between two limiting paradigms. On one side stand large generative models trained on vast corpora, capable of producing convincing musical surfaces while concealing the processes through which such surfaces are produced. On the other side stand interactive systems that privilege real-time response, immediacy, and performative control, often at the expense of structural depth. In both cases, the composer is placed in relation to a system whose internal musical reasoning is either inaccessible or subordinated to the speed of interaction.

The problem is not simply technical. It is aesthetic and epistemological. When AI functions as a black box, musical form appears as output rather than as inquiry. The composer may prompt, select, curate, or refine, but the generative logic that connects material to result remains largely hidden. Musical structure is displaced into the statistical interior of the model. Sound emerges, but the path by which it emerges is difficult to inspect, contest, or revise.

This article proposes *Latent Morphologies* as an alternative to black-box musical AI. The framework does not begin from a pretrained model, an external corpus, or an autonomous generative engine. Instead, it begins from the sound material itself. Audio events are segmented, analyzed, embedded, clustered, navigated, and recomposed

through a self-supervised process in which the morphology of the source material becomes the basis for musical form. Rather than imposing stylistic syntax from outside, the system discovers relations internal to the analyzed sound: patterns of similarity, resonance, instability, density, contour, and transformation. These relations are not treated as hidden computational states but as compositional resources.

The system is implemented within Praat AudioTools, combining two complementary environments. Praat provides an analysis-resynthesis layer in which sound can be represented through structured objects such as segments, pitch trajectories, intensity curves, formant structures, and annotations. Python extends this layer through feature extraction, clustering, latent representation learning, path generation, and adaptive planning. The purpose of this combination is not to replace compositional judgment with machine autonomy, but to construct a transparent environment in which computational processes remain available for inspection and intervention. The machine does not produce form alone; it participates in a compositional ecology in which analysis, listening, decision-making, and recomposition remain entangled.

In this sense, *Latent Morphologies* belongs less to the paradigm of generative replacement than to a renewed practice of digital lutherie. The "instrument" here is not a stable object with fixed affordances, nor a neutral interface awaiting human expression. It is a distributed analytical-computational environment that reshapes the conditions under which sound becomes musical material. Its strings are not stretched across wood, but across feature spaces, latent embeddings, similarity matrices, formant trajectories, and recomposition paths.

This shift has consequences for musical authorship. In conventional human-computer music discourse, agency is often divided between two poles: the human composer as intentional subject and the machine as tool or generator. *Latent Morphologies* complicates this division. The system is neither passive tool nor independent author. It functions as an actant: a participant that proposes, constrains, redirects, and reveals

possibilities within the sound material. Its agency is real but bounded; it acts, but its actions remain traceable.

The framework therefore operates in what may be called compositional deep time: unlike real-time systems, which value immediacy, this workflow unfolds through delayed, iterative cycles of analysis, navigation, rendering, and revision, treating latent space not as a hidden abstraction but as a terrain for musical decision.

The argument advanced here is that such a practice constitutes an infraction of several inherited musical paradigms: the idea of the instrument as a fixed interface, the idea of AI as autonomous generator, the idea of form as externally imposed syntax, and the idea of composition as either real-time control or post-hoc editing. In their place, *Latent Morphologies* proposes a model of self-supervised audio recomposition in which form emerges through the encounter between sonic morphology, computational analysis, and human listening. Its contribution is not a new universal model for AI music, but a situated framework for making musical structure auditable, mutable, and materially grounded in sound itself.

1. From Instrument to Meta-Instrument

To understand the compositional significance of *Latent Morphologies*, it is necessary to reconsider what is meant by an instrument. In many musical contexts, an instrument is understood as a relatively stable interface between bodily action and sound: a violin, a keyboard, a controller, a microphone, or even a software environment whose affordances can be learned, repeated, and mastered. The instrument mediates expression by translating gesture into sonic consequence. Digital music technologies have complicated this relation, but they have not always displaced it. Digital Audio Workstations, live-coding environments, modular synthesis systems, and performance

controllers all extend the field of instrumental action, yet they frequently preserve the assumption that musical agency is exercised through control over a known interface.

Latent Morphologies operates differently. It does not present itself as a single playable interface, nor as a fixed tool for applying predetermined transformations. Instead, it functions as a meta-instrument: a temporary, distributed, and material-specific environment in which the conditions of musical action are generated from the analyzed sound corpus itself. The composer does not simply play the system; the composer constructs, tunes, and navigates a field of relations extracted from the sound material. Instrumentality is therefore relocated from interface to process.

This relocation is central. In a conventional instrument, the material is shaped through an external mechanism: string, key, bow, breath, or controller mediates sound production. In *Latent Morphologies*, by contrast, the sound material contributes to the construction of the mechanism through which it will later be transformed. Audio events are segmented and analyzed; their internal features are mapped into spaces of similarity, distance, density, resonance, and motion. These spaces then become the terrain through which recomposition occurs. The instrument is not imposed upon the material from outside; it is partially derived from the material's own morphological structure.

This is why the framework belongs to digital lutherie, but not in the narrow sense of designing a new interface or controller. The lutherie at stake here is not the fabrication of a stable object but the fabrication of a compositional relation. The builder does not only design how a performer will act; they design how sound will become available to thought, decision, and transformation. The meta-instrument consists of analysis objects, feature vectors, latent embeddings, similarity matrices, navigation paths, and resynthesis procedures. These components are not merely technical infrastructure. They define what can be perceived, compared, altered, repeated, and recomposed.

Praat plays a crucial role in this shift because it treats sound not simply as a waveform but as a set of structured and editable representations. Pitch, intensity, formant

behavior, segmentation, and annotation can all be externalized as objects, available at a level between raw audio and symbolic notation. Python extends this layer by introducing clustering, embedding, self-supervised learning, and path generation, yet it does not erase the object-based structure: it acts upon it and returns new plans, trajectories, or recomposition instructions that can be rendered, inspected, and revised. The meta-instrument therefore has no single body; it is distributed across environments and representations. Its tuning is a matter of segmentation thresholds, feature selection, distance metrics, clustering parameters, latent dimensionality, navigation rules, and resynthesis choices, decisions that shape the system's musical behavior, so that to alter them is to rebuild part of the instrument.

This has consequences for performance and authorship. A meta-instrument does not demand virtuosity in the traditional sense of rapid embodied control, but a different musicianship: the capacity to listen structurally, understand how representations mediate perception, and evaluate how computational relations become audible form. Such an instrument also resists universality. Because its behavior depends on the analyzed corpus, each project generates a different instrumental condition: vocal fragments, instrumental attacks, or field recordings each produce different feature spaces, clusters, paths, and recompositional potentials. The system does not claim to be a general-purpose generator of style. It produces situated instruments from situated materials.

In this sense, *Latent Morphologies* proposes a post-digital model of instrumentality. The instrument is no longer a stable mediator between human intention and sound, but an evolving ecology of analysis, computation, listening, and reconstruction, built from the very material it transforms.

2. Latent Morphology as Musical Material

The central proposition of *Latent Morphologies* is that musical material does not begin only where sound is arranged, performed, or transformed. It also exists within the internal relations of the sound itself. A recorded event is not merely an audio object awaiting processing, nor a sample to be triggered, layered, or sequenced. It contains patterns of energy, resonance, articulation, stability, noise, brightness, contour, and temporal behavior. These features are not secondary technical descriptors. They are the conditions through which the sound becomes perceptually and compositionally available.

To treat latent morphology as musical material is therefore to shift attention from surface manipulation to relational structure. A sound is not approached primarily as an isolated object with fixed identity, but as a point within a field of possible relations. It may resemble another sound in spectral centroid, diverge from it in temporal envelope, share its resonance profile while differing in attack, or occupy an unstable position between voiced and noisy states. Such relations do not yet constitute musical form, but they create the terrain from which form can emerge.

This approach differs from both notation-based and timeline-based paradigms. In notation, musical material is typically organized through symbolic units: pitches, durations, dynamics, articulations, and formal divisions. In timeline-based digital production, material is often organized through placement, layering, looping, and automation. Both paradigms tend to position sound as something arranged within an external framework. By contrast, latent morphology begins from the proposition that the framework may be discovered within the sound material itself. The corpus does not simply fill a form; it generates the conditions of form.

Self-supervised learning plays an important role in this shift. Rather than applying categories supplied in advance, the system derives relations from the analyzed corpus, learning in a limited and situated way how events differ from and resemble one another. These learned relations are local, contingent, and material-specific: a corpus

of vocal fragments produces one morphology, a corpus of instrumental attacks another, a collection of environmental recordings yet another. Each corpus constructs a different latent terrain, and therefore a different compositional problem.

This is why the term “latent” should not be understood as a withdrawal from sound into abstraction. Latency here names a condition of potential audibility. The relations discovered by the system are not directly heard as such, but they can become audible through recomposition. A cluster, embedding, path, or similarity matrix is not itself music; it is a way of organizing possible musical movement. When the system navigates from one event to another according to proximity, contrast, attraction, or instability, latent relations are rendered as temporal experience. The hidden geometry of the corpus becomes perceptible as form.

Morphology also resists the reduction of sound to discrete musical categories. Many of the most compositionally significant features of sound are continuous rather than categorical: the gradual brightening of a spectrum, the slight destabilization of periodicity, the shift from pitched tone into noise, the thickening of resonance, the compression of gesture, the blurring of attack into texture. Such features are difficult to capture through traditional symbolic description, yet they are central to electroacoustic listening and sound-based composition. *Latent Morphologies* treats these continua not as ornaments or effects, but as structural materials.

The framework is therefore especially concerned with thresholds: where one sound becomes another, where resemblance begins to fail, where continuity becomes rupture, and where transformation remains perceptible as transformation rather than substitution. A path through latent space may produce smooth morphing, abrupt contrast, recursive return, or unstable oscillation between families of sound. Each strategy produces a different mode of musical time. Form arises from the way the system moves across relations of similarity and difference.

The three operations that organize the framework, discovery, navigation, and mutation, can be understood as three ways of activating latent morphology. Discovery identifies families, patterns, densities, and structural tendencies within the material.

Navigation turns these relations into paths, sequences, and trajectories. Mutation transforms the material by recombining or displacing features within the learned field. Together, these operations allow morphology to function not merely as analysis but as compositional matter.

This also changes the status of analysis. In many musical contexts, analysis follows the work: it explains, interprets, or validates structures that have already been composed. Here, analysis precedes and participates in composition without being a rigid preparatory stage; it remains active throughout the process. The analytical representation is not a report about the sound; it is a medium through which the sound can be re-entered and transformed. Analysis becomes generative because it creates new possibilities for listening and action.

At stake is a different understanding of musical form. Form is not imposed from above as a pre-existing schema, nor delegated entirely to an autonomous model. It emerges through the interaction between sonic morphology, computational relation, and compositional judgment. The composer listens not only to sounds, but to the relations between sounds: to distances, clusters, attractors, anomalies, and trajectories. In this sense, latent morphology is not an invisible technical layer beneath the music. It is the musical material itself, temporarily hidden, then disclosed through process.

3. System Architecture: Discovery, Navigation, Mutation

The architecture of *Latent Morphologies* is organized around three compositional operations: discovery, navigation, and mutation, understood not as sequential stages in a fixed pipeline but as recurring modes of engagement with sonic material. At the center of the framework is the hybrid architecture of Praat AudioTools, linking Praat and Python. Praat functions as the analysis-resynthesis environment: it segments sound, extracts and edits structured objects, and renders the resulting transformations. Python functions as the computational extension layer: it performs feature extraction,

clustering, representation learning, similarity analysis, path generation, and adaptive planning. The system preserves intermediate states, segments, feature vectors, latent embeddings, similarity matrices, paths, and tier structures, as inspectable and adjustable materials rather than letting them disappear into a hidden generative mechanism.

The first pillar, discovery, concerns the identification of morphological families and structural tendencies within the source material. The CNN Event Recomposer begins by segmenting audio into discrete events and representing each event as a temporal trajectory of acoustic features, including pitch, intensity, harmonic-to-noise ratio, voicing, spectral centroid, spectral spread, and zero-crossing rate. Rather than treating these features as isolated measurements, the module learns relations between events through a lightweight self-supervised process: multi-scale convolutional filters examine different temporal spans, allowing the system to register both short articulatory details and broader gestural shapes. The purpose of this learning process is not to classify the material according to external labels, but to discover a latent morphology internal to the corpus. Events that share similar contours, energy profiles, spectral tendencies, or voicing behaviors are drawn into proximity, while morphologically distinct events are separated. Clustering then organizes the learned embeddings into families that can be ordered, braided, phased, or walked through according to different recompositional strategies.

A second discovery module, the SSM Composer, approaches the corpus through self-similarity. Each event is represented through a compact feature vector, and pairwise similarity between events is used to construct a self-similarity matrix, a map of repetition, contrast, and structural affinity. This matrix is not treated as a neutral analytical image; it can be transformed. Blurring dissolves boundaries between families; sharpening intensifies motif-like relations; diffusion spreads similarity across neighboring zones; motif amplification emphasizes recurring diagonal structures; structural warping folds or bends the relational space. A path walker then traverses this modified matrix, producing new event sequences. These procedures turn

analysis into formal intervention: a blurred matrix may yield continuity and drift, a sharpened matrix recurrence and insistence, a warped matrix folded or recursive formal behavior. Discovery is therefore not passive recognition. It is the compositional shaping of what will count as relation.

The second pillar, navigation, concerns movement through the latent field. If discovery constructs a terrain, navigation defines how the system moves across it. The Self-Attention Latent module encodes events into compact latent representations and computes relations of attentional affinity between them. Attention here is not used as a universal model of musical coherence, but as one navigational force among others: the system balances geometric proximity in latent space with attentional attraction, producing paths that can drift, mutate, return, and settle. This four-phase model is compositionally significant. Drift allows relatively coherent movement through related materials; mutation permits more exploratory jumps and recombinations; return introduces a sense of orientation by directing the path toward a target region; settle reduces instability and allows the form to arrive at a provisional state of rest. Musical form emerges through the changing behavior of navigation itself. The Phase Space Composer offers another model of navigation, drawing on deterministic dynamical systems rather than learned attention. Attractors such as Lorenz, Rössler, Hopf, or logistic systems are mapped onto feature space, creating trajectories that are neither random nor conventionally linear, producing turbulent, spiraling, oscillatory, or chaotic movement through the corpus. The composer does not arrange events one by one; the composer selects and tunes a navigational force. Together, the navigation modules challenge the idea that musical form must be built from external syntax. They treat form as movement through a field of sonic relations: a sequence is not merely a line of events, but the audible trace of a path.

The third pillar, mutation, concerns the transformation of sound material through recombination, displacement, and hybridization. The Latent Barycentric module generates new sonic forms by interpolating among neighboring events in latent space, drawing on a local constellation of related events and combining them according to

distance-based weighting. The resulting sound is neither simple collage nor arbitrary synthesis, but a morphologically situated hybrid produced from the relational neighborhood of the corpus, occupying an intermediate zone between memory and invention, recognition and transformation.

The Formant Swarm Granulator approaches mutation through resonance. Grains are analyzed according to formant behavior, clustered into resonance families, and mapped into a projected feature space. A swarm-like process then organizes their movement according to attraction, repulsion, density, and spatial distribution. The resulting texture is a resonance field whose internal motion is shaped by the morphological relations of the grains themselves; formant behavior becomes a compositional force.

Finally, within Praat AudioTools, Praat Pbind functions as a declarative interface layer linking analysis, computation, and resynthesis. It allows patterns and schedules to be specified in a form that can be compiled into structured Praat objects such as PitchTiers and IntensityTiers. This is crucial because it keeps the system auditable. The result of computation is not only sound, but also an editable trace of the decisions that produced the sound, which can be inspected, modified, repeated, or reinterpreted. The architecture therefore does not aim to produce an autonomous AI composer. Its purpose is to construct an environment in which computational processes expose latent musical relations and make them available for compositional action. Across all three operations, the system maintains a commitment to transparency: the machine acts, but its actions remain situated, traceable, and open to intervention.

4. Compositional Deep Time

A central feature of *Latent Morphologies* is that it does not operate primarily in real time. This is not a technical deficiency awaiting optimization. It is a compositional

position, an expanded temporal relation to sound that I describe as compositional deep time.

Much contemporary music technology privileges immediacy. Real-time systems, live coding, interactive controllers, and responsive machine-learning instruments often value the speed with which an action produces audible consequence. This immediacy has considerable musical power: it supports performance, embodied control, risk, and responsiveness. Yet it also encourages a particular image of musical agency in which the value of the system is measured by its responsiveness to gesture.

Latent Morphologies proposes a different relation. Here, the composer does not primarily act through instantaneous control, but through delayed inquiry: a corpus is analyzed, events are segmented, features are extracted, latent relations are constructed, paths are generated, recompositions are rendered, the result is listened to, parameters are revised, the process is repeated. Musical time is distributed across multiple temporal layers: the duration of the source sounds, the computational time of analysis, the reflective time of listening, and the iterative time of compositional revision.

This slowness shapes what can be heard. When a system responds immediately, attention often moves toward gesture and effect. In compositional deep time, attention shifts toward relation and consequence: the composer asks not only what sound has been produced, but how this sound emerged from the corpus, which latent relations guided it, and what other trajectories might have been possible. Listening becomes investigative rather than reactive.

The offline workflow also makes possible a form of musical scrutiny that real-time systems often conceal. Because intermediate representations can be inspected before or after rendering, the composer may examine the structure of the process itself: clusters, similarity matrices, embeddings, path sequences, tier objects, barycentric weights, and swarm distributions. These are not merely diagnostic tools; they are part of the compositional material. To inspect them is to listen through representation toward possible sound.

This delay also changes the status of error. In real-time performance, an unexpected output may appear as interruption or failure. Within deep time, unexpected results can be inhabited: a strange transition, an unstable hybrid, an overly dense swarm, or an uncanny barycentric blend can be rendered, heard repeatedly, compared with its conditions of emergence, and either rejected or developed. The system does not demand immediate judgment; it allows the composer to remain with the sound long enough for its implications to become audible.

The concept is especially important for AI-assisted composition. Black-box generation often compresses the temporal relation between instruction and result: a prompt is given, an output appears, and the user evaluates the surface. Compositional deep time stretches this relation and reopens the middle of the process. The composer is not limited to accepting or rejecting an output, because the path from material to result remains available for intervention. In this sense, deep time is also an ethical relation to computation: it resists the acceleration of musical decision-making into automated production and insists that AI-assisted composition can remain slow, revisable, and accountable.

The musical consequence of this temporal model is a different kind of form, not only the arrangement of events in time, but the accumulated result of returning to sound through multiple layers of representation. Each iteration leaves a trace: in the rendered audio, in the adjusted parameters, in the altered path, and in the composer's changed listening. Compositional deep time names this expanded field of temporal engagement. It is the time required for latent morphology to become audible as musical thought.

5. AI as Actant, Not Author

The question of authorship is unavoidable in AI-assisted music. When a system segments a corpus, discovers relations, proposes paths, recombines events, and generates new sonic forms, it is not merely executing a neutral instruction; it participates in the formation of the work. At the same time, to describe such a system as an autonomous author risks overstating its independence. *Latent Morphologies* proposes a different position: AI functions as an actant, not as author.

An actant is not a sovereign subject. It does not possess intention in the human sense, nor does it originate musical meaning alone. Rather, it participates in a network of relations that shapes what can happen. In *Latent Morphologies*, the AI actant contributes by identifying proximities, generating paths, proposing recombinations, and revealing latent possibilities within the sound material. It does not decide what the music means, but it alters the field in which musical decisions become possible.

This distinction matters because much public discourse around AI music is organized around a false opposition between tool and creator. If AI is understood only as a tool, its agency disappears, even when its operations produce results the composer could not have predicted. If AI is understood as creator, human compositional responsibility is displaced into the machine, as though form had emerged from an independent musical subject. Both positions are inadequate.

The agency of *Latent Morphologies* is bounded by the corpus it can discover relations within, by the features selected for analysis, by the algorithms used to cluster, embed, navigate, and mutate the material, by parameters chosen or revised by the composer, and finally by listening, since no computational proposal becomes musically meaningful until it is heard and situated within compositional intention. This boundedness is precisely what makes the system musically useful. A fully opaque generative model may produce impressive results, but its agency is difficult to locate; *Latent Morphologies* instead makes the conditions of agency explicit: a cluster can be

inspected, a path retraced, a similarity matrix transformed, a recomposition rerun with altered parameters.

This traceability changes the composer's role. The composer is not reduced to a prompt-giver or curator of outputs, nor does the composer exercise total control in the traditional sense. Instead, the composer works by setting conditions, interpreting proposals, redirecting processes, and listening for emergent consequences. Compositional agency becomes distributed without becoming anonymous.

This is especially clear across the three operations. Discovery can produce clusters that surprise the composer, gathering sounds that share a contour but differ in timbre, revealing continuities within apparently heterogeneous material. Navigation generates paths through a relational field, and each movement is a proposition about how one sound might lead to another. Mutation makes authorship still more complex: a barycentric hybrid or formant-based swarm is generated from relations among source events rather than selected from the corpus, recognizable as belonging to the material yet identical to no original event, produced through a distributed agency in which the corpus provides morphology, the algorithm defines a recombinatory procedure, and the composer determines whether the result enters the work.

To call the AI an actant is therefore to acknowledge its participation without mythologizing its autonomy: it changes the compositional situation, introducing proposals the composer did not manually design, yet it possesses no taste, memory, or intention apart from the procedures and materials through which it operates. Treating AI as actant, rather than author or mere tool, allows for a more accountable position, in which responsibility includes understanding and shaping the agencies through which the work is produced. The question becomes not "Did the machine compose this?" but "What kinds of agency were allowed to participate, under what conditions, and with what audible consequences?"

6. Infractions of the Musical Paradigm

The significance of *Latent Morphologies* lies not only in its technical design, but in the way it unsettles several inherited assumptions about musical practice, taken together as a deliberate disturbance of the categories through which that practice is commonly organized. The instrument is no longer prior to the sound but partially derived from it, rebuilt each time from the analyzed corpus, the chosen features, and the recomposition paths. Analysis is no longer confined to a position before or after the work: segmentations, feature spaces, clusters, and trajectories are sites of intervention rather than mere description, so that to analyze is already to begin composing. Musical form is relocated from externally imposed syntax, thematic development, harmonic function, and metric structure to the internal morphology of the corpus itself, displacing formal decision into the negotiation between computational relation and listening judgment. Time is relocated from real-time responsiveness to a slow, iterative, recursive temporality in which the delay between action and result becomes productive rather than merely tolerated. Authorship is relocated from the opposition between human creator and machine generator toward the actant: a participant that proposes paths and hybrids the composer did not fully anticipate, while remaining auditable at every stage. And sound itself is relocated from file, sample, or signal toward morphology, its internal behavior becoming the ground of compositional action rather than raw matter awaiting an externally imposed structure.

None of this amounts to a rejection of instruments, analysis, form, time, or authorship, but a reconfiguration of each. In this sense, *Latent Morphologies* proposes a practice of sonic infraction: a way of making music by disturbing the inherited boundaries between tool and material, composer and system, analysis and composition, structure and sound.

Conclusion

Latent Morphologies proposes a model of AI-assisted composition in which computation does not replace musical thought, but reorganizes the conditions through which sound can be heard, analyzed, transformed, and recomposed. Against the opacity of black-box generation, the framework insists on transparency: events are segmented, features extracted, relations learned, paths generated, and transformations rendered through processes available for inspection and revision. Its musical value lies in exposing latent relations within sound and making them compositionally actionable, and the composer remains responsible throughout, not as absolute controller, but as listener, designer, evaluator, and respondent within a distributed compositional ecology.

Latent Morphologies is not, then, a universal solution for AI music, but a situated proposal for transparent, self-supervised audio recomposition, a way of composing with machine learning while preserving editability, traceability, and responsibility. Its aim is to show how AI can participate in musical practice without erasing the composer's listening, and how sound itself can become the source of the instrument that transforms it.

Bibliography

Paul BOERSMA and David WEENINK, *Praat: Doing Phonetics by Computer*, computer software, 2025.

Oliver BOWN, Alice ELDRIDGE and Jon MCCORMACK, *Understanding Interaction in Contemporary Digital Music: From Instruments to Behavioural Objects*, in «Organised Sound», vol. 14, n. 2, 2009, pp. 188–196.

Jean-Pierre BRIOT, Gaëtan HADJERES and François-David PACHET, *Deep Learning Techniques for Music Generation*, Springer, Cham, 2020.

Kim CASCONE, *The Aesthetics of Failure: 'Post-Digital' Tendencies in Contemporary Computer Music*, in «Computer Music Journal», vol. 24, n. 4, 2000, pp. 12–18.

Ting CHEN, Simon KORNBLITH, Mohammad NOROUZI and Geoffrey HINTON, *A Simple Framework for Contrastive Learning of Visual Representations*, in Proceedings of the 37th International Conference on Machine Learning, 2020, pp. 1597–1607.

Shai COHEN, *Praat AudioTools: An Offline Analysis–Resynthesis Toolkit for Experimental Composition*, software library and online documentation, 2026, https://mashav.com/sha/Praat%20AudioTools/; see also Shai COHEN, *Praat Audiotools: An offline analysis-resynthesis toolkit for experimental composition* in «Organised Sound», First View, 2026, pp. 1-9, https://doi.org/10.1017/S1355771826101204.

David COPE, *Computer Models of Musical Creativity*, MIT Press, Cambridge, MA, 2005.

Jesse ENGEL, Lamtharn HANTRAKUL, Chenjie GU and Adam ROBERTS, *DDSP: Differentiable Digital Signal Processing*, in Proceedings of the International Conference on Learning Representations, 2020.

Dorien HERREMANS, Ching-Hua CHUAN and Elaine CHEW, *A Functional Taxonomy of Music Generation Systems*, in «ACM Computing Surveys», vol. 50, n. 5, 2017, pp. 1–30.

Sergi JORDÀ, *Digital Lutherie: Crafting Musical Computers for New Musics' Performance and Improvisation*, PhD diss., Universitat Pompeu Fabra, 2005.

Diederik P. KINGMA and Max WELLING, *Auto-Encoding Variational Bayes*, in Proceedings of the International Conference on Learning Representations, 2014.

Bruno LATOUR, *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford University Press, Oxford, 2005.

Thor MAGNUSSON, *Sonic Writing: Technologies of Material, Symbolic, and Signal Inscriptions*, Bloomsbury Academic, New York, 2019.

Brian MCFEE, Colin RAFFEL, Dawen LIANG, Daniel P. W. ELLIS, Matt MCVICAR, Eric BATTENBERG and Oriol NIETO, *librosa: Audio and Music Signal Analysis in Python*, in Proceedings of the 14th Python in Science Conference, 2015, pp. 18–25.

Eduardo Reck MIRANDA (ed.), *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*, Springer, Cham, 2021.

Gerhard NIERHAUS, *Algorithmic Composition: Paradigms of Automated Music Generation*, Springer, Vienna, 2009.

Curtis ROADS, *Microsound*, MIT Press, Cambridge, MA, 2001.

Robert ROWE, *Interactive Music Systems: Machine Listening and Composing*, MIT Press, Cambridge, MA, 1993.

Pierre SCHAEFFER, *Treatise on Musical Objects: An Essay across Disciplines*, trans. Christine North and John Dack, University of California Press, Oakland, CA, 2017; original work published 1966.

Xavier SERRA and Julius O. SMITH, *Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition*, in «Computer Music Journal», vol. 14, n. 4, 1990, pp. 12–24.

Denis SMALLEY, *Spectromorphology: Explaining Sound-Shapes*, in «Organised Sound», vol. 2, n. 2, 1997, pp. 107–126.

Aaron VAN DEN OORD, Yazhe LI and Oriol VINYALS, *Representation Learning with Contrastive Predictive Coding*, arXiv preprint, 2018.

Ashish VASWANI, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N. GOMEZ, Łukasz KAISER and Illia POLOSUKHIN, *Attention Is All You Need*, in Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.

Trevor WISHART, *On Sonic Art*, Harwood Academic Publishers, Amsterdam, 1996.