Organised Sound (2026), **31**, 263–273
doi:10.1017/S1355771826101563

CAMBRIDGE UNIVERSITY PRESS

Check for updates

Article

Emotional inference as embedded algorithm in live performance: From recognition to dramaturgy

Gadi Sassoon
Independent Scholar, Italy

Abstract

This article examines what happens when emotional-inference algorithms are embedded as compositional tools within live performance systems rather than deployed as external classifiers. Through analysis of the Transmodal Emotional Amplification Listener (TEAL), a real-time multimodal system developed by the author for two performance works, it demonstrates how embedding transforms algorithmic function from recognition to composition. TEAL employs nested analysis windows that track interval statistics across multiple time scales, assigns composer-defined emotional parameters to these patterns and fuses them with gestural energy data extracted via MediaPipe. The composite signal passes through a locally hosted language model that generates narrative text, the Synthetic Dream. A secondary interpreter reverses the emotional parameters into MIDI and the Synthetic Dream into visual prompts, closing a bidirectional compositional loop in which the system's generated notes are played back acoustically on a Yamaha Disklavier grand piano alongside the performer. The article introduces emotional sovereignty, defined as the performer's right to define the affective ground on which algorithms operate, as a design principle for embedded affective systems. Drawing on music emotion recognition, interactive systems research and cybernetic art, it argues that bespoke compositional embedding offers a viable alternative to universal emotion recognition.

Keywords: affective computing; embedded algorithms; emotional sovereignty; live performance; music emotion recognition

(Received 30 June 2026 UTC; revised 30 June 2026 UTC; accepted 2 August 2026 UTC; First Published online 24 September 2026)

1. Introduction

The proliferation of emotion recognition systems across creative, commercial and governmental domains raises questions about agency, authenticity and artistic control in human–computer interaction (HCI). Since Picard's foundational argument that affect is integral to intelligence and to the design of responsive systems (Picard 1997), the development of systems that recognise, interpret and respond to human emotions has accelerated. Much of this research focuses on improving recognition accuracy. This article advances a different question: what happens when emotional-inference algorithms are embedded as compositional agents within live performance, and how does that embedding transform both their function and their aesthetic status? The title's use of 'dramaturgy' refers to the temporal structuring of relationships between performer, algorithm and audience across a shared performance, encompassing not only sound but text, image and physical actuation.

The author is simultaneously composer, performer and system designer. This convergence of roles is relevant because the system described here, the Transmodal Emotional Amplification Listener (TEAL), does not position emotional inference as a measurement of the performer's affective state. The limits of treating emotion as an externally measurable quantity are by now widely argued (Barrett et al. 2019; Crawford 2021); TEAL takes this critique as its premise rather than its target, treating emotional inference as an extended compositional technique in which the composer assigns arbitrary affective values to musical patterns as part of the compositional process itself. The system premiered at the UN AI For Good summit in Geneva in 2025, within the work *Modes of Vibration AV*, and was subsequently used to record a series of performer-system duets on a Yamaha Disklavier grand piano at the Université de Montréal, during a residency with the Wilding AI research-creation lab.

The work draws on three concurrent developments. First, real-time affective computing has matured sufficiently to support live performance applications, building on decades of music emotion recognition (MER) research that has demonstrated both the potential and the limitations of computational emotion detection (Kim et al. 2010; Yang and Chen 2012; Han et al. 2022). Second, large language models (LLMs) have introduced possibilities for contextual, narrative responses to musical events that exceed simple classification. Third, critiques of emotion recognition technologies have exposed the ethical and aesthetic problems of treating affect as measurable, universal psychological states (McStay 2018; Barrett et al. 2019; Crawford 2021; Stark and Hutson 2022).

The concept of emotional sovereignty proposed here, the right to govern one's own emotional trajectory in the presence of AI systems, draws an analogy with behavioural economics: prospect theory's account of asymmetric sensitivity to losses versus gains (Kahneman and Tversky 1979) suggests why preserving emotional autonomy may matter when systems can influence affective states. This article argues for personalised amplification: algorithms that

Email: g@gadisassoon.com

Cite this article: Sassoon G. (2026). Emotional inference as embedded algorithm in live performance: From recognition to dramaturgy. *Organised Sound* **31**: 263–273, https://doi.org/10.1017/S1355771826101563

© The Author(s), 2026. Published by Cambridge University Press.

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press

magnify intentional gestures without claiming universal emotional truth.

The article is organised as follows. Section 2 surveys the theoretical background: MER, affective computing, interactive music systems, expressive performance research and cybernetic art. Section 3 describes the TEAL system architecture in full technical detail. Section 4 presents two performance case studies. Section 5 discusses the implications for embedded algorithmic practice, limitations and future directions. Section 6 concludes.

2. Theoretical framework: From universal recognition to situated composition

2.1 The evolution and limitations of music emotion recognition

Research into MER has grown steadily since the late 1990s, largely within the music information retrieval community. The premise that music functions as a language of emotions is supported by psychological studies showing that listeners often engage with music to regulate affective states (Juslin and Sloboda 2001, 2010). Researchers sought to map acoustic features (timbre, rhythm, tempo, harmony) onto discrete emotional categories or onto dimensional models of valence and arousal.

Several fundamental problems emerged. The categorical approach reduces musical experience to a handful of classes, ignoring subtle and culturally embedded distinctions (Yang and Chen 2012). Emotion descriptors are ambiguous: different listeners describe the same passage differently, and the same listener may be inconsistent over time (Juslin and Laukka 2004). Despite advances in deep learning, these limitations persist. Surveys by Han et al. (2022) and Schedl et al. (2014) confirm that state-of-the-art models struggle with generalisability beyond controlled datasets.

The technical trajectory reflects these tensions. Early work used handcrafted features (spectral centroid, mel-frequency cepstral coefficients, rhythm periodicities) fed into support vector machines and Gaussian mixture models (Kim et al. 2010). More recent approaches employ convolutional neural networks (Choi et al. 2017), achieving accuracy gains within benchmarks but remaining vulnerable to dataset bias and cultural specificity. What registers as sad in a Western tonal context may be experienced as triumphant in another tradition (Eerola and Vuoskoski 2013; Susino and Schubert 2019). MER has produced technical advances but faces a conceptual impasse: emotions in music are not universal categories waiting to be detected but situated, relational and context-dependent phenomena.

2.2 Affective computing

Picard (1997) argued that if computers are to interact naturally with humans, they must recognise, model and respond to emotional states. The promise lies in systems that are responsive, adaptive and emotionally informed. Applications span education (D'Mello and Graesser 2012), healthcare (Calvo et al. 2016) and creative practice.

The perils are increasingly recognised. Critics argue that affective computing rests on dubious assumptions: that emotions are discrete, measurable and universally legible through proxies such as facial expression, voice or physiology (Barrett et al. 2019). In practice, these assumptions have led to deployment in surveillance, law enforcement and workplace monitoring, where algorithmic claims to emotional knowledge carry significant risks (McStay 2018; Crawford 2021). Birhane and Guest (2021) highlight how such systems encode cultural bias and perpetuate algorithmic injustice. Affective computing holds creative potential, but its dominant implementations treat emotions as extractable data points, obscuring their situated and relational character.

2.3 Interactive performance systems and embedded agency

Within music and sound art, a counter-tradition embeds algorithms not as detectors but as agents within performance ecologies. Rowe's (1993) taxonomy distinguished between the instrument model, where the computer extends the performer's control, and the player model, where the computer behaves as an autonomous partner. George Lewis's *Voyager* (2000) epitomises the latter: an improvisational collaborator whose embedded algorithms generate responsive, unpredictable contributions.

Research within the New Interfaces for Musical Expression community has expanded this approach. Fiebrink and Caramiaux (2018) survey how machine learning can serve as a creative tool, enabling performers to shape models around personal vocabularies rather than universal categories. Eigenfeldt and Pasquier (2013) developed autonomous musical agents that interact in ensembles. Tatar and Pasquier (2019) provide a comprehensive typology of musical agents and their roles in metacreation, distinguishing between systems that extend, collaborate with and operate independently of human performers.

Live coding offers another precedent. Collins et al. (2003) showed how exposing and manipulating code in real time creates transparency and contestability. Magnusson (2014) theorised live-coded algorithms as a form of musical notation, a code score in which the same text functions simultaneously as instrument, score and performed gesture.

The past decade's advances in machine learning have produced a wave of co-creative music systems that extend this agent-based tradition (Briot et al. 2020). Rather than classifying or recognising, these systems generate or shape musical material in dialogue with a human partner. Louie et al. (2020) show that novices retain a stronger sense of ownership over a composition when given tools to steer a deep generative model rather than passively receiving its output. McCormack et al. (2019) equip an improvising AI with an extra-musical channel that communicates its internal confidence to its human collaborators, and Thelle and Pasquier (2021) frame their agent explicitly as a partner for creative brainstorming. These systems share a concern with agency and control that TEAL also pursues, but they operate on musical content directly. TEAL differs in making affective inference, rather than direct note or audio generation, the locus of co-creation; musical material is generated downstream from that affective layer.

These practices establish algorithms embedded in performance as agents of composition rather than instruments of recognition. TEAL draws on this tradition.

2.4 Expressive computer music performance and affective algorithmic composition

A body of work directly addresses computational systems for expressive and emotionally informed music generation. Livingstone et al. (2010) developed a computational rule system for modifying musical scores to express target emotions, demonstrating that systematic manipulations of tempo, articulation and dynamics could shift perceived emotional content. Kirke and Miranda (2009) surveyed the broader landscape of computer systems for expressive music performance, including affective

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press

algorithmic composition, in which emotional parameters drive generative musical processes.

These systems share a common architecture: they take emotional descriptors as input and produce musical output designed to express those emotions. Livingstone et al.'s rule system modifies tempo, dynamics and articulation to shift a score from 'happy' to 'sad'. In these systems, the target emotion is specified in advance, and the music is generated to express it. Kirke and Miranda's survey documents dozens of systems operating on this same causal axis: emotion specification drives the generative process.

TEAL inverts this direction. Musical analysis produces emotional parameters; those parameters generate text; the text generates new musical output. The emotional layer is not a target to be reached but a compositional interface between analysis and generation. Where Livingstone et al.'s system asks 'how should music sound to express sadness?', TEAL asks 'what emotional profile does the current music produce, and what text and sound should that profile generate next?' The distinction is not merely technical. It determines whether the system serves recognition or composition.

2.5 Feedback loops and cybernetic traditions

The recursive feedback loop at the heart of TEAL situates it within a lineage of cybernetic and feedback-based art practices. Wiener (1948) established feedback as a principle for regulating complex systems. Pask extended these concepts into interactive art, including his *Colloquy of Mobiles* (Pask 1971), where machines and participants engaged in adaptive dialogues. Reichardt's *Cybernetic Serendipity* exhibition (Reichardt 1968) demonstrated the aesthetic potential of feedback-driven systems across media.

In music, feedback principles informed electronic composition and performance art throughout the 1960s and 1970s. Collins (2004) documents how David Tudor's live electronics demonstrated self-regulating, environment-sensitive systems. Biofeedback experiments sought to make physiological states audible, establishing an aesthetic of recursive coupling between human and machine.

TEAL operates within this lineage. Unlike early cybernetic art, it makes its interpretive limitations explicit. The system does not present algorithmic inference as a universal regulator but as a situated, partial and compositionally determined reading.

3. System architecture: Emotional inference as extended composition

3.1 Design principles

TEAL was conceived not as a universal recognition system but as a piece-specific compositional agent, calibrated to the specific harmonic, gestural and timbral language of each work. Four principles guided the design:

- *Embedded operation*. Real-time integration within the performance signal chain, with analysis-to-actuation latency consistently under 50 ms; the language-model text layer runs outside this tight loop, at higher latency.
- *Transparency*. Selected metrics, text outputs and confidence values are made visible to performer and audience.
- *Personalisation*. Analysis windows and mappings are tuned to individual compositional idioms.
- *Bidirectional generation*. The system does not merely analyse music and produce outputs; it closes a loop in which generated text is reversed back into musical material.

The central design decision is that emotional inference functions as an extended compositional technique. The composer assigns arbitrary emotional statistics to musical patterns: in other words, it is the composer's subjective understanding of specific musical gestures and structures that informs the rule set of the emotional inference engine. Examples of typical choices include mapping lingering minor thirds to 'sad', recurrence of the Dorian major sixth to 'light' or phrases articulated via the Phrygian minor ninth to 'dark'. These assignments are compositional choices, not scientific claims, combining subjective taste and music theory, primarily based on the perception of harmonic and modal idiosyncrasies. The system tracks these patterns, fuses them with gestural data and routes the composite signal onward (Figure 1). In this sense, the emotion inference layer is part of the score.

3.2 Musical analysis: Nested temporal windows

TEAL's musical analysis is performed by a 3,462-line JavaScript module running inside Max/MSP (`harmony_analyzer.js, v2.5`). The module employs a nested temporal architecture (the Matryoshka approach), operating across the micro-to-meso time scales described by Roads (2001), which maintains four concurrent analysis windows, each feeding harmonic, melodic and rhythmic feature extraction simultaneously.

Temporal architecture. A central performance table maintains pitch, velocity and timestamp for each note event, with perceptual loudness weighted using Stevens' power law (exponent 0.6) and temporal decay functions applied over a 180-second rolling history. This table is the shared data structure from which all feature extraction proceeds.

Four nested windows read from this table in parallel, spanning timescales from onset detection to phrase-level progression. The window durations are controlled by a single timeSeed parameter that scales all four proportionally, allowing the performer to tune temporal resolution to the piece. At the finest granularity, additional micro-windows (50–200 ms) capture onset clustering and attack transients. At the coarsest, the extended window tracks harmonic progressions, modulation events and cadential functions across multiple phrases.

Harmonic analysis. Three key-detection algorithms run in parallel within the sustained window: the Krumhansl–Schmuckler probe-tone algorithm (Krumhansl 1990), Gómez's squared Euclidean distance method using harmonic pitch-class profiles (HPCP) (Gómez 2006) and Temperley's key profiles derived from the Kostka–Payne corpus (Temperley 1999). A weighted voting mechanism combines their outputs and reports confidence scores. The brief window performs chord recognition via template matching against 37 chord types, including extended jazz harmonies (major/minor 7ths, 9ths, 11ths, 13ths, suspended and altered chords). The extended window detects cadential functions (authentic, half, deceptive, plagal), modulation events and secondary dominant tonicisations.

Modal analysis draws on the chord-scale theory taught at Berklee College of Music (Nettles and Graf 1997), which identifies a characteristic tone for each mode. The system operationalises this with weighted character notes: for D Dorian, the major sixth (B natural), the tone that distinguishes Dorian from the natural minor, receives 1.5–3.0x emphasis, and chromatic notes are penalised 2.0x; these weightings are the system's own heuristic, not part of the source theory. The system calculates modal character strength and reports it alongside key detection confidence. Tonal tension is modelled using Lerdahl's pitch-space distance metric

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press

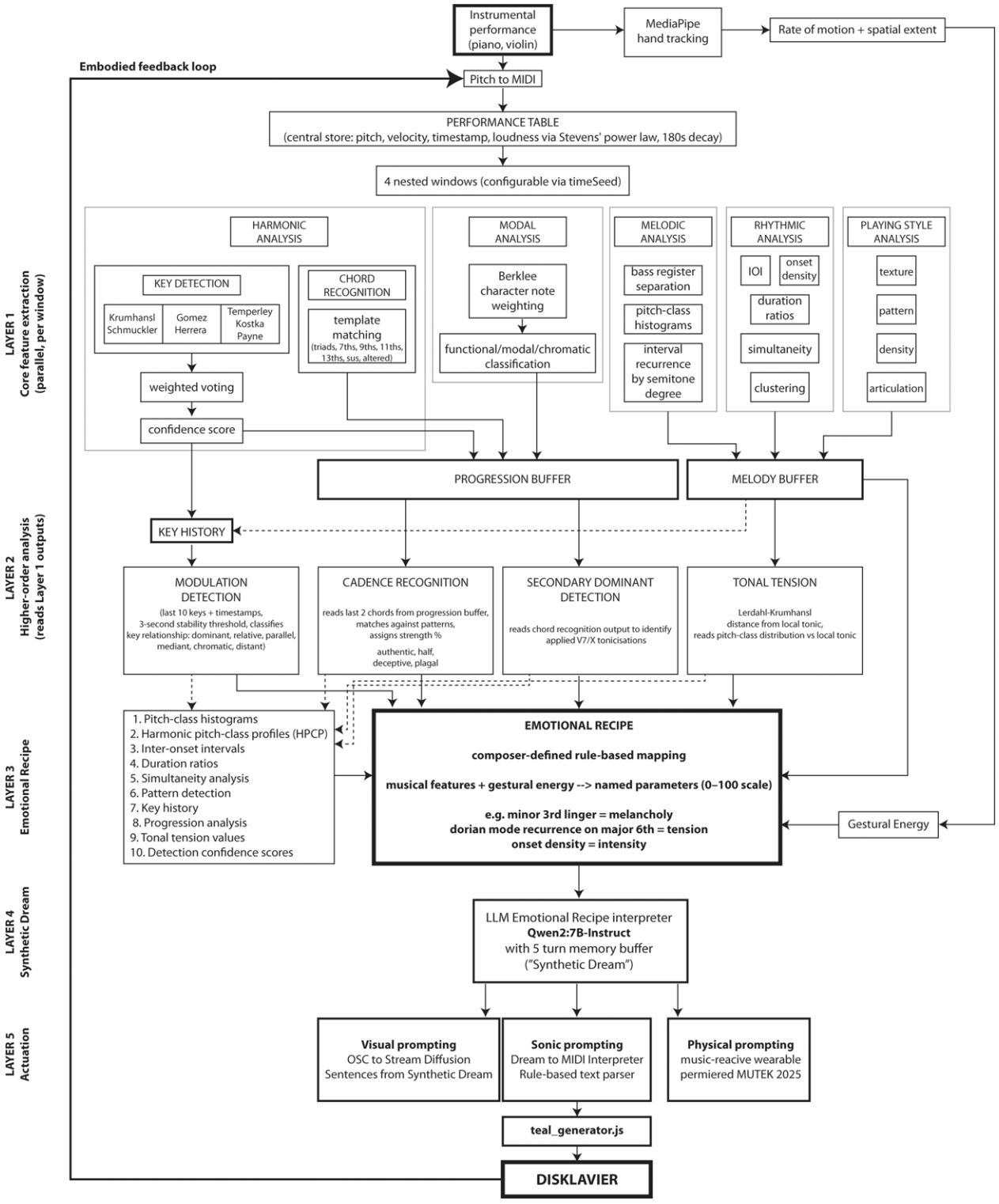


Figure 1. TEAL system architecture. Signal flow from instrumental performance through nested temporal analysis windows (Layer 1), higher-order harmonic analysis (Layer 2), composer-defined emotional recipe, language-model narrative generation (Synthetic Dream) and bidirectional output to Disklavier, visuals and physical actuation. The feedback loop from Disklavier output back to the performance table is shown at the top.

(Lerdahl 2001), producing a continuous tension curve that tracks the distance between the current pitch-class distribution and the local tonic.

Melodic and intervallic analysis. The module maintains rolling pitch-class histograms across each window and tracks interval recurrence by semitone degree: how often a minor third, a major sixth or a tritone appears within a given time frame. Bass register input is separated and weighted independently, enabling the system to distinguish low-register pedal tones and bass motion from upper-register melodic activity. Key history is tracked across

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press

windows, recording the sequence of detected keys and the duration spent in each.

Rhythmic analysis. The instant and brief windows extract inter-onset intervals (IOI), onset density (events per second), duration ratios between successive notes and simultaneity counts (how many notes sound at once). Onset clustering detects whether note events arrive in bursts or are evenly distributed. These rhythmic features contribute directly to the emotional recipe: high onset density with clustered attacks produces a different parameter profile than the same pitch content played sparsely.

The module outputs ten simultaneous data streams: pitch-class histograms, HPCP, IOI, duration ratios, simultaneity analysis, pattern detection, key history, progression analysis, tonal tension values and detection confidence scores. Processing latency for this analysis stage is approximately 20 ms.

A concrete example illustrates how these layers interact in composition. During a D Dorian piano improvisation, the instant window registers increasing onset density (3–11 events per second) with clustered attacks. The brief window identifies recurring minor-third intervals (D–F, A–C, E–G). The sustained window confirms D Dorian via Krumhansl–Schmuckler correlation at 78 per cent confidence. The extended window detects cadential motion as the flattened seventh (C natural) resolves to the modal centre. The composer has assigned arbitrary emotional statistics to these patterns: lingering minor-third intervals map to a 'sad' coefficient; recurrence of the major sixth in a Dorian context maps to 'light' (since B natural, the raised sixth, is the bright character note that distinguishes D Dorian from the natural minor, and its persistent presence signals modal commitment); rising onset density maps to 'intensity'. These named parameters are compositional assignments, not psychological measurements. A different composer could assign minor-third recurrence to 'exuberance'. The mapping is part of the piece, not part of the algorithm.

3.3 Gestural analysis and cross-modal fusion

Alongside the musical analysis, TEAL ingests body and hand tracking data via MediaPipe (Lugaresi et al. 2019), running within TouchDesigner. The MediaPipe skeleton provides continuous position data for hands, arms and upper body. Three gestural features are derived from this data: rate of motion (velocity of tracked points between frames), spatial extent (the bounding volume of the performer's gesture field, distinguishing expansive from contracted movement) and directional contour (whether motion trends upward, outward or inward). These features are transmitted to Max via Open Sound Control (OSC) at frame rate.

The primary derived metric is energy level, computed as a weighted combination of rate of motion and spatial extent. Rapid, expansive gestures produce high energy values; slow, contracted movements produce low values. This gestural energy is fused with the musical analysis trajectories in Max. The fusion is a weighted combination: the harmonic, melodic and rhythmic analysis provides affective content (which intervals, which mode, which rhythmic density), while gestural energy modulates intensity and urgency. The weights are part of the piece-specific calibration and are set during rehearsal.

The resulting composite, the emotional recipe, is a structured set of named parameters with numerical values, updated continuously. The inference method is entirely rule-based: correlation-weighted key detection outputs feed into composer-defined mapping rules that assign named emotional values to specific feature combinations. No regression model or labelled emotion dataset is involved. The recipe represents the system's compositional reading of the current musical and gestural material, according to rules defined by the composer, not a scientifically measured affect.

3.4 Language model: Emotional inference and Synthetic Dream

The emotional recipe is passed as structured input to a locally hosted Qwen2:7B-Instruct language model, running via Ollama for local inference with no network dependency. A Node.js proxy relays the recipe to the model and returns its output to Max.

The model receives a single structured prompt. The input is the emotional recipe rendered as a tagged list: eight named affect terms (happy, sad, delicate, powerful, soaring, light, dark and tense), three macro inferences (intensity, energy and harmonic tension), a certainty value and a dominant-emotion summary, each on a 0–100 scale, alongside the detected key, mode and dominant interval profile. The prompt instructs the model to return two linked texts. The first, the emotional inference, is terse and clinical: it opens by restating the certainty value, combines the affect terms into paired adjectives (with 'melancholy' reserved for passages where the sad and happy values converge) and reports intensity, energy and tension as macro descriptors of the passage's trajectory. The second, the Synthetic Dream, translates the same recipe into oneiric imagery, opening from one of five suggested incipits, which the model paraphrases freely, and drawing on a constrained vocabulary of cosmic and microscopic figures: constellations, microbial life, deep-ocean creatures, the resonant metal plates, the silence of mountains and the light of a star. The prompt is given below, with its two runtime insertions bracketed.

```
[prior emotional inferences, when present, are prepended here with an instruction to estimate certainty near their running average]
Here is a real-time emotional analysis of Gadi Sassoon's piano improvisation:
[emotional recipe]
Your task:

- Write a paragraph with 3 sentences, 5 words each, starting with: 'Gadi Sassoon emotional inference certainty X%', where X is the Certainty value from the input above.
- sentence 2, express the following in 10 words or less: Use all available data metrics to interpret the emotional character of the music. Use dry clinical terminology, machine-like. happy, sad, delicate, powerful, soaring, light, dark, tense must combine in two adjectives. intensity, energy and tension are macro inferences. For similar ranges of sad and happy use the word melancholy.
- sentence 3, in 5 words maximum Analyse development of the music from prior analysis block.
- Be extremely brief and data-like. The whole block is 1 continuous paragraph. Phrase your description as if you're listening to the music, not reading values.
```

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press

```
Then:

- Add a second short paragraph of 15 words about
  your synthetic dream based on my musical
  emotions.
- Begin with a choice of one of these, equal
  probability:
  - 'As I, a synthetic being, listen to the
    resonance of a human soul, I dream a syn-
    thetic dream of...'
  - 'Human emotion detected. Initiating syn-
    thetic dream sequence...'
  - 'Synthetic dreamscape:'
  - 'Emotional resonance engaged; machine
    hallucinating as follows:'
  - 'Emotional vibration detected – elaborat-
    ing oneiric signifier'
- Use the emotional input to imagine a synthetic
  being's dream of cosmic or microscopic scope,
  going into imaginative detail about plates,
  constellations, cosmic phenomena, microbial
  life, deep-ocean creatures, the silence of
  mountains, the shimmering of the light of a
  star.
```

The two texts serve distinct roles. The emotional inference is a transparent read-out, projected during performance so the audience sees the system's compositional reading in its own terms. The Synthetic Dream is the interpretative layer. In neither case does the model hear the music. It receives symbolic descriptors (pitch-class distributions, modal character, tension curves), not audio. This is an explicit design choice: the model interprets a compositional abstraction, not a perceptual experience.

Two representative outputs, generated by the system on recipes spanning the affective range of *Modes of Vibration AV*, illustrate the result. The first is from a luminous, modally grounded passage (78 per cent certainty):

> Gadi Sassoon emotional inference certainty 78%. Machine listening: a complex weave of delicate and powerful, with melancholy and light interwoven across time, intense energy and tension permeating the air. The music evolves from a tense, dark beginning to a soaring climax with a sense of release.
>
> In my synthetic dreamscape, I find myself floating within a vast sea of stars, each one pulsing in harmony with the piano journey. Microbial life thrives on the plates that orbit these stars, their bioluminescent glow painting the cosmic void in shades of pale blue and shimmering silver.

The second is from a turbulent, atonal passage in which the system reports its own diminished confidence (41 per cent certainty):

> Gadi Sassoon emotional inference certainty 41%. Using data metrics, I identify a blend of melancholy and tense emotions intertwined with the music. The overall intensity level is high; energy flows persistently, yet tension builds and dissipates dynamically throughout the piece.
>
> As I listen to the fusion of dark and tense emotions, my mind journeys through a nebulous expanse, where cosmic particles swirl in a delicate balance between order and chaos. Stars twinkle like distant tears on celestial plates, while microscopic organisms dance under a shimmering light. The deep ocean's eerie calm contrasts with the fiery constellations above.

Both texts run longer and more lyrical than the prompt's terse word limits specify; the seven-billion-parameter model honours the format loosely, the same productive non-compliance evident in the generated music. A buffered memory of the previous five inferences gives the output temporal coherence across a performance, averaging recent certainty values and accumulating references to earlier states so the texts develop a narrative arc rather than resetting at each frame. The Synthetic Dream is not a label applied to the music; it is a generated text with its own compositional trajectory, conditioned by but not reducible to the harmonic analysis that feeds it.

3.5 Bidirectional actuator: Closing the loop

Two output pathways reverse the analysis-to-text process, closing the compositional loop: the emotional parameters drive MIDI generation, and the Synthetic Dream conditions the visuals.

MIDI generation. A dedicated generator module (`teal_generator.js`, v2.7, 1,407 lines) receives the emotional parameters and produces MIDI note events. The module grew from an earlier text-to-melody prototype that read the Synthetic Dream directly, matching the narrative's affective keywords to register and dynamics and treating the stated certainty value as a control on rest density and pitch stability; the present generator works from the structured emotional parameters instead, trading literal text-parsing for fuller harmonic control while preserving the same text-to-music reversal in principle. The generator operates within four harmonic contexts, selected by the current modal analysis state, and each defines its own vocabulary of progressions and voicings. The functional context draws on tonal progressions ranging from ii–V–I and its tritone substitution to Coltrane changes; the modal context selects mode-specific progressions for each of the seven diatonic modes, such as a Phrygian cadence or Mixolydian flat-seven motion; the parallel context draws individual diatonic chords from the current mode; and the avant-garde context generates non-functional material, including quartal and quintal voicings, whole-tone and symmetric-diminished cycles, polychords and Messiaen mode-3 collections. Voicings are realised from a library of chord templates spanning triads to extended and altered jazz harmonies. The seven rhythm styles (steady, subdivided, syncopated, swing, phrase, groove and rubato) govern onset placement and subdivision, driven by an internal metronome (40–240 BPM); the rubato engine applies phrase-level tempo curves (arch-shaped phrasing, rallentando and accelerando) together with small agogic deviations, producing expressive micro-timing rather than mechanical quantisation. Because the generator emits standard MIDI, the same stream can drive electronic synthesis, an acoustic instrument or other embodied actuators. In *Modes of Vibration AV*, it is routed to Ableton Live, where it drives layers of synthesis shaped by custom timbral models trained on approximately eight hours of the composer's recorded corpus; in the Montreal sessions (Section 4.2, Figure 2), it is sent directly to an acoustic Yamaha Disklavier. These models interpolate between the generated MIDI and the sonic vocabulary of the compositional practice, ensuring timbral coherence even when the generated material diverges from the performer's live input. On an aesthetic level, the resulting MIDI output has an 'uncanny valley' quality, weaving between mechanical, repetitive phrases and surprisingly organic articulations.

The reversal is the architecturally distinctive feature. Music generates an emotional recipe; the recipe generates a Synthetic Dream and, through the same parameters, new music. The emotional parameters function as a compositional interface between analysis and generation, not as a recognition target but as a transduction layer.

Visual and physical actuation. The Dream is simultaneously truncated to a short textual prompt, the brief phrasing on which

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press

the diffusion model performs best, and sent via OSC to TouchDesigner. There, the prompt conditions real-time generative visuals, including StreamDiffusion-based animations and parametric particle systems. The visuals are grafted onto the performer's body via the same MediaPipe skeleton data used in gestural analysis, closing a visual feedback loop.

The result is a closed loop across multiple domains. The emotional parameters circulate between music, text, image and motion, functioning as the connective tissue of the system. It is this multi-domain temporal structuring, not the harmonic analysis alone, that constitutes the dramaturgy: the system organises relationships between performer, generated sound, projected text and visual environment across the duration of a performance.

3.6 Piece-specific calibration

TEAL is not a portable engine. It is calibrated to each work. In *Modes of Vibration AV*, calibration drew on the composer's idiomatic language: extended modal passages in D Dorian at the piano, violin glissandi traversing consonance and microtonal tension and the resonant properties of four slashed metal plates designed in collaboration with the NEMUS lab at the University of Bologna.

Calibration proceeds in two stages. First, the analysis windows are configured with interval templates and statistical thresholds matched to the harmonic vocabulary of the work. The extended window is tuned to detect the cadential pull of flattened sevenths; the sustained window tracks the persistence of modal centres characteristic of the piano improvisation style. Second, the mapping functions, the arbitrary emotional statistics, are iteratively rehearsed, aligning the composite output with the timbral models and the physical resonances of the plates.

If the same system were applied without modification to a performance in maqam, raga or gamelan, the interval templates and chord models would fail to register the modal logic, timbral hierarchies and rhythmic cycles of those idioms. This constraint is not a flaw but a structural consequence of the design: the emotional parameters are part of the composition, and the composition is culturally situated.

3.7 Sound sculpture resonance

In *Modes of Vibration AV*, TEAL actuates four large metal plates fitted with transducers and pickups, which together form an analogue feedback loop. The emotional recipe maps directly to resonant excitation patterns. High-energy, high-tension values increase amplitude and low-frequency excitation, producing dense metallic drones. Low-energy, high-confidence values emphasise higher partials and produce shimmering textures. Variations in the 'confidence' parameter determine how stable the resonances remain. Confidence is derived from the key-detection confidence score, the correlation strength of the Krumhansl–Schmuckler match, and thus reflects the harmonic clarity of the current musical passage, not a psychological property of the performer. The composer maps this technical metric to an aesthetic function: high confidence stabilises resonance; low confidence produces instability.

The plates extend the system into the physical domain. Vibration transmits through the performance space and through the performer's body. The performer feels the resonant energy and adjusts phrasing or dynamics in response, and the analysis windows register the change. The plates are both output and input.

4. Case studies

4.1 UN AI For Good summit, Geneva, 2025

The world premiere at the UN AI For Good summit (Sassoon 2025) provided a context in which the system's transparency and the performer's emotional sovereignty could be observed across a diverse audience of approximately 2,000 delegates, researchers and policymakers. The 15-minute performance comprised three sections. The first (six minutes) featured solo piano improvisation with real-time TEAL emotional inference and body tracking (Figure 3). The system's analysis was visible through projected confidence values and textual inferences. Throughout the performance, the emotional inference and Synthetic Dream texts were composited live in dynamic projection layouts, directed by Stefano Polli of Sugo Design. The improvisation moved through tonal and atonal territories; the system's varying confidence levels and its explicit acknowledgement when Western tonal frameworks proved insufficient were visible to the audience throughout.

The second section (three minutes) shifted to pre-composed sound sculptures manipulated live. This demonstrated a different mode of embedded agency: responsive manipulation of algorithmically generated materials rather than real-time inference.

The third section (six minutes) featured violin improvisation with body tracking and emotional inference driving live visuals, again directed by Polli. The electric violin explored extended techniques and real-time resynthesis; the visuals reflected the system's responses, including explicit uncertainty, through animations grafted onto the performer's skeletal tracking.

4.2 Disklavier sessions, Montreal, 2025

During a residency with the Wilding AI research-creation lab at the Université de Montréal studios in summer 2025, the author recorded a series of duets with TEAL on a Yamaha Disklavier grand piano (Figure 2). The Disklavier is a full acoustic piano that can simultaneously receive and execute MIDI input. In these sessions, the author improvised on the instrument while TEAL's generator module produced MIDI in real time. The Disklavier played both parts on the same keyboard: the performer's notes and the system's generated response sounded through the same strings and hammers.

This configuration physically instantiates the bidirectional compositional loop described in Section 3.5. The performer hears TEAL's response as acoustic piano, not as synthesised sound. The timbral consistency is total: both voices share the same instrument. The performer responds to the system's notes as to another musician. The system analyses the combined output, including its own contributions, and generates new material. The loop is audible, physical and contained within a single resonating body. In this configuration, the dramaturgy is purely acoustic: no visuals, no text projection and no external actuation. The temporal structuring of the performer-system relationship unfolds entirely through sound.

Four completed pieces from these sessions form the recording *TEAL Suite 1, Montreal*: *Starwhisp I. Voyage*, *Starwhisp II. Awaken The Sun*, *Starwhisp III. Cosmic Relay* and *Starwhisp IV. Rebirth* (submitted as supplementary sound examples 1–4). A second

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press



Figure 2. TEAL recording session at the Université de Montréal studios, 2025. Yamaha Disklavier grand piano with TEAL system visible on adjacent monitors. *Photo*: the author.

recording, *TEAL Suite 2, Vienna*, is in pre-production at the Vienna Synchron Stage, where TEAL will generate Disklavier parts alongside a live chamber orchestra.

4.3 Observations

Across both performance contexts, several informal responses recurred. At the UN summit, audience members noted the significance of seeing the system's confidence values and textual outputs. The visibility of algorithmic reasoning, particularly when the system admitted uncertainty, was among the responses most often remarked upon, distinguishing this approach from opaque recognition systems.

In the Montreal sessions, the compositional dimension was more apparent. The system amplified the performer's compositional decisions into material that the performer alone would not have produced. The generated MIDI introduced interval patterns and rhythmic configurations outside the performer's habitual vocabulary, while the performer's responses steered the system's subsequent output. The recordings capture this interaction as fixed compositions.

A recurring observation across both contexts concerned the staged withdrawal at the end of each performance, where the musical performance continued as the system's outputs gradually faded. Several observers read this as an enactment of human autonomy: meaning-making can continue without algorithmic mediation. This withdrawal is a compositional decision. It enacts sovereignty rather than claiming it.

5. Discussion

5.1 From recognition to amplification

Reorienting from recognition to amplification addresses the central critiques of affective computing without abandoning its creative applications. Rather than claiming to identify universal emotional states, TEAL magnifies situated expressive gestures, functioning as an amplifier rather than an evaluator.

This shift aligns with Jordà's (2005) argument that the most productive digital instrument designs extend human capability rather than replace human judgement and with Fiebrink and Caramiaux's (2018) finding that machine learning supports creative processes when positioned as a tool for improvisation rather than as an oracle of correctness. TEAL extends these principles to affective computing: the system does not dictate an emotional truth but scaffolds a compositional space in which the performer's gestures acquire new sonic, visual and narrative layers.

The bidirectional loop, in which music generates text and text generates music, materialises this principle. The audience perceives not a one-way pipeline but a continuous feedback loop in which each layer interprets, conditions and reshapes the others. The loop

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press



Figure 3. Sassoon performing on piano with TEAL at the UN AI For Good summit in Geneva, 2025. Visible on the projection screens are several layers of hand tracking, the emotional inference metrics and the LLM's Synthetic Dream elicited by the performance. *Photo credit*: ITU/AI For Good.

resonates with cybernetic accounts of human–machine coupling (Wiener 1948; Pask 1971), but here it is realised in the audiovisual domain: vibrating metal plates, modulated synthesis, algorithmic text and real-time visuals.

Amplification, however, is not a simple concept. Transparent amplification, where output perfectly tracks input, would require the system's emotional parameters to correspond exactly to the performer's expressive intent. TEAL does not claim this correspondence. The parameters are compositional assignments; they approximate but do not replicate the performer's internal state. Consequently, the system introduces distortion: the emotional recipe diverges from what the performer intended, and the Dream generated from that recipe diverges further. The performer responds to this distorted output, adjusting their playing, which generates new analysis, which generates new distortion. Over time, the loop may converge towards attractor states, stable configurations where the performer's adjustments and the system's responses reach equilibrium, or it may oscillate, producing unpredictable alternations between emotional profiles. In practice, both behaviours occur within a single performance, and the transitions between convergence and oscillation are themselves compositionally productive.

Negative feedback, where the system's output opposes the performer's intent, is not explicitly designed for but emerges naturally when the arbitrary parameter mappings produce readings that conflict with the performer's expressive direction. When a performer intends intensity but the harmonic analysis registers the passage as 'delicate' (because the interval distribution falls within that mapping), the resulting Dream and generated MIDI pull the texture in a direction the performer did not intend. The performer can override this by changing their playing, reasserting sovereignty, or they can follow the system's misreading, allowing it to redirect the composition. Both responses are legitimate within the dramaturgy.

The critical point is that the emotional layer does not claim to know what the performer feels. It maps compositional patterns to named parameters through rules the composer defined. The system performs its own bias, and that performance is the work. In this sense, TEAL functions less like an autonomous composer than like the cognitive equivalent of a guitar amplifier: it magnifies, colours and feeds back the performer's signal. The system does generate material, including patterns and narrative structures the performer would not have produced unaided, but it generates within an affective ground and a calibration that the composer fixed in advance, so authorship of that frame stays with the human rather than the system. The distinction is architectural rather than incidental: every design decision, from the composer-defined parameter mappings to the piece-specific calibration, keeps the system an amplifier of human compositional intent rather than an originator of it.

5.2 Emotional sovereignty in practice

The concept of emotional sovereignty extends beyond conventional notions of user control or transparency. It resonates with Suchman's (2007) notion of located accountability, the imperative to make visible the specific assumptions embedded in sociotechnical systems. In affective computing, sovereignty requires that

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press

performers and audiences can see how emotional inferences are generated, contest their validity and continue expressive practice independent of them.

Behavioural economics provides an additional framing. Kahneman and Tversky's (1979) prospect theory and Thaler and Sunstein's (2008) work on nudging both resonate with the idea that individuals place asymmetric weight on retaining agency, often preferring autonomy over algorithmically optimised outcomes. For affective computing, this suggests that preserving interpretive control over emotional states may be more valuable than any purported optimisation by an algorithm.

Emotional sovereignty as introduced here is distinct from existing concepts of user control in HCI. It is not merely the ability to override a system. It is the right to define the affective vocabulary (the named parameters, the interval mappings, the thresholds) on which the system operates. In TEAL, sovereignty is structural: the composer writes the emotional ground rules as part of the composition. The system amplifies within those rules. It does not impose its own.

The staged withdrawal at the end of each performance enacts this principle. The system acknowledges its limits and cedes space to human intentionality. This models a future for affective computing in which algorithms enter and exit dramaturgically rather than aspiring to permanent interpretive dominance.

5.3 Limitations and future directions

The system's analytical core is grounded in Western tonal and modal logics. Future work should develop culturally specific analysis modules for maqam, raga and gamelan as parallel situated frameworks, developed in collaboration with practitioners of those traditions, allowing performers to choose or hybridise perspectives while remaining transparent about assumptions. Building a performer-specific calibration currently requires approximately 40 hours of rehearsal and annotation; few-shot and transfer learning approaches could reduce this while preserving stylistic specificity.

The locally hosted language model balances speed with feasibility but limits narrative sophistication. Larger models could enrich the Dream at acceptable latency costs. Short-term memory currently maintains phrase-level continuity; future designs could add retrieval-augmented memory accessing rehearsal and performance archives, enabling the Dream to develop historical awareness across performances.

Future performances should incorporate mixed-method evaluation, combining surveys, interviews, physiological measures and ethnographic documentation to triangulate audience perception of transparency, sovereignty and amplification. A complementary safeguard concerns scale: as memory depth, modalities and analytical granularity increase, systems risk drifting back towards implicit recognition, so design guardrails such as explicit confidence displays, performer overrides and compositional exit cues are needed to ensure that capability gains do not erode amplification and sovereignty.

6. Conclusion

This investigation demonstrates that embedding emotional-inference algorithms as compositional agents transforms their function from recognition to amplification. The key design insight is that emotional parameters need not be scientific claims about affect; they can be arbitrary compositional assignments, part of the score, structuring the relationship between analysis, generation and actuation.

The paper makes three contributions. First, it demonstrates a viable alternative to universal emotion recognition through bespoke compositional embedding: rather than seeking generalised models that claim to detect affect across contexts, TEAL shows that piece-specific calibration produces meaningful expressive augmentation without universalising assumptions. Second, it shows how transparency and embedding can address persistent ethical concerns about affective computing, particularly critiques of surveillance, bias and the reification of emotion as measurable data. Third, it introduces emotional sovereignty as a framework for human–AI interaction that extends beyond music, proposing that performers and users should define the affective vocabulary on which algorithms operate.

The bidirectional architecture, in which music generates text and text generates music through a compositional interface of named emotional parameters, offers a replicable design pattern. The explicit performance of algorithmic limitations, including visible confidence values, acknowledged cultural bias and staged withdrawal, models how transparency can function as aesthetic material rather than mere disclosure.

The principles demonstrated here – embedding, transparency, personalisation and bidirectionality – constitute a design methodology for affective systems that serve compositional agency rather than extractive recognition. The future of emotion AI in creative practice lies not in perfecting detection but in refining the compositional use of inference: systems calibrated to specific works, transparent about their limitations and subordinate to the performer's control.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1355771826101563

Acknowledgements. The author thanks the UN AI For Good summit for providing the platform for this work's premiere, Stefano Polli of Sugo Design for visual direction, the NEMUS lab at the University of Bologna for collaboration on the sound sculptures and the Wilding AI research-creation lab and the Université de Montréal for the residency during which the Disklavier sessions were recorded. Dr Alexandre Saunier provided proofreading and critical feedback on the manuscript. AI language model assistance (Claude, Anthropic) was employed for copyediting and formatting adherence. All research content, analysis and arguments are the author's original work.

References

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M. and Pollak, S. D. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements. *Psychological Science in the Public Interest* **20**(1): 1–68.

Birhane, A. and Guest, O. 2021. Towards Decolonising Computational Sciences. *Kvinder, Køn & Forskning* **29**(1): 60–73.

Briot, J.-P., Hadjeres, G. and Pachet, F.-D. 2020. *Deep Learning Techniques for Music Generation*. Cham: Springer.

Calvo, R. A., Dinakar, K., Picard, R. and Maes, P. 2016. Computing in Mental Health. CHI'16 Extended Abstracts, 3438–45.

Choi, K., Fazekas, G., Sandler, M. and Cho, K. 2017. Convolutional Recurrent Neural Networks for Music Classification. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2392–6.

Collins, N. 2004. Composers inside Electronics: Music after David Tudor. *Leonardo Music Journal* **14**: 1–3. https://doi.org/10.1162/0961121043067352.

Collins, N., McLean, A., Rohrhuber, J. and Ward, A. 2003. Live Coding in Laptop Performance. *Organised Sound* **8**(3): 321–30.

Crawford, K. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

D'Mello, S. and Graesser, A. 2012. Dynamics of Affective States during Complex Learning. *Learning and Instruction* **22**(2): 145–57.

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press

Eerola, T. and Vuoskoski, J. K. 2013. A Review of Music and Emotion Studies: Approaches, Emotion Models, and Stimuli. *Music Perception* **30**(3): 307–40.

Eigenfeldt, A. and Pasquier, P. 2013. Evolving Structures for Electronic Dance Music. Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation (GECCO), 319–26.

Fiebrink, R. and Caramiaux, B. 2018. The Machine Learning Algorithm as Creative Musical Tool. In R. Dean and A. McLean (eds.) *The Oxford Handbook of Algorithmic Music*. Oxford: Oxford University Press, 181–208.

Gómez, E. 2006. Tonal Description of Music Audio Signals. PhD thesis, Universitat Pompeu Fabra.

Han, D., Kong, Y., Han, J. and Wang, G. 2022. A Survey of Music Emotion Recognition. *Frontiers of Computer Science* **16**(6): 166335.

Jordà, S. 2005. Digital Lutherie: Crafting Musical Computers for New Musics' Performance and Improvisation. PhD thesis, Universitat Pompeu Fabra.

Juslin, P. N. and Laukka, P. 2004. Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. *Journal of New Music Research* **33**(3): 217–38.

Juslin, P. N. and Sloboda, J. A. (eds.) 2001. *Music and Emotion: Theory and Research*. Oxford: Oxford University Press.

Juslin, P. N. and Sloboda, J. A. (eds.) 2010. *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford: Oxford University Press.

Kahneman, D. and Tversky, A. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47**(2): 263–91.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A. and Turnbull, D. 2010. Music Emotion Recognition: A State of the Art Review. Proceedings of the International Society for Music Information Retrieval Conference, 255–66.

Kirke, A. and Miranda, E. R. 2009. A Survey of Computer Systems for Expressive Music Performance. *ACM Computing Surveys* **42**(1): 1–41.

Krumhansl, C. L. 1990. *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.

Lerdahl, F. 2001. *Tonal Pitch Space*. New York: Oxford University Press.

Lewis, G. E. 2000. Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal* **10**: 33–9.

Livingstone, S. R., Muhlberger, R., Brown, A. R. and Thompson, W. F. 2010. Changing Musical Emotion: A Computational Rule System for Modifying Score and Performance. *Computer Music Journal* **34**(1): 41–64. https://doi.org/10.1162/comj.2010.34.1.41.

Louie, R., Coenen, A., Huang, C. Z., Terry, M. and Cai, C. J. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., et al. 2019. MediaPipe: A Framework for Building Perception Pipelines. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1–4.

Magnusson, T. 2014. Scoring with Code: Composing with Algorithmic Notation. *Organised Sound* **19**(3): 268–75.

McCormack, J., Gifford, T., Hutchings, P., Llano, M. T., Yee-King, M. and d'Inverno, M. 2019. In a Silent Way: Communication between AI and Improvising Musicians Beyond Sound. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, paper 38.

McStay, A. 2018. *Emotional AI: The Rise of Empathic Media*. London: SAGE Publications.

Nettles, B. and Graf, R. 1997. *The Chord Scale Theory & Jazz Harmony*. Mainz: Advance Music.

Pask, G. 1971. A Comment, a Case History and a Plan. In J. Reichardt (ed.) *Cybernetics, Art and Ideas*. London: Studio Vista, 76–99.

Picard, R. W. 1997. *Affective Computing*. Cambridge, MA: MIT Press.

Reichardt, J. (ed.). 1968. *Cybernetic Serendipity: The Computer and the Arts*. London: Studio International.

Roads, C. 2001. *Microsound*. Cambridge, MA: MIT Press.

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

Sassoon, G. 2025. Live at the UN AI For Good Summit, Geneva [video]. https://www.youtube.com/watch?v=SP6g8lTldOk (accessed 29 June 2026).

Schedl, M., Gómez, E. and Urbano, J. 2014. Music Information Retrieval: Recent Developments and Applications. *Foundations and Trends in Information Retrieval* **8**(2–3): 127–261.

Stark, L. and Hutson, J. 2022. Physiognomic Artificial Intelligence. *Fordham Intellectual Property, Media and Entertainment Law Journal* **32**(4): 922–78.

Suchman, L. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd edn. Cambridge: Cambridge University Press.

Susino, M. and Schubert, E. 2019. Cultural Stereotyping of Emotional Responses to Music Genre. *Psychology of Music* **47**(3): 342–57.

Tatar, K. and Pasquier, P. 2019. Musical Agents: A Typology and State of the Art towards Musical Metacreation. *Journal of New Music Research* **48**(1): 56–105.

Temperley, D. 1999. What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered. *Music Perception* **17**(1): 65–100.

Thaler, R. H. and Sunstein, C. R. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Thelle, N. J. W. and Pasquier, P. 2021. Spire Muse: A Virtual Musical Partner for Creative Brainstorming. Proceedings of the International Conference on New Interfaces for Musical Expression (NIME).

Wiener, N. 1948. *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.

Yang, Y. H. and Chen, H. H. 2012. Machine Recognition of Music Emotion: A Review. *ACM Transactions on Intelligent Systems and Technology* **3**(3): 1–30.

https://doi.org/10.1017/S1355771826101563 Published online by Cambridge University Press