STRATEGIES FOR MUSICAL ENTANGLEMENT

SHAPING CO-CREATIVE INTERPLAY THROUGH HARMONIC AND MELODIC VISUALIZATION OF INTENT IN SOMAX2

Marco FIORINI (marco.fiorini@ircam.fr) [1, *] and **Miles OKAZAKI** (milesokazaki@princeton.edu) [2, *]

[1] **Sorbonne Université, IRCAM, CNRS, Ministère de la Culture, STMS**, F-75004 Paris, France
[2] **Princeton University**, *Department of Music*, Effron Music Building, Princeton, NJ 08544 USA

[*] *The two authors contributed equally to this work*

ABSTRACT

This paper explores how visual feedback can actively shape real-time co-creative interaction between human improvisors and music AI systems. While corpus-based systems such as Somax2 offer high reactivity and sophisticated responsiveness, the lack of transparency regarding the machine's internal state and choices can limit the human response choices to a passive reactive dynamic. To address this issue, we have developed two new user interface (UI) modules that allow the performer to "sight-reading the machine": a chord-symbol-based module for harmonic navigation (used with Thelonious Monk's *Introspection* as the corpus of source) and a staff-based reading module for melodic anticipation (using the Max library `bach`). We discuss case studies from two live performances (ICMC2025 in Boston and the Improtech Festival at IRCAM in Paris), demonstrating how the visualization of harmonic and melodic intent transforms the human-machine relationship into a mechanics of interplay that unfolds in real time, enabling new strategies for counterpoint, synchronization, and shared action.

1. INTRODUCTION

The theme of "interplay" implies a bidirectional exchange of intent [1]. In human-AI improvisation, however, this exchange is often asymmetrical. The machine "listens" to the human's audio signal and reacts, but the human must wait for the machine's audio output to understand its decision. This "reactive lag" can restrict the interplay to a following game, where the human is constantly catching up to the machine's choices.

This leads us to the search for a dialectical relationship with the machine, where both parties are deeply involved in the information and intent of the other and are able to create spontaneous synchronized musical events, as opposed to the call and response model. We could think of this situation as a kind of creative human-machine entanglement.

Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

To reach true co-creative interaction, we propose the concept of "sight-reading the machine" [1]. By visualizing the intent of the generative agent in the form of transposable chord symbols or anticipated musical notation — concepts with which a professional improviser is highly comfortable — we reintroduce the musician's literacy into the performance's feedback loop, unlocking new layers of interaction and interplay.

This article, starting from the musicological perspective of working with a specific performer's language and repertoire, investigates notions of human-machine co-improvisation, details the technical implementation of dedicated new tools in the Somax2 environment [2] and evaluates how they influence improvisational strategies used in real-world professional live performances.

2. BACKGROUND AND MOTIVATION

This project sits at the intersection of the jazz improvisation tradition and interactive music systems design [3]. The improvisational approach used in this research is rooted in the lineage of the music known as *Jazz* and the various branches of music that came from artists such as Louis Armstrong, Charlie Parker, Thelonious Monk, Sonny Rollins, Bud Powell, John Coltrane, etc. Improvisation in this tradition is highly linked to language [4], and internalization of the performances of master improvisors. People in these circles of creative music have different levels of engagement with this tradition (and often are from an entirely different culture and tradition), but we can assess that this lineage of American improvisors is the source of the central discipline.

2.1 The Collaborative Framework

The research emerges from a dialogue between two expert improviser-technologists. Miles Okazaki is a guitarist and composer active in the contemporary jazz scene, whose performing career runs parallel to a sustained practice in creative coding [2]. Marco Fiorini is also a guitarist and improviser, as well as a researcher and computer music de-

[1] We use "sight-reading" in its conventional musical sense: visually reading notated (or, here, symbolically displayed) material and translating it into performance in real time, as distinct from a more general sense of "monitoring" or "supervising" an agent's status.

[2] https://www.milesokazaki.com/about/bio-long

signer in the Music Representation team at IRCAM, where he contributes to the development of Somax2 under the REACH project [3].

This dual perspective, where both authors are expert instrumentalists and coders, shaped the overall research trajectory and the implementation design of the system detailed here. The collaboration evolved across two distinct configurations:

- **June 9, 2025 – International Computer Music Conference** *(Fenway Center, Boston)* [4]: Fiorini acted as a computer music designer and consultant, assisting in the annotation of the Monk corpus and the implementation of the custom label module, while Okazaki performed solo during the concert, managing the real-time interaction with the machine.
- **December 7, 2025 – Improtech** *(La Dynamo, Paris)* [5]: The partnership expanded into a live duo. Fiorini performed as a live electronic musician, navigating the Somax2 agents and manipulating the corpus in real-time, while Okazaki improvised on guitar, using the staff-reading module, sight-reading the system's output displayed on a screen in front of him.

2.2 Improvisation as Algorithmic Thinking

For the improviser (Okazaki), the engagement with technology is driven by a need to transcend the mere "shuffling of previously established utterances." The goal is to identify systems that can reliably provoke entirely new musical territories within the constraints of real-time performance — what Okazaki describes as the "heat of the moment." In this context, improvisation is reframed not just as intuition, but as "algorithmic thinking mixed with emotion." The creative process involves articulating specific musical movements and then seeking variations or extrapolations of those movements that lie beyond the performer's immediate cognitive reach. Generative systems serve a crucial heuristic function here: they help "solve musical problems" by revealing options that the musician cannot see due to habitual pathways. Critically, this requires a rejection of "prefabricated software"—intended as general-purpose production environments—which, despite their flexibility, still embed default workflows and stylistic assumptions about the materials and methods of music-making. Instead, custom-built environments (such as Okazaki's own *Dogstar* patch, discussed in §3) allow the performer to define the logic of *agent–performer interplay*—as distinct from software interaction in the general HCI sense—from the ground up, ensuring that each tool is shaped around the specific musical language of its performer rather than around an inherited, external aesthetic.

2.3 The Computational Framework

We selected Somax2 as the primary computational environment for this project due to its sophisticated handling of multi-agent interactions and its capacity for high-level symbolic reasoning within a reactive framework. Somax2 is a multi-agent environment for computer-assisted co-improvisation, implemented as a hybrid architecture between Max and Python [2]. It operates through a reactive model where autonomous *players* (generative agents) interact with a human musician by navigating a multi-dimensional *corpus*. Crucially, this navigation is stochastic rather than deterministic: identical input from the human performer need not yield identical output from the agent, which distinguishes the system's behavior from simple mimicry even though its overall dynamic remains fundamentally reactive. This system is conceptually rooted in Corpus-Based Concatenative Synthesis (CBCS), a paradigm established by tools such as CataRT [5], where musical material is re-synthesized by navigating a database of sound atoms based on their descriptors. Several human-machine composition and improvisation systems have emerged from this paradigm, particularly through research conducted within the REACH project at IRCAM. Somax2 belongs to a well-established lineage of co-creative tools including Omax [6], Djazz [7], and Dicy2 [8]. While Somax2 proved to be an extensively robust and state-of-the-art framework for real-time co-improvisation, our practice-based research sought to move beyond its standard operational modes toward a higher degree of *symbolic interpretability*—that is, rendering the agent's internal symbolic state legible to the performer in real time, as distinct from the causal, model-internal sense in which "explainability" is typically used in machine learning.

To this end, we developed a suite of new, modular interface components designed to work alongside the standard Somax2 environment. These modules do not replace the core engine; rather, they function in parallel to externalize the agent's internal logic and symbolic metadata. By utilizing the features introduced in the latest 2.7.0 release [6], we expanded the corpus definition beyond signal-level features (such as pitch, chroma, and MFCC) to include custom symbolic *labels* as new *influences* [2]. This allows the performer to navigate the corpus not just as a sonic database, but as a structured architectural map of user-defined symbolic information, enabling real-time guidance of the agents as established in previous work [9].

3. CASE STUDY I: ICMC BOSTON

The first performance, titled *Introspectronica*, took place at the International Computer Music Conference at the Fenway Center in Boston, on June 9, 2025. The goal of this performance was to integrate Somax2 into a hybrid electric guitar and live electronics performance that was rooted in traditional Jazz language but could follow a completely new aesthetic path.

[3] https://reach.ircam.fr/

[4] https://icmc2025.sites.northeastern.edu/concerts-hall-part-1/

[5] https://improtech.ircam.fr/ikparis/pages/programme.html

[6] For technical specifications and release details of Somax 2.7.0, see https://forum.ircam.fr/article/detail/somax-version-270-is-out/.

3.1 Monk

A central focus of our first case study was the music of Thelonious Monk. Monk is viewed here as a pioneering musical architect, whose creations are structurally robust and amenable to abstract manipulation. His works often contain *sonic synecdoches*: single events or unique voicings that are so distinctive that they immediately signify the whole composition. The performer's history with this repertoire is extensive: Okazaki recorded the complete works of Monk for solo guitar in 2018 (a 6-CD set) [7], performed the entire cycle in a single concert in 2024 at The Jazz Gallery, in New York City. This deep familiarity informed the choice of *Introspection* as the Somax2 corpus source, a piece selected specifically for its navigational difficulty—that is, the unusually dense and unpredictable combination of harmonic and rhythmic information it requires an improviser to negotiate in real time, detailed below.

The composition presents unique challenges for both human and machine:

- **Melodic Density:** The initial ascending melody utilizes all 12 tones (from C# up to C).
- **Harmonic Volatility:** The progression traverses a serpentine, constantly modulating path that appears to land in $A\flat$ but abruptly flips to D major. The bridge features a slow seesaw between chromatically adjacent major key centers.
- **Polychordal Structures:** The piece contains standard chord types alongside "unnameable" vertical structures [8], such as the $A, E, B\flat, F$ cluster that resolves to D major at the end of the A sections.

To preserve the architectural integrity of the piece, we thus decided to annotate the corpus with a multi-layered labeling system, relying on the recent modules introduced in March 2025 in the 2.7.0 release of Somax2, which allow the use of any type of user-defined symbolic labels for corpus annotation to be taken into account by the system's cognitive model [10]. Events were described not only by standard chord symbols (providing a familiar "harmonic snapshot" for the improviser) but also by qualitative descriptors, such as *amen* for specific plagal cadences or *crunch* for high-dissonance events. In total, nine such qualitative types were defined for this corpus (*mi*, *ma*, *dom*, *aug*, *line amen*, *crunch*, *end* and *final*); the number and nature of these labels is not fixed a priori, but depends on the harmonic and gestural variety actually present in a given source corpus. The organization of the three labels we defined for each event in the corpus annotation process – index, type, chord symbol – is illustrated in Table 1.

Time (s)	Index	Type	Chord Symbol
0.000000	1	mi	Bmi7
0.664892	2	mi	Cmi7
1.388915	3	dom	Db7b5
...	...	...	...
7.339234	10	crunch	Bb7b9
10.174654	14	crunch	A7b9b13
10.758457	15	ma	Dma7
...	...	...	...
44.880194	32	amen	Ab7
46.189561	33	amen	Dbma7
48.111429	35	final	Db7

Table 1: Representative segmentation of the *Introspection* corpus showing the three used annotation families: index, qualitative type, and chord symbol.

3.2 Transposition Logic and Visual Feedback

In the preparation of the module, Okazaki identified transposition of the corpus events as the most crucial parameter for creating sonic variety. This draws on the defining feature of the Somax2 architecture in its capacity to transpose corpus slices in real-time to match the harmonic or melodic context of the live input. While this ensures a cohesive sonic result, it creates a semiotic discrepancy between the original label of a corpus slice and its sounding pitch. To resolve this for the performer, our custom module intercepts the agent's internal transposition logic to calculate and visualize the resulting chord symbol instantaneously. Let C_{orig} represent the root of the original chord label in the corpus (mapped to the integer range $0 \dots 11$), and T represent the transposition interval in semitones dynamically applied by the Somax2 agent. The displayed chord label root C_{disp} is defined as:

$$C_{disp} = (C_{orig} + T) \pmod{12} \quad (1)$$

By maintaining the original chord quality (e.g., "maj7", "min7") while updating the root, the interface provides an accurate harmonic roadmap. For example, if the system selects a $Cm7$ slice and applies a tritone transposition ($T = 6$), the interface immediately displays $F\#m7$.

The decision to enable these transpositions was a deliberate choice rather than an autonomous default setting of the system. We implemented it to reflect the fluid and non-linear nature of performance, forcing the system to recontextualize Monk's corpus based on its interplay with the human musician. This visual feedback transforms the role of the performer: in addition to following the machine by ear, the human improviser has the ability to react synchronously to the machine's harmonic intent thanks to the encoding of a language with which they are familiar in performance contexts. This multimodal engagement ensures that the musician can actively navigate and influence the machine's recontextualization of the material in real time. [11].

3.3 The Label Module Implementation

To address the informationally dense harmonic content of *Introspection*, we developed a custom Label Module in Max (see Figure 1). This interface serves as a real-time lead-sheet generator: it takes the Somax2 agent's currently active corpus metadata—chord symbol, qualitative type, and temporal index (§2.3)—and renders it in the familiar

[7] https://www.milesokazaki.com/albums/work

[8] By "unnameable" we mean structures that resist reduction to a single conventional jazz chord symbol, not that they are unclassifiable in principle—any such sonority can of course be described through pitch-class-set analysis, register, and voicing.



Figure 1: The Somax2 interface during the *Introspectronica* performance. The custom Label Module (top right) displays the instantaneous transposed chord symbol (here Eb7b9) and event type descriptor, allowing the performer (Miles Okazaki, bottom right) to visually navigate the harmonic intent of the machine in real time.

layout of a jazz lead sheet, i.e., transposed chord symbols displayed along a time-line, a format an improviser can read without additional decoding effort. The development was an iterative process conducted through remote prototyping sessions via Zoom, allowing immediate refinement of the visual hierarchy based on the feedback of the performer. The module functions by intercepting the agent's slice selection and processing it through a dedicated subpatcher, implementing the transposition logic established in Eq. 1, ensuring the displayed symbol matches the sounding pitch regardless of how the agent has transposed—and thereby recontextualized—the original corpus annotation. The interface provides parallel access to three specific descriptor streams:

1. **Chord Symbol:** (e.g., *Eb7b9*) for immediate harmonic navigation;
2. **Event Type:** Qualitative descriptors (e.g., *crunch*, *amen*) that reveal the structural function of the corpus segmentation;
3. **Temporal Indexing:** The *event_index*, allowing the performer to track the machine's position within the temporal structure of the piece and make deliberate structural choices.

These three descriptor streams are not merely informative: each can be individually gated—allowing, for instance, only events labeled *mi*, or only events within a given index range, to be selected—so that they also function as a control layer shaping the agent–performer interplay. This filtering mechanism, building on the one previously established in [9], enables the performer to use these streams of information to guide the co-creative agent in real-time. For example, a section of performance where the filter is gated to a single parameter (say, "minor 7th") with low transposition settings will differ drastically from a section with no filter (all events allowed) and high transposition settings. In *Introspectronica,* the arc of the performance was achieved by progressing through a series of saved states of the module filters and Somax2 settings, and the performer's reactions to the output and visual feedback.

4. CASE STUDY II: IMPROTECH PARIS

The second performance took place at the *Improtech*[9] concert at La Dynamo, Paris, on December 7, 2025. While the Boston case study prioritized harmonic navigation via symbolic labels, the Paris session explored detailed melodic interaction. Notably, one Somax2 player was loaded with a MIDI corpus consisting of a single-line improvisation generated by George Lewis's *Voyager* system [12].

4.1 The Staff-Reading Module

Translating real-time generative output into readable notation presents an immediate temporal paradox: a performer cannot sight-read an event that has not yet been computationally determined. To resolve this, we employed the *bach* library for Max [13] to construct a "Look-Ahead" interface (see Figure 2). The module operates through a four-step temporal buffering mechanism:

1. **Generation:** Somax2 calculates a stream of MIDI events based on the live instrumental influence.

[9] https://www.banlieuesbleues.org/dynamo/evenement/5318/2025-12-07-1800-dynamo-reachin-paris-improtech-paris-2025/

2. **Preview Buffer:** These events are routed to a lower `bach.roll` display (the Preview Staff"), rendering the machine's upcoming choices visually.

3. **Delay:** The corresponding audio/MIDI playback is held in a fixed delay line of 4000 ms (4 seconds).

4. **Execution:** As the delay elapses, the notes transfer from the lower Preview Staff to the upper Active Staff," guided by a vertical playhead indicating the "now" moment.

This 4-second window effectively simulates the cognitive process of look-ahead sight-reading, providing the human musician with the necessary visual lead-time to anticipate upcoming pitch content, prepare physical fingerings, and formulate intentional contrapuntal responses. While the software allows for real-time adjustment of this look-ahead duration, we found that maintaining a fixed value throughout the performance was essential for the performer's stability. A consistent temporal offset allows the musician to calibrate their internal pulse and predictive gaze, whereas a variable buffer would disrupt the established relationship between the visual preview and the delayed audio execution.

4.2 Interface Design

A critical design decision involved the use of stemless spatial notation. Because Somax2 preserves the micro-timing and organic groove of the original corpus, quantizing this data to a rigid metrical grid would generate unreadable clusters of nested tuplets and complex ties. Consequently, we adopted a spatialized representation where solid noteheads indicate pitch, and rhythmic duration is communicated entirely by the horizontal distance between events. This approach preserves pitch accuracy while leaving the micro-rhythmic interpretation open, encouraging the performer to feel the temporal flow rather than mathematically decode it.

4.3 Stage Layout

The module was deployed as a dedicated visual anchor on stage, displayed on a high-resolution monitor positioned within Okazaki's primary line of sight. This arrangement allowed the performer to maintain a natural instrumental posture while treating the digital staff as a dynamic, real-time score. A deliberate design choice was the functional decoupling of the performer's interface from the system's control parameters. While Fiorini actively played the Somax2 environment on stage, adjusting player parameters and influencing the performance through his musical decision making, these technical operations remained invisible to Okazaki. By isolating the visual feedback to the visual notation staff, we filtered out the underlying diagnostic data, ensuring that Okazaki could focus exclusively on the musical output. This setup reinforces the role of the module as a specialized tool for *embodied sight-reading*—the direct translation of a visual cue into an instrumental gesture, without an intervening step of abstract analysis—rather than a software controller. It maintains a clear division of labor: Fiorini's management of the generative engine's parameters on one hand, and Okazaki's creative curation of the resulting musical material on the other—that is, Okazaki's own moment-to-moment choices to align with, depart from, or withhold response to what appears on the staff.

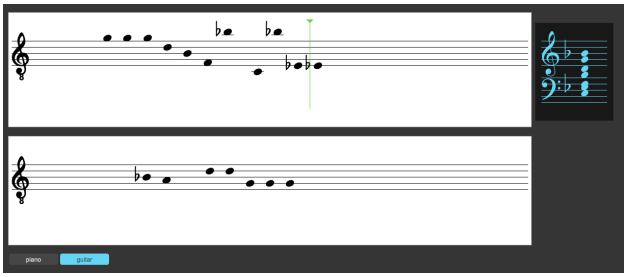


Figure 2: The Staff-Reading Module interface. The lower staff displays the preview material (4-second look-ahead), while the upper staff shows the active sounding material. The side visualizer provides a vertical harmonic snapshot of the current notes, while instrument toggles allow for real-time re-notation.

5. QUALITATIVE EVALUATION

Post-performance analysis and interviews with the performer revealed several key insights regarding the nature of interplay facilitated by these tools.

5.1 Performance Analysis: Boston Concert

The debut of *Introspectronica* at ICMC in Boston (June 9, 2025) was not recorded from the mixing board. For this reason, the performer made another version of the piece in his home studio using the same setup and roadmap as the live performance [10]. This video has the added benefit of the UI being fully visible so that the viewer can follow the workings of the human/machine interaction. There are a variety of stages to the performance, detailed as follows:

At [0:00] the performer initiates the Dogstar patch to play a rhythmic cycle of 21/8 that is also using the melody of Monk's "Introspection" as material, which in turn is triggering events from the corpus ("Introspection" from *Solo Monk,* 1964), which is tagged with metadata on each event. At [0:50] the window switches to reveal Somax2 with the corpus filter and some saved states to progress through the performance. These savestates were required because the performer was operating both the guitar and the computer.

At this point, the filter only allows chords with the "mi" tag, meaning any chord whose "type" is minor. At [1:55] the performer changes the filter to "ma." This is followed by an episode of improvisations by Dogstar. At [2:41] the filter is turned off, allowing events of all types through. The performer continues turning it on and off to create episodes of relative activity and inactivity. During this time

[10] *Introspectronica* performance available at https://www.youtube.com/watch?v=YKT1m8lf7k8

he also adjusts the channels so that Dogstar produces drum sounds instead of melodies. At [4:38] the performer begins improvising on guitar along with the machine. At this point the chord symbols are being produced too quickly for human processing of each one, so the improvisation is only skeletally related to the filter, and is more concerned with providing a rhythmic texture with the drum sounds.

For the purposes of this discussion, the most important part begins at [5:45], when the performer limits the output of the filter to a sparsely produced string of harmonies which are displayed on the UI. This continues for several minutes with gradually increasing velocity and complexity, reaching the end of the episode around [9:15]. During this section, Somax2 is responding to the performer's improvisation, which is in turn feeding off the chordal output of the machine.

At [11:00] the performer records another version of "Introspection" as a live corpus. This leads to a section where Somax2 players 1 and 2 are listening to and responding to each other, and the performer is improvising over the composite sound. At [14:30], the performer brings Dogstar back in as a Somax2 MIDI influencer, so that there are now three versions of the song talking to each other. He joins in on guitar to finish the piece off with fragments of the melody and improvisation.

5.2 Performance Analysis: Paris Concert

The performance in Paris at Improtech (December 7, 2025) marked the debut of the staff-reading module developed by Okazaki and Fiorini [11] . While the Boston concert utilized a form of shorthand notation (chord symbols and descriptions), the module implemented here provides precise notation unencumbered by assumptions regarding style or genre.

The performer begins with a brief improvisation to initialize the system, engaging directly with the Somax2 output at [0:34]. At this stage, the approach is a dialogue between guitar improvisation and "checking in" with the system by targeting upcoming phrases displayed in the look-ahead buffer. By [1:04], the strategy shifts toward counterpoint, with the performer crafting melodies that intersect with and depart from the Somax2 output; this is clearly audible at [1:34], for example.

At [3:09], the performer utilizes the visual cues to anticipate a shift in density, engaging an overdrive effect to match the machine's increased activity. This transition demonstrates the value of the look-ahead mechanism; without it, the timbral shift would have been reactive and likely delayed. The performance continues into a passage focused on matching phrasing, register, and activity rather than specific notated material.

From [5:42] to the conclusion, the performer returns to a focus on tethering the guitar lines to the machine. This follows a structural arc typical of traditional performance—moving from a theme to variations and returning to a unison statement. Crucially, as no material is prepared beforehand, the entire dynamic of linked and unlinked playing is generated spontaneously through the system.

Reflecting on the challenges observed in Boston regarding the difficulty of simultaneously managing Somax2 parameters and the instrument, the move to a duo format proved essential. This division of labor eliminates the need for switching attention, preventing interruptions in the flow and fostering a more organic performance.

5.2.1 Agency and Interplay Strategies

The introduction of the visual channel fundamentally shifted the perception of agency in performance. In typical human-AI interaction, agency is often viewed as a zero-sum game: if the machine leads, the human follows. However, the reading modules enabled a form of shared agency where the performer could interact with the machine's future. The performer noted three distinct modes of interplay enabled by the staff module—each reflecting a different degree of entrainment with the machine, rather than a fixed leader–follower role [14]:

1. **Synchronized Unison ("Playing With"):** By reading the preview accurately, the performer could play in near-perfect unison with the machine, creating the sonic illusion of a pre-composed, rehearsed passage.

2. **Counterpoint ("Playing Against"):** The performer observed that blindly following the notation could, at times, lead to stagnation: an anecdotally-observed feedback loop in which the system mimics the human and the human, in turn, mimics the system. The notation allowed the performer to consciously diverge, weaving counter-melodies around the machine's output to create a "richer composition."

3. **Leading/Lagging:** The visual buffer allowed the performer to deliberately place themselves in the temporal flow, giving the impression of being the leader or the follower. Because the performer has advance information, they can anticipate the machine's imminent output and adjust their own timing accordingly, reversing the usual call-and-response order—even though the system itself continues to operate reactively throughout.

6. DISCUSSION

In this section, we reflect on the artistic and technical implications of the Somax2 modular extensions. Drawing from our collaborative development process and the public performances in Boston and Paris, we identify key themes regarding the role of musical literacy in human-computer co-creation and the shifting boundaries of performer agency.

6.1 Performance-Led Design and Reflective Practice

To evaluate the impact of these modular extensions, we adopted a performance-led research methodology, functioning as an iterative art-research collaboration between the system designer and the performer. Rather than evaluating the system through rigid, quantitative metrics, our

[11] Improtech @ Paris performance available at https://www.youtube.com/watch?v=4NNpGoayLnE

insights emerged directly from the rehearsal process and live performances. Following the methodology of reflective practice in interactive music systems [15], we report on the feedback gathered through unstructured interviews, remote prototyping sessions, and written correspondence between the two authors. Analyzing this qualitative data revealed several key themes regarding how the visual interface altered the performer's cognitive load and creative agency, transforming the software from a technical tool into a negotiated *boundary object* [16]—an artifact flexible enough to be shared and interpreted differently across the distinct roles (system designer and performer) involved in its use.

6.2 Literacy as Embodied Interface

The outcomes of these case studies suggest that musical *literacy* serves as a highly effective, yet largely underexplored, interface paradigm in human-computer co-creation. By employing standard chord symbols and traditional staff notation, our approach draws directly upon the expert improviser's procedural memory. For a highly trained jazz musician, reading a transposed $Eb7b9$ chord or a melodic fragment on a staff is not a purely intellectual exercise; it triggers an immediate, embodied physical response on the instrument. This methodological shift dramatically reduces the cognitive overhead required to decode the machine's internal state. By aligning the system's output with the performer's existing instrumental literacy, we bypass the need for new cognitive mappings, freeing the musician's cognitive resources for higher-level structural, dynamic, and aesthetic expression.

It is worth distinguishing *why* cognitive load is reduced in each of the two modules, since the mechanism is not identical. In the staff-reading module, the reduction follows directly from the look-ahead buffer: the performer is given literal advance notice of material that has not yet sounded, which is precisely the kind of information an improviser needs to plan toward a point of arrival rather than merely react moment-to-moment. In the chord-symbol module, by contrast, the harmonic information the label displays is, in principle, also available by ear. What the display changes is not the availability of the information but the *rate* at which it can be processed: a trained improviser can hear and name individual chords in real time, but cannot reliably parse, plan, and physically respond to a continuous, rapidly-modulating stream of them from the sounding audio alone. The visual label effectively performs the same look-ahead function as the staff-reading buffer, only for harmonic rather than melodic information, letting the performer register upcoming harmonic territory quickly enough to plan a phrase toward it. This distinction also explains why the two modules take such different forms: the chord-symbol shorthand relies on a shared tonal vocabulary that already carries functional meaning for a jazz-trained performer, whereas the staff-reading module (built on unlabeled MIDI pitch data, §4) carries no such built-in semantics and does not need to—there, the notation is read purely as spatial pitch-and-timing information, without requiring the performer to reach for a harmonic interpretation of it.

6.3 The Option Generator

This transparency fundamentally alters the psychological dynamic of the performance, redefining the boundaries of artistic agency. Throughout the collaborative process, the performer conceptualized the system not as an autonomous entity possessing independent intentionality, but rather as a highly potent *option generator*. Reflecting on the interaction, Okazaki observed: "Humans usually make the best decisions, but machines come up with more options".

It is worth clarifying what "curation" means in this context, since Somax2's players remain, at base, reactive: they generate output in response to the human's audio or symbolic input rather than initiating material autonomously. One might therefore ask whether the human is simply curating the machine, or whether the machine is simply following (or "mimicking") the human—in which case the visual display would only ever be curating the performer's *own* playing, not the machine's. We do not think the distinction is quite so clean-cut. Because the agent's navigation of the corpus is stochastic (§2.3), the same human input does not reliably reproduce the same machine output; "mimicry" in the sense of a literal echo is not what is occurring. What the reading modules add is not a way of overriding the machine's reactivity, but a way of increasing *entrainment*: the performer can see the shape of what is about to sound and can choose, moment to moment, to converge with it, diverge from it, or hold back—choices that were previously only available after the fact, once the sound had already arrived. In this sense, "curation" describes the performer's real-time shaping of their own contribution in light of the machine's visualized intent, rather than direct editorial control over the generative engine itself (which, in the duo configuration, remained Fiorini's domain, §4). The result is less a leader–follower switch than a tightening of the loop between the two agents.

In this framework, the Somax2 agent excels at rapid, combinatorial exploration. It provides a continuous stream of unexpected harmonic propositions and permutations of the *Introspection* corpus that a human improviser might not instinctively execute due to ingrained stylistic habits or physical muscle memory. Crucially, the visual interface transforms how the performer handles these propositions. The performer described the visually mediated connection as a "revelation," noting that it allowed for comping and soloing over the machine's output with a level of structural connectivity that would be impossible to achieve purely by ear.

Instead of trailing slightly behind the machine to react to its audio output, the human is empowered to proactively curate the generated material. By seeing the harmonic or melodic intent before it fully materializes, the performer can elect to accept the machine's proposition (aligning in unison), reject it (employing strategic silence), or systematically recontextualize it (crafting a contrasting contrapuntal line). This shifts the paradigm of human-AI improvisation from a reactive listening exercise to a proactive, real-time curatorial dialogue.

7. CONCLUSIONS

In this paper, we have presented a suite of novel interface modules for the Somax2 environment, designed to facilitate a "sight-reading" approach to human-AI improvisation. By externalizing the machine's harmonic intent through transposed chord labels and its melodic trajectory via look-ahead staff notation, we have demonstrated that targeted visual feedback can effectively mitigate the black box nature of reactive generative systems.

Through a performance-led research methodology and iterative collaborations culminating in live public performances, we observed a fundamental shift in the human-machine dynamic. By aligning the system's output with the performer's embodied musical literacy, these modular extensions dramatically reduced the cognitive load typically associated with interpreting abstract AI interfaces. Consequently, the performer was able to engage with the Somax2 agent not as an opaque entity, but as a transparent "option generator." This visibility empowered the musician to transition from purely reactive, ear-based trailing to a state of proactive curatorial agency, being able to analyze the agent's generation with structural precision.

Ultimately, the implementation of these tools highlights a versatile design philosophy: rather than modifying the core generative engine, we provide an interpretative layer that bridges computational logic and professional artistic expertise. While the current implementation focuses on specific metadata families and notation styles, the modular nature of the system establishes an open framework for future research into the diverse, embodied ways performers can read, curate, and guide co-creative musical agents.

Acknowledgments

This research was supported by the European Research Council (ERC) as part of the Raising Co-creativity in Cyber-Human Musicianship (REACH) Project directed by Gérard Assayag, under the European Union's Horizon 2020 research and innovation program (GA #883313).

8. REFERENCES

[1] L. Goupil, P. Saint-Germier, G. Rouvier, D. Schwarz, and C. Canonne, "Musical coordination in a large group without plans nor leaders," *Scientific Reports*, 2020.

[2] G. Assayag, L. Bonnasse-Gahot, and J. Borg, "Cocreative Interaction: Somax2 and the REACH Project," *Computer Music Journal*, vol. 46, no. 4, pp. 7–25, Dec. 2022.

[3] R. Rowe, *Machine Musicianship*. Cambridge, MA: MIT Press, 2001.

[4] D. Bailey, *Improvisation: Its Nature and Practice in Music*. New York: Da Capo Press, 1993.

[5] D. Schwarz, G. Beller, B. Verbrugghe, and S. Britton, "Real-Time Corpus-Based Concatenative Synthesis with CataRT," in *9th International Conference on Digital Audio Effects (DAFx)*, Montreal, Canada, Sep. 2006, pp. 279–282.

[6] G. Assayag, G. Bloch, M. Chemillier, A. Cont, and S. Dubnov, "Omax brothers: a dynamic topology of agents for improvization learning," in *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, 2006, pp. 125–132.

[7] J. Nika, M. Chemillier, and G. Assayag, "ImproteK: introducing scenarios into human-computer music improvisation," *ACM Computers in Entertainment*, Jan. 2017.

[8] J. Nika, K. Déguernel, A. Chemla, E. Vincent, G. Assayag *et al.*, "DYCI2 agents: merging the "free", "reactive", and "scenario-based" music generation paradigms," in *International Computer Music Conference*. Shangai, China, 2017.

[9] M. Fiorini and N. Brochec, "Guiding Co-Creative Musical Agents through Real-Time Flute Instrumental Playing Technique Recognition," in *Sound and Music Computing Conference (SMC 2024)*, Porto, Portugal, Jul. 2024.

[10] N. Brochec, M. Fiorini, M. Malt, and G. Assayag, "Interactive Music Co-Creation with an Instrumental Technique-Aware System: A Case Study with Flute and Somax2," in *International Computer Music Conference (ICMC 2025)*, Boston (MA), United States, Jun. 2025.

[11] A. Faraco, A. Schwarz, C. Vincent, P. Susini, E. Ponsot, and C. Canonne, "Listening Behaviors and Musical Coordination in Collective Free Improvisation," *Music & Science*, vol. 7, Jan. 2024.

[12] G. E. Lewis, "Too many notes: Computers, Complexity and Culture in Voyager," *Leonardo Music Journal*, vol. 10, pp. 33–39, 2000.

[13] A. Agostini and D. Ghisi, "A Max library for musical notation and computer-aided composition," *Computer Music Journal*, vol. 39, no. 2, pp. 11–27, 2015.

[14] A. Golvet, L. Goupil, P. Saint-Germier, B. Matuszewski, G. Assayag, J. Nika, and C. Canonne, "With, against, or without? Familiarity and copresence increase interactional dissensus and relational plasticity in freely improvising duos." *Psychology of Aesthetics, Creativity, and the Arts*, vol. 18, no. 2, p. 182, Sep. 2021.

[15] S. Nabi, P. Esling, G. Peeters, and F. Bevilacqua, "Embodied exploration of deep latent spaces in interactive dance-music performance," in *Proceedings of the 9th International Conference on Movement and Computing*, 2024, pp. 1–8.

[16] S. L. Star and J. R. Griesemer, "Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in berkeley's museum of vertebrate zoology, 1907-39," *Social Studies of Science*, vol. 19, no. 3, pp. 387–420, 1989.