# Distributed Agency in Collaborative Improvisation with Intelligent Instruments: A Phenomenological Inquiry

Halla Steinunn Stefánsdóttir
Intelligent Instruments Lab
University of Iceland
Reykjavík, Iceland
hallasteinunn@hi.is

Robert Ek
GEMM))) Gesture Embodiment and Machines in Music
Luleå University of Technology
Luleå, Sweden
robert.ek@ltu.se

Thor Magnusson
Intelligent Instruments Lab
University of Iceland
Reykjavík, Iceland
thormagnusson@hi.is

## Abstract

Recent developments in artificial intelligence (AI) have been increasingly driven by technoscientific and corporate approaches that emphasise large-scale datasets and autonomous generation systems. In response, Human-Centred AI proposes an alternative framework foregrounding human agency, values, and creative control. This paper contributes to this discourse by examining collaborative performance with AI-augmented instruments through a practice-led experiment involving two musicians. In this way, we investigate how musical agency is distributed across humans and multiple AI systems. Conducted through a laboratory process leading up to a live performance, this project went through five phases: 1) introduction of interfaces, 2) curation of datasets, 3) training of neural audio synthesis models and applying corpus-based synthesis techniques, 4) working with the intelligent instruments in rehearsals and performance, and 5) analysing the outcomes. Through combining qualitative, phenomenologically grounded methods and practice-led artistic exploration, we identify emergent creative relationships between performers and AI-augmented instruments. By analysing the agency at play, we unpack how creative control is distributed between human and machine. Situating our work within the framework of professional collaborative performance, we address the lack of phenomenological research into intelligent instruments whilst contributing methodologies for accountable, artist-centred AI development in musical contexts.

## Keywords

Artificial intelligence, Music, Practice, Agency, Curation

## 1 Introduction

AI development has recently had the signature of being driven by approaches that emphasise large-scale datasets and autonomous generation systems, prioritising engineering metrics and corporate values [6]. In response, Human-Centred AI (HCAI) proposes an alternative framework that foregrounds human agency, values, and creative control throughout both the design and deployment of AI systems [15]. This perspective aligns with recent calls within the field of music that, in order to move beyond technoscientific and corporate paradigms, we need to promote approaches that centre artistic practice, embodiment, cultural specificity, and ethical considerations in musical AI design [17, 31, 38]. This move is characterised by a distaste for corporate AI with large datasets in favour of user-designed agential AI with small datasets provided by the artist. Of importance for this present paper is that the HCAI principles invite critical examination of methodologies, of relevance also for the growing field of AI music studies [40].

We approach this by combining work on intelligent instruments with practice-led strategies from music improvisation and analytical phenomenological methods. Central to the methodology is a focus on the staging of design processes, in particular data curation, which gives musicians prototyping agency [38]. Through an extended instrumental encounter [23], musicians may interrogate and explore AI co-creativity, revealing how agency and sociality emerge in AI-augmented musical practice, dimensions that remain difficult or even impossible to access only through workshop-based or technoscientific methods.

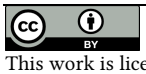
In this paper, we apply such methodology through a project instigated by two of the authors: Halla Steinunn Stefánsdóttir from the Intelligent Instruments Lab and Robert Ek from the GEMM Research Cluster. The project was designed as an experiment to explore what happens when two musicians collaborate, each having begun to develop performance practices with intelligent instruments. In this research, we ask: How is agency distributed across human performers and multiple AI systems when musicians collaborate using their individual intelligent instruments? How does this shape real-time interaction and musical outcomes?

By situating our experiment within the framework of collaborative improvisation, we open up the possibility of a phenomenological analysis that responds to the need to further address the lack of research into the phenomenological experience of intelligent instruments. This intersects with the call to further unpack what constitutes novelty in AI practices in the Arts [6] and conducting experiments that do not leave the agency of AI untroubled [43].

## 2 Background

### 2.1 Intelligent instruments, neural audio synthesis and corpus-based synthesis techniques

The aim of this project was to explore collaborative performance by two experienced instrumentalists working with AI-augmented instruments, combining three interfaces, differing datasets, and two




*NIME '26, June 23–26, 2026, London, UK*

generative AI approaches (neural audio synthesis and corpus-based synthesis) to examine phenomenological aspects and emergent creative relationships.

Whilst intelligent instruments build on a long tradition of interactive music technology systems, recent machine learning-based systems introduce a different mode of design through neural network approaches [17]. Rather than being explicitly programmed through top-down logic, these systems learn desired behaviours through training processes, representing a significant shift in redistributing the compositional agency anew between designer, composer, performer and instrument, opening up to novel forms of musical interplay [38].

Amongst the technologies applied in this experiment are RAVE (Realtime Audio Variational autoEncoder) and Somax2. RAVE has become a much-used approach in this domain, enabling the generation and transformation of audio through learned latent space representations [8]. The audio datasets get RAVE to function as a real-time sound engine, whereas through the training, the archive is left behind as residual waste, thanks to a "semantic depletion" [26]. The transformation affords navigation of multi-dimensional latent space, introducing elements of unpredictability, autonomy, and distributed agency into the creative process [28, 33, 38].

Another approach is presented in Somax2, an AI-driven multi-agent system designed for human–machine co-improvisation [3]. Instead of relying on deep neural networks, Somax2 employs corpus-based and probabilistic techniques to recombine segments of the analysed music within musical constraints, enabling it to generate coherent stylistic continuations. Somax2 operates as a memory-based system, generating new content by establishing a new route through learned material, either offline or live, using statistical sequence modelling [2, 11] as well as music information dynamics [48, 12].

Whilst large-scale datasets dominate much AI development, there is a growing focus on small dataset approaches in creative contexts [6]. Holistic in its stance as it avoids the computational requirements of large-scale training [6], it also represents a critical practice or that of resistance to "helicopter research" approaches, which colonise others [38] and acknowledging the embodied nature of data [30]. Whilst providing an entry into a deeper understanding of the workings of systems, the distributed agency at play and its ethics, it thereby affords a pathway into accountable AI practice through practice [38]. This intersection of neural audio synthesis, corpus-based techniques and small dataset methodologies creates fertile ground for investigating how performers develop relationships with intelligent instruments—a central concern of the present research.

### 2.2 Engaging in a phenomenological inquiry

Work focusing on phenomenological relationships in musical creativity has provided valuable insights into music research [10, 30, 39], revealing not only embodied perspectives but also the roles that technologies and environments play in phenomenological relationships.

An ongoing challenge in music research is the lack of a one-to-one relationship between music and verbal discourse [39]. Stimulated recall has emerged as a valuable methodological response: a recording-based approach that has gained significant momentum in music research over recent decades [5, 39]. This method of "doing phenomenology" serves dual purposes: it functions as an analytical tool whilst simultaneously enabling artistic creation, with intersubjective knowledge emerging through iterative engagement and phenomenological variation [35, 39]. Such approaches can deepen understanding of mediated processes, including their ethical dimensions.

In this experiment, audio and video recordings of testing/practice sessions and a live performance provided material for stimulated recall analysis. Think-aloud methods were also incorporated during testing/practice sessions through commentary during pauses in activity. The material was accompanied by qualitative coding [29] using MAXQDA software. This was done by the first and second author, first jointly on parts of it, then individually. The coding was done from a conceptual stance, focusing on the human-technology relations, accompanied by analytical memos. This approach facilitated examination of the documentation from shifting viewpoints, moving between first-, second-, and third-person perspectives. These findings were later subject to further thematic analysis and refinement.

### 2.3 Distributed agency

We align with Suchman's theorising, which posits agency as the result of "situated actions" [41, 42] through which creatives seek to develop their practice, distributed through culture, environment, and technology. Agency is therefore not something that exists only in the object itself or in the musician; it is a relational phenomenon that emerges through interaction. This necessitates paying attention to the material particularities and unruly contingencies in encounters [42].

To further unpack the relations that form, we apply post-phenomenological theorising [20, 46], which examines the role of technological mediation in shaping our relations to the world.

Building on the precedent perspectives, we argue that co-creativity (including co-improvisation) is not reserved for humans. Rather, creativity unfolds within an ecosystem in which humans, nonhuman entities, and technologies mutually constitute one another, forming a posthuman sociality [22, 49]. Emphasising sociality also challenges technoscientific agendas that downplay relational engagement in favour of technological novelty [42].

## 3 Collaborative experimentation with intelligent instruments

This section will present and analyse the work on the intelligent violin and intelligent clarinet from design to live performance.[1] Through this, we apply models that necessitate new interfaces. Meanwhile, it builds on machine learning that is "blackboxed", amplifying the musician's "tinkering" or "probing" role [44]. Such experimentation is guided by the curatorial, an agency which is reinvented through each situation, guided by contextual sensitivity as it negotiates material as well as immaterial formations [35].

[1] For excerpts from the live performance see https://doi.org/10.5281/zenodo.19865482.

The outcomes are framed by us as "cultural probes" [16, 44], which yield an effect within the wider ecosystem of music making [49]. This then enables assessing its effect. In what follows, we provide an episode-based account in sections 3.1–3.5, followed by section 3.6, which consolidates the analysis into a more explicit thematic synthesis.

### 3.1 Design of the experiment

The work proceeded through five phases: first, an initial encounter that led to a second phase of curating audio datasets. This was followed by training of new models in RAVE and constructing Somax2 corpora. In the fourth phase, the models were applied in performance by using a baroque violin and a sensor-augmented clarinet, engaging with a Max/MSP patch GLARE, customised Somax2 and a newly invented neural audio synthesis looper. This was done both in laboratory settings and for an audience at a concert. Finally, the complete process was documented from start to finish, and the fifth phase analysed the process and the findings that emerged.

### 3.2 Phase 1—Introducing prior instrument design

The project began with an initial encounter, which evolved around introducing systems we had been working with, both through talking, showcasing and playing.

The baroque violin was connected to the Living Looper, a neural audio synthesis looper created by Victor Shepardson at the Intelligent Instruments Lab [32]. This was done by employing a clip-on condenser microphone and a MIDI foot controller to record loops and react to their playback. The loops can be four in total and are represented both through audio and visual representation. The signal from the violin was sent to the system (as seen in Figure 1), but also straight to the speakers. By playing into the looper, sounds are transformed over time by mixing recorded loops through a statistical machine learning prediction algorithm [32, 33]. The looper affords a robust generative form of agency in performance, giving musicians only a certain amount of control and/or predictability, and expanding the experience of the looper beyond instrument and co-creator, towards that of a composition [38].

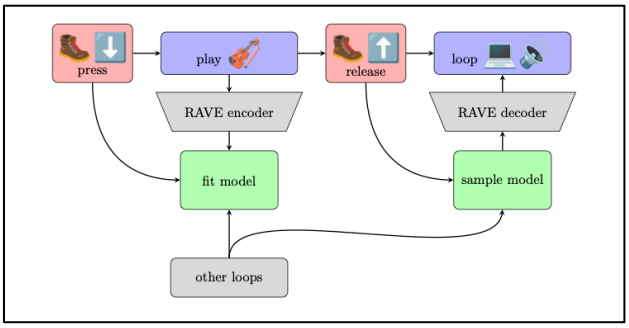



**Figure 1: The workflow of creating a living looper.**

Ek worked with his sensor-augmented clarinet [13], which enables access to motion qualities in musical performance. A combination of audio from the clarinet, captured by a pickup microphone, and movement data creates a multimodal system that enables navigation of a model's latent space using both sound and gesture. For the gestural component, Ek employed a newly developed machine learning tool that combines Laban Movement Analysis of Efforts [21, 4] along with motion-sensor data and machine learning to identify a musician's inner intentions during music performance. Consistent with Laban's Effort theory, the model's representation of Efforts continually fluctuates. Unlike the linear notation used by Certified Movement Analysts, the model captures the dynamism or "aliveness" of human motion, including minute nuances in Effort variations that human analysts would disregard.

Ek introduced the Somax2 environment (as seen in Figure 2), which produces stylistically consistent streams by continuously listening and adapting to an audio or MIDI input, as well as its own multi-agent system [3]. It analyses existing musical material, either audio or MIDI, to build an internal representation of style, structure, and musical gestures, and then generates new music in real-time that follows and responds to that material. Early tests revealed a strong bias towards favouring certain materials within the audio corpus. To counteract this, Ek had designed the system to use two "virtual players" with the same corpus, further detailed in section 3.5.3. Additionally, Ek's sensor-augmented clarinet enabled gestural modulation of Somax2 parameters.

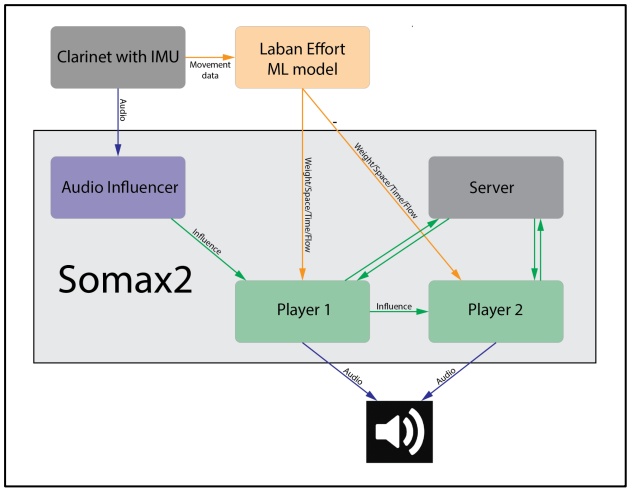



**Figure 2: The Somax2 setup employed, also showing the Laban Effort machine learning model driven by gestural sensor data.**

Another approach used the Max patch GLARE (as seen in Figure 3) developed by Ek, enabling gestural exploration of the latent space within a RAVE model. A combination of gestural control and automation manipulates the model's latent representation, moving away from direct timbre transfer between input and output. Blackboxing the latent space creates an artistically interesting system that enables creative resistance. Tuning the various aspects of the manipulations is crucial, finding a balance between control and unpredictability.

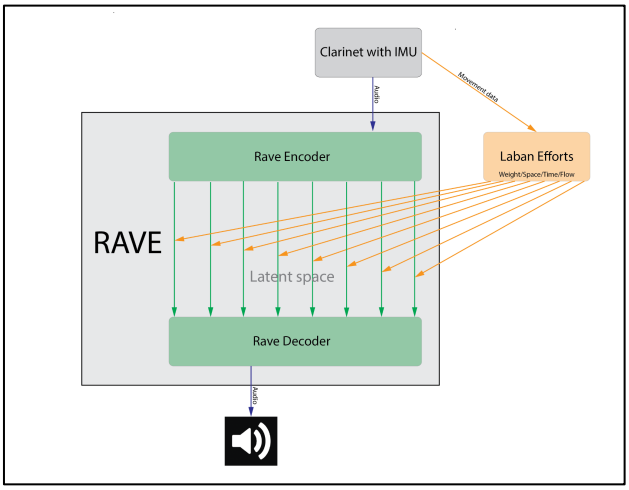



**Figure 3: The workflow of improvising with the Max patch GLARE.**

The difference in the design intentions, as manifested through the underlying architecture, became immediately clear. The Living Looper generates new audio by looping through latent spaces, thus morphing the sound over time. Somax2, on the other hand, affords closer connections to the temporal structures and phrase shapes present in the training material. In this way, we could feel the aesthetics of the original dataset, even when working with datasets by other colleagues.

From here, we entered a more structured approach, relating collaboratively to data curation as a form of instrument prototyping.

### 3.3 Phase 2—Curation of datasets

Corpus-based techniques and neural audio synthesis models are typically trained on audio datasets ranging from small, curated collections to large-scale corpora. We agreed that in this initial experiment, we should work through a dataset selection approach that contextualised AI practices in individual artistic intentions and sonic preferences. We decided on the following datasets:

1. Halla Steinunn Stefánsdóttir's baroque violin music stems, a discrete part of the tracks, from the album project *strengur*.
2. Robert Ek's clarinet music stems from a selection of audio files from live improvisations done by Ek.
3. Guitar improvisations by Victor Shepardson.
4. Halla Steinunn Stefánsdóttir's baroque violin music stems, a discrete part of tracks, from the recording of *Violin with Þytur*, a solo work by Kent Olofsson, created for Stefánsdóttir's nine mono-channel installation *Þytur*.
5. Robert Ek's clarinet music stems, consisting of the clarinet part from an improvised performance with the TCP/IP quartet.

All datasets exemplify a curation of generative training where models are trained on a personal body of work. The only exception is model number three, which was extended to work with guitar improvisations by Victor Shepardson involved in this project through the training of models. Stefánsdóttir's prior experience with the model had revealed that it was highly interesting in terms of response thanks to its metallic and shimmering sound, which contrasted with her violin model and playing.

The editing necessitated adding silences between edits to teach the model to recognise such transitions (for further reading on editing, see Stefánsdóttir and Magnússon, [2025]).

### 3.4 Phase 3—Training of neural audio synthesis models and applying corpus-based synthesis techniques

The first three datasets were trained on neural audio synthesis in RAVE. Datasets four and five were trained using the corpus builder inside Somax2.

RAVE's open-source training software was already installed on the Intelligent Instrument Lab's server; configurations were made using a text editor, and then the data preprocessing and training were run via Python scripts. The progress could be monitored in the TensorBoard application, which displays diagnostic graphs and audio samples from the model. In order to arrive at a high-quality model, they were trained on average for 5 days.

Somax2 uses a dual-layer learning approach: a general latent space trained on four centuries of Western, copyright-free MIDI [3], and musician's curated training that shapes each multi-agent "player"'s stylistic identity. Consequently, each player balances three elements: the overarching general knowledge focused on texture, navigating a musical memory subspace, and the immediate context by listening and responding [3]. Due to this fundamental design and the training material used for the overarching general knowledge, Somax2 has a clear bias towards the Western classical tradition, a fact that became very clear to us during the probing of the system.

### 3.5 Phase 4—Iterative development

In what follows, we will detail the analysis of our performance [37], focusing on our concert. We analyse the effect of each set, focusing on the configurations and reconfigurations between us and our systems.

**Figure 4: A still from the documentation video of the performance in Reykjavík.**

#### *3.5.1 The violin and clarinet models*

In the first set, Stefánsdóttir applied her violin RAVE model (no. 1) in the Living Looper, and Ek applied his sensor-augmented clarinet to the clarinet RAVE model (no. 2) in the GLARE Max patch.

We planned that each performer would begin one set with their system, starting with Stefánsdóttir. Her solo used a harmonics-based opening intended to form an interesting first loop via foot control.

The looper's architectural constraints were immediately felt, shaping the set's emergent form. Stefánsdóttir enacted only a few looper alterations, each producing significant effects, from looper-dominant to variation thereof, to transparent, to active, and back to transparency as one looper was deactivated.

Only late in the set did both systems receive joint focus, with Ek employing gestural control and Stefánsdóttir playing sporadic repetitive comments. The gestural tweaking of the latent space in GLARE proved too subtle to be consistently noticeable. Since the RAVE model was trained on Ek's own material and, due to RAVE's tendency to match the incoming signal, its output was often difficult to distinguish from the clarinet. We note moments of system-focused playing that drew down our individual expression: Stefánsdóttir sustained a repetitive looper aesthetic to make space for Ek's solo, while Ek foregrounded model output and used the clarinet mainly as a "feeder."

For the performance, we routed the systems to separate speaker pairs, one on stage and one at the back, helping the audience distinguish them aurally and spatially. This setup revealed shifting co-improvisatory relationships in the performers' choreography: Stefánsdóttir sometimes turned away from Ek toward the looper speaker, while at other moments they faced each other in a violin-clarinet duo that sidelined both systems. Video evidence reveals instances where Ek's interaction with the model shifted his gesturality from habitual body schema to probing manifested in his body language.

**Figure 5: A still from the documentation video of the performance in Reykjavík, revealing an altered choreography.**

#### 3.5.2 *The guitar models*

During rehearsals, combining our systems failed once. Ek applied the guitar corpus in Somax2 while Stefánsdóttir worked with the guitar RAVE model in the Living Looper, and their outputs created a caricature of post-modernist collage aesthetics that made joint improvisation impossible. Fixing this would have meant re-recording the Somax2 guitar dataset and further parameter prototyping, which was beyond the project. Ek instead moved the guitar model to his GLARE Max patch so we could improvise together.

In the concert, Ek opened with a solo, translating clarinet playing into gestural control. Steered partly through the keys, while those same keys still produced sound, the clarinet hovered between instrument and controller, foregrounding the RAVE output. This contrasted with the previous looper-led set and drew Stefánsdóttir towards non-pitched techniques, delaying the looper. When it entered, she used a single loop intended not to dominate, yet it still took over the improvisation, and Ek began to imitate it.

Gradually, we achieved equilibrium amongst the human improvisers and the intelligent systems. Stefánsdóttir at times paired with her system as a duo, with Robert joining; later, Robert's gesture-steered playing prompted what she described as creature-like outbursts. The set included an energy and dynamics climax between Ek and Stefánsdóttir that we later judged unnecessary, a return to familiar improvisational habits that pulled us away from making music "with" the systems.

#### 3.5.3 *The concert archive models*

In this set, we worked with Somax2 and exchanged audio corpora drawn from earlier live performances. We used a two-virtual-player setup: only Somax_player_1 received the live instrument signal, while Somax_player_2 listened to Somax_player_1, generating variation in selection and output.

Working with this AI-augmented system produced seemingly random fragments of one's past performance. In rehearsal, we recognised selection patterns, making the system somewhat predictable. Unlike RAVE's abstraction, Somax2 stays tied to recorded material, like a doppelgänger that fragments one's playing. Stefánsdóttir could even recall the spatial and atmospheric context of specific fragments from her solo within a 9-channel installation. With both musicians working through Somax2 and the systems, continually responding to input, creative agency was balanced amongst the four co-creators.

The aforementioned recognition may have obscured some of Somax_player_2's abstraction, only noticed later in stimulated recall. Through a "citational" effect, Stefánsdóttir could respond to, or even anticipate, her own fragments by playing variations earlier in the set. She also responded at times to Somax_player_2's unfamiliar material. With Ek's synthesised corpus, she wanted more time to familiarise herself with it and engage in mapping as a way of shifting its response. She did, however, learn that shifting the system required repetition, which strongly shaped her playing. In concert, this strategy failed because Ek could dominate the input stream through the microphone setup, suggesting the need for alternative miking, such as a clip-on piezo pickup.

### 3.6 Analysis

Working with machine learning as part of musical instruments, dataset curation grants musicians significant prototyping agency [38]. As a result, this collaboration required sharing insights into prior prototyping and performance through discussion, showcasing and improvisation. By familiarising ourselves with our new technological and human collaborators, we began to understand the embedded design values, how our intelligent instruments communicate with their human co-creator and each other and what expressivity and communicative possibilities it gives rise to. This also marked the start of familiarisation between Stefánsdóttir and Ek in terms of improvisation.

This open-ended curatorial sharing enabled our co-curation, a form of planning and projection, which is essentially a situated practice [42], where we bring our sociomaterially shaped skills and forge new ones as we respond to the situation at hand. Both musicians have carved out a practice of working with small, personalised datasets, respecting their origins and recognising that curatorial care and ideation will be felt in the model's output [38]. From an artistic

research laboratory perspective, we viewed this project as the start of longitudinal collaboration. At the same time, we needed to get to know each other and our systems. This had the effect that we opted for an initial curation that minimised unleashing too many agencies into our work: a strategy based on starting from familiar sociomaterial situations. In this way, our collaboration was in some way less experimental than our solo undertaking (see, for example, Stefánsdóttir and Magnússon [2025]). The findings were then a departure point for future formulation of the staging of design, including data gathering processes.

RAVE models and corpus-based synthesis represent fundamentally different approaches, although both involve "statistical flattening of the source material" [27]. This was immediately felt by performers: the RAVE approach provided a level of abstraction, whilst the latter preserved original material. This provided an entry into widely different practices, particularly regarding memory. Stefánsdóttir could recall her physical location and atmosphere during live performances whose fragments appeared in Ek's Somax2 performance, connecting these to the original score. Overall, such an effect brought forth traces of prior performances, spaces, and absent collaborators, introducing distinct spatial and temporal dimensions.

To further examine performance dynamics, we draw on Malafouris and Koukouti's [2022] account of how a potter's engagement with clay can shift from a "conscious" to an "immersive" mode. For intelligent instruments, the conscious mode involves analysing mediations and emergent agencies, followed by further testing. An example is when Living Looper and Somax2 proved incompatible: Somax2 operates as a citational system of recordings, while the looper functions as a composition, producing an aesthetic clash. We therefore abandoned combining them, and Stefánsdóttir instead worked to incorporate Somax2 into the collaboration. By the project's end, Stefánsdóttir was beginning to familiarise herself with the system, although this was made difficult in concert due to the clarinet signal disturbing the relations between her violin and the system.

If we refer this to Ihde's [2009] post-phenomenological theorising about how technologies make things "speak" to us differently, these systems produce different phenomenological variations. Stefánsdóttir has previously observed how the RAVE model in Living Looper makes her experience of the materials of her baroque violin, namely horsehair, wood and gut strings, appear in an altered way [38]. However, Somax2, given our material curation, creates phenomenological variations of specific events. This effect diminishes when working with another performer's recordings, shifting towards phenomenological variation of clarinet playing tied to certain aesthetics. Crucially, both approaches maintain ties to original embodied practices. For example, when playing with the guitar model, Shepardson's playing and its aesthetics remain recognisable. This brings to mind how datasets need to be seen as representing a dual concept of "embodied data", resulting in musicians "embodying data" [31].

Stefánsdóttir has earlier analysed work with intelligent systems by looking towards Verbeek's expansion of Ihde's hermeneutics into "composite intentionality," which entails a double intentionality, or "one of technology towards its' world, and one of human beings towards the result of this technological intentionality" [46]. This entails that during performance, two forms of human-technology relations are unfolding simultaneously:

performer → (violin-mic-speaker → world)

↑↓

performer → (violin-mic-foot–controller-interface-model-speaker → world)

Through the conjoining of practices, the composite intentionality relations became "coupled," leading to varying interplay across the ensemble. Such a shift in directedness can bring forth a different reality, in line with how technologies afford us to experience phenomena anew [46], as seen in Stefánsdóttir's attunement to Ek's Somax2 as it reworked and cited a recording of her past experiences. This coupling was also affected by breakdown, as when Ek took over Stefánsdóttir's Somax2 in response to limitations in her microphone setup.

To further unpack these relations, we may look at the differences between the systems. Whereas the looper and Somax2 can continue autonomously, the GLARE Max patch cannot continue without input. Furthermore, the looper's architectural constraints reconfigured the composite intentionality, meaning that the violinist had to respond and act as a "regulator," negotiating, redirecting, and sometimes halting the system's momentum to create space for herself and other collaborators. This can be understood as a form of "prototyping-through-performance," where altering parameters and permissions functions as ongoing curation of what kinds of actions the system can take and when. By contrast, the clarinet processing/Max layer afforded activation through clarinet gestures. Here, gesture operated as a form of "delegation," resulting in a redistribution of agency. Rather than producing musical direction primarily through conventional technique, the performer used embodied action to "grant" the system audibility, not merely via system autonomy but through selectively activating system responses.

The mechanisms of regulation, delegation and parameter curation exemplify how technology's structure affords "multistable variation" [19] or different trajectories. This is further evident in Stefánsdóttir's interplay with the foot controller, wherein it functioned variously as an adaptation tool or stop button curtailing the looper's constraints. Another multistable trajectory is located in the separation of speakers, a devising strategy staging them as co-creators. Here, Stefánsdóttir's altered choreography, turning towards her system, was at times spontaneous, at times intentionally signalling to the audience that relations differed from electroacoustic playback practices. Similarly, the clarinet hovered between the hybridity of being an instrument and a gestural controller.

Through this, we can talk about an aesthetic control being introduced through the distributed design of the system. Similarly, aesthetics are being constructed through continued prototyping and performance processes, in ways that are conjoined with the ethics of such work. For example, further tweaking or curating a system's parameters, intended in our case to reduce Somax2's predictability, and its trademark phrase "clipping," is a way to address situations in which it was difficult to play. Here, we may grasp that this work essentially revolves around a willingness to give up control and forge new relations, while iteratively exploring what instrumentality we are seeking.

## 4 Discussion

The aim of this paper is to better understand how agency is distributed across human performers and AI systems through interface design and engagement thereof, and how this shapes real-time interaction and musical outcomes.

The experiment seeks to counteract the common situation where musicians are relegated to the role of technology users or composers' assistants [35, 38]. Rather, musicians gain prototyping agency through data curation [38], in this present experiment as a joint endeavour. To arrive at that, the musicians, both of whom have begun carving out a performance practice with AI-augmented instruments, need to share their prior work. Such facilitation, unfolding through talking, showcasing, and improvising together, is essential to be able to engage in plans and projections of what becomes essentially a joint composition as a mode of instrument design.

Through this initial probing, the musicians start to grasp the new interfaces' functions and their differences. For example, the systems may present significantly different scales of co-evolvement, with one being conjoined with their interface to get it to participate, whereas the other can stop and listen to the system interpreting the sounds of a model with living loops bringing them to life and altering them. An effect which enables the instrument and co-creator to take on the form of "composition", a phenomenon that becomes problematic in this collaboration because it entails a reconfiguration of other co-creators. It draws down on the ambiguity valued by many improvisers and the possibilities of "co-tuning" in improvisatory practices with machines [34].

The different synthesis techniques are also a strong factor when it comes to the co-creative and instrumental agency of the system, and can result in aesthetics that are incompatible. Here, the musicians' artistic integrity made it impossible to work with them even in rehearsal. Through this process and continued iterations, we attune to each other, perceive our systems anew, and embed shared musical vocabularies and ethical sensibilities into the prototyping process. Some of its effects may be hard to perceive as they unfold in real time, and it is here that stimulated recall becomes crucial to further grasp the relations that form, essential to formulating strategies for future steps.

In this way, the extended encounter takes on the form of what Verbeek [2008b] describes as an "action-ethical" approach, intended to assess and focus on the quality of "practices" and their implications. Here we are dealing with a different kind of music making, where the materiality is partly blackboxed and producing an alternative machinic intelligence. It invites us to enter a different engagement, which bears the trace of attentive listening in environments similar to echolocation. Such listening attention has been theorised by Barry Truax [1984] as "listening-through-search." It is a shift in consciousness, one that evolves around participation and learning, a trait that shares an affinity with the practices of many environmental musicians [18].

Here, in the context of an extended encounter with a technological "other", we enter a hybridity where we embody curation, prototyping and performance as a form of composition, creating a gateway to interplay identified by learning, search, and negotiation. It explains our willingness to reconfigure our instruments and thereby our habitual engagement thereof, which may take on the form of a design that affords playing with the clarinet as a gestural control. It may invite a musician to stop her system, in order to try and create a new scenario that may be more dynamic for her human co-player and their system. But she may also turn her back to them both and engage in repetition, aimed at affecting the blackboxed function of her own system, backgrounding the other relations. Therefore, at times we follow system initiative, at times we intervene, and at times we step back to make space for other agencies.

This brings to mind how prototyping and making with intelligent instruments floats each time a new digital music-machine-ship[2] [38], inviting us to actively share autonomy with machine systems as we search for "different" outcomes [1]. Human-human improvisation can include practices that take the form described by Rowe as "non-listening" [9] that, as Frisk argues, can expand rather than diminish ethical capacity [2020]. With intelligent instruments, the analogy may sound nonsensical unless we specify what is being bracketed and how. This means that it does not revolve around "rudeness" but is rather grounded in design or interplay, which suspends responsiveness. For example, a looper can mimic the basic form of non-listening thanks to its design constraints that can work as superimposing composition for human co-creators. But the looper can also offer a structured form of "non-listening" that holds material steady so other relations can be foregrounded. Another example is how systems that remain continually responsive (e.g., Somax2) can appear "non-listening" in a different sense: they may persist in citational tendencies that do not register our shifting intentions, prompting us to intervene through mapping and parameter tuning. This aligns with other artistic research findings where mismatches between aesthetic expectation and system behaviour are negotiated through iterative adjustments, thereby pushing at pre-given notions of what AI is or may become [26].

From this, we may see how intelligent instruments can be framed as "cultural probes" [16, 44] as they decenter both the human subject, instrumental agency and the resulting intersubjective relations. In this way, it can open up to new practices, such as experienced with the Somax2 system, due to the materiality of the original dataset, which turned the performance into a virtual time travel and the performers even into "völva" (Icelandic for seeress) as she predicted her own performance of the past. This last point, however, cannot be separated from the fact that Somax2 also proved quite repetitious to the point of being predictable.

Bryan-Kinns et al. in their review of the arts in relation to HCAI observe that "the use of AI in the Arts raises questions not only about how creative outputs should be protected, but also what constitutes novelty in creative practice and production," and that (m)ethods and metrics for evaluating HCAI for the Arts likely also require a fresh perspective" [2025]. The combination of practice-led strategies from

[2] This is a modulation of Andersen et al. [2019] "digital craft-machine-ship," which is a concept similar to assemblage or apparatus. This term challenges the anthropocentric view of machines and suggests "a new kind of digital craftsmanship, one in which we may craft with the digital and find ways to make the machines craft along with us."

music improvisation and analytical phenomenological methods, including stimulated recall, holds great promise. Not only does it afford new research insights, but it is also essential to continued artistic iterations, as we search for the hows and whys of creative engagement with AI.

## 5 Conclusion

This paper introduced new AI-augmented instrument designs and explored the possibilities for co-creativity through an experiment involving two experienced musicians who have begun to carve out a performance practice with intelligent instruments. We detailed its five stages of design: introduction of interfaces, dataset curation, model training and application of corpus-based synthesis techniques, performance, and analysis. Our research asked how agency is distributed across human performers and multiple AI systems when musicians collaborate using individual intelligent instruments, and how this shapes real-time interaction and musical outcomes.

The introductory phase connects to the musician's prior experiences and preferences with datasets and interfaces. At the same time, the systems introduce architectural agency, including latent-space structures, corpus segmentation and stylistic priors, which come to affect rehearsals and live performances. We have control in dataset curation, however through editing and following training, the data is turned into a material agency which shapes and even dictates our co-creative response. The model making shifts the design towards specifying conditions for learning, conditions that are opaque or blackboxed. Here, machine learning produces a distinct model agency shaped by representational logic, rather than merely our intentions. As the models exit training, we embark on a performance as a phase of negotiation, sharing control in co-creative improvisation. Here, the models enact their own temporalities as seen in this experiment through RAVE's ability to transform and drift within learned latent constraints and Somax2's citational tendencies. Furthermore, this distributed agency bears the signature of learning and search, as altered listening relations become a site of real-time negotiation in which we resist, follow, and actively make space for co-creators, which may even lead to prototyping-through-performance.

In sum, agencies are continuously surfacing and disappearing. In this way, the notion of distributed agency might lead us to understand them as a singular phenomenon, but instead we are face to face with a network of agencies that are essentially dispersed and diffracted, forming through intentional design choices (selection, curation, staging), architectural and learned constraints (what the machine learning system can represent and how it behaves), and through situated musical improvisation, where we may follow the system's initiative but also intervene on the basis of ethical and aesthetic concerns. Performers therefore, do not simply "cede control" to intelligent instruments, for although they can offer rewarding co-creative moments, they also prompt regulation and, at times, withdrawal of system agency when interaction conflicts with musical aims. We also observed limits to co-improvisation, including breakdowns in performability when citational behaviour (Somax2) became predictable or when systems created nonsensical aesthetics. Enabling co-tuning, as a form of negotiable agency, requires ongoing work around the models and interfaces, and, when needed, a return to the archive through renewed curation and gathering.

The relations that arise then give way to active reconfigurations on-the-go, or provide insights that we take with us into the next phase of experimentation. Our next steps focus on gathering processes and interface curation, including parameter curation, as ways of probing intelligent instruments and unearthing the sociality they may offer. More broadly, our approach, combining qualitative, phenomenologically grounded methods with practice-led artistic exploration, can be extended to cross-disciplinary collaborations around AI-augmented instruments, with HCAI principles at its core. Work that does not leave AI untroubled, and does so by treating it as something we can stage, tune, and contest through design and performance.

## Ethical Standards

This paper complies with the ethical standards of the NIME conference [25] and does not present any conflict of interest.

## Acknowledgments

The Intelligent Instruments Lab is supported by the European Research Council (ERC) as part of the Intelligent Instruments project (INTENT), under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101001848). The authors would like to thank the anonymous reviewers and the meta-reviewer for their critical reading of this manuscript.

## References

[1] Kristina Andersen, Ron Wakkary, Laura Devendorf, and Alex McLean. 2019. Digital crafts-machine-ship: creative collaborations with machines. *interactions* 27, 1 (January - February 2020), 30–35. https://doi.org/10.1145/3373644

[2] Gérard Assayag and Shlomo Dubnov. 2004. Using Factor Oracles for Machine Improvisation. *Soft Computing* 8, 9 (August 2004), 604-610. https://doi.org/10.1007/s00500-004-0385-4

[3] Gérard Assayag, Laurent Bonnasse-Gahot, and Joakim Borg. 2022. Cocreative Interaction: Somax2 and the REACH Project. *Computer Music Journal*, 46, 4 (December 2022), 7–25. https://doi.org/10.1162/comj_a_00662

[4] Irmgard Bartenieff and Dori Lewis. 1980. *Body Movement: Coping with the Environment.* Langhorne: Gordon & Breach Science Publishers. (September 1980).

[5] Benjamin S. Bloom. 1953. Thought-processes in lectures and discussions. *J. Gen. Educ. 7*, 160–169.

[6] Nick Bryan-Kinns, Anna Wszeborowska, Olga Sutskova, et al. 2025. Leveraging small datasets for ethical and responsible AI music making. In *Proceedings of the 20th International Audio Mostly Conference (AM '25)*, Association for Computing Machinery, New York, NY, USA, 70–81. https://doi.org/10.1145/3771594.3771601

[7] Nick Bryan-Kinns, Phoenix Perry, Elizabeth Wilson et al. 2025. Human-Centred AI in Arts. In *Handbook of Human-Centred Intelligence.* Wei Xu (Ed.), Springer, 1-32.

[8] Antoine Caillon and Philippe Esling. 2021. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv:2111.05011 [cs, eess]* (November 2021). http://arxiv.org/abs/2111.05011 arXiv: 2111.05011

[9] Marcel Cobussen and Nanette Nielsen. (2016). *Music and Ethics*. Routledge. https://doi.org/10.4324/9781315596594

[10] Jonathan De Souza. 2017. *Music at hand: Instruments, bodies, and cognition*. Oxford: Oxford University Press.

[11] Ken Déguernel, Emmanuel Vincent, and Gérard Assayag. 2018. Probabilistic Factor Oracles for Multidimensional Machine Improvisation. *Computer Music Journal* 42, 2 (June 2018), 52–66. https://doi.org/10.1162/comj_a_00460

[12] Shlomo Dubnov. 2021. Machine Improvisation in Music: Information-Theoretical Approach. *In E. R. Miranda, ed. Handbook of Artificial Intelligence for Music*. Berlin: Springer.

[13] Robert Ek. 2024. Playing with resistance. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME' 24), Sept 2-6, 2024, Utrecht, NL. 154–159. https://doi.org/10.5281/zenodo.13904814

[14] Henrik Frisk. 2020. Aesthetics, interaction and machine improvisation. *Organised Sound 25* (March 2020), 33–40. doi: 10.1017/S135577181900044X

[15] Ozlem O. Garibay, Brent Winslow, Salvatore Andolina, et al. 2023. Six Human-Centered Artificial Intelligence Grand Challenges. *International Journal of Human–Computer Interaction*, 39 (January 2023), 391–437. https://doi.org/10.1080/10447318.2022.2153320

[16] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: Cultural probes. *Interactions* 6 (January 1999), 21–29. https://doi.org/10.1145/291224.291235

[17] Artemi-Maria Gioti, Aaron Einbond, and Georgina Born. 2022. Composing the assemblage: probing aesthetic and technical dimensions of artistic creation with machine learning. C*omput. Music. J.* 46 (December 2022), 62–80. http://doi.org/10.1162/comj_a_00658

[18] Bennett Hogg. 2013. The violin, the river, and me: Artistic research and environmental epistemology in Balancing String and Devil's Water 1, two recent environmental sound projects. Hz Journal, 18. https://www.hz-journal.org/n18/hogg.html

[19] Don Ihde. 2007. Technologies — musics — embodiments. *Janus Head*. 10 (Summer/Fall 2007), 7–24. https://doi.org/10.5840/jh20071012

[20] Don Ihde. 2009. *Postphenomenology and Technoscience*. The Peking University Lectures, State University of New York Press, New York, NY.

[21] Rudolf v. Laban. 1963. *Modern Educational Dance*. United Kingdom: MacDonald and Evans.

[22] Bruno Latour. 2005. *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.

[23] Thor Magnusson. 2026. The Extended Encounter: Kicking Against Methodology. *In Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME' 26), June 23-26, 2026, London, UK.

[24] Lambros Malafouris, and Maria-Danae Koukouti. 2022. Where the touching is touched: the role of haptic attentive unity in the dialogue between maker and material. *Multimodality Society* 2, 3 (June 2022), 265–287. https://doi.org/10.1177/26349795221109231

[25] Fabio Morreale, Nicolas Gold, Cécile Chevalier, and Raul Masu. 2023. NIME Principles & Code of Practice on Ethical Research.

[26] Teresa Pelinski, Adam Pultz Melbye, and Andrew McPherson. 2026. When AI don't sound like AI: Negotiating aesthetic expectations in technology-mediated musical practice. *Organised Sound* (April 2026) , 1–11. https://doi.org/10.1017/S1355771826101228

[27] Nicola Privato and Thor Magnusson. 2024. Querying the ghost: AI Hauntography in NIME. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME' 24), Sept 2-6, 2024, Utrecht, NL. https://doi.org/10.5281/zenodo.13904901

[28] Nicola Privato, Victor Shepardson, Giacomo Lepri, and Thor Magnusson. 2024. Stacco: exploring the embodied perception of latent representations in neural synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME' 24), Sept 2-6, 2024, Utrecht, NL. https://nime.org/proceedings/2024/nime2024_62.pdf

[29] Johnny Saldaña. 2013. *The coding manual for qualitative researchers* (2nd ed.). Sage Publications.

[30] Pierre Schaeffer. 2017/1966. *Treatise on musical objects* (trans. North, C., and Dack, J.). University of California Press, Oakland.

[31] Franziska Schroeder and Federico Reuben. 2025. Breathing Life into AI Querying AI through Tactile Imprints Using Breath and Sound. In *Artificial Intelligence in the Cultural and Creative Sectors*, Marta Massi, Marek Prokupek, Maria C. Ostillio (Eds.), Routledge.

[32] Victor Shepardson and Thor Magnusson. 2023. The Living Looper: Rethinking the Musical Loop as a Machine Action-Perception Loop. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME' 23), May 31-June 3, 2023, Mexico City, Mexico, https://doi.org/10.5281/zenodo.11189164

[33] Victor Shepardson, Halla S. Stefánsdóttir, and Thor Magnusson. 2025. Evolving the living looper: artistic research, online learning and tentacle pendula. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME' 25), June 24-27, 2025, Canberra, Australia, 255–260. https://doi.org/10.5281/zenodo.15698851

[34] Paul Stapleton and Tom Davis. 2021. *Ambiguous Devices*: Improvisation, agency, touch and feedthrough in distributed music performance, *Organised Sound*, 26, 1 (August 2021), 52-46. doi:10.1017/S1355771821000054

[35] Halla Steinunn Stefánsdóttir. 2023. *HÉR! An Exploration of Artistic Agency*. Ph.D. Dissertation. Lund University, Lund, Sweden. https://portal.research.lu.se/en/publications/hér-an-exploration-of-artistic-agency

[36] Halla S. Stefánsdóttir, and Davíð B. Franzson, D. B. 2025. Navigating an urban archive: an exploration of wave field synthesis as soundscape composition. *Organised Sound*. 30, 2 (June 2025), 143–151. doi:10.1017/S1355771825000056

[37] Halla S. Stefánsdóttir, and Robert Ek. 2026. Excerpts from the AURA concert in Reykjavík. Video. (2026, April 28). Zenodo. https://doi.org/10.5281/zenodo.19865482

[38] Halla S. Stefánsdóttir, and Thor Magnusson. 2025. Of altered instrumental relations: a practice-led inquiry into agency through musical performance with neural audio synthesis and

violin. *Front. Comput. Sci.* 7:1578-595 (October 2025). https://doi.org/10.3389/fcomp.2025.1578595

[39] Halla S. Stefánsdóttir and Stefan Östersjö, S. 2022. Listening and mediation: of agency and performative responsivity in ecological sound art practices. *Phenomenol. Practice* 17 (October, 2022), 116–136. https://doi.org/10.29173/pandpr29464

[40] Bob. L. T. Sturm, Ken Déguernel, and Ruhing S. Huang et al. 2024. AI music studies: preparing for the coming flood. *In Proceedings of the International Conference on AI and Musical Creativity,* 9-11 September, AIMC 2024, Oxford. https://aimc2024.pubpub.org/pub/ej9b5mv1

[41] Lucy Suchman. 1987. *Plans and situated actions: The problem of human-machine communication.* Cambridge University Press, Cambridge.

[42] Lucy Suchman. 2007. *Human–machine reconfigurations: Plans and situated actions.* (2nd Edn). Cambridge University Press, Cambridge.

[43] Lucy Suchman. 2023. The uncontroversial 'thingness' of AI. *Big Data & Society.* 10, 2 (July 2023), 1-5. https://doi.org/10.1177/20539517231206794

[44] Koray Tahıroğlu, Thor Magnusson, Adam Parkinson, et al. 2020. Digital musical instruments as probes: how computation changes the mode-of-being of musical instruments. *Organised Sound* 25, (March 2020), 64–74. https://doi.org/10.1017/S1355771819000475

[45] Barry Truax. 1984. *Acoustic communication.* Ablex Publishing Corporation, Norwood.

[46] Peter-Paul Verbeek. 2008a. Cyborg intentionality: rethinking the phenomenology of human-technology relations. *Phenomenol. Cogn. Sci.* 7, (Sept. 2008), 387–395. http://doi.org/10.1007/s11097-008-9099-x

[47] Peter-Paul Verbeek. 2008b. Morality in design: design ethics and the morality of technological artifacts. *In Philosophy and Design: From Engineering to Architecture*, Peter Kroes, Pieter A. Vermaas, Andrew Light et al. (Eds.), Springer, 91–103

[48] Cheng-I Wang, Jennifer Hsu, and Shlomo Dubnov. 2016. Machine Improvisation with Variable Markov Oracle: Toward Guided and Structured Improvisation. *Computers in Entertainment* 14, 3 (December 2016), 1–18. https://doi.org/10.1145/2905371

[49] Simon Waters. 2021. The entanglements which make instruments musical: rediscovering sociality. *J. New Music Res.* 50 (March, 2021), 133–146. https://doi.org/10.1080/09298215.2021.1899247.