Organised Sound (2026), **31**, 156–167
doi:10.1017/S1355771826101198

CAMBRIDGE UNIVERSITY PRESS

Check for updates

Article

Machine learning in electromechanical sound art

Fintan O'Hare
School of Arts and Creative Technologies, University of York, UK

Abstract

This article examines the use of neural networks in electromechanical sound art and music, where sound is materially enacted through physical means such as motors, solenoids, and physical resonators. It begins with a survey of documented works, outlining a range of current strategies and discussing how technical, material, and performative factors influence their design. Identifying natural language processing as underexplored in this domain, a practice-based case study, *Seven Studies for Electric Motors*, develops one such language-based approach. The project embeds a small language model for real-time sentence generation, extracts syntax structures, and translates these into patterns of motor-driven sound. Taken together, the survey and case study offer a picture of how machine learning has been integrated into electromechanical practices over the past decade and point to possible directions for further work.

Keywords: Electromechanical; machine learning; mechatronic; natural language processing

(Received 14 September 2025; revised 30 January 2026; accepted 2 March 2026; First Published online 6 April 2026)

1. Introduction

Over the past decade, neural networks have found wide adoption in a range of musical contexts, primarily focused on the symbolic and digital domains (Briot 2021; Wang et al. 2024), but there are comparatively few documented examples of their use within electromechanical sound-producing systems. Recent works illustrate this practice, from Gardener and Debicki's *Empty Vessels* (2019), in which mechatronic cellos perform generated MIDI (Musical Instrument Digital Interface) sequences, to Quadrature's *Fantasie#4* (2019), which used radio wave data as a seed for generative models and player piano. To document artistic approaches in this emerging field, this article presents a practice review of works from the past decade, identifying how physical constraints shape the design of these systems and discussing the varied forms of interaction that arise.

The review reveals that while audio classification and MIDI generation are relatively common, the use of natural language processing (NLP) in a musical context remains underexplored. In response, the article documents the creation of a do-it-yourself (DIY) motor-driven instrument, embedding a small language model to generate syntax sequences. The project, *Seven Studies for Electric Motors*, maps the resulting sequences to motors and found objects, demonstrating how linguistic structures can be translated into physical sound and exploring the strategies for improvisational interaction that result.

This paper was guided by three questions:

1. How are neural networks currently used in electromechanical music and sound art, and which areas remain underexplored?
2. How do material constraints shape the design and function of these systems?
3. What approaches to interaction are employed in these systems, and how can they inform new modes of creative practice?

Related work in this field includes Parkinson and Dunning's taxonomy of improvisational interaction with musical machines (Parkinson and Dunning 2024) and broader surveys of robotic musicianship (Bretan and Weinberg 2016). Several frameworks have emphasised the artistic dimensions of combining computational and material processes, including Valle's description of acoustic computer music in which the limitations of found objects and actuators are seen as a creative affordance (Valle 2013), and Pigott's emphasis on the electromechanical mechanism itself as central to creative exploration (Pigott 2017). Zareei's *Audiovisual Materialism* argues that the visual and material aspects of this practice are inseparable from the sonic (Zareei 2020), connecting with a broader field of practice, including kinetic sound art and sculpture.

This article contributes a practice review of neural network applications in electromechanical sound art and insights into artistic practice in machine learning through reflections on creative work. The following section outlines the broader context, tracing a renewed interest in mechanically-produced sound and its relevance to current forms of machine learning-based musical systems.

1.1. Electromechanical music

In recent years, composers and sound artists have made increasing use of motors, solenoids, and other electromechanical devices to enact computational processes through physical sound-producing bodies. This is driven in part by increased

Email: foh500@york.ac.uk
Cite this article: O'Hare F. (2026). Machine learning in electromechanical sound art. *Organised Sound* **31**: 156–167, https://doi.org/10.1017/S1355771826101198

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

accessibility to hardware – actuators, sensors, and microcontrollers – and software – a variety of open-source tools, educational resources, and forums (Flø 2018; Zareei 2020).

Current practitioners are drawing on a history of kinetic sound art and mechanical music, with automatic instruments paired with algorithmic procedures as early as Winkel's 1821 *Componium*, a self-playing organ capable of generating variations on a theme (Zemanek 1966). The late twentieth-century resurgence of interest in mechanical instruments saw the expressive potential of the player piano expanded notably by Nancarrow and later by Peter Ablinger and others (Drott 2004; Barrett 2009). Mechatronic ensembles including LEMUR (League of Electronic Musical Urban Robots), Machine Orchestra, and SolidNoise Ensemble represent a growing interest in the expressive potential of automated instruments, computational control, and DIY methods, alongside a recognition of the unique character and qualities of acoustic sound (Harriman et al. 2016; Kapur et al. 2011; Oh et al. 2016).

At the same time, kinetic sculpture and electromechanical sound art have been developing throughout the late twentieth and early twenty-first centuries, with notable examples including Trimpin's interactive sculptures and installations (Leitman 2011). Other artists working in this field include Peter Vogel, Pe Lang, Zimoun, and Marianthi Papalexandri, among many others.

Despite the widespread use of algorithmic control in electromechanical musical systems, the incorporation of machine learning into this field is a relatively recent practice, with exceptions including Baginsky's self-learning robotic rock band, *The Three Sirens* (1992-) (Audry et al. 2022). Unlike purely electronic systems, electromechanical systems introduce a materiality that shapes both the sound produced and the conceptual approach taken by the artist. These systems present a particularly compelling case for exploring machine learning algorithms through their acoustic and material properties, requiring consideration of the instruments' physical constraints and sonic affordances, from actuation mechanisms to the resonant properties of materials. At the same time, machine learning applications can extend approaches to interaction by allowing features to be learned from data rather than being explicitly programmed, as well as facilitating novel generative behaviours. To understand how artists are working within this space, the following section outlines the methodology used to survey existing works and structure a review of current approaches.

2. Methodology

This research was conducted in three stages: a review of the field, identification of a research gap, and a practice-based exploration in response to that gap.

First, a review of existing works from 2015 to the present was undertaken to provide a survey of the field over the past decade. The search began with academic databases,[1] journals,[2] and conference proceedings.[3] Terms covered three areas: machine learning, mechatronic music, and sound art, combining at least one machine learning term (such as 'AI' or 'neural network') with at least one mechatronic term (such as 'robotic', 'solenoid', or 'kinetic') and one sound art or music term (such as 'sound art' or 'installation'). To reduce the results to a manageable size, limitations were applied to the subject area or relevant keywords. This approach highlighted the limitations of keyword search strategies, e.g., *kinetic* is a common musical metaphor, which can obscure results. Results were supplemented with a search of gallery archives, artist websites, and repositories to identify further artistic projects and performances not otherwise documented.[4] This survey is partial and reflects available documentation rather than an exhaustive catalogue.

The discussion addresses the research questions on emerging applications, material constraints, and strategies for interaction. This revealed the application of NLP within electromechanical music as a comparatively underexplored area. While a few works were found to use NLP for narrative or conceptual purposes, its potential for generating abstract musical material – building on compositional approaches such as those developed by Clarence Barlow (Barlow 2021) – remained largely unaddressed.

To explore this application in greater depth, *Seven Studies for Electric Motors* was developed: a DIY motor-driven device that integrates a small language model for linguistic syntax-based sequence generation. The project serves as an exploratory case study for embedding language models in electromechanical instruments. Reflections on this practice aim to clarify insights emerging from the practical activity of developing new work – what Barbara Bolt describes as a 'reflexive knowing that imbricates and follows on from handling' (Bolt 2007).

3. Results

Surveyed works were manually organised into categories reflecting the function of a neural network. Many of the most prominent applications involve models that are highly specialised in their domain; MIDI generation models depend on the structure of symbolic music, and speech-to-text models are designed to process phonetic and linguistic data. In these cases, the categories are partly defined by the data itself (e.g., MIDI generation, speech transcription). However, a flexible tool like Wekinator is not defined by a specific domain, and so the category *Live Parameter Mapping* is used to describe its role in creating real-time, learned connections between input and output.

3.1. Audio classification

Software libraries for machine learning-based audio classification are becoming increasingly accessible, including Flucoma (Tremblay et al. 2021) and MLStar (Kopito et al. 2023). While neural networks have been applied to audio classification since the late 1990s, these recent tools allow practitioners without expert knowledge to train a neural network on a dataset of categorised sounds, using a set of user-defined audio features to teach the model to distinguish between different sound classes. These features can describe timbral characteristics, such as Mel-frequency cepstral coefficients (MFCCs), brightness, or roughness, or rhythmic qualities.

In practice, audio classification is often combined with acoustic instruments in performer-responsive works, with two works in this category exploring parallel human-robotic configurations.

- Ben Eyes' *Ghost in the Piano* (Eyes 2024) augments a piano with motors, solenoids, actuators, and loudspeakers, analysing live guitar audio using FluCoMa to classify playing techniques, which then inform a wavetable-based sequencer

[1]RILM, JSTOR, SCOPUS, IEEE Xplore.

[2]Organised Sound, Leonardo, Tempo, Neural, Journal of New Music Research, Computer Music Journal, Journal of Creativity.

[3]NIME, ICCC, ICMC, AIMC (keyword search terms).

[4]such as Instrumentinventors.org, Ars Electronica, Research Catalogue, V2_, Lab for the Unstable Media, and the Art Machines 2 Symposium (2021), alongside artists already known to the author.

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

and synthesiser. The resulting accompaniment is projected through the piano's acoustic resonance via synthesised sound and motor-driven actuators.

- Gioti's *Imitation Game* (Gioti 2019) paired a human percussionist with a robotic counterpart, controlled by a feed-forward neural network trained to classify the rhythms, instruments, and playing techniques of the human performer in three modes of interaction: *repeat, imitate*, and *initiate*.
- Franklin's *Robo-Cajon* (Franklin 2024) also features a parallel configuration, pairing a solenoid-actuated cajon with a human-played cajon equipped with a contact microphone, classifying rhythmic data from the human performer using a neural network in MLStar, with a hidden Markov model (HMM) used to predict and generate rhythmic patterns.

3.2. MIDI generation

Machine learning-based MIDI generation typically involves training a neural network, such as a recurrent neural network (RNN) or a transformer, on a large dataset of existing MIDI. Once trained, the models can generate new sequences that reproduce the stylistic and structural features of the source material, though challenges remain in capturing long-term form and expression (Mitra and Zualkernan 2025).

These applications are well-suited to existing mechanical instruments such as the Disklavier due to their compatibility with MIDI control and the availability of extensive training datasets (Wang et al. 2024), while projects involving custom-built instruments often require the creation of new datasets that reflect the specific affordances of the instrument. However, both approaches often feature other statistical or rule-based methods in tandem with generative models.

- Quadrature's *Fantasie#4* (Quadrature 2021) mapped live radio telescope signals to player piano, generating MIDI using Piano Genie (Donahue et al. 2019), and Magenta's Polyphony-RNN was trained on twentieth-century piano music. The work presents a gradual divergence of machine learning responses from direct mappings of the radio signals, as the models elaborate on the data.
- Nikrang's *Absolute Hallucination* (Nikrang 2024) used Ricercar, a collaborative software trained on public domain classical works, to control a self-playing piano. The work explores random compositional paths through the 'creative space' of the model.
- Thompson IV (Thompson IV 2022) created duets for Disklavier and human performer, training Polyphony-RNN on MIDI transcribed from speech and piano music from Thelonious Monk to Mara Gibson.
- Refree's *Interleggere ad affectum* (2025) repurposed GrooveTransformer, a transformer-based model for real-time drum accompaniment generation, to produce a rhythmic and harmonic accompaniment in a work for multiple pipe organs. The work used a Markov-based process to generate harmonic content and note durations, with a second performer supervising the generative accompaniment (Evans et al. 2025).
- Arca, Armitage, and Shepardson's *The Light Comes in the Name of the Voice* (Shepardson and Armitage 2024) employed a generative system built around the Notochord MIDI model to extend Arca's improvisations on the Magnetic Resonator Piano, an electromagnetically augmented grand piano (McPherson 2010), into 'uncanny' reinterpretations.
- Gardener and Debicki's *Empty Vessels* (Debicki and Gardener 2019) features a mechatronic cello trio performing generated MIDI sequences, using an HMM to generate melody parts from the MIDI model's harmonic output.
- Benjamin Bacon's *Little Sound Machines* (Bacon 2018) consists of multiple devices constructed from salvaged and repurposed objects. The work used separate Tensorflow models for rhythm, harmony, and melody construction, followed by stochastic parameter generation in Max-MSP. Throughout the installation, the models learn from and influence each other in a 'purely machine-made performance'.

3.3. Live parameter mapping

Mapping non-musical data streams to control mechanical actions is a common strategy in electromechanical sound art. In works not using machine learning, this is often achieved using rule-based systems, from Zareei's mapping of cellular automata to motor frequencies and light-emitting diode (LED) patterns in *Noise Square* (Zareei et al. 2015) to the mapping of a dancer's movement to mechanically actuated percussion in Hsu and Kemper's *Why Should Our Bodies End at the Skin?* (Hsu and Kemper 2019), and Eacott's submerged sensor in *Floodtide* (2008) for translating tidal data into parameters for a mechanical organ.

The advantage of machine learning in this context is the ease with which complex and high-dimensional mappings can be created, with the relationships learned from examples provided by the user. Wekinator is a prominent tool for real-time model training, supporting a range of algorithms – from neural networks to decision trees and support vector machines – and has found wide applications in interactive music (Fiebrink and Sonami 2020), as well as the single example identified below in the electromechanical domain.

- Macionis and Kapur's *Where is the Quiet* (Macionis and Kapur 2019) used *Wekinator* to analyse EEG data (*Muse* headband), mapping alpha waves to solenoid-actuated singing bowls in response to a calm state of the participant.

3.4. System modelling

System modelling involves training a model to reproduce the dynamics of a complex system. Unlike direct parameter mapping, where an input is directly mapped to an output, this method uses a neural network to generate data that simulates or predicts a system's behaviour. The underlying process is often similar to MIDI generation, in that a model learns sequential patterns from a dataset and is then used to make predictions – but the prevalence and specific focus of MIDI generation warrant a separate category.

- In *Supraorganism* (Emard 2020), Justine Emard trained a neural network on data from a honeybee colony. The network's predictions of the colony's behaviour are embodied by robotic glass and light sculptures that produce sound by rotating wires against their surfaces.

3.5. Analogue neural networks

Some artists have used analogue models of neural networks as the basis for physical installation works. This approach can be seen in

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

analogue electronic works such as Wolfgang Spahn's *Echo State Network* (Spahn 2019) and with sound enacted through motors in Faubel's generation of emergent rhythmic patterns using 'neurally inspired' analogue oscillators (Faubel 2021).

- Spahn and Faubel's *Symbolic Grounding* (Faubel and Spahn 2018) is an installation work in which kinetic sound and light patterns emerge from an analogue network of artificial neurons – a process the artists describe as rendering the emergent 'temporal structures and functions tangible'. The work's use of neural oscillators makes this a distinct approach to the algorithmic machine learning strategies discussed elsewhere.

3.6. Speech transcription

The transcription of speech to generate musical material has been widely used in the broader electromechanical field. Many works, from Peter Ablinger's *Quadraturen* series to Canuto's stepper motor-driven *Talking Motors* (Canuto 2023), focus on spectral features, mapping the acoustic properties of the voice to automatic instruments or motors. Other approaches include physical models of the acoustic vocal tract, as found in the resonators of Ali Miharbi's *Whispering* series (Miharbi 2016).

A different approach has emerged with the development of speech-to-text systems. They function by using an acoustic model to analyse an audio signal and map its sound features to phonemes. A language model then processes this sequence, statistically determining the most probable series of words. Artistic works using text transcription often centre around the misinterpretation or fragmentation of meaning.

- Benjamin Bacon's *PROBE II, Subaudition* (Bacon 2021) used DeepSpeech to transcribe audience speech, transmitted to a second machine which translates the data to a mirrored zoetrope projection. The work focuses on the 'translation, degradation, and misinterpretation of meaning'.

3.7. Text and language generation

Modern neural network models for text generation use an encoder-decoder framework to map an input sequence into low-dimensional vectors called embeddings. A decoder then uses these embeddings to generate new text (Li et al. 2024).

Text generation has found artistic uses across music and sound art to generate musical forms, from Walshe's *Text Score Dataset* (Walshe 2021) to the deliberate undertraining of models in Kıratlı's *Cacaphonic Choir* (Kıratlı et al. 2020). In the material domain, these models can be used to imbue physical objects with personality and narrative and model non-human forms of communication.

- Jan Zuiderveld's *Conversations beyond the ordinary* (Zuiderveld 2024) embedded large language models in appliances – a coffee machine, a microwave, and a photocopier – giving the devices specific personalities; the coffee machine requires the audience's sincere desire for a coffee in order to produce one.
- Savery developed singing capabilities for the robotic marimba player, *Shimon* (Savery et al. 2021), generating lyrics with a combination of WordNet to provide synonyms and antonyms from user-defined keywords, a Long Short-Term Memory network (LSTM) for text generation, and a Markov model to refine the generated rhymes. The authors synchronise the robot's physical gestures with the lyrics, including 'emotion-conveying eyebrow movements'.
- Thomas Grill's *Mutual Understanding* installation (Grill 2019) uses two agents communicating via horn loudspeakers, inspired by an experiment in which chatbots were directed to optimise their language for negotiation. The agents develop a vocabulary independent from human language with a goal to 'maximise the beauty of their own vocal expression'.

3.8. Current and emerging applications

Table 1 provides a summary of the key works identified. The seventeen works included are those which use a documented machine learning technique and in which a physical component is central to their realisation. The table categorises each work according to its technique and tool, instrument or physical embodiment, and mode of interaction, providing an overview that informs the following practice-based work and later discussion.

The 'Mode of Interaction' column considers the agents involved, whether an interaction is directional or reciprocal, and the primary artistic purpose (imitative, responsive, generative, etc.). This practical approach was chosen over more formal frameworks, such as those by Jung (Jung 2023) or Vear (Vear et al. 2023), to accommodate the diversity of the works, which range from human-machine collaborations to autonomous systems.

MIDI generation was the most widespread application. These works used a range of approaches, from creating custom models (*Empty Vessels*) to fine-tuning existing models (*Fantasie#4*, *The Light Comes . . .*). Generative MIDI models were often paired with player pianos, likely reflecting the availability of datasets and the compatibility of models with MIDI-enabled instruments. Some works in this category used non-musical data such as transcribed speech as input (*Seven Piano Etudes Speaks . . .*) but the majority used existing piano datasets. Custom-built instruments often repurpose or combine models (*Little Sound Machines*) for the particular affordances of the instrument.

A smaller number of works turned to alternative generative strategies. These included the modelling of environmental systems, as found in Emard's *Supraorganism*, where a neural network's predictions of a honeybee colony dictated the behaviour of the installation. In a distinct approach, the emergent behaviour of an analogue neural network was embodied in Spahn and Faubel's *Symbolic Grounding*, based on the spontaneous self-organising properties of such networks (Amari 1977).

Five works used neural networks for data interpretation, with sources ranging from audio to EEG streams. Here, the main affordance of machine learning (ML) is its ease of use and robustness relative to other algorithmic approaches. Audio classification was implemented with existing software libraries (FluCoMa, MLStar), while live parameter mapping often relied on flexible tools such as Wekinator (*Where is the Quiet*). These models typically formed part of larger algorithmic systems, incorporating performer-machine feedback loops (*Imitation Game*) and stochastic elements to generate electromechanical responses. While the role of the models themselves in these works was largely functional, Bacon's *PROBE II* used errors in the speech-to-text process as a central theme of the work.

NLP is an emerging area, with four works using models ranging from speech-to-text to large language models. Text generation strategies were used to explore machine personality (*Conversations beyond . . .*) or model non-human forms of communication (*Mutual Understanding*). These focus on language as a narrative or

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

Table 1. Summary of surveyed works

Work	ML-technique	ML-model/tool‘	Data source	Instrument/physical embodiment	Mode of interaction
Eyes (2024)	Audio classification	Flucoma	Live guitar, playing technique	Motor-augmented upright piano	Performer ⇔ machine: generative accompaniment
Gioti (2019)	Audio classification	Feed-forward neural network	Performer rhythms, instruments, technique	Servo-motors, multi-percussion	Performer ⇔ machine: imitation/initiation
Franklin (2024)	Audio classification	MLStar	Performer rhythm	Solenoid-actuated cajon	Performer ⇒ machine: imitation
Macionis and Kapur (2019)	Live parameter mapping	Wekinator	EEG bio-signal (participant)	Solenoid-actuated singing bowls	Participant ⇔ machine: biofeedback loop
Emard (2020)	System modelling	Custom neural network	Honeybee colony	Motors, wires and glass	Autonomous: corpus-based generation (participant responsive)
Faubel and Spahn (2018)	Analogue neural network	Pop-Neuron Oscillator	Network dynamics	Motors, analogue electronics	Emergent behaviour
Quadrature (2021)	MIDI generation	Piano Genie, Polyphony-RNN	Radio signals, classical piano corpus	Player piano	Live data stream ⇒ machine: data-driven generation
Nikrang (2024)	MIDI generation	Ricercar	Classical MIDI corpus	Player piano	Autonomous: corpus-based generation
Thompson IV (2022)	MIDI generation	Polyphony-RNN	Speech, MIDI piano corpus	Player piano	Autonomous: corpus-based generation *(movements II, IV, VI)*
Evans et al. (2025)	MIDI generation	Groove-Transformer	Performer MIDI	Pipe organs	Performer ⇒ machine: generative accompaniment (supervised)
Shepardson and Armitage (2024)	MIDI generation	Notochord	Performer MIDI	Magnetic Resonator Piano	Performer ⇒ machine: daily reinterpretation
Debicki and Gardener (2019)	MIDI generation	Custom model	Custom dataset (MIDI)	Mechatronic cello trio	Autonomous: corpus-based generation
Bacon (2018)	MIDI generation	Tensorflow	Rhythm, harmony and melody models	Multiple devices, found objects	Machine ⇔ machine: generative composition
Bacon (2021)	Speech transcription	DeepSpeech: (speech-to-text)	Human speech (audience)	Bespoke devices, light projections	Audience ⇒ machine ⇒ machine: data translation and mistranslation
Grill (2019)	Text and language generation	Custom model	Not documented	Horn loudspeakers	Machine ⇔ machine: emergent behaviour
Zuiderveld (2024)	Text and language generation	Mistral AI (LLM)	Pretrained LLM, custom prompts	Speakers/appliances	Participant ⇔ machine: Conversational interaction
Savery et al. (2021)	Text and language generation	LSTM, WordNet (lyric generation)	Custom lyric dataset (Verse Scraper)	Shimon (robotic marimba)	Autonomous: corpus-based generation

conceptual medium rather than as a structural source for composition, which suggests opportunities for further work in this area. One such approach is explored through the following practice-based case study.

Other notable gaps include machine vision, which, despite the maturity of tools in adjacent fields, remains underexplored in this domain. Similarly, while the potential for embedding inference directly onto microcontrollers is growing with frameworks like TensorFlow Lite and Edge Impulse, few works currently take advantage of completely embedded ML processes.

4. Discussion

The following discussion considers the second and third guiding questions of this article: how material constraints shape the design and function of electromechanical systems, and what approaches to interaction are employed within them.

4.1. Material constraints

Zareei's *Audiovisual Materialism* (Zareei 2020) points to the inseparability of the material, visual, and audible aspects of a work. This can be seen clearly in the alien forms of Bacon's *Probe II*, which are themselves the sound-producing bodies. Likewise, the light and sound patterns of Emard's *Supraorganism* are a direct visualisation of an emergent system, and in player piano works, the physical action of keys and hammers is a central part of the aesthetic.

Pigott, in foregrounding the electromechanical mechanism, distinguishes between works which celebrate the unpredictable qualities of materials and those which prioritise fine, expressive control (Pigott 2017). The most common approach was the latter, with existing instruments including the player piano and MIDI-enabled pipe organ and new augmented instruments such as the Magnetic Resonator Piano (MRP). The constraints of these systems often influenced the implementation of the machine learning

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

system: in *Empty Vessels*, Gardener and Debicki created a dataset tailored to the unique affordances of the mechatronic cello trio, finding that existing datasets did not suit the capabilities of the instrument. In Arca, Armitage, and Shepardson's *The Light Comes in the Name of the Voice*, the particular timbral and dynamic affordances of the MRP required a custom approach to model training, reusing aspects of the original performance to map between the expression of the instrument and the velocity values expected by the model.

Material constraints are, however, more immediately visible in Pigott's description of unpredictable sound materials. This approach, which Valle terms 'semi-random variations . . . in the electromechanical environment' (Valle 2013), typically features DIY instruments built from motors, servos, solenoids, and found objects. Here, the inherent limitations and erratic nature of the materials act as both affordance and constraint; the unpredictability of Emard's glass-enclosed wires or ringing school bell of Bacon's *Little Sound Machines* is central to those works, but the model and electromechanical mechanism are often less tightly integrated, with the model's output serving as a trigger for an action rather than a detailed performance instruction (as in the MRP). This can be seen in *Seven Studies*, where the sequences produced by the language model were mapped to precomputed grid sequences as an intermediate stage between the model and motors. Where instrumental affordances were particularly constrained – such as the solenoid-actuated singing bowls (*Where is the Quiet*) – a threshold was used to map complex data to physical action. This strategy occurs frequently in wider electromechanical sound art, such as Greg Kress' *Brain Noise Machine* (Kress 2010), which uses a threshold on the participant's level of mental focus to control a single DC motor.

A third distinct approach was the augmentation of existing non-musical appliances, with Zuiderveld embedding personality and narrative into coffee machines, microwaves, and other devices. Here, the object's pre-existing function shapes the artistic statement. This is an idea widely explored in electromechanical sound art outside of ML approaches: for example, Dunham's *CLICK:: TWEET* (Dunham 2021), which draws comparisons between modern and past methods of communication through pairing telegraph keys with social media feeds.

4.2. Interaction

The nature of interaction ranged across the surveyed works, from human-machine dialogue to fully autonomous systems. In directional (performer ⇒ machine) interactions, a performer's audio or MIDI data served as a one-way input to the system. The purpose of one-way interactions in the surveyed works ranged from direct imitation to using performance data as a seed for a generative MIDI model. While most works in this category involved live interaction, Arca, Armitage, and Shepardson's *The Light Comes in the Name of the Voice* presented a delayed interaction in which Arca's improvisations were reinterpreted by the model in an installation setting.

Reciprocal (performer ⇔ machine) interactions involved feedback loops. In Parkinson and Dunning's *Improvising with Machines* (Parkinson and Dunning 2024), these systems could be described as collaborative works in which a human performer responds to the output of the machine and, in turn, influences the machine. This dynamic is central to creating what Vear describes as *liveness*, 'a sensation that something is co-operating in the real-time flow, or not' (Vear et al. 2023). Gioti, in *Imitation Game*, describes balancing this real-time interaction with aspects of compositional control by using a *naïve rehearsal* format, in which a human improviser has no preconception of how the system will respond.

Other works minimised or removed the human from the interactive loop, focusing instead on data-driven or emergent behaviours. Autonomous systems (*Empty Vessels*, *Absolute Hallucination*) operated without direct human input during the performance. Quadrature's *Fantasie#4* presented an interaction between a non-human data source (live radio waves) and a non-human musical agent (generative model, player piano), with the creative act as the machine interpretation of data streams. Emard's *Supraorganism* combined both dataset-based generation and responsive approaches through a largely autonomous system, but with audiences influencing the installation's behaviour through their presence.

Machine-to-machine approaches explored emergent behaviour between multiple machine agents, as in Bacon's *Little Sound Machines*, where three models 'learn from and influence each other'. Grill's *Mutual Understanding* focused specifically on the outcomes of these machine-to-machine interactions in language processing, and Bacon's *PROBE II* examined interaction through the translation and mistranslation of information passed between human and machine and between machines themselves.

Jung proposes a three-axis model for intelligent music performance system design (IMPS) comprised of embodiment, participation, and autonomy – described as the 'extent to which non-human agents are integrated into a system, and . . . the autonomous contributions within a system'. The surveyed works as material objects demonstrate high levels of physical embodiment but vary widely in autonomy and participation. Performer-machine works such as Gioti's *Imitation Game* featured high participation with autonomy moderated through performer control or predefined modes of interaction. Corpus-based systems like Nikrang's *Absolute Hallucination* demonstrate a high degree of autonomy by eliminating the human from the immediate interactive loop, even while their output is shaped by a human-curated dataset.

From an audience's perspective, it may not be clear to what extent a work is autonomous (Couchot 2019: 243). This is examined in Anna Savery's work for the robot musician *Kierzo*, where an audience survey revealed questions about whether the device was pre-programmed or responding randomly to the performer – and suggests explaining the technologies before performances as a possible approach (Savery and Savery 2024). A degree of misinterpretation can also happen in the opposite context: in Stapleton and Davis' networked performance piece *Ambiguous Devices*, pre-concert talks explained that the devices were not autonomous but performed remotely by a human collaborator (Stapleton and Davis 2012).

5. Seven studies for electric motors

5.1. Overview

The practice review identified natural language processing as an underexplored application within musical systems. To explore this further, the following sections outline a practice-based approach through the design of a DIY electromechanical instrument, using a small language model to produce syntax sequences for musical output.

Syntax patterns as the basis for musical structure occur in several existing works. Barlow's *Progéthal Percussian for Advanced Beginners* (Barlow 2003) computationally mapped parts of speech

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

to a musical score for six percussionists, framed as a course in this new language. More recently, Ko et al.'s *Harmonic Words, Narrative Chords* (Ko et al. 2024) outlines a process of deriving harmonic sequences from parts of speech, word length, and sentiment. Other artists have explored formal grammar as a generative technique, such as Clemens Gadenstätter's *Semantical Investigations II* (Gadenstätter et al. 2015).

Seven Studies for Electric Motors develops Barlow's approach in *Progéthal Percussian* through the embedding of a small language model in a DIY motor-driven improvisation device (Figure 1). The work is built around the generation of sentences and the extraction of dependency relations, structures that describe the syntactic connections between words. These patterns are then mapped to sequences of motor and light activity. Motors are used in this realisation for three reasons: first, the algorithmic process of generation, parsing, and mapping is central to the work, providing a means to create structured sequences that would be impractical to compose or realise manually. Second, these devices allow testing of a mapping method with instant audible feedback. And finally, the use of found and recycled objects is an important part of this practice; as with the use of speech, this involves placing ordinary objects in a musical context, with their imperfections and limitations drawn out through the mapping. There is no requirement for the objects to be drawn from the text or designed to imitate any aspect of human vocal production – in fact, it is only through the difference between the source and material realisation that new and surprising musical material emerges.

The found objects and actuators, including springs, copper foil, sticks, paper, plastic, and broken terracotta, are attached magnetically and rearranged between each study, introducing a degree of unpredictability. I currently work with around sixty objects, gradually adding new materials over time.

The objectives of the project were as follows:

- **Embedded text generation.** The text was generated using a small language model running on a Raspberry Pi.
- **Extract dependency relations.** These syntactical labels describe relationships between words in a sentence and were extracted from the generated text using spaCy.
- **Map dependency relations to musical sequences.** A set of speech-derived sequence fragments was used as an analogue for dependency relation tokens.
- **Enact sequences through electric motors.** Sixteen DC motors performed the generated sequences via found objects and actuators.
- **Modular objects and actuators.** Found objects were mounted on magnets to allow easy adjustment and replacement during performance. Actuators were also removable to create a variety of sounds.
- **Amplification.** Two contact microphones were used to amplify the sounds produced.
- **Lights.** LEDs were used to make the resulting patterns and sequences visible.
- **Improvisation.** To interact with the sequences performatively, an interface was developed using a MIDI controller to adjust the speed, amplitude, and other parameters in real time.

5.2. Embedded text generation

A key objective for *Seven Studies* was to create a self-contained, portable system so that I could travel and perform with the device in different locations. This required embedding the language generation process; the computational requirements of large language models make them unsuitable for this kind of local deployment, and the need for a stable internet connection would limit venues for performance. The selection of a specific small language model – *Smollm2:135m* (Allal et al. 2025) – was led by this constraint. While small models lack the accuracy of their larger counterparts, their value in this project lies in their ability to efficiently generate novel, grammatically coherent sentences.

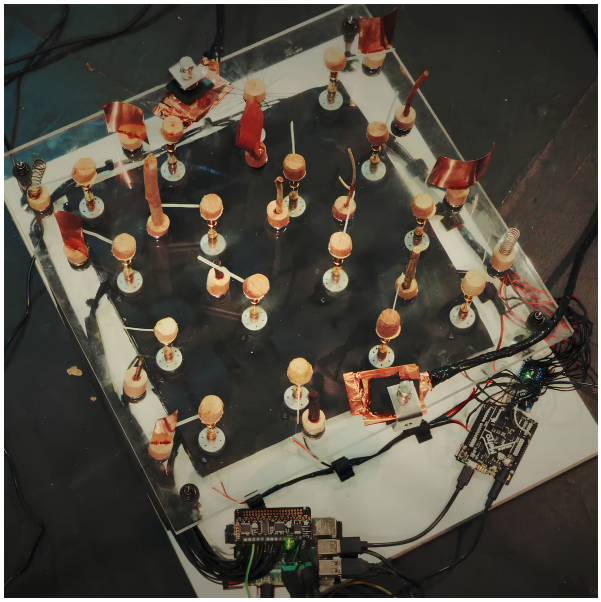


Figure 1. The seven studies for the electric motor device, showing the perspex surface, 16 DC motors, and an arrangement of found objects and actuators (copper, springs, plastic, and wood).

The model was prompted with the instruction 'Generate a sentence about x', where x is a random word supplied by the Wonderwords Python library. Each sentence is then parsed using the open-source spaCy library to extract the sequence of syntactic dependency relations, which form the input for the mapping process.

5.3. Mapping process

The mapping process converts the abstract syntactic structures from the language model into motion, sound, and light. This system was designed around an analogy: the machine-generated syntax provides the grammar to structure events in time, while approximations of human speech prosody provide the vocabulary for each event. To create this vocabulary, audio recordings of speech were segmented, analysed for pitch and amplitude, and translated into a set of over 250,000 4×4 motor activation grids, each representing a distinct sequence (Figure 2).

This was a two-stage process: each speech audio file was represented as an F0 pitch and amplitude curve and segmented on silence. Each curve was then reduced to a sequence of points in a grid, where x represents time, y represents F0 pitch estimation, and point size corresponds to average amplitude. This extreme simplification is related to Ablinger's grid-based *Quadraturen* method (Barrett 2009), applied here to generate a broad vocabulary of motor and light patterns. In performance, a randomised set of individual patterns corresponds to dependency relation labels, persisting until a remapping of the vocabulary (5.5).

This translation of a sentence to a motor sequence functions in the work as a hidden process. Its purpose is not for the audience to

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

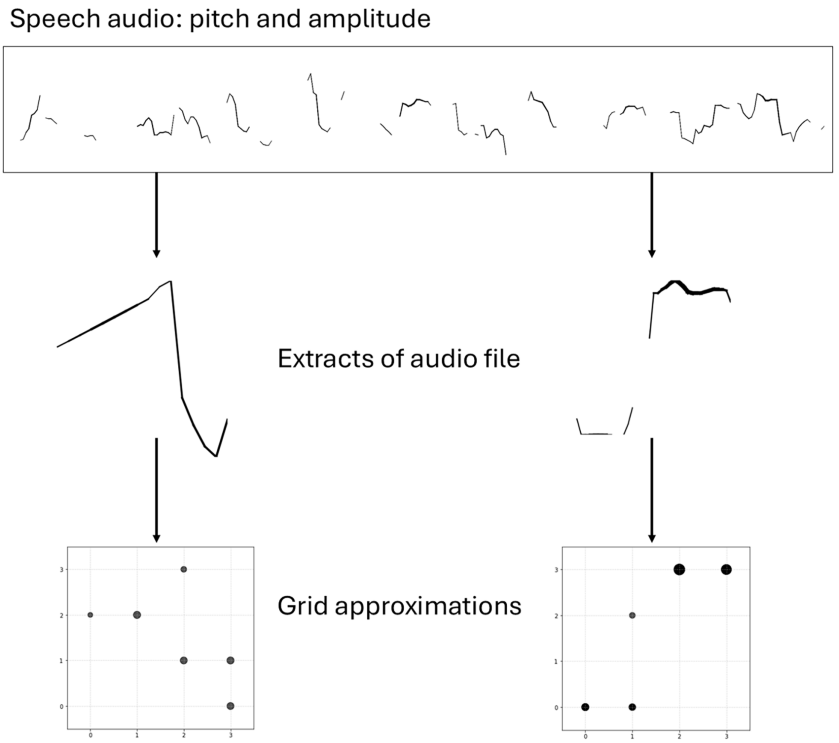


Figure 2. The mapping process: audio from speech recordings analysed for pitch and amplitude, translated into a 4 × 4 grid where point size represents motor speed.



Figure 3. A selection of objects and actuators for *seven studies for electric motors*.

consciously decipher the linguistic source but to experience the material results of this process. The brightness of each LED was mapped directly to the speed of the adjacent motor, producing sound-accompanying gestures to make the resulting patterns more easily discernible (Jensenius and Wanderley 2010).

5.4. Material approach

The material approach for *Seven Studies* aligns with the tradition of electromechanical works that celebrate the unpredictable qualities of found materials (Valle 2013; Pigott 2017). To facilitate this, the work uses a modular system, allowing objects and actuators to be added, removed, and reconfigured during performance.

Each object was mounted on a short wooden cylinder with a magnet at its base. This allowed for placement and adjustment on the perspex, with each object held in place by a second magnet from underneath. The material palette included porcelain, drum skin, thick copper foil, sticks, springs, stone, metal bells, and bars. Actuators were similarly designed for interchangeability. Each motor was fitted with a hexagonal shaft to hold carved wooden tops. These tops acted as mounts for a variety of striking and scraping implements fashioned from

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

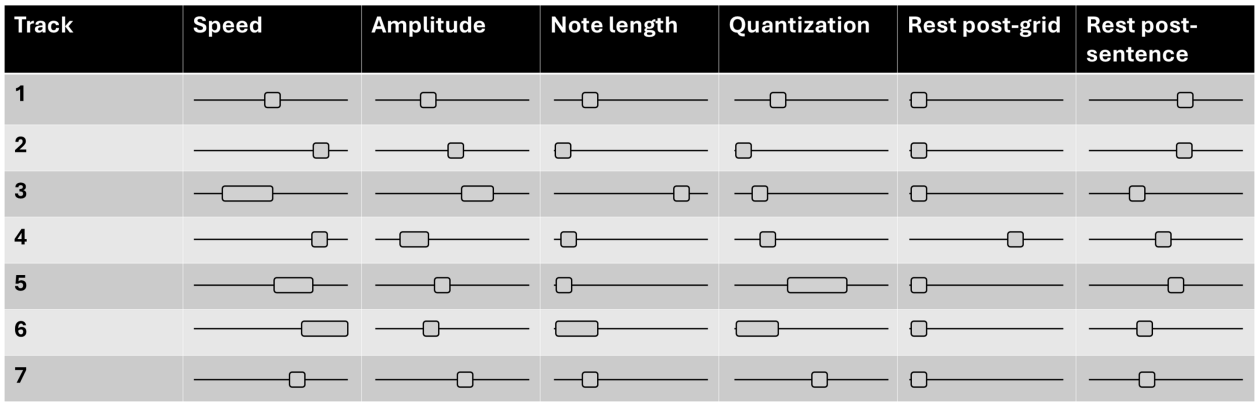


Figure 4. MIDI controller settings for each of the seven studies.

materials including plastic cable ties, springs, plastic bags, and brushes (Figure 3).

Amplification used two contact microphones, a preamp (metal marshmallow DIY phantom), and a Eurorack amplifier for high sensitivity and low noise. A perspex panel was used as the surface of the device, mounted on speaker spikes and rubber feet to reduce noise from the motors.

5.5. Interaction and improvisational control

Seven Studies was designed as an interaction between a generative system and a human improviser. While the system incorporates an autonomous process, it does not make musical decisions: the language model itself runs independently of the performer, generating sequences of dependency relations in response to random words. Instead, the interaction is built around the performer's real-time response to the system's output. Parkinson and Dunning describe this as a machine that *plays* rather than *runs*; this involves 'some element of intervention, unpredictability or randomness', and as a result, its 'machine-personality begins to materialise in the sound'. This places the performer in a curatorial role, responsible for interpreting and shaping the machine's output in real-time.

A rule-based interface was implemented using an ATOM MIDI controller, chosen for its combination of touch pads and rotary encoders. The 16 LED touch pads were configured as latching switches, allowing individual motors to be muted or unmuted. The four rotary encoders managed global parameters:

- **Overall speed** by multiplying the overall durations
- **Quantisation** allowing for rhythmic variation
- **Overall velocity** set the mean intensity of the motor strikes while preserving local variation
- **Rest post-sentence** allowing adjustment of the silence after sentences
- **Rest post-grid** in combination with a button press, this introduced silence between each dependency label for textural control

A final button allowed the performer to instantly remap the dependency labels to new motor patterns. This control scheme established the machine as a content generator, offering a continuous stream of motor and light patterns, while the human acted as an interpreter, responsible for shaping phrasing, dynamics, tempo, and texture. Three programs were run simultaneously on the device: a Python script to generate sentences using the model (via Ollama), a Pure Data patch to handle timing and controller input, and a final Python script to control the motors and lights.

5.6. Output

The video work, *Seven Studies for Electric Motors*, documents the instrument, capturing seven configurations of objects and actuators in a real-time performance using the improvisational interface.

- **1** copper, springs (video example 1)
- **2** wood, springs, copper (video example 2)
- **3** plastic, paper
- **4** copper, springs (video example 3)
- **5** terracotta, paper, wood
- **6** springs, copper, paper
- **7** copper, springs

Videos of studies 1, 2, and 4 are provided as supplementary material and online.[5] To differentiate each study, specific controller positions were either unchanged or modulated within a small range (Figure 4). Each position was arrived at through improvisation with the interface and materials. Texts generated throughout each study are not saved: the musical patterns are defined by the grammar, and linguistic meaning is lost in the process.

5.7. Reflections

The primary affordance of embedding the language model was the ability to generate novel musical material that adhered to an internal logic in a self-contained and portable instrument. The trade-off was a limitation in generative complexity compared to larger models, but even a small model can produce syntactically rich structures sufficient for musical sequence generation. The choice to generate text live, rather than drawing from a static corpus, prioritised an in-the-moment unpredictability which was central to the work's character. However, one limitation of the small model was the lack of control over sentence length, a feature often present in larger models (to a lesser extent) given the encoding of text (Fu et al. 2024).

Reflecting on the ML aspects of the system, it is clear that the language model is not an agent but a tool for producing novel, syntactically consistent text – a function that has more in common with works employing models for data interpretation than

[5]https://finohare.bandcamp.com/album/seven-studies-for-electric-motors

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

symbolic composition. In this context, the main affordance is ease-of-use relative to algorithmic approaches. While the model's ability to generate novel text drives the system's improvisational behaviour, the level of autonomy is low: the device produces material within narrow generative bounds, modulated through the interface. This level of intervention introduced a tension between data and performance: manipulation through the interface often resulted in an aesthetic decoupling from the source. This occasionally masked the patterned character of the dependency-relation sequences and shifted the focus toward linguistic data as a generative seed and to the stand-alone material qualities of the device. The machine-personality is therefore not a result of the model's creativity but an emergent property of the speech-analogy mapping and the translation of this logic into the physical domain.

This approach of repurposing a language model and dependency parser for musical generation aligned with the repurposing of materials. Because the translation from syntax to sound was a hidden process, the audience's attention was directed away from the abstract grammar and towards its physical enactment. While the objects themselves had no specific conceptual significance, the aesthetic of reuse was a key feature. Objects were chosen for their sonic qualities, picked from materials to hand and kept or discarded through improvisation with the device. A broad aim was to create a varied palette of noise-based sounds, from the pitched qualities of metal springs to the dry clicking of paper and plastic. Another was to explore variation within each material; objects were repeated in different sizes – similar pieces of broken terracotta and sheets of copper of varying dimensions. Several introduced a degree of change into performance through the variable wobble of a spring or the position of a spinning wire. This physical unpredictability is also explored in Graham Dunning's *Mechanical Techno*, in which he points to Ingold's non-hylomorphic approach to making; this involves 'collecting together raw materials and working with them', rather than working towards a fixed goal (Ingold 2010).

The compositional experience was characterised by the ability to shape and respond to this generative stream in real-time, and while the unpredictability of the output prompted the use of certain objects or actuators in response, these objects and actuators did not influence the machine's behaviour. The device could appear autonomous, generating a continuous stream of novel material without direct control. On the other hand, the rearranging of physical objects and occasional manipulations via a MIDI controller clearly shaped the musical qualities of the output. This is unresolved in its performance, leaving the audience to interpret the precise nature of the human-machine interaction. However, this one-way control contrasts with the reciprocal interaction seen in several surveyed works and could be addressed in the future by adding audio or visual input back into the system.

6. Conclusion

This article finds that the application of neural networks in electromechanical sound art has so far centred on symbolic generation, particularly through MIDI models paired with player pianos and other MIDI-enabled instruments. Beyond this, neural networks have been applied in more functional roles, such as audio or biosignal classification, and, in a smaller number of cases, to simulate environmental systems or emergent behaviours. Natural language processing, machine vision, and embedded inference remain underexplored in this domain. The case study in this article documents one possible approach to NLP and embedded inference, treating generated text as a structural grammar for physical sound production. In doing so, it extends earlier explorations of linguistic form in a live improvisation context.

Material constraints shaped the surveyed works both in terms of design and function. Those emphasising fine expressive control often adapt datasets or training methods to suit the specific affordances of the instrument, ensuring a close integration between model output and mechanical action. By contrast, DIY instruments often foreground the unpredictable qualities of materials, where limitations and irregularities become aesthetic features. The practice-based case study demonstrates one possibility in the latter approach, where modular objects and actuators allowed the improvisational shaping of a generative output. A further strand of practice augments existing appliances, where the pre-existing functions of objects shape the conceptual focus of the work.

Within these varied approaches to sound production, a range of interaction approaches were implemented, from directional performer-to-machine control through reciprocal feedback systems to autonomous and machine-to-machine interaction. Directional systems allow performers to seed or guide generative processes, as in *Seven Studies*, while reciprocal approaches involve feedback loops, where performers shape and are shaped by the machine's behaviour. Autonomous systems minimise human input, drawing on a large corpus of non-human data sources, and machine-to-machine interactions explore emergent behaviours between non-human agents.

In the specific area of the case study, future work could integrate the modular approach to acoustic sound production with the inputs to a neural network, as seen in existing physical interfaces for audio synthesis (Fiebrink and Sonami 2020; Privato et al. 2024). This approach could extend current improvisation practices with material objects. Other areas identified as underexplored through the survey include the area of machine vision, which, given the maturity of tools and approaches in adjacent fields (Ghelli Visi and Tanaka 2021), suggests space for further work.

The survey was not exhaustive, and many other applications and approaches will be in use across the field. The search was limited by the terms and available documentation, with several existing works not including detailed processes. This could be addressed in the future through formal interviews with artists. The practice-based study developed in response to the survey represents only a single possible approach, and reflections on the methods used are written from an artist-performer perspective, with more work needed on an audience perspective for a full evaluation.

Overall, the survey and case study presented here offer a partial view of the field, but one that suggests scope for further exploration of how machine learning systems can intersect with electromechanical sound-making.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S1355771826101198.

Funding statement

This work was supported by the Arts & Humanities Research Council (grant number AH/R012733/1) through the White Rose College of the Arts & Humanities.

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

References

Allal, L.B., Lozhkov, A., Bakouch, E., Blázquez, G. M., Penedo, G., *et al.* 2025. SmolLM2: When Smol Goes Big—Data-Centric Training of a Small Language Model. *arXiv Preprint arXiv:2502.02737*. https://doi.org/10.48550/arXiv.2502.02737.

Amari, S. 1977. Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields. *Biological Cybernetics* **27**(2): 77–87. https://doi.org/10.1007/BF00337259.

Audry, S., Drouin-Trempe, V. and Siebert, O. 2022. The Strangest Music in the World: Self-Supervised Creativity and Nostalgia for the Future in Robotic Rock Band "The Three Sirens". *Arts* **12**(12): 2. https://doi.org/10.3390/arts12010002.

Bacon, B. 2018. Little Sound Machines. https://www.benjaminbacon.studio/little-sound-machines (accessed 12 September 2025).

Bacon, B. 2021. PROBE II, Subaudition. https://www.benjaminbacon.studio/probe-1 (accessed 12 September 2025).

Barlow, C. 2003. Progéthal Percussian for the Advanced Beginners. https://clarlow.org/composition/from-progethal-percussian-for-advanced-beginners/ (accessed 12 September 2025).

Barlow, C. 2021. Music Derived from Other Sources. *Journal of Science and Technology of the Arts* **13**(1): 21–44. https://doi.org/10.34632/jsta.2021.9632.

Barrett, G. D. 2009. Between Noise and Language: The Sound Installations and Music of Peter Ablinger. *Mosaic: A Journal for the Interdisciplinary Study of Literature* **42**(4): 147–64. https://www.jstor.org/stable/44030532.

Bolt, B. 2007. The Magic is in handling. Practice as research. In E. Barrett and B. Bolt (eds.) *Approaches to Creative Arts Enquiry*. London: I.B. Tauris, 27–34.

Bretan, M. and Weinberg, G. 2016. A Survey of Robotic Musicianship. *Communications of the ACM* **59**(5): 100–9. https://doi.org/10.1145/2818994.

Briot, J. P. 2021. From Artificial Neural Networks to Deep Learning for Music Generation: History, Concepts and Trends. *Neural Computing and Applications* **33**(1): 39–65. https://doi.org/10.1007/s00521-020-05399-0.

Canuto, J. 2023. Talking Motors. https://www.jesuscanuto.com/talking-motors (accessed 12 September 2025).

Couchot, E. 2019. Automatism, Autonomy and Aesthetics in the Performing Arts. *Leonardo* **52**(3): 240–45. https://doi.org/10.1162/leon_a_01752.

Debicki, G. and Gardener, D. 2019. Empty Vessels. https://ars.electronica.art/newdigitaldeal/en/empty-vessels/ (accessed 12 September 2025).

Donahue, C., Simon, I. and Dieleman, S. 2019. Piano Genie. Proceedings of the 24th International Conference on Intelligent User Interfaces. California: ACM, 160–4. https://doi.org/10.1145/3301275.

Drott, E. 2004. Conlon Nancarrow and the Technological Sublime. *American Music* **22**(4): 533–63. https://doi.org/10.2307/3592992.

Dunham, P. 2021. Click:: TWEET. A Telegraphic Sound Installation. *NIME 2021*. https://doi.org/10.21428/92fbeb44.48c39866.

Emard, J. 2020. Supraorganism. https://justineemard.com/supraorganism/ (accessed 12 September 2025).

Evans, N., Haki, B. and Jordà, S. 2025. Repurposing a Rhythm Accompaniment System for Pipe Organ Performance. Proceedings of the International Conference on New Interfaces for Musical Expression 2025.

Eyes, B. 2024. Ghost in the Piano – An Investigation into AI, Robotics and Musical Performance. Proceedings of the International Conference on Artificial Intelligence and Music Creativity 2024. Oxford: AIMC. https://doi.org/10.5281/zenodo.15110248.

Faubel, C. 2021. Emergent Polyrhythmic Patterns with a Neuromorph Electronic Network. Proceedings of the 2021 International Conference on New Interfaces for Musical Expression. Shanghai: NIME, 74. https://doi.org/10.21428/92fbeb44.e66a8542.

Faubel, C. and Spahn, W. 2018. Symbolic Grounding. https://christian.faubel.derstrudel.org/symbolic-grounding/ (accessed 12 September 2025).

Fiebrink, R. and Sonami, L. 2020. Reflections on eight years of instrument creation with machine learning. Proceedings of the International Conference on New Interfaces for Musical Expression 2020. Birmingham: NIME, 237–42. https://doi.org/10.5281/zenodo.4813334.

Fløe, A. B. 2018. Materiality in Sound Art. *Organised Sound* **23**(3): 225–34. https://doi.org/10.1017/S1355771818000134.

Franklin, A. 2024. The Robo-Cajon: An Example of Live Performance with Musical Robotics. Proceedings of the 2024 International Computer Music Conference. Seoul: ICMA, 270–3.

Fu, T., Ferrando, R., Conde, J., Arriaga, C. and Reviriego, P. 2024. Why Do Large Language Models (LLMs) Struggle to Count Letters? *arXiv Preprint arXiv:2412.18626*. https://doi.org/10.48550/arXiv.2412.18626.

Gadenstätter, C., Mayer, D., Eder, T. and Nierhaus, G. 2015. Clemens Gadenstätter/Hidden Grammars. In G. Nierhaus (ed.), *Patterns of Intuition: Musical Creativity in the Light of Algorithmic Composition*. Dordrecht: Springer, 85–110.

Ghelli Visi, F. and Tanaka, A. 2021. Interactive Machine Learning of Musical Gesture. In E. Miranda (ed.) *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Cham: Springer, 771–98.

Gioti, A. M. 2019. Imitation Game: Real-Time Decision-Making in an Inter-Active Composition for Human and Robotic Percussionist. Proceedings of the 2019 International Computer Music Conference. New York: ICMA, 481–7. https://hdl.handle.net/2027/fulcrum.jq085n56r (accessed 12 September 2025).

Grill, T. 2019. Mutual Understanding. https://rottingsounds.org/threads/projects/mutual_understanding/ (accessed 12 September 2025).

Harriman, J., Bethancourt, M., Narula, A., Theodore, M. and Gross, M. 2016. SolidNoise: Making Musical Robots. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. California: ACM, 2504–10. https://doi.org/10.1145/2851581.2892539.

Hsu, A. and Kemper, S. 2019. Enacting Sonic-Cyborg Performance through the Hybrid Body in Teka-Mori and Why Should Our Bodies End at the Skin? *Leonardo Music Journal* **29**: 83–7. https://doi.org/10.1162/lmj_a_01069.

Ingold, T. 2010. The Textility of Making. *Cambridge Journal of Economics* **34**(1): 91–102. https://doi.org/10.1093/cje/bep042.

Jensenius, A. R. and Wanderley, M. M. 2010. Musical Gestures: Concepts and Methods in Research. In R. Godøy and M. Leman (eds.) *Musical Gestures: Sound, Movement, and Meaning*. New York: Routledge, 24–47.

Jung, M. 2023. Intelligent Music Performance Systems: Towards a Design Framework. *Studia Musicologica Norvegica* **49**(1): 28–44. https://doi.org/10.18261/smn.49.1.3.

Kapur, A., Darling, M., Diakopoulos, D., Murphy, J. W., Hochenbaum, J., Vallis, O. and Bahn, C. 2011. The Machine Orchestra: An Ensemble of Human Laptop Performers and Robotic Musical Instruments. *Computer Music Journal* **35**(4): 49–63. https://doi.org/10.1162/COMJ_a_00090.

Kıratlı, Ş. En, Wolfe, H. E. and Bundy, A. 2020. Cacophonic Choir: An Interactive Art Installation Embodying the Voices of Sexual Assault Survivors. SIGGRAPH 2020 Art Gallery. New York: ACM, 446–50.

Ko, J., Choi, Y. and Lee, J. 2024. Harmonic Words, Narrative Chords: Textual Sonification Using Part of Speech. Proceedings of the International Conference on New Interfaces for Musical Expression 2024, 28–32. https://doi.org/10.5281/zenodo.13904772.

Kopito, G., Schwartz, J., Amblard, J., Filman, R. and Rabern, L. 2023. MLStar: A System for Synthesis of Machine-Learning Programs. Proceedings of the 2023 Companion Conference on Genetic and Evolutionary Computation. Lisbon: ACM, 1721–26. https://doi.org/10.1145/3583133.3596367.

Kress, G. 2010. Brain Noise Machine. https://glkress.com/art-and-design/brain-art/brain-noise-machine/ (accessed 12 September 2025).

Leitman, S. 2011. Trimpin: An Interview. *Computer Music Journal* **35**(4): 12–27. https://doi.org/10.1162/COMJ_a_00088.

Li, J., Tang, T., Zhao, W. X., Nie, J.-Y. and Wen, J.-R. 2024. Pre-Trained Language Models for Text Generation: A Survey. *ACM Computing Surveys* **56**(9), 1–39. https://doi.org/10.48550/arXiv.2201.05273.

Macionis, M. J. and Kapur, A. 2019. Where is the Quiet: Immersive Experience Design using the Brain, Mechatronics, and Machine Learning. Proceedings of the 2019 International Conference on New Interfaces for Musical Expression, 335–8. https://doi.org/10.5281/zenodo.3672983.

McPherson, A. 2010. The Magnetic Resonator Piano: Electronic Augmentation of an Acoustic Grand Piano. *Journal of New Music Research* **39**(3): 189–202. https://doi.org/10.1080/09298211003695587.

Miharbi, A. 2016. Whispering. https://www.alimiharbi.com/work/whisper/ (accessed 12 September 2025).

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press

Mitra, R. and Zualkernan, I. 2025. Music Generation Using Deep Learning and Generative AI: A Systematic Review. *IEEE Access* **13**, 18079–106. https://doi.org/10.1109/ACCESS.2025.3531798.

Nikrang, A. 2024. Absolute Hallucination. https://ars.electronica.art/futurelab/en/projects-absolute-hallucination/ (accessed 12 September 2025).

Oh, H., Harriman, J., Narula, A., Gross, M. D., Eisenberg, M. and Hsi, S. 2016. Crafting Mechatronic Percussion with Everyday Materials. Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction. Eindhoven: ACM, 340–48. https://doi.org/10.1145/2839462.2839474.

Parkinson, A. and Dunning, G. 2024. Improvising with Machines: A Taxonomy of Musical Interactions. *Organised Sound* **29**(3), 340–51. https://doi.org/10.1017/S1355771824000268.

Pigott, J. 2017. Across Fields: Sound, Art and Technology from an Electromechanical Perspective. *Organised Sound* **22**(2), 276–85. https://doi.org/10.1017/S1355771817000188.

Privato, N., Shepardson, V., Lepri, G. and Magnusson, T. 2024. Stacco: Exploring the Embodied Perception of Latent Representations in Neural Synthesis. Proceedings of the 2024 International Conference on New Interfaces for Musical Expression, 424–31. https://doi.org/10.5281/zenodo.13904899.

Quadrature. 2021. Fantasie#4. https://zkm.de/en/exhibition/2022/04/fantasie4 (accessed 12 September 2025).

Savery, A. and Savery, R. 2024. Collaboration Between Robots, Interfaces and Humans: Practice-Based and Audience Perspectives. arXiv Preprint *arXiv:2407.16966*. https://doi.org/10.48550/arXiv.2407.16966.

Savery, R., Zahray, L. and Weinberg, G. 2021. Shimon sings-robotic musicianship finds its voice. In E. Miranda (ed.) *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Cham: Springer, 823–47.

Shepardson, V. and Armitage, J. 2024. Augmenting the Expressivity of the Notochord Generative MIDI Model for Arca's 'The Light Comes in the Name of the Voice' Magnetic Resonator Piano Installation. Proceedings of the International Conference on Artificial Intelligence and Music Creativity 2024. Oxford: AIMC. https://aimc2024.pubpub.org/pub/0lh6s86c/release/1 (accessed 12 September 2025).

Spahn, W. 2019. Echo State Network. https://wolfgang-spahn.de/doku.php/installation:echo_state_networks (accessed 12 September 2025)

Stapleton, P. and Davis, T. 2012. Ambiguous Devices. Installation Proceedings of the 2012 International Conference on New Interfaces for Musical Expression.

Thompson IV, W. A. 2022. Creating Musical Scores Inspired by the Intersection of Human Speech and Music Through Model-Based Cross Synthesis. Doctoral thesis, Louisiana State University and Agricultural & Mechanical College. https://repository.lsu.edu/gradschool_dissertations/5876 (accessed 12 September 2025).

Tremblay, P. A., Roma, G. and Green, O. 2021. Enabling Programmatic Data Mining as Musicking: The Fluid Corpus Manipulation Toolkit. *Computer Music Journal* **45**(2): 9–23. https://doi.org/10.1162/comj_a_00600.

Valle, A. 2013. Making Acoustic Computer Music: The Rumentarium Project. *Organised Sound* **18**(3): 242–54. https://doi.org/10.1017/S1355771813000216.

Vear, C., Benford, S., Avila, J. M. and Moroz, S. 2023. Human-AI Musicking: A Framework for Designing AI for Music Co-creativity. Proceedings of the International Conference on Artificial Intelligence and Music Creativity 2023. Sussex: AIMC. https://aimc2023.pubpub.org/pub/zd46ltn3 (accessed 12 September 2025).

Walshe, J. 2021. The Text Score Dataset 1.0. https://static1.squarespace.com/static/5300fd50e4b008058a48170a/t/6109150f3cd898236a5576fe/1627985177108/The+Text+Score+Dataset+1.0+-+Jennifer+Walshe.pdf (accessed 12 September 2025).

Wang, L., Zhao, Z., Liu, H., Pang, J., Qin, Y. and Wu, Q. 2024. A Review of Intelligent Music Generation Systems. *Neural Computing and Applications* **36**(12): 6381–401. https://doi.org/10.48550/arXiv.2211.09124.

Winkel, D. N. 1821. *Componium*. Brussels: Musical Instruments Museum, inv. 0456.

Zareei, M. H. 2020. Audiovisual Materialism. *Organised Sound* **25**(3), 362–71. https://doi.org/10.1017/S1355771820000321.

Zareei, M. H., Carnegie, D. A. and Kapur, A. 2015. Noise Square: Physical Sonification of Cellular Automata through Mechatronic Sound-sculpture. Proceedings of the 2015 International Symposium on Electronic Art. Vancouver: ISEA, 513–16.

Zemanek, H. 1966. Geschichte des Automaten: Das Componium von Winkel Ein Gerät mit einem Zufallsgenerator aus dem Jahre 1821. *It-Information Technology* **8**(1-6): 61–2.

Zuiderveld, J. 2024. Conversations Beyond the Ordinary. https://ars.electronica.art/hope/en/conversations-beyond-the-ordinary/ (accessed 12 September 2025).

https://doi.org/10.1017/S1355771826101198 Published online by Cambridge University Press