City Research Online

City, University of London Institutional Repository

Citation: Gioti, A. M., Einbond, A. & Born, G. (2023). Composing the Assemblage: Probing Aesthetic and Technical Dimensions of Artistic Creation with Machine Learning. Computer Music Journal, 46(4), pp. 1-43. doi: 10.1162/comj_a_00658

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/31551/

Link to published version: https://doi.org/10.1162/comj_a_00658

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

In this article we address the role of machine learning (ML) in the composition of two new musical works for acoustic instruments and electronics through auto-ethnographic reflection on the experience. Our study poses the key question of how ML shapes, and is in turn shaped by, the aesthetic commitments characterizing distinctive compositional practices. Further, we ask how artistic research in these practices can be informed by critical themes from humanities scholarship on material engagement and critical data studies. Through these frameworks, we consider in what ways the interaction with ML algorithms as part of the compositional process differs from that with other music technology tools. Rather than focus on narrowly conceived ML algorithms, we take into account the heterogeneous assemblage brought into play: from composers, performers, and listeners, to loudspeakers, microphones, and audio descriptors. Our analysis focuses on a deconstructive critique of data as contingent on the decisions and material conditions involved in the data creation process. It also explores how interaction among the human and nonhuman collaborators in the ML assemblage has significant similarities to – as well as differences from – existing models of material engagement. Tracking the creative process of composing these works, we uncover the aesthetic implications of the many nonlinear collaborative decisions involved in composing the assemblage.

Composing the Assemblage: Probing Aesthetic and Technical Dimensions of Artistic Creation with Machine Learning

Artemi-Maria Gioti,* Aaron Einbond,† and Georgina Born*

* Department of Anthropology and Institute of Advanced Studies

University College London
Gower Street, London WC1E 6BT, United Kingdom
{g.born, artemi.gioti}@ucl.ac.uk

† City, University of London
Northampton Square, London EC1V 0HB, United Kingdom
aaron.einbond@city.ac.uk

Research in music and artificial intelligence (AI) is, by definition, an interdisciplinary endeavor. Yet, the field of music AI seems to be leaning heavily towards the technoscientific side of its disciplinary genealogy, with methods from Computer Science (CS) and Human-Computer Interaction (HCI), and practices and discourse from research "on" or "in" music – that is, music or artistic research (Borgdorff 2006; Born 2021a) – being represented to substantially different degrees within the field. This asymmetry inevitably shapes how the discourse on topics such as datafication (that is, which aspects of music/sound are represented in training data and how), aesthetics, and the mediation of machine learning (ML) algorithms is framed. As a means of counterbalancing technoscientific approaches and framings within the field, in this article we explore the application of ML in music from an artistic research perspective, explicitly foregrounding questions of the aesthetic as they relate to the creation of data, the training of ML algorithms, and their interaction both with other software components and with humans in pieces of interactive electro-instrumental music. Building on recent work on the non-neutrality of music technology – from programming languages (McPherson and Tahıroğlu 2020; Snape and Born 2022), DAW software (Pardue and Bin 2022), and Music Information Retrieval (MIR) techniques (Holzapfel, Sturm, and Coeckelbergh 2018), to ML algorithms and data (Dahlstedt 2019; Gioti 2021a) – we aim to challenge

common understandings of data collection and the training and application of ML algorithms as "neutral" technical procedures; instead, we emphasize the inherently aesthetic, encultured, and material nature of these practices.

While a rich discourse has developed in past decades around aesthetic engagement with interactive music technology systems (Di Scipio 1998; Lewis 1999; Impett 2000), this discussion bears elaboration with respect to more recent technological contexts. Further, while the earlier systems considered from this perspective share qualities of interactivity and generativity with more recent systems incorporating ML, they differ from ML as understood in this article, which we define as the incorporation of supervised or unsupervised "learning" from a training dataset to inform interactive behaviors. Fiebrink and Sonami (2020) have reflected on the aesthetic concerns of creative collaboration with ML tools for instrument building and performance practice, but not composition, which is our focus. In what follows, ML becomes the object of practice-led research that engages critically and reflexively with its creative potential. Given the profuse criticisms of the risks and effects of ML in the wider culture, we propose that musical engagements with ML and related artistic research cannot be immune from the responsibility to reflect critically on ML's material and mediating properties.

In light of these goals, our principal research questions are: (1) how does ML shape aesthetic choices in distinctive compositional practices, and how is it shaped in turn by aesthetic commitments? (2) How can critical themes from relevant humanities scholarship inform artistic research in composition using ML? And (3) in what ways does the interaction with ML algorithms as part of the compositional process differ from that with other music technology tools? We address these research questions through two case studies undertaken by composer-researchers creating with ML. The second and third questions, in particular, elicit critical perspectives drawn out from each compositional project focused, respectively, on

issues of material engagement and datafication. Both compositional projects form part of the research program *Music and Artificial Intelligence: Building Critical Interdisciplinary Studies (MusAI)*, which investigates the cultural implications of AI through a set of critical interdisciplinary research projects focused on music and AI. Auto-ethnography, a method increasingly common in artistic creation and music research (Magnusson 2011, pp. 609-610; Findlay-Walsh 2018, pp. 121-122), is used here as a means of deepening critical reflection drawing on the ethnographic paradigm of Born (Born 1995; Born and Barry 2018). By reflexively tracking the creative process of conception, production, and performance of two new musical works (Donin 2016, 2018), we examine not only narrowly defined ML algorithms but the broader assemblage (Born 2005, 2012) in which each musical work is embedded: instruments, microphones, loudspeakers, audio descriptors, code, performers, and audience. The term assemblage, adapted from Deleuze, conceives of music as a constellation of heterogeneous mediations – sonic but also material and technological, discursive, social, and corporeal – each having a certain autonomy, where the interactions between them are non-linear and mutually catalyzing, "only contingently obligatory" (Deleuze 1988; DeLanda 2006, p. 12), and where "the assemblage's only unity is that of a co-functioning" (Deleuze and Parnet 1987, p. 69).

Uniting the compositional projects are their distinctive manifestations of distributed creativity and collaboration between human and nonhuman actors (Born 2005), including but not limited to ML algorithms. Eric Clarke and Mark Doffman propose that the term collaboration refers to when "the work of one person combines with, changes, complements or otherwise influences the work of another (or others), and is in turn influenced by it" (2017, p. 3). Not all distributed creativity is, then, collaborative, and they reserve the status of actor in musical collaborations to humans, whereas in the compositional projects discussed below it is extended to nonhuman actors – ML algorithms. The projects' collaborations are distributed in

time and space, from collaborative experimentation and sampling sessions with performers, to composers' iterative evaluations of and adjustments to the ML algorithms and the subsequent responses by those algorithms, to the multiplicity of interacting improvisations by performer and computer in the eventual performance. Both compositions also evidence "relayed creativity," resulting in what Gioti identifies as the "provisional" nature of the musical work (Born 2005, pp. 8, 30) and dramatizing contemporary experimentation with the very ontology of the musical work (Born 2005, 2022).

The two compositional projects also differ, since auto-ethnography led each composer-researcher to engage with a distinctive critical framework relevant to the work. In Einbond's case, critical reflexivity led the project to theories of material engagement (Malafouris 2004, 2013; Knappett and Malafouris 2008; Jones and Boivin 2010); in Gioti's case, to theories drawn from critical data studies (Markham 2013; Crawford and Paglen 2021; Poirier 2021). Each highlights differently the generative nature of critical reflection on creative practice with ML. At the same time, both projects innovate by introducing questions of the aesthetic into these theoretical domains. In what follows we emphasize the "aesthetic situatedness" (Snape and Born 2022, p. 223) of both composers' uses of ML as they participate in the broader compositional assemblage. We do this to counter a tendency in science and technology studies (STS) to overlook "the specific domain in which a technological assemblage is participating [… here,] music as an expressive, aesthetic and social art" (Snape and Born 2022, p. 222). In so doing we adapt Donna Haraway's critique of universalizing scientific epistemologies in which she argues for "politics and epistemologies of location, positioning, and situating, where partiality and not universality is the condition of being heard to make rational knowledge claims" (Haraway 1988, p. 589). In this light, any musician engages with ML with what might be called "situated musical knowledge – as the embodied bearer of a

particular musical history and culture" (Snape and Born 2022, p. 223). Moreover, by putting auto-ethnography in dialogue with theory, our research does not aim merely to illustrate existing theoretical precepts. Rather, we adopt an epistemological stance that Born terms "post-positivist empiricism" according to which ethnography can be "a subtle tool for the application and the amendment of theory" (Born 2010, pp. 197-198; 2022, pp. 12-13). In the central sections of this article, the necessarily situated and subjective accounts of creation with AI are narrated in the first person in the individual voice of each composer.

Case Study 1: *Prestidigitation* for Percussion and Interactive 3D Electronics

Einbond's composition *Prestidigitation*, written in close collaboration with percussionist Maxime Echardour, is scored for a custom-built percussion setup and specialized 3D microphone, and loudspeaker, all of which are intimately integrated in the conception of the work. (The performance video and score are available at this link: <<TYPE: Please insert URL>>.) Collectively these tools perform three interacting ML tasks: first, a k-nearest neighbor (k-NN) algorithm is used to select short samples from a prerecorded corpus of percussion, performed and recorded by Echardour and Einbond, and then to organize them into an electroacoustic texture that imitates a longer live or recorded target. This process may be referred to as concatenative synthesis, and the result as an audio mosaic (Schwarz 2007). Second, a Gaussian mixture model (GMM) associates each of these short samples with a database of 3D radiation patterns derived from acoustic instruments that are then used to diffuse the samples spatially (Einbond 2021). And third, an audio oracle (AO) models the sequence of these samples in time (Surges and Dubnov 2013). This serves as training data for computer improvisation by connecting samples based both on timbral and contextual similarity: samples are selected that are not only

similar themselves, but preceded or followed by similar segments. (For technical details, see Einbond et al. 2016 and the companion article by Einbond et al. in this volume.)

In order to delve more deeply into creating with ML, I (Einbond) bring auto-ethnographic reflections to bear on the process. My approach to composing for the ML assemblage – as noted earlier, and as for any artist working with technology – is situated by my embodied aesthetic history and culture, which I term "musique instrumentale concrète," adapted from Helmut Lachenmann's *musique concrète instrumentale* (Lachenmann 1996, pp. 378, 382; Einbond 2016, p. 156). It combines influences from the field of *musique concrète*, taking audio recordings as primary source materials, with experimental instrumental performance techniques centering on noisy sounds organized by their timbral characteristics. Through composer-performer collaborations, I begin work on each new project with experimentation focused on a specific performer's instrumental technique. I call this "radical personalization," derived from composer Richard Barrett's phrase "radically idiomatic," which in turn is translated from the guitarist Derek Bailey's "non-idiomatic improvisation" (Bailey 1992; Barrett 2003; Einbond 2013, p. 63). Barrett suggests that when performers work against the traditional, "idiomatic" performance techniques of their instruments, as Bailey did, they discover sounds and timbres suited to their particular expressive capabilities. I take this concept one step further, working with individual performers to tune in to techniques that might be unique to their bodies and instruments, which we preserve through detailed sampling. These samples, capturing the performer's and my collaborative creative labor at that particular moment, form the basis not only for the electronic materials for the work, but also the notated score, based on a process of audio mosaicking using timbral descriptors, as defined above. Significantly, as elaborated by Gioti below, I conceive of these descriptors not as objective data but expressive materials

subject to creative manipulation. The decisions surrounding the selection, analysis, and processing of these data are not transparent or neutral, but fundamentally aesthetic.

Machine Learning and Material Engagement

A scene from the final rehearsals of *Prestidigitation* typifies Echardour's and my work with ML as a component of the complex assemblage required for the piece. The rehearsal followed nine months of collaborative sessions, individual practice, and ongoing compositional work, documented regularly through auto-ethnographic recordings, videos, notes, and reflections. As we reached the final page, the computer improvisation sounded out-of-control, with too many layers of sound that prevented Echardour from hearing himself play. This judgment is both objective and subjective: when the density of the electronics saturates the audio output it causes unprofessional audio clipping; at the same time, Echardour and I judged that the busy texture did not fit the delicate, crystalline aesthetic that I sought in the work.

With the performance the next day, any changes to the electronics needed to be rapid and practical. I tried a few possibilities, tested out iteratively: raising the segmentation threshold for the incoming audio signal from the live percussion so that fewer sounds would trigger a response from the k-NN algorithm. I lowered the average length of the samples, reducing the amount of overlap, and therefore the density of the texture. I changed the parameters of the AO algorithm that guides computer improvisation to clear the training data from past events more frequently. This shortened the AO memory so it could only access sonic events that had occurred closer to the present musical context. Finally, as the bluntest solution, I turned off the computer improvisation output more often so that more silences were imposed.

Along with these changes to technical parameters, my fingers constantly sat on physical faders controlling levels in the work's Max patch, subtly adjusting the data input to and output from the ML algorithms in continuous feedback with my focused listening. These small shifts could lead to significant changes in the results, especially where threshold values are involved. Similarly, no two interpretations by Echardour are the same, each eliciting fine differences of color and intensity, both informing and responding to the ML assemblage. While we each have some expectation of how certain parameters could lead to specific musical results, several trials and careful listening are needed to evaluate how each change affects different possible ML reactions. Crucially, this process is iterative and speculative, feeding back through ears, eyes, and fingers, gradually leading through prediction and evaluation to an emergent aesthetic output.

Although the details may seem technical, the process is intuitive, based on rapid but sensitive interaction between performer, composer, and the various computerized elements of the assemblage. The overall experience can be summarized by analogy with anthropologist Lambros Malafouris's description of the interaction between a potter and the clay material she is working with when throwing pottery on a wheel:

> At one moment, movement is effortless and feels like happening to the potter rather than being done by the potter as if totally absorbed into the micro-structure of clay. At another moment, the potter is clearly conscious of moving clay around and shaping it, directing the flow of the clay and struggling to control the act and handle the clay. (Malafouris 2008, p. 34)

Malafouris defines this process as "material engagement," resulting from the interaction of human and nonhuman materials such that agency is emergent: "Agency is a property or possession neither of humans nor of nonhumans. Agency is the relational and emergent product of material engagement. It is not something given but something to become realized" (Malafouris 2008, p. 34).

Like potter and clay, Echardour, the computer, and I all engage in a gradual push-and-pull where we each, human and nonhuman, listen, react, then make iterative adjustments to reshape our interactions. Echardour interprets a passage in the score; the computer "listens" to train its ML model and outputs a computer-improvised response; I listen and adjust the ML algorithm parameters; Echardour listens and refines his performance; and we recursively repeat the process. While Echardour's and my actions may appear intentional, Malafouris draws a distinction between agency and intention as it is conventionally understood: referencing John R. Searle (1983), rather than "prior intention," a "premeditated or deliberate action where the intention to act is presumably formed in advance of the action itself," Malafouris cites Searle's concept of "intention-in-action," "non-deliberate everyday activity where no intentional state can be argued as being formed in advance of the action itself" (2008, p. 29). The latter better fits my experience of the immediacy of the creative workflow with ML, rather than it entailing premeditated decisions causing fully predictable outcomes. Intention-in-action recalls Donald Schön's (1983, pp. 49-50) notions of "reflection-in-action" and "knowing-in-action," which have been influential in HCI scholarship (Baumer 2015). However, Malafouris's account more effectively captures the multivalent interactions between composer, performer, and technical assemblage and the immediacy of feedback between the collaborators.

Andrew Pickering (1995) offers a similar account to Malafouris in the context of STS with his idea of the "dance of agency:" "an understanding of scientific engagement with the material world as a temporally extended back-and-forth *dance of human and non-human agency* in which activity and passivity on both sides are reciprocally intertwined" (2010, p. 195, original italics). His description of this process as "performative," "emergent," and "decentered" (2010, pp. 195-196) is echoed in Malafouris's account of pottery, or in my account of artistic creation with ML. Significantly, this could also apply to musical assemblages without ML, such as

the interactive and generative systems cited earlier from Di Scipio, Impett, and Lewis, as well as more recent applications with Max software. As Snape and Born (2022) write in their analysis of Max's mediation of the musician Mark Fell's creative practice, referencing Malafouris and Pickering among others: the nonhuman routines composing the "dynamic ecology of intertwined paths of data flow" within Fell's basic Max patch "are not only generative but exhibit a primitive self-organising quality, […] a capacity to produce a wide range of meaningful outputs based on a single general specification" (Snape and Born 2022, p. 230).

If even basic Max patches can be described in terms of reciprocal engagement between nonhuman and human agency, what new models of material engagement do more complex ML algorithms bring to the "dance of agency"? ML tools are often attributed with agency and intentionality: they appear to listen, react, and make decisions. Yet the emergent and decentered nature of these potential actions is often overlooked, as is the role of the human input that is frequently, but not always, "leading the dance." In both accounts of composing with ML in this article, we emphasize the centrality of human control on the part of composers and performers, allied to the aesthetic sensibilities they exercise. However, emphatically, returning to Malafouris's distinction: this control more closely resembles "intention-in-action" than "prior intention." Each time I adjust the parameters of the ML agents operating in the Max patch for *Prestidigitation*, I listen, react, adjust, and listen again. I cannot fully predict the decentered outcomes of the ML algorithms and their complex interaction with the multifaceted technological assemblage – but I can influence them. I neither control individual sonic events, nor the exact internal mappings "learned" from the training data (in contrast to a directly programmed system). But I do act on the statistical likelihood of certain outcomes based on my technical and musical experience and knowledge. This is part of my "background," to adapt another concept from Searle: "a set of skills, stances, pre-intentional assumptions

and presuppositions, practices and habits" (Searle 1983, p. 154, quoted in Malafouris 2008, p. 32) that form the grounds for intentionality. Malafouris, however, argues trenchantly that this "background" cannot be separated from material engagement: "the 'Background' becomes part of […] what we may call an extended intentional state" (2008, p. 32). In this light,

> The mediational potential of a certain artefact in a quite significant way shapes (both in the positive and negative sense of enabling and constraining) the nature of human intentions. […] The artefact should not be construed as the passive content or object of human intentionality but as the concrete substantiating instance that brings forth the intentional state. (Malafouris 2008, p. 33)

The ML algorithms, the performer, and I act together through the emergent material engagement constituting the extended creative process. Yet, compared to interactive systems that do not employ ML, the actions of the ML agents are even more decentered, unpredictable, and contingent not only on the performative inputs and outputs of the assemblage, but also on the training data and its condensation of artistic labor and situated aesthetics (as elaborated below). I do not know in advance which samples the computer will choose, or when, and if Echardour and I were to recreate the sample database the unpredictability would multiply. And yet, as we fine-tune these parameters, we are nevertheless capable of molding the details of the sonic outcome, even as these details "push back" by causing us to listen and readjust, again evoking the metaphor of clay's resistance. While the ML assemblage is more materially and conceptually complex and decentered than clay – or indeed other non-ML assemblages – my experience is still of "intention-in-action" in a dialogue (or "dance") of expressive agency.

Subjectification and Space

Material engagement with ML in *Prestidigitation* also led to specific aesthetic choices in the ways composer, performer, and listening subjects interact with the

performance space, aiming to address critically the neglect by music ML of embodied listening. The work was inspired by the concept of placing the listener metaphorically in the midst of the percussion instruments to experience the intimate sounds otherwise only heard by the percussionist. Echardour and I engaged with our hands and ears to shape the percussion instruments in a circle, like a mobile sculpture, around a specialized microphone capturing sound in 360°. The microphone, the mh acoustics Eigenmike 32-channel spherical microphone array, is connected via the ML assemblage to another specialized piece of 3D audio equipment, the IKO 20-channel spherical loudspeaker array. Unlike traditional loudspeakers surrounding the listeners, the IKO is situated among listeners in the midst of the space, more akin to an acoustic instrument. Like Malafouris's potter, Echardour and I "sense [...] and exchange vital tactile information necessary for a number of crucial decisions" (Malafouris 2008, p. 19) through a gradual process of selecting, positioning, and listening to the instruments, microphone, computer, and loudspeaker.

The resulting sculptural setup contributes a second angle on material engagement through its critical perspective on subjectification by ML – referring to the normative construction of the listening subject by music AI. Applications of AI to music so far show a limited awareness of the significance for sonic experience of the material, embodied presence of the listening subject situated in space. They tacitly assume a stereo or mono listening environment without consideration of the possible mediations of headphones, loudspeakers, or the interactions between the bodies of listeners and performers within a listening environment. They assume, that is, an unsituated and disembodied listener. This is especially surprising given that, in the wake of the development of the fields of sound art and sound installation, sound itself has been acknowledged as intrinsically "perspectival and relational […] in the sense that it is always experienced from particular subjective

and embodied, physical and social locations" (Born 2013, p. 17). In this light, *Prestidigitation* aims to "re-embody" sound synthesis in two ways: by learning from the spatial presence of live instruments and performer, but also by cultivating the audience's awareness of their capacity to orchestrate their listening through their embodied presence and movement around the performance space.

Technically, the first aim is implemented through a GMM as defined above: in response to the live input, the computer seeks to "learn" from the complex spatial characteristics of acoustic instruments and respond with synthesized sounds diffused spatially through the IKO. Significantly, percussion instruments are not included in the training database, so there is no "ground truth" against which to evaluate the machine learner's choices: this means the latter contributes its own creative "agency" in proposing novel combinations of timbres and spatial characteristics; it also represents my own aesthetic background favoring an experimental approach to electroacoustic sound over an attempt strictly to "reproduce" natural phenomena.

The second aim emerged from experiments in which Echardour and I further interrogated the performance setting of *Prestidigitation*. We decided to invite the audience to listen while moving freely around the performance environment, with open access to the space around, between and behind the sculptural percussion setup, the percussionist, and the IKO. If ML applications for music have tended to neglect the spatially situated, embodied, and relational experience of listening, then the concept of material engagement, with its stress on the relational and emergent nature of "intention-in-action," brings these aspects to the fore and has clear analogues with this conceptual paradigm for listening. Moreover, by participating, through their embodied movement in the space, in shaping their own listening experience, listeners engage interactively in the performance of *Prestidigitation*, in a sense becoming musical collaborators with the performer, composer, and

technological assemblage, and evoking sound art's commitment to "participation, interactivity [and] collaboration" (Born 2013, p. 18). However, this choice may challenge the norms of live concert performance, especially when it involves a complex technical setup, as we encountered during the second performance of the work at a different venue, where space did not permit an ambulatory audience.

Improvisation and Agency

Material engagement with the technological assemblage led to aesthetic consequences that would not have occurred without ML, through the developing relationship between improvised and fixed elements in *Prestidigitation*. Throughout the creation process, Echardour and I tested and rebalanced these elements, with profound implications for not only the proportions of the finished work but also its aesthetic language. In this way, our "dance of agency" with ML challenged and stretched our own creative identities, leading to a finished score featuring a unique combination of detailed musical notation and guided improvisation. Arguably, not only we humans but also the ML algorithms exercised "intention-in-action" in our dialogue of material agencies.

A key input to the process is human improvisation by Echardour, which I recorded during his initial explorations of the percussion setup. I analyzed one 3-min improvisation using audio-mosaicking: as described above, the choice of analysis parameters and subjective editing of the results were far from neutral, but rather represented expressive creative actions situated by my aesthetic roots in "radical personalization." The transcription served as central source material for *Prestidigitation*, which I transformed to produce the notated score, to be reinterpreted by Echardour in live performance. In this way, Echardour's performative labor and creative input are crystalized not only in the work's electronic materials but also in the score via processes of machine and human listening.

The electronic patch for *Prestidigitation*, in turn, features computer improvisation implemented with the AO algorithm. The oracle "listens" to Echardour's interpretation of the score, which is itself based on my detailed transcriptions of his initial improvisation, as live training data for computer improvisation as defined above. However, to what details the machine listener is attending is not transparent: in the context of *Prestidigitation*, the materiality of these choices decisively shaped the aesthetic result. Descriptors for relative specific loudness (RSL) facilitated the computer's comparison of noisy percussion timbres, allowing it to respond to the performer with what could be termed "computer noise improvisation." I choose these descriptors to exploit the wide timbral array of unpitched percussion instruments selected by Echardour and me, which in turn could be attributed to my own situated aesthetics of instrumental *musique concrète*. However, by responding with unexpected associations between sounds that may differ from those of the human composer and performer, the ML algorithm challenged us with its own material agency.

In this way, the "intention-in-action" of the ML assemblage required deep engagement and transformation, in turn, of our creative approaches. My first draft of the notated score several months before the performance left Echardour feeling constrained because it did not leave him space to improvise live in reaction to the unpredictable computer improvisations. While this followed from Echardour's own culture of improvisation, it challenged my aesthetic orientation toward precise notation, albeit based on detailed transcriptions of improvisation. In response, I added several passages of guided live improvisation, shaping the outcome by setting the timbral color of the instruments used in each section. The new version permitted Echardour to respond dynamically to the computer improvisation, while giving sufficient control of timbre and timing to blend smoothly with the fully notated sections. Our interaction with the novel conditions of creating with ML led,

then, to a significant shift in the final aesthetic results. Questioning the separation of "compositions for improvisers" from "work which 'incorporates' improvisation" (Lewis 1999, p. 102), Echardour's and my material engagement with the assemblage, situated by our aesthetic backgrounds, led to a unique juxtaposition of these approaches that would not have occurred without ML.

Case Study 2: *Bias II* for Piano and Interactive Music System

The second musical work discussed in this article, Gioti's *Bias II* for piano and interactive music system, is part of a series of works engaging with the materiality of ML algorithms and data (A documentation of the piece is available here: <<TYPE: Please insert URL>>). The piece uses ML to model interpretative choices made by pianists in past performances, setting performers in an explicit dialogue with the work's interpretative history. During its interactions with different pianists, the computer music system collects data pertaining to the way performers navigate a set of seven clusters, each consisting of a variable number of timbrally similar musical actions. Based on predictions made by a Recurrent Neural Network (RNN) trained on these data, the computer co-determines the form of the performance, by choosing to follow the musician or propose musical changes. Historical data, collected by the computer music system in past performances, influence future performances of the work, making *Bias II* a "provisional musical work" that "both retains and blurs the traces and boundaries of individual and collective authorship" (Born 2005, p. 30). Rather than being an independent, self-contained event, each performance of the work is part of a co-creative process that involves both humans and nonhumans and is dispersed in space and time. Machine learning in this context becomes the medium through which traditional notions of musical authorship and the ontology of the musical work are challenged and critically reflected upon.

At the same time, *Bias II* is an exploration of the materiality and aesthetic affordances of the ML algorithms used in it: a feedforward Neural Network (NN) that assigns incoming sounds to one of the seven timbral clusters in the score (classification), and a Recurrent Neural Network (RNN) that predicts possible continuations of the performance on the basis of these clusters. The predictions of the machine listening algorithm (feedforward NN) are processed to extract the predominant timbre/cluster over a time window of 1 sec and then fed into the RNN, which predicts which timbre is likely to follow next.

The score of the piece consists of a total of seven clusters of timbrally similar musical actions involving primarily inside-piano playing techniques and string preparations and exploring the extended capabilities of the piano. The performer is free to navigate this timbral space by transitioning freely between the clusters. During the performance, the outputs of the machine listening algorithm are used to match different timbres to different signal processing techniques, but are also fed into the RNN. If the prediction of the RNN differs from what the performer is currently playing, the system responds by playing back prerecorded samples of the predicted cluster.

In order to perform this piece, performers first interact with the system in the context of "training rounds." These are run-throughs of the piece in which the system reacts to the performer's actions using signal processing, but does not act proactively (i.e., it does not propose any sound material). Recordings of these "training rounds" are analyzed and added to the dataset used to train the RNN, influencing its behavior in future performances.

The compositional process for *Bias II* involved extensive experimentation with inside-piano playing techniques and string preparations, coding, training ML algorithms, creating a score, and working closely with pianists Magda Mayas and Xenia Pestova Bennett, whose distinctive interpretations of the score – crystallized in

the form of training data – are an integral part of the piece. Among the many critical perspectives arising from this interrogation of ML processes from an artistic research perspective, in the next sections I (Gioti) focus on what emerged as the main conflation point between the "situatedness" (Snape and Born 2022, p. 223) and individuality of my own aesthetic language, and ML algorithms: "data-making" (Vis 2013). Drawing from critical data studies, I examine the aesthetic dimensions of material contingencies of the data creation process, and use Poirier's (2021) denotative, connotative, and deconstructive readings of datasets to explore data-making as itself an encultured compositional act.

Training the Machine Listening Algorithm

Training the machine listening algorithm (feedforward NN) used in the piece involved an iterative process of data collection (recording examples of the seven timbral classes plus silence/background noise), data preprocessing, analysis (feature extraction), training, and testing. Each new iteration of data collection and training aimed at counterbalancing tendencies and biases in the network's predictions that could be traced back to the specificities of the training set – or in ML terms, improving the network's ability to "generalize" on previously unseen examples and avoiding "overfitting" (a phenomenon that consists in a ML algorithm performing well on the training data but poorly on previously unseen examples). In addition to this practical function, this process yielded tangible insights into the materiality of datasets, including the contingencies introduced by the analysis techniques, decisions, and hardware involved in the data collection process.

The mediation of the different microphones employed – with their different frequency responses, directionalities, and other material properties – seemed to be the most tangible of these contingencies. Training the feedforward NN with data collected using a single microphone quickly led to overfitting (i.e., reduced accuracy

on examples recorded with different microphones), highlighting an inherent tension between small-data, artistic approaches and the data-hungry nature of ML algorithms. What followed was an iterative process of data collection (using various microphones), training, and run-time testing, driven by qualitative judgments about the classification errors made by the model. While improving the overall (quantitative) accuracy of the model, these decisions, along with the six different microphones and three different pianos used to record the training examples, added to its material contingency, by shaping the qualitative properties of its errors.

Another layer of mediation in the data collection process was introduced by the analysis techniques used to extract features for ML. This step of the process involved the use of standard timbral descriptors – specifically Mel Frequency Cepstral Coefficients or MFCCs – which originate from and are widely used in speech recognition (Davis and Mermelstein 1980). In this layer, the digital audio signal was further mediated by mathematical processes grounded in a model of timbre derived from speech, introducing an additional source of material contingency and friction.

The decision to record examples of one timbre at a time was as consequential to the training process as were the material properties of the microphones and audio descriptors used in the data collection process. As a result of this decision, when performers superimpose musical actions from different timbral clusters during the performance, the machine listening algorithm is unable to make accurate predictions, producing a distorted representation of their interpretative choices. While originating from an "erroneous" expectation as to how performers might engage with the notated musical material, this decision led to sonically rich interactions and emergent behaviors, resulting from the RNN forging "false" associations between different timbres. In this case, I decided to exploit "errors" in the predictions of the machine listening algorithm for their potential to generate emergent behaviors – as opposed to eliminating them through further data

collection and training. This decision constitutes an example of "intention-in-action" and highlights the reciprocal relationship between aesthetic commitments and material engagement.

Data Materiality

The contingency of ML algorithms on the decisions and material conditions involved in the creation of the training data underscores the material dimension of datasets and suggests an understanding of data as an imprint of human decisions and material mediation. In contrast to the common usage and understanding of the term, data do not exist in nature, neither can they be "collected" (Markham 2013). The auto-ethnographic account presented here supports an understanding of data as process: as something that is made (and therefore artificial), situated, and contingent on material conditions, rather than "a priori and collectible" (ibid). Terms such as "data collection" or "data extraction" are indicative of a rhetoric of "naturalization" (Bowker and Star 2000, pp. 294-295), which renders the contingencies involved in the data creation process invisible (Denton et al. 2021). Rooted in objectivism and the "neutrality ideal" (Harding 1992), this rhetoric detracts from the materiality of data and their embeddedness in larger social-cultural-technical assemblages that shape what is rendered visible and invisible in the data, and through which (distorting) lenses. Data are not extracted, but rather constructed through processes of framing (Markham 2013), reduction and translation. In this data creation process, human decisions, tools, and practices take on a "data-making agency" (Vis 2013).

The data-making process described earlier involves successive layers of translation and reduction, driven by culturally grounded processes of framing. First, acoustic sound is translated into an electrical and then a digital signal, in a process that, far from being neutral, leaves an imprint of its material conditions – thus, two different microphones will produce two entirely different representations of the

same sonic environment. This representation of the acoustic sound is further mediated and abstracted through Music Information Retrieval (MIR) processes that produce a fundamentally "oligoptic view" (Kitchin 2014, p. 4) of the represented phenomenon (acoustic sound) – that is, a view from a specific perspective or "vantage point" (ibid). In *Bias II*, like in *Prestidigitation*, sound is analyzed from the perspective of timbre, a choice that is indicative of an aesthetic-cultural framing that prioritizes timbre – over, for example, pitch or rhythm – as the basis on which sound material and form are differentially and relationally constructed.

Crystallized in this framing is the "situatedness" (Snape and Born 2022, p. 223) of my own aesthetic language, as it has developed over many years of material engagement with the piano and the unique, rich sonic possibilities that its open mechanism affords, and as it has been shaped by its relation to a "musical past" (e.g., John Cage's works for prepared piano) and an anticipated "musical future" (Born 2005, pp. 63-64, 2015). Part of the work's dialogue with the history and repertoire of the piano is a process of defamiliarization of the instrument, through extended techniques and preparations that produce sounds that are uncharacteristic of a stringed percussive instrument – for example, glissandi produced by sliding a sharp plastic object between strings of the same pitch, or sustained sounds produced using an Ebow. This involves shifting the interface of the instrument from the keyboard, inscribed in which are several centuries of Western music theory (Magnusson 2009, p. 171), to its sound production mechanism, allowing for physical manipulation and interaction with the strings and other parts of the instrument, such as its metal and wooden frame.

While these processes of cultural and aesthetic framing drove the data-making process, their relation to mathematical abstraction was not free of friction. Timbral descriptors, such as MFCCs, for example, are a far from perfect approximation of human perception of timbre – if universal human auditory capacities can be

assumed to exist. Such frictions became apparent in a small experiment I conducted as part of the compositional process for *Bias II*. This experiment involved recording examples of the different musical actions involved in the score, analyzing them using MFCCs, and feeding the data derived through this analysis into a clustering algorithm. The k-means clustering algorithm, which was used in this experiment, is an unsupervised learning algorithm that groups data points together based on similarity as measured by a distance metric (e.g., Euclidean distance), rather than based on human labeling. In this experiment, the algorithm was used to group the musical actions involved in the score into seven clusters and its outputs were compared to the seven clusters in the score. Unsurprisingly, the clusters produced by the algorithm deviated significantly from my own (subjective) perception of timbral similarity and dissimilarity and seemed to be based on crude spectral differences (e.g., concentration of energy in the lower vs higher frequencies), rather than on timbral differences within more narrow pitch ranges. This experiment productively demonstrates the mediation of audio descriptors, while deconstructing the notion that spectral data can be "neutral" or "objective."

In *Bias II*, the predictions of the classification algorithm undergo further processing, driven by aesthetic considerations and my intention to balance generative and composed elements of the assemblage. Specifically, all predictions made by the feedforward NN within 1 sec are stored in an array and a majority rule is used to derive the predominant timbre for that second. This is an aesthetic decision meant to limit the maximum frequency of musical changes proposed by the interactive music system and arrived at through processual knowing and material engagement with the assemblage.

Data Semiotics

The decision to only trace the predominant timbre for each second of audio is an example of demarcation in the data-making process: a process of delimiting what gets traced from what doesn't. In an attempt to unveil such processes of demarcation and the politics behind them, Lindsay Poirier (2021) proposes three modes of reading datasets, derived from semiotics: a denotative reading, in which "the analyst momentarily assumes a neutral position" (p. 3) by focusing on the literal meaning of the values; a connotative reading, aiming to "situate data semantics historically and culturally" (p. 4); and a deconstructive reading, tracing what "gets Othered" (p. 1) in the data. Poirier's three modes of reading datasets are conceived as a pedagogical tool meant to draw attention towards the assumptions and politics manifest in data, in this way challenging naïve understandings of "datasets as essentially aperspectival structures for storing a priori truths and bias as an external force that contaminates them" (Poirier 2021, p. 2). To more closely investigate the aesthetic implications of processes of demarcation in the data-making process in *Bias II*, in the following I read the data fed into the RNN using Poirier's denotative, connotative, and deconstructive readings, albeit with a focus on their cultural and aesthetic rather than political "provenance" (ibid).

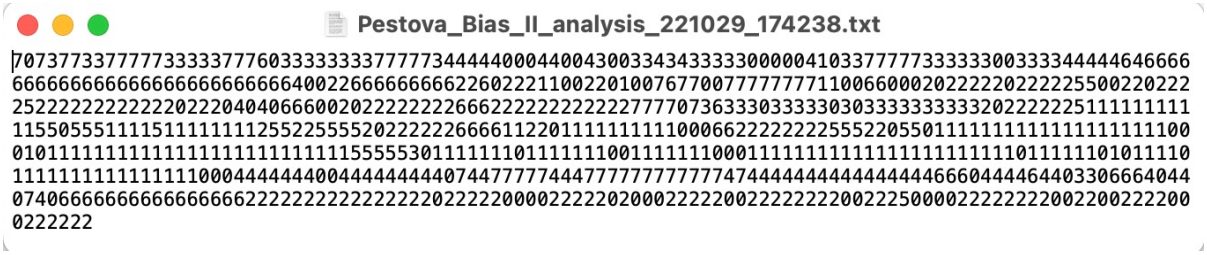

Figure 1. Analysis of a "training round" of *Bias II*, serving as a training example for the Recurrent Neural Network (RNN).

An example of the data fed into the RNN in *Bias II* is shown in *Figure 1*. A denotative reading of these data affirms that each number in this sequence corresponds to 1 sec of audio and represents one of the seven clusters in the score. Concretely, numbers from 1 through 7 represent the seven timbral clusters and 0 represents background noise/silence. The sequence as a whole represents the analysis of a single performance of the piece. The denotative reading of the dataset also states that these data were generated by a feedforward NN performing a classification task and using MFCCs as an input. Finally, for each second of audio only the predominant cluster is recorded.

Shifting perspectives from a denotative to a connotative reading reveals the aesthetic judgments driving the data-making process, as well as the broader cultural and music-historical context within which these judgments are situated. Concretely, the numbers in the data represent timbral clusters, that is, loosely defined clusters of sounds grouped together based on their subjectively perceived timbral similarity. Given that what is represented by this sequence of numbers is musical form, this suggests an understanding of form as a function of timbre, rather than pitch or harmony. Implied in the semantics of this dataset is a musical aesthetic that prioritizes timbre as an articulator of musical form and explores the extended capabilities of the piano through string preparations and inside-piano playing techniques. This information places the work within specific music-historical and

aesthetic contexts, for instance the tradition of the prepared piano and "sound-based" as opposed to "note-based" (Landy 2019) approaches to music. Crystallized in the data are socio-cultural (Bates et al. 2016) – and, in this case, aesthetic – values that powerfully frame and delimit what is represented from what is left out.

Not only the semantics of the data, but also their quantization is indicative of the crystallization of aesthetic values in the dataset. For instance, in the example shown above, time quantization has important aesthetic implications for the work, as it defines the shortest possible interval between two consecutive musical changes proposed by the interactive music system: 1 sec. The decision to quantize the data in this way was an aesthetic one, aiming to introduce some constraints on musical form by preventing the interactive music system from proposing what I (subjectively) perceived as "too frequent" musical changes.

The fact that the data fed into the RNN are themselves generated by another ML algorithm takes on additional meaning in the context of a connotative reading of the dataset. As ML algorithms are prone to errors, the data-making process itself can be assumed to be flawed and imperfect. Given that the performances used as training examples for the RNN could have been manually analyzed, this decision also suggests an aesthetic motivation: to explore the aesthetic potential of the materiality and limitations of the ML algorithms used in the piece. The classification errors made by the feedforward NN influence what the RNN learns and predicts, leading to emergent, unpredictable behaviors.

In addition to aesthetic judgments and decisions, a connotative reading of the data sheds light on the distribution of musical labor and the composer-performer relationship in the piece. Crystallized in the training data are the performers' individual interpretative strategies and choices. The "training rounds" (run-throughs of the piece during which training data are collected) destabilize the composition-interpretation binary: the performers' interpretations become acts of co-

authorship, influencing the future behavior of the interactive music system – albeit in non-linear and unpredictable ways.

This provides further insights into my aesthetic language, revealing an emphasis on interpretative individuality and multiplicity, an aestheticization of the process of interaction, and a displacement of the locus of the aesthetic from the sonic to the socio-sonic, which is to say, "the manifestation and materialization of social (e.g., composer-performer, human-technology etc.) relations in and through sound" (Gioti 2021b, p. 7). Rather than in sound alone, the aesthetic of this music lies in the social relations it materializes and is embedded within, and in the way these relations take form and are translated into real-time sonic interactions (Gioti 2021b). This emphasis on social and relational aspects of music-making is characteristic of "open work" practices, in which "every performance makes the work an actuality, but is itself only complementary to all possible other performances of the work" (Eco 1989, p. 15), evoking the practices of composers such as Cornelius Cardew, Pauline Oliveros, and Christian Wolff, among others. Yet, different performances of *Bias II* are not complementary and equivalent realizations of the work, but links in a chain of co-creative acts that both instantiate and rewrite the work, transcending the boundaries of the "open work," and making the work itself both a social and a musical process.

Equally important to what is represented in the data is what is left out of them, as illuminated by a deconstructive reading. In addition to the consequences of the "information loss" resulting from the quantization process described earlier, a deconstructive reading of the semantics of the dataset reveals that what is recorded in the training data is simply a path through the timbral topology of the score, leaving out a large number of alternative perspectives and framings, such as pitch, dynamics, rhythm, and/or density of musical events. Also left out is spectral information rendered redundant by the MIR tools used, through those tools' immanent processes of abstraction and reduction.

Finally, another process of demarcation in the data-making process concerns the medium of (audio) recording itself. Rendered invisible through the isolation of the aural from the embodied experience of the performance are the relationship between sound and the choreography of the performer's movements, the material conditions of sound production (such as the objects used for string preparation and extended playing techniques), and inaudible aspects of the performer's interaction with the instrument and interactive music system (notably moments in which the performer changes their mind shortly before performing a certain action, as a result of a musical change introduced by the computer music system). Through the mediation of recording, sound is stripped of its embodied, social, cultural, and experiential dimensions and reduced to a materially mediated representation of an acoustic phenomenon.

Data-Making as a Compositional Act

The material and semiotic dimensions of data discussed above underscore the temporal, spatial, social, cultural, and material conditions of the data creation process. While commonly understood as being "pre-analytical and pre-factual, that which exists prior to interpretation and argument" (Kitchin and Lauriault 2014), data – as argued earlier – are processual, situated, and contingent. They are created through processes of framing, translation, and abstraction and say as much about these processes as they do about the phenomena they represent. Crawford and Paglen (2021) and Denton et al. (2021) employ what they call an "archeology" and "genealogy" of data, respectively, to expose the values and politics on the basis of which image datasets are constructed. They "trace the provenance of skews and biases exhibited in working [AI] systems" (Crawford and Paglen 2021, p. 1114), and highlight the temporal dimension of datasets as "historically situated artifacts"

(Denton et al. 2021, p. 2). Their approaches denaturalize datasets by examining the political, social, and historical conditions of data-making processes.

Similarly, the two datasets used to train the feedforward NN and the RNN in *Bias II* are situated within music-historical contexts and aesthetic choices of the kind alluded to earlier. More specifically, enmeshed in the data are (from the broader to the narrower): broader music-historical and aesthetic contexts, my individual aesthetic language as shaped by and shaping these contexts, and more "narrow" compositional decisions, specific to this work, such as the decision to limit the maximum frequency of musical changes proposed by the interactive music system to one per second. Data-making in this context becomes a compositional act.

Along with such compositional decisions and their broader cultural and music-historical context, the mediation of mechanical, analog, and digital equipment (pianos, microphones, cables, interfaces, software plugins, and so on) also contributes to the contingent nature of the data-making process. Of course, exploring "the crucial importance of particular materiality" and "the impossibility of 'pure' transmission of a message from sender to receiver" (Impett 2021, p. 126) is neither new to composition and sound art, nor uniquely relevant to ML algorithms. In John Cage's 1977 composition *Inlets*, performers are instructed to tip seashells filled with water – a compositional strategy that relies on contingency in order to facilitate "improvisational activity without improvisational style" (McLaughlin 2021, p. 158). Composer Scott McLaughlin explores "material indeterminacy" as it relates to acoustic instruments, using "metastable states" – states that are only stable "under highly specific conditions" (McLaughlin 2021, p. 155) – as a means of introducing contingency and balancing the agencies of performer and instrument (McLaughlin 2021, p. 156). Similarly, Agostino Di Scipio's *Audible Ecosystemics* pieces explore the contingent materiality of feedback networks encompassing meticulously composed interactions between human agency, audio equipment, and the acoustic

environment (Di Scipio 2011). The *Audible Ecosystemics* pieces remain "*underdetermined* by the[ir] software components" in that "there is simply no way of determining from an examination of the software components in isolation what the details of a particular performance will be" (Green 2014, p. 62, original italics).

If mediation and contingency are not new themes in music composition, they have new significance in the context of compositional approaches encompassing ML. Unlike Di Scipio's carefully designed – i.e., *programmed* – parameter mappings, ML algorithms rely on "learned" input-output mappings, i.e., mathematical functions induced from the data. These introduce an additional layer of mediation – that of computational processes specific to ML (e.g., backpropagation), while delegating a higher degree of creative agency to software components of the assemblage and displacing the locus of the compositional process from explicitly designing (coding) relationships to influencing them in indirect and non-linear ways through processes of data-making. Data-making then becomes the main process through which the composer influences and interacts with the ML algorithm (Fiebrink and Sonami 2020, p. 239), as well as the process that stands between "the initial software construct with its non-specific, unrealised potential" and "the trained system that has constructed a model of its world and evolved an appropriate repertoire of responses" (Impett 2021, p. 126). In the case of the RNN used in *Bias II*, data perform another crucial conceptual function: data collected during performances of the piece modify the behavior of the interactive music system, challenging the distinction between instantiations and inscriptions of the work, and, therefore, composition and performance. Not only is the work "*underdetermined*" by its "software components" Green (2014, p. 62), but these components are themselves *underdetermined* and "in a constant state of becoming" (McLaughlin 2021, p. 158).

Conclusions

The questions guiding this research were: (1) how does ML shape aesthetic choices in distinctive compositional practices, and how is it shaped in turn by aesthetic commitments? (2) How can critical themes from relevant humanities scholarship inform artistic research in composition using ML? And (3) in what ways does the interaction with ML algorithms as part of the compositional process differ from that with other music technology tools? They inform the two compositional projects presented, and here we summarize salient points arising.

The most fundamental observation is that the ML algorithms employed in each compositional project, and how they act as components of the assemblage, are quite different. This points to a basic finding: about the heterogeneity of the systems within which ML operates, as well as the multiplicity of the category "machine learning" itself. Crucially, this identification of the heterogeneity of ML-informed compositional projects works against any tendency to universalize the nature and operations of ML algorithms in their distinctive compositional uses. Complementing this approach, by conceptualizing ML as a collaborator in assemblages of human and nonhuman actors, we resist attributing agency solely to human actors and depict the operations of ML as deeply relational.

In comparing how interaction with ML algorithms differs from that with other technological tools we are struck first by the continuities with non-ML-informed interactive music systems. Perhaps most obviously, ML shares features of material agency and engagement with technological systems in general (Pickering 1995), including those that do not use ML (e.g., interactive systems implemented in Max, cf. Snape and Born 2022).

However, our case studies also raise questions concerning differences that may be introduced into music-technology systems by ML. As we will suggest, this may be a matter of degree rather than kind. A first observation is that thanks to the wealth of parameters required in the ML data-making process, ML-informed

systems may tend to differ from non-ML systems in terms of the sheer density of actions and processes animated by human and nonhuman actors that together constitute the environment within which material engagement takes place. A second possible difference stems from ML's unique reliance on training data, which means that the timing of engagement with ML during the compositional process differs from that with other systems. It effectively multiplies the phases of creative decisions by human and nonhuman agents: from the performative labor crystalized in training sets, to the decisions involved in the data-making process, to the many later contingencies in the technological assemblage of live interaction. This suggests that these decisions may not produce consequences in immediate or continuous ways – as is often the case with non-ML interactive systems – but rather entail delays, iterations, and recursions in how they affect other elements of the assemblage. In turn this makes the process of engaging with ML more complex, non-linear, and unpredictably contingent on these other elements – yet this complexity can be stimulating and generative for the material engagement emerging through the creative process. A third possible difference turns on the distinction between programmed input-output mappings in non-ML systems and the learned mappings of ML. Composing with ML systems one may not have direct access to the mathematical functions learned by the algorithm but only to its inputs and outputs, which introduces additional forms of black-boxed mediation into such systems (Burrell 2016; Born et al. 2021); this requires a different approach to creative engagement – suggested by our auto-ethnographic accounts – than for a system to which the composer has explicit access to all internal parameters.

Our study also points to differences between creative and commercial or academic uses of ML with regard to the data-hungry nature of ML algorithms. The small datasets that artists tend to work with, as exemplified by our case studies, will often fall short of the kinds of datasets used in standard ML applications – yet this

very disjuncture can generate creative opportunities, for example, timbral, and spatial gestures that are a mismatch with training sources.

In both compositions described in this article, a guiding aesthetic sensibility permeates the assemblage – which is distributed technically, materially, socially, and temporally; yet the composer's aesthetic sensibility is likely to evolve in unforeseeable ways as an effect of the processual and distributed nature of the creative process we have described. Composing with ML is therefore a speculative act of interacting iteratively and recursively with processes that cannot be controlled or known but navigated, adapted, negotiated, and influenced, and with musical results that can only ever be described as emergent. Crucially, material engagement with ML is not limited to engaging with algorithms and data as part of the narrowly defined compositional and rehearsal process, but – as indicated above – extends to the conceptual phase of the creative process and the ways in which the perceived affordances of ML algorithms influence creative ideation. ML, then, becomes part of the composer's "background," affirming Malafouris's insight that "background" and material engagement are effectively inseparable.

At the same time, as we have shown, ML processes and datasets can be substantially shaped by distinctive aesthetic and conceptual commitments. A commitment to engaging with the materiality of ML and data, for instance, can lead to a questioning of common understandings of ML processes as closed tasks oriented to optimization, instead reframing them as open-ended experiments, serving purposes of aesthetic experimentation and imaginative critique. The ML algorithm in *Bias II* is an example of how an unconventional approach to data-making – a mismatch between the data used to train the machine listening algorithm and the inputs it is presented with during performance – can serve both as a source of sonically rich, emergent behaviors, and as a critique of prevalent approaches to data, highlighting data as contingent on aesthetic decisions and material mediation.

Similarly, in *Prestidigitation*, the absence of percussion in the training database of instrumental radiation patterns opened an opportunity for the emergence of novel, "unrealistic" spatial gestures. In both works ML is not simply a means to a predetermined goal (e.g., accurate prediction of different timbres or radiation patterns), but also the object of exploration and critique.

In a manner aligned with post-positivist empiricism, the theoretical frameworks drawn from relevant humanities scholarship both informed the work described here and were themselves extended in response to new insights gained, via auto-ethnography, from the two composers' artistic research. Einbond's work explored how material engagement with ML differs from engagement with other music technology tools. Gioti's work drew on critical data studies to reflect on the cultural and aesthetic dimensions of data and explored an additional – material – dimension, with particular relevance to machine listening. Our artistic research perspective played a crucial role in generating these insights, allowing us to engage in critique not from a theoretical distance but through material engagement with algorithms and data as part of the creative process. This perspective allows for a two-way movement between theory and artistic practice, the objective of which is not only the amendment and generation of theory, but also the transformation of practice, as no artistic practice exists in a theoretical vacuum and no theoretical interpretation of practice leaves it unaffected (Borgdorff 2006, p. 5). Despite the distinctive implications of our artistic research orientation – its positionality and first-person perspective – we contend that our findings have broader relevance for the Music AI artistic and research community, by drawing attention towards and calling for reflection upon immanent aspects of data-making and ML processes, including the inherently aesthetic nature of music data and the distinctive qualities of material engagement with ML.

Acknowledgments

This research was funded by the ERC advanced grant "MusAI - Music and Artificial Intelligence: Building Critical Interdisciplinary Studies" (European Research Council grant agreement no. 101019164, 2021-26).

References

Bailey, D. 1992. *Improvisation: Its Nature and Practice in Music.* London: British Library National Sound Archive.

Barrett, R. 2003. "DARK MATTER: Interview between Daryl Buckley and Richard Barrett." Available online at richardbarrettmusic.com/DARKMATTERinterview.html. Accessed February 2023.

Bates, J., Y. W. Lin, and P. Goodale. 2016. "Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows." *Big Data and Society* 3(2):1-12.

Baumer, E. P. S. 2015. "Reflective Informatics: Conceptual Dimensions for Designing Technologies of Reflection." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 585-594.

Borgdorff, H. 2006. "The Debate on Research in the Arts." In *Sensuous Knowledge: Focus on Artistic Research and Development*, Vol. 2. Bergen: Bergen National Academy of the Arts, pp. 7-28

Born, G. 1995. *Rationalizing Culture: IRCAM, Boulez, and the Institutionalization of the Musical Avant-Garde*. Berkeley, California: University of California Press.

Born, G. 2005. "On Musical Mediation: Ontology, Technology and Creativity." *Twentieth-Century Music* 2(1):7-36.

Born, G. 2010. "The Social and the Aesthetic: For a Post-Bourdieuian Theory of Cultural Production." *Cultural Sociology* 4(2):171-208.

Born, G. 2015. "Making Time: Temporality, History, and the Cultural Object." *New*

Literary History 46(3):361-386

Born, G. 2021a. "Artistic Research and Music Research." In A. Huber, et al., eds. *Knowing in Performing: Artistic Research in Music and the Performing Arts*. Bielefeld: transcript Verlag, pp. 35-50.

Born, G. 2021b. "On Genre, History, and Invention in the Analysis of Creative Processes in Music." In N. Donin, ed. *The Oxford Handbook of the Creative Process in Music*. Oxford: Oxford University Press.

Born, G. 2022. "The Dynamics of Pluralism in Contemporary Digital Art Music." In G. Born, ed. *Music and Digital Media: A Planetary Anthropology*. London: UCL Press, pp. 305-377.

Born, G., et al. 2021. "Artificial Intelligence, Music Recommendation, and the Curation of Culture." White paper. Toronto: Schwartz Reisman Institute for Technology and Society.

Born, G., and A. Barry. 2018. "Introduction: Music, Mediation Theories and Actor-Network Theory." *Contemporary Music Review* 37(5-6):443-487.

Bowker, G. C., and S. L. Star. 2000. *Sorting Things Out: Classification and Its Consequences.* Cambridge, Massachusetts: MIT Press.

Burrell, J. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data and Society* 3(1):1-12.

Clarke, E. F., and M. Doffman. 2017. "Introduction and Overview." In E. F. Clarke, and M. Doffman, eds. *Distributed Creativity: Collaboration and Improvisation in Contemporary Music*. Oxford: Oxford University Press, pp. 1-20.

Connor, S. 1997. "The Modern Auditory I." In R. Porter, ed. *Rewriting the Self: Histories from the Renaissance to the Present.* London: Routledge, pp. 203-223.

Crawford, K., and T. Paglen. 2021. "Excavating AI: The Politics of Images in Machine Learning Training Sets." *AI and Society* 36:1105–1116.

Dahlstedt, P. 2019. "Big Data and Creativity." *European Review* 27(3):411-439.

Davis, S. B., and P. Mermelstein. 1980. "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28(4):357-366.

DeLanda, M. 2006. *A New Philosophy of Society: Assemblage Theory and Social Complexity*. London: Continuum.

Deleuze, G. 1988. *Foucault*. London: Athlone Press.

Deleuze, G., and C. Parnet. 1987. *Dialogues*. London: Athlone Press.

Denton, E., et al. 2021. "On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet." *Big Data and Society* 8(2):1-14.

Di Scipio, A. 1998. "Questions Concerning Music Technology." *Angelaki: Journal of the Theoretical Humanities* 3(2):31-40.

Di Scipio, A. 2011. "Listening to Yourself through the Otherself: On Background Noise Study and Other Works." *Organised Sound* 16(2):97-108.

Donin, N., and C. Traube. 2016. "Tracking the Creative Process in Music: New Issues, New Methods." *Musicae Scientiae* 20(3):283-286.

Donin, N., ed. 2018. *The Oxford Handbook of Creative Process in Music*. Oxford: Oxford University Press.

Eco, U. 1989. *The Open Work*, trans. A. Cancogni. Cambridge, Massachusetts: Harvard University Press.

Einbond, A. 2013. "Subtractive Synthesis: Noise and Digital (Un)Creativity." In A. Cassidy, and A. Einbond, eds. *Noise In And As Music*. Huddersfield: University of Huddersfield Press, pp. 57-75.

Einbond, A. 2016. "Musique Instrumentale Concrète: Timbral Transcription in What the Blind See and Without Words." *Open Music Composer's Book Vol. 3*. Paris: Editions Delatour/IRCAM, pp. 155-171.

Einbond, A., et al. 2016. "Introducing CatOracle: Corpus-based Concatenative Improvisation with the Audio Oracle Algorithm." *Proceedings of the International*

Computer Music Conference, pp. 140-146.

Einbond, A., et al. 2021. "Instrumental Radiation Patterns as Models for Corpus-Based Spatial Sound Synthesis: Cosmologies for Piano and 3D Electronics." *Proceedings of the International Computer Music Conference*, pp. 148-153.

Fiebrink, R., and L. Sonami. 2020. "Reflections on Eight Years of Instrument Creation with Machine Learning." *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 237-242.

Findlay-Walsh, I. 2018. "Sonic Autoethnographies: Personal Listening as Compositional Context." *Organised Sound*, 23(1):121-130.

Françoise, J, et al. 2014. "Probabilistic Models for Designing Motion and Sound Relationships." *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 287-292.

Gioti, A. M. 2021a. "Artificial Intelligence for Music Composition." In: E. R. Miranda, ed. *Handbook of Artificial Intelligence for Music*. Springer, pp. 53-74.

Gioti, A. M. 2021b. "Agency and Distributed Creativity in Interactive Compositions." Dissertation, University of Music and Performing Arts Graz.

Green, O. 2014. "Audible Ecosystemics as Artefactual Assemblages: Thoughts on Making and Knowing Prompted by Practical Investigation of Di Scipio's Work." *Contemporary Music Review* 33(1):59-70.

Haraway, D. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14(3):575-599.

Harding, S. 1992. "After the Neutrality Ideal: Science, Politics, and 'Strong Objectivity.'" *Social Research* 59(3):567-87.

Holzapfel, A., B. L. Sturm, and M. Coeckelbergh. 2018. "Ethical Dimensions of Music Information Retrieval Technology." *Transactions of the International Society for Music Information Retrieval* 1(1):44-55.

Impett, J. 2000. "Situating the Invention in Interactive Music." *Organised Sound*

5(1):27-34.

Impett, J. 2021. "Dissociation and Interference in Composers' Stories about Music: The Renewal of Musical Discourse." In J. Impett, ed. *Sound Work: Composition as Critical Technical Practice*, Leuven University Press, pp. 117-135.

Jones, A. M., and N. Boivin. 2010. "The Malice of Inanimate Objects: Material Agency." In D. Hicks, and M. C. Beaudry, eds. *The Oxford Handbook of Material Culture Studies*. Oxford: Oxford University Press, pp. 333-351.

Kitchin, R. 2014. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data and Society* 1(1):1-12.

Kitchin, R., and T. P. Lauriault. 2014. "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." In J. Thatcher, J. Eckert, and A. Shears, eds. *Thinking Big Data in Geography: New Regimes, New Research*. Lincoln, Nebraska: University of Nebraska Press.

Knappett, C., and L. Malafouris, eds. 2008. *Material Agency: Towards a Non-Anthropocentric Approach*. Springer, pp. ix-xix.

Labelle, B. 2006. *Background Noise: Perspectives on Sound Art*. London: Continuum.

Landy, L. 2019. "Re-Composing Sounds ... and Other Things." *Organised Sound* 24(2):130-138.

Lachenmann, H. 1996. *Musik als existentielle Erfahrung: Schriften 1966-1995*. J. Häusler, ed. Wiesbaden: Breitkopf and Härtel.

Lewis, G. E. 1999. "Interacting with latter-day musical automata." *Contemporary Music* Review 18(3):99-112.

Magnusson, T. 2009. "Of Epistemic Tools: Musical Instruments as Cognitive Extensions." *Organised Sound* 14(2):168-176.

Magnusson, T. 2011. "Confessions of a Live Coder." *Proceedings of the International Computer Music Conference*, pp. 609-616.

Malafouris, L. 2004. "The Cognitive Basis of Material Engagement: Where Brain,

Body and Culture Conflate." In E. DeMarrais, C. Gosden, and C. Renfrew, eds. *Rethinking Materiality: The Engagement of Mind with the Material World*. Cambridge: McDonald Institute for Archaeological Research, pp. 53-62.

Malafouris, L. 2008. "At the Potter's Wheel: An Argument for Material Agency." In C. Knappett, and L. Malafouris, eds. *Material Agency: Towards a Non-Anthropocentric Approach*. Springer, pp. 19-36.

Malafouris, L. 2013. *How Things Shape the Mind: A Theory of Material Engagement*. Cambridge, Massachusetts: MIT Press.

Markham, A. N. 2013. "Undermining 'Data': A Critical Examination of a Core Term in Scientific Inquiry." *First Monday* 18(10). Available online at firstmonday.org/ojs/index.php/fm/article/download/4868/3749?inline=1. Accessed February 2023.

McLaughlin, S. 2021. "The Impossibility of Material Foundations." In J. Impett, ed. *Sound Work: Composition as Critical Technical Practice*, Leuven: Leuven University Press, pp. 153-169.

McPherson, A., and K. Tahıroğlu. 2020. "Idiomatic Patterns and Aesthetic Influence in Computer Music Languages." *Organised Sound* 25(1):53-63.

Pardue, L., and S. M. A. Bin. 2022. "The Other Hegemony: Effects of Software Development Culture on Music Software, and What We Can Do About it." *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Peeters, G. 2004. "A Large Set of Audio Features for Sound Description (Similarity and Classification) in the Cuidado Project." Technical report, IRCAM.

Pickering, A. 1995. *The Mangle of Practice: Time, Agency, and Science*. Chicago, Illinois: University of Chicago Press.

Pickering, A. 2010. "Material Culture and the Dance of Agency." In D. Hicks, and M. C. Beaudry, eds. *The Oxford Handbook of Material Culture Studies*. Oxford: Oxford University Press, pp. 191-208.

Poirier, L. 2021. “Reading Datasets: Strategies for Interpreting the Politics of Data Signification.” *Big Data and Society* 8(2):1-19.

Schnell, N., et al. 2009. “MuBu and Friends – Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP.” *Proceedings of the International Computer Music Conference*, pp. 423-426.

Schön, D. A. 1983. *The Reflective Practitioner: How Professionals Think in Action*. New York, New York: Basic Books.

Schwarz, D. 2007. “Corpus-based Concatenative Synthesis.” *IEEE Signal Processing Magazine* 24(2):92-104.

Searle, J. R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Snape, J., and G. Born. 2022. “Max, music software and the mutual mediation of aesthetics and digital technologies.” In G. Born, ed. *Music and Digital Media: A Planetary Anthropology*. London: UCL Press, pp. 220-266.

Surges, G., and S. Dubnov. 2013. “Feature Selection and Composition Using PyOracle.” *Proceedings of the Artificial Intelligence and Interactive Digital Entertainment Conference*, pp. 114-121.

Vis, F. 2013. “A Critical Reflection on Big Data: Considering APIs, Researchers and Tools as Data Makers.” *First Monday* 18(10). Available online at firstmonday.org/ojs/index.php/fm/article/view/4878/3755. Accessed February 2023.