Agency and Distributed Creativity in Interactive Compositions

Dissertation

Artemi-Maria Gioti

University of Music and Performing Arts Graz

Artistic Doctoral School

Composition and Music Theory (UV 795 500)

Supervisors:

Marko Ciciliani (University of Music and Performing Arts Graz)

Gerhard Eckel (University of Music and Performing Arts Graz)

Simon Emmerson (De Montfort University, Leicester)

Agostino Di Scipio (Conservatory of Music “Alfredo Casella”, L'Aquila)

May 2021

Abstract

This dissertation explores the aesthetic and composition-technical implications of *interactive compositions*, i.e., musical works that involve mutual real-time adaptation between musicians and interactive computer music systems. The latter are conceived as artificial "cognizers" capable of collecting and processing auditory information and acting both in response to human actions and independently of them, as a result of autonomous generative processes. The musical works described in this dissertation are rooted in a distributed notion of creativity that encompasses both human (composer and performers) and non-human actors (computer music system) and is manifested in the high degree of interpretative freedom and machine autonomy involved in them.

The use of Artificial Intelligence (AI) in this research extends beyond conventional applications of Machine Learning in machine listening tasks, to subversive and critical approaches to Machine Learning, as well as an exploration of the potential of AI as an *ideation tool*, i.e., its potential to shape musical thinking, by opening up new technical and conceptual possibilities.

The premise of the compositional practice described in this dissertation lies in a notion of work identity that encompasses, rather than excludes, diverse musical outcomes. As a result, the relationship between musical authorship and interpretative freedom lies at the center of this research. This relationship was explored using a series of experimentation methods based on guided improvisation tasks and conducted with the help of the musicians, and ethnographically informed data collection methods, such as observation, questionnaires and interviews with the musicians. The interpretation of qualitative data collected through these methods provided valuable insights into the works described in this dissertation and played a decisive role in their development.

Kurzfassung

Diese Dissertation untersucht ästhetische und kompositionstechnische Implikationen interaktiver Kompositionen. Dabei handelt es sich um musikalische Werke, die auf wechselseitigen Adaptionen zwischen Musiker_innen und interaktiven Computermusiksystemen in Echtzeit basieren. Diese Computermusiksysteme sind als künstliche „Cognizers“ konzipiert, die auditive Informationen sammeln und verarbeiten und Klangmaterial sowohl als Reaktion auf menschliche Aktionen als auch auf Basis autonomer Prozesse generieren. Die beschriebenen musikalischen Werke basieren auf einer verteilten Kreativität, die sowohl menschliche (Komponistin, Musiker_innen) als auch nichtmenschliche Akteure (Computermusiksystem) umfasst und mit einem höheren Grad an Interpretationsfreiheit und maschineller Autonomie verbunden ist.

Die Verwendung von Künstlicher Intelligenz (KI) geht in der vorliegenden Forschung über herkömmliche Anwendungen maschinellen Lernens hinaus und schließt subversive und kritische Zugangsweisen zu KI ein. Das Ziel dabei ist es, das Potenzial von KI als konzeptuelles Werkzeug zu erforschen. Verstanden als solches Werkzeug bietet KI Potenziale, kompositorisches Denken zu prägen, indem sie neue technische und konzeptionelle Möglichkeiten eröffnet.

Die Prämisse, die der in dieser Dissertation konzipierten kompositorischen Praxis zugrunde liegt, ist ein Verständnis von Werkidentität, das unterschiedliche musikalische Resultate eher fördert, als ausschließt. Somit steht das Verhältnis zwischen musikalischer Autorschaft und Interpretationsfreiheit im Mittelpunkt der vorliegenden Forschung. Dieses Verhältnis wurde mit einer Reihe von Experimentiermethoden erforscht, die auf angeleiteten Improvisationsaufgaben basierten und mit der Unterstützung von Musiker_innen durchgeführt wurden. Außerdem wurden ethnografische Methoden wie Beobachtung, Fragebögen und Interviews mit den Musiker_innen zu Datenerhebung eingesetzt. Die Interpretation der gesammelten qualitativen Daten lieferte wertvolle Erkenntnisse über die in dieser Dissertation beschriebenen Werke und spielte eine entscheidende Rolle bei deren Entwicklung.

Acknowledgments

I am extremely thankful to all the musicians who participated in this research for their valuable contribution to this work: Joel Diegert, Manuel Alcaraz Clemente, Margerethe Maierhofer-Lischka, Patrick Skrilecz, Nikolaus Feinig, Florian Müller and Szilárd Benes. This work wouldn't have been possible without them.

I would also like to express my immense gratitude to all my supervisors, who accompanied me and guided me in my research journey. To Marko Ciciliani, for his mentorship, encouragement and invaluable advice over the last five years, Gerhard Eckel, without whose support I wouldn't have embarked upon this journey, Simon Emmerson, for many fruitful and thought-provoking discussions, and Agostino Di Scipio for his insights and interest in this work.

I am grateful to all my colleagues at the IEM for creating an exciting and pleasant work environment and everyone I collaborated with in the context of the *Inter_agency* project: Luc Döbereiner, Alisa Kobzar, Jack Walker and Eduardo Miranda. I am extremely thankful to my former IEM-colleague Visda Goudarzi for involving me in the *COLLAB* project and for her support and mentorship. A special thanks goes to Andreas Pirchner for many interesting discussions and for proofreading the German abstract of this thesis.

Last but not least, I would like to thank my parents Katerina and Dimitris and my brother Christoforos for their love and continuous support, and my partner Aris for his patience, encouragement and support all these years and particularly the last few months of writing this thesis.

This research was funded by the Austrian Science Fund (FWF): AR 483-G24.

Declaration

Some of the work contained within this thesis has appeared in the author's following publications:

Gioti, Artemi-Maria. 2021. "A Compositional Exploration of Computational Aesthetic Evaluation and AI Bias." In *Proceedings of the 2021 International Conference on New Interfaces for Musical Expression*.

Gioti, Artemi-Maria. 2021. "*Converge/Diverge*: Collaborative Emergence in a Composition for Piano, Double Bass and Interactive Music System." In *Proceedings of the 2021 International Computer Music Conference*.

Gioti, Artemi-Maria. 2020. "From Artificial to Extended Intelligence in Music Composition." *Organised Sound* 25 (1):25–32. https://doi.org/10.1017/S1355771819000438.

Gioti, Artemi-Maria. 2019. "Imitation Game: An Interactive Composition for Human and Robotic Percussionist." In *Proceedings of the 2019 International Computer Music Conference*.

Gioti, Artemi-Maria. 2018. "Neurons: An Interactive Composition Using a Neural Network for Recognition of Playing Techniques." In *Proceedings of the 2018 Musical Metacreation Workshop*.

Gioti, Artemi-Maria. 2017. "Machine Listening in Interactive Music Systems: Current State and Future Directions." In *Proceedings of the 2017 International Computer Music Conference*, 216–20.

Table of Contents

1 Introduction

1.1 Scope and Research Questions

This research addresses the desideratum of human-computer symbiosis in composed electro-instrumental music and aims to enhance human-machine communication in compositions for acoustic instruments and electronics by incorporating Artificial Intelligence (AI) in them. The interaction model that serves as a frame of reference for this research is Rowe's (1993) 'player' paradigm – as opposed to the 'instrument' paradigm; a model in which the musician and the computer are co-actors in a reciprocal interaction (chap. 1). In the player paradigm, the software agent perceives human actions through machine listening and acts both in response to them and according to internal generative processes. In particular, the type of human-computer interaction that this research explores is:

(1) *sound-based*, i.e., based exclusively on (human and machine) listening[1],

(2) *composed* – as opposed to improvised – and

(3) *reciprocal*, that is, interaction in its literal sense: a process of mutual adaptation between the musician(s) and the software agent.

In this context, an agent is understood as an entity able to sense its environment and act in response to it, as well as autonomously (Wooldridge and Jennings 1996). The concept of autonomy points towards two of Young's (2008) attributes of 'live algorithms':

[1] This means that the computer collects and interprets audio data; other forms of data (e.g., video, MIDI data etc.) are effectively excluded.

(1) *empowerment*: the ability to make "decisions" that influence future actions, and

(2) *opacity*: the avoidance of linear input-output mappings and their replacement through complex generative processes.

The delegation of creative responsibility to the musicians and computer music system in the form of real-time decision-making during the performance destabilizes the dualism between composition and performance and challenges traditional notions of authorship and ontologies of the musical work. Rebalancing the relationship between authorship and interpretation, re-examining the locus of work identity and developing new compositional methods and strategies to deal with these conceptual shifts are some of the foci of this research.

This research therefore addresses a primary and secondary research question:

> *How can the focus of the compositional process be shifted from composing sounds to composing sonic interactions? And what are the implications of this shift for musical authorship and the work-concept?*

1.2 Interaction and Electronic Music Discourse

Interaction is one of the most broadly discussed, yet ambiguous terms in the discourse of electronic music (Paine 2002; Di Scipio 2003). Rowe's (1993) definition of interactive music systems as systems that adapt their behavior to musical input has been criticized for considering the performer's reaction as secondary and therefore viewing interaction as a unidirectional, rather than a reciprocal process (Drummond 2009). Similar criticism can be made of Chadabe's (1984) definition of interactive composing systems. Paine (2002) claims that the term interactivity is widely abused and joins Bongers (2000) in arguing that most "interactive systems" are in fact *reactive*, since they lack cognition.

Along with semantic ambiguities, these contradictory views on interaction are indicative of the wide spectrum of interaction models employed in live electronic music, ranging and blurring the boundaries between instrumental interactions, oriented towards the re-establishment of 'physio-sonic' (Brent 2012) relationships, and interactions with intelligent agents fulfilling Wooldridge and Jennings' (1996) criteria of pro-activeness and autonomy. The blurring of the line between instrument/interface and agent is evidenced in the terminology used to describe Digital Musical Instruments (DMIs) (Spiegel 1992). Gurevich and Fyans (2011) propose the term 'Digital Musical Interactions', instead of Digital Musical Instruments, and question the instrumentality of performer-DMI interactions, while Bown, Eldridge and McCormack (2009) use the term 'behavioural objects' to describe musical software that displays autonomous behavior. Magnusson (2009, 2019, chap. 3) argues that Digital Musical Instruments are as much extensions of the body, as they are of human cognition and refers to them as 'epistemic tools', i.e., 'systems of knowledge' with high 'symbolic pertinence' (Magnusson 2009, 168).

Rowe's *player paradigm* has so far been employed mainly by human-computer improvisation systems based on machine listening (Rowe 1999; Lewis 2000; Thom 2000; Pachet 2002; Bakht and Barlow 2009; Lévy, Bloch, and Assayag 2012; Hsu 2010; Collins 2011a; Young 2008; Leffue and Kestler 2016; Lepri 2016; Banerji 2016; Smith and Deal 2014; Van Nort 2009; Yee-King 2011; Linson et al. 2015). Interactive musical robots such as those developed by Weinberg and Driscoll (2006), Singer et al. (2003) and Jordá (2002) fall into the same interaction paradigm. Most of these systems are designed exclusively for human-computer improvisation, while fewer have been used in partly composed and partly improvised music (Rowe 1999; Bakht and Barlow 2009; Young 2008; Smith and Deal 2014). In the latter case, the focus in relevant publications still remains on technical aspects of the systems, while very little information is given regarding the compositions themselves.

This research aims to bridge this gap by exploring intelligent agent-based Human-Computer Interaction from a compositional perspective. What distinguishes *interactive compositions* from interactive improvisation systems are idiosyncratic, composition-specific *interaction scenarios*, delineated both by the interaction affordances of the computer music system and explicit performance instructions (i.e., a score), as opposed to the idiom-specific interaction capabilities of interactive improvisation systems and the improvisatory nature of the musical interactions they afford.

Admittedly, both human-computer improvisation systems (e.g., Lewis 2000) and interactive compositions challenge the composition/improvisation binary and can be conceptualized much more effectively with respect to a *composition-improvisation spectrum*. Interactive compositions clearly inhabit the space near the composition end of this spectrum. '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, 15) provide the broader aesthetic context for this research. These encompass a wide range of practices, from Earle Brown's, Cornelius Cardew's, Mauricio Kagel's and John Cage's diverse approaches to graphic notation (Hope and Vickery 2011), John Zorn's game pieces (Nesterenko 2017) and Pauline Oliveros' text scores (e.g., Oliveros 1974) to more recent approaches based on the integration of improvised musical actions in composed music by composers such as Cat Hope (2017), Richard Barrett (Barrett 2014) and Liza Lim (Clarke, Doffman, and Lim 2013).

1.3 Main Concepts and Terminology

In the following, some key terms relating to the interaction concept employed by this research are explained and defined with respect to their meaning and use within this dissertation. These are not meant as universal definitions of these concepts, but as clarifications of how they are to be understood in the following chapters of this thesis.

- *Interactive music system*: a computer music system capable of collecting and interpreting information from its acoustic environment and acting both in response to human actions and independently of them, as a result of autonomous generative processes.

- *Affordance*: first introduced by Gibson (1979, chap. 8), this term is used to refer to the interaction potentialities that a material or immaterial entity (such as an algorithm) affords, enables or privileges. In the next chapters, this term is used mainly to refer to the interaction affordances of various interactive music systems, i.e., what they *afford* musicians in terms of interaction. In this context, a distinction is often made between intended and 'perceived' (Norman 2013, 13) interaction affordances (i.e., what the designer of the system intended vs. what the users – or, in this case, the musicians – perceived). However, the term "affordance" is also used in other contexts (e.g., compositional affordance) and is broadly understood as being synonymous with potentiality.

- *Interaction scenario*: a sonic interaction concept defined by idiosyncratic interaction affordances and performance instructions.[2]

- *Interactive composition/interactive musical work*: a musical work that involves mutual real-time adaptation between human performers and an interactive computer music system, the design of which is idiosyncratic and composition-specific. An interactive composition might involve one or more interaction scenarios, each entailing distinct interaction affordances and performance instructions. Interactive musical works are generally associated with a higher degree of interpretative freedom than that involved in determinate, thoroughly notated works. This freedom can be manifested in the form of an open musical form, partially or ambiguously notated musical actions, improvised musical actions, text instructions etc. These compositional strategies allow musicians to make decisions in real-time and adapt to their human and virtual co-players' actions.

- *Performance instructions*: any type and form of instruction given to the performers. These can include different types of musical notation (e.g., metric or proportionate notation, determinate or indeterminate pitch notation, graphic notation etc.), as well as text instructions.

- *Action spaces*: the spaces of possibilities available to the performers in different interaction scenarios. These are delineated by the performance instructions and the interaction affordances of the computer music system and can include composed, partially composed or improvised musical actions.

[2] This definition differs from Nika, Chemillier and Assayag's (2017) definition of the term as a predefined temporal structure used to guide human-computer improvisation.

1.4 Interactive Compositions: Aesthetic and Ontological Implications

The interaction scenarios involved in an interactive composition can vary with respect to the interaction affordances of the computer music system (e.g., what the system listens to or listens *for* in the musician's input and how it responds to it) and the performance instructions given to the musicians. An obvious implication of this premise is that *composition*, both as a concept and practice, is effectively expanded to include the design of musical agents and their behaviors. A second implication is that the musical text (i.e., the score) is understood as having an evocative, rather than a directive function. It delineates a space of action for the musicians to explore, but does not – at least for the most part – describe concrete structures of sounds. Both action spaces and interaction affordances are *potentials for musical action,* the concretization of which into sound structures is highly dependent on the musicians' individual interpretative choices and their real-time interaction with the software agent.

One of the most consequential aesthetic implications of this compositional approach is its emphasis on *interpretative individuality and multiplicity*. Different instances (i.e., performances) of the work can vary significantly with respect to sound material and/or musical form. Spontaneous decision-making during the performance and deliberate interpretative choices can lead to varied musical outcomes, even across performances by the same musician(s). Aside from interpretative multiplicity, which can be a component of compositional practices that do not involve a division of musical labor between composer and performer, such as 'comprovisation' (Dudas 2010) or 'interactive composing' (Chadabe 1984), interactive compositions pursue an additional objective: that of *interpretative individuality*. In interactive compositions, interpretative choices can shape the performance in a decisive way, reflecting the musician's unique interpretative approach. The diverse interpretations of the work by different performers are constitutive of its identity and ontological status.

Related to the objectives of interpretative multiplicity and individuality is another aesthetic implication of interactive compositions: their prioritization of *process over product* and *ephemerality over permanence*. The focus of this compositional approach does not lie in concrete sound structures, but rather in the interaction spaces within which they emerge. Inherent to any interaction is the concept of *ephemerality*: interactions are momentary, fleeting and specific to the actors involved in them and the context within which they take place.

Bourriaud (2002) defines 'relational art' as 'an art taking as its theoretical horizon the realm of human interactions and its social context, rather that the assertion of an independent and private symbolic context' (14). The affinity of the compositional approach described here to relational aesthetics lies in its aestheticization of the sociosonic domain. The term *sociosonic* here refers to the manifestation and materialization of social (e.g., composer-performer, human-technology etc.) relations in and through sound.[3]

Consequently, this approach attributes aesthetic value not only to the musical outcome of the interaction between human and non-human agents, but also the *process of interaction* itself. This includes processes of real-time decision-making, the negotiation of intentions among actors, as well as the dynamics of their interaction (e.g., who is following and who is leading). These processes and their perception are rarely limited to the aural domain. The way a musician interacts with their instrument and other objects (e.g., mallets or objects used for instrument preparation) could be as indicative of processes of adaptation and in-the-moment decision-making – i.e., *interaction* – as is sound itself. Visual communication among musicians is another integral part of musical interaction. Musical interaction and its perception rely as much on aural information, as they do on visual cues. The attribution of aesthetic value to *interaction as a process* in the context of the compositional approach described in this dissertation indicates a non-reductionist understanding of musical performance as a multisensory lived experience. As audio recordings alone would fail to convey the richness of the interactions taking place during the performance, the compositions described in this dissertation were documented exclusively in the form of video recordings.

The web of interactions involved in an interactive composition is not limited to the immediate interactions taking place as part of the performance, but extends to a meta-level of interactions between composer and performer, as well as among different

---

[3] Rennie (2014) uses the term in a different context, to refer to artistic practices that view recorded sound – particularly field recordings – as embedded within a socio-cultural context, rather than strictly in spectromorphological terms.

performers. These interactions are technologically mediated, asynchronous and remote. The interaction between composer and performer revolves around the negotiation between compositional intent and interpretative freedom. The agency of the interactive music system is central to this mediated composer-performer interaction, as are the musicians' unique interpretative choices and strategies. The second kind of mediated interaction involved in an interactive composition is less obvious and involves the ways in which past performances can influence future interpretations of the work, facilitating a creative dialogue among its performers, mediated through recording technologies (Chapters 5 and 6).

The division of musical labor and the delegation of creative responsibility to the computer music system in this context are tied with an understanding of the musical work as the product of a distributed human-computer and human-human co-creativity. The social nature of creativity in general (Csikszentmihalyi 2014) and musical creativity in particular is broadly acknowledged, as is the role of material agency in creative processes. Impett (2000) considers the musical work as an activity that is 'distributed in space, technology, society and time' (27). Bown (2015) argues that 'all human creativity occurs in the context of networks of mutual influence' (17) and that cultural artifacts are produced by networks of interaction involving human and non-human actors. Brown (2016) proposes an understanding of creative acts as agency networks that encompass 'humans, tools, culture and environment' (140). Creative relationships within these networks are symmetrical with respect to influence, but asymmetrical with respect to contribution; for example, tools influence creative decisions, even though they might exhibit weaker agency than human actors.

Interactive compositions epitomize such notions of creativity. In them, creative responsibility is shared among the composer, the performers and the computer music system and is distributed in time and across the boundaries of the composition/performance and composition/improvisation binaries. Every instance (i.e., performance) of an interactive work is the product of sociosonic interactions that are ephemeral and specific to the actors involved in them. Admittedly, every musical performance is ephemeral; even performances of determinate and thoroughly notated works are never identical to one another. However, interactive musical works aim to destabilize the dualism between composition and performance, by delegating creative decisions traditionally belonging to the realm of composition to the performers and computer music system. Far from trying to eliminate the boundaries between composition and performance, this compositional practice seeks to redraw them, by drifting away from a directive and towards a co-creative relation between them.

The objective of rebalancing the relation between composition and performance has not only practical implications, relating to the compositional process and the techniques and methods employed in it, but also ontological ones. The ontology of the interactive musical work is linked to an understanding of work identity as constituted of interaction affordances and spaces of sonic possibilities, rather than concrete structures of sounds. In an interactive composition, sound material and musical form can vary significantly from one performance to another. The compositions described in this dissertation are based on a *dynamic form* (i.e., a form that is shaped through the real-time interaction between the musicians and the computer music system) and incorporate both composed and improvised musical actions. In such an approach, the locus of work identity shifts from the temporal organization of sound material to the interaction affordances and action spaces within which musical actions emerge.

The focus of interactive musical works on interpretative multiplicity and distributed notions of creativity suggests that both their ontological status and aesthetic intent lie in a sociosonic realm, rather than a purely sonic one. In line with that premise, in the research presented here, Artificial Intelligence is explored for its potential to expand the space of compositional possibilities, by enabling new types of sociosonic relations. Far from being a purely technical exercise, equipping computer music systems with machine listening capabilities, such as instrument and playing technique recognition (Chapters 3 and 4), or even aesthetic preferences (Chapter 7), aims to expand composition and performance practices and redraw the boundaries between composition and improvisation, composition and interpretation and human and machine agency. The role of machine intelligence in this approach extends beyond any agency that the computer music system might have during the performance to the various ways in which the affordances and specificities of AI algorithms can shape compositional thinking and influence creative ideation, i.e., their role as conceptual tools.

1.5 Artificial Cognition, Agency and the Posthuman Turn

The fact that machine intelligence differs from human intelligence is widely acknowledged. Yet, comparisons between human and machine intelligence are often central to the argumentation of AI critics. For instance, Searle's famous *Chinese Room* argument underscores the difference between using syntactic rules to manipulate symbols and comprehending the meaning of these symbols, pointing out that computers are only capable of the former. In this thought experiment, Searle sits in a room, receives questions in Chinese and responds by combining Chinese symbols according to a rule book (Searle 2004, 62–64). While to people outside the room it might appear as if Searle can speak Chinese, similarly to a computer program, he is simply following instructions to manipulate symbols the meaning of which he does not understand.

Searle's Chinese Room argument is directed against 'Strong AI' or 'computer functionalism', a theory that views mental states as computational states of the brain and likens the brain to a computer and the mind to a set of computer programs. Importantly, Searle's notion of 'Strong AI' is a theoretical construct that bears no resemblance to real-world applications of AI: according to his definition, 'Strong AI' aims to *create* a mind, as opposed to 'Weak AI', which aims to study the mind by simulating it (Searle 2004, 43–46). Searle's critique is therefore not a critique of AI *per se*, but rather a critique of *functionalism*: a theory according to which mental states are defined as functions and in terms of their causal relations to external stimuli and other mental states (43).

Dreyfus (1992), a prominent AI critic, claims that the conviction that artificial reason is possible is based on three false philosophical assumptions: a 'psychological assumption' that the human mind can be viewed as a machine that processes information following heuristic rules, an 'epistemological assumption' that all knowledge can be formalized in terms of such rules and an 'ontological assumption' that everything that is essential to intelligent behavior must be analyzable in terms of context-free determinate components. His argument is based on the fact that these assumptions fail to take into account the embodied and situated nature of human intelligence (i.e., the ability to judge which facts are relevant and essential in a given situation) and its dependence on indeterminate human needs and goals. According to Dreyfus, the inability of computers to deal with context suggests that they will never be capable of 'nonformal behavior', the type of intelligence involved in solving open-structured problems, which require identifying relevant facts and necessary actions, rather than appealing to rules.

Dreyfus' critique of AI is a critique of Good Old-Fashioned AI (GOFAI), also known as symbolic AI, an approach that dominated the first few decades of AI research and was based on expert rule-based systems. Yet, while some of his predictions were later disproved by developments in machine learning (e.g., recent successes in pattern recognition), his critique regarding the inability of computers to deal with context and open-ended tasks remains largely relevant today (Chapter 2).

Still, it is important to note that Dreyfus' definition of intelligence is a rather narrow one: for him "intelligence" is synonymous with human-level general intelligence. The question he poses is therefore not whether computers can be intelligent, but whether they can be *humanly intelligent*. While Searle and Dreyfus propose valid arguments and point out some important differences between human and machine intelligence, their arguments are largely rooted in anthropocentric bias, effectively equating intelligence with human intelligence.

Hayles (2017) offers a much more nuanced understanding of intelligence that encompasses both human and non-human cognition. In an effort to break away from anthropocentric views of cognition, she explicitly avoids using the term "intelligence" for non-human cognitions and considers *non-conscious cognition*, i.e., cognitive processes that are inaccessible yet essential to consciousness, as the link between human and non-human cognitions.

She proposes a definition of cognition that applies not only to humans and other life forms, but also to technical systems. According to Hayles (2017), cognition is a 'process that interprets information within contexts that connect it with meaning' (22). This definition contains three key concepts: *process*, *interpretation* and *context*. For Hayles, cognition is a *process*, rather than an attribute, as intelligence might be considered to be. This process involves *interpretation* of information, a concept that implies a choice. Choice in this context does not mean free will, but 'programmatic decisions' (25) that can be as simple as a binary choice between zero and one. Finally, interpretation and meaning are not context-agnostic, but specific to a certain situation.

Hayles (2017) acknowledges that human and technical cognitions have distinctive cognitive capacities (e.g., speed and computational intensity vs. empathic abilities and an 'encompassing world horizon' (140)), but views human and technical cognition as components of larger 'cognitive assemblages' that include material agents and forces. Regarding the relationship between human and technical congnizers within these assemblages Hayles comments:

> It is likely […] that the evolutionary development of technical cognizers will take a different path from that of Homo sapiens. Their trajectory will not run through consciousness but rather through more intensive and pervasive interconnections with other nonconscious cognizers. In a sense, they do not require consciousness for their operations, because they are already in recursive loops with human consciousness. […] It is now apparent that humans and technical systems are engaged in complex symbiotic relationships, in which each symbiont brings characteristic advantages and limitations to the relationship. The more such symbiosis advances, the more difficult it will be for either symbiont to flourish without the other. (Hayles 2017, 216)

While Searle's views of cognition are predominantly anthropocentric, his theory of 'biological naturalism' links consciousness and neurobiology in a way that mirrors Hayles' distinction between conscious and non-conscious cognition. According to Searle (2004, chap. 4), 'biological naturalism' provides a solution to the 'mind-body problem' that avoids both dualism (i.e., the Cartesian distinction between mind and body) and materialism (i.e., the reduction of mental phenomena to physical states of the brain). His theory argues that all conscious states are caused by lower-level neuronal processes and therefore are *causally*, though not *ontologically* reducible to these processes. That is, consciousness can be *causally* explained by neuronal functions but is not *ontologically* reducible to them, since it has a subjective, first-person ontology.

In addition to being rooted in human exceptionalism, anthropocentric views of intelligence and cognition are based on another disputed assumption: the categorical divide between human and non-human. Haraway's *Cyborg Manifesto* puts the distinctions between human and machine, as well as human and animal into question. Haraway (2016) argues that dichotomies between mind and body, animal and human, organism and machine, private and public, nature and culture, men and women, primitive and civilized have been persistent in Western thought and instrumental to the logic and practices of domination of all constituted as "others": women, people of color, nature, workers and animals. In her *Cyborg Manifesto*, the 'cyborg' represents ‘transgressed boundaries’, ‘potent fusions’ (Haraway 2016, 14) and rearrangements in social relations tied to science and technology. For Haraway, machines are not an “other” to be worshipped or feared; they can be ‘prosthetic devices, intimate components, friendly selves’ (61).

> The machine is not an it to be animated, worshipped, and dominated. The machine is us, our processes, an aspect of our embodiment. We can be responsible for machines; they do not dominate or threaten us. We are responsible for boundaries; we are they. (Haraway 2016, 65)

Barad (2007) also refuses to take the distinction between human and non-human for granted and examines the practices through which the boundaries between the differential categories of human and non-human are stabilized and destabilized. Her posthumanist account acknowledges the important role that non-humans play in social and technoscientific practices. As an alternative to representationalism, which assumes a tripartite distinction among entities awaiting representation, representations and subjects/knowers, she proposes a performative understanding of scientific practices that places the emphasis on the practices through which these representations are produced. For Barad (2007) ‘knowing does not come from standing at a distance and representing but rather from a *direct material engagement with the world*’ (49). Her concept of *intra-action* suggests that there are no distinct agencies that precede their interaction. Rather, distinct agencies emerge through their *intra-action* and exist only in a relational sense.

Latour (2005) also embraces a posthuman perspective, defining an agent as anything that is made to act and can ‘modify a state of affairs by making a difference’ (71). While this definition is very broad, Latour clarifies that his *Actor-Network-Theory* does not aim to establish some kind of ‘absurd’ symmetry between human and material agency (76).

Bennett (2010) uses the term ‘thing-power’ to refer to ‘the curious ability of inanimate things to animate, to act, to produce effects dramatic and subtle’ (6) and proposes an understanding of agency as distributed across an ‘ontologically heterogeneous field’ (23), rather than the result of human action alone. Similarly to Latour’s (2005) *Actor-Network-Theory*, Bennett’s ‘vital materialism’ does not aim to flatten the differences between humans and non-humans by suggesting that all actants are equal. Rather, it claims that ‘there is no necessity to describe these differences in a way that places humans at the ontological center or hierarchical apex' (11). While her account rejects anthropocentricism, it does not dismiss anthropomorphism; on the contrary, it claims that anthropomorphism can help reveal similarities and 'isomorphisms' across categorical divides (99).

The compositional approach described in this dissertation is rooted in posthuman notions of agency and cognition, viewing these phenomena as fundamentally distributed across the human/machine divide. To paraphrase Hayles' definition of cognition, the interactive music systems described in the following chapters are conceived as artificial congizers that *interpret auditory information within live performance contexts that connect it with musical meaning*. The *interpretation of information* by these systems involves a choice, which, in line with Hayles' definition, consists in choosing among possible courses of action and is not synonymous with free will. Finally, *musical meaning* in this context refers to the interpretation of auditory information with respect and in relation to specific musical ideas and interaction concepts within a given work.

In the distributed approach to creativity described in this dissertation, human and machine actors have *distinctive cognitive capacities, but no clear boundaries*, as they coalesce to form larger co-creative assemblages. While clearly influenced by materialist understandings of agency, this approach differentiates itself from the new materialisms of Barad (2007), Latour (2005) and Bennett (2010) in that it attempts to understand and describe, rather than simply acknowledge, the differences between human and non-human agency. To that end, materialist and anthropocentric views of intelligence and creativity are in dialogue with each other throughout this dissertation, in an attempt to avoid the pitfalls of what is yet another dualism. For instance, Chapter 2 examines musical creativity from an anthropocentric perspective with the purpose to better understand the distinctive capacities of human and computational creativity. The aim of this approach is to explore the relationship between human and computational creativity on the basis of their unique capacities, rather than negate the attribution of creativity to computers.

Interestingly, as far as the relationship between human and computational intelligence is concerned, both anthropocentric and posthuman accounts seem to arrive at the same conclusion, favoring human-machine cognitive assemblages over a competitive relationship between human and artificial cognition. Even Dreyfus (1992) does not seem to reject AI in general, but rather efforts to automate cognitive tasks, and seems to be in favor of cooperative, human-in-the-loop approaches to AI (304).

Reconciling posthuman notions of agency and cognition with the dualisms that pervade language is at present an open philosophical challenge that lies far beyond the scope of this dissertation. In discussing the difficulties involved in conceiving and describing agency in non-anthropocentric terms Bennett writes:

> In composing and recomposing the sentences of this book – especially in trying to choose the appropriate verbs, I have come to see how radical a project it is to think vital materiality. It seems necessary and impossible to rewrite the default grammar of agency, a grammar that assigns activity to people and passivity to things. Are there more everyday tactics for cultivating an ability to discern the vitality of matter? One might be to allow oneself, as did Charles Darwin, to anthropomorphize, to relax into resemblances discerned across ontological divides. (Bennett 2010, 119)

The language used to describe non-human agency in this dissertation might be perceived as somewhat anthropomorphic. Similarly to Bennett's approach, this is meant to highlight similarities that stretch across the human/machine divide and point out that the boundaries between human and non-human agency in the agentic assemblages that are the interactive musical works described here are porous and hard to define.

1.6 AI as a Secondary Agent

Along with Hayles' *non-conscious cognition*, Gell's concept of 'secondary agency' is central to how AI is conceptualized in this research. Gell (1998) defines an agent as someone or something that 'causes events to happen in their vicinity' (16), but makes an explicit distinction between 'primary agents', i.e., intentional (human) beings, and 'secondary agents' (objects, artifacts, works of art), through which 'primary agents' exercise and distribute their agency (chap. 2). Regarding the status of 'secondary agents' and their relation to 'primary agents', Gell explains:

> I describe artefacts as 'social agents' not because I wish to promulgate a form of material-culture mysticism, but only in view of the fact that objectification in

> artefact-form is how social agency manifests and realizes itself, via the proliferation of fragments of 'primary' intentional agents in their 'secondary' artefactual forms. (Gell 1998, 21)

Far from being simply a manifestation of 'primary agency', 'secondary agents' can also become its instrument, by influencing human actions and causing 'events to happen'. For instance, reappropriation of technologies seems to be an integral part of the life cycle of technological innovations (Latour and Venn 2002), with examples ranging from the typewriter, originally developed for blind people, to the internet, originally invented to enable communication among scientists (Hayles 2017, 36). The use of technologies for a different purpose than they were intended is one of many facets of the reciprocal relationship between 'primary' and 'secondary' agency and the dual role of 'secondary agents' as the 'outcome' and 'instrument' of social agency (Gell 1998, 15).

As far as compositional applications of AI are concerned, the role of machine learning as a 'secondary agent' lies in its potential to influence musical thinking, by opening up new creative possibilities and allowing for new artistic practices to emerge and take form. Novel interaction paradigms enabled by machine learning capabilities can lead to new conceptual paradigms that destabilize the dualisms between composition and performance and redefine musical authorship and performership.

Gell (1998) considers art as a 'system of action': a social process in which art objects play a 'mediatory role' (6), though his definition does not exclude other material and immaterial entities as social agents. The research described in this dissertation explores the role of machine learning algorithms as 'secondary agents' in the author's own work, focusing on their transformative potential for compositional thinking. Particularly, it explores how interaction paradigms enabled by machine learning capabilities can transform the sociosonic (human-human and human-technology) relations manifested in a musical work and establish new notions of authorship and ontologies of the musical work, rooted in interactivity and distributed creativity.

1.7 Compositional Methods

Balancing authorship and interpretative freedom in interactive compositions can be a challenging task and requires compositional methods and composer-performer collaboration practices that differ fundamentally from those involved in the 'directive' composer-performer collaboration paradigm (Hayden and Windsor 2007). In this paradigm, the composer might consult the musician regarding playing techniques and notation conventions in a meeting that usually takes place at the very beginning of the compositional process and have little to no contact with them until the first rehearsals of the finished piece. A compositional practice the premise of which lies in a notion of work identity that encompasses, rather than excludes, diverse musical outcomes calls for an experimentation-oriented compositional approach that takes into account and is in dialogue with potential interpretations of the compositional concept and performance instructions.

In the research presented here, experiments based on guided improvisation tasks and conducted with the help of musicians provided a fertile ground for creative discovery and helped refine compositional ideas and performance instructions. These experiments helped identify potential discrepancies between intended and perceived interaction affordances of the interactive music systems and devise performance instructions that effectively balance the trade-off between work identity and interpretative freedom in the works developed as part of this research.

Concretely, the methods employed in this research included three different types of guided improvisation tasks: exploratory, 'naïve' and 'informed' (Hsu and Sosnick 2009) rehearsals. Each of these methods serves a different purpose and was used in a different stage of the compositional process. Data from these improvisation sessions was collected using ethnographically informed research methods, such as observation, questionnaires and semi-structured interviews with the musicians. The interpretation of qualitative data collected through these methods provided valuable insights into the works described in this dissertation and played a decisive role in their development. Among these methods, questionnaires and interviews aimed at gaining insight into the musicians' perspective and experience of their interaction with the computer music system, while observation was performed from the composer's perspective. Data collected through these methods is presented throughout this dissertation in the form of 'thin' and 'thick descriptions', i.e., mere accounts and interpretations of the musicians' responses respectively (Geertz 1973).

The format of *exploratory rehearsals* was used in only one of the compositions developed as part of this research and had the purpose of exploring two abstract concepts: "convergence" and "divergence". Concretely, the musicians (pianist and double bassist) were asked to improvise on the concepts of "convergence" and "divergence" and then reflect on their interpretation of these concepts through questionnaires and a semi-structured group interview (Chapter 5). These experiments were conducted during the conception phase of the compositional process and had the purpose of exploring the evocative power of these concepts as musical metaphors and the degree of intersubjectivity involved in their interpretation in the context of musical improvisation.

While exploratory rehearsals focused exclusively on the interaction between (human) musicians, 'naïve rehearsals' (Hsu and Sosnick 2009) were centered around human-computer interaction. In these sessions, the musicians were asked to improvise with an interactive music system without being given any information on its interaction affordances and capabilities prior to the improvisation. The purpose of these sessions was to explore the relationship between intended and perceived interaction affordances of the computer music system and determine the extent to which interaction affordances can communicate compositional intent. Naïve rehearsals were followed by a questionnaire and semi-structured interview with the musicians, in which they were asked to describe different behaviors exhibited by the interactive music system and assess its degree of responsiveness, predictability and autonomy. While not originally intended as such, naïve rehearsals proved to be a valuable tool for creative exploration and discovery, as, in certain cases, creative misunderstandings and "misinterpretations" of the affordances of the computer music system ended up informing future revisions of the code (Chapter 4).

In 'informed rehearsals' (Hsu and Sosnick 2009), musicians were asked to improvise with the interactive music system after being given some general information regarding its auditory processing and interaction capabilities, but no performance instructions. These sessions provided an opportunity to observe the musicians' interpretative choices and devise performance instructions that would guide their actions towards the intended action spaces. The focus in these rehearsals therefore shifted from the exploration of intended and perceived interaction affordances of the interactive music system to the analysis and further refinement of the action spaces available to the musicians. Data from these sessions was collected through observation.

In general, the use of methods such as questionnaires and interviews in this research was meant to facilitate aesthetic reflection, by providing insight into the performers' perspective. Importantly, the research paradigm adopted by this research differs fundamentally from Human-Computer Interaction (HCI) approaches that share a similar methodology (Hsu and Sosnick 2009; Brown, Gifford, and Voltz 2017; Weinberg and Driscoll 2006). While the formats of 'naïve' and 'informed rehearsals' were borrowed from previous work in the evaluation of human-computer improvisation systems (Hsu and Sosnick 2009), the purpose of their use within the context of this research was not an evaluation of the interactive music systems by the musicians, but aesthetic reflection and creative experimentation as part of the compositional process.

As far as broader research paradigms are concerned, this research is aligned with the constructivist approach, in which investigator and object of investigation are interactively linked and knowledge is created as a result of and through that interaction (Guba and Lincoln 1994). In the context of practice-led artistic research, however, "knowledge" is understood in radically subjectivist and relativist terms. "Knowledge" here refers to creative insights grounded in culturally constituted and subjective aesthetic values and gained through the creative process. Consequently, the interpretation of data collected through the methods described above is explicitly subjective and serves the purpose of aesthetic reflection, rather than theory generation. Nevertheless, by describing the findings and insights gained through this research, this dissertation aims to make this knowledge and the methods through which it was attained available to others and contribute to the ongoing methodological discourse in artistic research in composition.

These methods were developed and revised during this research, based on insights gained through the processes of data collection and interpretation. For instance, informed rehearsals were initially viewed as complementary to naïve rehearsals and were followed by a questionnaire on the interaction affordances of the interactive music system, which had the purpose of determining whether the musician's perception of them changed after receiving information on the capabilities of the system. Later during this research, the purpose of these sessions shifted from identifying the perceived affordances of the interactive music system to observing the action spaces evoked by these affordances and devising performance instructions that would guide the musicians' actions towards more idiosyncratic sonic interactions. This shift was triggered by the realization that observation of these sessions led to valuable insights regarding the type and content of performance instructions that would be needed in order to delineate various interaction scenarios. As a result, in later stages of this research, data from informed rehearsals was collected through observation, rather than questionnaires, a shift that highlights the prominence of autoethnographic (auto- from Greek αὐτός, "self") over alloethnographic (allo- from Greek ἄλλος, "other") perspectives in this research.

The use of human-computer improvisation as a method for creative experimentation during the compositional process is indicative of a compositional approach that could be described as 'subtractive composition' (Marko Ciciliani, in discussion with the author, March 2019). This involves starting from an action space that is as open as possible, i.e., "free" improvisation, and gradually reducing this space to more idiosyncratic sonic interactions through the introduction of performance instructions. Admittedly, such an improvisation is "free" only as far as performance instructions are concerned, as the affordances of the interactive music system inevitably function as form of "constraint", by influencing the performers' actions and evoking specific types of responses.

1.8 Overview

The following chapters of this dissertation discuss the compositions developed as part of this research, the methods used in their development and the positioning of this research with respect to Computational Creativity and Music AI. The musical works are presented in a chronological order and discussed with respect to technical, methodological, conceptual and aesthetic implications of the concept of interactive compositions.

The presentation of the works in a chronological order is meant to highlight the autoethnographic foundation of this research and the reciprocal relationship between creative ideation and aesthetic reflection in artistic research. Particularly, the structure of this thesis aims to emphasize the various levels – conceptual, technical, methodological etc. – on which insights gained through each work informed the next. The dialogic relationship between creative ideation and practice resulted in various shifts in the foci of this research, the most notable of which is a shift from exploring the capabilities of machine learning algorithms to exploring their specificities and limitations. Another shift in the focus of this research made clear by the structure of this dissertation is the one from the description of technical and aesthetic aspects of the works discussed (Chapters 3 and 4) to the methods used in their creation, as well as their the broader sociosonic context and aspects of human-human and human-computer co-creativity in them (Chapters 5 and 7).

Concretely, Chapter 2 discusses the disparities and relationship between human and computational creativity, considering two distinct possibilities for the application of the latter: as a *simulation* and as an *extension* of human creativity. These two approaches are examined with respect to Boden's (2010) definition of creativity, as well as their underlying assumptions about the nature of creativity. This research is positioned within the paradigm of *distributed human-computer co-creativity*, in which computational creativity extends – rather than replaces – human creativity, and proposes an understanding of AI as an ideation tool, which has the potential to expand the space of creative possibilities and transform artistic practices.

Chapter 3 describes some first experiments in machine listening, involving a feedforward Neural Network trained to perform real-time recognition of different playing techniques on the soprano saxophone. This machine listening algorithm was integrated in a composition for soprano saxophone and interactive music system exploring different levels of aural attention in machine listening. The integration of a classification task in the auditory processing stage of the interactive music system had the purpose of shifting the focus of machine listening from sensory (signal-level features) to symbolic information (composer-defined sound classes), enabling the design of idiosyncratic agentive behaviors. The interaction scenarios involved in the composition are based on two of Truax's (2001, chap. 2) levels of aural attention ('listening-in-readiness' and 'listening-in-search') and expand on them with two additional attentional strategies: 'listening-in-context' and 'listening-at-will'.

Chapter 4 describes an interactive composition for human and robotic percussionist. Building on previous work, the auditory processing stage of the robotic percussionist incorporates a Neural Network trained to recognize different instruments and playing techniques. The robotic percussionist continuously assesses its interaction with the musician and chooses to either to "follow" them or take the "lead", by initiating musical changes. This decision is aesthetically driven and based on the evolution of three different metrics of musical contrast.

Chapter 5 discusses aspects of 'collaborative emergence' (Sawyer 2000, 183) in a composition for piano, double bass and interactive music system. In this work, the interactive music system analyzes and responds to the interaction between the musicians, rather than their individual actions. Concretely, the system monitors the timbral similarity between the two audio inputs and tries to identify moments of timbral convergence or divergence between them. As musical changes are initiated based on the relationship (i.e., similarity) between the two audio inputs, musical form in this piece emerges as a result of joint action and collective decision-making. This chapter contains an extensive description of the methods used in the compositional process and the concrete ways in which they informed creative decisions throughout it. In addition to autoethnographic aspects of this research, it highlights aspects of human-computer and human-human (composer-performer and performer-performer) co-creativity in interactive musical works and the broader sociosonic context within which they are embedded.

Chapter 6 examines the same work from a music-analytical perspective with the purpose to explore aspects of interpretative individuality in it and their relation to work identity. As a means of addressing some of the music-analytical challenges posed by interactivity – most importantly, the complex relation between the work and its performances as 'partial manifestations' (Young 2016, 96) of the possibilities it encompasses – this analysis adopts a comparative performance-centered approach. Two different performances of the composition by ensembles Schallfeld and Klangforum are analyzed and compared using a variety of methods, including a formal analysis, audio analysis and video-based interaction analysis (Jordan and Henderson 1995).

Chapter 7 describes a subversive approach to AI focused on the exploration of AI bias. In *Bias*, for bass clarinet and interactive music system, a computer music system using two Neural Networks trained to simulate the author's aesthetic judgments interacts with the musician by evaluating the sound input based on its "subjective" aesthetic judgments. Arbitrary assumptions about the training data made by the machine learning algorithm result in an aesthetic agency that deviates from the author's aesthetic preferences. The composition problematizes the discrepancies between aesthetic value as a non-measurable quality, which is subjective and socially constituted, and the concepts of error and accuracy normally associated with supervised machine learning, and aims to blur the boundaries between human and machine agency.

Finally, Chapter 8 discusses the significance of the individual works described in this dissertation to the research narrative and summarizes some of the most important findings and insights gained through this research. Along with conceptual shifts relating to musical authorship, the work-concept and interpretative freedom and individuality, these include interpersonal aspects of composer-performer collaboration and the tension between emerging human-human and human-computer co-creative practices and conventions surrounding the division of musical labor. The chapter ends with a discussion of the distributed nature of the musical work, as well as future challenges and directions for research in composition and AI.

2 Human and Computational Creativity

2.1 Defining Creativity

How well are computers currently performing at musical tasks? Could they potentially "outperform" human composers? Or, are *truly* creative computer music systems a figment of fiction? Questions such as these seem to dominate public discourse around AI and the arts, fuelled by the hype around Machine Learning and Data Science and their increasingly central role in public and private life. While definitive answers to these questions remain elusive, understanding the distinctive capacities of human and computational creativity is crucial to both the design of autonomously creative systems and human-computer co-creative approaches, in which creative decisions are distributed across human and non-human actors. A subcase of the latter are the interactive compositions described in this dissertation.

Creativity can be understood in a variety of contexts that span from human activity to biological processes, such as evolution (Bentley and Corne 2002). For the purposes of this discussion, however, creativity will be defined in anthropocentric terms and in relation to artistic production. This approach does not deny the attribution of creativity to non-human actors; rather, it seeks to understand the distinctive capacities of human and computational creativity and, by extension, the different types of relationships that are possible between them.

Boden (2004) defines creativity as 'the ability to come up with ideas or artefacts that are new, surprising and valuable' (29). In this chapter, Boden's criteria of *novelty* and *value* will be used as a basis for assessing autonomously creative music systems and will be juxtaposed with implicit and explicit assumptions about creativity that underlie the design of such systems. As far as novelty is concerned, the focus in this discussion will be placed on *H-creativity* (Historical Creativity), as opposed to *P-creativity* (Personal or

Psychological Creativity), that is, novelty with respect to music history, i.e., innovation (Boden 2010, chap. 2). Boden's third criterion, *value*, is arguably the hardest one to define, as aesthetic values are not only hard to describe in propositional terms, but also vary across cultures and time. Indeed, the focus of Boden's definition on a concept as intangible as *value* has been contested, with Bown (2012) distinguishing between two types of creativity (*generative* and *adaptive*) only one of which is concerned with value and Dorin and Korb (2012) rejecting discussions of value as irrelevant to creativity and effectively separating creativity from its attribution. The debate on value as a criterion for creativity will be examined more closely later in this chapter. For now, Boden's definition will be used as a starting point in assessing the capabilities of automatic composition systems and the assumptions about creativity underlying their design.

2.2 Automatic Composition Systems

In lieu of an exhaustive literature review of the plethora of automatic composition systems currently available, this chapter will consider three distinctly different approaches to the simulation of musical creativity and examine their successes and shortcomings as reported by their designers and users: David Cope's *Experiments in Musical Intelligence* (*EMI*), Nick Collins' *Autocousmatic* and *WaveNet*.

David Cope's *EMI* is a style imitation system designed to produce acoustic compositions (i.e., MIDI scores) in tonal musical idioms. The program performs a statistical analysis of a corpus of musical works, encoded as MIDI scores, with the purpose to identify patterns across them (Cope 1992, 1996). Based on this analysis and using Augmented Transition Networks (ATNs), it generates new pieces in the style of the sample works. Some of the shortcomings of the program, as reported by Cope (1992), include the inability of the program to recognize the minor mode and deal with chromaticism, cadences, phrase length and musical form (82).

Nick Collin's *Autocousmatic* is designed for a different musical genre (electroacoustic art music), but operates based on a similar principle. *Autocousmatic* generates audio mixes based on models of musical form derived from sample works, using a database of sound files provided by the user (Collins 2012). Both the input sounds and generated mixes are analyzed using audio descriptors. As part of the evaluation of the program's outputs, *Autocousmatic*-generated compositions were submitted to conferences and music festivals and evaluated by professional electroacoustic music composers. The professional composers' feedback suggested that the generated compositions lacked 'directionality' and the program seemed to lack the ability to deal with musical form and, particularly, transitions between different sections (Collins 2012, 18–19). Collins also reports that none of the submissions to calls-for-music have been successful so far.

Finally, *WaveNet* is representative of a much more recent trend in the automation of musical creativity, based on Deep Learning algorithms that learn from unstructured data (i.e., raw audio data, as opposed to features such as pitch or spectral descriptors). *WaveNet* is based on a probabilistic and autoregressive model, in which predictions for the next audio sample are conditioned on all previous samples (van den Oord et al. 2016). While originally developed for text-to-speech applications, the model has also been used in music generation tasks, mainly in tonal musical idioms (e.g., van den Oord et al. 2016; Manzelli et al. 2018). *WaveNet*-generated musical outputs are characterized by partially convincing local structure, but poor global structure, suggesting that the algorithm fails to learn mid- and long-range dependencies (Manzelli et al. 2018, 1).

Overall, human evaluation of AI-generated compositions seems to suggest that automatic composition systems are currently failing the Turing Test[4]. The systems described above differ fundamentally in terms of their implementation, yet demonstrate similar shortcomings – most notably an inability to deal with musical form. Therefore, the question that needs to be addressed is whether human-level computational creativity is just a matter of further refining and improving these models, or whether the shortcomings of automatic composition systems are indicative of more fundamental challenges involved in the simulation of human creativity. The next two sections try to address this question by examining the assumptions about musical creativity that underlie the design of autonomously creative systems, as well as some of the difficulties involved in the definition and attribution of creativity.

[4] Broadly defined, a Turing Test is meant to determine whether computer-generated outputs are indistinguishable from those produced by humans. An automatic composition system would pass the Turing Test, if a human evaluator was unable to determine whether its output was generated by a human or a machine.

2.3 Computers and Creativity

2.3.1 Novelty

Whether they are based on MIDI data, perceptual audio descriptors or raw audio data, the automatic composition systems discussed in the previous section are guided by a similar principle, essentially reframing the problem of musical creativity as one of style imitation. They analyze a corpus of works using features that are relevant for the musical style at hand (e.g., musical notes or perceptual descriptors) and subsequently generate outputs that resemble the sample works. As discussed above, these systems are currently failing the Turing test – i.e., they are failing to produce outputs that are indistinguishable from human-generated music. However, as automatic composition systems continue to improve their performance, this might no longer remain the case. But, would an automatic composition system succeeding in "mastering" the style of one or more human composers mean that computational creativity has reached or surpassed human creativity?

Creativity and, particularly, *H-creativity* involves skills beyond craftsmanship; namely imagination, resourcefulness and the ability to think beyond established norms and paradigms. Indeed, some of the most pivotal and influential works in music history are the ones that broke away from tradition, for example by establishing new styles (e.g., twelve-tone music), or by questioning the very ontology of music and the construct of the musical work (e.g., John Cage's (1952) *4′33″*). It seems then that human-comparable computational creativity is not just a matter of improving currently existing models, but rather re-examining the fundamental assumptions that underlie their design and scope.

Whether they are based on hand-coded rules or machine learning, automatic composition systems are designed to imitate already existing styles. Creativity, in the context of these systems, is understood as the ability to imitate, or conform to the constraints of a given style. Clearly, that leaves out a crucial side of human creativity: innovation, or 'transformational creativity' (Boden 2010, chap. 5).

Boden (2010, chap. 5) distinguishes between 'combinational', 'exploratory' and 'transformational' creativity. The latter two categories are crucial to understanding human creativity and its distinctive capacities from computational creativity. 'Exploratory' creativity involves generating new ideas – or artifacts – within already established conceptual spaces (i.e., 'exploring' already established styles of thought). Interestingly, Boden (2010) mentions David Cope's automatic composition system as an example of 'exploratory' creativity (37-38). Contrastingly, 'transformational' creativity involves transforming existing conceptual spaces and establishing new styles of thought. An example of this type of creativity, according to Boden (2010), is Schoenberg's twelve-tone system (74).

It appears then that the question "are computers capable of human-level musical creativity?" needs to be rephrased as "are computers capable of *transformational* creativity?"

2.3.2 Value

Boden (2010) seems to answer this question positively, using evolutionary algorithms that can randomly change their rules as an example of transformational computational creativity (38). However, the suggestion that by randomly modifying a set of rules evolutionary algorithms create new styles of thought seems to contradict Boden's third criterion for creativity: *value*[5].

Arguably, creativity is socially constituted and cannot be studied outside the historical and social context within which creative acts are carried out (Csikszentmihalyi 2014, 47). Therefore, whether a new composition system qualifies as a "style" can only be determined by its impact on and acceptance by the field of music composition (de Jager 1972; Meyer 1983). Schoenberg's twelve-tone system has had a considerable impact on western art music history, by influencing the work and musical thinking of his contemporaries and successors. Had Schoenberg been the only one to use it, the twelve-tone system would probably not be considered a "style", nor would it hold the cultural value it holds today.

[5] The term value here is used to refer to aesthetic or cultural value, rather than financial value. Admittedly, computational creativity is already producing high market value, particularly in the field of visual arts. For instance, in 2018 an AI-generated eighteenth-century-style painting was sold for over $400,000 (Cohn 2018).

The challenges posed by the concept of *value* have led Dorin and Korb (2012) to formulate an alternative definition of creativity that detaches it from notions of value. They argue that what makes an activity creative must be intrinsic to the activity, rather than related to the reception of its outcomes and use the example of artists that were recognized only posthumously as an argument for the detachment of creativity from persuasion. However, this begs the question: if it is not reception that determines whether something is creative or not, then what is it? Or, as Csikszentmihalyi (2014) puts it: 'if you cannot persuade the world that you had a creative idea, how do we know that you actually had it?' (102).

Csikszentmihalyi's (2014, chap. 4) systems model of creativity considers creativity as the product of the interaction between the *individual*, a *domain* and a *field*. *Individuals* produce variations of the corpus of knowledge contained within the *domain*, while the *field* (the institutions and individuals that can affect the structure of the *domain*) selects those variations that are worth preserving and incorporating into the *domain*. Csikszentmihalyi argues that each of these systems (*individual*, *field* and *domain*) both affects and is affected by the others. For him, creativity is inseparable from persuasion, as it is constructed through the interaction between the products of individuals and the judgments social systems make about them (Csikszentmihalyi 2014, chap. 8).

The same could be argued for art. Marcel Duchamp's *Fountain*, a readymade sculpture consisting of a urinal, and John Cage's (1952) *4′33″* are only art, because we (society, or a smaller social group within it) agree it is art. The ontological status of these works as artworks is not determined by intrinsic but extrinsic qualities and is grounded in social agreement.

Still, to simply acknowledge the inseparability of creativity and persuasion would be to oversimplify the role that social systems play in the attribution of creativity. For instance, the absence of female composers from history textbooks and the Western music canon (Oliveros 1970; Rodgers 2010; Criado Perez 2019, 16–19) suggests that recognition is not simply a matter of persuasion, as some individuals might be excluded from the canon for reasons that have nothing to do with the value of their contributions, e.g., due to their gender, race or ethnicity. While this example does indeed support Csikszentmihalyi's argument that social systems play an important role in the attribution of creativity, it also proves that the relationship between creativity and recognition is far from straightforward.

The close entanglement between creativity and value and their dependence on social agreement seem to pose another significant challenge for computational creativity, concerning the evaluation of computer-generated artifacts. The evaluation of AI-generated compositions is commonly based on Turing Tests, meant to determine whether they are distinguishable from works created by human composers. Ariza (2009) proposes a series of variations of the Turing Test, designed specifically for evaluating the outputs of generative music systems. The rationale behind the use of Turing Tests is that the ability of an automatic composition system to produce outputs that are indistinguishable from human-composed musical works should be evidence that the system is in fact creative.

An obvious limitation of the Turing Test as a method for evaluating computer-generated artifacts is that the criterion of indistinguishability from already existing artworks is almost a perfect antithesis to the criteria by which human creativity is usually judged, i.e., novelty and individuality. Additionally, Turing Tests can only be performed *post factum*. While some automatic composition systems integrate machine listening processes as part of a formative evaluation taking place during the compositional process (e.g., Collins 2012), such processes fail to simulate listening as an analytical and evaluative act, based on culturally informed and – most importantly – *subjective* aesthetic criteria.

In addition to Turing Tests, which are performed by human listeners, computer-generated artifacts can be evaluated through computational means. Computational aesthetic evaluation encompasses a wide range of approaches, from formulaic theories to biologically inspired fitness measures and empirical aesthetics and can be based both on human-defined and software-generated aesthetic criteria (Galanter 2012). Yet, the challenges involved in the evaluation of computer-generated artifacts through computational means seem to be even greater.

Kalonaris and Jordanous (2018) criticize approaches using formulaic measures to evaluate musical works (e.g., Manaris, Romero, and Machado 2005; Manaris et al. 2007) for equating pleasantness and popularity with aesthetic value, as well as for assuming that aesthetic value can be judged based on universal aesthetic principles. McCormack (2012) is also critical towards universal aesthetic values – even those derived through empirical studies (e.g., Martindale 1988) – stressing out the dependence of aesthetic judgments on cultural and subjective values and questioning the relevance of 'surface aesthetic qualities' (44) for the appreciation of modern art. The use of computer-generated aesthetics as an alternative to simulated human aesthetics has both been praised (Dorin and Korb 2012) and criticized (McCormack 2012; Galanter 2012). All in all, the debate on computational aesthetic evaluation is far from settled.

2.4 Human Creativity

Understanding how (human) composers innovate seems crucial to determining whether computers are capable of transformational creativity. The work of music pioneers such as Pauline Oliveros, John Cage and Arnold Schoenberg exemplifies the complex nature of musical creativity and can potentially help shed some light on the challenges involved in the simulation of transformational musical creativity.

Pauline Oliveros' (1974) *Sonic Meditations* is a revolutionary work characterized by a participatory approach to music-making that aims to 'erase the subject/object or performer/audience relationship by returning to ancient forms which preclude spectators' (1). *Sonic Meditations* is a collection of text scores for a group of participants – both musicians and non-musicians – meeting regularly over a longer period of time. The goals of this activity, as described by Oliveros (1974), are sharing a common experience with other members of the group, expanding one's sonic awareness and releasing physiological and psychological tension, while music *per se* is only 'a welcome by-product' of this process (1). Oliveros' process-over-product approach prioritizes music-making as an *experience* over the aesthetics of its *outcome*, highlighting another fundamental issue relating to the evaluation of both human and computer-generated compositions: the relation between process and product and their relative prioritization within different artistic approaches.

Similarly, in his work *4′33″*, a piece that consists entirely of silence, John Cage (1952) questions the ontology of music and the construct of the musical work. Cage's work illustrates how musical meaning is constructed from cultural and historical contexts. The value of the work lies in the position it takes with respect to the debate on musical ontology – for instance, Varèse defined music as 'organised sound' (Varèse and Wen-Chung 1966, 18) – rather than its intrinsic qualities. Therefore, taking the work out of its context would be stripping it of its cultural value. For instance, one could not expect to feed this piece into a machine learning algorithm and generate music in Cage's style.

Both Oliveros' and Cage's examples demonstrate that art can neither be produced nor perceived outside a sociocultural and historical context, standing in stark contrast to models of human creativity based exclusively on domain-specific knowledge. Admittedly, these examples involve *innovation in goals*, that is, innovation with respect to extra-musical ends, in addition to *innovation in means* (de Jager 1972). However, even in Schoenberg's case, in which innovation concerns mainly the means (i.e., a composition system) rather than the ends of the creative process, transformational creativity required an extensive knowledge of music history, which led him to the judgment that the tonal system had reached and exceeded its limits and would need to be replaced by new composition systems.

All of these examples highlight an aspect of human creativity that automatic composition systems seem to either consider as inconsequential or implicitly acknowledge that it is impossible to simulate: its *situated* nature. The situated nature of human creativity is evidenced by the influence extra-musical factors, such as technological advances, have on artistic practices. For instance, musique concrète and laptop performances would not have been possible, if it were not for the development of recording technologies and the invention of the microchip (and later the personal computer) respectively. Artistic practices such as network performances, live coding and performances with sensor-based interfaces are all made possible by the low cost, small size and high computational power and speed of modern computers. Yet, the realization that these technologies held creative potential for musical thought and practice was arrived at by 'minds-in-the-world' (Kohn 2013, 34), that is, *situated* cognizers that were able to envision new and, in many cases, subversive uses of these technologies.

Importantly, there is one aspect of human creativity that was intentionally left out of this discussion: *consciousness*. The reason for this is that consciousness is a fundamentally unsolved philosophical problem to which theories such as solipsism, epiphenomenalism, materialism, behaviorism, physicalism and functionalism have all produced largely contradictory answers (Searle 2004). Solving the problem of consciousness is beyond the scope of this thesis. However, failing to acknowledge that experience and inter-human communication are important aspects of the production and reception of art would be failing to understand its nature.

O'Hear (1995) defines art as inter-human communication, effectively negating the possibility of computational creativity. Axiomatically rejecting computational creativity as an impossibility would indeed be counterproductive to the discussion on human and computational creativity. Yet, it is important to acknowledge the fundamental role that inter-human communication plays in aesthetic experience. John Cage's (1952) *4′33″* is an excellent example of this communication: by subverting the listeners' expectations, the work invites them to challenge their most fundamental assumptions about music and the musical work.

Even in light of an epiphenomenalist view of consciousness, according to which consciousness exists but is 'causally inert' (Searle 2004, 21) – it is an *epiphenomenon*[6] – the challenge still remains: in order for *transformational* computational creativity to become possible, our models of human creativity would need to be expanded to reflect its *situated* nature. Currently, this appears to be beyond computational means. This is not to suggest that transformational computational creativity is fundamentally impossible, but that while our models remain focused on domain-specific features, e.g., MIDI data or spectral descriptors, computational creativity will probably not be able to challenge human creativity.

2.5 Artificial Intelligence vs. Intelligence Augmentation

Far from suggesting that computational creativity is not worth pursuing, the distinctive capacities of human and computational creativity invite us to rethink the relationship between them. Machine learning algorithms produce impressive results when applied to closed-ended tasks within controlled environments with clearly defined inputs; for instance, in applications such as image recognition. Open-ended tasks, such as creative tasks, on the other hand, in which goals are not defined in advance and often gain their meaning through reference to broader cultural and historical contexts, seem to rely on situated forms of cognition and, as a result, resist computational reduction and contextual detachment. The distinctive capacities of human and computational creativity suggest that a complementary, rather than a competitive, relationship between them might hold greater potential for artistic innovation.

[6] A by-product or symptom of another process or phenomenon.

Questions regarding the relationship between human and computational intelligence are not new and can be traced back to the early days of AI and the debate on *Artificial Intelligence* (computers simulating human cognition) versus *Intelligence Augmentation* (computers augmenting human cognition) (Ashby 1964; Licklider 1960; Engelbart 1962). The objective of creativity augmentation could be an interesting alternative to that of automation. In the context of such an approach, computational creativity could contribute to larger, co-creative human-machine assemblages, through its distinctive capacities and affordances.

Bruno Latour's *Actor-Network-Theory* provides a potentially useful conceptual framework for such an approach. Latour (2005) proposes a non-anthropocentric notion of agency, by defining an actor as anything that 'is made to act' (46) and can 'modify a state of affairs' (71). Posthuman notions of agency and creativity can open up new conceptual and technical possibilities, by viewing human and computational creativity as actors interacting within and contributing – albeit in an asymmetrical way – to larger co-creative networks.

2.6 Distributed Human-Computer Co-creativity

Another interesting conceptual framework for distributed human-computer co-creativity are McCormack's (2012) *creative ecosystems*, which encompass 'humans, technology and the socially/technologically mediated environment' (56). McCormack's ecosystemic approach to creativity does not seek to automate creative tasks, but rather open up new creative possibilities and enhance human creativity – i.e., in the ecosystemic approach creativity is understood as an exploratory, rather than an optimization process.

Co-creative approaches in which computational means are used to expand and enhance, rather than replace, human creativity are many and an exhaustive review of them is beyond the scope of this chapter. In its place, the following two sections examine examples of distributed human-computer co-creativity in two different domains: human-computer co-exploration and interactive performance systems. The latter are discussed with respect to their use within interactive compositions and its implications for the compositional process and its product.

2.6.1 Human-Computer Co-exploration

McCormack (2012) views creativity as a process of exploration and search within spaces of possibilities. In human-computer co-exploration, this exploration of creative spaces (e.g., all possible settings of a sound synthesis algorithm or outputs of a generative music system) is assisted through computational means. The purpose of human-computer co-exploration is to facilitate creative discovery and enhance the artist's ability to think beyond their established creative habits. The role of computational creativity in this process is to generate outputs that the user/artist might otherwise not have created, guiding the creative process towards new paths. For instance, Jones, Brown and D'Inverno (2012) describe a co-exploration approach in which the artist makes high-level aesthetic decisions and curates computer-generated material. In addition to breaking creative habits, this approach is meant to facilitate reflection on the artist's own artistic practice and aesthetic stance.

The type of human-computer co-creativity involved in co-exploration tasks is concerned mainly with transforming the creative process, the product of which might fall within already existing paradigms (e.g., a fixed-media composition). In this approach, the computer functions as a 'compositional prosthesis' (Impett 2000, 31), i.e., an extension of the (human) composer, who defines the parametric space within which the algorithm can generate outputs and curates (i.e., selects and further refines) computer-generated material. While in compositional approaches based on human-computer co-exploration computational creativity is primarily explorative and human creativity is mainly evaluative, creativity is nevertheless distributed between human and machine agency.

An example of a co-exploration tool capable of learning from and adjusting its outputs to user preferences is *Sonic Xplorer*. *Sonic Xplorer* uses Neural Networks to build correlations between six different adjectives ('warm', 'bright', 'stable', 'thick', 'noisy' and 'evolving') and four perceptual audio descriptors based on examples provided by the user (Tsiros 2017). After training the system, the user can use six sliders, each corresponding to one of the adjectives, to describe the qualities of the sound they want to generate. By transitioning between the *Sonic Xplorer* interface and the parameters of the synthesis engine, users can fine-tune and further experiment with the generated sounds.

Scurto, Bevilaqua and Caramiaux (2018) describe a similar approach to human-computer co-creativity, based on reinforcement learning. In their study, participants were asked to evaluate their collaboration with software agents in the completion of a closed-ended task. The task involved the exploration of a Virtual Studio Technology (VST) with the purpose of finding the parameter settings that produce the brightest sound possible. At each iteration the agent would produce a sound and receive positive or negative feedback by the user, based on whether the new sound was brighter than the previous one. This study involved a synthesis task with only two discrete control parameters and a predefined goal (i.e., finding the brightest sound possible) and is therefore not representative of the high-dimensionality of real-world synthesis applications, or the open-ended nature of creative tasks. Nevertheless, it shows the potential of co-exploration tools to assist the creative process and expand human creativity.

While both of the examples mentioned above involve synthesis processes, human-computer co-exploration can also be applied to larger-scale generative processes, e.g., to the generation of musical textures. The open-ended nature of the creative process, the subjective nature of aesthetic judgments and goals and the high-dimensionality involved in real-world creative exploration tasks suggest that there are still significant challenges to be overcome in sonic human-computer co-exploration. However, early experiments in this field demonstrate significant potential for future applications.

2.6.2 Interactive Music Systems

Interactive music systems are computer music systems that can sense their environment, by collecting and interpreting sensing data, make decisions and act both in response to human (or non-human) actions and independently of them (i.e., as a result of autonomous generative processes). While the majority of interactive music systems are designed for human-computer improvisation (Gioti 2017), this section will focus on their application in composed music and its implications for musical authorship and the musical work.

Doug van Nort's (2018) *Genetically Sonified Organisms* (GSOs) is an example of such an application. *Genetically Sonified Organisms* is a work of environmental sound art based on a set of artificial agents ('artificial creatures'), capable of interacting with and adapting to their acoustic environment. Each of these agents is equipped with a vocabulary of twenty sounds produced using physical modeling synthesis to imitate the sounds of animals that inhabit the site of the installation. The *GSOs* analyze and compare incoming sounds to their vocabulary and, using a nearest neighbor approach, match them to the closest (i.e., the most similar) synthesis model. Rather than instantly converging to their acoustic environment, the agents update their synthesis parameters incrementally, coming a bit closer to the input sounds with each response. This process of interaction with and convergence to a continuously changing acoustic environment is responsible for the evolution of the work over long periods of time.

Similarly, in Jennifer Walshe's and Memo Akten's *Ultrachuck* a Neural Network interacts with a vocal performer (Walshe) based on processes of listening and learning. The Neural Network was trained by Akten using video recordings of solo vocal improvisations performed by Walshe over the period of one year and interacts with her in a live performance setting, functioning as her 'AI doppelganger' (Akten 2018). The algorithm generates audio and video frames in real-time, while listening and responding to Walshe's improvised performance.

The interactive compositions described in this dissertation involve similar processes of interaction and adaptation between software agents and (human) musicians. In these works, composing does not entail designing concrete structures of sounds, but rather interaction affordances and action spaces, the exploration of which by the musicians and the computer music system during the performance can lead to varied musical outcomes. As a result, the act of musical interpretation is expanded to include real-time decision-making and adaptation to a non-human partner, a premise that challenges the composition/improvisation binary. This compositional approach is based on a distributed and posthuman notion of agency and views creative responsibility as dispersed in time ("offline" compositional decisions vs. real-time decision-making during the performance) and across actors (composer, performer and interactive music system). As a result of this shift towards distributed notions of agency and creativity, both the compositional process and its product (i.e., the work itself) are effectively redefined and re-conceptualized.

2.7 Conclusions

Xu, Wang and Bhattacharya (2010) argue that design research on artificially intelligent systems has focused primarily on goal-oriented problem-solving, ignoring the problem creation phase that should precede problem-solving and addressing only the *how* and not the *why* of the design process. A similar approach seems to prevail in research on automatic composition systems: composition is considered as problem-solving – the "problem" being one of style imitation – rather than problem creation. As a result, autonomously creative music systems tend to produce outputs with limited aesthetic value and virtually no innovation potential. This is not to say that research on automatic composition systems is not valuable: indeed, it can provide interesting insights into creative processes and help us understand and appreciate the complex phenomenon that is musical creativity. It is not the epistemic value or relevance of this research that is questionable, but its aesthetic potential.

As far as the latter is concerned, the disparities between human and computational creativity seem to suggest that an 'ecosystemic' (McCormack 2012) approach to musical creativity, encompassing both humans and machines, holds significantly more potential than approaches aiming to automate or simulate human creativity. In human-machine co-creative networks high-level aesthetic decisions can be made by humans, while computational intelligence can be used to enhance human creativity, by assisting creative exploration and discovery, or enabling novel forms of human-technology interaction that lead to new artistic concepts and practices. The rationale behind this approach is that artistic production can benefit from a synergetic – rather than a competitive – relationship between human and computational creativity.

3 Experiments in Machine Listening

Neurons, for Soprano Saxophone and Interactive Music System

3.1 Introduction

This chapter describes *Neurons*, a composition for soprano saxophone and Interactive Music System (IMS) incorporating machine learning.[7] The work involves real-time interaction and mutual adaptation between the saxophonist and the computer music system and entails four interaction scenarios, which are distinct in terms of sound material and interaction affordances.

The computer music system incorporates a feedforward Neural Network trained to recognize four different playing techniques: single tones, multiphonics, air tones and slap tones. This allows the computer to interact with the musician on the basis of symbolic music information (a dictionary of sound classes defined by the composer), obtained through lower-level sensory information (signal-level descriptors). The results of this instant recognition are stored, enabling the IMS to deduce information regarding the timbral variability of larger sections of the performance. Information collected in the auditory processing stage of the system informs its decision-making stage, influencing its responses to the musician's live input (Figure 3.1). The musician adapts to the sound output of the computer in real-time, by interpreting a non-linear score.

[7] A video documentation of the piece is available at:
https://www.artemigioti.com/demos/Neurons.html.

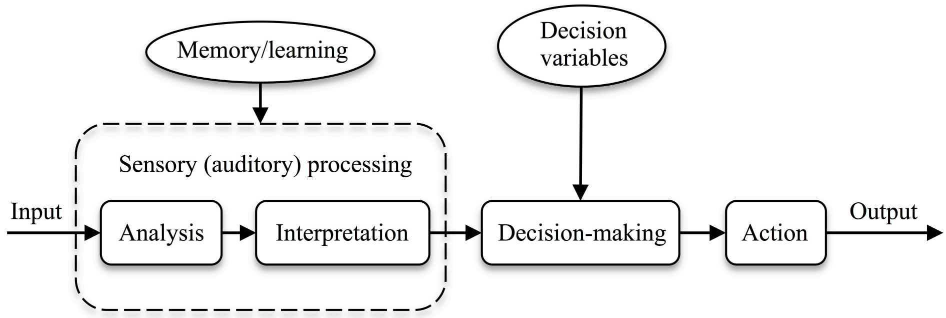


Figure 3.1 *Neurons*: architecture of the interactive music system.

While the auditory processing stage of the computer music system is based on a classification task, the score of the piece largely explores the spaces in-between the categories (i.e., playing techniques) recognized by the machine learning algorithm. Transitions from air tones to pitched tones or from single tones to multiphonics, and unstable multiphonics are used to challenge the categorical divide that forms the basis of the classification task performed by the computer and push the recognition process to its limits. Categorically ambiguous sounds (e.g., transitions from air tones to pitched tones) and "unstable" sounds lead to both human and computational – i.e., classification – errors and affect the timbral variability measure calculated by the computer, shaping the form of the performance.

The starting point for this composition was a series of machine learning experiments aimed at exploring the compositional potential of machine learning-based approaches to timbre recognition. The recognition of different timbral categories served as the basis for designing various listening and attentional strategies that explore the role of listening as an agentive process in the context of musical performance.

3.2 The Neural Network

The composition described in this chapter employs a supervised machine learning algorithm, particularly a feedforward Neural Network (NN), trained to recognize five different sound classes: single tones, multiphonics, air tones, slap tones and background noise (i.e., silence).

The training data for the NN was collected in four recording sessions with the help of saxophonists Joel Diegert and Matej Bunderla. The recording sessions were conducted in two different rooms with microphones of different directionality (one super and one hyper-cardioid) and placement (clip-on and stand respectively). Each of the musicians used a different soprano saxophone. Collecting samples from more than one musicians/instruments and recording setups aimed at ensuring adequate variability in the training set and avoiding overfitting (the problem of a machine learning algorithm fitting the training set very well, but failing to generalize on previously unseen examples). For the same reason, initially synthetic data was generated by applying filters and artificial reverberation to some of the recorded examples. However, the use of synthetic data did not seem to improve the performance of the NN and was abandoned later in the training process.

The recorded examples were edited manually to remove any ambiguities that could lead to data mislabeling (e.g., unstable multiphonics) and analyzed using a window size of 2048 samples and 50% hop size. The data set was partitioned into three separate sets: a training set consisting of 23889 examples (about 60% of the data set), a cross-validation and a test set (each about 20% of the data set). Each example consisted of a feature vector and a label between 1 and 5 (e.g., 1 for single tones, 2 for multiphonics etc.). The feature vector included 13 Mel Frequency Cepstral Coefficients (MFCCs)[8] and a few additional features, such as spectral flatness, onset, pitch variation (ratio between the current and previous frequency value) and frequency beats. The latter is a binary-valued feature used to signal amplitude periodicities that are indicative of interference between frequency components of a multiphonic.

The NN consisted of an equal number of input and hidden units using the logistic sigmoid as an activation function and was trained using backpropagation. During the training process several feature sets were tested and evaluated both on a separate test set and live with the collaboration of saxophonist Joel Diegert. These run-time tests were crucial to the development process, since they helped identify the weaknesses of the algorithm and provided valuable feedback for the ongoing training process. Each time the network consistently failed to identify certain examples, similar examples were recorded

[8] Mel Frequency Cepstral Coefficients (MFCCs): the coefficients of a Mel-Frequency Cepstrum, which is used to analyze periodical structures in a frequency spectrum. In a Mel-Frequency Cepstrum, frequency bands are spaced on a Mel-frequency scale, which approximates human perception of frequency. MFCCs are timbral descriptors often used in speech recognition and Music Information Retrieval tasks such as instrument classification.

and added to the training test, or new features were added to the feature vector and the training process was repeated (Figure 3.2).

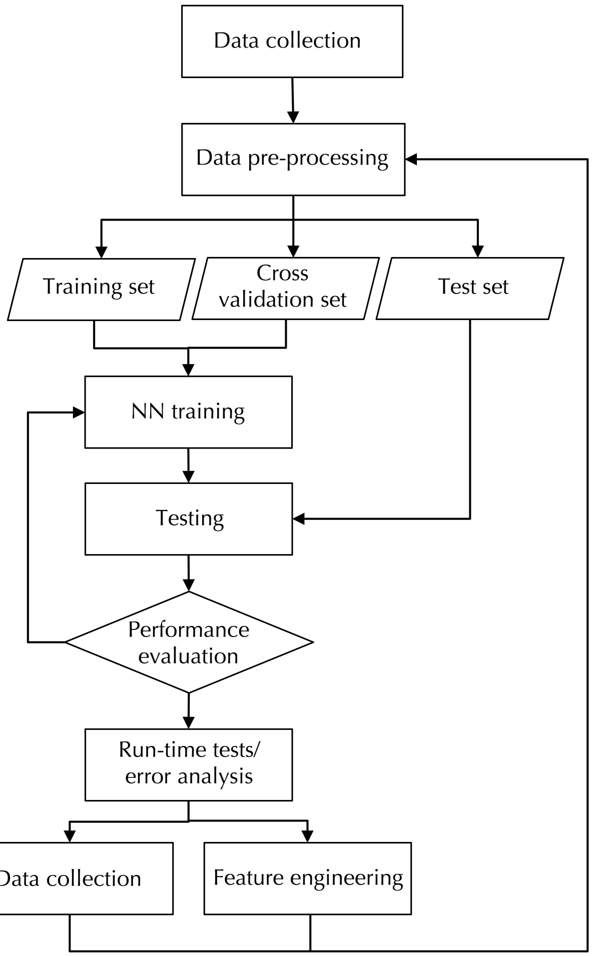


Figure 3.2 Neural Network: data collection and training.

For instance, in one of these run-time tests, the NN seemed to only recognize single tones in mid- and high-range dynamics and consistently misclassified quieter tones (*pp* and below) as multiphonics. A closer analysis of this error revealed that the reason for this was that the training set contained a large number of quiet muliphonics (i.e., multiphonics that can only be played using low air pressure), but very few single tones in similar dynamics. This problem was resolved by recording more single tones in low-range dynamics and adding them to the training set.

One of the main challenges of the training process was finding a workaround for polyphonic pitch detection. Several polyphonic pitch detection tools were tested and rejected due to their poor performance. Instead, fluctuations in the detected pitch values (resulting from the presence of more that one pitches) and beat frequencies were used to facilitate the recognition of multiphonics. However, the detection of beat frequencies made use of a larger FFT window size, resulting in a delay in the detection of multiphonics.

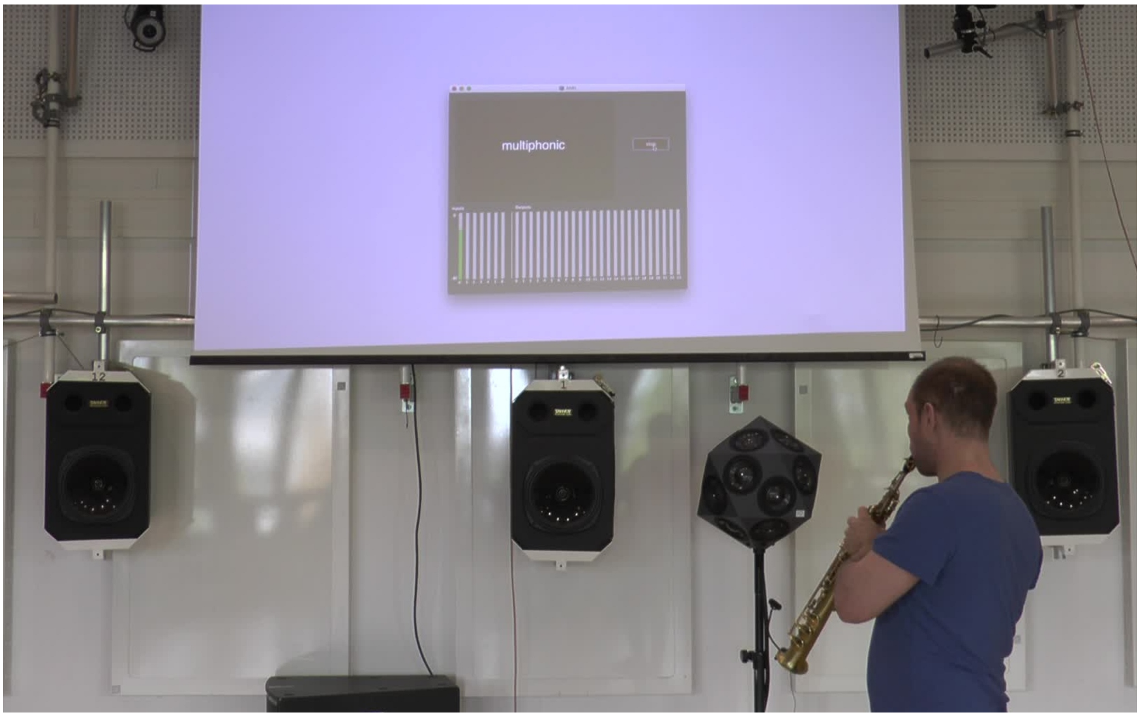


Figure 3.3 Saxophonist Joel Diegert testing the Neural Network in real-time.

After training, the accuracy of the network on the test set reached 91%.[9] In order to further improve the performance of the algorithm in run-time, two common machine learning strategies were explored: averaging the predictions of the network over a certain time span (e.g., averaging every 2-5 predictions) and filtering the output of the network based on its confidence (i.e., outputting only predictions with a probability higher than a certain threshold). The first method made the system less flexible by increasing its response time, a weakness particularly noticeable in denser musical textures, and was therefore rejected. The second method improved the performance of the network significantly, by filtering out some false predictions and increasing its overall accuracy. An additional gain from the use of the confidence filter was the integration of sound source separation in the recognition process. Concretely, the NN seemed to only output

---

[9] A video demonstration of the machine listening algorithm is available at: http://www.artemigioti.com/demos/soprano_sax_sound_event_recognition.html.

predictions for the four classes it was trained to classify, "ignoring" any other sounds (i.e., the electronics).

3.3 Interaction Scenarios

In *Neurons*, the classification algorithm described in the previous section is embedded in various interaction scenarios, entailing different sonic interaction affordances and performance instructions. Two of Truax's (2001, chap. 2) levels of aural attention ('listening-in-readiness' and 'listening-in-search') are referenced as metaphors for some of the listening modes involved in these scenarios.

3.3.1 Scenario 1: Listening-in-readiness, Listening-in-context

In this scenario, the occurrence of each of the four sound classes (single tones, multiphonics, air tones and slap tones) causes a different response (*listening-in-readiness*). Single tones and slap tones trigger textures of synthesized sounds, air tones are processed by a signal processing chain and multiphonics are resynthesized using a "spectral freeze" effect.

In parallel to this instant recognition process, a measure of timbral variability is calculated every second, providing information on the variability/uniformity of the sound material played by the saxophonist over the last ten seconds. The value of estimated timbral variability is used to control the amplitude of low frequency components of the electronics, which become louder as timbral variability increases. When timbral variability reaches a certain threshold value, the system temporarily switches off its input and enters a non-listening state. The musician is instructed to make sure that the system does not enter the non-listening state before indicated in the score. When the low frequency components of the electronics become considerably louder, the performer has to intervene by taking some control action (i.e., playing less variable sound material, or introducing rests). Air tones are ignored by the variability measure and can also be used as a regulatory measure, in order to prevent the system from entering the non-listening state.

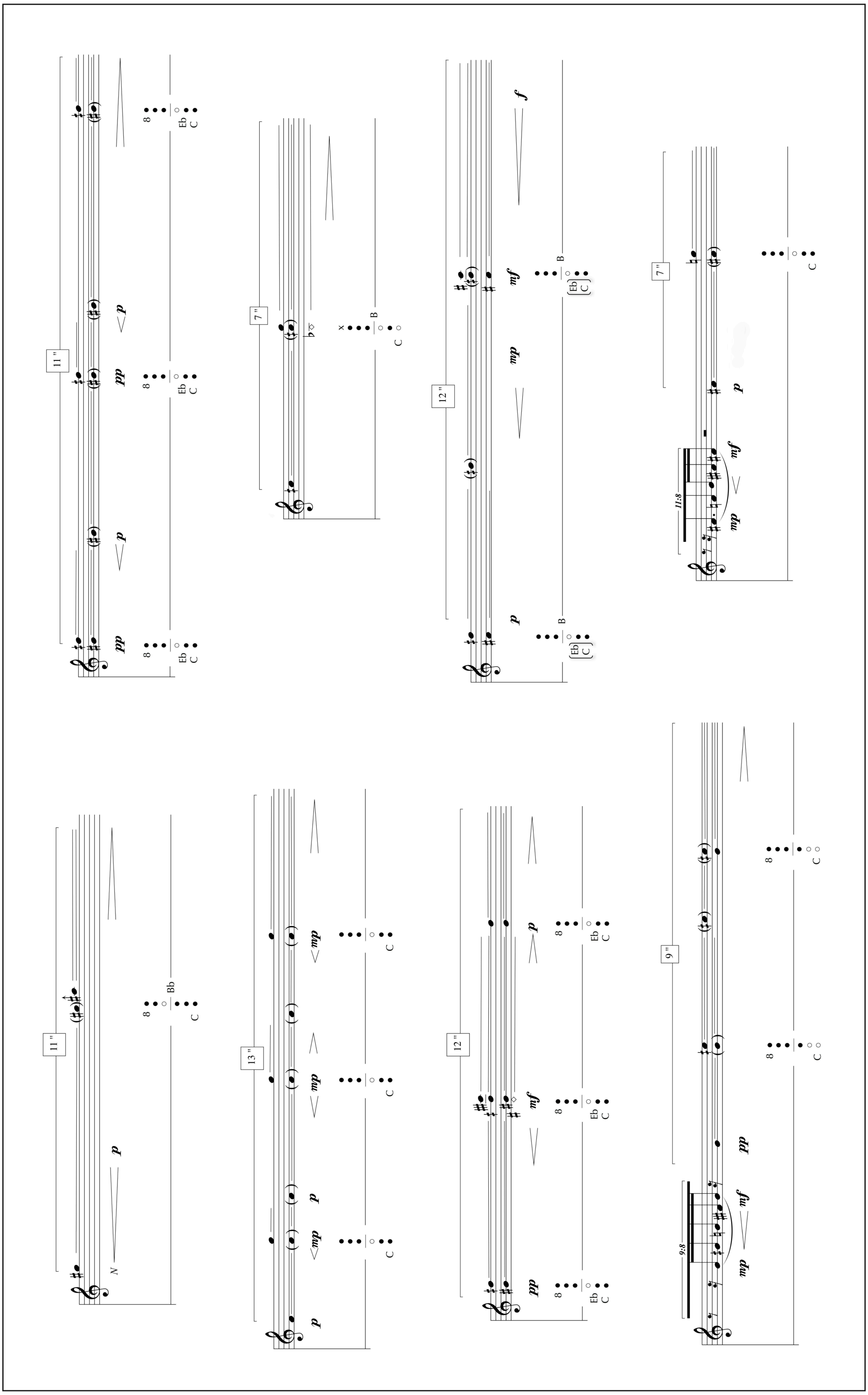


Figure 3.4 *Neurons*: score excerpt.

In this scenario, the value of timbral variability is key to the interaction between the musician and the computer. The value of this measure can increase both due to classification errors and due to "human error" (e.g., an unsuccessful execution of an unstable multiphonic, in which the second pitch is hard to obtain). This part of the score contains a large number of such unstable multiphonics, along with other material, organized in short fragments the order of which is left to the performer (Figure 3.4).

3.3.2 Scenario 2: Non-listening State

In the non-listening state, the sound output of the computer is controlled exclusively by algorithmic processes involving sound synthesis and feedback. During this part of the piece, the musician is instructed to stop playing and wait for an auditory cue signaling that the computer music system is listening again.

3.3.3 Scenario 3: Listening-in-search, Listening-at-will

In scenario 3, the input of the machine listening algorithm is switched on and off in search of multiphonics. The IMS randomly "chooses" when to switch its input on (*listening-at-will*) and provides the musician with an auditory cue when doing so. The term "at-will" in this context is suggestive of the changed dynamics of the interaction between the musician and the software agent, rather than "free will": by choosing when to listen, the IMS transitions from a mostly reactive to a proactive role (Table 3.1).

When the IMS indicates that it is "listening", the musician responds by playing a single multiphonic, selected from a pool of multiphonics provided in the score. If the execution of the multiphonic is evaluated as "stable" by the computer, the algorithmic synthesis processes initialized in the previous scenario are temporarily interrupted by a "spectral freeze" effect. The musician is instructed to repeat this process until the computer responds (i.e., until the multiphonic is "stable" enough). Sound events other than multiphonics (e.g., single tones, air tones etc.) are ignored by the IMS (*listening-in-search*).

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Performer	Proactive	Inactive	Reactive	Proactive
Computer	Reactive	Proactive	Proactive	Reactive

Table 3.1 Interaction dynamics between the musician and the computer in the 4 interaction scenarios of *Neurons*.

3.3.4 Scenario 4: Listening-in-search

In scenario 4, the computer responds only to air tones. In addition to signal processing, the detection of air tones in this mode triggers the playback of resynthesized spectra of multiphonics played by the saxophonist earlier in the performance. In this interaction scenario, the performer can choose from a number of actions in the score, which can be executed in any order, one or more times.

In the interaction scenarios described above, different listening and attentional strategies create distinct interaction affordances and musical action spaces. The *selective listening* and *non-listening* modes constitute behavioral elements of the computer music system, which exhibits varying degrees of autonomy and responsiveness (Table 3.1). The concept of (human and computational) "error" is also explored for its potential to produce idiosyncratic sonic interactions. For example, in scenario 3, the computer music system evaluates the "stability" of the execution of multiphonics and responds only when a multiphonic is "stable", while in scenario 1, classification errors can cause the system to enter a non-listening state.

The listening modes of the IMS are part of idiosyncratic, composition-specific interaction scenarios, delineated both by the interaction affordances of the system and the performance instructions. In scenario 3, the computer music system "listens-in-search" of multiphonics and ignores any other sound events. Similarly, in scenario 1 the musician "listens for" increases in the amplitude of low frequency components of the electronics, signaling that the IMS is about to enter a non-listening state. These sounds carry extrinsic information related to the compositional idea and the rules of the interaction between the musician and the IMS and are anticipated by the performer – hence "listens for" – as part of that interaction scenario.

3.4 Summary

This chapter described a composition for soprano saxophone and interactive music system using supervised learning to perform real-time recognition of different playing techniques. The integration of machine learning in the auditory processing stage of the IMS had the purpose of shifting the focus of machine listening from *analysis* to *interpretation* and from 'sensory' to 'symbolic' (McAdams and Bigand 1993) information. The term *sensory information* is used here to denote signal-level features extracted in the analysis stage of the system, while *symbolic information* refers to higher-level representations of the human input (in this case, the four classes recognized by the NN), obtained by interpreting analysis data. This approach aimed at integrating composer-defined sound classes in the listening task and allowing for the design of idiosyncratic agentive behaviors and interaction scenarios.

The research objective of this work was to explore the compositional potential of machine learning-based approaches to timbre recognition. The NN served as a basis for the design of idiosyncratic machine listening strategies beyond one-to-one input-output mappings, including 'listening-in-search' of specific timbral categories and basing long-term musical decisions on the degree of timbral variability of larger musical textures.

In *Neurons*, the concepts of "error" and categorical ambiguity are exploited compositionally, questioning the categorical divides that are inherent to any classification task. Categorically ambiguous sounds, classification errors and "errors" in the execution of unstable multiphonics affect the value of the timbral variability measure calculated by the IMS, influencing the course of the performance and causing both the saxophonist and the computer music system to adapt through real-time decision-making.

4 Experiments in Open Form and Aesthetically Driven Decision-making

Imitation Game, for Human and Robotic Percussionist

4.1 Introduction

Musical robotics is a fast expanding research field, covering a wide range of musical instruments, from percussion to string and wind instruments (Kapur 2005), as well as interaction paradigms: from interactive musical robots to laptop orchestras (Kapur et al. 2011). The growing interest for musical robotics can be attributed to both sound- and interaction-driven design and compositional choices. The complexity of acoustic sound, the expressive potential of physical actions and the role of visual communication in anticipating and coordinating performers' actions, as well as establishing cause-effect relationships are some of the most commonly cited advantages of musical robotics over electronically produced sound (Weinberg and Driscoll 2006, 28; Weinberg 2007, 423).

Research in musical robotics encompasses a large variety of applications, from 'robotic musical instruments' played by human musicians or triggered by predetermined sequences, to 'anthropomorphic musical robots' designed to imitate (physical) human actions, and 'perceptual robots' (Weinberg and Driscoll 2006; Weinberg, Driscoll, and Parry 2005). The last category refers to autonomous musical robots able to perceive and interact with their sonic environment, suggesting an overlap with the field of interactive music systems.

The musical work described in this chapter falls under the latter category, incorporating both hardware components and software agency. *Imitation Game* is an interactive composition for human and robotic percussionist based on a dynamic form, which is shaped by decisions made by both the musician and the robotic percussionist in real-time. The robotic percussionist interacts with the human based exclusively on machine listening, particularly a feedforward Neural Network trained to recognize different instruments and playing techniques. Decisions are made by the robotic percussionist both on a meso and macro time scale, based on metrics of rhythmic, timbral and dynamic contrast.[10]

4.2 Interaction, Agency and Musical Meaning in Human-Robot Musical Interactions

4.2.1 Interaction Modes

Most interactive music systems – whether hardware of software-based – incorporate one or more interaction 'modes' (Weinberg and Driscoll 2006) or 'modules' (Hoffman and Weinberg 2011), which entail specific sonic interaction affordances. In the case of interactive robotic percussionists, these modes can differ with respect to rhythmic material, interaction timing (e.g., synchronous vs. asynchronous interaction) and/or the sensory processing and decision-making processes involved in them.

For example, *Haile* is a perceptual robot equipped with six different interaction modes, some of which are synchronous and some sequential (Weinberg and Driscoll 2006). These modes are not selected in real-time, but are activated in predetermined sequences. Real-time decision-making processes are employed mainly on the phrase level: the robotic percussionist calculates the stability of an input rhythm and then chooses from a database of rhythms based on similarity metrics and a target stability value (Weinberg, Driscoll, and Parry 2005). Another perceptual robot, *Shimon*, is based on three interaction modules, i.e., segments with a fixed or condition-dependent duration (Hoffman and Weinberg 2011), while the *CIM software* is based on a model of duet interaction centered around six different types of musical activity: 'imitate', 'initiate', 'loop', 'restate', 'shadow' and 'silence' (Brown, Gifford, and Voltz 2017).

---

[10] A video documentation of the piece is available at:
http://www.artemigioti.com/demos/Imitation_game.html.

4.2.2 Generating Meaningful Responses

The integration of different interaction modes and complex decision-making processes in the above mentioned systems is indicative of an interaction design oriented towards a 'conversational' (Paine 2002) – i.e., *reciprocal* – model of interaction, rather than one based on cause-effect relationships. Emmerson (2013) distinguishes between 'causing' a reaction and 'provoking' a response – particularly a 'meaningful response' – using the example of two musicians improvising in a call-and-response fashion as a model for the second (2-3). However, as he points out, 'meaningful' is a musical judgment (2).

In interactive musical robotics, musical meaning is – not unjustifiably – linked to 'higher-level percepts' (Weinberg 2007) and subjective concepts. What Weinberg (2007) refers to as 'higher-level percepts' are musical meta-parameters (e.g., metrics of rhythmic stability, melodic similarity etc.), which are used to describe the meso and macro time scale, rather than the sound event level (425). Meaning is, therefore, not only subjective but also context-dependent. Furthermore, these 'higher-level percepts' are in most cases specific to the instrumentation, the musical idiom and/or the compositional idea.

4.2.3 Can the Computer Say "No"?

Another key distinction between a reciprocal interaction based on decision-making processes and a mere input-output mapping is that of *intention*, as well as *negotiation of intentions* between actors. Or, as Emmerson (2013) puts it: can the computer say 'no, thanks' (3)? A behavior that is strictly reactive and not pro-active falls under causality, rather than interactivity. A meaningful response does not entail just following, but also leading, a behavior that is often incorporated in the decision-making stage of interactive music systems (e.g., Weinberg 2007; Lewis 2000).

4.3 Imitation Game

Notions of musical intention and meaning – particularly a meaning that is constructed through context (i.e., on a meso and macro time scale, rather than on a sound event level) – are some of the central concepts explored in *Imitation Game*. This meaning is not universal, but composition-specific and constructed – *composed* – based on the composer's subjective criteria.

Auditory processing in *Imitation Game* extends beyond the sound event level (instrument and playing technique recognition), to the phrase level (calculating metrics of musical contrast) and form level (monitoring the evolution of contrast metrics over time). Similarly, decision-making extends beyond the selection of single actions to the initiation of various interaction scenarios, in which the agent assumes different roles (e.g., following and leading). The auditory processing, decision-making and action stage of the robotic percussionist in *Imitation Game* are described in detail in the following sections.

4.3.1 Auditory Processing

The auditory processing stage of the robotic percussionist is based on a feedforward Neural Network (NN) trained to recognize different instruments (cymbals, bongos and cowbells) and playing techniques (strokes, scraping and bowing). In order to train the NN, several examples of each class were recorded using a large number of different mallets and various microphones to ensure variability in the data set and prevent overfitting. The recorded examples were analyzed using a window of 2048 samples and 50% hop size (sampling rate: 44100 Hz) and divided into three sets: a training set (60% of the data set), a cross-validation and a test set (each 20% of the data set). The final set of features used for machine learning was selected through an iterative process of training and testing and consists of the following features: onset, spectral centroid, spectral spread, spectral slope, spectral flatness, spectral roll-off and Mel Frequency Cepstral Coefficients (MFCCs).

The strategy used in approaching this classification problem included testing various approaches, such as breaking the task down to two classification problems (e.g., using one NN for instrument recognition and another for playing technique recognition).

Both single-label classification (assigning a single label to each sample) and multi-label classification (assigning multiple labels to a single sample, e.g., an instrument and a playing technique label) performed equally well on a balanced training set (i.e., a training set in which none of the classes are significantly over- or underrepresented). Eventually, single-label classification was preferred over multi-label classification due to its practical advantages (e.g., lower computational cost in run-time).

In its final form, the NN consisted of one hidden layer with an equal number of units as the input layer and 11 output units corresponding to the following classes/labels: "bongo, stroke", "cymbal, stroke", "cowbell, stroke", "bongo, scraping", "cymbal, scraping", "cowbell, scraping", "cymbal, bowing", "cowbell, bowing", "cymbal, resonance", "cowbell, resonance" and "background noise". Background noise was added as a separate class in order to integrate noise gating in the classification task. The activation function used was the logistic sigmoid.

The accuracy of the NN on the test set reached 85%, with one of the main weaknesses of the algorithm being the low accuracy of the onset detection algorithm[11] on cymbal strokes, presumably due to the characteristic envelope shape of the instrument (slow attack). Finally, a confidence threshold was introduced to filter out some false predictions and improve the overall accuracy of the algorithm.

4.3.2 Decision-making

The decision-making stage of the robotic percussionist processes data collected in the auditory processing stage and chooses among three different interaction scenarios:

(1) *repeat* (play the exact same material as the human percussionist),

(2) *imitate* (play similar material to that played by the human) and

(3) *initiate* (introduce new sound material).

The terms *imitate* and *initiate* were borrowed from Brown, Gifford and Voltz (2017) and adapted to describe specific interaction scenarios used in the composition.

[11] "Onsets" SuperCollider UGen (Collins 2011b) using the rectified complex deviation onset detection function (Stowell and Plumbley 2007).

Particularly, *imitate* is used to refer to the generation of similar material, using high-level percepts such as rhythmic contrast as similarity measures, rather than the reuse of material within a short time frame (Brown, Gifford, and Voltz 2017, 3).

It has been suggested that musical changes are key to designing meaningful musical interactions (Ravikumar, McGee, and Wyse 2018; Young 2008, 339). This is presumably because the ability of an interactive music system to propose changes (e.g., introduce new sound material) is indicative of a high level of music understanding, as well as a high level of autonomy. In line with that view, interaction scenarios in *Imitation Game* are not selected randomly by the robotic percussionist, but based on metrics of *rhythmic, timbral* and *dynamic* contrast, which are calculated as follows:

- Rhythmic contrast: standard deviation of (detected) Inter-Onset-Intervals (IOIs).
- Timbral contrast: standard deviation of the (detected) timbre probability distribution (where timbre x is treated as a random variable that can take 8 possible values: "bongo, stroke", "cymbal, stroke", “cowbell, stroke” etc., excluding background noise).
- Dynamic contrast: standard deviation of the (detected) dynamics probability distribution (where “dynamic” x can take 3 possible values: p, mp/mf and f).

These contrast metrics are calculated on a phrase basis and their values are stored in arrays, allowing the robotic percussionist to make decisions based on their evolution over time. Specifically, if the estimated rhythmic contrast has been constant (i.e., around the same value), or monotonic (i.e., constantly increasing or decreasing) for the last few phrases, the robotic percussionist is less likely to follow the musician’s lead (“imitate”) and more likely to introduce new, contrasting sound material (“initiate”).

From the three interaction scenarios mentioned above, “imitate” and “initiate” are based on a call-and-response interaction, while “repeat” is the only scenario entailing synchronous action (i.e., both the human and the robotic percussionist playing simultaneously). In this scenario, auditory processing and decision-making are based on short- rather than long-term memory functions. Instead of calculating contrast metrics and generating responses on a phrase level, the robotic percussionist interacts with the musician on a sound event level, freely repeating some of the actions performed by the musician. The conditions for initiating this scenario are not dependent on contrast metrics, but a record of past scenarios kept to ensure that it is not repeated too often.

The musician can alternate among the same scenarios as the robotic percussionist, while “navigating” a non-linear score that consists of both descriptive and prescriptive notation. The composed fragments/phrases used in the “imitate” and “initiate” scenarios are organized in three concentric rectangles according to pre-calculated contrast metrics as follows:

- *From the center outwards:* in order of decreasing rhythmic contrast,
- *From the center upwards:* in order of decreasing timbral contrast, with strokes being the predominant playing technique,
- *From the center downwards:* in order of decreasing timbral contrast, with scraping being the predominant playing technique.

This “topological” organization of the sound material facilitates real-time decision-making and interaction, allowing the musician to adapt to the robotic percussionist’s actions (Figure 4.1).

The material used in “repeat” is less thoroughly notated. In this scenario, instead of playing composed musical phrases, the musician is instructed to improvise on a set of notated actions with variable or open instrumentation and duration. This scenario has two variations depending on “who” is leading the improvisation: the musician or the robotic percussionist. In the former case, the musician can improvise freely, while in the latter, they are instructed to structure their improvisation around repetitions of the robotic percussionist’s actions.

The beginning and end of the piece are fixed and based on two differentiated instances of “repeat” in which the musician is leading and the robotic percussionist is following. Concretely, the beginning of the piece is based on a mapping of the amplitude of the musician’s input (bowing on the cymbal) to the frequency of two computer-controlled electromagnets and has the character of an instrumental interaction, rather than an interaction with an autonomous agent. The ending sequence of the piece, which is initiated by the robotic percussionist, is based on a repetition of detected strokes (onsets) initially with a variable delay, which is progressively reduced until only the latency of the onset detection algorithm and the actuation mechanism remains.

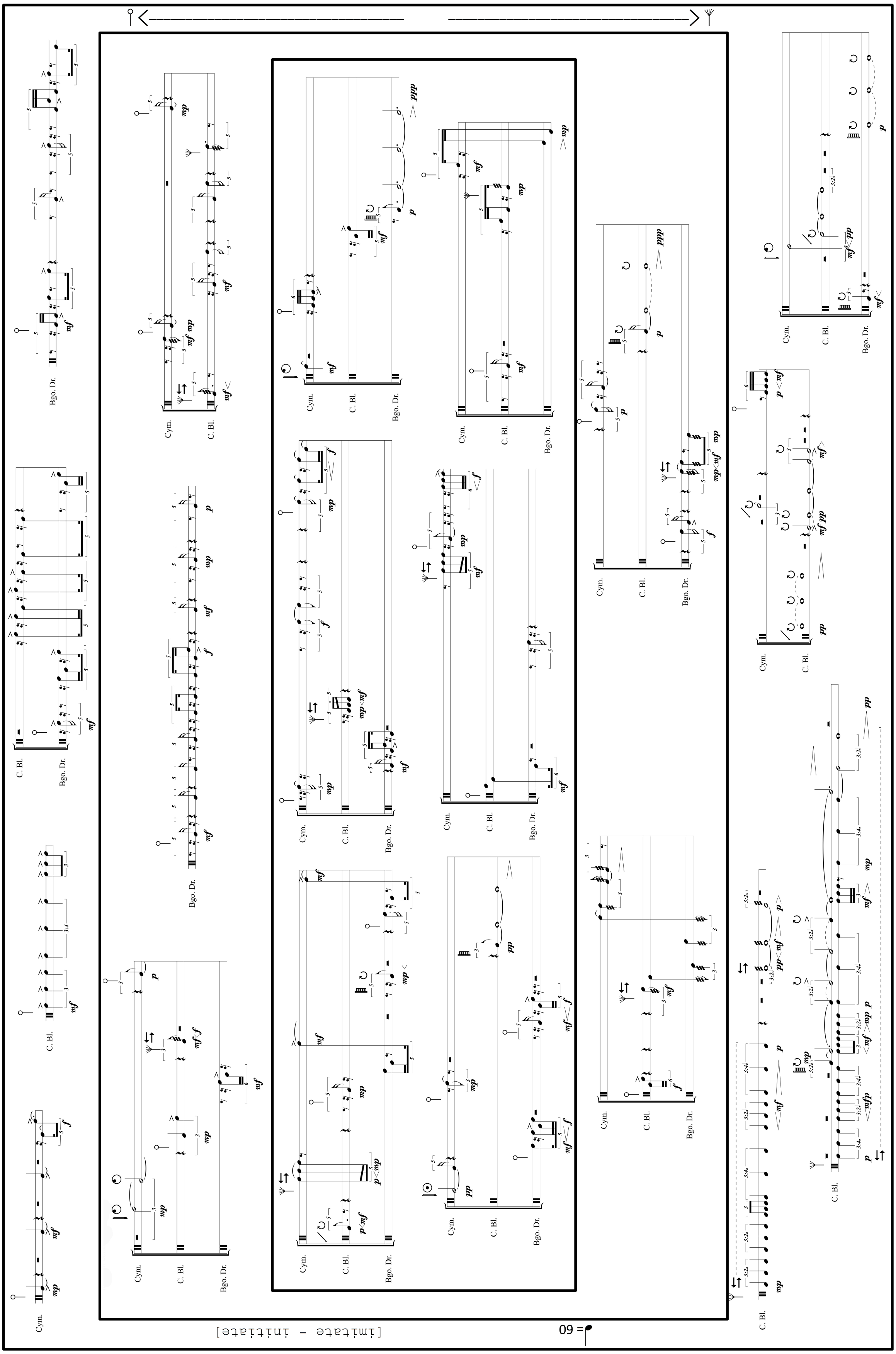


Figure 4.1 *Imitation Game*: score excerpt.

Figure 4.2 *Imitation Game*: instrument and stage setup.

4.3.3 Action

The responses generated by the robotic percussionist are based on pre-composed sequences of inter-onset-intervals. The specific actions (instrument and playing technique) to be performed and their durations are chosen on the fly based on the current scenario (i.e., according to whether the current response is an "imitation" or an "initiation", the robotic percussionist might choose similar or different actions than those performed by the human).

The actions employed by the robotic percussionist include strokes and scraping and are implemented through the use of servo-motors, controlled by an Arduino UNO micro-controller, and two permanent magnets suspended over one of the cymbals and set into motion by two computer-controlled electromagnets, which are placed directly underneath the cymbal.

4.4 System Autonomy and Responsiveness

Decision-making in *Imitation Game* is centered around two seemingly contradictory agent attributes:

- *Responsiveness* or reactivity: an agent's ability to act in response to its environment, including human actions, and
- *Autonomy*: an agent's ability to act independently of human actions.

Arguably, balancing responsiveness and autonomy is a key factor and, at the same time, a major challenge in designing meaningful sonic human-computer interactions (Brown, Gifford, and Voltz 2017, 5–6). A high degree of responsiveness coupled with a low degree of autonomy is associated with linear input-output mappings and therefore cause-effect relationships, rather than complex decision-making processes. Conversely, high autonomy and low responsiveness are suggestive of erratic, rather than intelligent behavior. Balancing agent responsiveness and autonomy is therefore key to designing intelligent behaviors – or at least behaviors perceived as such.

4.4.1 "Imitate": Establishing System Responsiveness

In *Imitation Game*, system responsiveness is established through the "imitate" mode. The robotic percussionist's ability to play similar material to that played by its human counterpart (e.g., by choosing similar rhythms, instruments and playing techniques) suggests that the agent is not only collecting auditory information, but also interpreting it in a musically meaningful way (i.e., understanding human/musical concepts such as instrument and playing technique categories), while confirming that the agent is in fact responding to the human percussionist and not acting based exclusively on generative processes.

4.4.2 "Initiate": Establishing System Autonomy

Along with responsiveness, the robotic percussionist also exhibits a high degree of autonomy, demonstrated mainly in the "initiate" scenario. Based on a complex decision-making process involving aesthetic criteria (i.e., musical contrast), the robotic percussionist might choose to steer the interaction in a different direction, by introducing new sound material.

4.4.3 "Repeat": Introducing Musical Tension

"Repeat" differs from the other two interaction scenarios both with respect to interpretative freedom (improvised vs. notated material) and interaction timing (synchronous vs. asynchronous interaction). The higher density of sound events (2 "voices" instead of one) and the condensed call-and-response intervals that are characteristic of this scenario function as a source of musical tension in the piece.

Overall, the wide range of behaviors exhibited by the robotic percussionist is suggestive of an *instrument-agent continuum* – rather than a dichotomy – in which system responsiveness and autonomy are alternately established and questioned.

4.5 Naïve Rehearsals as a Framework for Artistic Experimentation

Evaluation is becoming a topic of increasing relevance to human-computer improvisation, with evaluation frameworks often being borrowed from Human Computer Interaction (HCI). Linson, Dobbyn and Laney (2016) argue that qualitative evaluation by experts is the most appropriate evaluation method for freely improvising interactive computer music systems and a preliminary literature review reveals that it is indeed the most commonly used method.

Brown, Gifford and Voltz (2017) adopt an iterative design process based on evaluation by expert musicians, during which they collect both quantitative and qualitative data in the form of open-ended feedback. Hsu and Sosnick (2009) focus on usability, interaction and 'musical results', combining expert evaluation ('naïve' and 'informed' rehearsals, as well as questionnaires) with audience surveys. In Weinberg and Driscoll's (2006) user study, expert users were asked to interact with a robotic percussionist, participate in a 'perceptual experiment' and answer a questionnaire.

While in the case of human-computer improvisation systems, these evaluation methods seem to provide interesting insights, by helping identify and subsequently address shortcomings of the system, the question of their applicability to composed music is undoubtedly a complex one. For instance, usability and interaction – both important aspects in HCI evaluation frameworks – may be irrelevant and even undesirable in the context of a specific composition. For example, in Mark Applebaum's (2014) *Aphasia* the performer ('singer') is asked to synchronize highly detailed hand gestures to an audio tape. Since there are no sensors involved, the synchronization is left entirely to the performer's ability to execute the score as accurately as possible. This creates a carefully composed illusion of interaction, which leaves the audience wondering whether the performance was in fact based on some kind of sensor technology. In this example, there is essentially no interaction – at least not in an HCI sense. In fact, evaluating parameters such as usability and interaction would contradict the very premise of the composition.

In addition to aspects of usability and interaction, the evaluation of human-computer improvisation systems often includes aesthetic components (Weinberg and Driscoll 2006; Brown, Gifford, and Voltz 2017; Hsu and Sosnick 2009). An application of such a framework to composed music would be rather problematic, as appealing to crowd-sourced aesthetics stands in stark contrast to the highly subjective nature of the aesthetic judgments and goals involved in compositional practices.

Nevertheless, some of the evaluation methods mentioned earlier can be a useful tool when used in the context of *creative experimentation* instead of a formal evaluation. In the case of interactive compositions, in particular, balancing authorship and interpretative freedom remains a significant challenge and one that can only be addressed through extensive experimentation in collaboration with the musicians.

In the development of the composition described in this chapter, 'naïve rehearsals' (Hsu and Sosnick 2009) were used as a framework for artistic experimentation throughout the creative process. In these sessions, percussionist Manuel Alcaraz Clemente was asked to improvise with the robotic percussionist, without being given any information on its interaction capabilities prior to the improvisation. The purpose of these experiments was to identify the perceived interaction affordances of the robotic percussionist. Some of the interactions that emerged during these sessions were considered as undesirable and led to revisions of the score and/or software, while others were considered as musically interesting and were later integrated in the composition. It is important to clarify that what constitutes an "undesirable" or a "musically interesting" interaction component in this context was determined by the composer and not the user/performer, since the purpose of these sessions was not an evaluation of the composition, but rather *aesthetic experimentation as part of the compositional process*.

These experimentation sessions started with a naïve rehearsal, followed by a semi-structured interview in which the musician was asked to describe his experience and the ways in which the system responded to his actions in each scenario. Following this short interview, the musician was asked to fill-in a questionnaire regarding the degree of controllability, responsiveness and autonomy of the system, the degree of influence that the generated responses had on his actions, as well as the timing (synchronous vs. asynchronous) and time-scale of the interaction (i.e., whether the responses were based on short or long-term changes). Finally, the musician was asked to fill-in a similar questionnaire after participating in an informed rehearsal.

The musician's responses to the questionnaires and interview suggested that his perception of the degree of responsiveness and autonomy of the robotic percussionist, as well as the influence its responses had on his actions differed substantially in the naïve and informed rehearsals. Interestingly, during the interview the musician tended to personify and assign a male gender to the robotic percussionist, a tendency also observed in other experimentation sessions conducted as part of this research (Chapters 5 and 7). In addition to questionnaires and interviews with the musician, data from these sessions was collected through observation by the author. This helped identify discrepancies between the intended and perceived interaction affordances of the robotic percussionist and devise performance instructions that would guide the musician's actions towards the intended action spaces.

These experimentation sessions fed back into the compositional process, informing revisions of the score and code and, in some cases, leading to entirely new ideas. For instance, the inspiration behind "repeat" was a naïve rehearsal in which the musician mistakenly thought that the robotic percussionist was repeating his actions one by one. This misinterpretation of the robotic percussionist's actions resulted in an interesting counterpoint between the human and the robotic percussionist, which was later integrated in the composition as a separate interaction scenario.

4.6 Discussion

As suggested by its title, the work described in this chapter aims to establish isomorphisms and equivalencies between human and machine agency. Concretely, the decision-making stage of the robotic percussionist is based on aesthetically driven decisions incorporating high-level percepts, while its action stage involves acoustic sound sources and actuators used to simulate human actions (e.g., "strokes" and "scraping"). Similarly, its auditory processing stage is based on a dual classification task involving (human) musical concepts such as "instrument" and "playing technique".

As part of the compositional process for this piece, evaluation methods from human-computer improvisation were adapted into a framework for creative experimentation, fostering composer-performer collaboration. Particularly, the formats of 'naïve' and 'informed rehearsals' (Hsu and Sosnick 2009) were used to explore perceived interaction affordances in composed interaction scenarios. Data collected through observation, questionnaires and a semi-structured interview with the musician was used to inform the compositional and software development process, with the objective to balance compositional control with real-time interaction and decision-making.

Naïve and informed rehearsals provided valuable insights into the perceived affordances of the robotic percussionist and were crucial to the compositional process. While the experiments described in this chapter aimed mainly at exploring the interaction affordances of the robotic percussionist and informing revisions of the code, in subsequent works/case studies, these methods were developed further to address issues related to the trade-off between musical authorship and interpretative freedom in interactive musical works, and provided the main methodological basis for the rest of this research.

5 Experiments in Group Decision-making and Collaborative Emergence

Converge/Diverge, for Piano, Double Bass and Interactive Music System

5.1 Introduction

Compositional strategies aiming to blur the boundaries between composition and improvisation and expand the space of possible interpretations of a musical work are many and diverse. Over the last century of music history, open, graphic and text scores have been employed to allow for a higher degree of interpretative freedom, leading to a new understanding of the musical work as a space of possibilities, as opposed to a thoroughly composed structure of sounds. In the last few decades, interactive music systems, i.e., computer music systems that use machine listening and generative algorithmic processes to interact with human musicians, have added to the complexity of interactions that can take place as part of musical performance and, by extension, to the creative possibilities available to composers.

Performances shaped by decisions made in real-time, whether by human or virtual performers, share a common objective: allowing for emergent musical phenomena resulting from collective and spontaneous creativity. In interactive compositions, in particular, real-time decision-making takes place in the context of concrete interaction scenarios and is guided both by the interaction affordances of the computer music system and some form of performance instructions. Comprising both composed and improvised musical actions, interactive musical works showcase yet another type of collaborative creativity: one that is distributed between the composer and the performers.

The concept of 'collaborative emergence', arising through group behavior and decision-making, is the focus of the composition described in this chapter. 'Collaborative emergence' refers to emergent group behavior that arises in improvisatory contexts in which there is no structured plan or a 'leader' guiding the group (Sawyer 2000, 183). *Converge/Diverge* is a composition for piano, double bass and Interactive Music System (IMS) based on a dynamic form, shaped by decisions made by the musicians and the IMS in real-time.[12] The dynamic form of the piece allows for emergent musical phenomena, resulting from collective spontaneous decisions. Another central concept in this work is that of joint agency. In order for any musical change to happen during a performance of *Converge/Diverge*, all actors involved (i.e., both musicians and the IMS) have to act jointly. As a result of this "constraint", during the performance, intentions are being continuously negotiated and adapted to group dynamics and momentary stimuli, leading to widely varied interactions and musical outcomes.

5.2 Converge/Diverge

In *Converge/Diverge*, the two musicians (pianist and double bassist) are free to explore three different states of the interactive music system: "converge", "diverge" and "negotiate". By playing similar or dissimilar sound material (i.e., "converging" or "diverging"), the musicians can initiate different interaction scenarios, entailing diverse sonic interaction affordances. The terms convergence and divergence in this context refer exclusively to the degree of timbral similarity between the two audio inputs (piano and double bass), measured by calculating the Euclidean distance between Mel Frequency Cepstral Coefficients (MFCCs) extracted from the input signals. The interaction dynamics between the musicians are both sonified and influenced by the IMS, which, in addition to monitoring the interaction between the two musicians and responding accordingly, can initiate two additional states ("cooperate" and "compete").

[12] Performances of the piece by ensembles Schallfeld and Klangforum are available at: https://www.artemigioti.com/demos/Converge_Diverge.html.

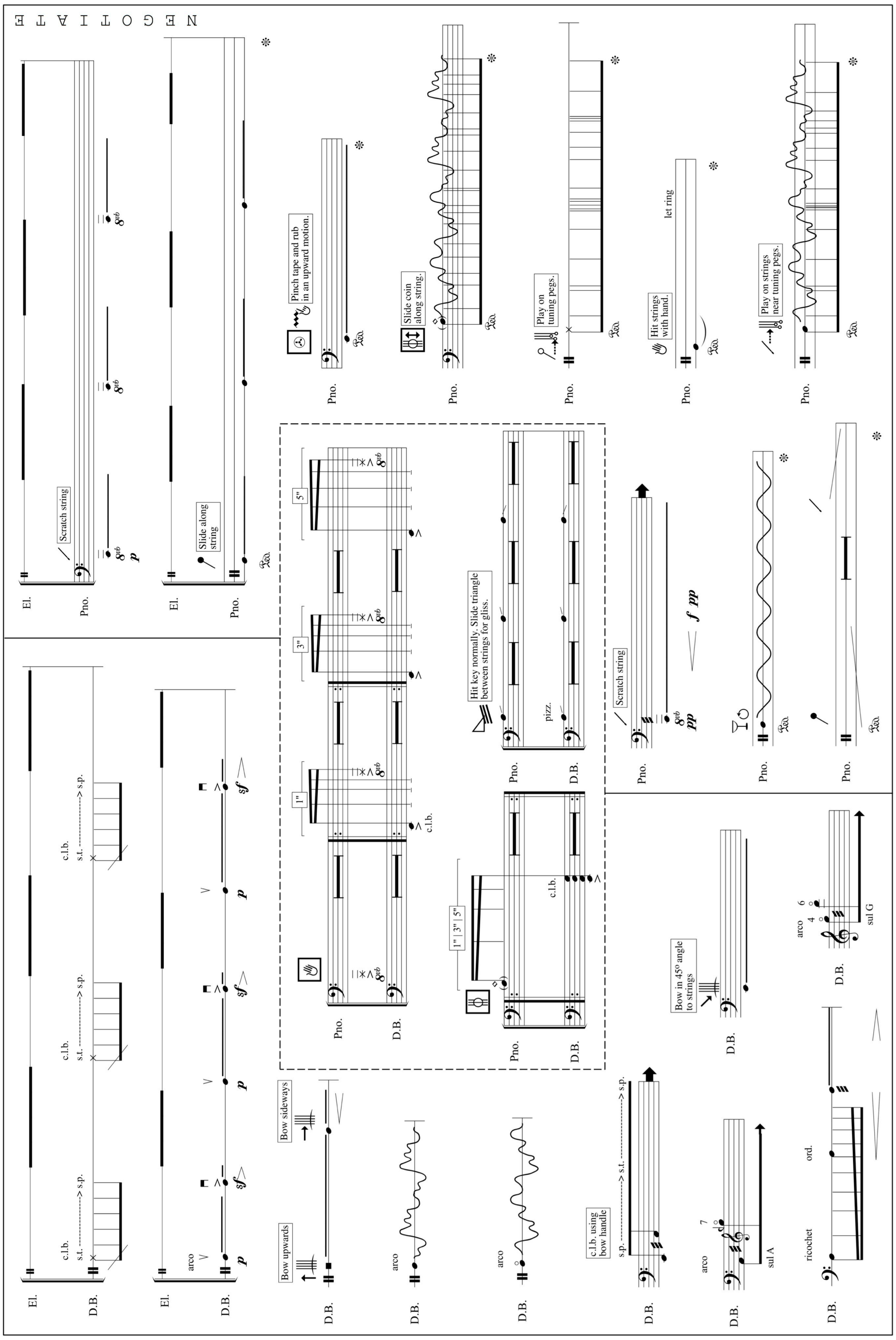


Figure 5.1 *Converge/Diverge*: score excerpt.

The default state of the IMS is “negotiation”. In this interaction scenario, the musicians take turns, choosing sound material from a pool of notated actions (Figure 5.1). The response of the IMS in this scenario consists in generating spectrally compressed variations of the input signal, using a series of band pass filters and envelope followers to analyze it and additive synthesis to resynthesize it. As only a small number of frequencies are used for resynthesis, the electronic sound resembles a resonance, rather than an exact imitation of the human input.

Convergence and divergence can only be initiated by both musicians jointly, making interaction with the IMS a matter of negotiation, collaboration and joint action between the two musicians. Two separate pools of synchronous actions are provided as sound material for “convergence” and “divergence”. By playing sound material from one of these pools, a musician extends an invitation to their co-player to “converge” or “diverge”. As such an invitation can either be accepted or rejected, joint agency plays a central role in shaping the form of the performance. If the second musician decides to accept the invitation and join their co-player (that is, if both musicians start playing simultaneously), the IMS begins to assess their interaction with the purpose to determine whether they are in “convergence” or “divergence” with each other.

The IMS responds to convergence by generating spectrally richer responses (i.e., increasing the number of individual frequencies used by the synthesis algorithm) and updating synthesis parameters with a longer delay. The system remains in this state for as long as the spectral distance between the two inputs remains under a certain threshold – i.e., as long as the musicians remain in “convergence” – meaning that the duration of this scenario is up to the performers.

When divergence is detected, the IMS responds by initiating one of two additional scenarios: “compete” or “cooperate”. In the latter, the system responds by generating a static spectrum, essentially becoming unresponsive. In order for this spectrum to be dissolved, the musicians have to “cooperate” (i.e., “converge”).

A pulsating electronic sound (the result of amplitude modulation with a square wave) is an indication that the system has entered the “compete” mode. In this scenario, the musicians compete for the computer’s attention, which only responds to the musician currently playing the most “novel” sound material. “Novelty” in this context is judged by calculating the spectral distance between currently and previously played sound material for each musician. In this interaction scenario, the musicians can use the notated material as a starting point and/or improvise freely, introducing new sounds of their own choosing. The duration of this scenario is determined by the IMS.

The IMS has no preconception of convergence or divergence, meaning that there are no hand-coded thresholds or machine learning involved in identifying certain sonic interactions as convergent and others as divergent. These states are understood as relative to the overall sonic interaction between the two musicians. The computer is essentially "learning" on the fly, by observing the interaction between the two musicians and comparing the current spectral distance between the two audio inputs to previously observed values. Whether a certain sonic interaction constitutes a "convergence" or a "divergence" is determined by comparing the current distance value to the standard deviation of previously observed values. If the current value falls outside the standard deviation in either direction, the IMS responds accordingly by activating either "converge" or "diverge".

This constitutes an additional interaction feature of the IMS, which though originally not intended as such, adds to the idiosyncracy of the piece. As convergence and divergence are understood by the IMS relative to a specific sonic interaction and determined with respect to previously observed values, the ability of the IMS to successfully identify these states is based on data collected during the performance. Consequently, for the first few minutes of the performance, the response of the IMS might be less reliable and predictable, as its decisions are based on a small amount of collected data. This feature only comes into play if the musicians try to initiate "convergence" or "divergence" within the first few minutes of a performance and is irrelevant if they remain in "negotiation" during this time.

In their interpretation of the piece, Nikolaus Feinig and Florian Müller (Klangforum Wien) deliberately attempted to initiate "convergence" and "divergence" early into the performance, with the purpose to induce unpredictable responses. This interpretative choice is an interesting example of the degree of interpretative freedom involved in the performance of interactive musical works, as well as the interplay between intended and perceived interaction affordances in them.

Besides interpretative freedom, another aspect of interactive music that is exemplified in different instantiations (i.e., performances) of the piece is that of interpretative individuality. In the rehearsals leading to another performance of the piece in New York, pianist Jana Luksts and double bassist Evan Runyon suggested that they were interested in finding ways to differentiate their performance from previous performances by ensembles *Schallfeld* and *Klangforum* (Jana Luksts and Evan Runyon, in discussion with the author, October 2019*)*. Jana Luksts, in particular, suggested that she intended to play exclusively on the piano keyboard (as opposed to inside the soundboard) in "compete", as means to demarcate this scenario from other interaction scenarios involved in the piece. This interpretative choice reveals another way in which the creative responsibility delegated to the performers manifests itself in the piece: by informing and influencing its future performances. The work seems to evolve as different musicians develop diverse interpretative strategies and explore new areas of the action spaces available to them. Of course, the documentation and dissemination of different performances of the piece in the form of video or audio recordings is instrumental to this process.

Finally, central to any performance of this piece is the aural and visual communication between the musicians and their interpretation of each other's intentions. As mentioned earlier, in order for any musical change to happen in the piece, the intentions of all agents involved – including the IMS – have to be aligned. Not only do both musicians need to be on the same page – both metaphorically and literally, as the pool of sound material for each interaction scenario occupies a single page! – but also the IMS needs to correctly interpret their interaction. "Misunderstandings", both on behalf of the IMS and the musicians, are rare but possible, while intentions are constantly negotiated, modified and adapted to the current interaction.

5.3 Compositional Process and Methods

The compositional process for this work involved a series of experiments centered around improvisation tasks and conducted with the help of double bassist Margarethe Maierhofer-Lischka and pianist Patrick Skrilecz. In these experiments, improvisation was used to explore and refine both abstract compositional ideas and concrete interaction scenarios. Exploratory, 'naïve' and 'informed' rehearsals (Hsu and Sosnick 2009) were used in different stages of the compositional process and data from them was collected through observation, questionnaires and semi-structured interviews with the musicians.

5.3.1 Exploratory Rehearsals: Defining Convergence and Divergence

The purpose of exploratory rehearsals was to explore the evocative power of the concepts of convergence and divergence as metaphors for musical interaction, as well as the degree of intersubjectivity involved in their interpretation by the musicians. The musicians were given a total of 4 different improvisation tasks and were asked to reflect on various aspects of their improvisation (e.g., form, sound material, interaction etc.) in semi-structured group interviews following each task.

In the first task, the musicians were asked to improvise for an approximate duration of 10 minutes. They were then asked to reflect on their interaction during the improvisation and try to identify any moments of convergence and divergence. This was the first instance in which the concepts of convergence and divergence were introduced to the musicians (i.e., the musicians were asked to reflect on these concepts only after the improvisation, rather than take them into account while improvising). Interestingly, both musicians agreed that their actions were highly "convergent" and could not identify any moments of divergence in the session. When asked which element of the improvisation was most suggestive of convergence, they responded that their playing was centered around specific pitch centers.

In the second task, the musicians were instructed to explore the concept of convergence in an improvisation of approximately 10 minutes. In the discussion following this session, they commented that their actions were convergent with respect to pitch, timbre ('playing techniques') and loudness. Elements of musical form, such as different textures and musical gestures were also mentioned as aspects suggestive of convergence. The musicians agreed that both aural and visual communication played an important role in their interaction and pointed out that they perceived not only similar but also complementary actions as convergent, citing as an example a part of the improvisation in which loud chords on the piano were followed by sustained tones on the double bass, creating an artificial 'resonance'.

In the third improvisation task, the musicians were asked to explore the concepts of divergence and competition. When asked to describe this session, they mentioned that it was characterized by a higher level of activity, more frequent musical changes and a wider range of dynamics, pitch and rhythms. They commented that they consciously tried to avoid imitating each other's actions, but disagreed on which musical parameter was most suggestive of divergence, with opinion being split between rhythm and dynamics. Both musicians agreed that their interaction was not antagonistic and pointed out that they still tried to 'make music together'. Reflecting on their reluctance to explore more antagonistic forms of interaction, the musicians suggested that instructions to 'play faster or louder' than their co-player could potentially be helpful. Similarly to the previous improvisation task, visual communication was considered a crucial aspect of music-making.

Overall, the musicians repeatedly used the terms 'harmony' and 'harmonic' to describe the session exploring the concept of convergence and the term 'counterpoint' to describe the session on the topic of divergence, while they associated complementarity with both convergence and divergence.

Finally, the musicians were asked to improvise for another 10 minutes, this time incorporating both concepts in their improvisation. They were later asked to listen to a recording of this session and assess the degree of convergence between their actions on a scale from 1 (very low) to 5 (very high) for every 15′′ of the improvisation. Their responses were very similar, as shown in Figures 5.2 and 5.3.

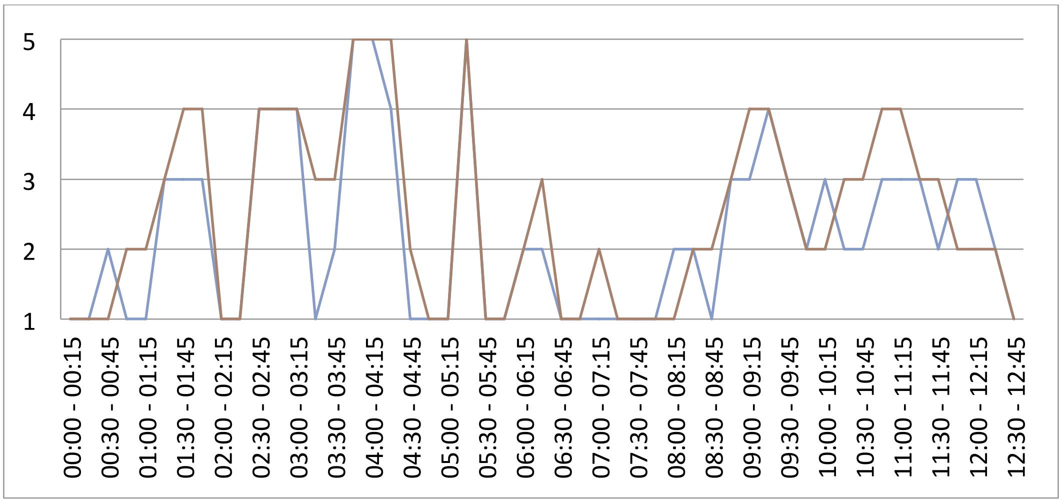


Figure 5.2. Exploratory rehearsal: individual responses. Degree of perceived convergence from 1 ("very low") to 5 ("very high").

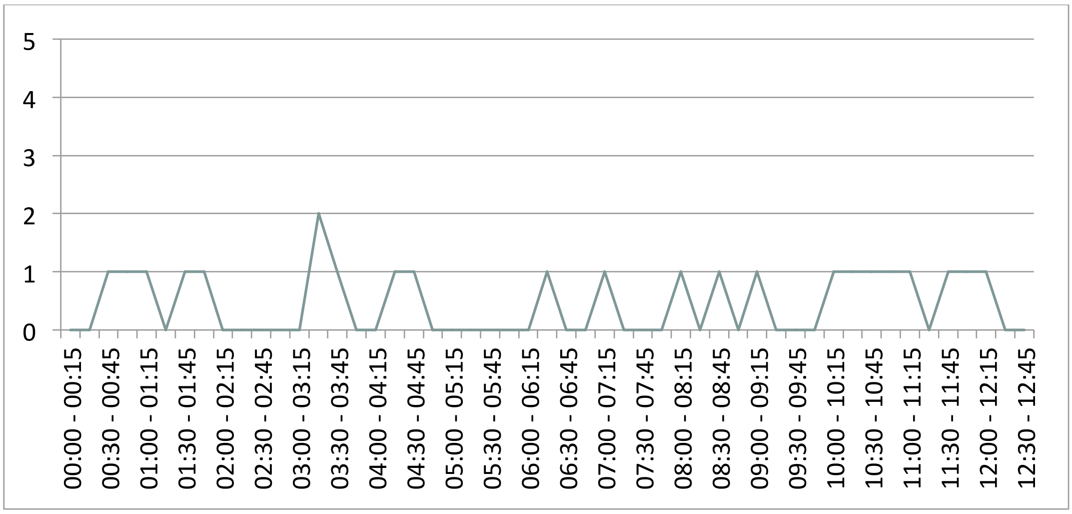


Figure 5.3. Exploratory rehearsal: absolute difference between the musicians' responses.

The exploratory rehearsals helped shed some light into the concepts explored by the piece and the challenges involved in their adaptation into sonic interaction scenarios. The two main challenges identified through this process were:

1) the musicians' reluctance to explore antagonistic forms of interaction, a concept that was central to the compositional idea, and
2) that convergence and divergence can potentially be understood with respect to a variety of musical parameters (e.g., pitch, rhythm, timbre etc.) and behaviors (e.g., complementarity can be associated with both convergence and divergence).

The first point was addressed by designing responses that encourage the musicians to explore divergent sound material. While the response of the IMS to convergence is hardly distinguishable from its default mode, consisting solely in increasing the number of frequencies and response time of the additive synthesis algorithm, divergence can trigger more diverse and less predictable sonic interactions. Concretely, when divergence is detected, the IMS can trigger either "compete" or "cooperate", a decision over which the musicians have no control. And, while "cooperate" consists in a simple error-like behavior (i.e., a "spectral freeze" effect), which can be resolved through prescribed actions, the sound material for "compete" is effectively left to the musicians, who can choose to use (some of) the notated actions or improvise. Additionally, the IMS only responds to the musician playing the most "novel" sound material, a feature that was implemented specifically to encourage the musicians to experiment sonically.

While "convergence" and "divergence" can be understood in relation to a variety of musical parameters (e.g., pitch, rhythm, timbre etc.), in *Converge/Diverge* the focus lies on timbre. This was partly a sound-driven decision, dictated by the broader aesthetic context of the piece (i.e., sound-based as opposed to note-based music), and partly a form-driven decision, aiming to make different interaction scenarios and behaviors more distinguishable.

The decision to use the Euclidean distance between MFCC vectors as a measure of timbral similarity, as opposed to machine learning models built from human-labeled data, also had implications for the sound material used in the composition. First experiments with this approach revealed significant differences between human perception of timbral similarity and the computer music system's perception of spectral convergence and divergence. Spectral convergence was identified rarely by the IMS and seemed to be correlated with high-pitched, sine-wave-like sounds – i.e., overtones, lacking the characteristic timbre of the instrument. This led to the use of a number of unconventional playing techniques, such as rotating a glass on top of the piano strings or sliding a triangular ruler between them (Figures 5.4 and 5.5). As the system's estimation of spectral similarity can at times deviate from human perception, the IMS has the potential to surprise the musicians, by behaving in unpredictable ways, a feature that adds to its idiosyncracy.

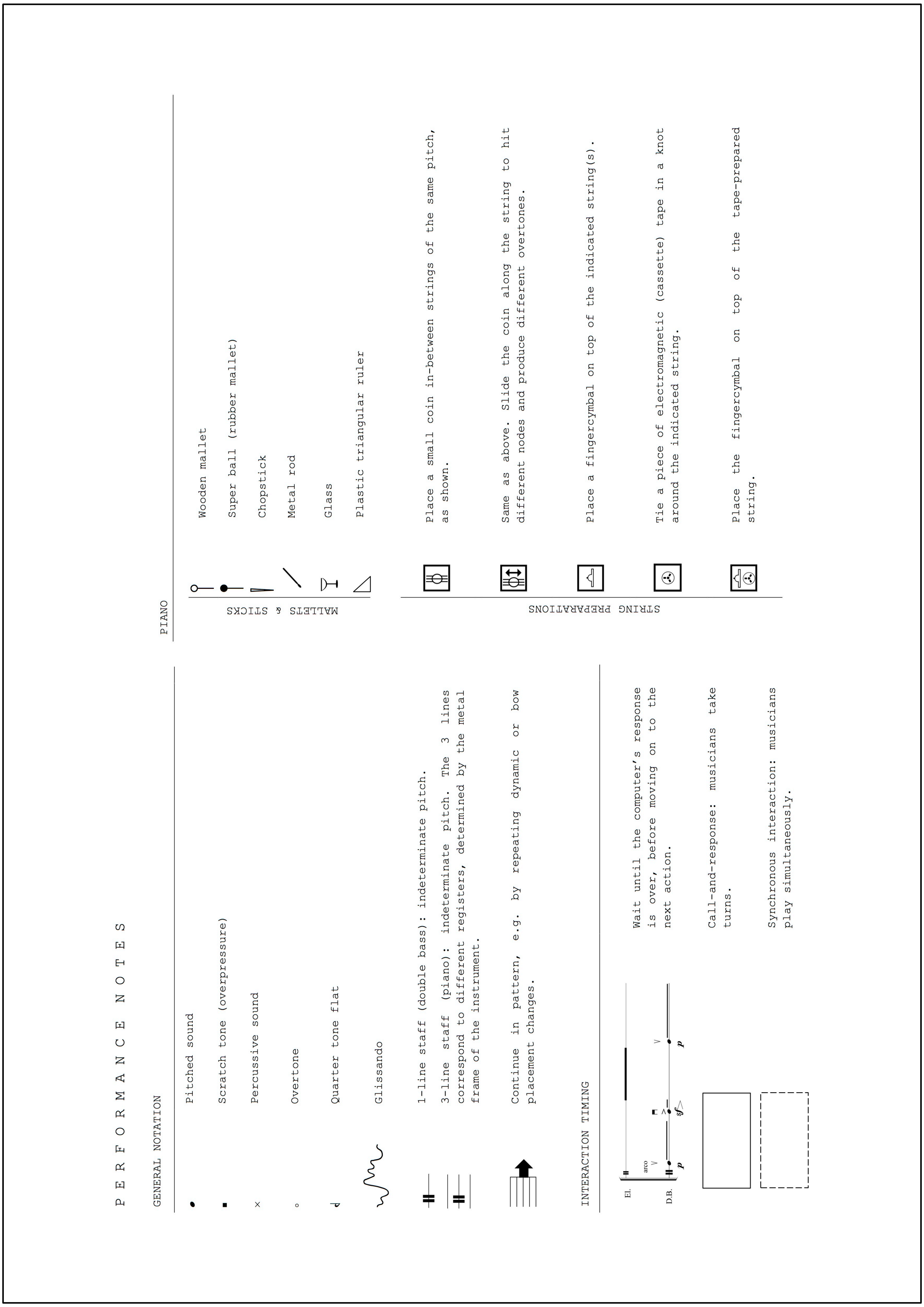


Figure 5.4 *Converge/Diverge*: performance notes.

Figure 5.5 *Converge/Diverge*: extended playing techniques.

5.3.2 Naive Rehearsals: Balancing Authorial Responsibility and Interpretative Freedom

The exploratory rehearsals described in the previous section informed the compositional process and played a decisive role in both compositional and design choices. The use of qualitative research methods in their context (e.g., semi-structured interviews with the musicians) enabled a more systematic and productive composer-performer collaboration and helped explore an abstract compositional concept and gain insight into some of the challenges relating to its interpretation and implementation. Both the interaction affordances of the IMS and the sound material used in this composition were greatly influenced by insight gained through these sessions.

Exploratory rehearsals were employed mainly the conceptual stage of the compositional process. Later in the creative process, when first drafts of the score and code were written, the musicians were asked to participate in a 'naïve rehearsal' (Hsu and Sosnick 2009), a format meant to explore the *perceived* – as opposed to *intended* – interaction affordances of the IMS and inform further compositional decisions. In this session, the musicians were asked to improvise with the IMS without being given any information regarding its affordances – although at this point the musicians already knew that the concepts of "convergence" and "divergence" would play a central role in the piece. The purpose of this experiment was to identify unintended affordances of the IMS and explore strategies for balancing the trade-off between musical authorship and interpretative freedom.

After the improvisation session, the musicians were asked to fill-in a questionnaire about the system's behavior and responsiveness. Interestingly, the musicians failed to identify most interaction scenarios and only one of them identified amplitude modulation (i.e., "compete") as a response to 'divergent and chaotic sounds'. When asked to describe the system's different behaviors, they focused mainly on its response to different dynamics and registers, rather than the degree of similarity between the sounds they played. They correctly observed that in some parts of the improvisation the IMS was listening to both of them, while in others it was only listening/responding to one musician at a time. They agreed that the system was able to act independently of their actions, but thought that the influence its actions had on the course of the improvisation was limited. Overall, the musicians' responses suggested that the system's interaction affordances alone were ineffective in communicating compositional intent and that further performance instructions and knowledge of its capabilities and intended affordances would be necessary in guiding their actions towards the intended action spaces.

After filling-in the questionnaire, the musicians were given some general information about the system's sonic interaction affordances and listening capabilities and were asked to improvise with it for another 10 minutes. Data from this 'informed rehearsal' (Hsu and Sosnick 2009) was collected through observation and video analysis, as the focus in this session shifted from the musicians' to the composer's perception of the improvisation. Observing and analyzing the musicians' 'informed' interaction with the IMS helped compare the intended and perceived affordances of the IMS and devise performance instructions that would help bridge the gap between the two. This compositional strategy is described by Marko Ciciliani as 'subtractive composition' and involves starting from an action space that is as open as possible and gradually introducing performance instructions, until it is reduced to an aesthetically narrower, yet as far as concrete musical actions are concerned, still open space of sonic possibilities (Marko Ciciliani, in discussion with the author, March 2019).

The purpose of this method was to balance the trade-off between authorial responsibility and interpretative freedom in the work through revisions of the score and/or code. For instance, one of the main discrepancies between the intended interaction scenarios and the way the musicians chose to interact with each other and the IMS during the informed rehearsal concerned interaction timing. Concretely, the musicians played simultaneously for most of the improvisation and opted for textures of high density, which meant that were there virtually no moments of silence. While this is in no way meant as criticism, these choices deviated significantly from the interaction concept of *Converge/Diverge*, namely a dialogue-like, call-and-response interaction in which synchronous interaction would be the exception rather than the rule and would signify specific states (i.e., convergence and divergence). The reasons behind this compositional decision were both conceptual and aesthetic. As the piece is based on a conversational metaphor, the call-and-response paradigm seemed more fitting, inviting the musicians to listen and respond to each other in a dialogue-like way. From an aesthetic viewpoint, this interaction paradigm allowed more space for the electronics, as well as for silence, a concept of central importance in the author's work.

5.4 Composition and Improvisation

The tension between authorial responsibility and interpretative freedom in interactive musical works points towards the complex and dynamic relationship between composition and improvisation in them; a relationship that goes far beyond the composition/improvisation binary. Admittedly, the use of improvisation in composed music is not specific to interactive works and can take various forms depending on the composer's artistic goals and aesthetic stance. For instance, Scelsi famously used improvisation as a compositional method, by recording his own improvisations on tape and then transcribing them with the help of musicians (Uitti 1995). In Scelsi's practice, improvisation was a means rather than an end in itself; it was a method used to produce scores that would ensure the reproducibility of the notated material.

Composers such as Mauricio Kagel and Cornelius Cardew, on the other hand, viewed improvisation as a compositional strategy and incorporated it in their work in varying degrees. The use of ambiguous graphic notation by composers such as Cardew is a compositional strategy aiming to increase interpretative freedom. Composer Cat Hope (2017) uses graphic and animated scores to allow musicians to make decisions on how to engage with their instruments (both acoustic and electronic) in an approach that views improvisation as part of interpretation. In her works, some aspects of the performance are left to the musicians while others are clearly defined, ensuring that, despite the high degree of interpretative freedom involved in them, they are always identifiable as the same work.

Similarly, Richard Barrett (2014) views notation and improvisation as compositional strategies and often combines precise notation with free improvisation within the same work. He uses the term 'seeded improvisation' to describe works in which precisely notated passages are interspersed with improvisatory passages, providing a form of overall structural context, while allowing the musicians to focus on spontaneous improvisatory actions (64-65). The argument behind this approach is that it can give rise to emergent musical phenomena, which would not have resulted from notation or free improvisation alone.

In *Converge/Diverge*, improvisation was used both as a compositional strategy and as a method for artistic experimentation. The combination of action-based notation with what could be partly described as a 'mobile' score, i.e., a score in which the order of notated material is decided during the performance (Hope and Vickery 2011, 225), suggests that improvisation is an essential aspect of the interpretation of the work. However, “improvisation” in this context is not synonymous with “free improvisation”, but rather improvisatory musical actions and decisions within “composed” interaction scenarios.

Additionally, in the work described here, improvisation was integrated in a series of experiments designed to explore and refine an abstract compositional idea and, later on, identify the perceived interaction affordances of the IMS and inform compositional decisions. As interactive musical works challenge the composition/improvisation binary and, along with it, traditional compositional practices, such experiments can be helpful in dealing with the high degree of unpredictability involved in this type of music and deciding which aspects of the performance should be determined through performance instructions and which should be left to the performers.

5.5 Composer-Performer Collaboration

Hayden and Windsor (2007) identify three different and, at times, overlapping types of composer-performer partnership: *directive*, *interactive* and *collaborative*. In the *directive* paradigm, the performance is completely determined through the score, while the relationship between composer and performer is hierarchical, with any collaboration between them being limited to issues of technical nature. In an *interactive* partnership, compositional decisions are informed by the performers’ and/or technicians’ input, while some aspects of the performance might be open, but the composer is still the single author. Finally, the *collaborative* approach involves co-authorship and collective decision-making. In pieces created through the collaborative approach, the macro-structure of the performance is not determined by a single composer, but rather by group decisions made in real-time.

The type of composer-performer collaboration described in this chapter falls under the *interactive*, rather than the *collaborative* paradigm, even though the form of the piece is the result of group decisions and can vary from one performance to another. While compositional decisions were made by the author, each performance of the piece is a unique and unrepeatable event resulting from collaborative and distributed creativity. Creativity in this work is distributed across actors (composer, performers, computer music system), different types of activities (composing, programming, performing) and in time (“offline” compositional decisions vs. real-time interpretative choices).

5.6 Discussion

The concept of the interactive musical work poses a number of conceptual and technical challenges, not the least of which is reconciling its ontological status as a product of a co-creative process involving human and non-human actors with traditional compositional strategies. In interactive musical works compositional intentions, interpretative freedom and machine agency stand in a discursive relation to each other, as is evidenced by their widely varied instantiations in different performances.

This chapter presented a series of methods used to navigate the tension between work identity and interpretative freedom in a composition for piano, double bass and interactive music system. These experiments aimed at exploring compositional ideas and their early implementations and informing further compositional and design decisions. Admittedly, these experiments were designed to meet the needs of a specific composition and are far from universally applicable. Nevertheless, similar experimentation frameworks could provide a fertile ground for composer-performer collaboration and creative experimentation within a broader range of ‘open work’ (Eco 1989) musical practices.

6 Interpretative Individuality in *Converge/Diverge* for Piano, Double Bass and Interactive Music System

6.1 Analyzing Interactive Music

Interactive musical works present a unique set of challenges for music analysis, which are closely intertwined with their aesthetics; particularly their emphasis on spontaneity, unpredictability, ephemerality and relational aesthetics. Not the least of these challenges is defining the object of analysis. Since a representation of an interactive composition through notation alone is not only practically impossible, but also antithetical to its premise and aesthetics, analysis cannot focus on the score alone, but needs to take into account the interdependencies between performance instructions (notated or otherwise), the interaction affordances of the computer music system and the agency of human performers.

The tradition of electroacoustic music analysis offers an obvious alternative to score-based music analysis. Smalley's (1997) *spectromorphology* approaches music analysis from a phenomenological perspective, focusing primarily on the listening experience of acousmatic music, though applications of the theory to acoustic works with textural and spectral complexity are not excluded. Smalley specifically mentions the work of Xenakis, Grisey, Saariaho, Murail and Dillon as examples of instrumental music to which this type of analysis could be applicable. His subsequent *space-form* concept (Smalley 2007) privileges space as an articulator of musical form – specifically, a transmodal perception of space that relies on spectromorphological relations and *source-bonding*, i.e., the perceived relations between acousmatic sound and its (imagined) sources. Other phenomenological approaches to the analysis of acousmatic music include Fischman's (2007) framework of *mimetic space*, based on Emmerson's (1986) *language grid* and Smalley's (1997) concept of *source-bonding,* and the precursor of Smalley's *spectromorphology*: Schaeffer's *typomorphology* (2017) and its subsequent adaptations (e.g., Thoresen 2001).

While these approaches have substantially expanded the conceptual frameworks and analytical tools available to the musical analyst, their focus lies in the analysis of fixed and determinate musical works (i.e., fixed media compositions or fully notated acoustic works). Therefore, they are mostly suitable for product-based compositional approaches, which prioritize the permanence of notation and/or recording over the ephemeral nature of real-time musical interactions.

Interactive musical works, on the other hand, are examples of 'open work' (Eco 1989) practices that prioritize embodiment, interaction and interpretative freedom, and view sound as a (dynamic) relation, rather than a (static) object. Indeed, in interactive compositions sound results might be secondary to the interaction that produced them and can differ significantly from one instance/performance to another.

Interpretative multiplicity in interactive works is a result of the evocative, rather than representational function of both the score and the software/code and the high degree of interpretative freedom they afford performers. Score and code delineate the agentive space within which the musicians and the computer can make decisions, but don't accurately describe the sonic structures that will emerge through their interaction. Depending on the level of abstraction involved in the score and the degree of autonomy and idiosyncracy exhibited by the computer music system, an interactive composition may allow multiple and diverse interpretations and musical results. The sonic trajectories that are explored in a given performance depend on the performers' interpretative choices and their real-time interaction with the computer.

In essence, in an interactive composition the relationships between score and performance, as well as score and work are questioned and redefined. The score or a performance alone is insufficient as a representation of the work. Admittedly, no musical work can be reduced to a single instantiation, e.g., the score or a performance (Born 2005), but particularly in the case of interactive works these instantiations are highly convoluted, as the distinction between composition and performance is effectively destabilized.

Similar challenges apply to the analysis of generative music systems, which produce multiple and diverse musical outputs using the same generative functions but different initial conditions (e.g., a different random seed). In an approach that aims to explore the relation between process (i.e., the algorithm) and product (i.e., its potential musical outputs) in generative music systems, Collins (2008) proposes a 'white-box' (as opposed to black-box) analysis method, which involves analyzing the code itself using comments and pseudo-code written in natural language. This approach is admittedly useful, but can only provide a partial account of the processes involved in an interactive performance, as it focuses explicitly on non-interactive generative music systems.

In interactive musical works, the analysis of the code could potentially be complemented by a performance-centered analytical approach. Still, as a single performance of an interactive work is only a 'partial manifestation' (Young 2016, 96) of the possibilities it encompasses, any performance-centered analysis of an interactive work would have to take into account different performances/realizations. Analyzing multiple performances of an interactive composition could potentially help trace the agentive space and sonic affordances it encompasses and shed light on the complex relationship between work identity and interpretative individuality in it.

The overarching question that connects all the analysis scenarios described above is that of perspective. An analysis might adopt a purely phenomenological stance by approaching the work exclusively from a listener's perspective, combine a phenomenological approach with an analysis of the score and/or code, or do all of these things in different stages (e.g., Emmerson 2016). The latter approach is perhaps the most comprehensive, as it takes into account all possible instantiations of the work that is being analyzed. Yet, one last challenge, pertaining specifically to interactive works, remains: analyzing the real-time interaction taking place during the performance; that is, the process of interaction itself, rather than its musical outcome. This interaction does not encompass only computational (e.g., generative) processes, but also the way in which human performers interact with the computer music system, each other, their instruments and any other objects that might be involved in the performance. It includes the dynamics of the interaction (e.g., who is leading and who is following), the interaction timing between the performers and/or the computer music system, as well as non-aural (e.g., visual) forms of interaction, through which the intentions of individual actors are conveyed and negotiated. This aspect reveals some of the limitations of aural information in analyzing sonic interaction and calls for a transmodal phenomenological approach to the analysis of interactive music, which takes into account the multisensory nature of musical performance as a lived experience.

This chapter describes an exploratory approach to the analysis of an interactive composition, focusing mainly on the interaction between human performers and the various manifestations of interpretative individuality in the work. The author's composition *Converge/Diverge* for piano, double bass and interactive music system is used as a case study. This analysis is written from the composer's perspective, which means that it assumes a thorough knowledge of the score and code, both of which have been detailed elsewhere (Chapter 5). The analysis presented in this chapter is centered around two different performances of the work by ensembles *Schallfeld* and *Klangforum*[13], and employs a combination of quantitative and qualitative analysis methods tailored to the specificities of the work and used to address three different analytical foci: form, sound material and interaction.

The objective of this analysis is to outline the sonic possibilities involved in the work and understand the role of interpretative individuality in it and its relation to work identity (i.e., understand which aspects of the piece differ significantly and which remain constant across different performances). This analysis is therefore primarily a compositional tool, meant to facilitate aesthetic reflection and inform compositional decisions, both in the context of a potential revision of the work at hand and the author's future work.

6.2 *Converge/Diverge*, for Piano, Double Bass and Interactive Music System

The work analyzed in this chapter is an interactive composition with a dynamic form, shaped by the real-time interaction among the musicians and the computer. The title of the piece, *Converge/Diverge*, refers to the degree of timbral convergence or divergence between the two audio inputs (piano and double bass), which drives the responses of the

[13] Both performances are available at:
https://www.artemigioti.com/demos/Converge_Diverge.html.

Interactive Music System (IMS) and determines the form of the performance. By playing spectrally “convergent” or “divergent” sound material, the musicians can evoke different responses and initiate interaction scenarios with diverse sonic interaction affordances.

In the default state of the IMS (“negotiation”), the musicians interact with each other in a call-and-response fashion, choosing sound material from a pool of notated actions. Convergence and divergence can only be initiated by both musicians jointly, meaning that musical form in the piece emerges as a result of negotiation, aural and visual communication and collaboration between the two musicians, as well as between the musicians and the computer. Two separate pools of synchronous actions are provided as sound material for “convergence” and “divergence”. By playing sound material from one of these pools, a musician extends an invitation to their co-player to “converge” or “diverge”. If the second musician decides to join the first, the IMS responds accordingly.

As a response to “divergence”, the IMS can initiate one of two interaction scenarios: “cooperate” or “compete”. In the former, the musicians have to “converge” (i.e., play spectrally similar sound material) in order to dissolve a static spectrum produced using a “spectral freeze” effect. “Converge” and “cooperate” are essentially the same interaction scenario, the only difference between the two being that “converge” is initiated by the musicians and “cooperate” by the computer. Finally in “compete”, the musicians engage in a more competitive form of interaction, as the computer only responds to the musician currently playing the most “novel” sound material, i.e., one musician at a time.

The spectral similarity between two sounds is measured by the IMS by calculating the Euclidean distance between Mel Frequency Cepstral Coefficients (MFCCs) extracted from them. This distance metric is used both to determine whether the musicians are in “convergence” or “divergence” with each other and to assess the novelty of the sound material played by them in “compete” (by comparing each musician’s current input to previously played sounds).

The thresholds for “convergence” and “divergence” are set by the IMS during the performance and are not defined based on human-labeled data or observations from past performances. Concretely, the computer keeps a record of all observed spectral distance values and determines whether the current spectral distance between the two audio inputs lies within or outside the standard deviation for this specific performance.

The IMS generally responds to the two instrumentalists by producing spectrally "compressed" versions of their signal, using additive synthesis. In each of the interaction scenarios mentioned above, this response is modified. In "converge", the number of frequencies used by the synthesis algorithm is increased and the response time is decreased (i.e., the computer responds more slowly to spectral changes in the input signals), while in "compete", the amplitude of the electronic sound is modulated by a square wave, an element that adds a rhythmic/percussive quality to the computer's sound output.

The sound material played by the musicians is partly improvised and partly notated, using action-based notation. Concretely, there are three separate sound pools for "negotiate", "diverge" and "converge"/"cooperate", each consisting of several partially notated musical actions, i.e., musical actions one or more aspects of which (e.g., duration, pitch, dynamics etc.) are open. In "compete", the musicians can use the notated actions as a stimulus for their improvisation or improvise freely.

6.3 Analysis Methods

The selection of methods used to analyze the two performances of the piece by ensembles *Schallfeld* (double bass: Margarethe Maierhofer-Lischka, piano: Patrick Skrilecz) and *Klangforum* (double bass: Nikolaus Feinig, piano: Florian Müller) was generally based on the specificities of the piece; for instance the concept of "composed" interaction scenarios and the integration of partially notated and improvised musical actions in different degrees in each of these scenarios. The formal analysis of the two performances was based on the discrete interaction scenarios involved in the piece: specifically, the succession of interaction scenarios observed in each performance, as well as the duration and number of individual instances of each scenario.

Another aspect of musical form that was examined separately is the transition between different interaction scenarios in each performance. The musicians (and computer) can freely navigate the interaction scenarios involved in the piece by taking virtually infinite different paths, so identifying and comparing these paths was considered as an essential component of a formal analysis of the piece. The different ways in which the two ensembles navigated the space of possibilities offered by the composition were traced and compared by calculating transition probabilities between different scenarios for each of the two performances.

As far as the analysis of sound material is concerned, aside from an analysis of the playing techniques used by the musicians in "compete", spectral data extracted from the two recordings and juxtaposed with the formal analysis of the respective recording was used to identify and compare the spectral "profiles" of different interaction scenarios, as well as the two ensembles' unique interpretations of the compositional concept and performance instructions. To that end, self-similarity matrices calculated from FFT data were computed to visualize spectral patterns throughout each and across the two performances.

Park (2016) suggests that the analysis of electroacoustic music could be made more compelling and convincing through the use of quantitative musical features, which are 'measurable', 'consistent' and 'present in the music itself, as opposed to being excessively imagined' (124). While audio descriptors can be a useful tool in the analysis of electroacoustic music and were used extensively in this analysis, the analytical approach presented here views such claims of measurability and objectivity rather critically and adopts a more constructivist view of music analysis, in which interpretations are constructed subjectively, albeit with the help of quantitative data. Audio descriptors can deviate significantly from human auditory perception and are not necessarily more accurate or objective than a manual analysis (e.g., automatic pitch and onset detection can be significantly less accurate than a manual analysis of the same musical features).

Another limitation of audio descriptors concerns the nature of the music examined here. In live electronic music, unlike in fixed media works, what is being analyzed is the recording of a live performance and therefore any audio features are contingent to the recording equipment and conditions (e.g., background noise), a limitation that must be taken into account when interpreting data collected through music information retrieval techniques.

The analysis of the interaction between the two musicians was conducted using video-based interaction analysis (Jordan and Henderson 1995), a method borrowed and adapted from ethnographic research. Interaction analysis is an interdisciplinary empirical research method that investigates the way humans interact with each other – both verbally and non-verbally – as well as with objects, such as artifacts or technologies. Video-based interaction analysis, in particular, can be used to analyze complex, multi-agent and technology-mediated settings and environments. Some of the advantages of this method that are relevant for this analysis are the permanence of the primary record, the complexity and richness of interaction data, the reproducibility of recorded interaction sequences, which allows the analyst to keep track of and analyze overlapping activities, and its low rate of information loss in comparison to other data collection methods, such as field notes (Jordan and Henderson 1995). Video-based interaction analysis has been applied in musical contexts before and, specifically, in the evaluation of musical tabletops (Xambó et al. 2013).

The analysis conducted here was based on three of Jordan and Henderson's (1995) analytical foci:

- the temporal organization of the activity (performance),
- turn-taking,
- the use of artifacts (in this case, the two acoustic instruments and other objects used for string preparation), and
- 'trouble and repair', a category that in this analysis is understood as pertaining to the way in which musicians adapt to each other's actions and deal with misunderstandings and conflicting intentions.

The first two categories (temporal organization of the activity and turn-taking) were analyzed using primarily quantitative data and the second two (use of artifacts and trouble and repair) using qualitative data.

As far as interaction timing is concerned, the performance instructions specify that "compete", "diverge" and "converge"/"cooperate" involve synchronous interaction, while "negotiate" is based mainly on asynchronous interaction (i.e., call-and-response). Nevertheless, the exact timing of this call-and-response interaction is left to the musicians, who can choose to introduce shorter or longer rests between each call and response (non-overlapping call-and-response), or play synchronously for a few seconds (overlapping call-and-response). The interaction timing between the musicians in each of the two performances was analyzed manually and represented graphically. Finally, the same data was used to calculate the total playing time for each musician and the average duration of rests between call and response.

6.4 Results

6.4.1 Form

Just by listening to the two performances by ensembles *Schallfeld* and *Klangforum* (performance 1 and 2 respectively), it is easy to spot a rather significant difference in the way the musicians handle musical form. With the exception of the beginning and end of the performance, performance 2 is characterized by more frequent musical changes, which create the impression of a more fluid musical form in which materials and behaviors succeed each other without forming larger homogenous sections. Additionally, the level of musical (rhythmic, timbral etc.) contrast between different sections seems to be higher in performance 1, creating the impression of a more clearly articulated form in which transitions between different sections are clearly noticeable (e.g., due to the use of distinctively different sound material, or noticeable changes in the interaction timing between the musicians, the dynamics etc.).

The comparative formal analysis of the two performances shown in Figure 6.1 confirms this impression and reveals some additional differences between the two interpretations. Concretely, in performance 2, the transitions between different interaction scenarios are more frequent and therefore the durations of individual instances of these scenarios are significantly shorter.

In quantitative terms, the average duration in seconds of each section (i.e., individual instance of an interaction scenario) in performance 1 is 83 seconds (std. dev. 51), while in performance 2 it is 63 seconds (std. dev. 54) (Figure 6.2). The count of individual instances of almost all interaction scenarios is higher in performance 2 than in performance 1 (Table 6.1).

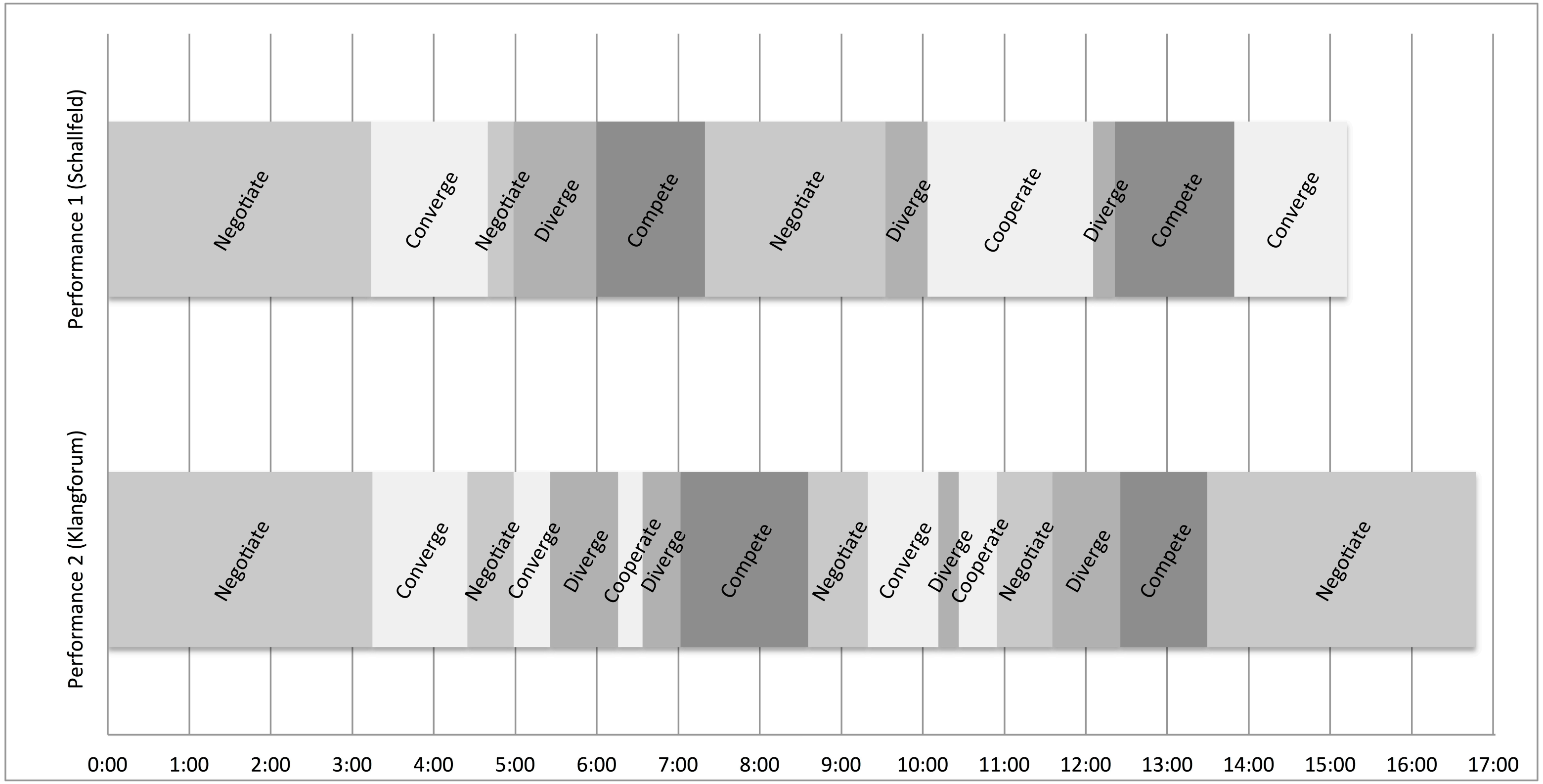


Figure 6.1 Comparative formal analysis of performance 1 (Schallfeld) and 2 (Klangforum). Note: "converge" and "cooperate" are depicted using the same color.

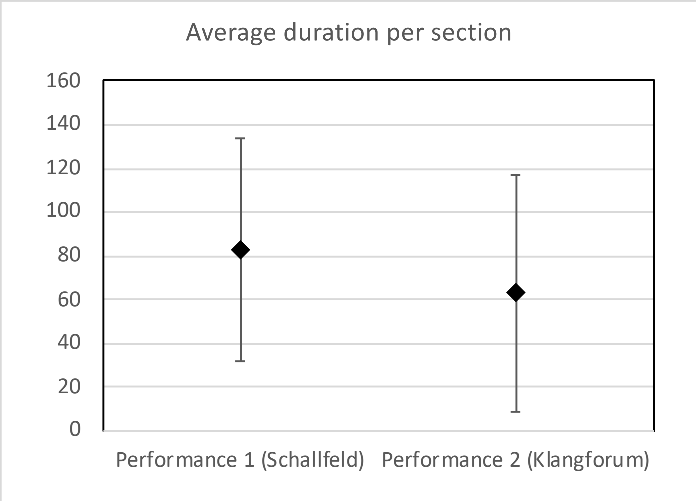


Figure 6.2 Average duration of individual scenario instances/sections in performance 1 (Schallfeld) and 2 (Klangforum).

	Negotiate	Converge	Diverge	Compete	Cooperate
Performance 1 (Schallfeld)	3	2	3	2	1
Performance 2 (Klangforum)	5	3	4	2	2

Table 6.1 Number of individual instances of each interaction scenario in performance 1 (Schallfeld) and 2 (Klangforum).

The transition probabilities between different interaction scenarios for performances 1 and 2 are shown in Figures 6.3 and 6.4 respectively. As “compete” and “cooperate” can only be initiated by the computer, transitions involving any of these two scenarios are depicted using dashed lines, to help differentiate between decisions made by the performers and those made by the IMS. A first examination of the two transition diagrams reveals that certain “paths” are only observed in one performance but not the other. For example, the transition from “compete” to “converge” is observed in performance 1 but not in performance 2. Conversely, transitions from “cooperate” to “negotiate” and from “converge” to “diverge” are only observed in performance 2.

The weights of the transitions convey some additional information regarding the differences between the two performances. For instance, the highest weight in performance 1 (0.2) corresponds to the transition from “negotiate” to “diverge”, while in performance 2 (0.2) from “negotiate” to “converge”. The most frequently observed transition in performance 1 (from “negotiate” to “diverge”) is among the least frequently observed transitions in performance 2, while a commonly observed transition in performance 2 (from "converge" to "diverge") does not appear at all in performance 1.

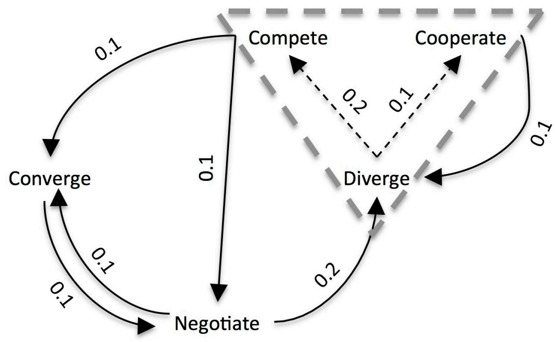


Figure 6.3 Performance 1 (Schallfeld): transition probabilities between different interaction scenarios. Dashed lines represent transitions initiated by the computer.

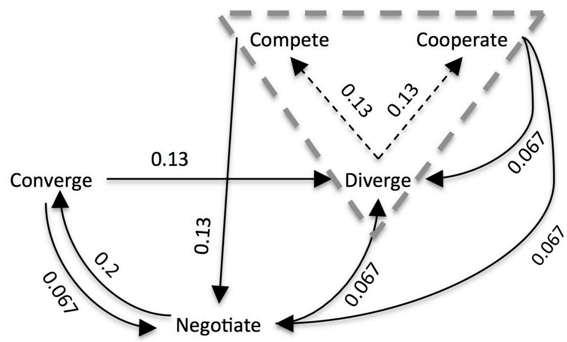


Figure 6.4 Performance 2 (Klangforum): transition probabilities between different interaction scenarios. Dashed lines represent transitions initiated by the computer.

Overall, the scenario-based formal analysis of the two performances reveals differences regarding both the frequency of musical changes (i.e., the duration of different musical sections) and the transition probabilities between different interaction scenarios. The "paths" followed by the two ensembles in their exploration of the interactive space of the composition differs both in their notion of musical form (larger musical sections vs. frequent musical changes) and their unique trajectories through sound material and interaction affordances.

6.4.2 Interaction

From the video-based interaction analysis of the two performances, it is evident that visual communication plays an important role in the interaction between the musicians, yet it differs significantly from one performance to the other. In performance 1, visual communication is used predominantly to establish turn-taking, coordinate synchronized actions and acknowledge musical changes, or the intention to introduce musical changes. The musicians often look at each other to signal that they have completed a musical gesture, or to make sure that their co-player is done playing. Nodding is only observed in one occasion, when the pianist tries to set the tempo for a short sequence of synchronized actions. The musicians also look at each other to acknowledge musical changes, or communicate their intention to initiate a scenario change. Adaptation to each other's actions seems to be another important aspect of their interaction. In one instance, the pianist visibly changes his mind as to which action to perform next, when he realizes that the double bassist is extending an invitation to "diverge". This is indicated both by his gaze and his interaction with physical objects used for specific extended playing techniques (putting aside one object and picking up another). In "compete", the musicians seem to be more focused on the interaction with their instruments, rather than their co-player and visual communication between them is very rare.

In performance 2, visual communication is used mainly to acknowledge musical changes and set the tempo for synchronized actions, but seems to play a very small role in turn-taking. During "negotiate" the musicians' gaze is mostly focused on the score and their instrument rather than their co-player and turn-taking seems to be based almost exclusively on aural information. The musicians seem to establish mutual recognition of musical changes through both gaze and nodding. They exchange very few, if any, gazes in "compete" and seem to rely much more on visual communication in "converge".

In both performances, the visual interaction between the two musicians is asymmetrical, with one musician watching the actions of their co-player more closely than the other. In performance 1 this person is the pianist and in performance 2 the double bassist.

A limitation of this analysis that needs to be acknowledged is that the videos used here were produced for artistic purposes, rather than for data collection purposes. As a result, data collection was hindered by what Jordan and Henderson (1995) refer to as 'limits of the operator' (53-54). While both videos were shot using multiple cameras (both stationary and operated), in the final, post-produced videos only one camera angle is available at any given moment and, therefore, some information is inevitably lost.

As far as interaction timing is concerned, the main difference between the two performances is the degree of overlap between the two instruments in "negotiate" (Figure 6.5) and, even more so, in "compete". In performance 1, the musicians play mostly simultaneously in all instances of the scenario, creating textures of high density that stand in stark contrast to the call-and-response interaction paradigm of "negotiate". By contrast, in performance 2, the musicians mainly respond to each other by playing variations of short rhythmic motives in what seems to be a "time-compressed" call-and-response sequence.

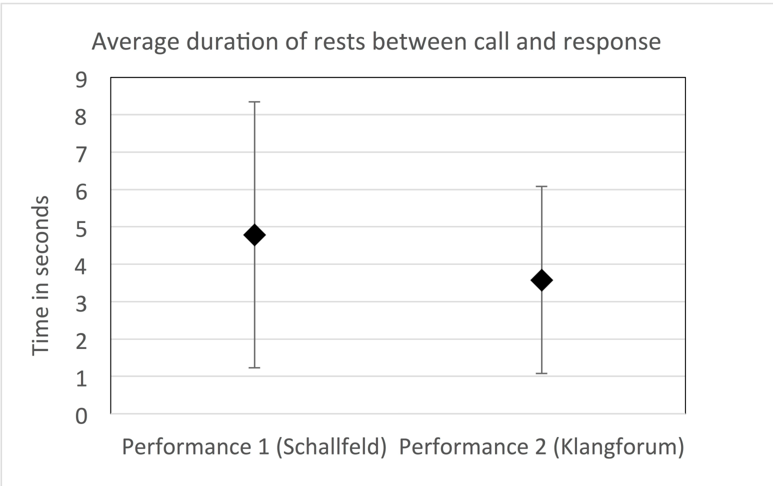


Figure 6.5 Rests between call and response in performance 1 (Schallfeld) and 2 (Klangforum).

The timing of the interaction between the musicians in each of the two performances is shown in greater detail in Figures 6.6 and 6.7. The two performances were analyzed manually, using both audio and video-based analysis, and the start and end times of musical actions were rounded to the closest second. Aural and visual information were used as complementary sources, particularly in cases where determining the start and end time of a musical action proved to be challenging (e.g., due to spectral masking between different sound sources).

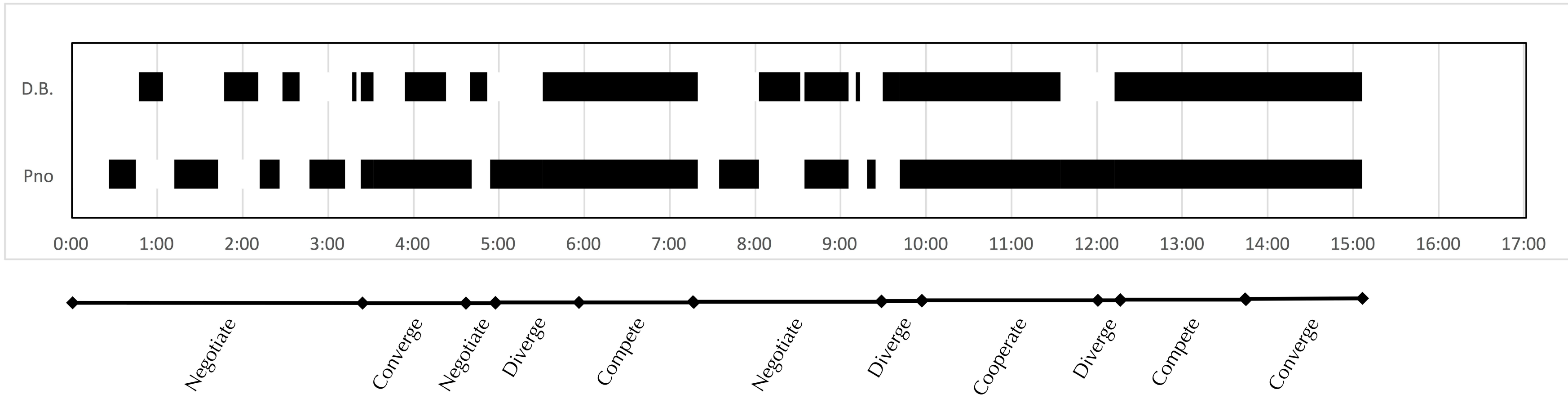


Figure 6.6 Performance 1 (Schallfeld): interaction timing.

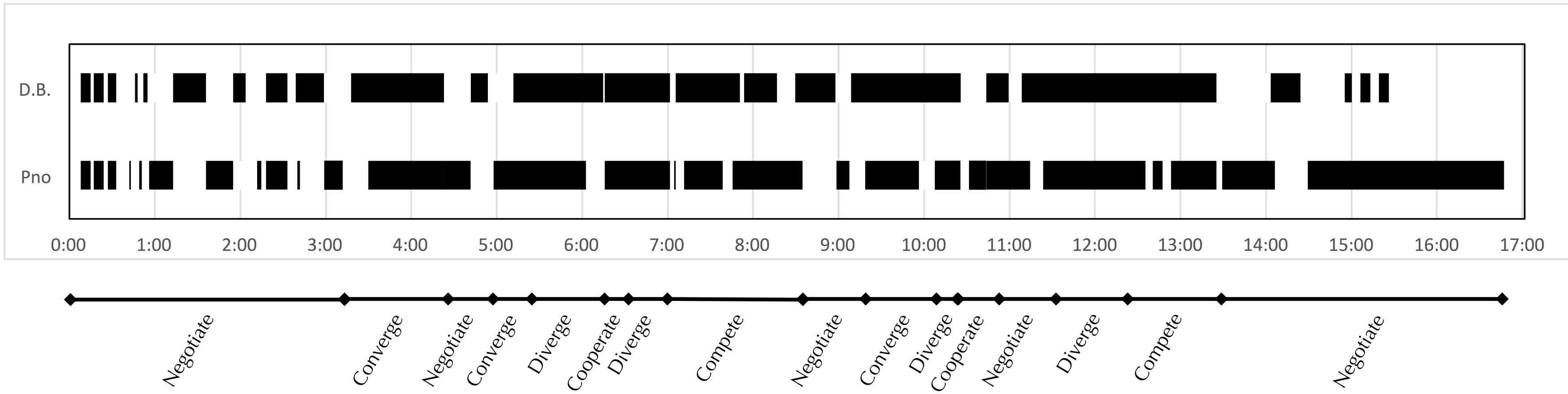


Figure 6.7 Performance 2 (Klangforum): interaction timing.

In addition to technical challenges, such as determining the exact duration of sound events, the analysis of the interaction timing between the two musicians posed a series of music-analytical challenges that had to be addressed along the way. The most important of these challenges was defining the musical "unit" that would serve as a basis for this analysis. A form-agnostic approach using sound events, i.e., sounds with a clear beginning and end, as a basis was considered but rejected due to its lack of music-analytical relevance. Instead, the analysis was based on criteria such as phrasing and turn-taking. For instance, the musical phrase depicted in Figure 6.8, which consists of a repeated timbral and rhythmic pattern, was treated as a single musical action.

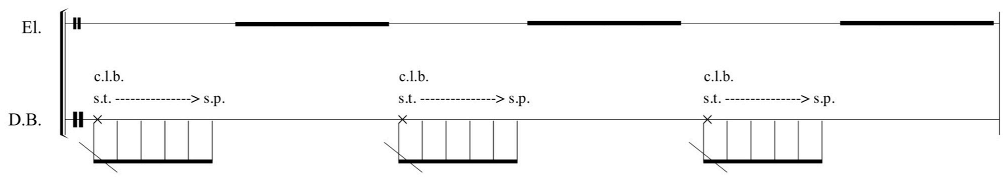


Figure 6.8 Musical phrase consisting of more than one sound events.

The interaction analysis of the two performances revealed some similar patterns across them. One of the similarities concerns the distribution of performance time between the two musicians (Figure 6.9). In both cases, the piano "solo" segments seem to occupy significantly more time than the double bass "solos", with the difference between the two being more extreme in performance 1.

Interestingly, this asymmetry in the interaction between the two instrumentalists seems to extend to the initiation of musical changes: 80% and 85% of the musical changes in performances 1 and 2 respectively were initiated by the pianists. This is particularly noteworthy, as there is nothing in the score or performance instructions that would suggest or evoke such an asymmetry. Collecting data from more than two performances could help determine whether this is a coincidence or an unintended emergent property of the IMS, the performance instructions or the affordances of the two instruments.

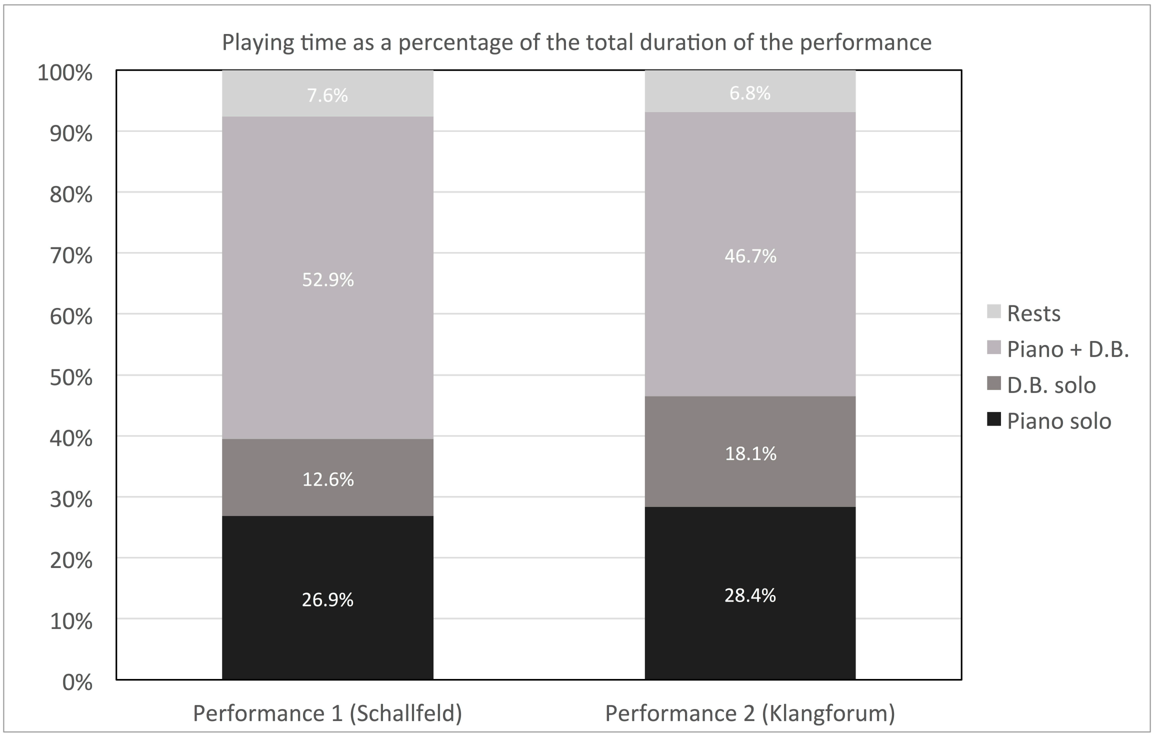


Figure 6.9 Distribution of playing time between the musicians in performance 1 (Schallfeld) and 2 (Klangforum).

6.4.3 Sound Material

As far as sound material is concerned, the call-and-response sections of both performances are characterized by a high degree of timbral and textural similarity between consecutive musical actions. The similarities between the two performances extend to the sound material used in "compete". Concretely, both ensembles focused primarily on variations of the musical actions provided in the score and only occasionally introduced some new playing techniques or variations of musical actions from other interaction scenarios, such as sliding a rubber mallet on the piano strings (performance 1), hitting the piano keys while applying preparations to the strings and playing *col legno battuto* or scratch tones on the double bass (performance 2).

Video-based interaction analysis played a central role in identifying these playing techniques, a task that would have been considerably more challenging if only audio recordings were available. This points to yet another advantage of video analysis as a music-analytical method: its usefulness in analyzing not only human-human, but also human-object interaction (in this case, the interaction between the musicians and their instruments).

To illustrate and analyze the relationship between sound material and form, spectral descriptors were computed and their values were compared to the formal analysis presented in section 6.4.1. Figure 6.10 shows the Spectral Centroid and RMS Amplitude, juxtaposed with the sonogram of the recording of performance 1.[14] The differently colored regions of the graphs correspond to different interaction scenarios (for details see figure legend). In this performance, "diverge" and "compete" seem to be associated with louder dynamics, while "converge"/"cooperate" is associated with softer dynamics. "Negotiate" seems to have a less narrowly defined profile, covering a wider range of spectral centroid and amplitude values. The increase in the value of the spectral centroid in "converge"/"cooperate" is mainly a result of the type of sound material provided for this scenario (i.e., mostly overtones).

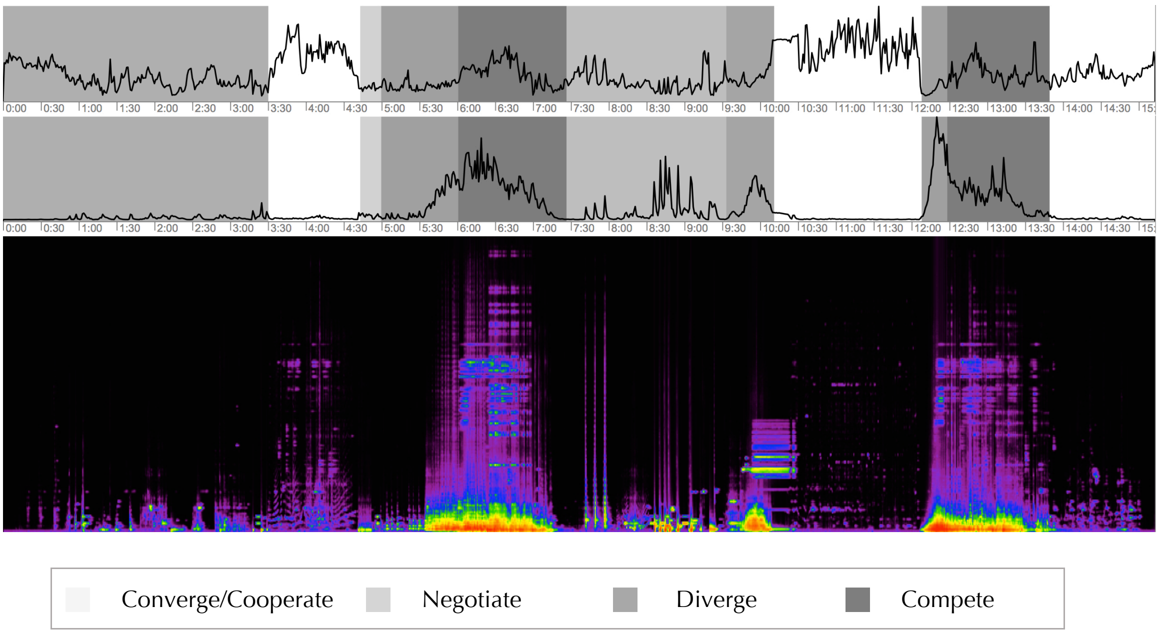


Figure 6.10 From the top down: spectral centroid, RMS amplitude and sonogram of performance 1 (Schallfeld).

In performance 2, the spectral and dynamic profiles of different interaction scenarios seem to generally follow the same patterns as in performance 1 (e.g.,

[14] Spectral descriptors were computed and plotted using the software *EAnalysis* (Couprie 2016).

“compete” is generally associated with louder dynamics than “converge”/”cooperate”), though the differences among them are less extreme (Figure 6.11). Interestingly, in both interpretations “compete” is characterized by louder dynamics, a performance choice that is not in a direct causal relation with the instructions given to the performers. In this scenario, the musicians are instructed to play “novel” sounds (i.e., sounds that are spectrally different from what they have played so far). The IMS only responds to the musician playing the most novel sound material (i.e., the sound that differs the most from previously played sounds) and does not take into account or respond to the amplitude of the two audio inputs.

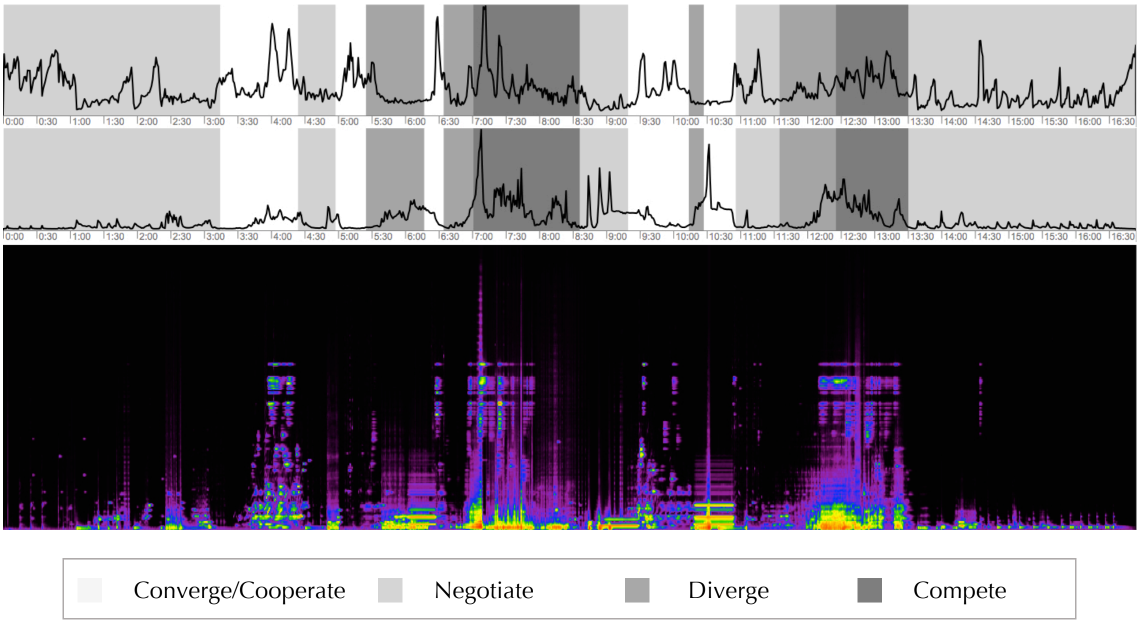


Figure 6.11 From the top down: spectral centroid, RMS amplitude and sonogram of performance 2 (Klangforum).

Besides these general observations regarding some crude differences between the interaction scenarios involved in the piece, the insights gained through this descriptor-based analysis are admittedly limited. As audio descriptors can be more contingent to the recording than descriptive of the performance *per se*, caution must be taken in order not to draw conclusions based exclusively on them. The values of spectral descriptors, such as the spectral centroid, can depend largely on the types of microphones used to record the performance, their distance from the sound sources, the level of background noise in the location and many other factors. This is not specific to the descriptors used in this analysis, but concerns spectral analysis in general, as well as the 'intentionality' (Ihde 1978) of recording technologies, i.e., their directedness towards sound.[15]

In some cases, even the reliability of the descriptor at hand can be an issue (e.g., in the case of pitch or onset detection). Indeed, an analysis of the density of sound events in the two performances was attempted but abandoned due to the insufficient accuracy of available onset detection algorithms.

As, for the purposes of this analysis, a direct comparison between spectral data obtained from two different recordings proved to be rather problematic, self-similarity matrices were computed instead to visualize the degree of spectral similarity within the same recording. In the self-similarity matrices shown below (Figures 6.12 and 6.13), each sound file is depicted as a square with both axes representing time, i.e., the bottom left corner corresponds to the beginning of the recording, while the top left and bottom right corners correspond to the end. For each 1-second-long segment of the recording the matrix depicts its spectral similarity to every other 1-second-long segment. The degree of similarity between the spectral content of two different segments is inversely proportional to the brightness of the corresponding point in the matrix, i.e., the darker the color the more similar the spectral content. As a result, there is always a black line running diagonally from the bottom left to the top right corner, where each segment is compared to itself.

[15] According to Verbeek (2008), a tape recorder's intentionality towards sound differs fundamentally from a human listener's intentionality, as it is unable to focus on the sonic foreground and suppress background noise.

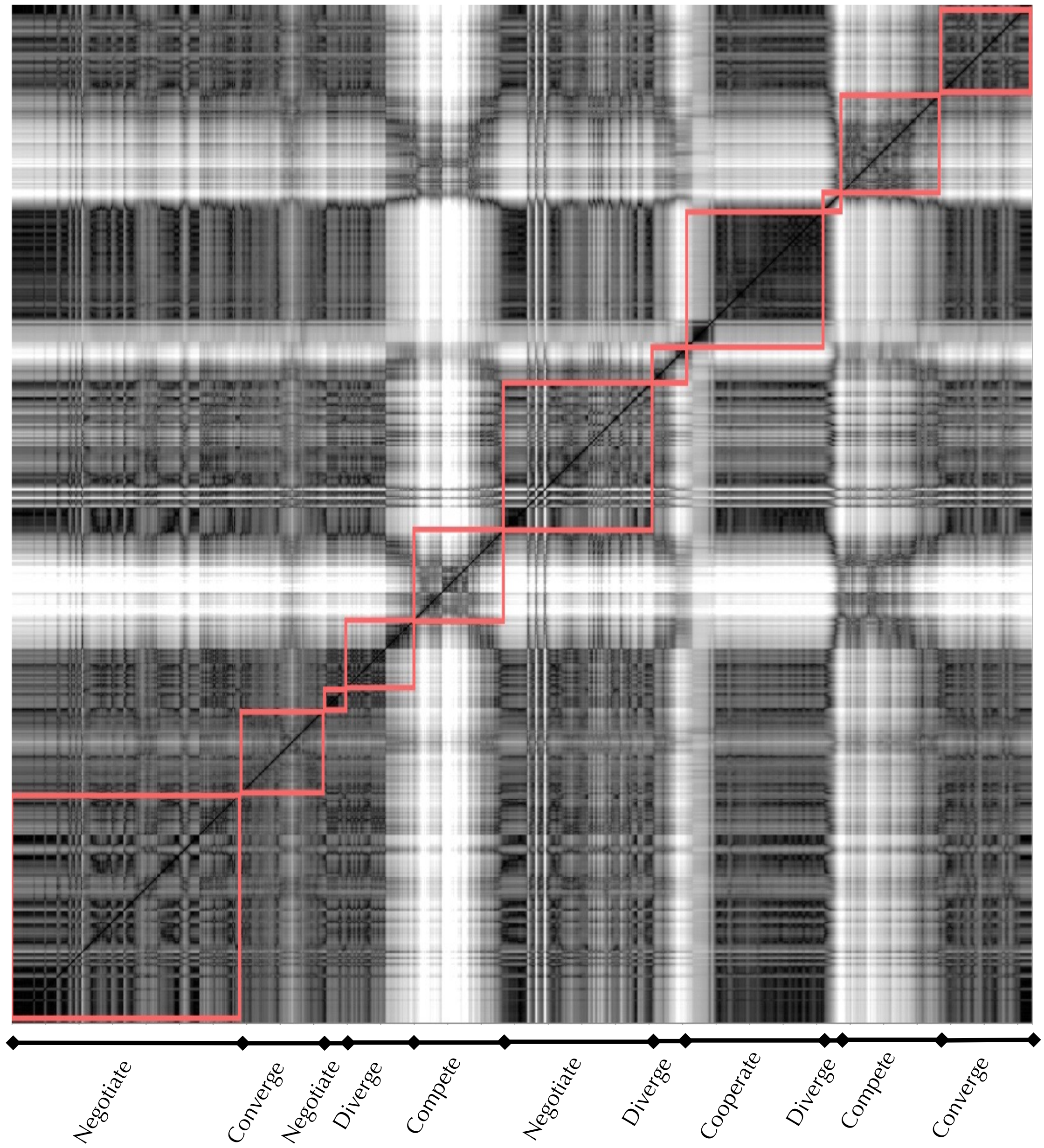


Figure 6.12 Performance 1 (Schallfeld): spectral self-similarity matrix.

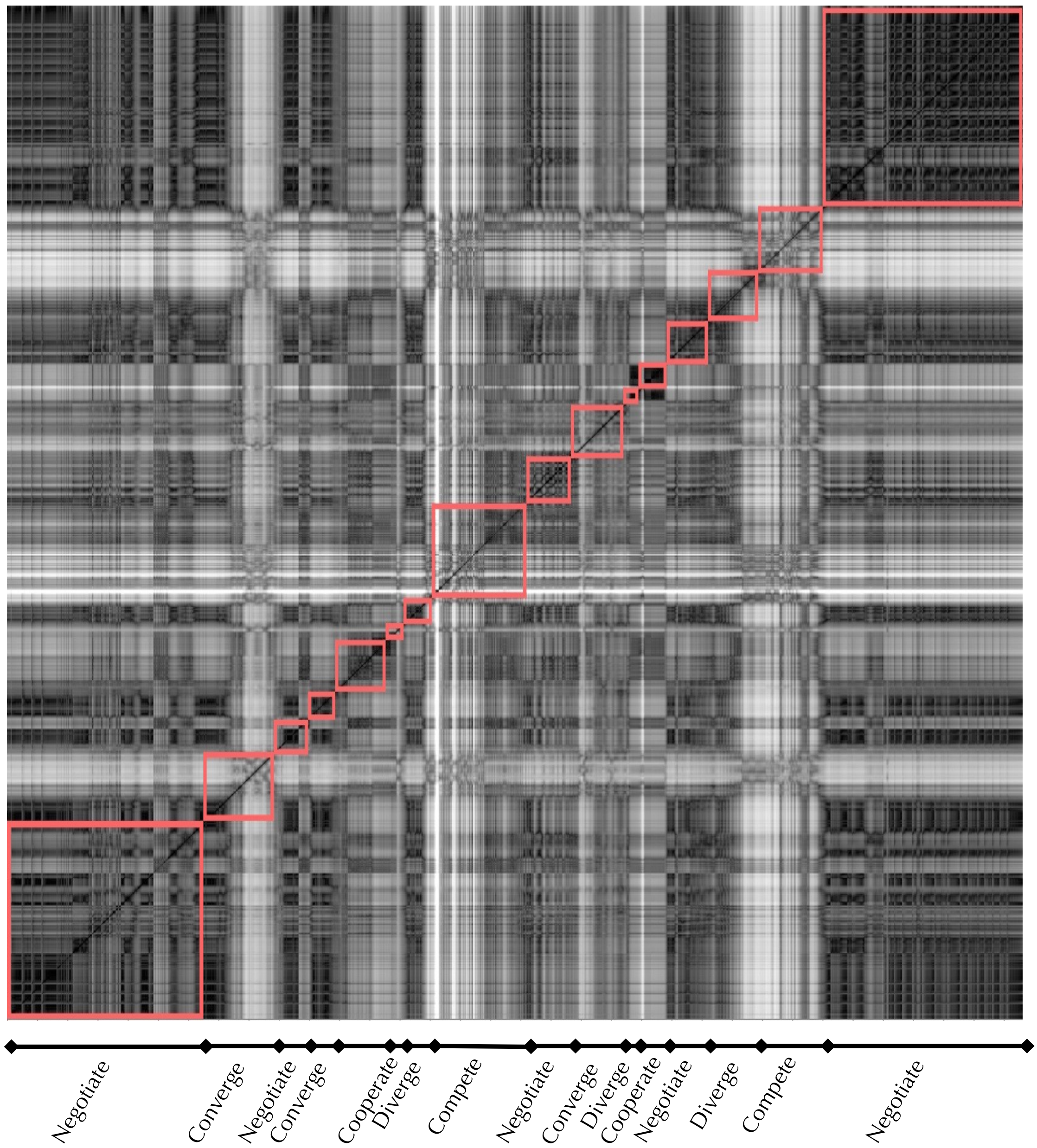


Figure 6.13 Performance 2 (Klangforum): spectral self-similarity matrix.

A juxtaposition of the self-similarity matrix of performance 1 with its formal analysis unveils some interesting relationships between form and sound material. All instances of "negotiate" seem to involve repetitive similarity (checkerboard patterns), which can be attributed to the dialogue-like interaction paradigm that this scenario is based on (i.e., musicians taking turns and "responding" to each other by playing similar textures and timbres). Instances of "converge"/"cooperate" appear to be generally homogenous in terms of sound material, while repetitive similarity is a rare occurrence in them and appears only as a result of amplitude fluctuations (e.g., in the last instance of "converge"). "Diverge" and "compete" display a lower degree of self-similarity, which means that they are generally less homogenous with respect to spectral content. This is anticipated, as the performance instructions for "compete" explicitly call for "novel" sound material. These two scenarios are not only less self-similar, but also the least similar to any of the other interaction scenarios, as evidenced by the white stripes running both vertically and horizontally across the matrix.

The self-similarity matrix of performance 2 (Figure 6.13) shows both similarities and differences to that of performance 1. A consistent pattern across the two performances is the presence of repetitive similarity in "negotiate", as well as the low degree of self-similarity of "compete", in comparison to the other interaction scenarios. "Diverge" is less consistent in comparison to performance 1 and occasionally displays repetitive similarity (e.g., in the first instance of the scenario). Finally, the spectral self-similarity of "converge"/"cooperate" seems to vary significantly across different instances of the scenario, with the first instance being among the least self-similar sections of the whole performance and comparable to the two instances of "compete". Overall, the self-similarity matrix of performance 2 seems to be consistent with its formal analysis (i.e., qualitative and quantitative data seem to agree): larger sections are hard to identify and the form seems to consist of shorter episodes succeeding each other in an almost seamless manner. The different instances of "compete", while less self-similar than instances of other interaction scenarios, are less clearly delineated and more nuanced with respect to their spectral proximity to other sections of the piece.

6.5 Discussion

This chapter described an exploratory performance-centered approach to the analysis of interactive compositions, using the author's *Converge/Diverge* for piano, double bass and interactive music system as a case study. This analysis encompassed both traditional music-analytical foci, such as form and sound material, and some less conventional ones, which are nevertheless essential to the perception and experience of interactive performances (e.g., human-human and human-object interaction).

Some of the most interesting insights gained through this analysis concern the different ways in which the two ensembles approached some of the more "open" and indeterminate aspects of the piece. The comparative analysis of the two performances revealed that the two interpretations materialize distinctively different notions of musical form and interaction, manifested in the interaction timing and the visual communication between the musicians.

The similarities between the two performances are as noteworthy as their differences and reveal some potentially emergent, unintended properties of the interaction concept and its implementation. For instance, in both interpretations the notion of a "competitive" interaction seems to be associated with louder dynamics. Additionally, the interaction analysis of the two performances suggests that the piece might privilege an asymmetric interaction between the two instrumentalists, as in both performances the distribution of playing time and the initiation of musical changes are skewed towards the pianist. This is a striking finding that is relevant not only from a music-analytical, but also from a compositional perspective (e.g., this asymmetry could potentially be counterbalanced through a revision of the performance instructions or code).

Nevertheless, it is important to note that these observations are context-bound and any generalization should be made with caution. Each performance of an interactive work is only a partial realization of the possibilities it encompasses. Therefore, any performance-centered analysis of an interactive work should be focused on gaining insights into the specific interpretative choices made in a given performance, rather than generalizing from these to other (past or future) performances – even ones by the same performers.

In one of the most "open" interaction scenarios of the piece ("compete"), the musicians of both *Schallfeld* and *Klangforum* based their interaction on variations of the musical actions proposed in the score. Yet, this was not the case in a subsequent performance of the piece by Jana Luksts (piano) and Evan Runyon (double bass), who explicitly stated their inclination to explore different sound material than what *Klangforum* and *Schallfeld* had used in their own interpretations (Jana Luksts and Evan Runyon, in discussion with the author, November 2019). Jana Luksts, in particular, shared her intention of playing exclusively on the piano keys, as opposed to inside the soundboard, an interpretative decision that aimed to sharpen the contrast between "compete" and the other interaction scenarios. Unfortunately, Evan and Jana's performance of the piece was not recorded and could therefore not be included in this analysis. Nevertheless, their interpretative choices point to a very interesting aspect of interactive works: the way in which past realizations can inform future performances and contribute to the exploration of new sonic territories and the expression of interpretative individuality. An analysis of this phenomenon based on interviews and focus group discussions with musicians who have performed the piece would be interesting in its own right, but is beyond the scope of the analysis presented here.

While the findings described in previous sections are specific to the analyzed work, the analytical methods presented here could be relevant for a broad range of interactive, generative and participatory musical works. This analysis employed a variety of methods, from a traditional formal analysis to an audio-driven analysis and an exploratory adaptation of video-based interaction analysis. The latter, along with functioning as a complementary tool to aural analysis (e.g., by helping identify different playing techniques and estimate the duration of sound events more accurately), provided crucial insight into the interaction between the musicians. Concretely, it helped illustrate the in-the-moment decision-making, adaptation and negotiation of intentions between the musicians – all aspects of central importance for interactive aesthetics. Some of the challenges associated with the use of this method relate to technical aspects of video production and post-production, such as the selection of appropriate post-production techniques (e.g., "split screen"), to ensure that all camera angles are visible at all times and minimize information loss.

7 Experiments in Computational Aesthetic Evaluation and AI Bias

Bias, for Bass Clarinet and Interactive Music System

7.1 Introduction

This chapter describes a subversive approach to Music AI, focused on the exploration of a specificity of machine learning algorithms. Specifically, the composition described here explores the concept of AI bias, a phenomenon normally viewed as a limitation of machine learning models, in an approach that is suggestive of a critical perspective on AI, yet acknowledges the creative potential of such limitations. In *Bias*, for bass clarinet and Interactive Music System (IMS), a computer music system using two Neural Networks trained to develop "aesthetic bias" interacts with the musician by evaluating the sound input based on its "subjective" aesthetic judgments. The composition problematizes the discrepancies between the concepts of error and accuracy associated with supervised machine learning, and aesthetic value as a non-measurable quality, the attribution of which is not only highly subjective, but also reliant on social agreement and extra-musical (e.g., historical) context.[16]

Floridi (2020) suggests that in anticipation and prevention of a new AI winter, i.e., a period of funding cuts and decreased interest in AI research, we need to ask ourselves whether AI solutions are going to *replace* previous solutions, *diversify* them, or *complement* and *expand* them. Expanding the musical work-concept, as well as notions

[16] A video documentation of the piece is available at:
https://www.artemigioti.com/demos/Bias.html.

of musical authorship are among the objectives of the critical and subversive approach to Music AI described in this chapter. The premise behind this approach is that the affordances of machine learning algorithms can open up creative possibilities that challenge us to question some fundamental constructs around music composition, performance and perception (e.g., the "authorship" construct, the composer/performer divide, modernist ontologies of the musical work etc.). Machine learning algorithms in this context are understood as 'secondary agents', i.e., material entities through which 'primary agents' (intentional beings) exercise their agency and are as essential to action as intentionality (Gell 1998, chap. 2). The 'secondary' agency of machine learning algorithms lies in the particular ways in which their affordances shape musical thinking and transform compositional and performance practices.

7.2 *Bias*, for Bass Clarinet and Interactive Music System

Bias, for bass clarinet and interactive music system, explores the concept of computational aesthetic evaluation as a decision-making mechanism in human-computer music interaction. The question around which the work is centered is twofold: how can computers make aesthetically informed decisions in their interaction with human musicians, and how can the machine autonomy afforded by computational aesthetic evaluation shape notions of musical authorship?

A symmetrical human-machine interaction, in which not only the musician but also the computer can make decisions that change the course of the performance, lies at the core of interactive music. The work described here explores computational decision-making, focusing on the concept of computational aesthetic evaluation as a parallel for the aesthetically-driven decisions made by musicians in interactive and improvised musical contexts. The basis for this composition was a series of experiments aimed at developing a computer music system with idiosyncratic behavior, "subjective" aesthetic preferences and capable of communicating intentions and "cognitive states" through musical actions.

Concretely, the interactive music system performs an aesthetic evaluation of the musician's input in real-time and imitates sounds and textures it finds "interesting", but remains silent or proposes new sound material when it loses interest in the musician's input.

The aesthetic evaluation of the musician's input is performed by two Neural Networks trained on data collected with the help of clarinetist Szilárd Benes and evaluated by the composer based on her subjective aesthetic judgments. Recordings of improvisation sessions made with the help of the clarinetist were segmented and evaluated by the composer using a Likert-type scale from 1 ("not at all interesting") to 5 ("extremely interesting") and were used as training examples for the Neural Networks. Two separate pools of data were collected and used as training sets for two separate Neural Networks: one performing aesthetic evaluation on a sound event basis and the other on a texture basis.

The features used for sound event evaluation include the Mel Frequency Cepstral Coefficients (MFCCs), spectral flux and amplitude of the sound event averaged over its duration. In the case of the amplitude, the standard deviation is used as well in order to track amplitude fluctuations. The features used for texture evaluation include the mean spectral distance between consecutive sound events, measured by calculating the Euclidean distance between averaged MFCC vectors, the mean and standard deviation of Inter-Onset-Intervals (IOIs) and the mean and standard deviation of the durations of individual sound events. Texture evaluation is performed every second for the last five seconds of audio, while sound event evaluation is performed continuously, using an FFT window of 1024 samples and a hop-size of 0.5. Features are averaged over the (up-to-this-moment) duration of the sound event, i.e., if a sound event is in progress, features are averaged between its onset time and the current time point. The start and end time of individual sound events are determined using a k-nearest neighbor algorithm, trained to distinguish between clarinet sounds and background noise and using MFCCs as an input.

Unlike machine learning applications that involve objective ground-truth labels (i.e., "correct" answers), in this experiment the process of data labeling was explicitly focused on exploring the annotator's/composer's subjective bias, revealing some interesting aspects of intra-rater reliability relating specifically to aesthetic judgments. Intra-rater reliability refers to the consistency with which a single rater labels data over several trials. The issue of intra-rater reliability was brought to the foreground accidentally, due to the need to repeat the data labeling process for technical reasons. Concretely, the labeling and feature extraction process had to be repeated in order to test the efficiency of different sets of features, by comparing the accuracy of the resulting machine learning models. However, intra-rater reliability seemed to be an issue even within the same trial, for instance, due to fatigue caused by listening to similar sound material for a long time. While consciously rating sounds with similar spectral characteristics with similar scores could have helped resolve this issue, such an approach was considered as contradictory to the premise of this work, which lies in the exploration of aesthetic judgments as manifestations of complex value systems and psychological processes that are intangible and subject to change. Consequently, any apparent lack of consistency in the labeling process was treated as an integral part of the phenomenon being modeled (i.e., aesthetic judgments), rather than a limitation that needed to be overcome.

The training process and subsequent testing of the obtained machine learning models revealed that the Neural Networks had indeed developed some interesting forms of "bias". For instance, the Neural Networks seemed to prefer low frequency sounds over high frequency ones and static, drone-like textures consisting of sustained sounds over fast and virtuosic melodic passages. These preferences represent reasonable, though somewhat exaggerated assumptions about the author's aesthetic preferences, demonstrating that the machine learning models did in fact "learn" some interesting correlations between the features and evaluations of individual sounds and textures, yet failed to capture the subtleties of the author's aesthetic judgments.

At this stage, the machine learning models could have been improved further by collecting more examples or adding new features. However, as the premise of this piece was not to simulate the author's aesthetic judgments as accurately as possible, but rather to explore the artistic potential of AI bias, any "creative" or "distorted" (i.e., exaggerated) interpretations of the training data were instead exploited for their aesthetic potential. For instance, the preference of the machine learning model for slowly evolving, drone-like textures influenced the design of the generative processes of the IMS, largely determining the aesthetic direction of the piece.

In addition to "mimicking" the musician's input and remaining silent, in certain cases the IMS may try to "redirect" the musician's attention towards specific types of sound material. An example of this behavior is its response to detected onsets (i.e., key clicks). This includes the use of a series of signal processing techniques (e.g., convolution, comb filters etc.) applied only to the onset segment of the signal and meant to deter the musician from playing dense melodic passages (detected as frequent fingering changes) and encourage them to explore key clicks and other percussive sounds instead.

Aside from decisions made on a sound event basis, which generally involve a choice between responding and remaining silent, the computer monitors and influences the formal development of the piece, by occasionally taking the "lead" and introducing new sound material. This behavior indicates that the computer has lost "interest" in the musician's input for a while. The choice between "following" and "leading" is based on a relative evaluation of the last 20 seconds of the performance in relation to previous 20-second sections.

The score of the piece consists of a pool of partially notated musical actions that are open with respect to pitch and duration and can be played any number of times and in any order. Durations are relative and given in "breaths", rather than in seconds or through meter and tempo indications. For example, the following excerpt depicts a musical action that consists in transitioning repeatedly from air tone to pitch and back, while playing a multiphonic. In this example, there are no pitch or fingering indications, meaning that the musician is free to play *any* multiphonic, while the duration of the action is specified as "4 breaths".

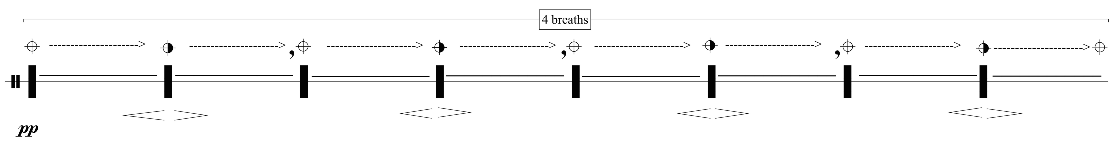


Figure 7.1 *Bias*: score excerpt.

The high level of abstraction involved in the score means that the musician's actions are guided – at least in part – by the interaction affordances and idiosyncratic behaviors of the IMS through sonic stimuli: the concrete sounds played in a given performance emerge as a result of a negotiation between the musician's choices and the computer's aesthetic preferences.

The creative agency of the performer in the piece is underscored by the fact that all sound material generated by the IMS during the performance is collected during its interactions with musicians – that is, all musicians that have performed the piece up to the present moment. Specifically, the IMS stores the spectral data of sounds it finds "interesting" in a sound database, which is continuously updated. These updates consist in both adding and removing sounds from the database based on their overall evaluation (i.e., keeping the most "interesting" sounds in each iteration). This effectively means that none of the electronic sounds heard in the performance were "composed", a feature that adds to the high degree of autonomy of the IMS.

This sound database functions as a form of musical memory, connecting past instances of the piece to the present and maintaining continuity beyond a single performance. By "echoing" past performances, the IMS facilitates a mediated and asynchronous dialogue among performers, whereby each musician both contributes to and interacts with a collectively assembled sound corpus.

The ability of the IMS to autonomously collect and update its own sound database has yet another implication for the identity of the work. Namely, the electronic sounds heard in the piece can change significantly over a large number of instances (i.e., performances), a process over which the composer has no control. This process is suggestive of a meta-generative approach to music composition, in which the object of composition is not a space of sonic possibilities, but rather the behavior that generates it. The IMS and, by extension, the work evolves autonomously through "experience" (i.e., real-time interaction with human musicians), questioning traditional notions of authorship and ontologies of the musical work.

The recorded instrumental sounds are analyzed by the IMS using a series of band pass filters and envelope followers and resynthesized using additive synthesis. Instead of an exact resynthesis, the computer creates spectral variations of the initial sound, the relation of which to the original can be more or less recognizable. This is achieved by reducing the spectrum to a small number of frequencies (e.g., reproducing only the most prominent frequencies, or resynthesizing a filtered version of the original sound). This allows the algorithm to generate sound material that, though originally derived from instrumental sounds, is still distinct from the acoustic sound and, due to the use of additive synthesis, has a certain degree of plasticity. The computer can generate and interpolate between a virtually infinite number of spectral variations of a single sound and, by changing the degree of spectral "compression" applied to it, interpolate across the recognizability spectrum.

Figure 7.2 *Bias*: rehearsal with Szilárd Benes.

7.3 Computational Aesthetic Evaluation, Aesthetic Experience and Aesthetic Theory

In addition to AI bias, the composition described in this chapter explores computational aesthetic evaluation in an approach that implies a critical perspective towards reductionist approaches to aesthetic evaluation and comments on the gap between computational aesthetic evaluation and aesthetic experience and theory.

While artistic applications of computational aesthetic evaluation in generative systems generally seem to acknowledge the complex and subjective nature of aesthetic judgments (McCormack and Lomas 2021; Galanter 2014), applications of computational means and crowd-sourced aesthetics in the evaluation of artworks often appear to be based on rather simplistic assumptions about both aesthetic experience and theory. A common approach to the aesthetic evaluation of musical works involves the use of formulaic aesthetic measures such as Zipf's law (1949), which states that the occurrence frequency of an event is inversely proportional to its statistical rank, and Birkhoff's (1933) aesthetic measure, which is expressed as the ratio between order and complexity. Applications of Zipf's law in the evaluation of musical works (e.g., Manaris, Romero, and Machado 2005; Manaris et al. 2007) seem to equate concepts such as 'pleasantness' or 'popularity' with aesthetic value and have been criticized for assuming that aesthetic value can be judged based on universal aesthetic principles (Kalonaris and Jordanous 2018). Furthermore, the relevance of Zipf's law for musical styles that favor repetition or stasis (e.g., minimalism and noise music) has been challenged (Kalonaris, Gifford, and Brown 2019).

Galanter (2012) suggests that the fields of psychology and neurology could provide useful insights for computational aesthetic evaluation. He specifically cites psychological models of human aesthetics, such as Arnheim's (1974) law of *Prägnanz*, which states that perceptual cognition prioritizes wholes and clarity of structure over individual components, Berlyne's concept of *arousal potential* and its relation to *hedonic response* (Berlyne 1960, 1971), and Martindale's (1988) neural network model of aesthetic perception that relates preference with *prototypicality* (i.e., the degree to which a stimulus is typical of its class).

However, the assumption that aesthetic experience can be reduced to perception is debatable. In a discussion on the 'gap' between empirical aesthetics and aesthetic experience, Makin (2017) criticizes what he calls the 'reductive psychophysical approach' to aesthetic science, which involves varying a stimulus dimension x and measuring some subjective experience y. His criticism concerns the assumption that stimulus dimensions are orthogonal and their effects independent, as well as the nature of the responses that can be evoked in a lab setting (i.e., 'cold' cognitive evaluations, as opposed to 'hot' emotional reactions). As Makin points out, an artwork is the opposite of a controlled stimulus: it is a 'labyrinth' of interacting perceptual and semantic dimensions which cannot be easily isolated or quantified (188).

Similarly, Leder and Nadal (2014) criticize Berlyne's (1971) psychobiological aesthetics as 'weak and overly simplistic' (455) and argue that the psychological mechanisms involved in the appreciation of art extend beyond the perception of aesthetic qualities to 'grasping an artwork's symbolism, identifying its compositional resources, or relating it to its historical context' (445), and that an aesthetic episode consists in feedback and feedforward interactions among cognition, perception and emotion. Their approach is based on an information-processing model of the aesthetic experience of art that takes into account declarative knowledge, domain-specific knowledge and personal taste, and acknowledges the ambiguities involved in the perception and interpretation of art (Leder et al. 2004). This model suggests that aesthetic experience begins before perception, with the social discourse and context that shape expectations and contribute to the artistic status of the work. In line with Dewey's (1934) view of experience as interaction with the physical, cultural and institutional environment, Leder et al. (2004) argue that contextual factors, such as presentation formats, play an important role in aesthetic experience.

The importance of domain-specific knowledge for aesthetic experience is evidenced in a study by Kozbelt (2006), in which non-artists and art students were asked to rate 22 in-progress states of Henri Matisse's *Large Reclining Nude*. The study revealed significant differences in aesthetic judgment criteria between the two groups. Art students valued originality, while non-artists seemed to prioritize technique and realism and judged the painting as getting worse over time, as the abstraction of the image increased. To make matters more complex, the aesthetic value of an artwork might not lie in its physical manifestation, but rather in its concept (e.g., conceptual art) or the social relations it materializes.

The ambiguity surrounding the concept of aesthetic experience and its complex, overlapping dimensions have been a ground for debate not only in aesthetic science and empirical aesthetics, but also in aesthetic theory. Shusterman (1997) identifies four dimensions of aesthetic experience: an *evaluative*, a *phenomenological*, a *semantic* and a *demarcational-definitional* one, which concerns the demarcation of art from other domains of human activity. He attributes the marginalization of the concept of aesthetic experience in analytic philosophy, which led to an 'anaesthetization of aesthetics' by philosophers like Goodman and Danto, to tensions generated by these four dimensions and a 'deep confusion about this concept's diverse forms and theoretical functions' (29).

Galanter (2012) claims that computational aesthetic evaluation is a difficult and fundamentally unsolved problem. This claim is supported by the divergent and often contradictory assumptions about the concept of aesthetic experience both across different fields and within the same discipline (e.g., Shusterman 1997). Far from trying to solve this problem, the composition described here attempts a 'meta-aesthetic exploration' (Galanter 2012, 256), which involves artificially created aesthetic standards rather than simulated human aesthetics, while acknowledging that aesthetic preferences are culturally grounded, highly subjective and hard to rationalize and predict. By trying to do exactly that, i.e., predict and simulate aesthetic judgments, it attempts a *reductio ad absurdum* (Latin: "reduction to absurdity") of the concept of aesthetic evaluation. It questions whether it is possible to simulate aesthetic judgments or trace the criteria on which they were based using computational means. Considering that aesthetic preferences are subject to change – both on a cultural and individual level – and are often hard to describe in propositional terms, what is being simulated here is at the same time ephemeral, erratic and intangible; essentially: impossible to simulate.

Another contradiction that is made apparent in this process concerns the focus of supervised learning algorithms on closed-ended tasks, i.e., tasks that have "right" answers, as contrasted with the open-endedness of artistic practices. Particularly in artistic practices that prioritize interactivity and, by extension, unpredictability and emergence, the intended role of machine agency is not to predict the "right" or most "accurate" answer, but rather to produce "creative" and even "unlikely" answers that the composer-programmer might not have envisioned. A concept as impalpable and ambiguous as that of (perceived) aesthetic value offers an interesting ground for artistic experimentation, gravitating away from right/wrong dichotomies (or spectra) and towards autonomous and idiosyncratic agentive behaviors that can produce unexpected musical outcomes.

In *Bias*, the discrepancies between the subjectively and culturally grounded attribution of aesthetic value, on the one hand, and the concepts of error and accuracy normally associated with supervised learning algorithms, on the other, are problematized and brought to the foreground. The work aims to draw parallels between aesthetic judgments as inherently "biased" (i.e., subjective) and AI bias, a phenomenon that consists in machine learning algorithms making arbitrary assumptions about data, or amplifying any bias present in the data. The composition takes a critical and subversive approach to machine learning, the aim of which is not to simulate the composer's aesthetic preferences as accurately as possible, but rather to use them as a departure point for the development of AI bias. What is essentially a specificity of machine learning algorithms and normally viewed as an unwanted outcome of the training process is explored for its potential to produce idiosyncratic agentive behaviors.

7.4 Distributed Human-Human and Human-Computer Co-creativity

The use of computational aesthetic evaluation in *Bias* serves yet another purpose: the ability of the IMS to autonomously collect and update its own sound material in its interactions with human musicians signifies a notion of the musical work as an ever-evolving process, shaped by the mediated sociosonic relations among composer, performers and IMS.

The musical interactions at play here extend beyond the real-time interaction between the performer and the IMS to the technologically mediated interactions between composer and performer, as well as among different performers. The IMS is central in this process, as it is through its agency that compositional intentions and interpretative choices are communicated, negotiated and re-contextualized. Compositional intentions are mediated through the aesthetic preferences and idiosyncratic interaction affordances of the IMS, while interpretative choices are "echoed" in its autonomously collected sound database and negotiated among different performers and across various instances of the work.

The result of this process is a musical work that is collaboratively constituted by the composer, the interactive music system and its performers. While any musical practice that involves a division of musical labor between composition and interpretation can be considered as collaborative, the approach described here involves a rebalancing of the relation between composition and interpretation, as well as a more nuanced approach to co-creativity that extends beyond co-located and synchronous forms of collaboration to a technologically mediated collaboration that is distributed in space and time.

7.5 Compositional Process and Methods

Related to the aesthetic ends of the work are the methods used in its creation. In order to balance the trade-off between authorship and interpretative freedom in the piece, a series of improvisation experiments were designed, conducted with the participation of the musician and analyzed using ethnographically informed research methods. These experiments included a 'naïve' and several 'informed rehearsals' (Hsu and Sosnick 2009), the difference between the two being whether the musician is given information regarding the interaction affordances of the IMS prior to the improvisation. Data from these improvisation sessions was collected using a combination of methods, including observation, a questionnaire and a semi-structured interview. These methods were selected for their complementarity in terms of perspective, with observation focusing on the composer's perspective and the questionnaire and interview on the performer's, and their potential to facilitate a creative dialogue focused on open-ended questions/problems and creative discovery. These experiments were conducted with the participation of clarinetist Szilárd Benes. A repetition of these experiments with other musicians would potentially benefit this research, but was not possible due to time constraints and limited resources.

The methods mentioned above are considered as ethnographically informed or inspired rather than purely ethnographic, as their use within an artistic research context inevitably meant that they had to be adapted considerably. The intent behind the selection of these methods is strongly aligned with the 'transactional' and 'subjectivist' epistemology of the constructivist research paradigm, in which investigator and object of investigation are interactively linked and knowledge is created as a result of and through that interaction (Guba and Lincoln 1994). Yet, in the context of artistic research, "knowledge" is understood in radically relativist and subjectivist terms: knowledge here is simply *insight* gained through and feeding back into the compositional process.

The purpose of the naïve rehearsal was to identify the perceived interaction affordances of the IMS and determine their effectiveness in communicating compositional intent. The question driving this experiment was: how effective are interaction affordances in guiding the performer into an action space that is aligned with the aesthetics of the piece? The broader context within which this question was explored was that of a 'subtractive' approach to the compositional process, which involves starting from an improvisational context and gradually introducing a series of constraints or instructions, until arriving at an aesthetically narrower yet, as far as concrete musical actions are concerned, still open space of sonic possibilities (Marko Ciciliani, in discussion with the author, March 2019). In the experiment described here, the starting point was a naïve improvisation with an interactive music system and therefore the perceived interaction affordances of the computer music system functioned as an additional initial "constraint", by influencing the musician's actions and the course of the improvisation. The informed rehearsals provided an opportunity to further refine these performance instructions, as well as the code, and decide which aspects of the performance should be determined through the score and which should be left to the musician (Figure 7.3).

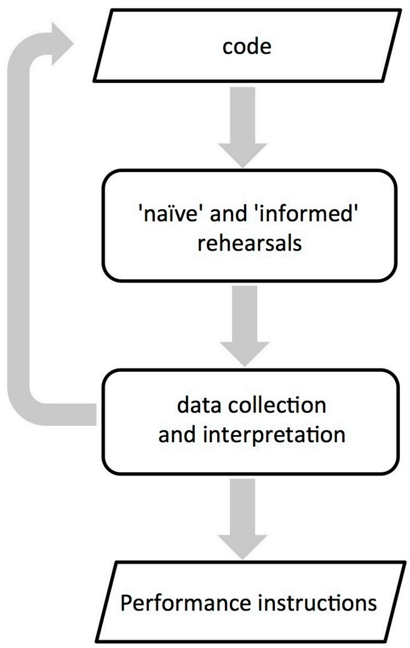


Figure 7.3 *Bias*: compositional process.

After the naïve rehearsal, the musician was asked to fill-in a questionnaire including questions on the degree of responsiveness, autonomy and agency of the IMS. In the interview that followed, the musician was asked to elaborate on some of his responses and comment on additional aspects of the improvisation that came up during this discussion.

The musician's responses to the questionnaire suggested that he was uncertain as to whether the system changed its behavior in response to his actions, as well as to whether its responses were predictable, but assessed its degree of autonomy as rather low. He agreed that musical changes introduced by the system influenced his actions and changed the course of the improvisation, but thought that there were no moments in which the computer was "leading" the improvisation. He correctly identified that the system was listening only some of the time. When asked to describe different behaviors exhibited by the system, he focused mainly on the description of different types of sound material and textures (e.g., drone-like sounds vs. percussive sounds).

In the interview that followed, the musician specified that the computer responded to percussive sound material, such as key clicks and slap tones, through some sort of 'imitation' and that its responses were varied with respect to their register and the use of effects such as 'echo'. He seemed to make a clear distinction between immediate and delayed responses to his actions (i.e., signal-processing and playback of manipulated recordings of the input), but based this distinction on the timbre of these responses rather than their timing.

The musician identified the textures of sustained 'tones' played by the computer when taking the "lead" as a distinct behavior and described them as a 'nice moment' in the improvisation. He described these sounds as 'soft' (i.e., quiet) and mentioned that, even when he tried playing something more 'aggressive', there was no discernible response from the system in terms of dynamics or timbre.

He commented that the system's degree of responsiveness was probably higher than what he had suggested in the questionnaire, but its degree of autonomy was rather low. He pointed out that in some cases the same sound material (e.g., key clicks) caused different responses and suggested that playing with the system longer and trying out different types of sound material could help him better assess its degree of predictability. Along with the system's degree of autonomy, the musician expressed criticism towards the lack of timbral and rhythmic variability in the sound material used by the computer.

When asked to explain in what ways the IMS influenced his actions and changed the course of the improvisation, he responded that it was by introducing new sounds, causing him to adapt his own sound material to the computer's output. He also mentioned that the computer responded to some, but not all of his actions, but suggested that he was uncertain whether that was because the computer was not listening all of the time, or whether it was intentional, and implied that the system might produce responses on different time-scales.

Overall, the musician was able to identify many, though not all of the behaviors exhibited by the system. He was able to distinguish between behaviors such as "following" and "leading" – though he did not use these terms to describe them. He correctly observed that the system produced different responses for different types of sound material and that its decision-making was driven by non-linear processes (i.e., the same action did not always cause the same response).

Also noteworthy is an apparent contradiction in the musician's responses. Concretely, the musician suggested that sound material introduced by the IMS caused him to adapt his actions and changed the course of the improvisation, yet he could not identify any moments in which the computer was "leading". This discrepancy could be indicative of a reluctance to associate the term "leading" with the interactive music system, despite recognizing and describing instances in which the system initiated musical changes, causing the musician to follow its "lead".

During the interview, the musician repeatedly referred to the IMS using a male pronoun ("he"), not only anthropomorphizing it, but also assigning a gender to it, a choice that does not reflect the author's/interviewer's wording. As the interviewee was more fluent in German than in English – though neither of these languages was his native tongue – this could be due to a direct translation from German (*Der* Computer; masculine). However, similar tendencies have been observed in previous experiments conducted by the author with different musicians (e.g., Chapter 4). In two cases, the gender assigned to the IMS was "male" and only in one case the performers decided to refer to the IMS as a "she" (Jana Luksts and Evan Runyon, in discussion with the author, November 2019).

Observation of the naïve rehearsal helped identify some further issues with the design of the IMS and assess how effective its interaction affordances were in communicating compositional intent. The sections of the improvisation in which the computer was "leading" seemed to be particularly effective in guiding the musician's actions towards specific timbral and textural qualities, yet allowing sonic exploration and experimentation. Already in this first rehearsal it was clear that this interaction scenario would not require any performance instructions. Similarly, the system's response to key clicks seemed to guide the musician away from highly virtuosic and dense melodic passages and towards the exploration of percussive material, such as key clicks and slap tones. In this case, however, the space of sonic possibilities created by the system's interaction affordances was still too vast and would need to be reduced further through some form of performance instructions. The perceived lack of autonomy of the system was also identified as an issue that needed to be addressed, an observation that was in agreement with the musician's comments. In the version of the code that was used in this experiment the IMS tended to remain silent, rather than propose different sound material, when it lost interest in the musician's input. The code was later revised, in order to increase the agency of the computer music system and facilitate a more symmetrical interaction between the clarinetist and the IMS.

The relationship between interaction affordances and performance instructions, as well as compositional and interpretative decisions was further refined through a series of informed rehearsals. In these sessions, the musician was asked to improvise with the interactive music system after being given some general information regarding its design and interaction capabilities, but without being given any performance instructions. Data from these sessions was collected through observation, as, in this part of the compositional process, the focus shifted from the exploration of intended and perceived interaction affordances of the IMS to the analysis and further refinement of the action space available to the performer.

One of the creative decisions inspired by such an informed rehearsal concerned the use of "key releases" instead of key clicks, as a means to create more delicate and less controllable/virtuosic pointillistic textures. The ability of the IMS to detect and respond to onsets (i.e., "attacks") prompted the musician to explore pointillistic textures consisting of percussive material, such as key clicks and slap tones, which, however, favored higher densities and dynamics. As these dense and opaque textures deviated significantly from the compositional idea behind this scenario (delicate and sparse sounds on the verge of the inaudible), the technique of "key releases" was introduced as a sort of physical "constraint" to the density and dynamics of these percussive textures. This technique consists in pressing the keys as quietly as possible and then releasing them, letting only the "release" segment of the gesture sound/get detected by the IMS. In order to make sure that pressing the keys does not activate the system's onset detection, the musician has to press the keys quietly and slowly, which means that playing high-density textures using this technique is effectively impossible.

In addition to the exploration of playing techniques and various forms of performance instructions, the informed rehearsals provided an opportunity to improve the design of the IMS and refine its decision-making processes. For instance, after a few improvisation sessions it became obvious that the IMS handled musical form in a way that lacked context-awareness. Aesthetic evaluation alone seemed insufficient in determining the duration of larger sections of the piece and the balance between different types of sound material and textures. The system never got "bored" of sounds it "liked" and, as a result, kept playing the same material for long stretches of time. As a means to increase its context-awareness, the decision-making stage of the IMS was enhanced with a "memory" that kept track of the duration of different types of sound textures during the performance, as well as a preference regarding the overall duration ratio between "drones" and "onsets", favoring the former.

The data collection methods described above (questionnaire, interview and observation) served different purposes, providing complementary perspectives on questions regarding the agency and interaction affordances of the IMS. It is important to note that, despite the fact that some of these methods are commonly employed in the evaluation of human-computer improvisation systems (e.g., Hsu and Sosnick 2009), their use in this context had a completely different purpose. The musician's contribution was valuable in identifying some shortcomings in the design of the IMS and devising effective performance instructions, yet the purpose of these experiments was not an "evaluation" of the IMS by the performer, nor a revision of the code or performance instructions based on crowd-sourced aesthetics. Far from "grounding" compositional decisions in qualitative data, this approach sought to facilitate aesthetic reflection as part of the compositional process and help crystallize the author's ideas and the aesthetic values manifested in them.

7.6 Discussion

At the time of writing this thesis, *Bias* has been premiered by Szilárd Benes at the 2020 *Ars Electronica Festival* (Linz, Austria), but has not received any further performances. In this first performance, the preferences of the IMS appeared to have a strong influence on the musician's actions, who seemed to repeat sound material that consistently evoked a response from the system. As a result, the initially vast space of possibilities available to the musician was effectively reduced to what could be described as a "common language" between the clarinetist and the IMS. Interestingly, the musician seemed to consciously avoid playing sounds that did not evoke a response from the IMS, even though the score does in no way limit the selection of sound material to sounds that the IMS responds to. Indeed, sounds that the IMS finds "uninteresting" can be employed by performers as a source of musical contrast and tension. Whether other performers will follow a similar approach remains to be seen.

As a central aspect of this piece is the "sound memory" of the IMS and its evolution over a large number of instances, more performances, particularly ones by different performers, will be necessary in order to better understand its role in shaping the identity of the work. Of particular interest for future analysis could be the frequency in which this memory gets "overwritten" and the contributions of individual performers to it. Aspects of interpretative freedom and individuality in the piece could also be studied using ethnographic methods.

8 Conclusions

8.1 Reconstructing the Narrative

Each of the musical works described in this dissertation emerged as a continuation of or response to the previous one and deals with new research questions that arose through the compositional and research process. In chapters 3 through 7, these works were presented in a chronological order, with the purpose to help the reader reconstruct the narrative that connects them and trace the iterative research process from creative ideation to implementation and aesthetic reflection.

The first piece described in this dissertation, *Neurons*, is a study on *listening*. The agentive behaviors exhibited by the interactive music system in this piece are a comment on the complexity of human auditory perception and the polysemy of the concept of *listening*. Listening is understood as the recognition of perceptual (i.e., timbral) categories, as a music-analytical process and as an active, conscious process that is agentive in its own right.

The starting point for this composition was a series of machine learning experiments revolving around the recognition of different timbral categories: specifically, different playing techniques. The research objective of this work was to explore the compositional potential of computer music systems capable of recognizing musically relevant timbral categories. Or, formulated as a research question: how can the capability of machine learning algorithms to identify human-level perceptual categories, such as different playing techniques, be exploited from a compositional perspective?

This recognition process served as a basis for the design of idiosyncratic machine listening strategies, besides the obvious one-to-one input-output mappings (i.e., producing a different response for each perceptual category). The interactive music system has the ability to recognize all timbral categories (i.e., playing techniques) played by the performer, yet, based on the interaction scenario, it might choose to respond only to some of them or none at all. The system responds "at-will" and listens "in-search-for" specific sound qualities, guiding the performer's actions through its affordances (Chapter 3).

The idiosyncratic interaction affordances of the interactive music system produce diverse sonic interactions and play a decisive role in shaping the form of the piece. The system's perception of musical form is based on a metric of timbral variability, used to guide long-term decisions. This attribute of the interactive music system points to yet another aspect of listening: listening as an analytical process informing real-time decision-making in musical performance.

In *Imitation Game* these concepts were elaborated further, with the notion of perceptual categories being expanded to include both playing techniques and different instruments. Yet, the focus in this piece does not lie in the design of different listening "modes", but rather decision-making processes that shape the form of the performance in an adaptive and dynamical way.

The integration of musical robotics in this work introduced unique compositional constraints and challenges. Due to its physical, embodied presence, a robotic agent requires a radically different design approach than a software agent, both in terms of sound production and interaction. As far as sound production is concerned, the use of robotically controlled acoustic instruments instead of electronic sound meant that the agent would be sonically responsive but not dependent on the musician, as the use of 'transformative techniques' (Rowe 1993, chap. 1) that depend on an input signal (e.g., digital audio effects and signal processing) was effectively excluded.

The artistic and research goals behind this seemingly arbitrary technical constraint pertain to the way the robotic percussionist's agency is conceptualized in the piece. In *Imitation Game*, the robotic percussionist essentially mirrors the human, by operating in the same domain as them and playing an almost identical instrument setup. This mirroring implies a symmetrical relationship between human and robot, in which the robotic agent can be assumed to have (close to) human-level agency. Affording the same degree of agency to human and machine was both an aesthetic desideratum and a composition-technical challenge in *Imitation Game*.

"Imitation" then refers not only to the robotic percussionist's "embodied" presence, but also to the symmetrical relationship between human and machine agency in the piece. The form of the performance is shaped by and depends equally on human and computational decision-making. Based on this premise, the composition attempts to answer the question: how can machines make *aesthetically driven* decisions?

The robotic percussionist continuously assesses its interaction with the musician and chooses to either "follow" or "lead", by introducing musical changes. This decision is based on the evolution of three different metrics of musical contrast (rhythmic, timbral and dynamic contrast) over time. By introducing musical changes with the purpose to increase or decrease (perceived) musical contrast, the robotic percussionist exercises an agency that is inherently *aesthetic* – if not by intention, then at least in effect (Chapter 4). The robotic agent's ability to introduce musical changes means that human and robot essentially share agency over musical form, which emerges as the result of negotiated intentions and mutual adaptation between them.

The concepts of negotiated intentions and 'collaborative emergence' (Sawyer 2000, 183) were explored in greater depth in the subsequent composition/case study: *Converge/Diverge*. In *Converge/Diverge*, for piano, double bass and interactive music system, form emerges as a result of group decision-making, while musical changes can be initiated only through joint action. The Interactive Music System (IMS) analyzes and responds to the inter-action between the two instrumentalists, rather than their individual actions. In contrast to previous works, the machine learns on the fly, by observing the interaction between the two musicians and comparing the current timbral distance (i.e., dissimilarity) between the two audio inputs to previously observed values.

In this work, a musician can propose a musical change to their co-player, who can either accept or reject this invitation. Musical interaction is conceptualized as a dialogue among three parties (musicians and IMS), in which ideas are proposed, negotiated, accepted or rejected and in some cases mis- and re-interpreted.

In addition to 'collaborative emergence' as a result of joint action during the performance, collaboration in this work takes place in a distributed and asynchronous manner and involves various actors (composer, performers, IMS), time spans and activities (e.g., composition and performance). Each performance is the product of a co-creative process that expands well beyond real-time interpretative decisions to "offline" compositional decisions and even past interpretative choices, differentiation from which might be a decisive factor in subsequent interpretations of the piece (Chapters 5 and 6). Balancing the trade-off between compositional and interpretative decisions was one of the focal points of the compositional process for this piece, leading to the development and exploration of compositional methods such as exploratory and naïve rehearsals, which were meant to foster creative experimentation and composer-performer collaboration (Chapter 5).

The concept of aesthetically driven decisions is a constant throughout the works described in this dissertation, though in each one of them it is approached from a different perspective. In *Bias*, the concept of aesthetic judgments was the focus of a series of machine learning experiments exploring the author's own aesthetic preferences. The title *Bias* has a dual meaning, referring both to the subjective nature of aesthetic judgments and AI bias, a phenomenon that occurs when machine learning algorithms make arbitrary or erroneous assumptions about data, resulting in "biased" predictions. The ambiguity of the title aims to draw parallels between aesthetic judgments and machine learning predictions on the basis of their susceptibility to prejudice and is meant as a comment on the unattainability of objectivity (both in human aesthetic judgments and data-driven models).

Bias explores the relationship between human and machine agency by blurring the boundaries between them and blending human and computational decision-making. The machine learning experiments described in Chapter 7 aimed at creating a new hybrid agency that is based on, yet deviates from the author's own aesthetics. The decisions made by the IMS during its interaction with the musician cannot be entirely attributed to the author's aesthetic preferences, as these are distorted through AI bias. The result is a hybrid human-machine agency that shapes the performance in a decisive and, at times, unpredictable way.

In this work, in addition to more ephemeral interpretative decisions that affect the course of a single performance, the IMS makes some "compositional" decisions, as it autonomously collects its own sound material in its interactions with the performers. The sound material generated by the IMS during the performance is the result of various resynthesis techniques applied to this continuously evolving sound database, collected by the system based on its "own" aesthetic preferences.

This collectively assembled sound corpus is the result and manifestation of the different agencies involved in the work and the co-creative relations among them. The sounds *per se* are provided by the performers – that is, all the performers who have performed the piece to date – while the decision which of these sounds will be recorded and used in the current and future performances of the piece is made by a hybrid, human-derived machine-mediated aesthetic agency. By echoing past instantiations of the work, this sound database facilitates a mediated and asynchronous collaboration among its performers, whose contributions shape its ever-changing sonic identity.

8.2 Interactive Compositions: from Conceptual Shifts to Practical Implications

Far from being a purely theoretical construct, the concept of the interactive musical work has practical implications for both compositional and performance practices. The conceptual shifts resulting from the integration of interactivity and distributed human-computer and human-human co-creativity in electro-instrumental compositions pose a number of practical challenges, including the need to reframe the rehearsal process and develop new experimentation methods and notation strategies.

8.2.1 Compositional Methods

Undoubtedly, traditional compositional methods have limited, if any, applicability in interactive musical works, as the focus of the compositional process in them shifts from composing sound structures to designing interaction potentialities, i.e., fields of possibilities which are open enough to produce varied musical outcomes. In this context, the task of the composer is to determine the conditions that will allow aesthetically consistent, yet varied sonic interactions to emerge.

As interaction affordances are by definition suggestive rather than prescriptive, the full range of interactions they might evoke cannot always be predicted. Additionally, idiosyncratic interaction design or obscure performance instructions might lead to perceived interaction affordances that deviate significantly from those intended by the composer. To bridge the gap between expectations and reality, interaction scenarios have to be tested with the help of the musicians *as part of* – rather than *subsequent to* – the compositional process. This opens the road for a new paradigm of composer-performer collaboration and the development of new compositional methods, designed to address the challenges associated with the integration of interactivity in composed music.

This dissertation presented a variety of methods employed to explore different interaction scenarios and balance the trade-off between work identity and interpretative freedom in interactive compositions. Methods such as 'naïve', 'informed' (Hsu and Sosnick 2009) and exploratory rehearsals were used to compare the intended and perceived affordances of various interactive music systems and gain insight into the different ways in which musicians might interpret compositional concepts and performance instructions. The insight gained through this process played a formative role in the compositional process, by helping refine and revise compositional ideas.

In exploratory rehearsals, improvisation was used to explore the evocative power of abstract concepts, such as "convergence" and "divergence". Comparing the musicians' interpretation of these concepts to the author's intentions and expectations helped find effective ways to bridge the gap between the two. This enabled the design of interaction scenarios that allow for a high degree of interpretative freedom, yet are idiosyncratic and aesthetically consistent.

In naïve rehearsals, musicians were asked to improvise with a virtual co-player without being given any information on its interaction affordances and capabilities. The purpose of these improvisation sessions was to allow the musicians to discover the capabilities of their virtual partner on their own and, in doing so, reveal whether and to what extent interaction affordances alone can successfully communicate compositional intent. Data from these sessions was collected through observation, in combination with questionnaires and semi-structured interviews with the musicians.

In these sessions it became clear that, while interaction affordances may be carriers of compositional intent, they are no replacement for the score – at least not in the context of the works described here. Most of the interaction affordances explored in these rehearsals made little sense without further performance instructions and guidance. Informed rehearsals helped identify the best strategies for communicating compositional intent and find the right balance between prescribing musical actions through notation and evoking them through meticulously designed interaction affordances.

Additionally, naïve and informed rehearsals provided a fertile ground for aesthetic experimentation and, in some cases, led to new creative ideas, playing an instrumental role in the compositional process. In one case, a creative misunderstanding led to the emergence of a new interaction scenario that was later integrated in the composition. When improvising with the robotic percussionist as part of a naïve rehearsal for *Imitation*

Game, percussionist Manuel Alcaraz Clemente mistakenly thought that the robot was repeating some of his actions. While this was in fact not true, this false interpretation led to a unique sonic interaction between the two "players", characterized by varied repetitions of musical actions and dense call-and-response intervals. This behavior was later integrated in the composition as a distinct interaction scenario.

Admittedly, not all of these sessions led to significant insights and therefore not all of them produced results. A result, in this context, is understood as the elaboration or revision of a compositional idea or interaction concept, influenced by insight gained through artistic experimentation. Clearly, these insights are highly subjective in nature and concern the assessment of compositional concepts and their implementation from the composer's perspective.

While some of these methods were borrowed and adapted from research in the evaluation of human-computer improvisation systems (Hsu and Sosnick 2009), their use in the context of this research served the purpose of aesthetic reflection and creative exploration as part of the compositional process. The musicians' contribution to this process did not consist in an evaluation of the computer music systems, but rather an exploration of musical action spaces and the interpretative choices they encompass. The evaluative dimension of these experiments therefore concerns the author's aesthetic criteria and goals, reflecting the inherently subjective nature of aesthetic decisions.

The degree of subjectivity involved in the design of these experiments and the interpretation of data collected from them highlights some of the challenges involved in developing methods for artistic research. One of these challenges is that no single method is universally applicable: methods usually have to be devised specifically for a compositional concept/idea and based on concrete objectives relating to the compositional process (e.g., exploring whether and to what extent interaction affordances can communicate compositional intent). Most importantly, the experiments conducted have to be open-ended enough to allow for the emergence of new creative insights and the discovery of new creative paths. Often, insights might be gained in an area different than originally planned or expected. For that reason, creative experiments should allow for a certain degree of flexibility and adaptability.

The focus on experimental methods in this dissertation is meant to provide an insight into the creative process, as well as its product, with the hope that this might be helpful to other composers working with similar concepts and approaches. While the likelihood that these methods can be transferred unaltered to a different compositional framework is rather low, the rationale behind their design and application could be of relevance to a wide range of compositional approaches gravitating around the concepts of interactivity and human-computer or human-human co-creativity.

8.2.2 Music-analytical Challenges

A concept that is pertinent to interactivity and co-creativity is that of interpretative individuality: the distinctiveness of interpretations of a musical work produced by different individuals. In the context of an interactive composition, interpretative individuality takes on a different meaning, as interpretative decisions extend beyond parameters such as dynamics and phrasing to the sound material and form of the piece. During the performance, the musicians are (inter-)actively shaping compositional ideas through interpretative decisions. Some of the most interesting findings regarding interpretative individuality in this research are discussed in Chapters 5 and 6.

Nevertheless, interpretative individuality and multiplicity in the context of interactive compositions pose a number of music-analytical challenges related to both the object of analysis (e.g., score, code, recordings of performances or some combination of the above) and its methods. In an interactive work, the score and code have an evocative, rather than a descriptive function and individual performances are only 'partial manifestations' (Young 2016, 96) of the sonic possibilities encompassed by it.

Along with more traditional analytical foci, such as form and sound material, an analysis of an interactive work should arguably address the process of interaction itself, i.e., the real-time decision-making and adaptation taking place during the performance (Chapter 6). In this dissertation, the interaction between the musicians in two different performances of *Converge/Diverge* was analyzed using video-based interaction analysis (Jordan and Henderson 1995), a method borrowed and adapted from ethnographic research (Chapter 6). This analysis addressed four of Jordan and Henderson's (1995) analytical foci: the temporal organization of the performance, turn-taking, the use of artifacts and 'trouble and repair' (i.e., cases of miscommunication and the ways in which the musicians dealt with them).

Importantly, the analysis of interactive musical works can be of both music-analytical and compositional interest: the comparative study of different performances of interactive works can provide valuable insights into the degree of interpretative freedom involved in them, as well as potential gaps between intended and perceived interaction affordances, and inform future revisions of the code and/or performance instructions.

8.2.3 Notation and Performance Instructions

The works described in this dissertation entail both composed and improvised musical actions, communicated through notation and text-based instructions. Some of the strategies used to communicate performance instructions in these pieces are summarized in the following subsections.

Composed Musical Actions

Composed musical actions include both thoroughly notated musical fragments and partially notated musical actions. The former are composed musical "phrases" of various lengths that can be played in any order during the piece or a certain interaction scenario. All musical parameters of these fragments are thoroughly composed, including pitch, durations, articulation and dynamics. Partially notated musical actions refer to musical actions one or more parameters of which are left to the performers. These parameters can include the duration, pitch, dynamics and instrumentation of the notated action. The order in which these actions can be performed is open as well.

When viewing the compositions described in this dissertation in a chronological order, a shift away from composed fragments and towards partially notated actions is clearly noticeable in the scores. Partially or ambiguously notated actions seemed to allow for a much higher degree of interpretative freedom, compared to the limited agency of (re-)arranging composed fragments in a specific order, and were preferred as a means to increase the performers' agency and enable diverse readings of the score.

Improvised Musical Actions

The integration of improvised musical actions in the works described in this dissertation took the form of *stimulus-guided*, *goal-guided* and *affordance-guided improvisation*, as well as various combinations among the three.

In *stimulus-guided improvisation* proposed musical actions are used as stimuli for improvisation. The musicians are free to use these actions as a starting point or simply as inspiration for their improvisation. These actions are thought of as suggestions, rather than instructions, and might consist in the use of certain objects, suggestive of specific playing techniques or preparations (e.g., coins used to pluck or "prepare" piano strings), or other idiosyncratic sonic possibilities.

In *goal-guided improvisation* musicians are instructed to work towards a specific goal, such as dissolving a "frozen" spectrum or competing over the computer's attention. These "goals" are evocative of distinctive sound aesthetics, despite leaving the selection of "means" (i.e., the actual sound material) to the performers.

Affordance-guided improvisation is driven by the affordances of the interactive system itself, i.e., what the interactive music system *affords* the performers in terms of sound and interaction potentialities. The idea of affordance-guided improvisation is based on the premise that the affordances of objects – both physical and digital – can, to some extent, determine their use (e.g., chairs are for sitting, even though in the context of conceptual art they can be turned into art objects). Affordance-guided improvisation can be an effective compositional strategy, particularly when the goal is spontaneous and in-the-moment interaction with the computer music system. However, as in the works described here affordances alone often proved to be insufficient in guiding the musicians' actions, affordance-guided improvisation was used mainly in combination with other compositional and notation strategies (e.g., in combination with stimulus-guided improvisation).

This applies to other compositional strategies as well. For instance, in *Converge/Diverge* goal-guided improvisation (i.e., the musicians competing for the computer's attention) was used in combination with a set of proposed musical actions meant as stimuli for the improvisation. Indeed, most types of instructions described above were used in some kind of combination and as complementary components of a broader compositional strategy, rather than mutually exclusive alternatives.

8.2.4 Composer-Performer Relationship and the Rehearsal Process

Interactive musical works challenge the traditional roles of composer and performer and, by extension, the dynamics of the composer-performer relationship. Since such compositional practices are relatively uncommon in contemporary instrumental music, interacting with an intelligent computer music system while following a non-linear score is something that most musicians have little, if any, experience with. The disparities between the composer-performer collaboration paradigm assumed by this research and traditional assumptions regarding the composer-performer relationship were particularly evident in the rehearsals leading to performances of the works described here.

In these rehearsals, after an initial introduction into the compositional concept, the musicians were usually asked to attempt a first run-through of the piece, focusing on understanding and identifying different behaviors of the interactive music system, rather than delivering an aesthetically polished performance. Despite this instruction, musicians generally seemed to find it hard to prioritize exploration and experimentation over the aesthetics of the musical outcome. A reaction that usually arose as a response to the dynamic form of the piece was a tendency to determine the form of the performance beforehand. This might have been partly due to the musicians' prior experience with 'mobile scores' (Hope and Vickery 2011, 225), which offer performers the ability to determine the order of the notated material prior to, or even during the performance. While this approach might work for certain pieces, it does not only defy the purpose of interactive musical works, but is also nearly impossible in them, as the responses of the computer music system cannot be fully predicted, requiring constant adaptation and in-the-moment decisions. After these circumstances were explained to the musicians, they were usually more willing to attempt a first "off-script" run-through of the piece, focusing on an exploratory interaction with the computer music system.

Discussions around notation were another important component of first rehearsals. Depending on their prior experience with open and experimental notation strategies, musicians were more or less familiar with the aesthetic goals of such strategies and their relation to interpretative freedom and multiplicity. As a result, the musicians' reactions to these strategies varied widely, requiring a highly individualized approach to each situation.

While open notation strategies are neither new nor innovative, they remain the exception rather than the rule in contemporary music creation. Historically, the concept of *Werktreue* (being faithful to the work), along with an understanding of the score as the most genuine representation of the compositional idea, has shaped compositional and performance practices, leading to increasingly complex and detailed notation systems – at least within certain strands of contemporary music. In light of this score-centric understanding of the musical work, open notation strategies can potentially be interpreted as a failure to communicate compositional intent or a lack of clear intent altogether.

The centrality of the score in Western art music relates to what Hayden and Windsor (2007) refer to as the 'directive' model of composer-performer collaboration. This is based on a hierarchical composer-performer relationship, in which all aspects of the performance are determined through the score and any collaboration between the composer and performers is limited to issues of technical nature (Hayden and Windsor 2007, 33). The directive paradigm of composer-performer collaboration is centered around compositional intent: this intent is expressed through the score and actualized through the performance.

The centrality of compositional intent and the score as its representation has dominated musical practice for several decades, establishing the directive composer-performer collaboration paradigm as the norm. While alternative approaches have existed for a long time, this norm is – still today – shaping the expectations with which both composers and performers enter the rehearsal process. Most composers would be unhappy with "creative" interpretations of their works that deviate significantly – or, indeed, even slightly – from the score. Likewise, musicians expect clear-cut and detailed instructions regarding the playing techniques involved in the piece, the notation used etc.

By sharing creative responsibility with the performers, the composer effectively redefines not only their own role, but also that of the performer. This has implications for the composer-performer relationship and the rehearsal process itself, as it is in direct conflict with conventional rehearsal practices. The objective of the rehearsal is no longer to fine-tune and perfect a single reading of the score, but rather to produce different readings in each run-through. Such compositional concepts are demanding and, to some extent, disruptive for the rehearsal process and are often met with mistrust and skepticism by the musicians.

This reaction is justified, when considering the discrepancies between the concepts of interactivity and co-creativity, on the one hand, and traditional notions of authorship and interpretation, on the other, and goes to show how intertwined compositional concepts are with the value systems that produced them. Distributed co-creativity is not merely an aesthetic that begins and ends with sound, but a new way to view music-making and the social relations and interactions within which it is embedded.

Undoubtedly, such a premise can pose a number of challenges for performers, requiring a wide range of skills beyond instrumental technique. Focused listening, in-the-moment decision-making and adaptation are only a few of the musical skills involved in the performance of interactive musical works and indicative of the dissolution of yet another binary in them: the one between interpretation and improvisation.

Nevertheless, the rehearsals conducted as part of this research showed that these challenges were overcome rather quickly by the musicians. After a few run-throughs, the musicians seemed eager to explore the capabilities of the IMS, including its degree of autonomy and controllability, as well as its limits. After feeling confident that they could identify different behaviors of the system, their focus usually shifted towards aesthetic exploration. In this phase, their efforts were focused on experimenting with different interpretative approaches and subsequently comparing and assessing their musical outcomes.

This change was reflected in the interpersonal communication that took place during the rehearsal. While, at first, the musicians' questions were centered around compositional intent, partly due to assumptions about authorship derived from a different paradigm of music-making, later on, the dialogue evolved mainly around different interpretative approaches and their musical outcomes. In cases where there were more than one musicians involved (e.g., *Converge/Diverge*), the dialogue seemed to evolve more and more between them and around their subjective evaluation of different interpretative strategies and the development of new ones to be tried out in the next run-throughs.

In general, musicians seemed to gradually take more risks with every run-through. Indeed, individual and idiosyncratic interpretative strategies seemed to emerge as a result of multiple iterations of the piece and the exploration of different possibilities in them. In this phase of the rehearsal process, the author's involvement as a composer was virtually non-existent. This was, at least in part, an intentional decision, meant to facilitate and encourage interpretative individuality and multiplicity. As a rule, what was not determined through the score was left to the musicians, leaving space for creative interpretations, which in certain cases revealed entirely new possibilities.

For instance, in *Converge/Diverge*, for piano, double bass and interactive music system, the computer "learns" to detect timbral convergence and divergence between the two instruments by collecting and analyzing spectral data during the performance. When rehearsing the piece, Florian Müller and Nikolaus Feinig (*Klangforum Wien*) decided to bring this real-time "learning" process to the foreground, by forcing the IMS to make predictions about their interaction before allowing it to collect enough data. This created a unique dramaturgy, woven around what was originally a technical detail, rather than an intended interaction affordance of the IMS.

While the composer-performer collaboration taking place in these compositions does not fall within the *directive* paradigm, clearly, high-level aesthetic decisions are still made by the composer: the performers are asked to explore a pre-defined space of possibilities, the boundaries and aesthetics of which have been determined by the composer. Yet, this space is large enough for them to be able to take a virtually infinite number of highly differentiated and individualized paths though it. This type of co-creativity embraces and benefits from the expertise of all actors involved in it, allowing them to collaborate in a distributed and asynchronous manner. The composer contributes by creating idiosyncratic interaction scenarios that allow diverse, yet aesthetically consistent interpretations, while the performers contribute through improvisatory, in-the-moment decisions, informed by years of musical training, along with culturally informed and subjective aesthetic values.

8.3 Work Identity and Dynamic Form

The question whether performances of an interactive composition should be identifiable as the same work is a complex one with no right answer. For interactive music systems that populate the improvisation end of the composition-improvisation spectrum, work identity might be of limited, if any, relevance, though even the design of such systems involves "compositional" decisions with non-trivial aesthetic implications. For instance, whether a human-computer improvisation system operates based on pitch or spectral information is an inherently aesthetic decision that imposes specific constraints on the way performers can interact with it in a live performance. Nevertheless, in the case of improvisation systems these constraints are usually style or idiom-specific, rather than composition-specific. That is, such systems are conceived as improvisation partners able to produce responses within a broader musical style, rather than a concrete compositional concept entailing idiosyncratic interaction scenarios.

The works described in this dissertation clearly inhabit the space near the composition end of the composition-improvisation spectrum. They are based on specific compositional ideas/concepts and entail highly idiosyncratic sound material and interaction affordances. The concept of work identity is of central importance to this approach, though its definition diverges from its traditional conception. The working hypothesis here is that work identity can lie beyond (recognizable) sequences of sounds, in the behaviors and interactions that produced them.

All but one of the compositions described in this dissertation are based on a dynamic form, leading to widely varied temporal configurations of sound material in each performance. In these compositions, work identity lies in identifiable behaviors, interaction dynamics and sound qualities, rather than form. Highly idiosyncratic interaction affordances are responsible for distinctive sonic interactions and musical outcomes. For example, in *Imitation Game* the performer is instructed to improvise while the robotic percussionist selectively repeats some of their actions. This behavior alone leads to distinctive musical outcomes, which, though improvised, are easily recognizable as instances of the same interaction scenario.

The objective of relocating work identity from the temporal organization of sound material to distinctive interaction affordances and sound qualities has implications for the ontological status of the interactive work. The concept of a dynamic form as a source of multiple and diverse interpretations of the score points towards an understanding of the musical work as a space of sonic possibilities. The course the performers – both human and virtual – take while exploring this space might differ every time, yet the space itself is always recognizable.

8.4 Distributed Creativity and the Interactive Musical Work

Boden (2010) defines creativity as 'the ability to come up with ideas or artefacts that are new, surprising and valuable' (29). According to Boden, novelty can have two different meanings: an idea can be new to the person who had it or it can be new with respect to the history of human knowledge. She refers to the first type of creativity as *P-creativity* (where "P" stands for "person" or "psychological") and the second as *H-creativity*, or historical creativity. Similarly, "surprise" can have three different meanings: it can refer to something unexpected, previously unknown or thought of as impossible. Boden's third criterion, *value*, is much more complex, as it is socially constituted.

The three different types of surprise described above relate to yet another distinction: the one among 'combinatorial', 'exploratory' and 'transformational' creativity (Boden 2010, chap. 5). 'Combinatorial' creativity involves combining otherwise familiar ideas in new and interesting ways. 'Exploratory' creativity involves the exploration of an already established 'conceptual space' (i.e., a 'structured style of thinking'). Finally, 'transformational' creativity involves defining new conceptual spaces, i.e., establishing new paradigms of thinking. Boden considers style imitation systems (i.e., computer programs that generate new compositions in the style of established composers or entire historical periods) as examples of exploratory creativity (37-38). By contrast, transformational creativity involves creating new conceptual spaces and changing the "rules" of established styles of thought. Both the concept of value and that of 'conceptual spaces' (commonly accepted styles of thought) that underlie Boden's definition of creativity highlight the social nature of creativity and its attribution.

Csikszentmihalyi (2014) suggests that creativity cannot be studied outside the social and historical milieu within which creative actions are carried out (47). His systems model of creativity views creativity as the product of the interaction between a *field*, a *domain* and the *individual*, rather than individual actions alone (Csikszentmihalyi 2014, chap. 4). The *domain* represents a corpus of knowledge that is preserved and passed down to the next generations. This domain is continuously expanding through a process of variation and selection. The role of the *individual* in this system is to produce variations of the information contained within the domain, while the role of the *field* (the social institutions and individuals that can affect the structure of the domain) is to select the variations/contributions that are worth preserving and incorporating into the domain.

The domain of music, for instance, consists of various notation systems, musical styles, existing musical works etc. By producing new musical works, a composer produces variations within the domain, which, if deemed worthy by the field (other composers, musicians, critics, curators etc.), will be incorporated into the existing domain (Csikszentmihalyi 2014, 128).

According to Csikszentmihalyi, each of these three systems – individual, domain and field – affects and is affected by the others and is an integral part of the creative process. It follows then that any attribution of creativity is grounded in social agreement and, conversely, social agreement is a constitutive aspect of creativity (Csikszentmihalyi 2014, 49). For Csikszentmihalyi, creativity is constructed through the interaction between social systems and the products of individuals – specifically, the judgments the former make about the latter.

Similar views are echoed by Gell (1998), who views art as 'a system of action' (6) and argues that artworks both embody and function as mediators in social relations. For Gell, the very nature of the artwork is a function of the social-relational milieu within which it is embedded, rather than some independent intrinsic qualities. Brown (2016) and Bown (2015) both consider creative acts as the products of networks of agency that include human and non-human actors. Similarly, Impett (2000) describes the musical work, in particular, as an activity that is 'distributed in space, technology, society and time' (27).

Born (2005) considers music in general as a distributed object that destabilizes the dualisms between subject and object, present and past, individual and collectivity, authentic and artificial and, finally, production and reception. For Born, musical creativity is not only social, but also distributed in time, as each work exists in continuation of a musical past and in anticipation of a musical future – even though its past- and future-oriented agencies are not symmetrical (i.e., anticipation is speculative and therefore uncertain) (23). In addition to the concepts of social and distributed creativity, she proposes the term 'relayed creativity' to describe the circulation, composition, 'decomposition' and 're-composition' of musical materials by different producers/authors, enabled in part by electronic and digital technologies (26). She uses jazz as an example of a musical practice that is based on relayed creativity and negates the distinction between the ideal musical object and its instantiation, as well as the hierarchical assemblages that are characteristic of the concept of the musical work (e.g., composer over interpreter, conductor over instrumentalist and interpreter over listener). In jazz, performances disseminated through recordings educate other musicians, enabling them to create something new, to 're-work' the original in a series of 'successive re-creations' (27).

In work-based practices that involve a division of musical labor between composition and performance, creativity is distributed not only across time and individual works, but also within the same work and across different actors. Live music (i.e., music that is performed) can only exist within the context of a distributed and collaborative creativity, since compositional ideas can only materialize through the mediation of performance. In composed instrumental music, or any other practice characterized by a division of musical labor, an instantiation of the compositional idea is the product of distributed and collaborative creativity between the composer and the performers. Even in electronic performances in which composer and performer are the same person (the composer/performer/programmer model), creativity is dispersed across time and different activities (composing/programming and performing), if not across different individuals.

Born's interpretation of Goehr (1992) suggests that musical meaning does not reside in a single instantiation of the musical work (e.g., the score, a performance or its experience), but is distributed across all these mediations and constructed by the relations among them (9). It should follow then that any musical work that is performed is the product of a distributed creativity involving different actors – both human and non-human (composer, performer, software agent etc.). The musical work does not transcend its instantiations, nor does it exist independently of the agencies that are responsible for its materialization.

While this last statement is true for all musical works, it is epitomized in interactive works. Interactive musical works do not only destabilize the dualism between the work and its instantiations, but also between composition and interpretation. In the absence of a fixed linear score, interpretative decisions determine aspects of the performance that are traditionally defined by the composer. The degree of creative responsibility delegated to the performers can vary depending on the compositional concept. Yet, whether it is the selection of sound material or the form of the piece that is left to the performers – or in some cases both – interpretative decisions in interactive compositions extend far beyond the scope of those associated with the performance of a determinate musical work.

The type of distributed creativity involved in the performance of interactive works is neither hierarchical nor lateral, but rather interactional and integrative of different types of knowledge (e.g., embodied and theoretical) and expertise (e.g., performance and composition). Instantiations of an interactive work in different performances emerge through the interaction and negotiation among compositional intentions, interpretative choices and intended and perceived interaction affordances.

Finally, in the performance of interactive works creativity is distributed across both human and non-human actors. Whether computers are capable of the same type of creativity as humans has been discussed elsewhere (Chapter 2). Still, in the context of interactive compositions software agents make decisions (i.e., they collect, analyze and interpret information from their environment and decide among possible courses of action) that can change the course of the performance and influence the musicians' actions. In this particular setting, human and machine agency are *symmetrical*, as they co-determine the course of the performance.

8.5 Artificial Intelligence and Music Composition: The Road Ahead

In the research described in this dissertation, Artificial Intelligence (AI) is viewed as a 'secondary agent' (Gell 1998, chap. 2): a material, non-intentional entity, which nevertheless has the potential to contribute to the development of new artistic concepts and practices. This approach is suggestive of an understanding of Computational Creativity as part of larger human-computer co-creative networks, rather than a replacement for human creativity. Exploring the potential of AI as an ideation tool had a transformative effect on this research, which is evidenced in the shift in the works described here towards a higher degree of machine autonomy and interpretative freedom, as well as an increased awareness for and reflection on the sociosonic milieu within which these works are embedded, and their ontology as products of co-creative human-human and human-technology relations.

In the compositional approach outlined in the previous chapters, co-creative relations between humans and machines are not grounded in an anthropocentric definition of intelligence. Rather, in an approach that echoes Hayles' views on cognition, AI is considered a non-conscious cognizer with distinctive capacities from human cognition, which synergizes with it as part of larger 'cognitive assemblages' (Hayles 2017). AI is understood not as a reflection or simulation of human intelligence, but as a distinct form of cognition, the value of which lies in its potential to redefine notions of musical authorship, performership and the concept of the musical work. While in this research the emphasis was placed on the sociosonic realm and the relationship between musical authorship and interpretation in interactive works, as AI-based creative tools become increasingly available and accessible to artists, other possibilities will emerge from the specific needs of various compositional approaches. The impact that this technology will have on musical thought and practice remains to be seen.

Admittedly, applications of AI in music pose a series of new, domain-specific challenges. Perhaps the most crucial of these challenges is the discrepancy between the focus of supervised learning algorithms on quantitative evaluation metrics and closed-ended tasks (i.e., tasks that have "right" and "wrong" answers) and the nature of artistic practices as open-ended processes of exploration and discovery. Even unsupervised learning algorithms, which do not involve "right" and "wrong" answers, do not seem to provide a much better alternative, as they perform the same types of tasks as supervised learning algorithms (e.g., clustering algorithms group data points together based on similarity, essentially performing a classification task, albeit without reference to human-labeled data).

Generative machine learning models such as *Generative Adversarial Networks* (*GANs*) (Goodfellow et al. 2020) and *WaveNet* (Engel et al. 2017), which generate musical outputs based on style imitation, pose different and, arguably, more fundamental challenges. Both the user-system interaction afforded by these models and their optimization objectives are suggestive of an understanding of computational creativity as a replacement, rather than an extension of human creativity, making their use within the context of human-computer co-creativity particularly challenging. Such algorithms are essentially "black boxes": the user's interaction with the system is limited to providing a database of sample works, while the "fitness" of the generated outputs is judged by the algorithm based solely on their proximity to the sample works. Developing generative models for human-computer co-creativity would require a fundamentally different, human-in-the-loop design approach, in which the user's feedback would replace similarity with some sample works as an optimization objective. Indeed, as arguably one of the objectives of human-computer co-creative approaches is to break one's creative habits and discover new creative possibilities (Jones, Brown, and D'Inverno 2012), style imitation as an optimization objective is probably incompatible with human-computer co-creativity.

Last but not least, the design of machine learning-based tools for music composition needs to take into account the aesthetic implications of the affordances of these tools. Music-theoretical and other assumptions underlying the design of AI-based tools for music composition are rarely aesthetically neutral. For instance, tools using musical notes as a basic unit (Roberts et al. 2018; Hawthorne et al. 2017; Engel et al. 2017) are suitable for note-based music, but not for sound-based musical idioms and practices. In the development of tools for creative and artistic practices, all design decisions are – and should be viewed as – inherently aesthetic.

Despite the increasing availability and accessibility of AI-based tools for music, the discrepancy between the aesthetic orientation of such tools and contemporary art music points towards the need for a closer collaboration between machine learning developers and composers and the involvement of the latter in the development process. As part of this collaboration, new machine learning algorithms might have to be developed, to fit the needs of artistic practices as open-ended and exploratory processes. In order to ensure that developments in the field of Music AI are aligned to the needs of current and emerging artistic practices, along with questions of technical nature, fundamental philosophical questions regarding the role of AI in the creative process – i.e., *Intelligence Augmentation* vs. *Artificial Intelligence* (Engelbart 1962; Licklider 1960) – and the relation between human and computational creativity will have to be re-examined.

Interdisciplinary collaborations between artists and AI developers could be beneficial not only for the arts, but also for AI. With the mystification and hype around AI fuelling both alarmism and over-inflated expectations and leading to predictions about a new AI winter (Floridi 2020), artistic applications of AI can help shed light on the capabilities, limitations and inner-workings of these algorithms, allowing us to see AI for what it is: neither a 'panacea' nor a 'plague', but a technology that comes with both potential and challenges.

Epilogue

When I started this research, my motivation was to explore and develop human-computer interaction concepts that would re-conceptualize the role of the computer in pieces for acoustic instruments and live electronics. These interaction concepts would establish the computer as a *co-actor*, i.e., a virtual musician that interacts with human musicians in a reciprocal way and co-determines the outcome of the performance. In order to develop computer music systems capable of processing auditory information – i.e., *listening* to their human counterparts – and producing musically meaningful responses, I turned to Machine Learning and explored its applications in machine listening.

The capabilities of Machine Learning algorithms opened up new technical and conceptual possibilities for my work and led me to an entirely new set of research challenges and possibilities. Equipping computer music systems with machine listening and music understanding capabilities enabled me to delegate a higher degree of autonomy and creative responsibility to the computer, as well as the musicians.

As I experimented with interaction concepts and performance instructions that allowed for a higher degree of interpretative freedom and machine autonomy, I found myself fascinated by the levels of engagement demonstrated by the performers in rehearsal and performance settings, the individual subtleties in their interpretations and the variety and richness of musical outcomes that could be produced by the same set of performance instructions and interaction affordances. Undeniably, these performances were engaging for me as well, as in each one of them I discovered new aspects of the work and new sonic possibilities, unveiled by different interpretative choices and approaches.

Indeed, each of the musical works I created as part of this research was a bit more "open" than the previous one, both in terms of interpretative freedom and machine autonomy. By presenting these works in a chronological order in this dissertation, I aimed at creating a narrative that emphasizes this shift in my musical thinking, along with another, secondary shift, which concerns the writing itself: from detailed technical descriptions of the computer music systems to ethnographic and auto-ethnographic accounts of composer-performer collaboration as part of the compositional process.

This notion of a shared creative responsibility that spans across the composition/performance and human/machine divides had a radically transformative effect on my musical thinking. My interests as an artist and researcher shifted from the sonic to the *sociosonic* domain and from sound as object to sound as 'relation' (Born 2019). What was initially an investigation of human-computer interaction became – at least in part – an investigation of human-human interaction and the fundamentally distributed and collaborative notions of creativity that are inherent to any work-based musical practice that involves a division of musical labor.

In the works I created as part of this research, every sound produced during the performance is the result of a dialogue and negotiation among compositional intention, interpretative individuality and technological intentionality (i.e., the directedness and specificities of technological artifacts such as Machine Learning models, Music Information Retrieval tools etc.). In these works, composer, performers and computer are all parts of a bigger co-creative assemblage, while their agencies are closely entangled and intertwined.

Of course, this shift in my musical thinking did not come without its challenges. Most of these challenges related to tensions between the concept of a distributed composer-performer and human-computer co-creativity and the hierarchical relationship between composition and performance, as well as the primacy of the score in Western art music tradition. These tensions gave rise to a series of questions that led me to redefine my role as a composer, as well as my understanding of the musical work: Where does work identity lie when interpretative choices can lead to different musical outcomes in every performance? What is the role of the musical text (i.e., score) in this process? And, what kind of techniques and methods can help navigate the trade-off between musical authorship and interpretative freedom in composed interactive music?

As is often the case with artistic research – or any type of research for that matter (Barad 2007; Latour 2005) – I found that the development of my work and thinking was largely contingent on the tools I used. Artistic means and ends became closely entangled in this process, leading my work to entirely new territories. After exploring a few conventional applications of Machine Learning algorithms, I became increasingly interested in their specificities and limitations. I started exploring critical and subversive approaches to Machine Learning and viewing AI as a *conceptual tool*, rather than the means to solving problems of a purely technical nature. This approach is exemplified in

Bias (Chapter 7), which explores a limitation of Machine Learning algorithms (AI Bias), taking a critical stance towards Machine Learning, while exploring the aesthetic and creative potential of this limitation.

Exploring the potential of AI for music composition was an important part of this research. In my own practice, this potential seemed to lie less in finding novel, AI-based solutions to existing technical problems and more on the new conceptual spaces this technology opens up through its capabilities and limitations. In my compositional work, these new conceptual spaces were opened up by shifting the object of composition from sound itself to the social (human-human and human-technology) relations it materializes. In this process, AI was not just a means to predefined artistic ends, but *a means in creating new ends*; a catalyst for musical thinking; an actant capable of producing 'effects dramatic and subtle' (Bennett 2010, 6).

List of Figures

List of Tables

References

Akten, Memo. 2018. "Ultrachunk." 2018. Accessed March 3, 2021. https://www.memo.tv/works/ultrachunk/.

Applebaum, Mark. 2014. "Aphasia: For Singer and Tape."

Ariza, Christopher. 2009. "The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems." *Computer Music Journal* 33 (2):48–70. https://doi.org/10.1162/comj.2009.33.2.48.

Arnheim, Rudolf. 1974. *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley: University of California Press.

Ashby, William Ross. 1964. *An Introduction to Cybernetics*. London: Methuen (orig. pub. Chapman and Hall, 1956).

Bakht, Salman, and Clarence Barlow. 2009. "PAPAGEI: An Extensible Automatic Accompaniment System for Live Instrumental Improvisation." In *Proceedings of the 2009 International Computer Music Conference*, 521–23.

Banerji, Ritwik. 2016. "Balancing Defiance and Cooperation: The Design and Human Critique of a Virtual Free Improviser." In *Proceedings of the 2016 International Computer Music Conference*, 48–53.

Barad, Karen. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham and London: Duke University Press.

Barrett, Richard. 2014. "Notation as Liberation." *Tempo* 68 (268):61–72. https://doi.org/10.1017/S004029821300168X.

Bennett, Jane. 2010. *Vibrant Matter: A Political Ecology of Things*. Durham and London: Duke University Press.

Bentley, Peter J., and David W. Corne, eds. 2002. *Creative Evolutionary Systems*. London: Academic Press.

Berlyne, Daniel Ellis. 1960. *Conflict, Arousal, and Curiosity*. New York: McGraw-Hill.

———. 1971. *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.

Birkhoff, George David. 1933. *Aesthetic Measure*. Cambridge, MA: Harvard University Press.

Boden, Margaret. 2004. *The Creative Mind: Myths and Mechanisms*. London: Routledge.

———. 2010. *Creativity and Art: Three Roads to Surprise*. Oxford: Oxford University Press.

Bongers, Bert. 2000. "Physical Interfaces in the Electronic Arts. Interaction Theory and Interfacing Techniques for Real-Time Performance." In *Trends in Gestural Control of Music*, edited by Marcelo Wanderley and Marc Mattier, 41–70. IRCAM - Centre Pompidou. http://emerald.tufts.edu/programs/mma/emid/IRCAM/Bon.pdf.

Born, Georgina. 2005. "On Musical Mediation: Ontology, Technology and Creativity." *Twentieth-Century Music* 2 (1):7–36. https://doi.org/10.1017/S147857220500023X.

———. 2019. "On Nonhuman Sound - Sound as Relation." In *Sound Objects*, edited by James A. Steintrager and Rey Chow, 185–208. Durham and London: Duke University Press. https://doi.org/10.1515/9781478002536-012.

Bourriaud, Nicolas. 2002. *Relational Aesthetics*. Translated by Simon Pleasance, Fronza Woods, and Mathieu Copeland. Dijon: Les Presses du réel.

Bown, Oliver. 2012. "Generative and Adaptive Creativity: A Unified Approach to Creativity in Nature, Humans and Machines." In *Computers and Creativity*, 361–81. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-31727-9_14.

———. 2015. "Attributing Creative Agency: Are We Doing It Right?" In *Proceedings of the 6th International Conference on Computational Creativity*, 17–22.

Bown, Oliver, Alice Eldridge, and Jon McCormack. 2009. "Understanding Interaction in Contemporary Digital Music: From Instruments to Behavioural Objects." *Organised Sound* 14 (2):188–96. https://doi.org/10.1017/S1355771809000296.

Brent, William. 2012. "Perceived Control and Mimesis in Digital Musical Instrument Performance." *eContact!* 14 (2). https://econtact.ca/14_2/brent_mimesis.html.

Brown, Andrew R. 2016. “Understanding Musical Practices as Agency Networks.” In *Proceedings of the 7th International Conference on Computational Creativity*, 144–51.

Brown, Andrew R., Toby Gifford, and Bradley Voltz. 2017. “Stimulating Creative Partnerships in Human-Agent Musical Interaction.” *Computers in Entertainment* 14 (2):1–17.

Cage, John. 1952. *4′33″*. Leipzig: Edition Peters.

Chadabe, Joel. 1984. “Interactive Composing: An Overview.” *Computer Music Journal* 8 (1):22–27.

Clarke, Eric, Mark Doffman, and Liza Lim. 2013. “Distributed Creativity and Ecological Dynamics: A Case Study of Liza Lim’s ‘Tongue of the Invisible.’” *Music and Letters* 94 (4):628–63. https://doi.org/10.1093/ml/gct118.

Cohn, Gabe. 2018. “AI Art at Christie’s Sells for $432,500.” New York Times. October 25, 2018. https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html.

Collins, Nick. 2008. “The Analysis of Generative Music Programs.” *Organised Sound* 13 (3):237–48. https://doi.org/10.1017/S1355771808000332.

———. 2011a. “LL: Listening and Learning in an Interactive Improvisation System.” http://w.sussex-ucu.org.uk/Users/nc81/research/ll.pdf.

———. 2011b. “SCMIR: A SuperCollider Music Information Retrieval Library.” In *Proceedings of the 2011 International Computer Music Conference*, 499–502.

———. 2012. “Automatic Composition of Electroacoustic Art Music Utilizing Machine Listening.” *Computer Music Journal* 36 (3):8–23. https://doi.org/10.1162/COMJ_a_00135.

Cope, David. 1992. “Computer Modeling of Musical Intelligence in EMI.” *Computer Music Journal* 16 (2):69. https://doi.org/10.2307/3680717.

———. 1996. *Experiments in Musical Intelligence*. Middleton, WI: A-R Editions.

Couprie, Pierre. 2016. "EAnalysis: Developing a Sound-Based Music Analytical Tool." In *Expanding the Horizon of Electroacoustic Music Analysis*, edited by Simon Emmerson and Leigh Landy, 170–94. Cambridge: Cambridge University Press.

Criado Perez, Caroline. 2019. *Invisible Women: Exposing Data Bias in a World Designed for Men*. London: Penguin Random House.

Csikszentmihalyi, Mihaly. 2014. *The Systems Model of Creativity: The Collected Works of Mihaly Csikszentmihalyi*. Dordrecht: Springer.

Dewey, John. 1934. *Art as Experience*. New York: Penguin Publishing Group.

Dorin, Alan, and Kevin Burt Korb. 2012. "Creativity Refined: Bypassing the Gatekeepers of Appropriateness and Value." In *Computers and Creativity*, edited by Jon McCormack and Mark D'Inverno, 339–60. Berlin, Heidelberg: Springer.

Dreyfus, Hubert L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.

Drummond, Jon. 2009. "Understanding Interactive Systems." *Organised Sound* 14 (2):124–33.

Dudas, Richard. 2010. "'Comprovisation': The Various Facets of Composed Improvisation within Interactive Performance Systems." *Leonardo Music Journal* 20:29–31. https://doi.org/10.1162/LMJ_a_00009.

Eco, Umberto. 1989. *The Open Work*. Translated by Anna Cancogni. Cambridge, MA: Harvard University Press.

Emmerson, Simon. 1986. "The Relation of Language to Materials." In *The Language of Electroacoustic Music*, edited by Simon Emmerson. London: Palgrave Macmillan. https://doi.org/10.1007/978-1-349-18492-7_3.

———. 2013. "Rebalancing the Discussion on Interactivity." *Proceedings of the Electroacoustic Music Studies Network Conference - Electroacoustic Music in the Context of Interactive Approaches and Networks*.

———. 2016. "The Analysis of Live and Interactive Electroacoustic Music: Hans

Tutschku - Zellen-Linien (2007).” In *Expanding the Horizon of Electroacoustic Music Analysis*, edited by Simon Emmerson and Leigh Landy, 333–54. Cambridge: Cambridge University Press.

Engel, Jesse, Cinjon Resnick, Adam Roberts, Sander Dieleman, Douglas Eck, Karen Simonyan, and Mohammad Norouzi. 2017. “Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders.” *arXiv*. https://arxiv.org/abs/1704.01279.

Engelbart, Douglas. 1962. “Augmenting Human Intellect: A Conceptual Framework.” *SRI Summary Report AFOSR-3223*.

Fischman, Rajmil. 2007. “Mimetic Space: A Conceptual Framework for the Discussion, Analysis and Creation of Mimetic Discourse and Structure.” In *Proceedings of the 2007 Electroacoustic Music Studies Network Conference*.

Floridi, Luciano. 2020. “AI and Its New Winter: From Myths to Realities.” *Philosophy and Technology* 33 (1):1–3. https://doi.org/10.1007/s13347-020-00396-6.

Galanter, Philip. 2012. “Computational Aesthetic Evaluation: Past and Future.” In *Computers and Creativity*, edited by Jon McCormack and Mark D’Inverno, 255–93. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-31727-9_10.

———. 2014. “XEPA - Autonomous Intelligent Light and Sound Sculptures that Improvise Group Performances.” *Leonardo* 47 (4):386–93. https://doi.org/doi:10.1162/LEON_a_00844.

Geertz, Clifford. 1973. “Thick Description: Toward an Interpretative Theory of Culture.” In *The Interpretation of Cultures*. New York: Basic Books.

Gell, Alfred. 1998. *Art and Agency: An Anthropological Theory*. Oxford: Oxford University Press.

Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gioti, Artemi Maria. 2017. “Machine Listening in Interactive Music Systems: Current State and Future Directions.” In *Proceedings of the 2017 International Computer Music Conference*, 216–20.

Goehr, Lydia. 1992. *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Oxford: Oxford University Press.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. "Generative Adversarial Networks." *Communications of the ACM* 63 (11):139–44. https://doi.org/10.1145/3422622.

Guba, Egon G., and Yvonna S. Lincoln. 1994. "Competing Paradigms in Qualitative Research." In *Handbook of Qualitative Research*, edited by Norman K. Denzin and Yvonna S. Lincoln, 105–17. Thousand Oaks, CA: Sage.

Gurevich, Michael, and A. Cavan Fyans. 2011. "Digital Musical Interactions: Performer-System Relationships and Their Perception by Spectators." *Organised Sound* 16 (2):166–75. https://doi.org/10.1017/S1355771811000112.

Haraway, Donna J. 2016. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." In *Manifestly Haraway*, edited by Donna J. Haraway. Minneapolis: University of Minnesota Press.

Hawthorne, Curtis, Erich Elsen, Jialin Song, Adam Roberts, Ian Simon, Colin Raffel, Jesse Engel, Sageev Oore, and Douglas Eck. 2017. "Onsets and Frames: Dual-Objective Piano Transcription." *arXiv*. https://arxiv.org/abs/1710.11153.

Hayden, Sam, and Luke Windsor. 2007. "Collaboration and the Composer: Case Studies from the End of the 20th Century." *Tempo* 61 (240):28–39. https://doi.org/10.1017/S0040298207000113.

Hayles, N. Katherine. 2017. *Unthought: The Power of the Cognitive Nonconscious.* Chicago: The University of Chicago Press.

Hoffman, Guy, and Gil Weinberg. 2011. "Interactive Improvisation with a Robotic Marimba Player." In *Musical Robots and Interactive Multimodal Systems*, edited by Jorge Solis and Kia Ng, 233–51. Berlin, Heidelberg: Springer.

Hope, Cat. 2017. "Electronic Scores for Music: The Possibilities of Animated Notation." *Computer Music Journal* 41 (3):21–35. https://doi.org/10.1162/comj_a_00427.

Hope, Cat, and Lindsay Vickery. 2011. "Screen Scores: New Media Music Manuscripts." In *Proceedings of the 2011 International Computer Music Conference*, 224–30.

Hsu, William. 2010. "Strategies for Managing Timbre and Interaction in Automatic Improvisation Systems." *Leonardo Music Journal* 20:33–39. https://doi.org/10.1162/LMJ_a_00010.

Hsu, William, and Marc Sosnick. 2009. "Evaluating Interactive Music Systems: An HCI Approach." In *Proceedings of the 2009 International Conference on New Interfaces for Musical Expression*, 25–28.

Ihde, Don. 1978. *Technics and Praxis*. Vol. 24. Boston Studies in the Philosophy of Science. Dordrecht: Springer. https://doi.org/10.1007/978-94-009-9900-8.

Impett, Jonathan. 2000. "Situating the Invention in Interactive Music." *Organised Sound* 5 (1):27–34. https://doi.org/10.1017/S1355771800001059.

de Jager, Hugo. 1972. "Some Sociological Remarks on Innovation in Music." *International Review of the Aesthetics and Sociology of Music* 3 (2):252–58.

Jones, Daniel, Andrew R. Brown, and Mark D'Inverno. 2012. "The Extended Composer." In *Computers and Creativity*, edited by Jon McCormack and Mark D'Inverno, 175–203. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-31727-9_7.

Jordá, Sergi. 2002. "Afasia: The Ultimate Homeric One-Man-Multimedia-Band." In *Proceedings of the 2002 International Conference on New Interfaces for Musical Expression*, 102–7.

Jordan, Brigitte, and Austin Henderson. 1995. "Interaction Analysis: Foundations and Practice." *Journal of the Learning Sciences* 4 (1):39–103. https://doi.org/10.1207/s15327809jls0401_2.

Kalonaris, Stefano, Toby Gifford, and Andrew Brown. 2019. "Computational Aesthetics and Music: The Ugly, the Small and the Boring." *Proceedings of the 2019 Musical Metacreation Workshop*.

Kalonaris, Stefano, and Anna Jordanous. 2018. "Computational Music Aesthetics: A Survey and Some Thoughts." In *Proceedings of the 3rd Conference on Computer*

Simulation of Musical Creativity.

Kapur, Ajay. 2005. "A History of Robotic Musical Instruments." In *Proceedings of the 2005 International Computer Music Conference*, 21–28.

Kapur, Ajay, Michael Darling, Dimitri Diakopoulos, Jim W. Murphy, Jordan Hochenbaum, Owen Vallis, and Curtis Bahn. 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.

Kohn, Eduardo. 2013. *How Forests Think: Toward an Anthropology Beyond the Human*. Berkeley: University of California Press.

Kozbelt, Aaron. 2006. "Dynamic Evaluation of Matisse's 1935 Large Reclining Nude." *Empirical Studies of the Arts* 24 (2):119–37. https://doi.org/10.2190/a2vy-tebw-vh45-285e.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Latour, Bruno, and Couze Venn. 2002. "Morality and Technology: The End of the Means." *Theory, Culture and Society* 19 (5–6):247–60. https://doi.org/10.1177/026327602761899246.

Leder, Helmut, Benno Belke, Andries Oeberst, and Dorothee Augustin. 2004. "A Model of Aesthetic Appreciation and Aesthetic Judgments." *British Journal of Psychology* 95 (4):489–508. https://doi.org/10.1348/0007126042369811.

Leder, Helmut, and Marcos Nadal. 2014. "Ten Years of a Model of Aesthetic Appreciation and Aesthetic Judgments: The Aesthetic Episode - Developments and Challenges in Empirical Aesthetics." *British Journal of Psychology* 105 (4):443–64. https://doi.org/10.1111/bjop.12084.

Leffue, Steven, and Grady Kestler. 2016. "AIIS: An Intelligent Improvisational System." *Proceedings of the 2016 International Computer Music Conference*, 430–33.

Lepri, Giacomo. 2016. "InMuSIC: An Interactive Multimodal System for Electroacoustic Improvisation." In *Proceedings of the 2016 International Computer Music*

Conference, 1–5.

Lévy, Benjamin, Georges Bloch, and Gérard Assayag. 2012. "OMaxist Dialectics: Capturing, Visualizing and Expanding Improvisations." In *Proceedings of the 2012 International Conference on New Interfaces for Musical Expression*, 137–40.

Lewis, George E. 2000. "Too Many Notes: Computers, Complexity and Culture in Voyager." *Leonardo Music Journal* 10:33–39. https://doi.org/10.1162/096112100570585.

Licklider, Joseph Carl Robnett. 1960. "Man-Computer Symbiosis." *IRE Transactions on Human Factors in Electronics* HFE-1 (1):4–11. https://doi.org/10.1109/thfe2.1962.4503337.

Linson, Adam, Chris Dobbyn, and Robin Laney. 2016. "Critical Issues in Evaluating Freely Improvising Interactive Music Systems." In *Proceedings of the 2016 International Conference on Computational Creativity*, 145–49.

Linson, Adam, Chris Dobbyn, George E. Lewis, and Robin Laney. 2015. "A Subsumption Agent for Collaborative Free Improvisation." *Computer Music Journal* 39 (4):96–115. https://doi.org/10.1162/COMJ_a_00323.

Magnusson, Thor. 2009. "Of Epistemic Tools: Musical Instruments as Cognitive Extensions." *Organised Sound* 14 (2):168–76. https://doi.org/10.1017/S1355771809000272.

———. 2019. *Sonic Writing: Technologies of Material, Symbolic and Signal Inscriptions*. New York: Bloomsbury.

Makin, Alexis D.J. 2017. "The Gap Between Aesthetic Science and Aesthetic Experience." *Journal of Consciousness Studies* 24 (1–2):184–213.

Manaris, Bill, Juan Romero, and Penousal Machado. 2005. "Zipf's Law, Music Classification, and Aesthetics." *Computer Music Journal* 29 (1):55–69. https://doi.org/10.1162/comj.2005.29.1.55.

Manaris, Bill, Patrick Roos, Penousal Machado, Dwight Krehbiel, Luca Pellicoro, and Juan Romero. 2007. "A Corpus-Based Hybrid Approach to Music Analysis and

Composition." In *Proceedings of the National Conference on Artificial Intelligence*, 839–45.

Manzelli, Rachel, Vijay Thakkar, Ali Siahkamari, and Brian Kulis. 2018. "Conditioning Deep Generative Raw Audio Models for Structured Automatic Music." In *Proceedings of the 2018 Musical Metacreation Workshop*.

Martindale, Colin. 1988. "Relationship of Preference Judgments to Typicality, Novelty and Mere Exposure." *Empirical Studies of the Arts* 23 (2):79–96.

McAdams, Stephen, and Emmanuel Bigand. 1993. "Introduction to Auditory Cognition." In *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by Stephen McAdams and Emmanuel Bigand. Oxford Science Publications. New York: Clarendon Press/Oxford University Press.

McCormack, Jon. 2012. "Creative Ecosystems." In *Computers and Creativity*, edited by Jon McCormack and Mark D'Inverno, 39–60. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-31727-9_2.

McCormack, Jon, and Andy Lomas. 2021. "Deep Learning of Individual Aesthetics." *Neural Computing and Applications* 33 (1):3–17. https://doi.org/10.1007/s00521-020-05376-7.

Meyer, Leonard B. 1983. "Innovation, Choice, and the History of Music." *Critical Inquiry* 9 (3):517–44.

Nesterenko, Oksana. 2017. "'Open Work' One Step Further: From John Cage to John Zorn." *Perspectives of New Music* 15 (2):199–217.

Nika, Jérôme, Marc Chemillier, and Gérard Assayag. 2017. "ImproteK: Introducing Scenarios into Human-Computer Music Improvisation." *ACM Computers in Entertainment* 14 (2). https://doi.org/10.1145/3022635.

Norman, Don. 2013. *The Design of Everyday Things: Revised and Expanded Edition*. New York: Basic Books.

Van Nort, Doug. 2009. "A System for Musical Improvisation Combining Sonic Gesture Recognition and Genetic Algorithms." In *Proceedings of the 2009 Sound and Music*

Computing Conference, 131–36.

———. 2018. “Genetically Sonified Organisms: Environmental Listening/Sounding Agents.” In *Proceedings of the 2018 Musical Metacreation Workshop*.

O’Hear, Anthony. 1995. “Art and Technology: An Old Tension.” In *Philosophy and Technology*, 143–58. Cambridge University Press. https://doi.org/10.1017/CBO9780511524080.010.

Oliveros, Pauline. 1970. “And Don’t Call Them ‘Lady’ Composers.” *New York Times*, September 13, 1970.

———. 1974. *Sonic Meditations*. Baltimore: Smith Publications.

van den Oord, Aaron, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. “WaveNet: A Generative Model for Raw Audio.” https://doi.org/10.1109/ICASSP.2009.4960364.

Pachet, Francois. 2002. “The Continuator: Musical Interaction with Style.” *Journal of New Music Research* 31 (1):1–9.

Paine, Garth. 2002. “Interactivity: Where To From Here?” *Organised Sound* 7 (3):295–304. https://doi.org/10.1017/S1355771802003096.

Park, Tae Hong. 2016. “Exploiting Computational Paradigms for Electroacoustic Music Analysis.” In *Expanding the Horizon of Electroacoustic Music Analysis*, edited by Simon Emmerson and Leigh Landy, 123–47. Cambridge: Cambridge University Press.

Ravikumar, Prashanth Thattai, Kevin McGee, and Lonce Wyse. 2018. “Back to the Experiences: Empirically Grounding the Development of Musical Co-Creative Partners in Co-Experiences.” In *Proceedings of the 2018 Musical Metacreation Workshop*.

Rennie, Tullis. 2014. “Socio-Sonic: An Ethnographic Methodology for Electroacoustic Composition.” *Organised Sound* 19 (2):117–24. https://doi.org/10.1017/S1355771814000053.

Roberts, Adam, Jesse Engel, Colin Raffel, Curtis Hawthorne, and Douglas Eck. 2018. "A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music." *arXiv*. https://arxiv.org/abs/1803.05428.

Rodgers, Tara. 2010. *Pink Noises: Women on Electronic Music and Sound*. Durham and London: Duke University Press. https://doi.org/10.1007/s13398-014-0173-7.2.

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

———. 1999. "The Aesthetics of Interactive Music Systems." *Contemporary Music Review* 18 (3):83–87. https://doi.org/10.1080/07494469900640361.

Sawyer, R. Keith. 2000. "Improvisational Cultures: Collaborative Emergence and Creativity in Improvisation." *Mind, Culture and Activity* 7 (3):180–85. https://doi.org/10.1207/S15327884MCA0703.

Schaeffer, Pierre. 2017. *Treatise on Musical Objects. An Essay across Disciplines*. Translated by Christine North and John Dack. Berkeley: University of California Press. https://doi.org/10.1525/9780520967465.

Scipio, Agostino Di. 2003. "Sound Is the Interface." *Organised Sound* 8 (3):269–77. https://doi.org/10.1017/S1355771803000244.

Scurto, Hugo, Frédéric Bevilacqua, and Baptiste Caramiaux. 2018. "Perceiving Agent Collaborative Sonic Exploration in Interactive Reinforcement Learning." In *Procedings of the 2018 Sound and Music Computing Conference*, 72–79.

Searle, John R. 2004. *Mind: A Brief Introduction*. Oxford: Oxford University Press.

Shusterman, Richard. 1997. "The End of Aesthetic Experience." *The Journal of Aesthetics and Art Criticism* 55 (1):29–41. https://doi.org/10.2307/431602.

Singer, Eric, Kevin Larke, and David Bianciardi. 2003. "LEMUR GuitarBot: MIDI Robotic String Instrument." In *Proceedings of the 2003 International Conference on New Interfaces for Musical Expression*, 188–91.

Smalley, Denis. 1997. "Spectromorphology: Explaining Sound-shapes." *Organised Sound*

2 (2):107–26. https://doi.org/10.1017/s1355771897009059.

———. 2007. "Space-form and the Acousmatic Image." *Organised Sound* 12 (1):35–38. https://doi.org/10.1017/S1355771807001665.

Smith, Benjamin D., and W. Scott Deal. 2014. "ML.∗ Machine Learning Library as a Musical Partner in the Computer-Acoustic Composition Flight." In *Proceedings of the 40th International Computer Music Conference and 11th Sound and Music Computing Conference*, 1285–89.

Spiegel, Laurie. 1992. "An Alternative to a Standard Taxonomy for Electronic and Computer Instruments." *Computer Music Journal* 16 (3):5–6.

Stowell, Dan, and Mark Plumbley. 2007. "Adaptive Whitening for Improved Real-Time Audio Onset Detection." In *Proceedings of the 2007 International Computer Music Conference*.

Thom, Belinda. 2000. "BoB: An Interactive Improvisational Music Companion." In *Proceedings of the 4th International Conference on Autonomous Agents*, 309–16. http://dl.acm.org/citation.cfm?id=336595.337510.

Thoresen, Lasse. 2001. "Spectromorphological Analysis of Sound Objects: An Adaptation of Pierre Schaeffer's Typomorphology." In *Proceedings of the 2001 Electroacoustic Music Studies Network Conference*.

Truax, Barry. 2001. *Acoustic Communication*. 2nd ed. Westport, Connecticut: Ablex Publishing.

Tsiros, Augoustinos. 2017. "Sonic Xplorer: A Machine Learning Approach for Parametric Exploration of Sound." In *Proceedings of EVA London*, 144–49. https://doi.org/10.14236/ewic/eva2017.34.

Uitti, Frances-Marie. 1995. "Preserving the Scelsi Improvisations." *Tempo*, 194:12–14. https://doi.org/10.1017/s0040298200004484.

Varèse, Edgard, and Chou Wen-Chung. 1966. "The Liberation of Sound." *Perspectives of New Music* 5 (1):11–19.

Verbeek, Peter Paul. 2008. "Cyborg Intentionality: Rethinking the Phenomenology of Human-Technology Relations." *Phenomenology and the Cognitive Sciences* 7 (3):387–95. https://doi.org/10.1007/s11097-008-9099-x.

Weinberg, Gil. 2007. "Robotic Musicianship – Musical Interactions Between Humans and Machines." In *Human Robot Interaction*, edited by Nilanjan Sarkar, 424–44. InTech. http://www.intechopen.com/books/human_robot_interaction/robotic_musicianship_-%0A_musical_interactions_between_humans_and_machines.

Weinberg, Gil, and Scott Driscoll. 2006. "Toward Robotic Musicianship." *Computer Music Journal* 30 (4):28–45. https://doi.org/10.1162/comj.2006.30.4.28.

Weinberg, Gil, Scott Driscoll, and Mitchell Parry. 2005. "Musical Interactions with a Perceptual Robotic Percussionist." In *Proceedings of the 14th IEEE International Conference on Robot and Human Interactive Communication*, 456–61.

Wooldridge, Michael, and Nicholas Jennings. 1996. "Intelligent Agents: Theory and Practice." *The Knowledge Engineering Review* 10 (2):115–52. https://doi.org/doi:10.1017/S0269888900008122.

Xambó, Anna, Robin Laney, Chris Dobbyn, and Sergi Jordà. 2013. "Video Analysis for Evaluating Music Interaction: Musical Tabletops." In *Music and Human-Computer Interaction*, edited by Simon Holland, Katie Wilkie, Paul Mulholland, and Allan Seago, 241–58. London: Springer. https://doi.org/10.1007/978-1-4471-2990-5.

Xu, Dongming, Yonggui Wang, and Sukanto Bhattacharya. 2010. "Thinking Beyond Means-Ends Analysis: The Role of Impulse-Driven Human Creativity in the Design of Artificially Intelligent Systems." In *Information Systems Foundations: The Role of Design Science*, edited by Denis Hart and Shirley Gregor, 212–32. Canberra: ANU E Press.

Yee-King, Matthew John. 2011. "An Autonomous Timbre Matching Improviser." In *Proceedings of the 2011 International Computer Music Conference*, 122–25.

Young, Michael. 2008. "NN Music: Improvising with a 'Living' Computer." In *Computer Music Modeling and Retrieval. Sense of Sounds*, edited by Richard Kronland-

Martinet, Sølvi Ystad, and Kristoffer Jensen, 337–50. Berlin, Heidelberg: Springer.

———. 2016. “Interactive and Generative Music: A Quagmire for the Musical Analyst.” In *Expanding the Horizon of Electroacoustic Music Analysis*, edited by Simon Emmerson and Leigh Landy, 80–101. Cambridge: Cambridge University Press.

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press.

Appendix 1: Performances

Video documentations of the musical works discussed in this dissertation can be found in the following links.

Neurons

Performance by Joel Diegert:

https://www.artemigioti.com/demos/Neurons.html.

Imitation Game

Short documentary and performance by Manuel Alcaraz Clemente:

https://www.artemigioti.com/demos/Imitation_game.html.

Converge/Diverge

Performances by *Schallfeld Ensemble* (Margarethe Maierhofer-Lischka and Patrick Skrilecz) and *Klangforum Wien* (Nikolaus Feinig and Florian Müller):

https://www.artemigioti.com/demos/Converge_Diverge.html.

Bias

Short documentary and performance by Szilárd Benes:

https://www.artemigioti.com/demos/Bias.html.

Appendix 2: Scores

N E U R O N S

for soprano saxophone and interactive music system

[2017]

Artemi-Maria Gioti

Score

P E R F O R M A N C E   N O T E S

NOTATION

- 1/8 tone higher
- 1/8 tone lower
- 5/8 tone higher
- 3/8 tone lower
- 3/4 tone higher
- 1/4 tone higher
- 5/8 tone lower
- shadow sound (in multiphonics only)
- 5" — proportional time notation; perform the notated fragment within the indicated duration
- duration line
- transition line
- held note
- slap tone
- air tone (no pitch)
- half air, half pitch
- pitch (no air)
- 8va — harmonics glissando
- accelerating tremolo
- Solid line rectangle: Fragments inside this box can be played in any order. All fragments inside the box must be played once before moving on to the next section. Repetitions are not possible.
- Dashed line rectangle: Fragments inside this box can be played in any order, one or more times.
- Dotted line rectangle: The performer can choose any of the fragments inside this box. Repetitions are possible.

INSTRUCTIONS

Neurons is an interactive composition for soprano saxophone and interactive music system, utilizing machine learning. The title of the piece refers to a feedforward Neural Network trained and used for real-time recognition of four different sound classes (single note, multiphonic, air tone and slap tone). The computer's response to the saxophonist is the result of both this instant recognition, as well as short and long-term memory processes. System characteristics and behavior change during the piece, calling for mutual adaptation between the musician and the computer as their co-actor.

Listening mode 1: In this mode, which corresponds to sections A, B and C (pages 1, 2 and 3 of the score respectively), each of the four sound classes (single note, multiphonic, air tone and slap tone) triggers a different response. The amplitude of low frequency components of the electronic sound is used to signal changes in texture variability, as "perceived" by the computer. As texture variability increases, so does the amplitude of these low frequency components. When texture variability reaches a threshold value, the system enters a non-listening state. This should happen no sooner than the last staff of page 3. Silence and air tones can be used by the performer as control actions in order to prevent the system from entering the non-listening state too soon. Section C can also be interrupted by control actions, when necessary. Listening mode 1 is activated again in section E (lower half of page 4).

Non-listening state: While in the non-listening state, the system is unresponsive. The performer should stop playing and wait for listening mode 2 to be activated.

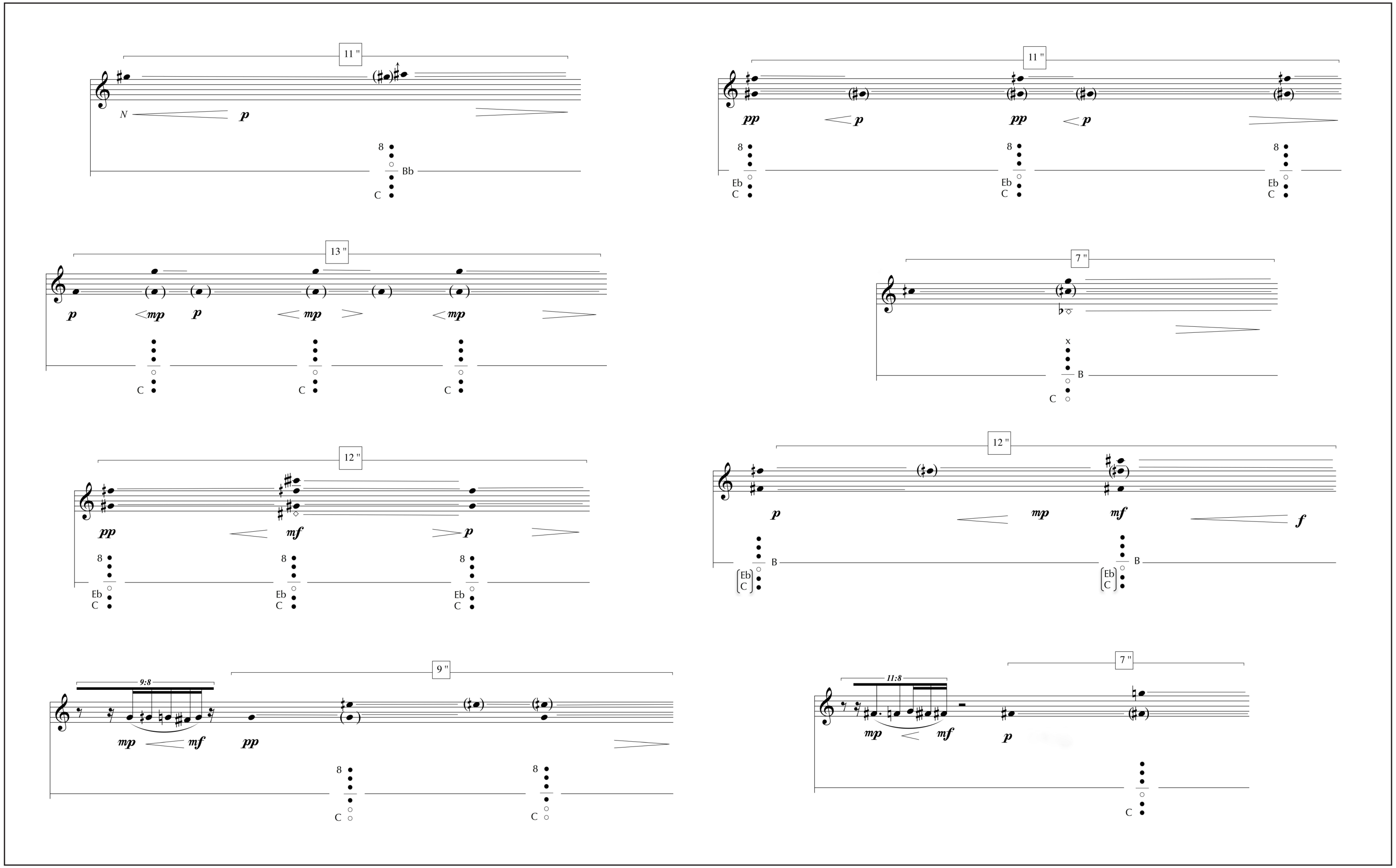
Listening mode 2: In listening mode 2, the system switches its input on and off at will, in search of multiphonics. A texture of high-pitched percussive sounds is used as a cue, when the input is switched on. When this happens, the performer should play a multiphonic from section D (upper half of page 4). If the execution of the multiphonic is evaluated as "stable" by the computer, a "spectral freeze" effect is activated. Otherwise, the musician should choose a different multiphonic and try again. This process is repeated 3 times.

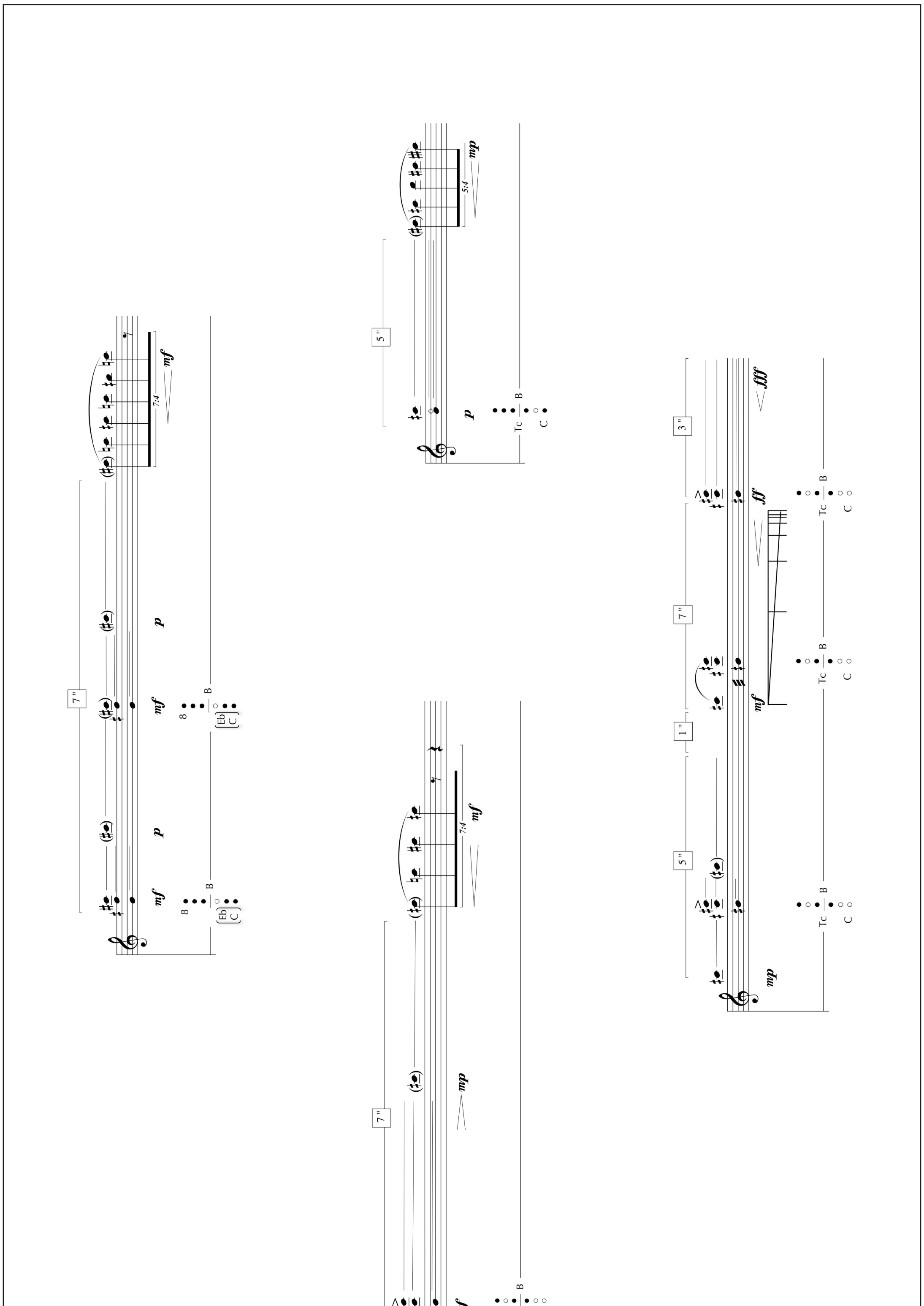
Neurons

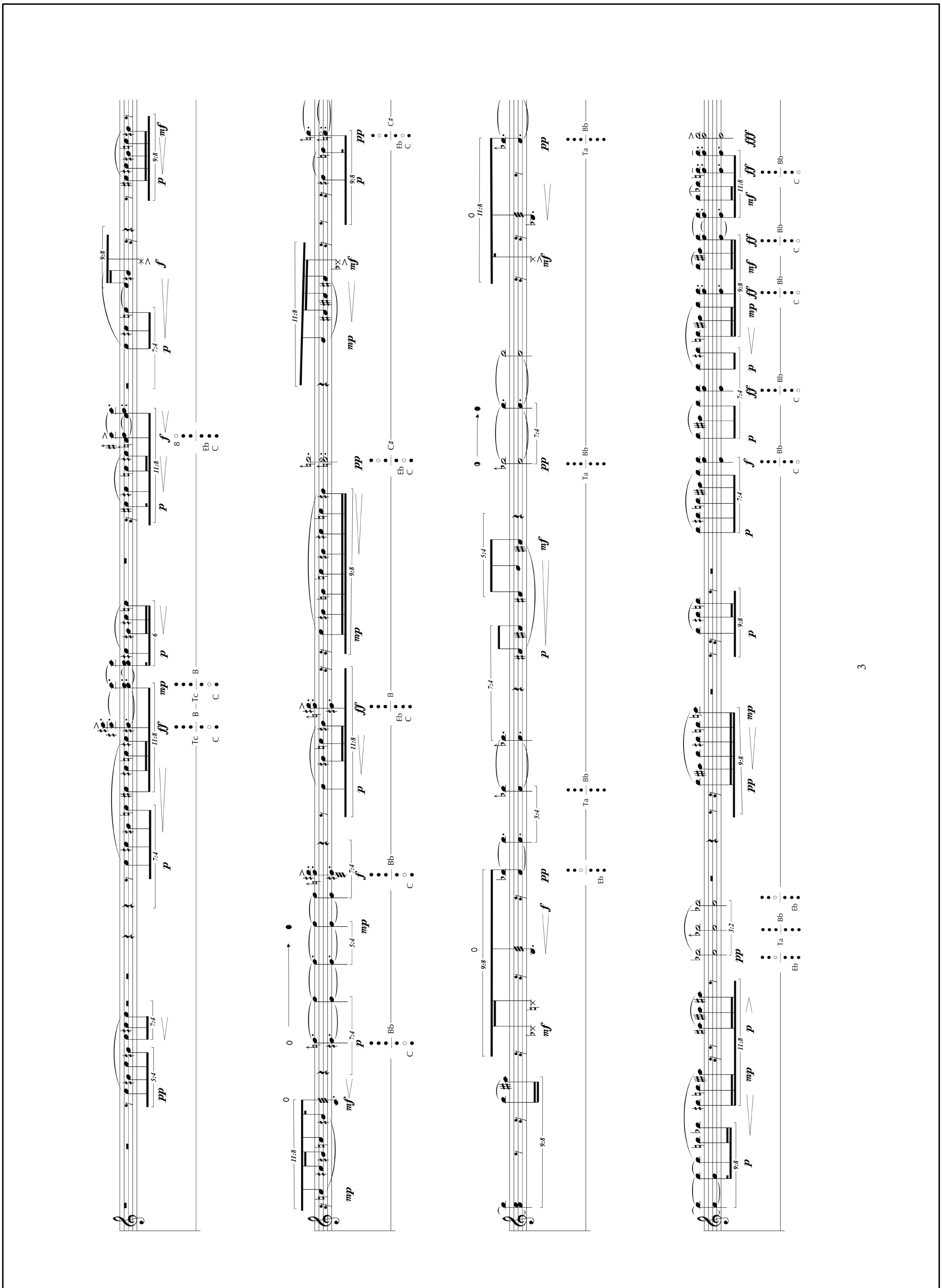
for soprano saxophone and interactive electronics

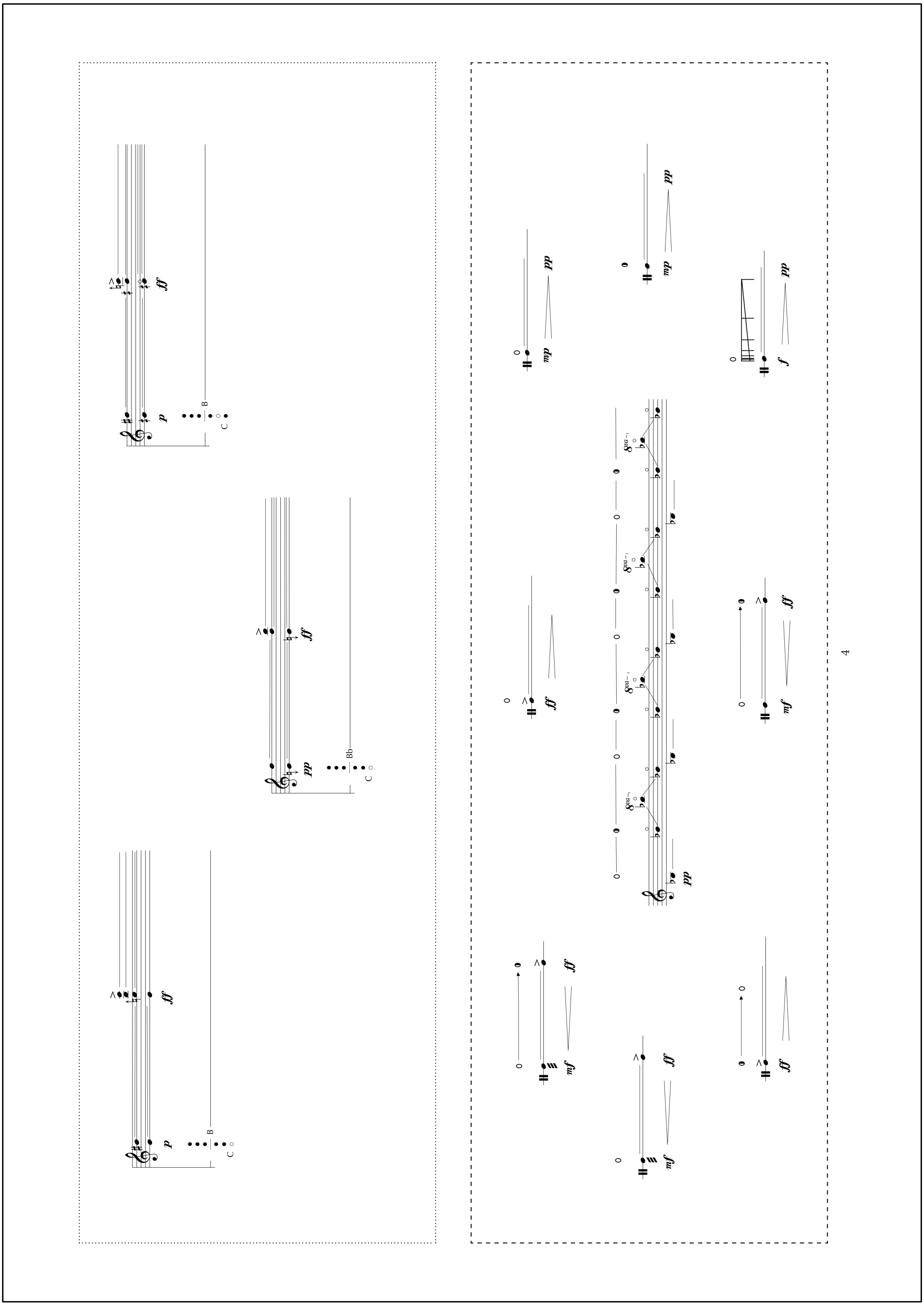
Artemi - Maria Gioti
(2017)

♩ = 60/𝅗𝅥 = 30




7"
mf
p
Eb
C
B
8
5"
Tc
f
mp
7:4
5:4
1"
3"
fff
ff
2





I M I T A T I O N   G A M E

for human and robotic percussionist

[2018]

Artemi-Maria Gioti

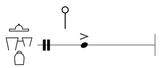
Score

P E R F O R M A N C E   N O T E S

MALLETS & STICKS

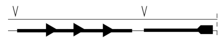
	Bow
	Soft yarn mallet
	Wire brush
	Metal stick
	Brush

NOTATION

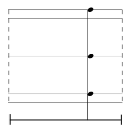


Perform notated action on any of the indicated instruments

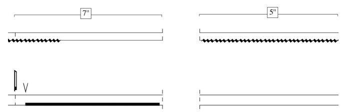
	Magnet with chain
	Magnet (without chain)
	(Any) cymbal
	(Any) bongo drum
	(Any) cowbell
→	Continue indicated action
⇆	Zig-zag motion
↻	Circular motion
	Slightly touch the edge of the cymbal using one finger
⊙	Touch the bell of the cymbal with your palm



Duration/bow pressure lines



Sound event expected within the indicated duration



Wait for notated action

INSTRUCTIONS

The opening and closing sequences of the piece can be found on pages 1 and 4 of the score respectively. The second and third page (not numbered) consist of musical fragments which can be executed in any order. The performer can alternate between the two pages.

The responses generated by the robotic percussionist are based on three different interaction scenarios:

(1) imitate (play similar sound material),
(2) initiate (introduce different sound material), and
(3) repeat (improvise, occasionally repeating some of the human percussionist's actions).

Imitate – Initiate:

In these scenarios the robotic percussionist interacts with the human in a "call-and-response" fashion, playing either phrases that "imitate" those played by the musician (e.g., by using the same instruments and playing techniques), or phrases that introduce new sound material (e.g., a new timbre or rhythm).

The robotic percussionist selects interaction scenarios based on metrics of rhythmic, timbral and dynamic contrast. These metrics are calculated on a phrase basis and their values are stored, allowing the robotic percussionist to make decisions based on their evolution over time. For example, if the estimated rhythmic contrast has been constant (i.e., around the same value), or monotonic (i.e., constantly increasing or decreasing) for the last few phrases, the robotic percussionist is less likely to play similar material to that played by the human percussionist ("imitate") and more likely to introduce new, contrasting sound material ("initiate").

The page of the score that corresponds to these two scenarios is marked with "imitate - initiate" and consists of 19 fragments organized in three concentric rectangles in order of rhythmic, dynamic and timbral contrast as follows:

- From the center outwards: in order of decreasing rhythmic contrast,
- From the center upwards: in order of decreasing timbral contrast, with strokes being the predominant playing technique,
- From the center downwards: in order of decreasing timbral contrast, with scraping being the predominant playing technique.

Repeat:

The "repeat" scenario is the only interaction scenario that entails synchronous action (i.e., both the musician and the robotic percussionist playing at the same time) and has two variations depending on who is "leading" the improvisation:

- Robot "repeats" human:

This variation of the "repeat" scenario is initiated by the robotic percussionist, when it breaks the call-and-response cycle and starts repeating some of the musician's actions. When this happens, the musician should start improvising on the material found on the page marked with "repeat". The notated actions on this page can be performed any number of times, in different variations (e.g., on various instruments) and with variable durations. The scenario is terminated by the robotic percussionist with a simultaneous stroke on all three instruments.

- Human "repeats" robot:

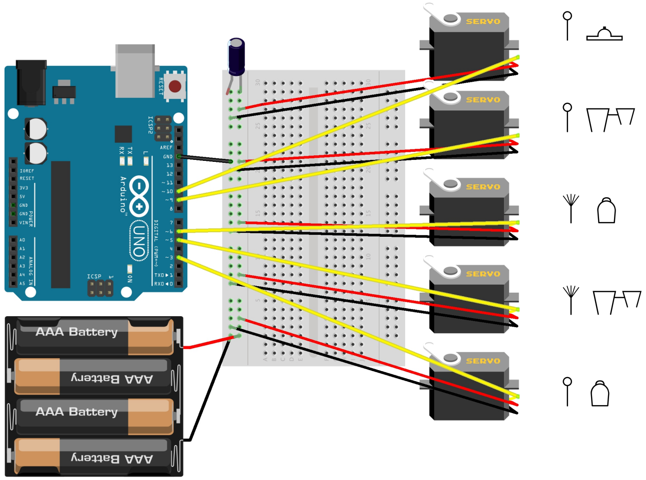
This variation of the "repeat" scenario is initiated by the musician. Similarly to the first variation, the musician improvises incorporating "repetitions" of the robotic percussionist's actions in their improvisation. The scenario is terminated a few seconds after the musician stops playing.

The closing sequence of the piece is initiated by the robotic percussionist.

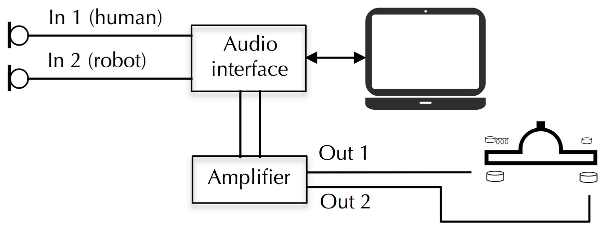
Approximate duration: 16 min.

H A R D W A R E

CIRCUIT



AUDIO

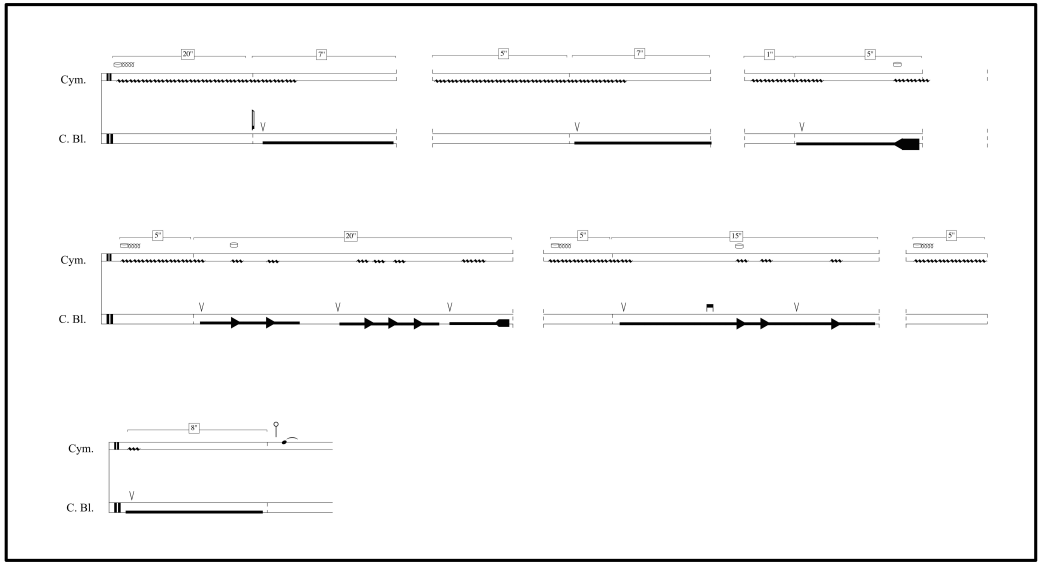


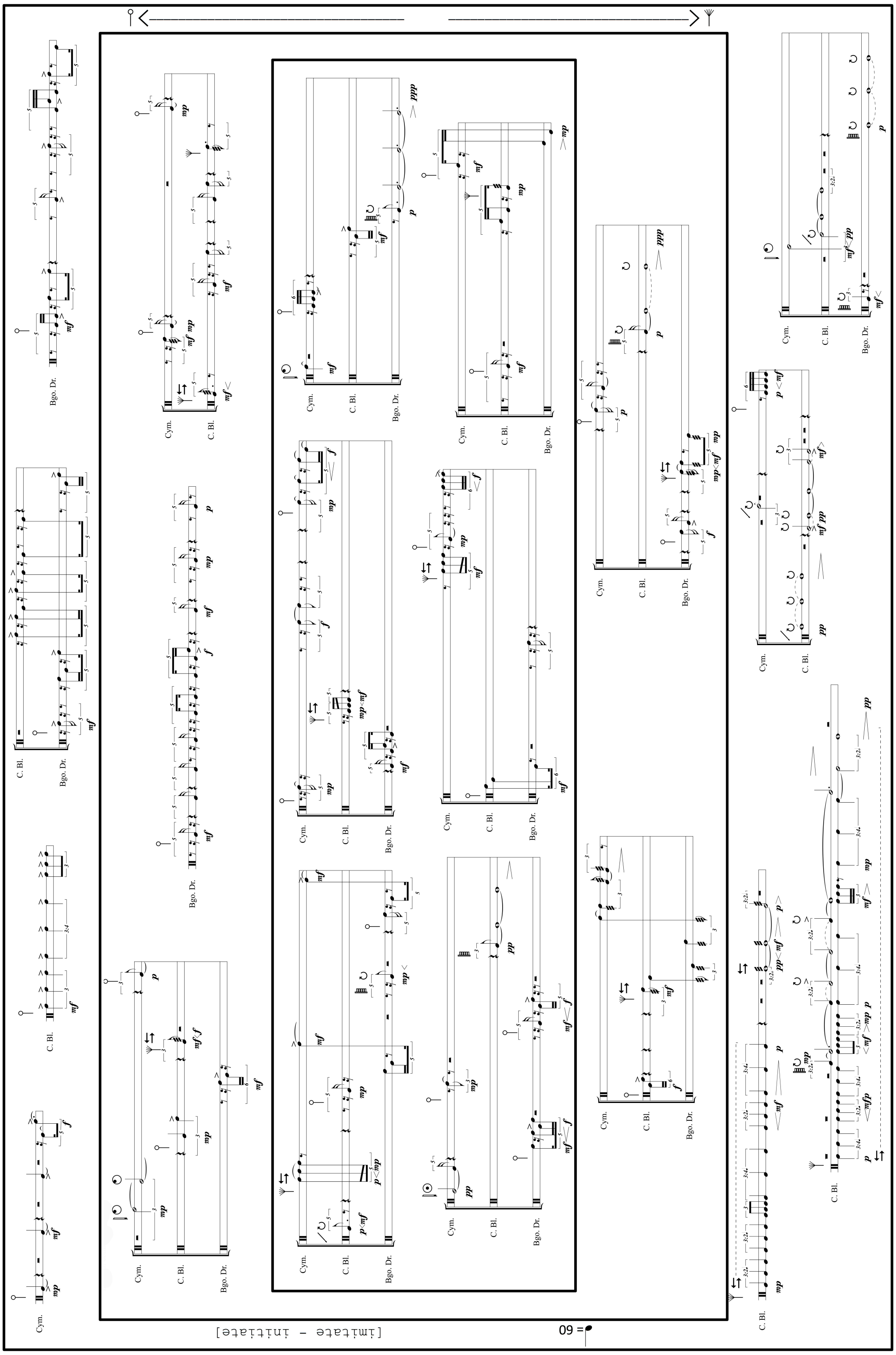
STAGE DISPOSITION

IMITATION GAME

for human and robotic percussionist

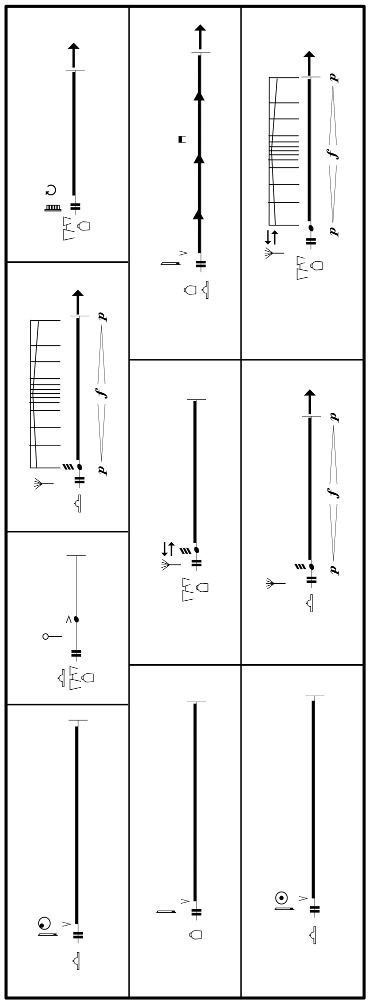
Artemi-Maria Gioti
[2018]

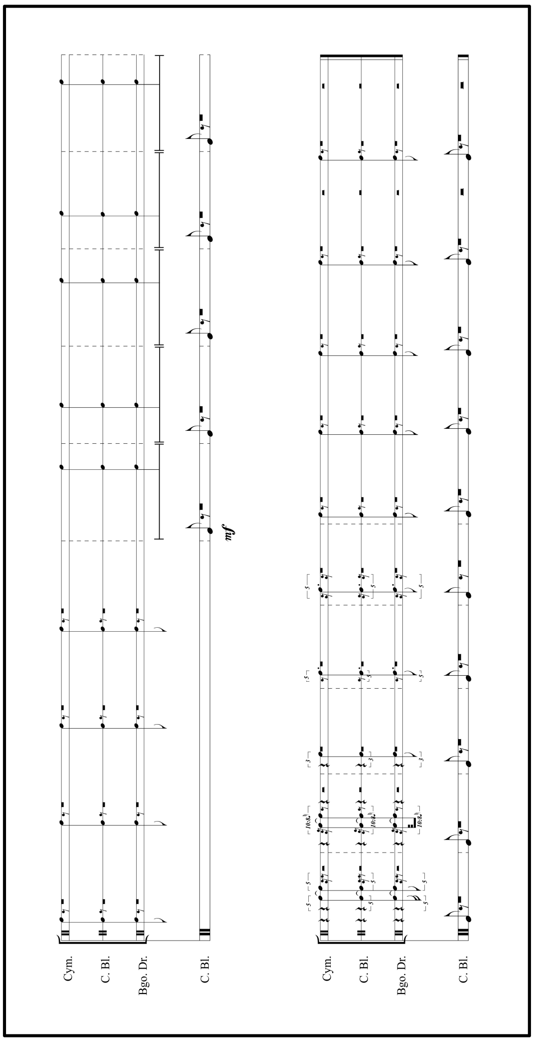




[imitate - initiate]
= 60
Cym.
C. Bl.
Bgo. Dr.

[repeat]





4

C O N V E R G E / D I V E R G E

Artemi-Maria Gioti

for piano, double bass and interactive music system

[2019]

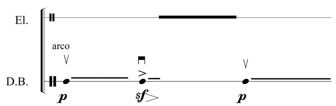
Score

P E R F O R M A N C E   N O T E S

GENERAL NOTATION

- Pitched sound
- Scratch tone (overpressure)
- Percussive sound
- Overtone
- Quarter tone flat
- Glissando
- 1-line staff (double bass): indeterminate pitch.
- 3-line staff (piano): indeterminate pitch. The 3 lines correspond to different registers, determined by the metal frame of the instrument.
- Continue in pattern, e.g., by repeating dynamic or bow placement changes.

INTERACTION TIMING



Wait until the computer’s response is over, before moving on to the next action.

Call-and-response: musicians take turns.

Synchronous interaction: musicians play simultaneously.

PIANO

MALLETS & STICKS

- Wooden mallet
- Super ball (rubber mallet)
- Chopstick
- Metal rod
- Glass
- Plastic triangular ruler

STRING PREPARATIONS

- Place a small coin in-between strings of the same pitch, as shown.
- Same as above. Slide the coin along the string to hit different nodes and produce different overtones.
- Place a finger cymbal on top of the indicated string(s).
- Tie a piece of electromagnetic (cassette) tape in a knot around the indicated string.
- Place the finger cymbal on top of the tape-prepared string.

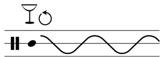
PLAYING TECHNIQUES

Mute the indicated string with your hand.

Place a bamboo chopstick between 2 strings of the same pitch.

Play a glissando in the indicated register.

Place the glass on top of the indicated strings and rotate it counterclockwise around its axis.



Same as above. While rotating the glass, make bigger circular movements, producing slow glissandi.

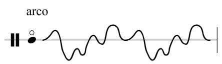
DOUBLE BASS

PLAYING TECHNIQUES

s.t. Sul tasto

s.p. Sul ponticello

c.l.b. Col legno battuto



Overtone glissando

INSTRUCTIONS

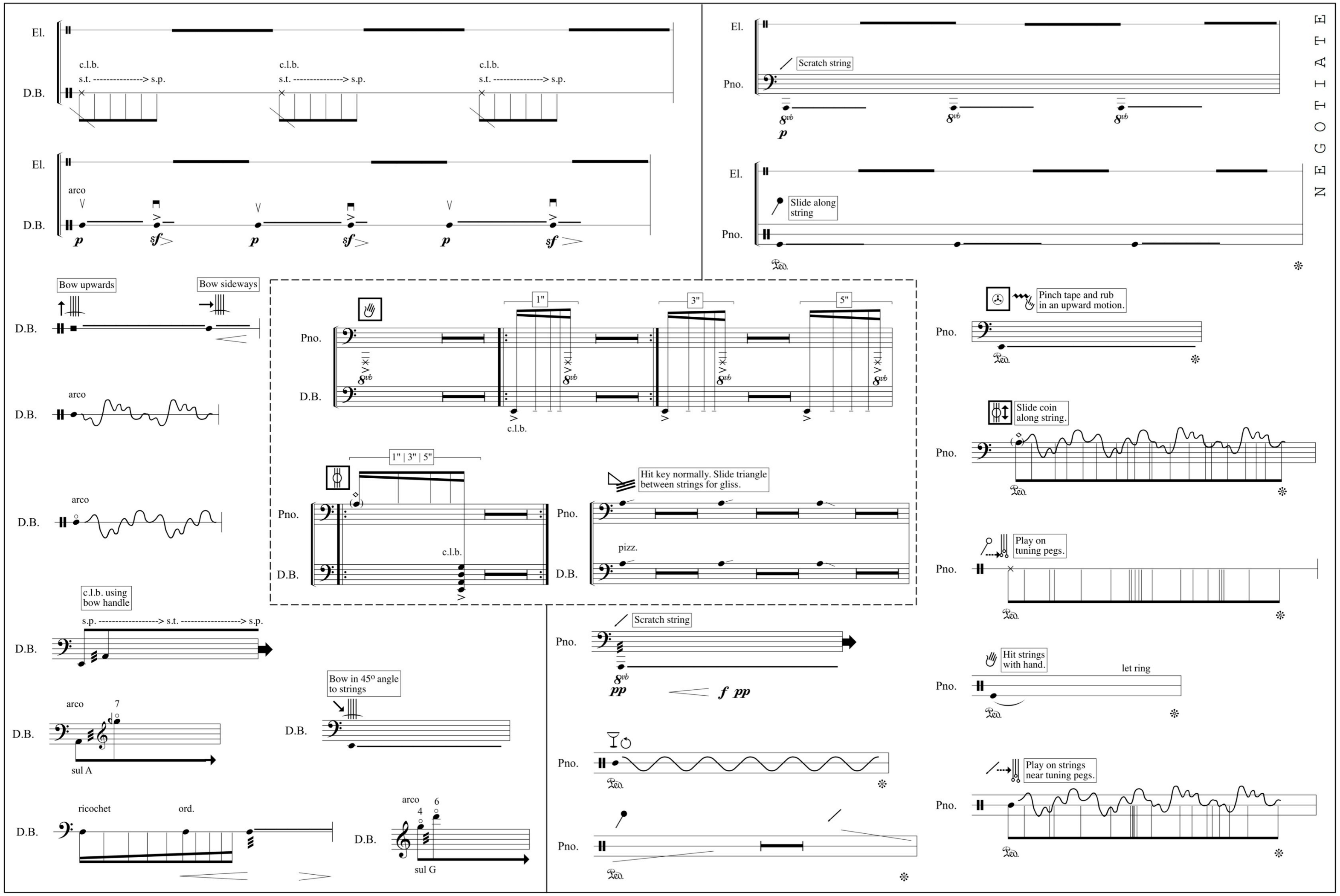
Converge/Diverge is a study on collaborative emergence and joint agency in music-making. During the performance, the 2 musicians are free to explore three different states of the Interactive Music System (IMS): "converge", "diverge" and "negotiate". The terms convergence and divergence in this context refer to the degree of timbral similarity between the two inputs (piano and double bass). The interaction dynamics between the musicians are both sonified and influenced by the IMS, which, in addition to monitoring the interaction between the two musicians and responding accordingly, can initiate two additional states ("cooperate" and "compete").

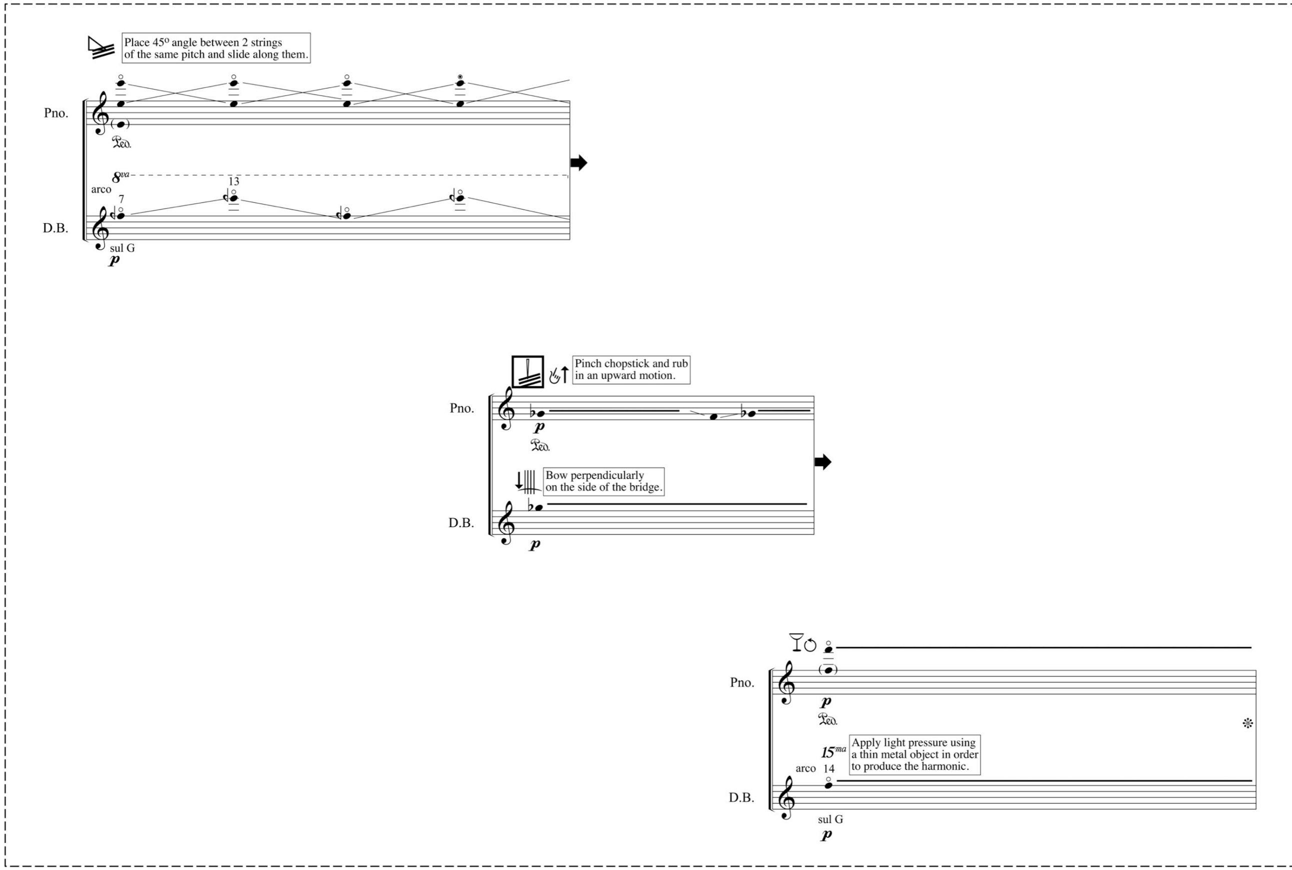
The default state of the IMS is negotiation. In this state the musicians interact with each other by taking turns. The response of the IMS in this scenario consists in generating spectrally compressed variations of the input signal. As only a small number of frequencies are used for resynthesis, the electronic sound resembles a resonance, rather than an exact imitation of the human input.

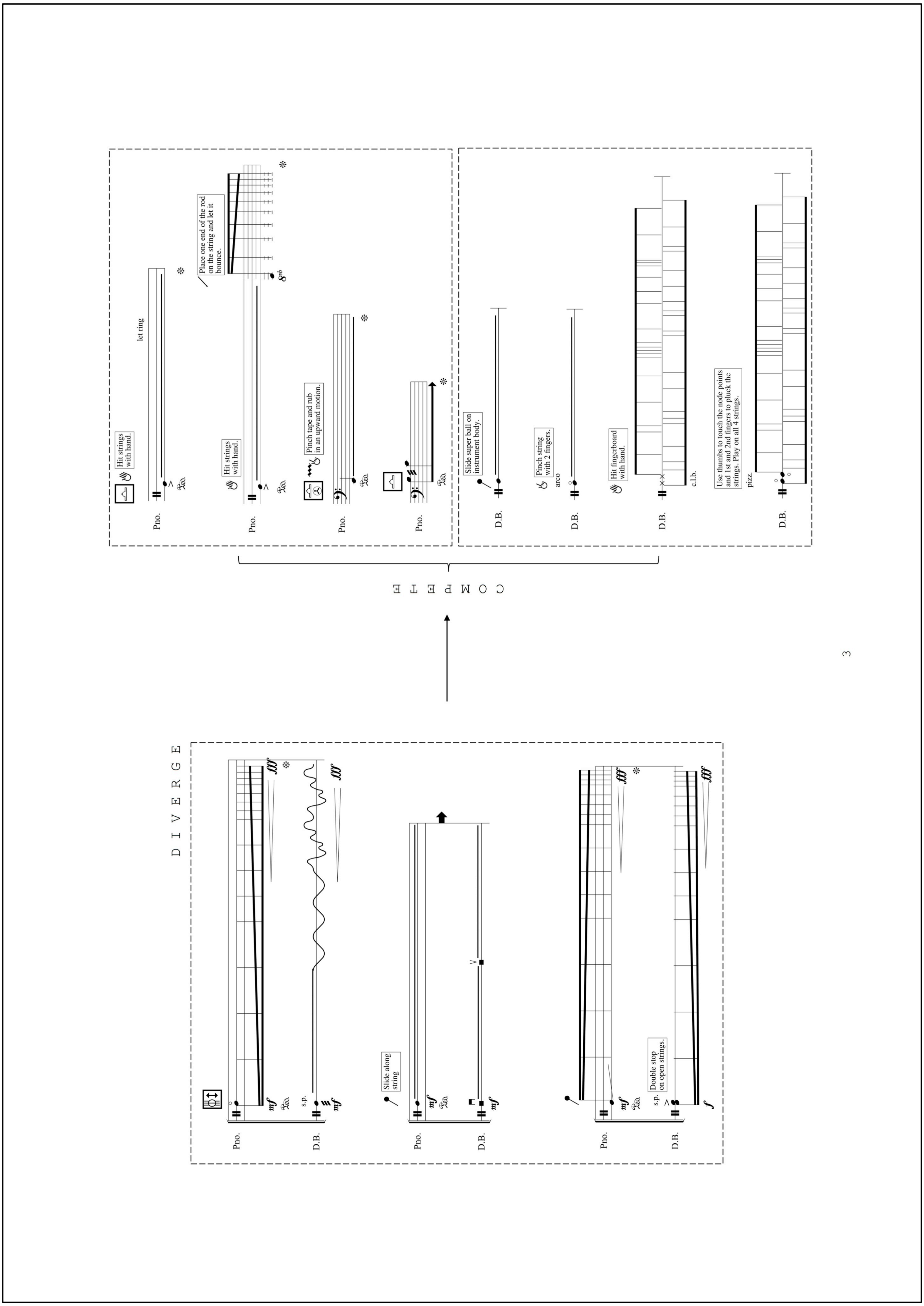
Convergence and divergence can only be initiated by both musicians jointly. Two separate pools of synchronous actions are provided as sound material for "convergence" and "divergence" (pages 2 and 3 of the score respectively). By playing sound material from one of these pools, a musician extends an invitation to their co-player to "converge" or "diverge". If the second musician decides to accept the invitation and joins their co-player, the IMS begins to assess their current interaction with the purpose to determine whether they are in "convergence" or "divergence" with each other.

When divergence is detected, the IMS responds by initiating one of two additional scenarios: "compete" or "cooperate". In the latter, the system responds by generating a static spectrum, essentially becoming unresponsive. In order for this spectrum to be dissolved, the musicians have to "cooperate" (i.e., "converge"). A pulsating electronic sound (the result of amplitude modulation with a square wave) is an indication that the system has entered the "compete" mode. In this scenario, the musicians compete for the computer's attention, which only responds to the musician currently playing the most "original" sound material. "Originality" in this context is judged by calculating the spectral dissimilarity between currently and previously played sound material for each musician. In this interaction scenario, the musicians can use the notated material as a starting point and/or improvise freely, introducing new sounds of their own choosing. The duration of this scenario is determined by the IMS.

Approx. duration: 15 min.


NEGOTIATE
El.
D.B.
c.l.b.
s.t. --------------> s.p.
arco
p
sf
Scratch string
Pno.
Slide along string
Bow upwards
Bow sideways
Pinch tape and rub in an upward motion.
1"
3"
5"
c.l.b.
Slide coin along string.
1" | 3" | 5"
Hit key normally. Slide triangle between strings for gliss.
pizz.
Play on tuning pegs.
c.l.b. using bow handle
s.p. -----------------> s.t. -----------------> s.p.
f pp
pp
Hit strings with hand.
let ring
Bow in 45º angle to strings
sul A
Play on strings near tuning pegs.
ricochet
ord.
sul G


Place 45º angle between 2 strings
of the same pitch and slide along them.
Pno.
D.B.
arco
8va
7
13
sul G
p
Pinch chopstick and rub
in an upward motion.
Bow perpendicularly
on the side of the bridge.
15ma
arco
14
Apply light pressure using
a thin metal object in order
to produce the harmonic.


DIVERGE
Pno.
D.B.
s.p.
Slide along string
Double stop on open strings.
COMPETE
Hit strings with hand.
let ring
Place one end of the rod on the string and let it bounce.
Pinch tape and rub in an upward motion.
Slide super ball on instrument body.
Pinch string with 2 fingers.
arco
Hit fingerboard with hand.
c.l.b.
Use thumbs to touch the node points and 1st and 2nd fingers to pluck the strings. Play on all 4 strings.
pizz.
3

B I A S

for bass clarinet and interactive music system

[2020]

Artemi-Maria Gioti

Score

NOTATION

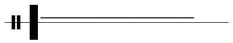
Pitched sound

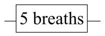

Air tone


Half air; half pitch

Key release

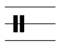
Any multiphonic


Duration line



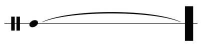
Duration indication; the notated fragment should be performed within the duration of 5 breaths. Breathing rests should be as unnoticeable as possible.


1-line staff: indeterminate pitch.


3-line staff: indeterminate pitch. The 3 lines represent different registers and are used to show the distribution of the notated material across the range of the instrument.


Continue in pattern by playing variations of the notated material.


Microtonal glissando.


Transition from a single note to a multiphonic containing that note.

INSTRUCTIONS

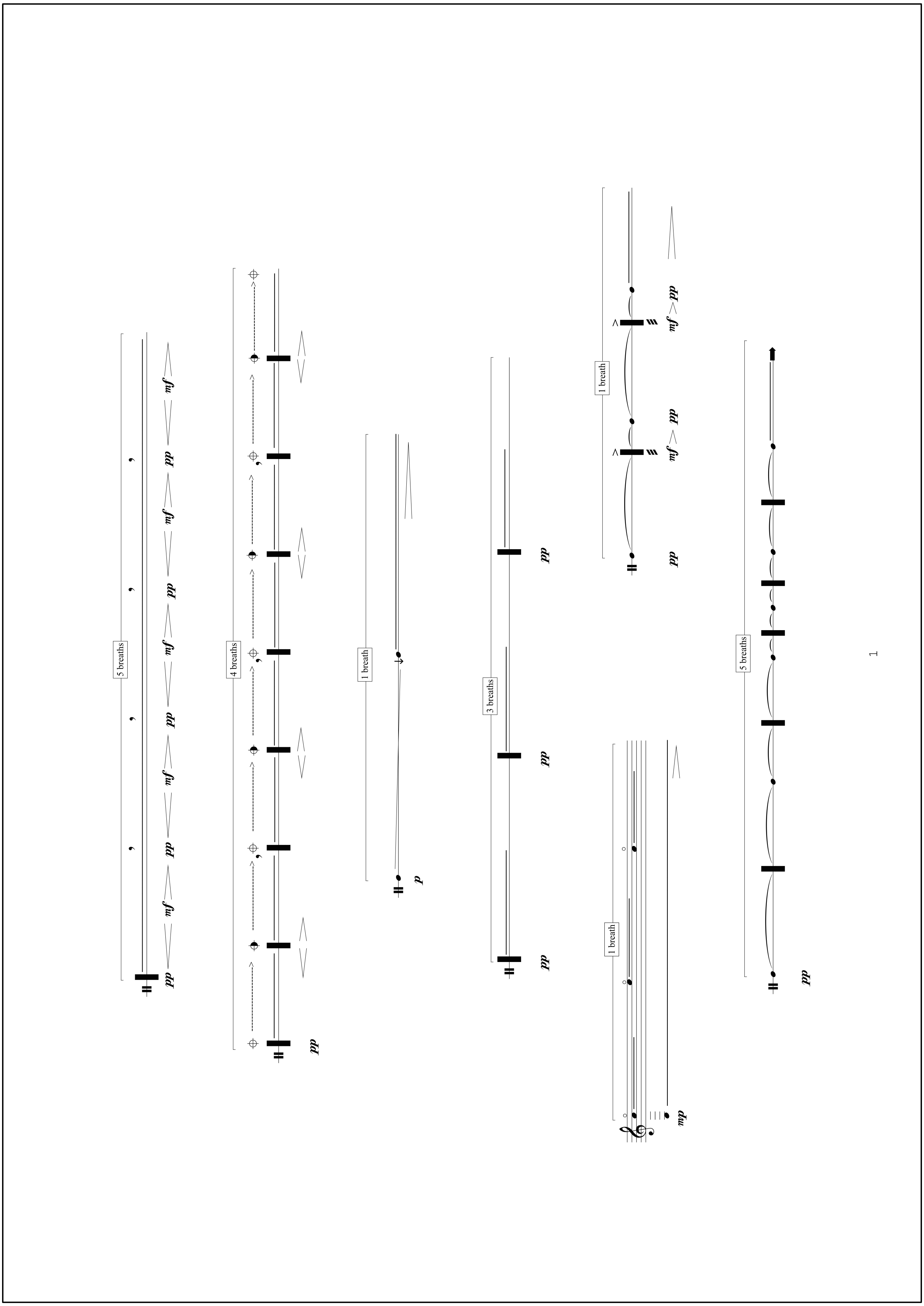
Bias, for bass clarinet and interactive music system, explores the concept of AI bias, a phenomenon that consists in machine learning algorithms making arbitrary assumptions about data, or amplifying any human bias present in the data. In *Bias*, two Neural Networks trained to simulate the composer's aesthetic judgments determine how the interactive music system will respond to the musician's input. Specifically, the interactive system imitates sounds and textures it finds "interesting", but remains silent or proposes new sound material when it loses interest in the musician's input. The composition takes a critical and subversive approach to machine learning, the aim of which is not to simulate the composer's aesthetic preferences as accurately as possible, but rather to use them as a departure point for the development of AI bias.

The interactive system can choose between two different interaction scenarios: "follow" and "lead". In the former, the system evaluates the sound material played by the musician and responds to sounds it finds "interesting" by imitating them. Evaluation is performed both on a sound event and a texture basis (over a window of 20 sec). When the interactive music system loses interest in the musician's input, it takes the "lead", by proposing new sound material. In this state, the interactive music system is no longer responsive, i.e., it is no longer listening to the musician. When the system enters the "lead" mode, performers should "follow", by improvising in response to the sound material played by the computer.

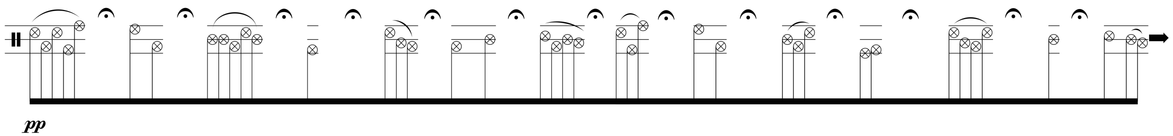
Performers are free to explore different types of sound material, including sounds that the computer might not respond to, and should try to demonstrate the system's idiosyncracies and capabilities in their interpretation.

All sound material generated by the computer during the performance is collected during its interactions with musicians. The interactive music system stores sounds it finds "interesting" in a sound database, which is updated every time it interacts with a musician (i.e., during rehearsals and performances of the work). This effectively means that none of the electronic sounds heard in the performance are "composed". By performing the piece, musicians interact with and contribute to this collectively assembled sound database, shaping the identity of the work and influencing its future performances.

Approx. duration: 12 min.


5 breaths
4 breaths
1 breath
3 breaths
1 breath
1 breath
5 breaths
1

Press the keys as slowly and as quietly as possible, making sure that only the "release" segment of the gesture sounds/is detected by the onset detection algorithm. Rests between individual phrases/gestures can vary in length. Play only for as long as the computer is following.



Appendix 3: Questionnaires

Imitation Game: Naïve Rehearsal Questionnaire

*Required

1. The system was listening: *

Mark only one box.

☐ All of the time

☐ Some of the time

☐ Never

2. Which of the following statements is true? *

Mark only one box.

☐ The system was responsive to human input, but not able to autonomously generate sound material

☐ The system was both responsive to human input and able to autonomously generate sound material

☐ The system was fully autonomous and did not respond to human input

3. The system was responsive to short-term changes. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

4. The system was responsive to long-term changes. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

5. The system responded to specific parameters of the human input *

Mark only one box.

☐ Yes

☐ No

6. If yes, please specify which parameters (e.g. pitch, dynamics etc.):

____

____

____

____

____

____

____

7. The response of the system was *

Mark only one box.

☐ Synchronous to human actions (i.e., the computer was playing at the same time as the musician)

☐ Asynchronous to human actions

8. How would you describe the degree of controllability of the system? *

Mark only one box.

	1	2	3	4	5	
very low	☐	☐	☐	☐	☐	very high

9. How would you describe the degree of predictability of the system? *

Mark only one box.

	1	2	3	4	5	
very low	☐	☐	☐	☐	☐	very high

10. How would you describe the influence that the system had on your actions? *

Mark only one box.

	1	2	3	4	5	
very weak	☐	☐	☐	☐	☐	very strong

11. Comments:

Imitation Game: Informed Rehearsal Questionnaire

*Required

1. The system was listening: *

Mark only one box.

☐ All of the time

☐ Some of the time

☐ Never

2. Which of the following statements is true? *

Mark only one box.

☐ The system was responsive to human input, but not able to autonomously generate sound material

☐ The system was both responsive to human input and able to autonomously generate sound material

☐ The system was fully autonomous and did not respond to human input

3. The system was responsive to short-term changes. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

4. The system was responsive to long-term changes. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

5. The response of the system was *

Mark only one box.

☐ Synchronous to human actions (i.e., the computer was playing at the same time as the musician)

☐ Asynchronous to human actions

6. How would you describe the degree of controllability of the system? *

Mark only one box.

	1	2	3	4	5	
very low	☐	☐	☐	☐	☐	very high

7. How would you describe the degree of predictability of the system? *

Mark only one box.

	1	2	3	4	5	
very low	☐	☐	☐	☐	☐	very high

8. How would you describe the influence that the system had on your actions? *

Mark only one box.

	1	2	3	4	5	
very weak	☐	☐	☐	☐	☐	very strong

9. Comments:

Converge/Diverge: Naïve Rehearsal Questionnaire

*Required

1. I was able to identify ___ different system behaviors (please insert a number). *

2. Please describe these behaviors. *

3. By changing its behavior the system influenced the musicians' actions and changed the course of the improvisation. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

4. In different parts of the improvisation, the system responded to: *

Select all that apply.

☐ Both musicians

☐ One musician at a time

☐ None of the musicians

5. The system changed its behavior in response to the musicians' actions. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

6. The system changed its behavior independently of the musicians' actions. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

7. The system responded to the musicians' actions in a predictable way. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

8. The system responded differently to the states of "convergence" and "divergence". *

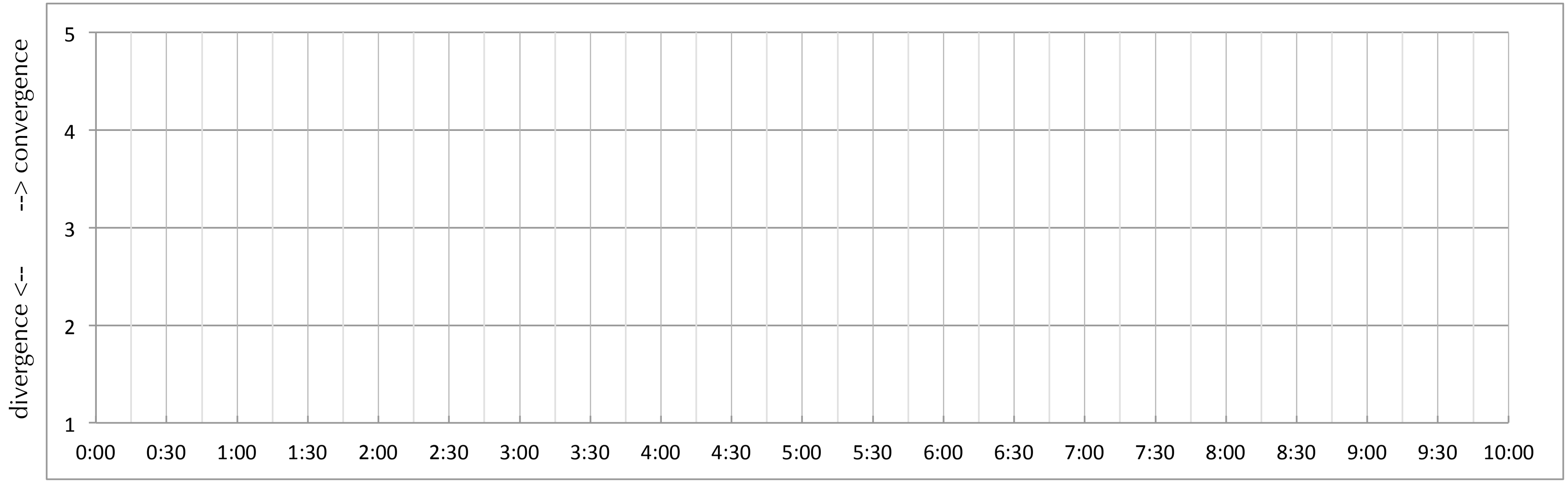
Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

9. How did the system respond to "convergence"? *

10. How did the system respond to “divergence”? *

Please evaluate the degree of convergence between the actions of the 2 musicians, as follows:
1. very low,
2. low,
3. moderate,
4. high,
5. very high.
Mark one value for every 15 seconds of the improvisation. You can watch the video as many times as necessary.



Comments:

Bias: Naïve Rehearsal Questionnaire

*Required

1. I was able to identify ___ different system behaviors (please insert a number). *

2. Please describe these behaviors. *

3. The system changed its behavior in response to my actions. *
Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

4. The system changed its behavior independently of my actions. *
Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

5. The system responded to my actions in a predictable way. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

6. By changing its behavior the system influenced my actions and changed the course of the improvisation. *

Mark only one box.

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

7. The system was listening *

Mark only one box.

☐ All of the time

☐ Some of the time

☐ Never

8. How did the system respond to sounds it found "interesting"? *

____

____

____

____

____

____

9. How did the system respond to sounds it found "uninteresting"? *

10. Were there moments in which the computer was "leading" the improvisation? *

Mark only one box.

- [ ] Yes
- [ ] No

11. If yes, please describe these moments.