Mixed-Initiative Music Making

Collective Agency in Interactive Music Systems

Notto J.W. Thelle

NMH Publications 2022:4

Norges musikkhøgskole
Norwegian Academy of Music

NORDART
Arne Nordheim-senteret

Mixed-Initiative Music Making

Collective Agency in Interactive Music Systems

Notto J.W. Thelle

Dissertation for the PhD degree
Norwegian Academy of Music, Oslo, 2022
NMH Publications 2022:4

NMH Publications 2022:4

ISSN 0333-3760
ISSN 2535-373X (online)
ISBN (print) 978-82-7853-304-8
ISBN (online) 978-82-7853-305-5

Norwegian Academy of Music
PO Box 5190 Majorstua
0302 OSLO
Tel.: +47 23 36 70 00
E-mail: post@nmh.no
nmh.no

Setting and printing: Aksell, Oslo, 2022

Acknowledgements

Having spent more than three years researching collective agency, I am well aware that many people's labor, wisdom, talent, compassion, and enthusiasm have contributed to this thesis. My deepest gratitude goes to the two people who have provided me with support and guidance from beginning to end:

- Dr. Sidsel Karlsen, my PhD advisor, whose capability for seeing the wide interdisciplinary picture while maintaining a keen eye on disciplinary rigor always kept me on track. I truly enjoyed discussing ideas with Sidsel, and her feedback on my progressing work was consistently balanced between firmness and encouragement. I always felt more confident and determined about my work after meetings with Sidsel.
- Dr. Rebecca Fiebrink, my PhD advisor, who has provided me with invaluable insights into the domains of creative computing, machine learning, and HCI. Rebecca's ability to quickly identify problems and efficiently communicate ways to address them is remarkable. I am also grateful for having been included in Rebecca's weekly online PhD group meetups, first at Goldsmiths and later at UAL. These meetups were always inspiring and helped me stay focused.

At the Norwegian Academy of Music, I would like to thank the PhD Committee for their tireless efforts to maintain the superb quality of the academy's PhD program. In particular, I would like to mention Solveig Christensen for her impressive work ethics and kind support to all PhD fellows, and Øivind Varkøy and Sigrid Røyseng for heading the committee during my fellowship period. I am also grateful to Mats Claesson, Anders Eggen, Tore Simonsen, and of course, all the PhD fellows at the academy. Too many to mention, but you all know who you are!

Several tutors, collaborators, and friends have influenced and supported different parts of the project. Many thanks to: Kıvanç Tatar, Philippe Pasquier, Sebastian Gregorius Winther-Larsen, and Bernt Isak Wærstad for invaluable conceptual and technical insights. Dr. Andrew McPherson for being my opponent in the trial disputation. Jan Wiese, Morten Qvenild, Labdi Ommes, and Gyrid Nordal Kaldestad for trying Spire Muse and providing feedback. Vidar Fineide at Finde AS for quick and efficient design of web application for the final study. Henrik Sundt, Eirik Blekesaune, and Bálint Laczkó for conversations that helped me along. All the (anonymous) participants in my studies deserve mention as well.

Finally, thanks to my family and friends for continuous support throughout this endeavor. Especially, a huge thanks to Kristine for her love and patience, and to Leo, Helene, and little Notto Tobias for always reminding me of what is important in life.

Abstract

A *mixed-initiative* user interface is one where both human and computer contribute proactively to a process. A *mixed-initiative creative interface* is the same principle applied in the domain of computational creativity support, such as in digital production of music or visual arts. The title "Mixed-Initiative Music Making" therefore implies a kind of music making that puts human and computer in a tight interactive loop, and where each contributes to modifying the output of the other. Improvisational collective music making is often referred to as *jamming*. This thesis focuses on jamming-oriented approaches to music making, which takes advantage of the emergent novelty created by group dynamics. The research question is: *How can a mixed-initiative interactive music system aid human musicians in the initial ideation stage of music making?*

Starting from a vantage point of dynamical systems theory, I have addressed this question by adopting a Research through Design approach within a methodological framework of triangulation between theory, observation, and design. I have maintained a focus on the activity of collective music making through four studies over a period of two years, where the gradual development of a mixed-initiative interactive music system has been informed by findings from these studies. The first study was a focus group with musicians experienced in collective music making, where the goal was to establish commonalities in musical interaction and idea development with a focus on viable conceptual frameworks for subsequent studies. The second study was a case study of two improvising musicians engaged in an improvised session. They were separated in two rooms, and could only communicate instrumentally or through preset commands on a computer screen. The session was analyzed in terms of how the musicians dynamically converged and diverged, and thus created musical progression. In the third study, several musicians were invited to jam with a prototype of an interactive music system. Unbeknownst to them, they had been recruited to a *Wizard of Oz* study—behind the scenes was a human keyboard player pretending to be a computational agent. The purpose of this arrangement was to obtain empirical data about how musicians experience co-creativity with a perceived computational agent before the implementation of the computational agent had begun in earnest. In the final study, two different implementations of a mixed-initiative interactive music system were developed for a comparative user study, where the tradeoff between user control and system autonomy was a central premise.

Combined, the studies show that a mixed-initiative interactive music system offers musicians freedom from judgement and freedom to explore their own creativity in relation to an unknown agency. Social factors make these kinds of freedom difficult to attain with

other musicians. Hence, playing with interactive music systems can lead to different kinds of musical interaction than can be achieved between people. An acceptance of machine aesthetics may lead to surprising creative results. Repeated exposure to mixed-initiative interactive music systems could help cultivate attitudes that are valuable for collective music making in general, such as maintaining a process-oriented approach and accepting the loss of idea ownership.

Sammendrag

Begrepet *mixed-initiative* (blandet initiativ) hentyder til brukergrensesnitt der både menneske og datamaskin bidrar proaktivt til en prosess. Et *mixed-initiative kreativt brukergrensesnitt* er samme prinsipp overført til domenet *computational creativity support* (datastøttet kreativitet), som for eksempel i digital musikkproduksjon eller elektronisk kunst. Tittelen «Mixed-Initiative Music Making» innebærer dermed en form for musikkskaping som plasserer menneske og maskin i en tett interaktiv sløyfe, og der begge påvirker hverandres bidrag. Improvisasjonsbasert kollektiv musikkskaping omtales ofte som «jamming». Denne avhandlingen fokuserer på jamming-orienterte tilnærminger til musikkskaping som drar nytte av den fremvoksende nyskapingen som skjer i gruppedynamikk. Forskningsspørsmålet er: *Hvordan kan et mixed-initiative interaktivt musikksystem hjelpe musikere i den tidlige idemyldringsfasen av musikkskaping?*

Med et utgangspunkt i dynamisk systemteori har jeg tatt fatt på dette spørsmålet ved å bruke en Research through Design-tilnærming innenfor et metodologisk rammeverk med triangulering mellom teori, observasjon og design. Jeg har opprettholdt et fokus på kollektiv musikkskaping som aktivitet gjennom fire studier over en toårsperiode, der den gradvise utviklingen av et mixed-initiative interaktivt musikksystem har blitt informert av funn fra disse studiene. Den første studien var en fokusgruppe med musikere erfarne i kollektiv musikkskaping. Her var målet å etablere fellestrekk i musikksamspill og ideutvikling med tanke på gunstige konseptuelle rammeverk for påfølgende studier. Den andre studien var en case studie av to improviserende musikere som spilte et improvisert stykke musikk. De ble adskilt i to forskjellige rom, og kunne kun kommunisere via instrumentene eller gjennom forhåndsdefinerte beskjeder via en dataskjerm. Samspillet ble analysert i forhold til hvordan musikerne dynamisk konvergerte og divergerte, og på den måten skapte musikalsk progresjon. I den tredje studien ble flere musikere invitert til å jamme med en prototype av et interaktivt musikksystem. Uten at de var klare over det hadde de blitt rekruttert til en såkalt *Wizard of Oz* studie – bak fasaden var det et menneske på elektroniske tangenter som simulerte en data-musiker. Formålet med dette oppsettet var å skaffe empirisk data på hvordan musikere erfarer samskaping med en datamaskin før programmeringen av det interaktive systemet hadde blitt påbegynt. I den siste studien ble to forskjellige implementasjoner av et mixed-initiative interaktivt musikksystem utviklet til en komparativ studie, der avveining mellom brukerkontroll og system-autonomi var et grunnleggende premiss.

Til sammen viser funn fra studiene at et mixed-initiative interaktivt musikksystem tilbyr musikere frihet fra å bli kritisk vurdert og frihet til å utforske sin egen kreativitet i relasjon til et ukjent aktørskap. Sosiale faktorer gjør denne typen friheter vanskelige å oppnå med andre musikere. Dermed kan det å spille med interaktive musikksystemer føre til andre typer

musikksamspill enn man får til mellom mennesker. Det å akseptere maskinens estetikk kan føre til overraskende kreative resultater. Gjentakende eksponering til mixed-initiative interaktive musikksystemer kan bidra til å kultivere holdninger som er verdifulle for kollektiv musikkskaping generelt, som for eksempel det å opprettholde en prosessorientert tilnærming og akseptere tap av idemessig eierskap.

Contents

Figures

Tables

Videos

1 Introduction

In my experience as a musician, new ideas often emerge as a result of interactions with other musicians. Particularly fascinating are contexts where it seems that an idea "just happens", and it is impossible to attribute ownership of that idea to one single musician. To some degree, I have had similar experiences when playing with music software. Choosing a wrong process or glitches in the software makes unexpected things happen. Sometimes, "happy accidents" are pivotal to a composition, and a result could leave me incredulous. Who made this? It was not just me. It was the interaction. The more I grew interested in music technology, the more I came to realize that we do not have enough software that specializes in open-ended creative human-computer interaction—particularly the kind of interaction that is not goal-oriented, but where the interaction is an end in itself. In short, I wanted to jam with a computer in the same way as musicians improvise together to cultivate their collective creativity. This led me to my current interest in *interactive music systems*.

The idea that computers can make music may seem alien or even threatening to some musicians. However, music and technology are both quintessentially human phenomena. The tools we make change us, because that is what we want. We want challenge; we want surprise; we want new ideas. Even when dealing with *artificial intelligence* (AI), humans are at the inputs and outputs of the computational system, curating the entire process. This is also the case with music AI. A human defines a musical concept and manifests it in code for the computer to execute. At the other end, a human engages with the outcome. In between there is a model. The model is a part of the work, in the same way that a traditionally notated score is part of a symphony. Interactions are at the core of the creative process in both cases. A human interacts with instruments (technology) and the conceptual framework of a music system. As long as there are people who enjoy the outcome, those who do usually call it music.

In this PhD thesis, I will be focusing on a particular hybrid of systems that combine the activity of human instrumental performance and computer-generated sound. The context I am interested in is human-computer *co-creativity*, where a musician views the computer as a virtual musical partner.

1.1 Definitions

Before posing the research question, I will define the core terms in the title of the thesis.

1.1.1 Mixed-initiative music making

A *mixed-initiative user interface* is one where both human and computer contribute proactively to a process (Horvitz, 1999). This requires an interface design where the user and computational system share decision-making, and where the style and pace of turn taking is adapted to the task. A *mixed-initiative creative interface* is the same principle applied in the domain of computational creativity support, such as in digital production of music or visual arts (Deterding et al., 2017). The term music making is chosen to deemphasize the rigid "division of labor" associated with composition, performance, improvisation, and listening—music making encompasses them all. The title "Mixed-Initiative Music Making" implies a kind music making that puts human and computer in a tight interactive loop, and where each contributes to modifying the output of the other (Deterding et al., 2017, p. 628). This is different from *algorithmic composition*, which generates music according to a predefined set of rules.

1.1.2 Collective agency

Agency is the capacity of an entity to act in a given environment. While this definition itself is quite straightforward, there are major disagreements as to who or what agency applies. Historically, the dominant view has been that agency can only be ascribed to intentional beings. This delimits agency to human beings, because people are the only creature that can make their inner mental representations known. Some sociologists argue that agency could also be ascribed to non-humans and even objects (e.g. Latour, 2005). Misselhorn (2015) points out that this debate has a blind spot. Individual agency cannot account for the type of behavior that emerges from complex hybrid systems involving both human and artificial agents. *Collective agency* accounts for the behavior of systems that are irreducible to the intentions of the individuals that constitute them (Misselhorn, 2015, p. iii). A *sociotechnical perspective* proposes a gradual concept of distributed agency that can be used to "describe and discriminate between different levels and grades of action without any regard to the ontological status of the acting unit, may it be human-like or machine-like" (Rammert, 2008, p. 3). According to this view, a computational system may be a non-intentional entity that one can relate to not as an object or tool, but as a collaborator with some degree of agency. In Chapter 2, I will elaborate further on these different conceptions of agency.

1.1.3 Interactive music systems

According to Robert Rowe, interactive computer music systems are "those whose behavior changes in response to musical input" (Rowe, 1992). Rowe's seminal work provides further classification of such systems, built on the combination of three dimensions: (i) drive—a binary classification into score-driven or performance-driven; (ii) response method—a ternary classification into transformative, generative, or sequenced, and (iii) paradigm—a continuous spectrum from "instrument" to "player" (Gifford et al., 2018).

As detailed in Rowe (1992), the first dimension distinguishes between *score-driven* and *performance-driven* systems. Score-driven systems rely on predetermined event collections or musical material programmed to match the musical input. An example of this category is score-following software that can provide performers with accompaniment in a score-based fashion, based on cue points in the input. Performance-driven programs, on the other hand, have no score representation and are designed to respond to the input in a more general, open-ended fashion.

The next category dimension divides the response methods of interactive systems into *transformative*, *generative,* and *sequenced*. Transformative systems apply transformations to the received input—a sort of advanced or "intelligent" processor. Generative systems use sets of rules or models to generate new output based on features extracted from the input. A system in the sequenced category will use prerecorded music fragments in response to different kinds of input.

The final dimension distinguishes between the *instrument* and *player* paradigms. Systems in the instrument paradigm are ones designed as extended instruments, and may include both completely new interfaces and traditional instruments with added sensors that contribute toward producing output exceeding normal instrument response. Conversely, systems following a player paradigm are a separate entity from the performer. In other words, a performer can *play together with* an interactive music system following the player paradigm, whereas a performer *plays* a system in the instrument paradigm.

The distinctions described here should not be seen as discrete categories, but rather as points along continua between extremes. Any system may show a combination of attributes belonging to either end of the continua, but in general it is purposeful to classify a system as "more like" one category than another. The interactive music systems that are the focus of this thesis are ones that are *performance-driven*, *generative,* and belonging to the *player* paradigm.

1.2 Research questions

The vantage point of this PhD is a motivation to help towards developing interactive music systems that can act as collaborators in the initial, explorative ideation stage of collective music making. This ideation stage is often characterized by improvisations starting with basic musical ideas—a sonic equivalent to sketching, i.e. *jamming*. In my personal experience as a musician and composer, many creative ideas originate in a social context, during interactions with an agency opposed to or tangential to one's own.

In interactions between humans, the distinction between actions and decision-making is barely noticeable—they are intrinsically interwoven. In human-computer interaction (HCI), on the other hand, the human user often acts as a substitute for the computational system's lack of autonomy. Most software interfaces are essentially a submission of decision-making power to the human user. An effect is that the user may become preoccupied with handling this aspect of the interaction, perhaps sometimes to the detriment of creative engagement. Collective agency is missing.

The notion of a mixed-initiative creative interface where both human and machine contribute to musical ideation—a virtual jamming partner—led to the formulation of my main research question:

- *How can a mixed-initiative interactive music system aid human musicians in the initial ideation stage of music making?*

In view of the issues described above, what seems to be missing is an autonomous computational agent that can contribute to a collective creative agency. On the other hand, too much autonomy is undesirable. The system should be both *reactive* and *proactive*—truly interactive. The major challenge underpinning the main research question is not a technological one—it is deeply human. The dynamics between human musicians when engaged in creative musical interaction, although extensively researched, remains poorly understood. Without a comprehensive understanding of these interaction dynamics between humans, addressing the main question would be a quixotic experiment. Therefore, even before getting to the technological issues, the following research sub-questions must be addressed:

- *What can be learned about the interaction dynamics between musicians in the ideation stage of collective music making?*
- *To which degree can these interaction dynamics serve as a model for an interactive music system?*

1.3 A dynamical systems approach

I have opted to address the research questions through the lens of a conceptual framework based on dynamical systems. Although *dynamical systems theory* (DST) originated in mathematics, it has garnered much interest from disparate fields such as economics, law, psychology, and music. One reason for the popularity of DST can be attributed to some of the visual representations of phenomena and intuitively mind-capturing concepts emanating from the field. Most of us are familiar with aesthetically pleasing fractals and the butterfly effect. However, applying knowledge across disciplines could also lead to bad science. The caveats of borrowing knowledge and conceptual displacement are thoroughly discussed in Chapter 3.

In the context of this PhD, DST has served as a conceptual tool to navigate between the human-oriented sub-questions and the more technologically oriented main research question. The framework functions as a metaphor for the high-level dynamics emerging from a complex system of interactions including both physical and psychological phenomena. It would be incorrect to claim that musical interaction *is* a dynamical system, although in many cases it *appears to behave as one*, and musicians express themselves in ways that lend belief to it *feeling like being part of one*. Therefore, I have been interested in finding measurable correlates to findings supported by DST concepts. Rather than serving as proof, correlates serve to strengthen the rationale behind choosing this framework as a way to understand the complex dynamics of collective musical interaction.

1.4 Structure of the thesis

A literature review in Chapter 2 places the research of interactive music systems into a historical and scientific context, and introduces the research fields of HCI, machine learning, agency, creativity, and music making—all of which are important background for subsequent chapters.

Chapter 3 introduces the theoretical framework of dynamical systems theory, and presents a case for why such a framework is useful for the study and development of mixed-initiative interactive music systems.

Chapter 4 presents the overall methodological structure of the research reported in this thesis. It introduces a Research through Design approach and proposes a methodological framework involving a triangulation between theory, observations, and design. A four-study plan for empirical data gathering is put forward, and ethical considerations for the studies are taken into consideration.

Chapter 5 presents the first study, which is a focus group discussion with musicians experienced in collective music making. The goal is to establish commonalities in musical interaction and idea development with a focus on viable conceptual frameworks for subsequent studies.

Chapter 6 details a case study of two improvising musicians engaged in an improvised session. The musicians are physically separated, and may only communicate through musical sound or commands using a computer interface. The session is analyzed in terms of how the musicians create musical progression by converging to collective sequences and articulating transitions between sequences.

Chapter 7 describes the third study, where several musicians are invited to jam with a prototype of an interactive music system. Behind the scenes is a human keyboard player pretending to be a computational agent. The purpose of this arrangement is to obtain empirical data about how musicians experience mixed-initiative interaction before developing a functioning interactive music system.

Chapter 8 describes the development and evaluation of a mixed-initiative interactive music system. Two different implementations of the system are used in a comparative user study, where the tradeoff between user control and system autonomy is a central premise.

Chapter 9 discusses the findings from the four studies within the framework of addressing the research questions, and by drawing upon the theoretical underpinnings of the thesis.

Finally, Chapter 10 concludes the thesis by summarizing contributions, limitations, and future work.

2 Literature review

An interactive music system comprises humans and machines. The communication in between—the interaction—can be objectively observed or subjectively described. In terms of research focus, this broaches a wide range of scientific domains, including engineering, computer science, sociology, and psychology. Methods can be quantitative, qualitative, or both. A field that straddles these domains is human–computer interaction (HCI), which studies the design of computer technology and how it is used. As for what interactive music systems are designed for—musical interaction—several fields have informed research into this subject. This includes research into machine learning, creativity, and music making. After providing a historical background on interactive music systems and HCI, this chapter will provide a brief overview of these fields, and demonstrate how they can illuminate different aspects of musical interaction and interactive music systems.

2.1 Interactive music systems in a historical context

The earliest documentations of musical automata—instruments that play by themselves—can be traced back to the Islamic golden era in the 9th to the 13th centuries. The Banū Mūsā brothers devised a programmable, self-playing flute driven by a hydraulic pump as early as the 9th century, and the 12th century engineer and inventor Ismail al-Jazari was famous for various mechanical automata, some of which featured robotic sculptures of musicians sounding instruments such as flutes, harps, and drums (Abattouy et al., 2016). In Europe, mechanical automata began appearing after the Renaissance. This long history of mechanical musical automata suggests that the human desire to use technology—including automation—for artistic purposes is nothing new. Mostly, however, these automata were not interactive—once set in motion, they mostly did not respond to anything from outside of their deterministic constructions. A merging of the programmability of automata and the interactivity afforded by musical instruments became more feasible with the advent of electronic synthesizers and digital computers. This review therefore begins at this point in history.

2.1.1 Interactive composing systems

Experiments with computer-generated music and computer-assisted composition can be traced back to the latter half of the 1950s—within the first decade of the invention of general-purpose computers. Max Mathews, with the help of his team at Bell Laboratories, is widely recognized as being the first person who wrote a computer program that generated melodies in 1957 (Roads, 1996, p. 830). In the same year, Lejaren Hiller and Leonard Isaacson (1959) developed the world's first compositional score generated algorithmically by a computer program, *The Illiac Suite for String Quartet*.

While similar projects started mushrooming around North America, and later in Europe and the rest of the world, they were almost exclusively tied to research institutions that could afford computers at that time. The early systems were rule-based and deterministic, and did not operate in real-time. In parallel developments, other initiatives utilizing relatively cheaper analog electronic circuitry were underway. The first modular voltage-controlled synthesizers, pioneered by Robert Moog and Donald Buchla, became commercially available in the mid-1960s. Analog synthesizers, with their opportunities for real-time control of sound, had a larger immediate impact on music culture in general than computer music, which developed more exclusively in the academic realm.

The drawback of not being able to use computers to generate sound in real-time was addressed by using computer programs to control analog synthesizers algorithmically. In 1966–69, Joel Chadabe received funding to build an automated synthesizer system, and he commissioned from Robert Moog the construction of what became the *Coordinated Electronic Music Studio* (CEMS) system at the State University of New York in Albany. Describing the system as "the real-time equivalent of algorithmic composition" (Chadabe, 1997, p. 286), the system eventually also featured joysticks that allowed a human to influence a complex network of modular interconnections in the CEMS (J. V. Albert, 2013). Chadabe's own description of performing with CEMS highlights a qualitative step away from the purely reactive electronic instruments of the day to a more interactive experience:

> Because I was sharing control of the music with sequencers, I was only partly controlling the music, and the music, consequently, contained surprising as well as predictable elements. The surprising elements made me react. The predictable elements made me feel that I was exerting some control. It was like conversing with a clever friend who was never boring but always responsive. I was, in effect, conversing with a musical instrument that seemed to have its own interesting personality. (Chadabe, 1997, p. 287)

Meanwhile, at the University of Illinois, Salvatore Martirano was building another interactive composing instrument. Completed in 1972, the *SalMar Construction* (SMC) featured a synthesis control system with almost 300 switches that could control four concurrently running computer programs, all of which were producing its own composition layer. The system allowed the human performer to move between several levels of control hierarchy, "from the micro-structure of individual timbres to the macro-structure of an entire composition" (Walker et al., 1992, p. 190). Upon asked what it was like to perform with SMC, Martirano replied:

> It was too complex to analyze. But it was possible to predict what sound would result, and this caused me to lightly touch or slam a switch as if this had an effect. Control was an illusion. But I was in the loop. I was trading swaps with the logic. I enabled paths. Or better, I steered. It was like driving a bus. (Chadabe, 1997, p. 291)

Later in the 1970s, one of the most influential scenes related to electronic music performance emerged around Mills College in Oakland, California. In particular, a group that called themselves the League of Automatic Music Composers began using newly acquired KIM-l microcomputers—the first commercially available personal computers—to experiment with improvised sounds and interactive algorithmic composition. Jim Horton, John Bischoff and Tim Perkins were the group's core members, although it often featured several others (C. Brown & Bischoff, 2002). Each composer programmed their own computer so that it could produce music by itself, but the programs also accepted the output of another group member's computer as input (Dean, 2003). A chain of one computer's output affecting the performance of another resulted in networked music performances, with a circular data flow between the composers. The group put on regular performances at Mills College, often casting the computer in the role of independent composer-performer (Lewis, 2017). Performances amounted to "letting the network play", and the humans performing became a part of the network as well (Born et al., 2017).

The League of Automatic Music Composers would have a significant impact on George Lewis, at the time an aspiring young jazz trombonist, later to become a pioneer himself in the domain of human–computer improvisation.

> [The League] sounded like a band of improvising musicians. You could hear the communication between the machines as they would start, stop, and change musical direction. Each program had its own way of playing. I hadn't heard much computer music at the time, but every piece I had heard was either for tape or for tape and people, and of course none of them sounded anything like this. I felt like playing, too, to see whether I could understand what these machines were saying. (Roads, 1985, p. 75)

Lewis' encounter with the League composers inspired him to acquire his own computer and learn programming for himself. In contrast to the League, however, his primary interest was to make a program that could improvise together with himself as an instrumental performer (Roads, 1985).

The lack of a universal communication protocol to describe musical events meant that computer music programs were designed to work with specific hardware. With each computer music system being unique to the equipment where the software was developed, their impact was geographically limited. This began to change in 1983 when some of the world's largest manufacturers of electronic music instruments agreed to adopt the MIDI protocol (IMA, 1983). Working at Bell Laboratories under the sponsorship of aforementioned Max Mathews, Laurie Spiegel utilized the advantages of MIDI to program the *Music Mouse*—one of the first interactive music systems for general use—in 1986. Best described as a system to control a musical automaton (Gifford et al., 2018), the program featured embedded knowledge of scales and chords, including "all of Bach's favorite manipulations—retrograde, inversion, augmentation, diminution, transposition" (Cope, 1991). Activated notes were harmonized and stylistically transformed through the selection of preferred modes from a computer window. By these means the user could control a four-part MIDI stream that played a synthesizer with the aid of switches, knobs, pushbuttons and keys (Dean, 2003).

At the same time as Spiegel was working with developing her software, Joel Chadabe started the company *Intelligent Music*, with the mission of developing an interactive instrument for the home entertainment market (Chadabe, 1997). Although Chadabe later admitted that Intelligent Music was founded with "abundant vision and no capital" (Chadabe, 1997, p. 315), the company did achieve moderate success when David Zicarelli, a promising young programmer of music software, joined the company. In particular, the two programs *M* and *Jam Factory*—both commercially released in 1986—made somewhat of an impact. M and Jam Factory could be run on Apple Macintosh computers and provided a platform for composing and performing with MIDI using graphic and gestural interfaces (Zicarelli, 1987). Based on Chadabe's conceptualization of controlling an interactive instrument like "flying a plane through a musical space" (Chadabe, 1997, p. 316), the program was loosely modelled on a flight simulator where the user had access to a range of algorithms in the form of graphical control objects. The user could record sequences with a MIDI keyboard and perform various transformations to the recorded material while it was being played back. Jam Factory consisted of four Players that independently transformed recorded sequences using transition tables, or Markov models. Zicarelli's motivation behind Jam Factory was to have the program "improvise" with the recorded input, while allowing the user to add notes or adjust the parameters of the algorithmic process while listening to the results in real-time (Zicarelli, 1987).

2.1.2 From instruments to agents

The interactive music systems described so far belong to the instrument paradigm in Rowe's classification system (see Chapter 1.1.3). They were like intelligent instruments that produced formal musical structures in real-time, controlled by the user. Some of the first accompanying systems that musicians could *play together with* as duo partners came with Peter Beyls' *Oscar*, Robert Rowe's *Cypher*, and George Lewis' *Voyager*. All three emerged independently in 1988 on the back of years of development and experimentation.

Using the computer as a vehicle for what he called "conceptual navigation" for the exploration of ideas, Beyls developed the software Oscar (*Osc*illating *Ar*tist) as a musical companion in live performance (Chadabe, 1997). Describing the software in anthropomorphic terms, Beyls explains that Oscar's personality "exhibits an inclination to impose his own individual character while at the same time expressing a wish for 'social' integration into a whole" (Beyls, 1988, p. 223).

Rowe's Cypher was a multi-agential interactive system that could make sense of what it heard and create an interesting response (Chadabe, 1997). Cypher had two major components: a listener and a player. The listener listened to MIDI streams and identified the beginnings and ends of phrases by noting discontinuities or sudden changes in register, loudness and density. By classifying the behavior of the current phrase and comparing it with the overall musical context, the listener mapped the combined features of the input to various response methods in the player. The player used a range of different algorithmic techniques to produce new musical output (Rowe, 1992).

As mentioned previously, Lewis (2000) was inspired by the aesthetics of The League of Automatic Composers to create his own computer system, which resulted in Voyager:

> A computer program analyzes aspects of a human improviser's performance in real time, using that analysis to guide an automatic composition (or, if you will, improvisation) program that generates both complex responses to the musician's playing and independent behavior that arises from its own internal processes. (p. 33)

Voyager combined stochastic selection methods and musical constraints to create an interactive dialog between musician and machine (Gifford et al., 2018). The program integrated Lewis' framework of African-American cultural practice and embodied the aesthetics of *multidominance*, which runs counter to the Western music philosophy of avoiding "too many notes" (Lewis, 2000). Hence, Voyager is more like a "virtual improvising orchestra" consisting of many agents who combined constitute the computer musical improviser (Gifford et al., 2018).

The level of autonomy these three systems exhibited made them stand out as something new compared with their predecessors. They were *musical agents*—artificial agents that tackle musical creative tasks, partially or completely (Tatar & Pasquier, 2018). Although I will continue to use the term interactive music systems, it should be noted that my focus moving forward is on systems that qualify as musical agents. From the 1990s and onward, the appearance of interactive music systems displaying autonomous or semi-autonomous behavior has proliferated. Tatar and Pasquier (2018) identified close to a hundred systems that have been presented in peer-reviewed publications. It is unfeasible to present the full scope of this work in this chapter. To set the historical context for this thesis, I find it sufficient to describe briefly a few systems that represent the development in the domain of improvisation and collaborative performance. Chapter 8 will also feature some relevant systems in more detail.

The aforementioned early systems were mostly rule-based and specifically adapted to the aesthetics of the developer-practitioner (usually the same person). As systems began taking advantage of learned models instead of predefined rules, specificity gave way to generality. *GenJam* (Biles, 1994) used genetic algorithms to "breed" stylistically appropriate jazz solos to be played over predetermined sections of jazz standards. *The Reactive Accompanist* (Bryson, 1995) provided chord accompaniment of unfamiliar melodies using a subsumption architecture methodology. With *Band-out-of-the-Box* (BoB) (Thom, 2000), the human user could trade four-bar solos in the style of blues/jazz with the machine agent. The agent utilized unsupervised machine learning techniques to adapt to the musical sense of its user. *The Continuator* (Pachet, 2003) produced musical continuations to phrases introduced by users with the help of Markov models, allowing for a stylistically coherent back-and-forth interaction. *OMax* (Assayag et al., 2006) pioneered the use of the *Factor Oracle* for music purposes. Both Markov models and the Factor Oracle will be covered in more detail later in this chapter.

This historic overview represents only a fraction of interactive music systems developed to date. Overlapping terminologies can make researching the field confusing. Other partially overlapping terms are *musical agents* (Tatar & Pasquier, 2018), *virtual improvisation partners* (Nika & Chemillier, 2012), *interactive music improvisers* (Blackwell & Bentley, 2002), *live algorithms* (Blackwell et al., 2012), *improvisational music systems* (Gifford et al., 2018), and *music response systems* (Ravikumar & Wyse, 2019), to name a few.

2.2 Human-Computer Interaction

Human-computer interaction (HCI) is not a unified disciplinary field in the traditional sense. Rather, it is a collective, multifaceted scientific endeavor to understand and contextualize emerging interactive technologies, with contributions from many disparate fields. Music has a relatively small place in this universe. However, a brief historical overview of HCI provides a necessary context, because general developments in HCI considerably influence the field of music technology. Many of the trends from HCI have visited pursuits in music interaction design, including research related to interactive music systems. After the historical overview, I will present some concepts from HCI that have an enduring influence on the music technology field.

2.2.1 The three waves of HCI

HCI emerged from diverse roots in computer science, human factors, ergonomics, and industrial engineering as an applied science that used theories and methods from cognitive psychology for software development, with the primary goal of making systems and devices efficient and easy to use (Rogers, 2012). The "first wave of HCI" characterized by cognitive models reached its heyday in the late 1980s and early 1990s. A "second wave" of HCI happened as researchers in the field realized that the scope of cognitive psychology was too narrow and often failed to address real world contexts (Barnard, 1991). In addition to alternative cognitive models such as *distributed cognition* (Hutchins, 1995) and *ecological psychology* (Gibson, 1979), the second wave of HCI was characterized in particular by a "turn to the social": "Sociologists, anthropologists and others in the social sciences joined HCI, bringing with them new frameworks, theories and ideas about technology use and system design" (Rogers, 2012, p. 32). One of the most influential frameworks for studying interaction design in social contexts during this period was *situated action* (Suchman, 1987), which called for a perspective where the manner in which people relate to computational devices was deemed integral to the entire workplace environments where artifacts are used. A social orientation also took place with the adoption of *activity theory* into HCI (Bødker, 1989; Kaptelinin & Nardi, 2006). Furthermore, methodological approaches with "theory-free" outsets gained traction. For instance, ethnography and grounded theory played an integral role in the development of new HCI concepts and theoretical structures (Rogers, 2012, p. 6).

HCI diversified even further with a "third wave" in the early 2000s. New perspectives from feminism, multiculturalism and globalization were brought in to question current ideals, methods and practices in HCI. Ethical and political considerations pertaining to design gained influence. "In the wild" studies turned the tables on research focus; instead of designing products as solutions for existing practices, researchers began looking for possibilities for technology to enhance or even disrupt behavior (Rogers, 2012). A "turn to embodiment" saw an inroad for researchers arguing that technology and practice are inseparable, and that they are "coextensive and will coevolve" (Dourish, 2001, p. 204). In short, HCI today is an eclectic, interdisciplinary field. This is evident by the sheer number of specialized conferences under the HCI umbrella, with dozens of large annual international events and hundreds (possibly thousands) of peer-reviewed publications every year.

2.2.2 User experience

In first-wave and second-wave HCI, user tests would primarily focus on the concept of usability. The majority of researchers viewed the computer as a tool for enhancing productivity, and evaluated computer interfaces with quantifiable metrics such as task completion and error rates. Although such metrics are still relevant within certain scopes of interaction studies, they are largely irrelevant to creative activities (E. A. Carroll & Latulipe, 2012). In fact, productivity metrics could be directly misleading when evaluating technology for creative use, because people who experience periods of high creativity often describe losing track of time completely and can spend large amounts of time focused on one task (Csikszentmihalyi, 1996). Thus, while a metric showing a long completion time could indicate a problem with the tool, it may just as likely be explained by a user's deep engagement with it (E. A. Carroll & Latulipe, 2012)—a desirable outcome for both the designer and the user of the product.

With third-wave HCI, there was a shift in the focus of HCI from task-oriented towards experience-oriented (Wu, 2018). For instance, Hassenzahl et al. (2000) introduced the notion of *hedonic quality* (e.g. novelty, originality) as an additional dimension to *ergonomic quality* (e.g. simplicity, controllability). Instead of a singular focus on usability, HCI began paying attention to *user experience* (UX), and this ushered in a range of new—often qualitative—methodologies. A review of evaluation methods used to collect data regarding the user experience of interactive products has been provided by Bargas-Avila and Hornbæk (2011).

In his PhD thesis, Swift (2012) offers an account of how a UX perspective is beneficial in the context of evaluating interactive music systems:

> In a word processor, the ultimate goal of the user is the production of a high-quality document. […] However, in an improvisational computer-music environment, for instance, the goal of the participant(s) is to have an experience: of flow, connection, or "groove". The musical output of the system is merely a means to that end. In these two different contexts the role of the created artefact and the experience of making

> it are reversed. The highest goal of an interface or an interaction context is no longer to be functional, it must be a joy to use. These two outcomes are not independent (indeed they are strongly correlated in many cases) but the fundamental shift is in their prioritisation: traditionally usability and task performance were the ultimate goal, with experience being used as a "hygiene factor" (Hassenzahl et al., 2010), whereas the third wave assertion is this binary has been inverted. UX is the ultimate goal, and usability is a proximate one. (Swift, 2012, p. 19)

It is important to note that a shift towards experience does not imply that usability studies are no longer fashionable. UX offers HCI an additional evaluative dimension, but usability requirements are still relevant—also for music tools. For instance, audio processing tools within digital workstation environments need to be efficient, reliable, and well adapted to the overall user workflow. Notwithstanding, the scales are undoubtedly tipped towards evaluating UX in recent literature pertaining to interactive music systems.

2.2.3 Iterative design

One of the most established design frameworks within HCI is the cyclic process of planning, prototyping, testing, and evaluating the product in several rounds or iterations (Hewett et al., 1992; Nielsen, 1993). This is referred to as an *iterative design approach*, where each iteration refers to the complete design-test-evaluation cycle. At the end of each iteration, the designer takes what is learned from the evaluation stage and incorporates this knowledge in the planning stage of a new iteration. The number of iterations is usually predetermined in a project outline, but the outcomes are open-ended to a degree deemed permissible by the project owner. This allows the designer to generate different ideas for each iteration and successively improve on them (Beaudouin-Lafon & Mackay, 2003, p. 1010). One of the advantages of iterative design approaches is the option to make prototypes that are not fully functional artifacts, but which are focused on testing specific aspects of an envisioned artifact. A prototype in the first iteration can be cheap and quick to design, i.e. cardboard replicas, mockups, or video demonstrations (Beaudouin-Lafon & Mackay, 2003). For large-scale projects, addressing fundamental questions from a user-perspective at early design stages may save both time and expenses compared with going through a large development stage before knowing anything about the product's usability or user experience.

A considerable portion of the literature on interactive music systems describes iterative approaches to artifact design. Yet other literature describes one iteration where the final product is referred to as "a work in progress" or where the concluding chapter dedicates a section about "future work". This demonstrates a philosophy of iterative design that is integral to the field, even for projects where an iterative approach is not explicitly specified. The design thread of this thesis should be considered in the same way—as the first iterations of a longer-term project.

2.2.4 Affordances and constraints

The notion of *affordance* was originally coined by Gibson (1979) and is rooted in *ecological psychology*. An affordance, in the Gibsonian sense, is the potential action capabilities that things in the environment offers an actor. For example, a horizontal, flat, extended, and rigid surface may *afford support*. According to Gibson, affordances are invariant—they exist whether the actor perceives the affordance or not. Thus, an affordance in its original meaning is a property of the action *capabilities*—not the *experience*—of the actor. Affordance is widely adopted in HCI in a slightly different sense, which often includes the experience of the actor. An influential understanding of affordance is attributed to Norman (1988), who made a distinction between "the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used" (p. 9). According to this view, affordances can be both real and perceived. Norman's stance is summed up by Rogers (2012):

> On the one hand, physical objects were considered to have real affordances, […] which are perceptually obvious and do not have to be learned. User interfaces that are screen-based, on the other hand, do not have these kinds of real affordances. Importantly, this means that users have to learn the meaning and function of each object represented at the interface before knowing how to act. Norman argued that screen-based interfaces have perceived affordances, which are based on learned conventions and feedback. (p. 43)

Gaver (1991) argued that the notion of perceptible affordances implies that some affordances may be *hidden* or even *false*. Design errors may arise when functions intended by the designer are undiscovered by the user (hidden in a negative sense), or when apparent affordances perceived by the user turn out to be incorrect assumptions about functionality (false, again in a negative sense). From the perspective of creative exploration in digital musical instruments or interactive music systems, the notion of hidden affordances has taken on a positive meaning, usually referring to the discovery of surprising possibilities inherent in the design. From this vantage point, Magnusson (2009) points to the scarcity of hidden affordances in digital tools as a problem rooted in the fact that the computer only offers knowledge inscribed by humans:

> It does not contain much hidden affordances or mysteries. The computer mediates: in itself it is never the object of engagement as that object is virtual. The digital

> instrument maker therefore gets nothing for free, unlike the maker of the acoustic instrument who receives the gift of sonic timbre from the physical properties of the materials he works with. Instead of presenting, the computer represents. (p. 154)

In his PhD thesis, Martin (2016) expresses the discovery of hidden affordances outside the design intensions of his digital instruments as a desirable outcome. Mudd (2017) describes the discovery of hidden affordances not intended by design as a key to developing a personal style with a given instrument. In complex interactive music systems including machine learning models, it is reasonable to assume that hidden affordances are prevalent due to emergent effects provided by the models, and that documenting the discovery of these should be an integral part of the user evaluation.

It is worth pointing out that affordance carries yet a different meaning in musicology, where music is claimed to provide affordances for actors to socially organize and make sense of the world (DeNora, 2000). This sociological take on affordance aligns with the HCI stance that affordance is rooted in the experienced world of the actor, but differs in the sense that it pertains to social constructions as opposed to the properties of designed objects.

In HCI, the notion of *constraints* is often invoked in complement to affordances. The two terms have the appearance of being opposites: Affordances suggest possibilities, and constraints limit them (Norman, 1988). An intuitive assumption to make, therefore, is that less constraints leads to more affordances. In actual practice, this is not usually the case. From a design perspective, the thoughtful use of constraints makes it easier for the user to discover affordances: "Constraints map out a territory of structural possibilities which can then be explored" (Boden, 1990, p. 95). Anyone who has experienced writer's block can testify that the blank page's most menacing feature is that options are limitless. Possibilities crystalize as text starts filling the page. The emerging structure imposes limits, but seems to propel creativity. This dynamic is widely recognized in HCI. Interaction design generally focuses on keeping interfaces as complex as necessary, but as simple as possible. According to Norman (1999), constraints can be *physical*, *cultural*, or *logical*. Physical constraints prevent users from performing certain things in the environment—real or virtual—such as plugging in contacts upside down or moving the mouse cursor outside the screen. Cultural constraints are ones imposed by cultural norms, such as a red light signifying "stop" and a green light signifying "go". Finally, logical constraints are conceptual guides to behavior, such as mapping buttons in the same order and layout as corresponding burners on a kitchen stove.

Undoubtedly, appropriate constraints are key to guiding users' attention toward affordances intended by the designer. However, technology designed for creative tasks differ from many other kinds of software in that users may find pleasure in discovering effects not foreseen during development. In interactive music systems, the ability for users to explore and experiment at length could be equally as important as having an interface that is easy to navigate. These demands are not mutually exclusive. It is possible to design an interface that is cognitively undemanding while simultaneously generating complex results. This is the theme for the next section.

2.2.5 Mapping

Many operations in computer software have clear relationships between the user's actions and the resulting effects. When clicking a button, something happens. When dragging a scroll bar, the page moves. The relationship between the action and the effect is the *mapping*. In the examples above, it makes sense to have an explicit, singular mapping from one action to one effect. This is a so-called *one-to-one mapping* (Hunt & Kirk, 2000). An early popular electronic music instrument—the *Theremin*—is an example of the use of one-to-one mapping strategies for instrument control. The proximity of the right hand to the rightmost antennae controls the pitch, and the proximity of the left hand to the leftmost antennae controls the amplitude. Even the combination of these two one-to-one-mappings becomes a surprisingly complex task when performed simultaneously. Through practice, Theremin players are able to internalize the operation of these two controls and commit it to muscle memory. However, there is a limit to how many trivial mappings humans are able to control at the same time.

In real environments, complex mappings are the norm (Hunt & Kirk, 2000, p. 234). To the point, music is a rich multi-dimensional signal, and most musicians are not used to attending to all of these dimensions separately—they perform and listen holistically as opposed to analytically. Hunt and Kirk (2000) uses the violin as an example. For instance, the sound volume can be controlled by any combination of bow speed, bow pressure, choice of string, and finger position. An equivalent relationship in the software domain would be a *many-to-one mapping*. Conversely, the bow controls not only volume, but also timbre (sound character), articulation, and pitch (as an artifact of bow pressure). The design of such a relationship would amount to a *one-to-many mapping*. From a holistic perspective, a violin player continuously applies many-to-one and one-to-many strategies simultaneously, where variables are cross-coupled in dynamic and complex ways. An attempt to approximate such intricate cross-couplings would entail designing a *many-to-many mapping* system (Hunt et al., 2002).

Jensenius (2007) points out that in acoustic instruments, action-sound relationships are *coupled*—meaning that they have a natural and inextricable relationship as opposed to being deliberately and independently designed. Because of our life-long experience with natural action-sound couplings, we are used to them—we do not spend energy understanding the causality. New, artificial relationships between action and sound stand out—they grab our attention. It means that designed mappings between action and sound run the risk of being experienced as arbitrary or even counter-intuitive. The relationships may also be poorly designed. A poor mapping strategy, for instance, may place a high cognitive burden on the user to generate relatively simple sounds. Conversely, a clever mapping strategy may present a relatively simple interface that has the potential of creating interesting, nuanced, and complex results.

Miranda and Wanderley (2006) describe two main directions in mapping. *Explicit mapping strategies* are concretely defined. This is often referred to as a *rule-based* approach. Early interactive music systems such as Lewis' Voyager or Rowe's Cypher are examples of systems that rely predominantly on explicit mapping strategies. *Model-based mapping strategies*, on the other hand, use machine learning techniques to create input-output relationships that are complex, yet manageable from a user's perspective. Compared with explicit strategies, computer programs required to create and run models can be relatively small. For example, Fiebrink's (2011) *Wekinator* system allows the user to train models that map input values (such as the position of a slider) to desired output sounds by providing examples of action-sound relationships. Through regression—a supervised machine learning technique—the resulting trained model can replicate not only these relationships, but also generate "in-between sounds" along the entire slider range. As these "in-between sounds" consist of (potentially high-dimensional) interpolations between the example sounds, the exploration of the resulting sonic range driven by this one parameter (the slider) can yield surprisingly complex results. Furthermore, because the machine learning model is a "black box" (i.e. the inner workings of it are not known), the model-based mapping is often more interesting from a creative standpoint.

There is an interesting link between the kind of dynamics afforded by model-based systems and *cybernetic* approaches to music performance. Wiener (1948) characterized cybernetics as concerned with "control and communication in the animal and in the machine". This is one of a large number of definitions—it does not by far capture the full essence of cybernetics and how it is applied as a framework within various academic disciplines. Beer (1959) defined cybernetics as "the art of effective organization", and made a distinction between systems that are knowable and predictable from a scientific perspective, and exceedingly complex systems, which are not. Pickering (2010) points out that the aspect of cybernetics that appeals to him is that it assumes an "ontology of unknowability" and "tries to address the [problem] of getting along performatively with systems that can always surprise us" (p. 23). Eno (2004) adequately expresses this sensibility in the context of experimental music performance by quoting the above mentioned cybernetician Stafford Beer: "Instead of trying to specify it in full detail, you specify it only somewhat. You then ride on the dynamics of the system in the direction you want to go." (Eno, 2004, p. 230) Chapter 3 is dedicated to exploring such themes in further detail.

In the kind of interactive music systems that are the focus of this thesis, with complex relations between inputs and outputs, explicit mapping could be a daunting task. It seems quite clear that model-based mapping strategies—where one can leave it to the machine to find efficient cross-coupling between the input and output parameters—may often be preferable to explicit mapping strategies for interactive music systems designed for co-performing with musicians.

2.3 Machine learning

Machine learning (ML) is a range of computer-based statistical analysis methods that achieve tasks by learning from examples (Caramiaux & Tanaka, 2013). ML is considered by some to be a subfield of *artificial intelligence* (AI), but in popular culture, the two terms tend to be used interchangeably. In order to avoid any confusion, it is appropriate to make the distinction clear. AI is concerned with using computers to build entities with apparently intelligent behaviors or decision-making capabilities (Russell & Norvig, 2010). Because learning and adaptability are capacities deemed integral to the appearance of intelligence, a large share of AI artifacts today have ML at the core of their architecture. Historically, this has not always been the case. Although computational *neural networks*—a type of machine learning—were invented as early as in the 1950s, they proved to be impractical due to a lack of sufficient computer power to run anything but simple models (Russell & Norvig, 2010). For several decades, rule-based systems based on explicit computer code dominated the AI field. ML made a comeback as a field in its own right in the mid-1980s, as a "natural outgrowth of the intersection of Computer Science and Statistics" (Mitchell, 2006, p. 1).

Some of the areas where ML has a legacy of success are *speech recognition*, *computer vision*, *bio-surveillance*, *robot control*, and *accelerating empirical sciences* (Mitchell, 2006). In general, ML methods are particularly apt in applications where the tasks are too complex for people to manually design the algorithm, or where there is a need for the software to be able to adapt to its operational environment after it is fielded (Mitchell, 2006, p. 3). In the following sections, I will introduce some of the most common categories of ML methods, and provide examples of how they have been applied in interactive music systems.

2.3.1 Supervised machine learning

The group of ML methods that fall into the category of *supervised machine learning* uses *training datasets* consisting of inputs paired with their corresponding outputs. The computer program essentially runs through multiple examples of "if input is x then output is y". Given sufficient examples, the algorithm produces a *model* that subsequently reliably can predict useful output values from new inputs. The process of creating the model is the *training*. The input can consist of a single *feature* (a one-dimensional input) or a set of several features (a multidimensional input, often referred to as a *feature vector*). All of the features need to be numbers. If a feature is originally a text string (e.g. the name of a country), it needs to be encoded as a number (e.g. 0=Andorra, 1=Belgium, etc.). The output column in the training dataset is usually referred to as *labels*, or sometimes *targets*.

Supervised learning is used to solve two kinds of problems. *Classification* problems require models that predict outputs within a finite range of discrete values. *Regression* problems are used to predict outputs along a range of continuous values. An example of a classification model is one that can predict whether an incoming email is spam or not. An example of a regression model is one that can predict the profit of a company given a feature vector containing data about investment levels in various departments of the company and the company's geographical location. There are a range of different ML methods to tackle both classification and regression problems. To name a few, classification can be performed with *k-nearest neighbors* (KNN), *support vector machine* (SVM), *Naïve Bayes*, *decision tree classification*, and *random forest classification*. Regression problems can be addressed using *linear regression*, *polynomial regression*, *support vector regression*, *decision tree regression*, *random forest regression*, or other. Explaining all of these methods is outside the scope of this chapter. A comprehensive introduction to a wide variety of ML methods is provided by Bishop (2006).

An important part of becoming an ML expert is obtaining the experience of knowing which methods to use in various contexts and for different kinds of datasets. Some of the methods are good at generalizing from data, but may perform less well at predicting *outliers*—output values that appear outside the general trends in the dataset. Other methods may be better at capturing outliers at the expense of poorer generalization. There is always a risk that the training process could lead to *overfitting*, which happens if the model performs very well with input values from the existing training set, but turns out to perform significantly worse with new input vectors. In such cases, the training has caused the model to learn the training data too well, but failed to extrapolate the means to predict well on unseen data.

The ML methods mentioned so far are sufficient for many kinds of datasets. However, for very large feature vectors or otherwise complex datasets with highly nonlinear relationships between the inputs and outputs, it is common to train models using *artificial neural networks* (ANNs). ANNs are modelled on the function of interconnected neurons in the human brain. The algorithmic equivalent of neurons are computational nodes that take the weighted sum of its inputs and produces an output based on an *activation function*. It is common to use nonlinear activation functions, which is useful for seeking out patterns in complex data. In Figure 1, the nonlinearity of the activation function is featured with the S-shaped *sigmoid* curve.

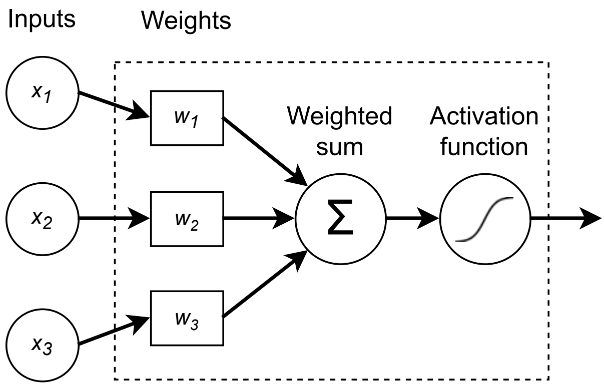


Figure 1. The inside of the rectangle shows the kind of computation taking place within each node in the ANN.

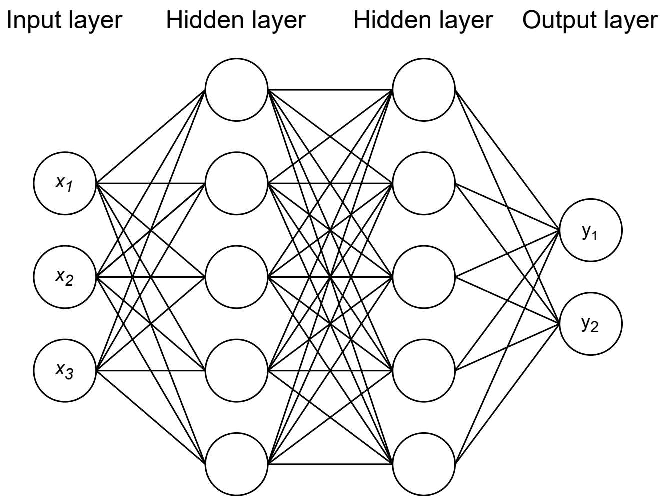


Figure 2. ANN with interconnected nodes.

ANNs are usually arranged in a way that values from the inputs are propagated through one or several *hidden layers* of nodes before arriving at an output layer (Figure 2). The term *deep learning* is used to characterize ANNs that have multiple hidden layers. Training an ANN consists in repeatedly calculating the outputs from the inputs, comparing the results with the correct labels in the training dataset, and using the error rate—a measure of the discrepancy between the label and the value derived from the model—to adjust the input weights for each node. The adjustment of the weights is called *backpropagation*, because the weights are fine-tuned in reverse order of the directionality of the network (Figure 3). Each run of this entire process is called an *epoch*. When the error rate (the loss) is acceptably low, the training is considered successful.

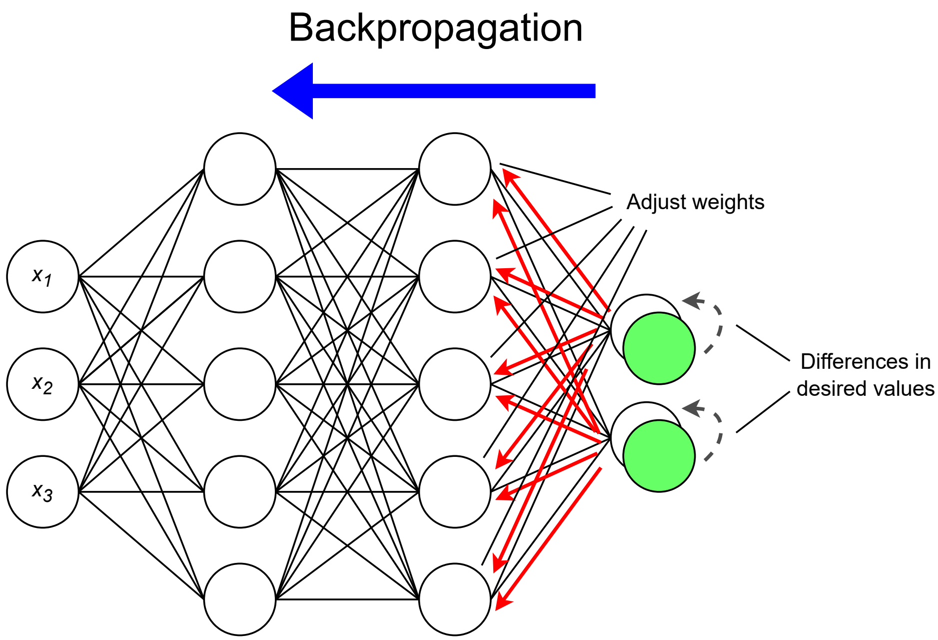


Figure 3. The weights for each node are adjusted in reverse order of the directionality of the network.

ANNs are commonly associated with complex classification problems, but they can be used for regression problems as well. One issue with conventional ANNs is that they do not take into account temporal aspects of the input (Caramiaux & Tanaka, 2013, p. 4). For this reason, they are not well suited for making inferences about sequential data where each new state is dependent on the previous state. Obviously, music falls into this category. Extensions to ANNs that attempt to remedy this include the *recurrent reural retwork* (RNN), which introduces short-term memory. *Long Short-Term Memory* (LSTM) networks are a type of RNN that include "memory cells", which make them capable of maintaining memory over longer periods of time. More recently, *Transformer models* have demonstrated a unique capacity to hold important information over long periods through a so-called attention-mechanism (Vaswani et al., 2017).

Supervised ML has many useful applications in music. Model-based approaches to dealing with mapping problems have already been mentioned in the previous section. Supervised methods are also important for recognition problems such as distinguishing between different sound classes. The interactive music system *NN Music* (Young, 2008) uses two supervised neural networks in series to model an improvising co-performer. The first network takes the audio input of the human performer and classifies it according to a set of predefined statistical representations of audio features. The second network takes the output of the first network and maps them to synthesis parameters. *Zamyatin* (Bown, 2011) is an example of a system where supervised ML is used as part of a decision-making process, but where the system as a whole appears to operate in an unprescribed manner. To achieve this level of perceived autonomy, a decision tree model is trained to activate appropriate *behavioral objects*, which operate within a complex system network.

Although supervised ML methods are widely used as tools in music technology, they are not always optimal for systems focused on exploratory, open-ended co-performance with human musicians. One limitation is that putting together a good quality training set is time-consuming. Furthermore, many aesthetical choices are made in the process of curating the dataset and defining how features and labels should relate to each other. Once trained, the models will tend to behave in quite predictable ways. Thus, models based on supervised ML tend to reflect the user's sensibilities and may lack surprising or contrasting responses to the user's input. Such factors may explain the prevalence of other approaches to designing interactive music systems—particularly ones geared toward improvisation.

2.3.2 Unsupervised machine learning

Algorithms associated with the *unsupervised machine learning* approach do not need any predefined feedback mechanism. The goal is unknown, so there are no labels or target values corresponding to the input vectors. Instead, an unsupervised algorithm uncovers internal patterns in the input data. The most common type of unsupervised ML is *clustering*, where the input data is grouped into regions called clusters based on feature similarity. For example, a clustering algorithm given a dataset containing samples of different instruments may successfully group most or all of the samples in separate clusters according to the instrument type. Figure 4 shows an example of clustering done on a dataset containing two-dimensional feature vectors, which are easy to visualize. Multidimensional vectors are more difficult to visualize, and for this reason, it is common to compress multidimensional data into two or three dimensions before clustering if the goal is to visualize the clusters. This preprocessing is called *dimensionality reduction*. A common method for dimensionality reduction is *principle component analysis* (PCA), which in itself is a form of unsupervised learning.

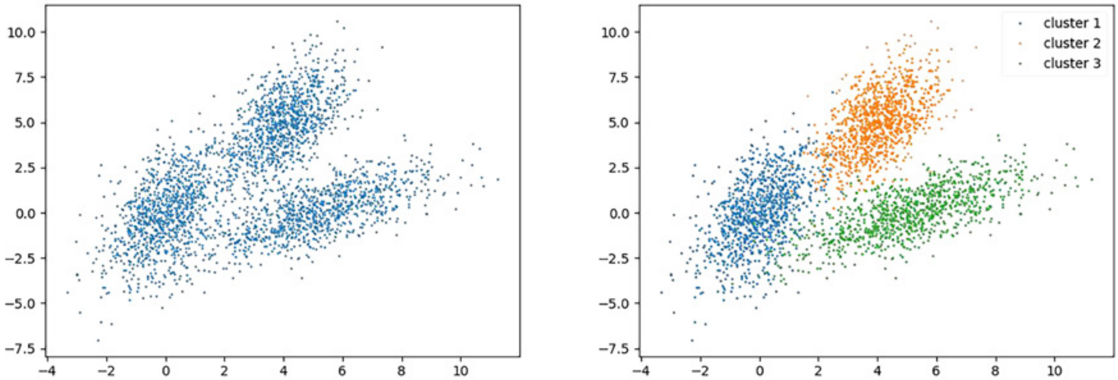


Figure 4. An example of a dataset before and after clustering.

Clustering can be a useful way to make musical sense of input data without making prior assumptions about it:

> A musician might employ unsupervised learning simply to discover structure within the training set, for example to identify latent clusters of perceptually similar sound samples within a large sample database, or to identify common chord progressions within a database of musical scores. A musician might employ this learned structure to generate new examples similar to those in the database. Or, she might use this learned structure to provide better feature representations for further supervised learning or other processing. (Fiebrink & Caramiaux, 2018, p. 5)

Another class of unsupervised ML is the *self-organizing map* (SOM), which is a type of ANN that utilizes unsupervised learning to map high-dimensional feature vectors onto a two-dimensional topological grid (Kohonen, 1990). Given a set of n-dimensional feature vectors, the learning algorithm organizes these vectors such that the resulting two-dimensional feature space is qualitatively aligned with the input. Each coordinate in the SOM, called a node, is a feature vector that represents approximations of a varying number of input vectors.

Although the SOM is a compressed representation of the input data, the data itself is left as is. Through a *best matching unit* (BMU) function, each feature vector in the original dataset is associated with the node that is most similar to it. Thus, a SOM node will tend to contain perceptually similar vectors from the input space. If one thinks of the SOM as a kind of self-organizing library, the SOM nodes are analogous to shelf numbers containing the original data. The input vectors are unevenly distributed across the SOM, so some nodes may be empty (see Figure 5).

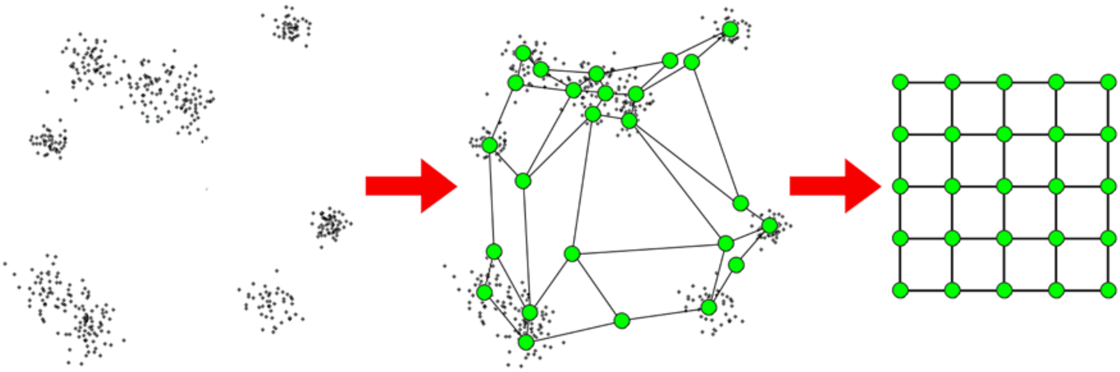


Figure 5. A self-organizing map is essentially a 2D grid "stretched across" and adapted to the input data.

A final category of unsupervised ML worthy of mention is the *autoencoder*, which is yet another type of ANN. The goal of an autoencoder is to compress high-dimensional data input into a low-dimensional *latent space* by throwing away insignificant data (redundant information or noise). The encoding is validated through iterative attempts at reconstructing the input from the encoded representation. This validation involves comparing the decoded output with the original input. Thus, the input feature vector functions as labels in much the same way as in a supervised ANN—the difference being that the original input is the desired output. The symmetrical encoder-decoder parts give autoencoder networks a characteristic bowtie shape where the middle (the "bottleneck") is the encoded latent space (Figure 6).

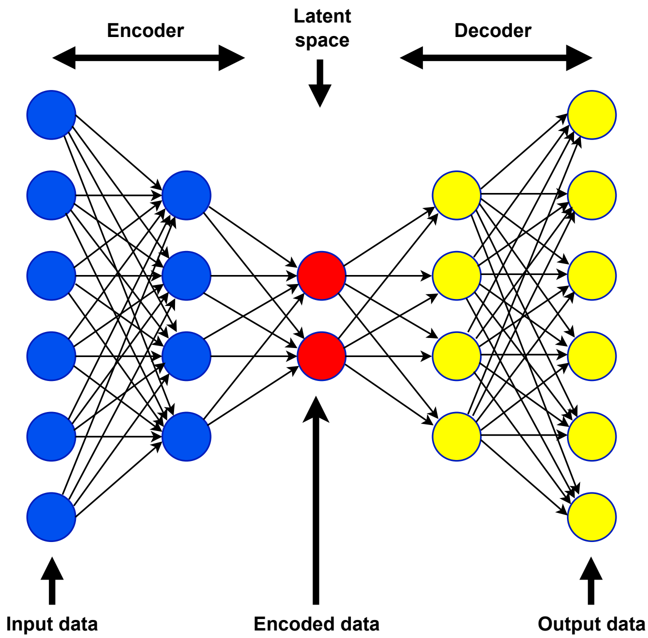


Figure 6. An autoencoder with six inputs and outputs and a two-dimensional latent space.

A growing number of interactive music systems rely on unsupervised ML to perform computational decision-making based on current audio input. *ListeningLearning* (Collins, 2011) uses a pre-trained clustering model to discern between ten different timbral states in the input. Each timbral state activates its own dedicated response agent, which takes on certain synthesis and processing resources. Smith and Deal (2014) describe a co-improvising system that trains a SOM in real-time. Based on how close a new input is to an existing SOM node, a decision module triggers a musical response. If the input is close to an existing node, it is taken as being predictable, and the agent will seek more complexity and increase its divergence from the input. If the input is further away from an existing node, the agent will steer toward unison with the human, giving the initiative back to the musician. Bretan et al. (2017) developed a system that treats a measure of performed music as a fundamental unit. A deep autoencoder encodes each performed measure into a latent space, and outputs the closest melody from a library that is embedded into the same latent space. Thus, the system engages in a call-and-response type behavior when in improvisation mode.

Smith and Garnett (2012) argue that the self-organizing aspect of unsupervised algorithms map well to models of human perception. From the perspective of mixed-initiative creative interfaces, the idea of algorithms that give rise to models with their own internal logic is appealing. The ideas provided above show that unsupervised models have the capacity to link machine listening and generative processes in ways not prescribed directly by humans, but rather guided by patterns computationally discovered in the data. This potentially gives the computational agent more agency, and can lead to exploratory forms of human-computer interaction.

2.3.3 Reinforcement learning

Reinforcement learning (RL) is a category of ML methods where the learning agent receives feedback about how suitable its actions are in a given environment. Instead of receiving examples of optimal outputs as in supervised learning, the agent discovers appropriate actions by utilizing a *policy* that maximizes a *reward* value (Bishop, 2006). The reward is accumulated through trial and error. Actions that bring the agent closer to achieving a given goal will yield an increase in the reward. Some undesired actions may receive a penalty and make the reward decrease. Typically, RL models are used in real-world environments where many variables are unknown. RL has demonstrably outperformed other ML methods in robotics, where the robots can learn how to navigate in an environment through exploration instead of relying on examples of correct behavior. Although RL is reminiscent of unsupervised ML, RL algorithms focus strictly on maximizing a reward and do not try to find hidden structure. Hence, RL is considered its own ML paradigm separate from supervised and unsupervised approaches (Sutton & Barto, 2018).

There are a few examples of interactive music systems that have architectures that rely on RL. *Improvagent* (Collins, 2008) takes MIDI input, captures time-limited frames of music, and extracts a feature vector from each frame consisting of data about the number of onsets, pitch class, key type, and a range of other higher level features. These feature vectors are treated as environment states. Using an online (real-time) clustering algorithm, Improvagent creates and continually updates a database of observed state-action pairs. The RL part of the process produces ratings for each state-action pair based on both the quality of its predictions of the next state and on the degree to which it influences the interaction. In other words, the reward is calculated computationally. Another human-machine improvisation system using RL is *Pock* (Beyls, 2018), which also uses a self-optimizing approach based on the implied motivation of the human user. Two competing parameters—*integration* and *expression*—indicate whether the user is more interested in attuning to the output of the system or attempting to play independently from it. If the musical distance (i.e. melodic similarity) is low, the integration parameter increases in opposite proportion to the expression parameter. Vice versa, if the musical distance is high, expression increases and integration decreases. The RL algorithm maximizes the reward when there is a relatively stable balance between the two parameters, indicating that there is a give and take between the human and the machine.

The use of RL in interactive music systems is a promising concept in theory. In particular, the focus on finding optimal transitions between states seems to fit the domain of music quite well. However, RL algorithms typically need a considerable quantity of feedback data to become efficient. This can be difficult to achieve in a musically sound manner. Computational methods of calculating a reward tend to oversimplify the creative problem domain. On the other hand, manual feedback in the form of approval or disapproval from the human user is also problematic for several reasons. Firstly, the data is sparse even if the user signals feedback every few seconds. Secondly, an onus on the user to analyze the quality of the co-performance is probably detrimental to the capacity for the user to get into a state of creative flow. Conversely, users may tend to forget to give feedback when they become deeply engaged with the interaction. Thirdly, this form of manual feedback is a lopsided form of interaction. Arguably, a computational agent that bends to the will of a human performer may not necessarily be creatively stimulating. For the time being, the impact of RL in interactive music systems is marginal compared to other approaches.

2.3.4 Statistical modeling of musical sequences

Complexity in music presents itself in a sequential manner. Even with grossly simplified MIDI notation, there are millions of potential melodic, harmonic, and rhythmic variations within a single measure of music (Bretan et al., 2017). The production of an indefinite number of semantic forms from a limited set of symbols is a problem domain traditionally associated with linguistic theory. In fact, the study of grammars for music has a long history and is heavily influenced by natural language modelling and speech recognition (Conklin, 2003). Long before ML made any significant impact in music, researchers in music informatics were experimenting with probabilistic models for music generation. The general principle behind *context models* is that events can be predicted from the sequence of preceding events (Assayag et al., 2006). In particular, *Markov models* have been widely popular, and variations of these are still used in many interactive music systems. In the final section of this chapter, I will briefly introduce *Markov chains*, *hidden Markov models*, and the *Factor Oracle* due to their significant impact in computer music and their prevalence in interactive music systems. These modelling approaches have in common that they are computationally cheap, and they can learn new musical phrases incrementally and generate stylistically faithful variations of these phrases within just a few milliseconds. However, their robustness is limited to short time windows. The problem of recognizing longer formal structures is demonstrably better solved with RNN-based methods such as LSTM and Transformer models (Huang et al., 2018), albeit at a steep computational cost.

A Markov chain is a model consisting of a sequence of states, where the next state is given by the transition probabilities from the current state. The transition probabilities can be depicted in a diagram (Figure 7) or in a *transition matrix* (Tables 1 and 2). Musical sequences such as notes, note lengths, and chord sequences can be modelled with Markov chains, affording efficient ways for computer programs to learn the musical style of examples (Fiebrink & Caramiaux, 2018). The *Markov property* assumes that the next state is dependent only upon the current state. This is a *first-order Markov chain*. A *multi-order Markov chain* relaxes the Markov property and allows for some memory. For example, a third-order Markov chain considers the probabilities of possible next states based on the three most recent states.

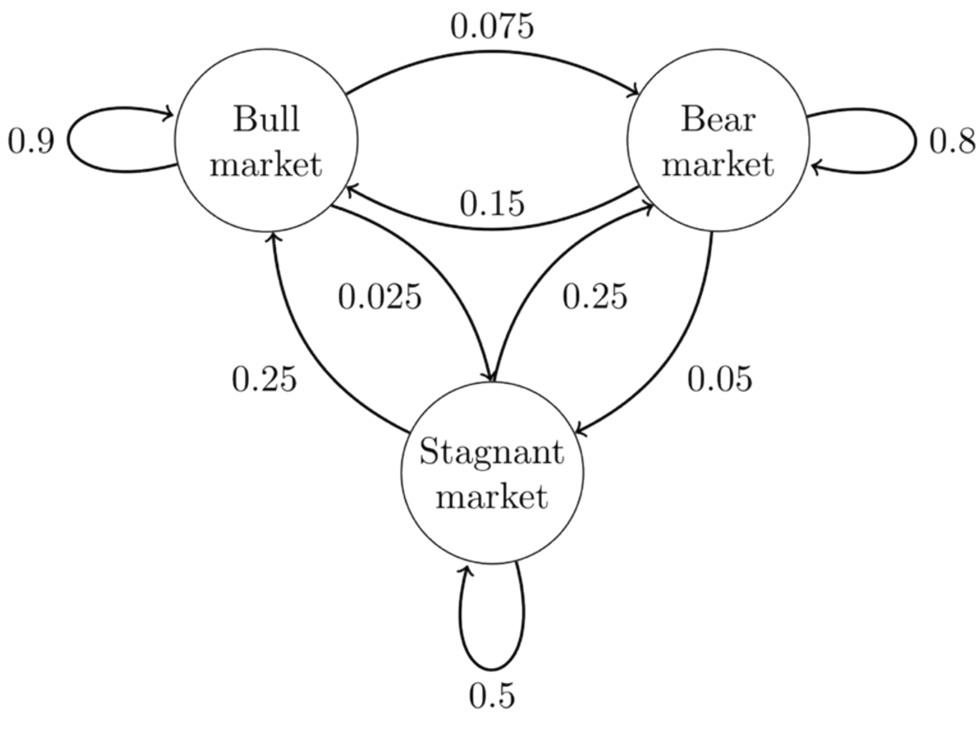


Figure 7. A Markov chain depicting the probabilities of next states in a hypothetical stock market. The model assumes states at discrete time steps (e.g. once a month). CC BY-SA 3.0 https://creativecommons.org/licenses/by-sa/3.0/.

Note	C	D	E
C	0.2	0.4	0.4
D	0.3	0.2	0.5
E	0.4	0.4	0.2

Table 1. Example transition matrix for a first-order Markov model, where each cell value denotes the probability of transitioning to each new note (column) given the last note (row).

Notes	C	F	G
CC	0.1	0.4	0.5
CF	0.2	0.3	0.5
CG	0.5	0.3	0.2
FF	0.4	0.1	0.5
FC	0.4	0.2	0.4
FG	0.5	0.3	0.2
GG	0.6	0.3	0.1
GC	0.3	0.3	0.4
GF	0.3	0.2	0.5

Table 2. Example transition matrix for a second-order Markov model. Here, the cell values denotes the probability of transitioning to a new note based on the last two notes.

Higher-order Markov models are much more powerful than first-order models. However, the processing requirement increases exponentially for each order. Thus, there is a tradeoff between computational efficiency and predictive capacity when dealing with multiple-order Markov chains. One solution is relying on a *variable-order Markov models* (VMM), where the number of orders is adapted to the context in a dynamical fashion. Pachet's *Continuator* uses a form of variable-order Markov chain where the system builds a corpus based on the human performer's real-time input and is able to continue phrases in a seamless fashion (Pachet, 2003). Continuator also handles the Markovian drawback of time-limited generative power in a pragmatic manner. The human musician is left in control of defining the long-term formal structure of the performance—the system only "fills in the gaps" in a call-and-response interaction with the user who provides the musical context (Pachet, 2003, p. 334).

In a hidden Markov model (HMM), the actual states of the system are not known—they can only be inferred from a series of observed events. An HMM has two types of probabilities. *Transition probabilities* determine how the next hidden state is chosen from the current hidden state. *Emission probabilities* determine how observable states are derived from hidden states. This two-way dependency makes HMMs a powerful tool for predicting musical accompaniment to a given input. For example, Farbood and Schöner (2001) present an HMM-based interactive music system that provides the most probable Palestrina-style counterpoint to a cantus firmus. More subtly, Hamanaka et al. (2003) describe a jam session system where an HMM model extracts the groove (i.e. variations in onset times) of human performers and stores this in a "personality database". Thus, the system is capable of responding with material that preserves the idiosyncrasies of a performer's groove.

A Factor Oracle (FO) is a finite state automaton that efficiently learns internal relationships between components of a string, and was originally developed as a technique for string matching and compression (Allauzen et al., 1999). Researchers affiliated with IRCAM discovered that a FO could be used as a model for machine improvisation (Dubnov & Assayag, 2005; Assayag et al., 2006; Cont et al., 2007). Inside the FO, an input string is represented as a string of events where repeated substrings (called *factors*) are interconnected through *factor links* (pointing forward in time) or *suffix links* (pointing backward). Navigating through this structure generates a musical sequence that is both different from the original one but coherent with its internal logic (Bonnasse-Gahot, 2014, p. 4). Thus, the FO reassembles the events in a manner that claims to yield a *stylistic reinjection* of the original sequence (Assayag et al., 2006).

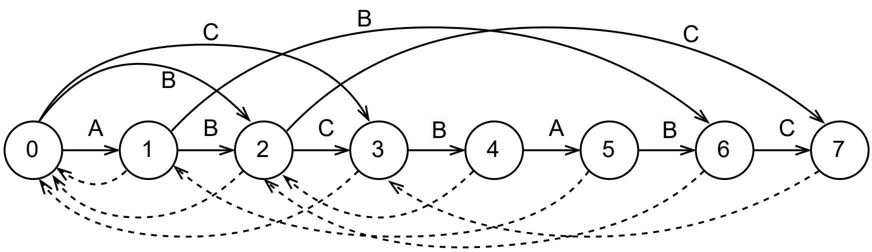


Figure 8. Factor oracle for the sequence ABCBABC. Solid arrows are factor (forward) links and dotted arrows are suffix (backward) links.

The first interactive music system to use a FO was OMax in 2006. Since then, FOs have featured in several systems designed as interactive improvisation partners, such as *Audio Oracle* (Dubnov et al., 2011), *ImproteK* (Nika & Chemillier, 2012), *PyOracle* (Surges & Dubnov, 2013), *SOmax* (Bonnasse-Gahot, 2014), and *MASOM* (Tatar & Pasquier, 2017).

Machine learning and statistical modelling are powerful tools, which in some cases give machines the appearance of autonomy. In the remainder of this chapter, I will examine the notions of agency, creativity, and music improvisation in light of the transformational dimensions of human-computer interaction and machine learning.

2.4 Agency

As mentioned in the beginning of this thesis, *agency* is the capacity of an entity to act in a given environment. It is a highly charged concept, because its application and ensuing discussions often runs deeper than semantics—the controversy is ontologically rooted. The dominant post-Enlightenment view has been that agency is a more or less fixed property belonging to human beings. From this perspective, having agency also implies accountability, and therefore has political, financial, legal, and ethical dimensions. In modernity, the attribution of agency to moving objects, forces of nature, punishing gods, and helpful angels has been relegated to the world of fetish and fiction (Rammert, 2008, p. 2). From the latter half of the twentieth century, however, increased automation and the advancement of artificial agents is once more changing the way agency is conceptualized. Rather than being an inherent property of entities, I present several alternative theories suggesting that agency is a phenomenon emergent in relations, and which take into account the agency of tools and artifacts.

2.4.1 Actor-network theory, material agency, and agential realism

Actor-network theory (ANT) is an approach to social theory that emerged from Science and Technology Studies (STS). A comprehensive presentation of ANT is provided by Latour (2005). Proponents of ANT radically oppose the existence of any preexisting structures in society. The only thing that exists, they claim, is what can be traced as dynamically shifting relations between actors in a network. Actors can be any entity acting as a mediator or a node in a network. ANT posits a flat and symmetrical concept of agency: it is not an attribute located within an actor, but manifests itself in heterogeneous associations of actors, whether they be human, nonhuman, or even objects. Thus, the ANT view of agency does not presuppose individual intentionality, but is rather distributed across the entire network of actors.

ANT scholars accuse mainstream sociology of forcing non-existing structures, such as social ties and hierarchies, upon the subjects of their studies, grouping people together according to arbitrary preconceptions in order to fit some powerful explanation. According to Law (1992), such assumptions preclude the possibility of addressing the most interesting questions about the origins of power and organizations. Rather, he proposes "we start with interaction and assume that interaction is all that there is" (p. 2). ANT maintains that there are no groups—only group formations (Latour, 2005). Established relations must be continuously upheld by some means in order to keep on existing. All the actors in a network must be identified and their relationships meticulously traced in order to see how the formations are held in place. ANT also eschews the differentiation between micro and macro. One must begin in the middle of things—*in medias res*—and faithfully document traceable connections by "following the actors" (Latour, 2005, p. 12). Only after this effort of mapping an actor-network will it be possible to see which nodes have more or less connections, and only then is it possible to identify the means by which the network is upheld.

A common criticism toward ANT is that it takes the notion of symmetry between actors too far. In the context of this thesis, the most relevant criticisms have come from theorists who share the general outlook of agency as attributable to both humans and nonhumans, but nonetheless reject symmetry between the two. Pickering (1993) defends the ANT position of agency as a phenomenon emergent in interactions, and he acknowledges a role for nonhuman agency, which he calls *material agency*. He notes, however, that material agency is *temporally emergent* on short time scales in practice. Human agency, while also temporally emergent, generally appears to work on a different time scale than material agency. Pickering attributes this to the intentional structure of human agency: "We humans differ from nonhumans precisely in that our actions have intentions behind them, whereas the performances (behaviors) of quarks, microbes, and machine tools do not" (Pickering, 1993, p. 565). Thus, the ANT view of symmetry only makes sense when maintaining a momentary view of practice. At this level, Pickering agrees with proponents of ANT that agency is evenly distributed between human and nonhuman actors: "Material and human agencies are mutually and emergently productive of one another" (p. 567). He would later coin the term *dance of agency* to describe this co-dependency (Pickering, 1995). When disregarding human intentionality, however, ANT fails to account for the capacity of humans to bring about future states determined by longer-term goals, which Pickering claims requires a view of human agency that has no material counterpart.

Kaptelinin and Nardi (2006) also favor the attention that ANT brings to the agency of things, and point out that this comes at a good time in this age of smart machines (Kaptelinin & Nardi, 2006, p. 237). Siding with Pickering, they too reject perfect symmetry due to the existence and importance of human intentions. They introduce different *levels of agency*, where agency is understood as a dimension rather than a binary attribute. According to this model, tools may have *conditional agency* (the capacity to produce unintended effects) or *delegated agency* (the capacity to realize intentions delegated to them by somebody or something else). A third kind of agency—*need-based agency*—relates to the capacity to form intentions based on cultural or biological needs, and to act upon these. Need-based agency is attributable not only to humans and animals, but also social entities such as organizations. Kaptelinin and Nardi thus maintains asymmetry between (human) subjects and (nonhuman) objects while accepting that some types of agency are manifested by nonhuman entities:

> Actor-network theory has drawn attention to the power of things, but its suspension of the concepts of human intentionality and creativity to attain symmetry is too limiting. Our position is not a "humanist" one in which humans are seen as superior by virtue of tipping the agentic seesaw over to their side; we simply seek a characterization that draws attention to the particular potency of human agency. (Kaptelinin & Nardi, 2006, p. 241)

Barad's theory of *agential realism* starts from the vantage point of quantum physics, and introduces the concept of *intra-action* to call attention to the inseparability of agencies within an entangled whole (Barad, 2007). Whereas *inter*action presupposes separate entities that do the acting, intra-action is the process whereby both objects and their agencies are realized—they are effects of the intra-action. According to Barad, what we experience as reality is continuously formed and reformed through the process of intra-action. As with ANT, agency is seen as a distributed, emergent phenomenon: "Crucially, agency is a matter of intra-acting; it is an enactment, not something that someone or something has. Agency is doing/being in its intra-activity" (Barad, 2007, p. 235). In contrast to ANT, however, Barad's notion of entanglement—the one-ness of phenomena in the world—is fundamentally different from the node-based discreetness of the actor-network:

> Practices of knowing and being are not isolable; they are mutually implicated. We don't obtain knowledge by standing outside the world; we know because we are of the world. We are part of the world in its differential becoming. The separation of epistemology from ontology is a reverberation of a metaphysics that assumes an inherent difference between human and nonhuman, subject and object, mind and body, matter and discourse. (Barad, 2007, p. 185)

The framework of ANT, the views of material agency advocated by Pickering and Kaptelinin and Nardi, and Barad's agential realism—although different from each other—all share the commonality that agency is placed outside individual actors. It is emergent-in-action as opposed to present-in-matter. This provides useful ways to consider the interplay between humans and technology.

2.4.2 Dividuals and collective agency

A durable legacy of the European-American Enlightenment is the rise of the individual. Modern Western democracies are founded upon the idea of a social contract between the individual and the state, defining the rights and duties of each. This social contract as expressed in documents such as the Declaration of Independence and the Declaration of

Rights of Man were integral to the American and French revolutions respectively. It forms the moral, secular backbone of Western culture at large. However, individualism coupled with the dominant philosophical notion of agency as a potency residing within every individual begets the question whether there is an infinite amount of agency in the world. Fuller (1994) challenges the notion of agency as an indivisible quality of individuals when taken to the extreme:

> The specter of 5 billion people simultaneously exercising their "inalienable rights" as Lockean agents to "life, liberty, property" gives one pause. If all the world's people were to succeed in aspiring to the powers of agency enjoyed by the average American citizen, then the resulting counterfinality effects to the environment would soon subvert anyone's ability to enjoy such agency. Thus one may suspect that agency is really a divisible good whose value is governed by ordinary economic principles of scarcity. (p. 742)

Fuller suspects that the philosophical disinclination of granting agency to nonhumans arises from an underlying suspicion that agency is, in fact, a "fixed pie"—the more agents there are in the world, the less agency each can have (Fuller, 1994, p. 741). On this note, it is worth reiterating that individualism is by all accounts a predominantly Western idea with its own historical and cultural contingencies. A contrary view of personhood is the notion of the *dividual*, a term coined by Marriott (1976) during field work on the Indian subcontinent:

> Single actors are not thought in South Asia to be "individual", that is, indivisible, bounded units, as they are in much of Western social and psychological theory, as well as in common sense. Instead, it appears that persons are generally thought by South Asians to be "dividual" or divisible. To exist, dividual persons absorb heterogeneous material influences. They must also give out from themselves particles of their own coded substances, essences, residues, or other active influences that may then reproduce in others something of the nature of the persons in whom they have originated. […] What goes on between actors are the same connected processes of mixing and separation that go on within actors. (p. 111)

The dividual has been recognized in subsequent anthropological studies in different regions of the world. Bown (2015) argues that this kind of agential fluidity is apparent in all world views to some extent, but particularly evident in computationally creative systems:

> A piece of software is itself an assemblage of subsystems and may communicate beyond its nominal boundaries to form supersystems, including with humans.

> We should expect that in some cases it is clear that agency is more strongly associated with a specific subsystem than with others, whereas in other cases, agency takes the form of interaction between subsystems or the system and its environment. (p. 19)

The notion of agential fluidity in complex sociotechnical systems emphasizes the system or the collective as the acting unit. Misselhorn (2015) points out that individual agency cannot account for the type of behavior that emerges from complex hybrid systems involving both human and artificial agents. The term *collective agency* is adopted to account for the behavior of systems that are irreducible to the intentions of the individuals that constitute them (Misselhorn, 2015). Collective systems display emergent coordination that hinges on *we-intentions* (Tuomela & Miller, 1988) and on the discovery of *joint affordances* (e.g. a seesaw as an affordance for two people).

The foregrounding of the individual in much of modern Western philosophy has apparently stolen focus from worldviews that accommodate much more nuanced and dynamic notions of agency, which arguably have been present all along. When assembly line factory workers clock in, they leave their sense of agency waiting in their lockers. The king of the cocktail party can feel like nobody the next day. The self has always been malleable, but the hive-mind of the Internet and the increased offloading of both manual and cognitive tasks to technological systems has brought a challenge to the traditional static view of agency.

2.4.3 Agency in interactive music systems

The usefulness of ontologies where agency is viewed as a potency throughout a heterogeneous plurality rather than one within a select few entities is apparent in research domains engaging in computationally creative systems. For example, the concept of mixed-initiative creative interfaces presented in Chapter 1 is based on a *spectrum of agency*, along which computers can be reactive tools, interactive collaborators, or proactive creators (Deterding et al., 2017). Mixed-initiative creative interfaces are ones in which creative agency is shared between the human and the computer.

Bown et al. introduce the term *behavioral object* to denote “software that has the capacity to act as the musical and social focus of interaction in digital systems” (Bown et al., 2009, p. 188). Behavioral objects are akin to the idea of dividuals in a heterogeneous collective, and the introduction of the term is a deliberate step away from relying on metaphors drawn from “the classical triumvirate of composition, performer and instrument” (Bown et al., 2009, p. 188). The authors argue that digital environments engender a fluid, modular approach to music creation where traditional roles associated with the acoustic paradigm are enmeshed in the software medium. It is therefore counterproductive to view music making through the lens of the acoustic paradigm with its rigid distinctions between objects, activities and roles. They propose that behavioral objects can contribute actively to musical culture through *performative agency* and *memetic agency*. Whereas performative agency refers to the here and now of musical performance, memetic agency refers to the influences of software on musical styles over historical time. An example of performative agency is a particular behavioral object (e.g. a machine learning model) that contributes to a piece of music proactively during an interaction. An example of memetic agency is a type of behavioral objects (e.g. step sequencers) that become a part of musical culture over time. While these agencies are different from agencies associated with the traditional roles of human performers and composers, they nonetheless affect music performance in real-time and the evolution of musical culture over historical time.

Beyls reasons that systems that are truly interactive as opposed to merely responsive engage in a symbiotic relationship with a hybrid performance agency where "communication is characterized as a form of common initiative and shared control" (Beyls, 2018, p. 238). Once more, a fluid conception of agency is tied to the idea of humans sharing the creative initiative with computational modules. Beyls presents a multi-agent system reminiscent of Minsky's "society of mind", where intelligence is seen as emergent from the interaction between many agents which in and of themselves are not necessarily powerful (Minsky, 1986). The agency arises from the coordination between a variable number of player agents held by a Player Agency, recruited from a pool of potential agents held by a Policy Agency (Beyls, 2018, p. 240).

The concepts and terminologies introduced in the above examples demonstrate a nascent recognition of a "distinctly new kind of agency in the creative loop" (Jones et al., 2012, p. 199). By reducing their own agency, musicians can share agency with interactive music systems and discover new forms of musical expression (Mudd, 2017). This sentiment is neatly summed up by Jones et al. (2012):

> Interaction with a semi-autonomous music system inhabits an unfamiliar midpoint on the spectrum of creative relationships. It resides somewhere between tool usage and human collaboration, inheriting some characteristics of each and adding some of its own. (p. 182)

The discussion of agency has foreshadowed its inextricable relationship to the concept of creativity—a term that has pervaded this chapter. It is therefore fitting to turn the focus toward creativity in the next section.

2.5 Creativity

Upon tracing the myriad conceptions of creativity throughout history, it is striking how the trajectory of discourse seems to have been informed by the same ontological shifts as the one related to agency. A concise summary of this historic development is provided by van der Schyff et al. (2018). Early records explain creativity as an external force bestowed upon humans. In Ancient Greece, creativity was thought to be delivered by the muses (Dacey, 2011). In the European Middle Ages, the muses were replaced by a Christian God, who gifted certain people with "divine inspiration" (Albert & Runco, 1999). Then, coinciding with the turn to individualism during the Enlightenment and the ensuing Romantic era, the "giftedness" of creativity came to be viewed as coming from a special kind of individuals (R. S. Albert & Runco, 1999). Now an internal human property, albeit one reserved for people with innate talents, the notion of uniquely creative geniuses propelling human culture forward has stayed in strong currency ever since. What all of these views have in common—whether founded in spiritual, divine or human realms—is that creativity is an inscrutable force off-limits from scientific inquiry. They also suggest that creativity is unattainable for ungifted people. Nonetheless, the domain of creativity research offers a multitude of theories that demystify these persevering notions.

2.5.1 Contemporary views of creativity

Apparently, there are more than a hundred different definitions of creativity (Meusburger, 2009), out of which the common elements are based in novelty and value (Loughran & O'Neill, 2018). One definition that attempts to accommodate these variations is provided by Boden: "Creativity is the ability to come up with ideas or artefacts that are new, surprising and valuable (Boden, 1990, p. 1). Boden distinguishes between two senses of being creative. One sense is psychological: *P-creative* ideas are ones that are new to a person. The other sense is historical: *H-creative* ideas are new to the world. In other words, H-creative ideas are special cases of P-creative ideas that lead to innovations for humankind (Boden, 1990).

The process of creativity has been studied in earnest since the early twentieth century, and is generally accepted to include both conscious and subconscious mental processes. One of the most widespread accounts of the creative process is the four-stage model first described by Wallas (1926). The four stages are:

1. *Preparation.* Collection of data or acquisition of knowledge about a problem, perhaps without successfully understanding it.

2. *Incubation.* A period of unconscious deliberation and restructuring of the problem.

3. *Illumination.* A sudden flash of insight when a solution of the problem becomes apparent.

4. *Verification.* The development of the idea into a form that can be applied in practice and shared.

Much debate revolves around what actually happens during the incubation stage leading up to the "eureka" moment of the illumination stage. A number of authors have proposed that the abandonment of actively thinking about a problem allows the brain to superimpose new information from the environment upon the latent problem (Meusburger, 2009). This background activity can help toward seeing the problem from new perspectives.

Indeed, the power of analogy is prevalent in first-hand accounts of creative processes. A famous example is Kekulé's description of how he came upon the possibility of the existence of ring molecule structures as he sat dozing by the fireplace. In his lucid mind, the flames took on the form of snakes. As one of these "snakes" bit itself in the tail, he immediately recognized this as the potential structure of the benzene molecule which had been troubling him for a long time. The result of seeing this tail-biting snake in the fire was *aromatic chemistry*—a whole new branch of science (Boden, 1990).

The example of Kekulé and countless other stories about innovators experiencing epiphanies "out of the blue" contribute to keeping the myth alive that exceptionally creative people have an ability to peer into the unknown and conjure up ideas as if by magic. This image belies the amount of hard work required to attain such experiences. Pasteur famously claimed that "chance favors only the prepared mind" (Pasteur, 1854). Serendipity apparently befalls people with extensive knowledge in their field, and who have spent large amounts of time embroiled in a problem. Added to this, the culmination of creative insights requires everyday psychological abilities such as remembering, noticing interesting things and recognizing analogies (Perkins in Boden, 1990, p. 35).

The potency of analogies is typically realized in what Boden calls *combinational creativity*, which is an effective, unfamiliar juxtaposition of known ideas (Boden, 1990). This is essentially the same as Koestler's (1964) earlier description of creativity as the *bisociation of previously unrelated matrices*. Boden (1990) describes two other types of creativity which are non-combinational. *Exploratory creativity* occurs when new ideas are discovered by traversing a conceptual space (a structured style of thought). An example of exploratory creativity is composing never before heard music within the constraints of an established musical genre. *Transformational creativity* goes further than exploration of a conceptual space and redefines it altogether. The abandonment of functional harmony in favor of the new conceptual framework of atonal music in the early twentieth century is an example of transformational creativity. Artists who are described as "pushing the envelope" is an apt description of artists who frequently engage in exploratory creativity, perhaps occasionally verging on the transformational (it is implied that "the envelope" is the confines of a known conceptual space).

In recent decades, views countering the focus on creativity as an individual capability have been proposed by researchers who highlight creativity as an emergent phenomenon that occurs in the interaction between multiple agents, as well as the broader sociotechnical milieu they are situated in (van der Schyff et al., 2018, p. 1). According to Csikszentmihalyi, this interaction involves the evaluation and verification of new ideas by a domain:

> There is no way to know whether a thought is new except with reference to some standards, and there is no way to tell whether it is valuable until it passes social evaluation. Therefore, creativity does not happen inside people's heads, but in the interaction between a person's thoughts and a sociocultural context. It is a systemic rather than an individual phenomenon. (Csikszentmihalyi, 1996, p. 11)

Simonton (2003) also promotes a systemic view of creativity (focusing on the scientific domain) when arguing that it can be modelled as a stochastic process across a population. This model is based on the perspective of products that emerge from scientific communities instead of studying creative individuals or creative processes.

Both Csikszentmihalyi and Simonton focus on what Boden would categorize as H-creative ideas. Perhaps then, a way of looking at the social dimension of creativity is as a filter through which P-creative ideas are input, only seldom resulting in ideas that have an impact on larger communities. While this may be the case, it leaves out an important social aspect of creativity which is much more common. Collective pooling of ideas through group thinking, brainstorming, or even informal "water cooler talk" is a well-documented source of creativity. There is a dynamic at work in collaborative contexts that cannot be explained by adopting a solely individual view of creativity. This is certainly the case for musicians who spend large amounts of time together in rehearsals, traveling, eating, drinking, and generally socializing. The flux of influence between individuals in such contexts makes it difficult to maintain a purely individual view of creativity. Once more, Bown's adoption of the anthropological notion of the "dividual" offers a useful perspective:

> From this alternative point of view it is argued that artistic behaviour has a significant generative creativity element by which new forms "spring up", not because individuals think of them, but through a jumble of social interaction. Such emergent forms may have structural properties related to the process that produced them, but they were not made with purpose. (Bown, 2014, p. 116)

Creativity, in other words, is not always intentional. Bown's use of the term *generative creativity* captures non-cognitive and even nonhuman processes such as biological evolution and new ideas that emerge in complex sociotechnical systems. The only criterion for generative creativity is that it results in something new and valuable. The process by which the result is produced is secondary. On the other hand, *adaptive creativity* concerns ideas and artifacts created intentionally by an intelligent agent in response to a need or opportunity (Bown, 2014, p. 116). In collaborative contexts such as the ones mentioned above, both generative and adaptive creativity could be at work, and it can become exceedingly difficult to trace idea ownership.

Sawyer refers to idea generation in creative group dynamics as *emergent novelty*, which cannot be understood by trying to reduce the interaction to studies of the psychology of individual actors (Sawyer, 1999, p. 449). According to Sawyer, emergent novelty alone is not the same thing as creativity. A second property he calls *appropriateness* involves the curation of the combination of ideas that amounts to the creative product: "Whereas emergent novelty is a bottom-up process in complex systems, appropriateness requires that we also consider top-down effects in systems with multiple levels of emergent process" (Sawyer, 1999, p. 449).

So far, this review has focused on the longer-term dynamics of creativity. There are two psychological states occasionally at work during in-the-moment creativity that should be mentioned due to their significance in musical creativity. One is the concept of *flow*, which Csikszentmihalyi (1990) defines as "the state in which people are so involved in an activity that nothing else seems to matter; the experience itself is so enjoyable that people will do it even at great cost, for the sheer sake of doing it" (p. 15). Characteristics of the flow experience include feeling "at one" with the task at hand and the loss of self-consciousness. The second relevant psychological state is the *creator/witness phenomenon* (Berkowitz, 2010), which refers to the momentary loss of a conscious connection to the body's actions. Music improvisers report the sense of becoming a spectator of one's own playing during performance peaks. The idea of "letting go" while improvising entails the allowance of automated skills to take precedence over cognitive deliberation (Walton et al., 2015), leading to moments where the performer can be surprised by their own performance. Pressing (2001) succinctly explains:

> The accompanying feeling of automaticity, about which much metaphysical speculation exists in the improvisation literature, can be simply viewed as a natural result of considerable practice, a stage at which it has become possible to completely dispense with conscious monitoring of motor programs, so that the hands appear to have a life of their own. [...] In a sense, the performer is played by the music. (p. 139)

Both flow and the creator/witness phenomenon involve the reduction or loss of self-consciousness, and hence seem interrelated. The latter appears to be momentary, singular, and intermittent, whereas the former could describe more sustained periods involving a range of disparate tasks. Importantly, the colloquial notions of "going with the flow" and "letting go" point back to the interactive view of agency described in the previous section. If attaining these states of heightened creativity indeed entails a sense of shared agency with one's environment, it follows that tools with some level of interactive autonomy have the potential to help users reach these states more often.

2.5.2 Creativity support tools, computational creativity, and co-creativity

Creative music technologies can be roughly divided into two main categories. *Creativity support tools* (CSTs) are designed to support human creativity (Shneiderman, 2007), whereas systems with *computational creativity* (CC) are built to generate artifacts that are judged by unbiased users to be creative (Colton & Wiggins, 2012). The difference between CST and CC is the attribution of creativity: who or what is the creative part. The former category includes the most commercially widespread music technologies to date. Digital audio workstations, notation software, sound synthesis systems, and most commercially available music production technologies in general are designed to support humans in creative processes. More recently, however, artificially intelligent systems designed to generate music or contribute to music making according to users' specifications have gained traction. Such technologies are computationally creative.

It is important to emphasize that the qualifying factor for computational creativity is the *appearance* of creativity. Researchers in the field are not necessarily arguing that computational systems are *actually* creative (at least not as traditionally defined). From a cognitive-social approach, some of the more interesting questions revolve around what computational ideas can teach us about how human creativity is possible, whether such creativity can be simulated, and whether computers can recognize creative ideas coming from other agents (Boden, 1990). From a mathematical-engineering approach on the other hand, relevant questions revolve around the construction of artifacts that are appealing for an audience (Pérez y Pérez, 2018). Neither of these approaches necessarily engage in the philosophical conundrum of actual nonhuman creativity.

Apart from CST and CC, a third, hybrid category is *co-creative systems* where computers and humans collaborate with each other to make shared creative artifacts (N. Davis, 2015). It can be helpful (albeit a simplification) to think of the field of HCI as the proponent of CST, the field of AI as the driver of CC, and that co-creativity draws from both these fields. The term *co-creativity* was proposed by Candy and Edmonds (2002) to account for a move away from purely computer-generated "computer art" toward more diverse forms of human-computer co-creation. Other authors have characterized co-creativity as *human-computer creative partnerships* (A. R. Brown et al., 2017). Karimi et al. (2018) claim that there needs to be interaction between at least one artificially intelligent agent and at least one human in order for a system to qualify as being co-creative.

It is reasonable to assume that co-creative systems are equivalent to the notion of mixed-initiative creative interfaces introduced earlier. However, Karimi et al. (2018) make a distinction between the two terms:

> Mixed-initiative systems are by definition co-creative, but not all co-creative systems are mixed-initiative. In many systems there is an explicit turn-taking process, but this is not a requirement: some systems are machine-initiative dominated, operating as a kind of "wizard" interface in which the user is consulted during a highly scripted process, while others are user-dominated, with the system jumping in only infrequently with suggestions or critique. (p. 105)

In other words, the creative agency in mixed-initiative creative systems is designed to be shifting dynamically between the human and computational agents. While this may also be the case for some co-creative systems, it is not by definition so. The distribution of agency in co-creative systems may be more or less fixed or static. Regardless of whether a system is co-creative, mixed-initiative, or both, these forms of distributed creativity can yield emergent, dynamic, and unexpected results with an unclear authorship: "The mixed initiative co-creativity […] which emerges from this human-computer interaction cannot be ascribed either to the human or to the computer alone" (Liapis et al., 2016, p. 137). Jordanous (2017) points out that this unclear authorship requires new methods for evaluating the creativity of such systems:

> Although there have been many advances in computational creativity evaluation tools to date, typically these tools tend to assume we are evaluating a single piece of software, working without creative contribution from other entities. In co-creativity, there are more than one participants contributing to the creative process, but often we cannot delineate the specific contribution of each participant. (p. 159)

It is safe to assume that the conglomeration of human and computational creativity is only just beginning. Similar tendencies are already far advanced in other practices, such as translation and map reading. The prevailing belief that creativity is an untouchable human capacity is questionable. Although qualitatively different from other practical domains, the arts are not exempt from aesthetic and ethical challenges in the face of AI, machine learning, and big data. Creativity viewed as a sociotechnical phenomenon is a step toward accepting these challenges.

2.6 Music making

This chapter has focused much on the implications of technology on music, agency, and creativity. The red thread in this composite review is a shift toward an interactional view of music, which highlights music as an unfolding social process. I have chosen the term *music making* as a way to deemphasize the rigid "division of labor" associated with composition, performance, improvisation, and listening. In *The Improvisation of Musical Dialogue*, Benson (2003) writes:

> The binary schema of "composing" and "performing", which goes along with the construal of music making as being primarily about the production and reproduction of musical works, doesn't describe very well what musicians actually do. In its place, I wish to suggest an improvisational model of music, one that depicts composers, performers, and listeners as partners in dialogue. From this perspective, music is a conversation in which no one partner has exclusive control. (p. x [preface])

As outlined in Chapter 1, the theme of this thesis is the implications of including computational agents as creative partners in the initial ideation stage of music making. To conclude this chapter, I will therefore elaborate on the facets of music making at work in such contexts. Rather than seeing these facets as separate activities, I will make a case that they are all different modes of the activity of music making.

2.6.1 Improvisation as the catalyst for music making

Unless one accepts the possibility of creation *ex nihilo*—that it is somehow possible to make something out of nothing—it makes sense that anything new is in some way a variation of something pre-existing. It could be a new combination of ideas, or the result of exploring or restructuring existing conceptual frameworks (cf. combinational, explorative, and transformational forms of creativity: Boden, 1990). This is the essence behind Benson's (2003) claim that composition is at its very core improvisatory in nature:

> To improvise is to rework something that already exists (that is, "conveniently on hand") and thus transform it into something that both has connections to what it once was but now has a new identity. "Composing" is not simply a matter of bringing elements together; rather, they are brought together in a way that transforms those elements. (p. 45)

Claiming that all music making is improvisatory at its core is not the same as proposing that all music is improvised. This is clearly not the case. However, impromptu decisions are necessary for novelty to occur. The outcomes of these decisions are then refined. No matter how much theory that goes into the preparation of music making, and regardless of how elaborate the framework is for arranging novel ideas, there are key points at which spontaneous choices cause "ignitions" or "chemical reactions" that transform the material. In chemistry, a *catalyst* is a substance that causes or increases a reaction without itself being consumed. Metaphorically, improvisation is the catalyst that brings elements together and transforms them in the process.

This phenomenon is more readily observable in collaborative music making. Musicians bring their own set of experiences and expertise to a collaboration. They present ideas and begin playing together. They pick up on each other's contributions, and experiment by testing different complementary approaches. At some points, transformations occur—new ideas "pop out" that could not have happened by any other means than bringing these elements together in an improvisatory manner. This is what Sawyer (1999) refers to as emergent novelty, described in the previous section. In additional to the unpredictable outcomes that occur in such collaborative contexts, there is moment-to-moment contingency. Each path taken entails ruling out countless alternatives. Thus, a piece of music is a realized potential among many others.

Music group dynamics are often characterized by a tug-of-war between individual and collective subjectivities. This is described by Seddon as "a tension between individual performance and awareness of other musicians" (Seddon, 2005, p. 47). He elaborates:

> During improvisation jazz musicians deal with the internal constraints of the knowledge base acquired through deliberate practice (e.g. musical materials, excerpts and problem solving routines). But they also deal with the external constraints of cultural referents (e.g. in jazz the 32-bar cycle of jazz standards, chords, and characteristic rhythmical patterns). (p. 48)

According to Seddon, musicians aspire to transcend their knowledge base and attain inter-subjective engagement through *empathetic attunement*, which happens when they are able to decenter and see things from other musical perspectives (Seddon, 2005). The decentered subject is a poststructuralist notion whose proponents claim that the subjectivity is a construction with unclear and variable boundaries between the individual self and society (T. Davis, 2011). For example, Guattari (1995) writes:

> We know that in certain social and semiological contexts, subjectivity becomes individual: persons, taken as responsible for themselves, situate themselves within relations of alterity governed by familial habits, local customs, juridical laws, etc. In other conditions, subjectivity is collective—which does not, however, mean that it becomes exclusively social. (p. 9)

Based on Guattari's notion of collective subjectivity, Davis states that "[t]he performance ecosystem […] represents the dynamical and emergent structural characteristics that form in the moment of encounter—in the social interaction found in collective music-making" (T. Davis, 2011, p. 122). With this understanding, perhaps flow experiences are, in fact, a form of decentering (cf. loss of self-consciousness, feeling "at one" with the task) that enables collective subjectivity, which in turn may lead to emergent novelty. The tension created by individual subjectivities, on the other hand, may be viewed as a regulatory force that guides the musical interaction into an inter-subjective space where collective creativity becomes possible.

2.6.2 The dynamics of collective music making

Improvisational collective music making is often referred to as *jamming*. This thesis focuses on jamming-oriented approaches to music making, which takes advantage of the emergent novelty created by group dynamics. A jam often starts with a simple idea that serves as the starting point for musicians to elaborate on. This simple idea could be a melodic phrase, a chord structure, a certain sound object, a rhythmical pattern, or any variation of such basic musical elements. Benson (2003) refers such an origin as the *Ursprung* of a piece of music. Elsewhere, the original idea is referred to as the *seed* (Pressing, 1984). The process of jamming may be an efficient method to get from a basic musical idea to larger formal structures. An apt metaphor is thinking of a musical phrase as an elementary kernel. Interactions may "fertilize" this kernel and larger forms can "grow" from it (Thelle & Pasquier, 2021).

A wide range of musical genres utilize jamming or collective improvisation as a method to develop music from basic ideas to larger forms, and there are, of course, many differences in what kind of musical strategies musicians employ depending on the aesthetic frameworks they operate within. On a higher conceptual level, however, there are also many similarities in regards to the overall dynamics involved in these open-ended forms of creative interaction. A significant number of proposed models in the literature focus on a balance between phases of "pulling together" and "pushing apart" (Thelle & Pasquier, 2021).[1]

Wilson and MacDonald (2016) report that musicians are faced with an ongoing choice of whether to *change* or *maintain* what they are doing. A change can be either an *initiative* (something new) or a *response* (to what another musician is doing), and three emergent response categories are *adoption*, *augmentation*, and *contrast*. Borgo (2005) describes how forms emerge in collective improvisation through *positive feedback*—a mutual reinforcement of a particular idea, and how interest is simultaneously maintained through *negative feedback*—an exploration of new ideas diverging from the current one.

Similar concepts have been presented in models for interactive systems. Dubnov and Assayag (2005) introduce a *flow model* where improvisation occurs along the axes of *replication*, *recombination*, and *innovation*. Replication implies following another musician's initiatives, while someone engaging in innovation is taking an initiative. Beyls (2018) presents a model for human-machine interaction where the system's behavior follows from the competition between the opposing forces of *expression* (output generated irrespective of or contrasting to current context) and *integration* (output that is complementary to the prevailing context and contributes to its further existence). Canonne and Garnier (2012) invoke a model for collective free improvisation where strategies range from *stabilization* (attempts to converge to a "collective sequence") to *densification* (deliberately creating complexity to provoke a transition).

1 Parts of the following review of some of these models were presented in the paper "Spire Muse: A Virtual Musical Partner for Creative Brainstorming" (Thelle & Pasquier, 2021).

Although these models present an apparent terminological jungle, Thelle and Pasquier (2021) propose that all these concepts are attempts to explain the mentioned forces of "pulling together" and "pushing apart". In essence, they are musical strategies that may be grouped along a *musical similarity axis* ranging from *converging* to *diverging*, as depicted in Figure 9.

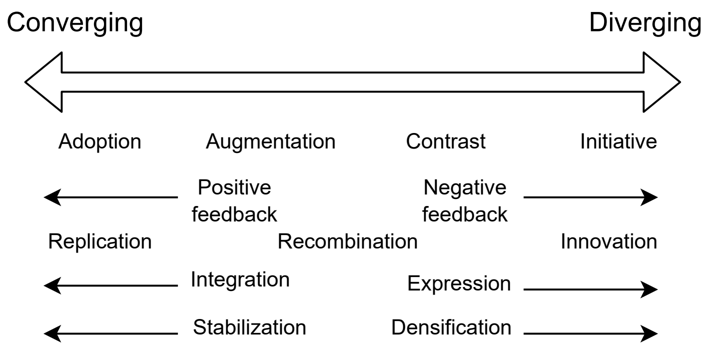


Figure 9. Musical strategies mapped onto the musical similarity axis.

In order for musicians to engage in these converging and diverging dynamics, they must first attain an inter-subjective understanding of the ongoing situation. When co-improvising musicians reach such an understanding—in colloquial terms, a feeling of being "on the same page"—they have a *shared representation* of the interaction (Murray-Rust & Smaill, 2011; Canonne & Garnier, 2012). During the course of jamming, a musician will listen to musical output either as something expected according to the shared representation, or something which is unexpected or novel (Murray-Rust & Smaill, 2011).

According to Pelz-Sherman (1998), musicians have a dual focus on the musical signal while improvising together. One is the acoustic content of the musical signal, and the other is the *semantic content*, which conveys the perceived intentions of the musicians. It is through the semantic channel that musicians are able to engage in the interactive dynamics that lead to evolving musical forms. The semantic content may contain different types of cues that the co-improviser can choose to respond to or ignore. For example, through a given gesture a musician can communicate to their co-improvisers objectives relating to formal development (*formal content*¸ e.g. a break away from a thematic cycle). Alternatively, a musician may convey intentions to engage in certain forms of interaction (*interactional content*, e.g. call-and-response, accompaniment, counterpoint, etc.). A musician may also communicate evaluations of the prevailing situation (*evaluative content*, e.g. signaling approval by playing in unison or giving space).

2.7 Summary

In this chapter, I have scratched the surface of themes that cannot be treated comprehensively within a limited scope. However, I hope I have provided the reader with enough to sustain an appropriate background understanding of the issues to be covered in the following chapters. First, I introduced interactive music systems and presented the historical context for their development into the current state of the art. Next, I provided a review of HCI and some concepts within this domain that are important for this thesis, followed by an introduction of the field of machine learning. Where applicable, I will refer back to the relevant sections of this chapter as concepts are reintroduced in the upcoming chapters. In the latter half of the chapter, I made forays into the concepts of agency and creativity with an emphasis on contemporary sociotechnical perspectives that align well with the theme of mixed-initiative music making. Finally, I presented research into music making with a particular focus on collective improvisation as a means to develop music from basic ideas to larger formal structures.

3 Dynamical systems

Collective music making is a universal phenomenon. Some music *collectives* (duos, groups, bands, ensembles, orchestras, etc.) may have an authoritative leader—a composer or songwriter—who writes the music and directs the other musicians. Other collectives have less hierarchy. In the most flat-structured collectives, all members may contribute with ideas, and music making may happen through the process of *jamming*, as described in Chapter 2. Usually, there are bouts of negotiation between the members, including breaks with verbal communication and individual elaboration of sequences in between sessions. This thesis focuses on the nonverbal jamming part of this compartmentalized process.

Meeting the challenge posed in the research question requires an interactive music system that can model some of the converging and diverging strategies outlined in Chapter 2.6.2, and engage in both taking and following initiatives in an open-ended, exploratory fashion. Some of the systems reviewed in Chapter 2 do take such an approach. However, I find that they are mostly designed for live performance, which is somewhat different from the more introspective context of musicians making music in a rehearsal space scenario. From the outset, I envisioned a session-based musical brainstorming environment to help musicians find emerging musical structure through playful interaction, without any notions of performing in front of an audience.

There is a gap in our knowledge of the creative processes in such collective, yet personal and intimate contexts. As outlined in Chapter 2, collaborating music makers do not follow a destined path toward a goal. It is often a messy process, contingency-based and highly unpredictable. A small contribution from one musician may trigger a radical shift to a musical idea. The flow of influences between the musicians tends to dissolve individual subjectivity. Saxophonist Evan Parker formulates this eloquently:

> However much you try, in a group situation what comes out is group music and some of what comes out was not your idea, but your response to somebody else's idea…. The mechanism of what is provocation and what is response—the music is based on such fast interplay, such fast reactions that it is arbitrary to say, "Did you do that because I did that? Or did I do that because you did that?" And anyway the whole thing seems to be operating at a level that involves… certainly intuition, and maybe faculties of a more paranormal nature. (Evan Parker cited in Borgo, 2005, p. 183)

Describing how ideas develop in musically creative contexts can be challenging due to the intrinsic knowledge and intuitive choice making that musicians apply in the creative process.

Several qualitative studies point to a lack of awareness on the performer's part during optimal performance, and neurological research seems to confirm that typical flow experiences are accompanied with the suppression of central processes associated with self-monitoring and conscious volitional control (Walton et al., 2015). Parker's attribution of a paranormal dimension to the group mind phenomenon is an illuminating example of this lack of awareness.

In this chapter, I propose that *dynamical systems theory* (DST) (Strogatz, 2015) is a useful framework to explain the behavior of the group mind in musical interaction, which in turn may influence design. First, I will introduce the scientific foundations of DST. Next, I will present both challenges and potential benefits of applying concepts from one domain to explain phenomena in another unrelated domain, and make a case for adopting DST as a conceptual framework for this thesis. Finally, I will review a number of examples of how DST has been applied in other domains, including economics, psychology, music, and interactive music systems.

3.1 A brief introduction to DST

A *dynamical system* can be described as any set of equations giving the time evolution of the state of a system from a knowledge of its previous history (Ott et al., 1994). In the past few decades, DST has been associated with *complexity theory*—the study of complex systems—and is frequently categorized as its subfield. However, the mathematical principles behind the science of dynamical systems was established as early as the 17th century (Strogatz, 2015)—long before the relatively new focus on complex systems. In fact, it could be argued that DST "midwifed its parent" through the discovery of complex behavior and chaos in deterministic nonlinear systems in the latter half of the 20th century. Borgo (2005) makes a distinction between the two that I find pertinent with respect to the topic of this thesis. Whereas researchers of dynamical systems have been mostly interested in how mathematically simple systems can produce complex and chaotic behavior, complexity theory deals with complex and highly interconnected systems that may, under certain conditions, self-organize and produce emergent forms of order. Borgo refers to this distinction as "two sides of a turbulent mirror" (Borgo, 2005, p. 83).

Although originating from the domain of physics, both DST and complexity theory have been discovered as useful conceptual and mathematical modelling sources across disciplines today (Mainzer, 2004; Strogatz, 2015). In particular, the surge of interest in chaos theory following James Gleick's *Chaos: Making a New Science* (1987) ushered in a wave of research from the social sciences and humanities that began applying chaos, dynamical systems, and complexity theory as explanatory models for all kinds of activity or behavior in their fields. This may also explain why terms such as "dynamical systems theory", "nonlinear dynamics", "complex systems", "complex dynamic systems", "chaos theory", and "complexity theory" are often used interchangeably, sometimes to a confusing degree.

3.1.1 Terminology

The following terminology forms the foundation of DST. I will present both the traditional definitions and briefly explain how I am applying the concepts in the conceptual framework of this thesis. Structurally, this part follows the terminological overview presented by Mudd (2017).

Phase space is a space that represents all possible solutions to a given problem. In classical dynamical systems, this means the set of possible solutions to the equations that models the phenomenon being studied. For systems with up to three dimensions, the phase space can be visualized as a graph or a computer animation. Higher-dimensional phase spaces are more difficult to visualize, but are useful conceptual tools. In this thesis, phase space is used conceptually, and I will adopt Borgo's understanding of phase space as a system's "geometry of possibilities" (Borgo, 2005, p. 69). This includes the idea of the phase space of musical interaction—a space of "musical possibilities" too multidimensional and complex to model in mathematical terms.

At the heart of all dynamical systems is the concept of *iteration*: the future state of any dynamical system is dependent on its current state. Hence, the set of values of the current state functions as inputs to the future state, whose results in turn become inputs to calculate further the subsequent state. The first set of parameters are the system's *initial conditions*. A *trajectory* in phase space represents the set of states that the system undergoes when starting from one particular initial condition. The combination of trajectories from *any* initial condition amounts to the entire phase space. In the framework of this thesis, a trajectory can also represent a musical progression through the conceptual phase space.

An *attractor* in phase space is a set of values toward which all neighboring trajectories converge when the system's underlying functions are iterated (Kaplan & Glass, 1995). There are several types of attractors. If the solution set converges to a single coordinate, this coordinate is called a *fixed-point attractor*. Systems whose trajectories in phase space tend to oscillate between two or more coordinates is said to have a *periodic attractor*. A trajectory that ends up cycling through the same coordinates converges to a *limit cycle*. Among several other types, there is also the *strange attractor*, which will be explained in Section 3.1.3. A system may have different attractors depending on its parameters, and typically, the reconfiguration of attractors in phase space can happen abruptly as parameters are gradually changed and cross certain thresholds. Subsequent sections in this chapter also introduces *perceptual attractors*, and extends the concept of attractors to emergent mental phenomena.

Nonlinearity is another important concept in DST. A nonlinear system is one whose output is not proportional to its input (Borgo, 2005). Most phenomena in nature are nonlinear, and cannot be understood as a sum of its parts (Strogatz, 2015). This acknowledgment has led several branches of science away from a reductionist to a more holistic systemic view. Social sciences are also realizing that the greatest problems of humanity are complex and nonlinear. Small changes in local ecological, economic or political systems may trigger unpredictable, radical changes on a global scale (Mainzer, 2004). Nonlinearity is a prized property of most acoustical music instruments. For instance, minute variations of control parameters such as air pressure and embouchure in wind instruments can yield rich expressive nuances. Therefore, developers of digital musical instruments often build nonlinearity into sound synthesis modules to enhance expressivity (Miranda & Wanderley, 2006).

Nonlinear systems may display chaotic behavior. The mathematical definition of *chaos* is widely misunderstood because it is very different from its colloquial use. Although it is common to equate chaos with randomness, the process that produces chaotic behavior is completely deterministic. A system is chaotic if iterating a mathematical function yields a trajectory that never repeats itself (infinitely aperiodic), stays within a bounded range, and displays *sensitive dependence on initial conditions*. The latter implies that varying the initial conditions with only a miniscule amount will result in completely different trajectories. Hence, chaos is actually wholly predictable if the model is complete and one knows the *exact* initial conditions, although the behavior is seemingly random. Chaos has fascinated artists and musicians because many emergent properties, when visualized or sonified, often reveal complex patterns that are aesthetically pleasing.

Bifurcations are one of the hallmarks of nonlinear systems. A bifurcation is a sudden shift from one form of qualitative behavior to another as a parameter is gradually changed (Kaplan & Glass, 1995). When a bifurcation occurs, attractors in the system's phase space undergo a reconfiguration, which can help explain this shift in behavior. The manner in which the human voice abruptly transitions between the chest and head voice (voice breaking) can be modelled as a bifurcation (Bader, 2018). A bifurcation may also describe how a small event at some critical point may define a long-term trajectory on a perceptual level. For example, changing just one note in a scale during a song may cause a transition in overall tonality to a great emotional effect. Viewed in terms of DST, one can say that the key is an attractor (the notes "gravitate" towards it), and that the subtle note change (a parameter variation) causes a bifurcation that makes a new attractor (the new key) appear in phase space.

3.1.2 The application of DST

DST offers a way to explain the emergence of certain macroscopic phenomena via the interactions of several microscopic elements (Mainzer, 2004). For instance, meteorologists may be able to predict the emergence of a hurricane system through careful measurement and analysis of changes in temperature, humidity, and air pressure at various locations. However, they cannot be certain of the hurricane's strength or its path; its behavior is unpredictable. This is because the elements that combine to create the hurricane have sensitive dependence on initial conditions. This sensitivity is famously exemplified in the *butterfly effect*—the idea that the flapping of a butterfly wing may cause a major weather event in another part of the world (Kaplan & Glass, 1995). DST is also well suited to explain and model qualitative changes in the behavior of systems. For example, at a very low flow rate, water may come out of a faucet as a periodic dripping at a certain rate. Increasing the flow, the dripping rate increases until it reaches a threshold, whereupon the dripping no longer happens in a uniform fashion, but still with a certain periodic pattern. Here, a bifurcation has occurred, and the phase space changes from having a fixed to a periodic attractor. As the flow increases further, the dripping will go through several bifurcations until the dripping becomes chaotic. DST has demonstrated that such qualitative transitions happen in all kinds of systems governed by nonlinear dynamics, and in strikingly common ways (Strogatz, 2015).

Musical interaction, and indeed any creative interaction, could be conceptualized as a nonlinear complex system, or more accurately, a psychophysical system of systems with both physical and mental components. As with purely physical systems, the interaction may be impossible to solve analytically by looking at variables in an isolated fashion. However, it is possible to study emergent behavior in the system as a whole with the aid of DST concepts. Musicians frequently use spatial terms such as "exploring spaces", "going places" and "finding the pocket" in ways that bear a fascinating resemblance to phase spaces, trajectories and attractors. Furthermore, certain modes may cause sudden transitions in the interaction that cause new ideas to form. Accounts of "sudden, radical shifts" and the "opening up of new spaces" sound much like the description of bifurcations. The premise of this thesis is that creative experiences through musical interaction may be explained in such terms. Sections 3.2.3 and 3.2.4 feature several examples where DST has been applied in similar research contexts.

According to Borgo (2005), the conditions for complexity requires two components: an *irreversible medium* and *nonlinearity*. The irreversible medium in the study of dynamical systems is time; systems evolve in time. The nonlinear aspect makes many complex systems impossible to solve analytically, but as Mainzer (2004) states: "Understanding complex dynamics is often more important for our practical behavior than computing definite solutions, especially when it is impossible to do so" (p. 1). Consequentially, DST often focuses on the approximation of phenomena through developing mathematical models. Elements deemed most important in the manifestation of phenomena are treated as variables or parameters in the models, and other minor factors are either ignored or treated as noise. Although the resulting models present simplified views of real-world phenomena, contemporary work with nonlinear dynamical systems have proven very useful in discovering qualitative features of systems and possible shapes of processes, even though the exact numerical values of parameters cannot be ascertained (Borgo, 2005).

3.1.3 Mathematical foundations of DST

Two main categories of mathematical equations are utilized to model the behavior of dynamical systems: *differential equations* and *iterated maps*. Differential equations model systems whose parameters are measured continuously (as curves), whereas iterated maps define systems that have been measured at sampled intervals (at discrete time steps).

The general form of using differential equations to calculate the future state of a dynamical system, as described by Mudd (2017), is

$$\dot{x} = f(x)$$

where:

- $x = (x_1,\ldots,x_k)$ denotes k components, considered as a vector in k-dimensional phase space,
- $f(x) = (f_1(x),\ldots,f_k(x))$ is an k-dimensional vector function of x, and
- $\dot{x}$ denotes the system's time derivative $\frac{dt}{dx}$ as calculated by the function $f(x)$.

For iterated maps, the function to calculate a future state takes the form

$$x_{(n+1)} = g(x_n)$$

where $g(x)$ is a k-dimensional function of g.

Both equation types share the characteristic that if one knows the exact initial parameters, it is possible to calculate all possible future states of a system. However, as models only approximate real-world phenomena, prediction is only practical for relatively short time scales, and are treated as estimates.

3.1.4 Examples of dynamical systems

This thesis does not cover the theoretical foundations of DST in depth. Readers who are interested in a comprehensive introduction may refer to cited literature in this section. However, I will provide two examples of dynamical systems: one based on an iterated map (a discrete system) and one based on differential equations (a continuous system).

The logistic map is often used to demonstrate the complex dynamics of population growth (Strogatz, 2015), given by the equation

$$x_{(n+1)} = rx_n\ (1{-}x)$$

where *r* denotes the intrinsic growth rate and *x* is the measure of the population in the nth generation. *r* is normalized to the range $0 \leq \times \leq 1$, and *r* is restricted to the range $0 \leq r \leq 4$, so that the equation maps into itself (always keeping *x* within its range). Although the equation is deceptively simple, it is a powerful demonstration of complex dynamics. When *r* is below 1, the population will always go extinct (drop toward zero) when the function is iterated, regardless of the initial value of *x*. When *r* is in the range $0 \leq r \leq 3$, x will settle to a steady state after several iterations. Any initial value of *x* will yield the same result; *x* is attracted to a fixed point. Therefore, we say that the system's phase space has a fixed-point attractor when *r* is between 1 and 3.

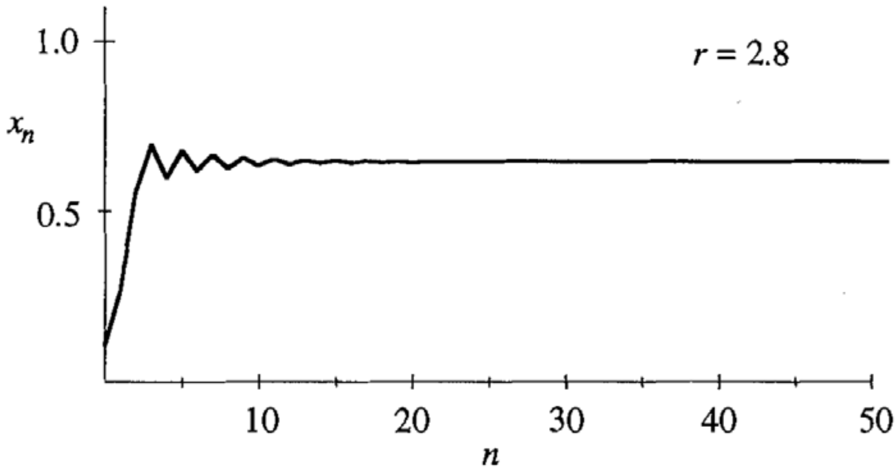


Figure 10. After several iterations, x settles to a fixed point of 0.643 when r = 2.8.[2]

2 Copyright (2015) From (Nonlinear Dynamics and Chaos) by (Steven H. Strogatz). Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc. This permission does not cover any third party copyrighted work which may appear in the material requested. User is responsible for obtaining permission for such material separately from this grant.

If *r* is increased above 3, the system's behavior starts changing. Instead of *x* settling to a fixed point, it ends up oscillating between two different values of *x*. Now the phase space is said to have a period-2 attractor.

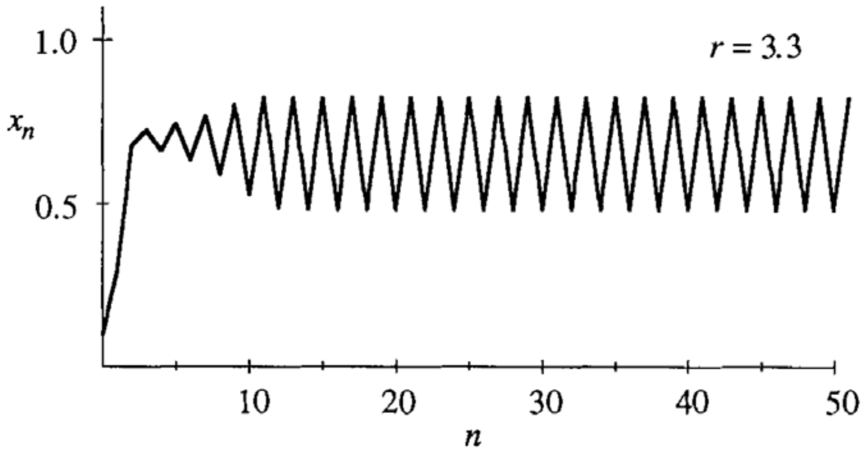


Figure 11. At r = 3.3, x ends up oscillating between the values 0.48 and 0.82. (See footnote 2)

Increasing *r* further leads the system through several bifurcations, each time doubling the amount of periods in the attractor (from 2 to 4 to 8, etc.). At *r* = 3.569946, the number of periods grows infinitely, and for many values of r, the sequence $\{x_0 \ldots x_n\}$ is completely aperiodic. We also find that two nearly identical initial values of x will cause completely different trajectories—all the properties of chaos are in display (see Section 3.1.1).

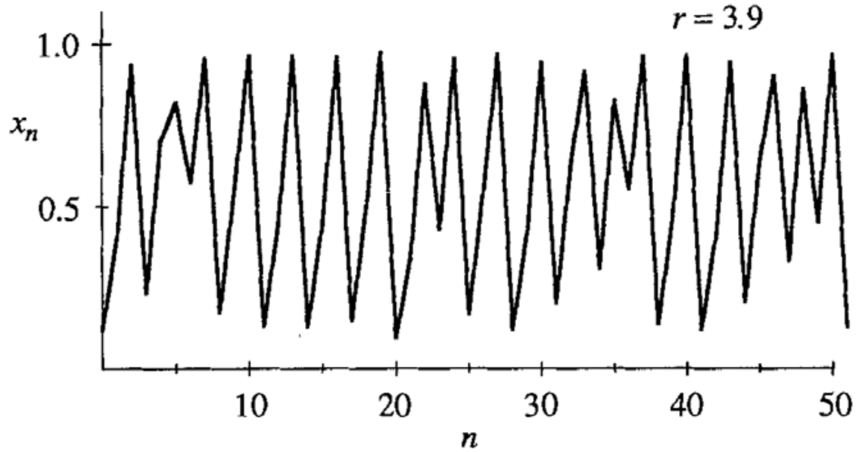


Figure 12. At r = 3.9, the system's behavior is chaotic. (See footnote 2)

We can see the period doubling and subsequent chaotic behavior as *r* increases in the bifurcation diagram below. The diagram shows the long-term behavior of all values of *r* at once. The diagram also shows there are several intervals of *r* where the long-term behavior temporarily goes back to periodic oscillations (the white gaps in the diagram). At $r = 4$, the system displays chaos in the full range $0 \leq x \leq 1$. Interestingly, if one were to zoom in forever on the chaotic parts of the diagram, one would find the same "forks" in miniature, ad infinitum. This self-similarity is a fractal property.

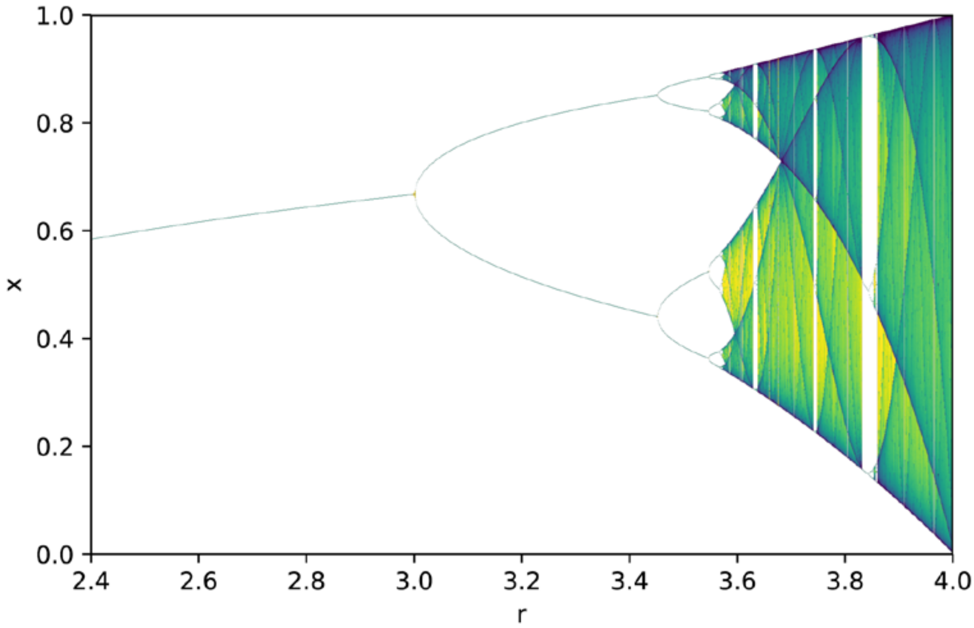


Figure 13. Bifurcation diagram. The lines trace the attracting points for any value of r between 2.4 and 4. Morn [CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0)]

The logistic map demonstrates that there is a difference between the nature of a process and the nature of the behavior it produces. The generating process can be simple and deterministic, but produce complex, aperiodic or chaotic behavior. It also shows that even the slightest change of a parameter can cause great leaps of behavioral change at certain transition points (bifurcation points).

The Lorenz equations is a set of differential equations devised by Edward Lorenz in 1963, and this provides our example of a continuous dynamical system. Lorenz was studying a simplified model of convection rolls in the atmosphere (circular displacement of air due to thermal conditions) in an attempt to account for the unpredictability of weather systems. This led him to the discovery of chaotic motion on a strange attractor (Strogatz, 2015). A strange attractor is so named because of its fractal dimension—instead of a point, a curve or a surface, the attractor has a fractal dimension between 2 and 3. The Lorenz equations

$$\dot{x} = \sigma(y - x)$$

$$\dot{y} = rx - y - xz$$

$$\dot{z} = xy - bz$$

where $\sigma, r, b > 0$ are constants describes a three-dimensional system with extremely erratic dynamics. At some values of σ, r and b, the system spirals into a fixed point (a fixed attractor), and at other values into various complex circular orbits (*limit cycles*). The most complex behavior, however, is what is known as the *Lorenz attractor*. The solutions to the equations in this regime are infinitely irregular, meaning that they are never exactly repeating. In spite of this, the trajectories are bounded into a narrow region of phase space. When plotted in three dimensions, these trajectories form a geometrical figure that is the visual representation of the attractor. As this is a three-dimensional representation projected into two dimensions, the trajectories are only seemingly crossing. In reality, the trajectories never cross, even when repeated infinitely.

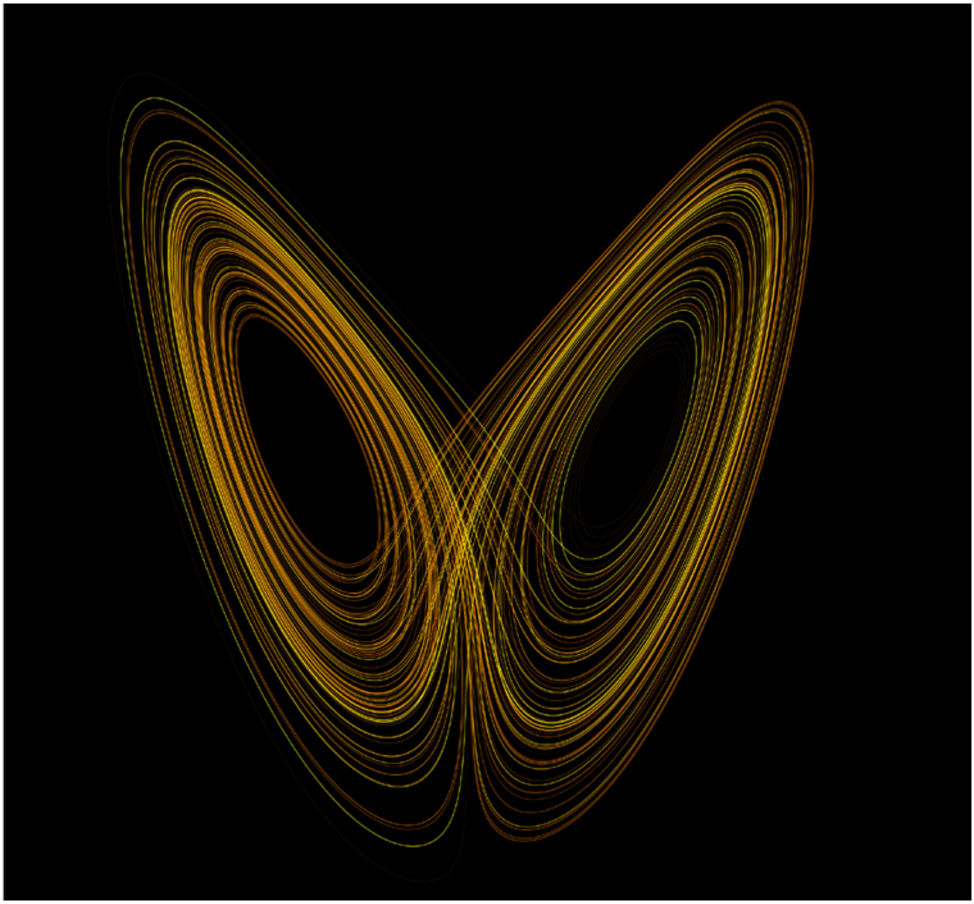


Figure 14. The Lorenz attractor. Wikimol, Dschwen [CC BY-SA(https://creativecommons.org/licenses/by-sa/3.0/)].

Another property of the Lorenz attractor is the system's extreme sensitivity to initial conditions when the Lorenz attractor is present in the phase space. Miniscule variations in the first set of x, y and z will yield widely varying trajectories as the functions are iterated, even though they are confined to the attractor. This apparently paradoxical dual display of order and chaos has been extensively researched. The Lorenz equations have later successfully been used to explain the dynamics of lasers and dynamos (Strogatz, 2015). The properties of this and other dynamical systems with chaotic behavior generated by deterministic functions have also been linked to aspects of music, such as the structures of waveforms, patterns of notes or events, or longer structural forms (Mudd, 2017).

3.2 A conceptual displacement: DST applied in other scientific domains

In the previous section, I explained the principles on DST as applied in its original field. In this section, I will present how DST has been applied in other fields, including in social sciences and music. However, I will begin with a discussion of what happens to concepts when they are transferred from one scientific domain to another.

3.2.1 Metaphors

In *Borrowed Knowledge*, Kellert (2008) conducts a comprehensive examination of what happens to scientific knowledge when researchers from the social sciences and humanities apply theory from the natural sciences for their purposes. His vantage point is the uses (and abuses) of chaos theory particularly in the domains of economics, law and literature, during and after the surge of interest the phenomenon of chaos received from the late 1980s and onward. As chaos theory is a subset of the science of nonlinear dynamics, Kellert's work has a significant bearing on the application of DST in this thesis. At any rate, the implications of transferring concepts, methods and results across disciplines outlined in *Borrowed Knowledge* is relevant for any cross-disciplinary enterprise.

While providing several caveats along the way, such as "tenuous leaps of reasoning" (Kellert, 2008, p. 5) or the "slavish imitation" (p. 13) of technical scientific language carrying the risk of dilettantism, Kellert's account is above all an acknowledgement of the virtues of academic borrowing. One central theme is the *metaphor*. When quantitative techniques from the source field cannot be applied, Kellert writes, the conceptual framework may be used metaphorically (Kellert, 2008). Hence, just as Newtonian physics has proven useful to explain forces and counterforces in society, reconceptualizations of the physical world may lead us to the restructuring of knowledge in fields outside of the natural sciences. Indeed, science from the source field need not be in any way new or revolutionary to structure or restructure knowledge in the target field. Donald Schön (1993) illustrates the generative function of metaphors in the description of a group of industrial designers discovering that a paintbrush is "a kind of pump". Frustrated with how the design of a paintbrush with synthetic bristles tended to spread the paint unevenly, the pump metaphor led the group to gradually transform their perception of how a brush works. The space between the bristles became "channels" and the angle of bend accounted for the pumping pressure. Using this knowledge, they could refocus their attention to details that would ultimately lead to a better-designed product. Kellert (2008) classifies such as view of metaphor as the *interaction view*. Rather than regarding metaphor as merely a trick with words (the *dismissive view*) or a device to attract attention to the similarity between things (the *comparison* view), metaphors can have a distinctive cognitive function and content, which cannot be replaced by a literal statement of comparison (Kellert, 2008, p. 105).

Kellert proposes the following organization of criteria for evaluating metaphorical borrowings:

1. *Fit*—how well the elements and relations of the source field can be mapped onto the target.

2. *Utility*—how well the metaphor performs its structuring or restructuring function.

3. *Need*—whether the metaphor serves a purpose otherwise unmet.

4. *Awareness*—whether the metaphor pays explicit attention to its own figurative status. (Kellert, 2008, p. 126)

Kellert notes that fit and utility in some sense are competing criteria. According to the interactive view, a cognitive restructuring in the target field can only take place if the original idea is somehow displaced. In other words, overfitting may reduce the utility of a metaphor. In emphasizing this he leans on Kittay (1987), who claims that a metaphor is rendered decorative rather than cognitive if the match is too easy. A good metaphor involves an interaction between the source and target fields with important differences in their structure, but where "many of the affinities and oppositions are carried along in the transfer of meaning" (Kittay, 1987, p. 154). In regards to utility, an effective metaphor may also provide further implications in the target field, allowing for extensions beyond concepts in the source field. Yet, Kellert warns that "metaphors with adequate fit and reasonable utility may still be utterly unnecessary" (Kellert, 2008, p. 144). A dispensable metaphor may not serve an adequate purpose. A metaphor is only necessary when "B allows us to learn about A in ways that could not be accomplished without modeling A as B" (Kirchhoff & Matheson, 1997, p. 41).

I have elaborated on the functions and implications of cross-disciplinary knowledge transfer and metaphors, because the usage of DST as a theoretical framework in researching musical interaction is relatively unconventional. Looking back at Section 3.1.1, it is clear that the terminology I borrow from DST qualifies as a metaphorical remapping when applied to the domain of musical interaction. In light of Kellert's list of criteria, I have already put forward some arguments with regard to the fit between the source and target domains in the terminology section earlier in this chapter. The utility and need of DST consist in providing a holistic view of collective music making not afforded by other models such as the ones presented in Chapter 2. DST cannot account for all aspects of musical interaction, but it is apt when focusing on the theme of this thesis—the dynamics of change, initiative taking and creativity. Creative musical interaction is not a linear process. Sometimes, everything seems to "fall into place" and new ideas may seem to "jump out at you". Other times, the interaction seems to become "stuck", and there is no "flow". In flash moments of inspiration, a small contribution from one musician may cause a large thematic shift in the overall interaction. Concepts from DST may account for these dynamics in a coherent and intuitive way.

The aspect of intuitiveness brings me to the concept of the *cognitive metaphor*, which is an even more fundamental kind of metaphor than the interactive type of metaphors advocated by Kellert. According to the *cognitive metaphor theory*, a basic property of human cognition is the ability to understand one conceptual domain in terms of another (Lakoff & Johnson, 2003). This is referred to as *metaphorical projection*. The directionality of projection tends to go from concrete and physical concepts to more abstract concepts. Lakoff and Johnson suggest that all cognition relies on many layers of metaphorical projections, with roots in sensorimotor experiences. We feel our thoughts viscerally first, and language comes after. This adds an important dimension to how we may think about musical interaction in terms of DST. It allows us to use DST, not to *describe* musical interaction objectively, but we may use phenomenological accounts of people's experiences of musical interaction and claim that these experiences create associations to manifestations of dynamical systems in the physical environment—it *feels like* a dynamical system. When saying, for instance, that a chord is "pulling" the listener toward the release of tension in the tonic, we are projecting the concept of a physical pulling force to the more abstract concept of an "unstable" chord transitioning to a more "stable" chord.

Thus, applying the framework of complex dynamical systems to musical interaction is, indeed, a metaphorical projection of measurable auditory features in the music onto the listener's experience of the dynamical forces in the musical interaction. In such an account, we are not only describing a system from the outside. We are giving agency to the trajectory itself—the trajectory being the subjective experience of being a part of the interactive system. This is referred to as an *endopsychological* view: the observer is regarded to be a part of the system observed (Masterpasqua & Perna, 1997). An *exopsychological* view, on the other hand, would require a quantitative approach and disregard the subjective experiences of the participants of an interactive system.

3.2.2 DST in social sciences: economics and psychology

It is interesting to trace the degree of figurativeness in the ways DST is applied in various fields. Early adoptions of chaos theory occurred in the field of economics—arguably one of the social sciences closest to the natural sciences in terms of methods and presentation of results. Several economists began seeing chaos as a way to explain seemingly random fluctuations in stock markets, and the mathematics of nonlinear dynamics was used directly to create theoretical models demonstrating these dynamics (Kellert, 2008). Until the 1980s, unpredictable economic fluctuations were typically modeled as *exogenous shocks*—factors arising outside of the economic system and imposing turbulence. However, markets tend to fluctuate also in the absence of any apparent outside influence, and this left some economists unsatisfied with existing models. A string of publications in the wake of the popularization of chaos theory presented economic models with strange attractors that could account for complex behavior "on the basis of internal forces alone, with no resort to unexplained outside influences" (Day, 1993, p. 28). Such claims sparked some controversy regarding the real-world implications of the theoretical results (Kellert, 2008). Randall Bausor (1994) pointed out the lack of empirical evidence, noting that there is no laboratory that allows one to examine the behavior of entire stock markets or national economies. Economists must go "straight to the wild" of real-world economics—equivalent to telling a student of fluid dynamics to "begin with Niagara" (Bausor, 1994, p. 123) instead of a controlled experiment in the research lab. Bausor concluded that all models of economic chaos remain "wholly metaphorical" (p. 144).

Over time, the gap between theory and real-world data has made economists less inclined to prove the existence of chaos and strange attractors in economic markets. However, by now economists generally accept that fluctuations are due to both nonlinear internal forces and exogenous shocks. Thus, theory behind nonlinear dynamics is at least partly responsible for bringing about a methodological shift in economics, with less reliance on modeling randomness in terms of external shocks, and a general redirection toward nonlinear economic dynamics (Kellert, 2008, pp. 89–90).

The "chaos wave" visited several other fields. In legal theory, for example, Reynolds (1991) compared the decision-making patterns of the US Supreme Court with a chaotic system, and suggested that this and similar analogies may be a useful in "clearing away some cobwebby ideas" in legal thinking that "remains thoroughly rooted in a nineteenth century linear determinism" (Reynolds, 1991, p. 116). Even more than in the case of economics, however, the early optimism for the applicability of nonlinear dynamics and chaos in various fields gave way to a gradual realization that analogies alone are not enough: qualitative theories and quantitative methods are no easy match when applying DST outside of the natural sciences. This is an important lesson for any field vying to apply DST as a theoretical framework.

One field where researchers have continued to apply DST is psychology. A possible explanation for this could be that theoretical models in this field are fundamentally figurative, making the adoption of DST less controversial than in "harder" scientific fields. Furthermore, the results of psychological undertakings—therapeutic outcomes, management and team building models, design, cognitive heuristics, etc. are often evaluated qualitatively. This makes the theory less vulnerable to empirical discredit. I venture the adoption of DST is more effective in psychology for a third (and possibly the most important) reason: DST harbors a complete set of concepts that seems to resonate with the generative aspects of human perception—especially concerning change or perceptual shifts. A thematic line may be drawn from the Gestalt principles of perception established a century ago, which demonstrated the generative aspects of human visual perception in a variety of different ways. The principle of *multivariance*, for instance, is famously exemplified by Rubin's vase, which may also be perceived as two faces in profile according to which colors are seen as background or foreground. The switching of percepts happens abruptly and can be described as a bifurcation: the function of the vase as an attractor suddenly changes to the faces. The concept of perceptual attractors is illustrated rather elegantly by Tschacher (1997) in the figure below.

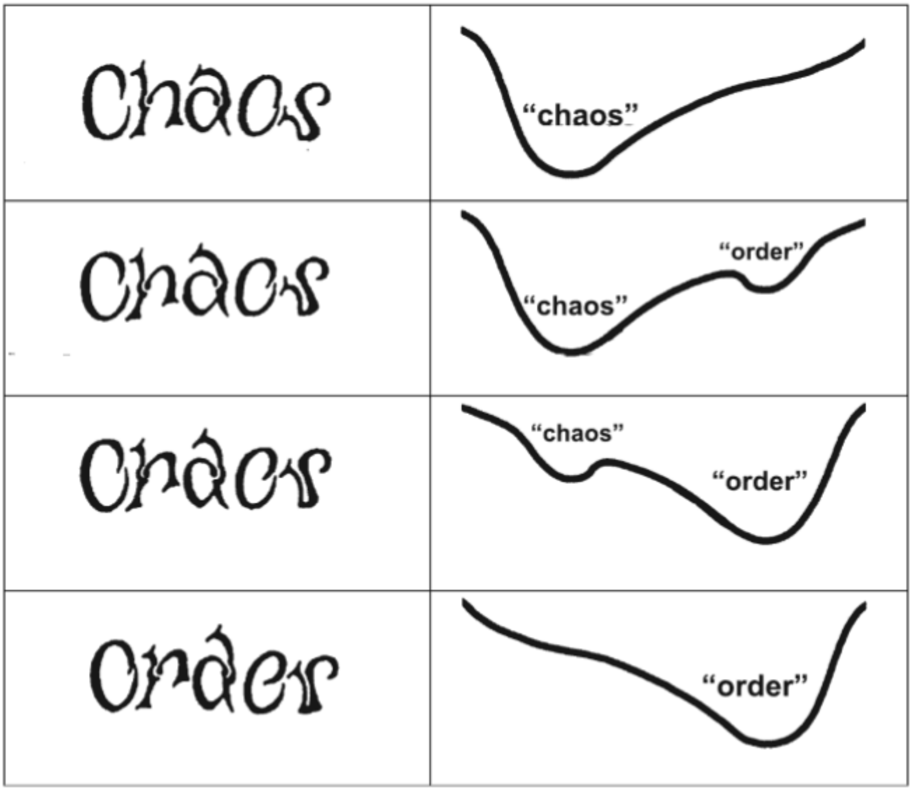


Figure 15. Left: the transition between two perceptual attractors. Right: the landscape of potential attractors. Copyright © 1997 by Hofgrefe-Verlag. Reproduced with permission. (Tschacher, 1997)

In Figure 15, the depictions of "chaos" and "order" represent two different perceptual attractors. As the features of "chaos" are gradually altered, a bifurcation occurs and "order" suddenly emerges as the new attractor. To the right of each word we see troughs representing the attractors in what Grawe (2004) refers to as "the landscape of potentials" (p. 372). This brings to mind Borgo's (2005) characterization of phase space as a system's "geometry of possibilities". In this thesis, the notion of perceptual attractors is applied in the context of musical interaction, and refers to musical sequences that musicians are attracted to in an interactive setting. Going forward, the perceptual aspect is implied, and such sequences will simply be referred to as attractors.

Tschacher and Scheier (1997) neatly tie together DST and complexity theory in the approach of *synergetics*. We recall that the premise of a dynamical system is one of a simple process producing complex behavior. Obviously, biological systems are the opposite of simple. The synergetics approach proposes that complexity theory may account for perceptual pattern formation: the human mind self-organizes and produces emergent patterns out of the highly complex psychosocial and cognitive-emotional systems within which it is embedded. Through this self-organization, a set of macroscopic "order variables" are established. A clear distinction between the process of dimensionality reduction taking place in self-organization and the subsequent perceptual dynamics is extremely important. DST can only account for the latter, because low-dimensional dynamics occurs after self-organization has taken place: "Only the macroscopic level of complex psychological systems is accessible to observation; macrodynamics constantly emerges from microscopic complexity" (Tschacher & Scheier, 1997, p. 274).

The synergetic relationship between self-organization and the dynamics of representational phenomena is also the topic of two studies by Stephen et al. (2009). In these studies, participants were asked to solve gear-system problems by predicting the rotational direction of a target gear based on the directionality of a driving gear, with several interlinking cogs. The gear systems were presented as static images on a computer screen, so participants had to solve the problems analytically. At first, participants would solve the problems by manually tracing the force across the systems. After solving a few problems, they spontaneously discovered that the gears form an alternating sequence, and they subsequently dropped any references to the physics of the system (Stephen et al., 2009, p. 1814). Through video recordings, an intricate coding system and time-series analysis methods, Stephen et al. quantified the transition between the two different methods of solving the problems (finger tracing vs. rhythmic pointing to each gear). The studies showed that the lead-up to the "penny-drop moments" when the participants discovered the more effective method of alternation (a new attractor) was accompanied by a breakdown of the finger-tracing regime. This was measured in levels of entropy. Results consistently showed that a peak in entropy (indicating critical instability), followed by a sharp decrease in entropy (indicating the onset of a new attractor), predicted the discovery of alternation (Stephen et al., 2009, p. 1811). According to complexity theory, such a peak in entropy is consistent with the principles of self-organization. The second study also showed that introducing random perturbations on the computer screen (making finger-tracing more difficult) led to an earlier discovery of alternation, i.e. the process of self-organization speeds up with the introduction of entropy. These findings are consistent with Skarda and Freeman (1987), who postulated that the brain must be driven into a chaotic state in order for new patterned activity to take place. In a journal article titled "How brains make chaos in order to make sense of the world", the neurological underpinnings of the process of discovery and the role of self-organization in learning is thoroughly discussed within the framework of nonlinear dynamics. Their research on the recognition of odor suggests that if a new odor occurs with no reinforcement, it goes unnoticed because it falls into the basin for familiar background odors; habituation takes place. With reinforcement, however, the system is forced into a "chaotic well" which "enables the system to avoid all of its previously learned activity patterns and to produce a new one" (Skarda & Freeman, 1987, p. 171).

The above are excellent examples of the utility of DST in the context of researching creativity. The process of brainstorming encourages the suspension of judgement and welcomes wild ideas (Osborn, 1953). In ideation workshops, "thinking outside the box" is a virtue. When musicians jam or improvise, it is common to introduce elements that force the collective into unfamiliar or challenging territory, increasing the chances of new discovery. All these activities are motivated by the same objective: bringing new ideas to fruition. In the next sections, we will finally see DST applied in research on musical interaction and interactive music systems.

3.2.3 DST in music

An early example of a dynamical systems approach to musical composition was presented by Beyls (1991). Beyls suggested that complex dynamical systems are an alternative to the traditional approach of creating musical structures according to explicit rules. Noting the contemporary vogue for expressing fractals as a static generator of visual design, Beyls suggested that the zooming in and out of details in these popular images could be viewed as a metaphor for the exploration and discovery of emergent patterns in complex dynamical system of interacting agents. The emerging patterns are a product of self-organization according to a set of generative principles. "Composition becomes navigation in attractor fields, the interactive conversational exploration of levels of stability and sensitivity" (Beyls, 1991, p. 15). Beyls' approach to composition described here aligns with the theoretical models of perception presented in the previous section to a remarkable degree, and clearly demonstrates the creative potential of DST in artistic contexts.

In *Sync or Swarm*, David Borgo (2005) explores the complex dynamics of musical improvisation and group creativity through the twin lenses of dynamical systems and complexity theory. He draws attention to that nearly all analytical terms for music describe emergent properties:

> Harmony and rhythm describe qualities that emerge as tones and silences are combined simultaneously and in succession, and melody appears to be an emergent phenomenon that draws on harmony, rhythms, contour, and other musical and cognitive dimensions as well. Our difficulties in approaching the emergent qualities of music in general, and of improvised music in particular, are not dissimilar to those of traditional physicists who tended to steer clear of the complexity that is readily apparent in daily life. (Borgo, 2005, p. 65)

Emergent musical forms as experienced by a listener give rise to what Borgo terms as *qualia*—phenomenological units of experience (Borgo, 2005, p. 66). Qualia with high degrees of *salience* (experienced significance) coupled with particular musical sequences can become attractors in the phase space of improvisation, especially if reinforced through repetition (the saliencies of qualia will gradually decay if left alone). Borgo contends that contemporary improvisers tend to favor "strange" musical attractors to those that rely on cyclic or predictable interactions. High complexity regions, he argues, yield a higher return of new patterns. Improvisers "surf the edge of chaos" to ensure continual development and excitement throughout the improvisation (Borgo, 2005, pp. 74–75). We recall that one of the defining behaviors of strange attractors is sensitivity to initial conditions. Borgo carries this notion into the phase space of musical improvisation. The initial conditions of an improvisation are not, as one would perhaps assume, the initial musical gestures of the performed piece. Rather, the initial point is "already implicated by feedback processes in a complex network dynamic" (Borgo, 2005, p. 72) involving the personal backgrounds, moods and musical abilities of all the performers involved and the context of the improvisation. Hence, the complexity of this dynamic quickly reveals itself as the system reiterates, and the smallest variations of input may yield widely different performances. Each musician explores his or her own phase space, and the group phase space is the combination of these (Borgo, 2005, p. 70). Bifurcations are characterized by the branching off from a distinct segment of an improvisation to a new segment. This happens when current phase space "buckles" under the collective influence of the performers' experimentations on the "edge of chaos". A critical point is reached, the regime collapses, and new attractors may emerge.

In addition to the qualitative analytical approach described above, one chapter of *Sync and Swarm* is also dedicated to quantitative time-series analyses of various improvisation pieces (Borgo, 2005, pp. 83–122). The method involves mapping the fractal dimension of audio recordings of improvisation pieces and plotting the results as graphs on a horizontal time axis. The fractal dimension corresponds to the relative sonic complexity of the sound signal, based on three subsystems: harmonic overtone components, inharmonic frequencies and large amplitude modulations (Borgo, 2005, p. 83). Borgo used these fractal dimension graphs to explore the level of correspondence between sharp transitions in fractal dimension and phenomenological accounts of experiencing these sequences from a listener's perspective. Borgo found that abrupt changes in fractal dimension are often accompanied by dramatic shifts in perceived complexity, related to experiences such as tension/release or less/more interest.

The collective dynamics of improvisation is also the research focus of Canonne and Garnier (2011; 2012). Starting with a fully theoretical approach, they proposed a model for *collective free improvisation* (CFI), a form of improvisation they define as *referent-free*—i.e. having no underlying formal scheme to guide the musical progression on an intermediate time scale (Pressing, 1984). The mathematics behind the model is rather involved, and I will not present the details here. In summary, Canonne and Garnier (2011) use a system of equations to model the non-linear dynamics of intention evolving on a short time-scale and objective forming on the long time-scale. Several cognitive-emotional phenomena are quantified and used as parameters in the model (*self-sensitivity*, *cognitive load* and *"boreness"*, i.e. the level of disinterest a musician displays toward a particular sequence), as well as the level of influence and degree of dependency/independency between the musicians. Experiments with the model showed two types of collective behavior in CFI:

- *Collective sequence*: a stable behavior corresponding to a fixed point attractor in the phase space of the system
- *Phase of discoordination*: an oscillating behavior corresponding to a limit cycle or the absence of a fixed attractor in the phase space of the system

The results also demonstrated that self-organization (emergent and sustained structures) of the collective was possible despite the absence of a priori structures. Interestingly, the model shows a spontaneous alteration of musicians taking the lead on the intention scale, indicative of a natural dynamic of initiative turn taking. A subsequent paper (Canonne & Garnier, 2012) reported on a study made with free improvisers. The participating musicians self-reported segments deemed as collective sequences (stable and interesting), and were asked to pay particular attention to strategies they employed in the articulation (transitions) between the segments. The authors identified two main challenges facing free improvisers: "the first one is about establishing and/or identifying an attractor (convergence problem), while the second one is about the transition from one attractor to another (articulation problem)" (Canonne & Garnier, 2012, p. 7).

DST has also been used as a framework for the study of gestures in music. In his PhD thesis, Demos (2013) examined the intricate relationship between ancillary body movements of performers, their expressive intent and sonic features in the music within a nonlinear dynamical systems framework. Ancillary body movements, as classified by Jensenius et al. (2010), are gestures that do not directly produce sound, but accompany sound production in ways that are linked to musical expression, including postural sway, flourishes of the hands, head movements, and more. Through several time-series analysis methods from DST, Demos produced experimental evidence to suggest that the systematic correspondence between the movements of the performer and the musical structure is more complex than a 1:1 correspondence between a particular musical idea and body movement (Demos, 2013, p. 216). Demos incorporated a synergistic view of the motor system, using it to argue that the distinction between ancillary and sound producing movement is arbitrary (Demos, 2013, p. 45) and context-dependent.

Another PhD thesis dedicated to a nonlinear dynamics approach to music was published by Mudd (2017). The theme is the interactive relationship between musicians and their musical instruments/interfaces, and Mudd employed both quantitative and qualitative research methods to explore these relationships. In the quantitative part, Mudd conducted comparative surveys among musicians interacting with four different digital interfaces with both linear and nonlinear mappings between the inputs and outputs. Results showed that interfaces with nonlinear dynamics increased the sense of surprise (the good kind) and the ability to explore new modes of expression. Notably, there did not seem to be a link between nonlinear interfaces and the lack of control or inability to repeat sequences. This suggests that although interfaces with nonlinear dynamics may lead to complex or chaotic behavior, they are not inherently uncontrollable (Mudd, 2017, p. 164). The qualitative part of Mudd's thesis is devoted to an ethnographically informed study of musicians engaged in improvisatory practices. This offers a different perspective on some of the findings from the comparative surveys. In particular, the qualitative study yields rich details in regards to exploration and surprise. Interviews with musicians revealed two main approaches to exploration when interacting with musical instruments or tools: *Divergent approaches* are associated with "novelty generation" and characterize an outward *problem finding* search, whereas *convergent approaches* constitute a narrowing in on optimal "solutions", or *problem solving*. This bears resemblance to Canonne and Garnier's (2012) dichotomy of convergence (focusing on an attractor) and articulation (transitioning between attractors). Undoubtedly, the divergence approach is also related to Borgo's notion of "surfing the edge of chaos" in search for new attractors to emerge in the phase space. Concerning surprises, Mudd identified two main kinds. *Combinatorial surprises* refer to unexpected effects of known elements when combined, for example two sounds merging to create something surprisingly expressive. This model of surprise is associated with compositional aspects of the engagement with the instrument or interface. *Interactional surprises*, on the other hand, emerge directly in the interaction with the instrument or tool, and apply to unforeseen effects of the interaction itself—the instrument "pushes back". Mudd's thesis is a good example of Kellert's (2008) notion of the utility of knowledge transfer. Instead of applying terminology from DST to a "slavish" degree, Mudd keeps the conceptual structure in place while extending the theoretical framework with additional concepts discovered in the empirical data.

A final example in this section about DST in music research is van der Schyff et al. (2018). They explore musical creativity from the perspective of embodied cognition, and suggest that DST provides useful tools for research and theory that align closely with this perspective (van der Schyff et al., 2018, p. 1). The authors devote some space to explain how the phenomenon of creativity has developed through history, from being viewed as an inscrutable external (divine) force, off-grounds from scientific query, into now being conceived of an emergent process in the context of complex and distributed systems of interactions, with unpredictable outcomes and moment-to-moment contingency. This view classifies creativity as a collaborative phenomenon, highly susceptible to the actions of each participant and the interactional context (van der Schyff et al., 2018). With this vantage point, van der Schyff et al. propose a "4E approach" (*embodied*, *embedded*, *enactive* and *extended*) where "the mind may be understood as an emergent property of organism-environment interactivity, which involves biological, non-biological, cultural, social, technological, and historical dimensions" (van der Schyff et al., 2018, p. 7). The DST perspective is proposed to account for musical systems as dynamically self-organizing phenomena in the context of such complexity:

> In doing so it reveals aspects of the system that tend to converge and diverge as patterns of relative stability and instability. These are referred to as *attractors* and *repellors*, respectively, and are often represented on a topographical space, or a *phase portrait*. (van der Schyff et al., 2018, p. 8)

Further, the authors refer to Stephen et al. (2009), presented in 3.2.2, who documented spikes of entropy just before the "ah-ha!" experience associated with moments of understanding (van der Schyff et al., 2018, p. 8). Fruitful bifurcations that yield such moments are the result of divergence—a key factor in musical creativity. Divergence is the introduction of entropy to the system, either because of the willful activity of an agent or due to perturbations in the environment that the agents must deal with (van der Schyff et al., 2018, p. 8).

3.2.4 DST in interactive music systems

In review, one of the most remarkable aspects of the adoption of DST in music research is its versatility. DST has been used to research composition, improvisation, musical body movement, performer-instrument interaction, and musical creativity. Despite the wide range of application areas, the frameworks are impressively coherent. This shows that DST may be a promising framework for researching and developing interactive music systems. The advantages of the framework for this purpose includes the possibility of correlating quantitatively measurable musical parameters with qualia experienced by musicians, and terminology that lends itself to a generalization of musical actions and perceptions, which frees the discussion from associations to specific genres. Furthermore, a DST-inspired modelling approach may translate well into programming code and machine learning techniques due to the theory's foundation in mathematics and the underlying deterministic nature of the processes involved. However, out of all the interactive music systems that have been developed and presented through peer-reviewed publications over the past few decades, only a few have explicitly promoted a dynamical systems approach.

Peter Beyls' pioneering work with the development of interactive music systems since the 1980s was mentioned in Chapter 2. As seen in the current chapter, he was also one of the first to promote a dynamical systems approach in music composition. It is therefore not surprising that he, indeed, does incorporate this approach in his work with multimedia arts and interactive music systems. One of his most recent creations is *Pock* (Beyls, 2018), a system designed as a fully interactive improvising duo partner. He invokes the enaction paradigm as a conceptual point of departure: interaction between agents and their dynamic environment gives rise to autonomy and cognitive functionality. The system's machine learning algorithm is founded on this principle, and involves a reinforcement learning strategy where rewards are derived from the implied motivation of the human interactor. Two contrary dimensions of motivation are tracked by the algorithm: *integration* (aiming to connect with the machine generated material) and *expression* (independent activity) (Beyls, 2018, p. 238). The machine listening module (called The Ear) infers the level of these motivational dimensions by calculating the consecutive musical distance (i.e. melodic similarity) between the human and machine. Depending on the current context and the motivation index, a "reference agent" is recruited from a pool of "potential agents" (individual complex MIDI player objects with access to a large library analysis and processing functions). A range of high-level system parameters allows the user to customize the system into particular behavioral niches:

- *connectedness-level*: how much the system should be influenced by the human input
- *responsiveness-level*: how much and how fast the system should respond
- *continuity-level*: the degree to which the system should generate new material vs. reinjecting variations of existing material
- *complexity-level*: the number of parallel agents actively generating material
- *autonomy-level*: how independently the system should behave in the absence of human input
- *learning-factor*: relative value of delayed vs. immediate rewards
- *exploration-rate*: likelihood of deviating from the currently optimal policy (i.e. how "disloyal" the system should be to the above parameters) (Beyls, 2018, p. 241)

The latter can be compared with a divergence approach, whereas several of the other parameters, especially connectedness and responsiveness are more closely associated with convergence. For more technical details about the machine learning algorithm in Pock, the reader is referred to the original paper.

Blackwell et al. (2012) propose that one of the strengths of an open dynamical systems framework for *live algorithms* (autonomous machines that interact with musicians in an improvised setting) is its ability to adapt to an unknown input due to its independence from a priori explicit rules. True autonomy in a system, the authors argue, requires not only output that is reactive, but also the ability to *negotiate*:

> A system that negotiates constructs an expectation of the collective musical output and attempts to achieve this global target by modifying its output. Since the collective musical output depends on the performer as well, negotiation, as the name suggests, may involve attempts to manipulate the behaviour of the performer, or equally, to adjust one's expectations in light of the direction of the music. (Blackwell et al., 2012, p. 163)

They propose a so-called *PfQ model* as a way to approach the design of live algorithms. *PfQ* is shorthand for a modular approach to designing computer music systems, where the modules are *P* (listening/analysis), *Q* (performing/synthesis), and *f* (patterning, reasoning or even intuiting) (Blackwell et al., 2012, p. 152). Viewed as a dynamical system, the parameter stream from *P* (the listening module) constitutes the state of "the outside world" (i.e. the iterated input to the dynamical system), whereas the internal, hidden functioning of *p* (the patterning module) constitutes the trajectory or current state in the phase space of the system.

An example of the implementation of this theoretical model is *Zamyatin*, an interactive music system designed by Oliver Bown (2011, 2018), one of the three authors above. Also conceived of as a duet partner in a free improvising context, the system requires no control by an operator during a live performance. It engages autonomously in interaction and development during the piece, although compositionally the decision-making process is flavored by the author's aesthetic preferences, to his own admission (Bown, 2018, p. 41). Zamyatin is not implemented with sophisticated statistical models of the performer's actions, does not track note sequences, tempo or meter, and has no machine learning algorithm or musical rules programmed into it. It simply tries to find interesting pattern-producing couplings that satisfy two constraints: that it stimulates the interaction in a way that makes it seem to have some kind of agency, and that it can be easily adapted by the performer creatively (Bown, 2018, p. 41). Thus, instead of emulating human musicianship, Zamyatin is more like a simple artificial organism:

> [...] a complex non-linear system that embodies dynamic properties that make for musically engaging interaction. One can describe Zamyatin's behaviour by appeal to cybernetic principles of dynamic systems, specifically the liquid analogy used in liquid state machines or the liquid brain model: like a bucket of water, the system has a number of natural resting states when it is not being driven by an input. But when something stimulates it it jumps into action, rippling and oscillating, before settling down again, possibly in a new state. The system might resonate with its input or jump into different dynamic modes. (Bown, 2018, p. 41)

The system extracts low-level audio features from the input (the *P* module in the *PfQ* model), and sends this into different dynamic models (the *p* module). The dynamic model is chosen from a pool which the composer can search by trial and error. This is similar to the "pool of potential agents" described by Beyls (2018), although in the latter case the recruitment of agents is done by the system with the aid of a machine learning algorithm. By the author's own admission, it remains a challenge for Zamyatin to "lock into a moment" during performance, i.e. convergence to attractors are not easily achieved.

In the few examples of interactive music systems that use a DST framework provided in this section, I have not gone into technical detail. My objective has been to show how the theoretical framework has been invoked to build design concepts. The number of ways these concepts can be implemented is vast, ranging from collections of simple models to complex multi-agent systems amounting to a "society of mind" (Minsky, 1986). While later chapters are devoted to the technical side of interactive music systems, this chapter is solely focused on theory and design concepts.

3.3 Summary

In this chapter, I have introduced DST as a conceptual framework for collective music making. DST can help us glimpse something fundamental and universal about many kinds of dynamics—whether pertaining to physical, social or cognitive systems. Musical interaction is a system of systems with elements of all these. I devoted a section to discuss the implications of borrowing theory from another scientific domain with a particular focus on the concept of metaphors, and argued the usefulness of DST in this thesis consists in a remapping of concepts that provides holistic view of collective music making while maintaining an awareness of the framework's figurative status. I presented several examples of how DST has been applied in social sciences, in particular in economics and psychology. Finally, I reviewed cases where DST has been applied in a variety of ways in music research in general and interactive music systems in particular. Combined, this is a useful backdrop for the next chapter, where I will explain how derivations of DST ended up as useful frameworks for several studies covered in Chapters 5–8.

4 Methodology

In Chapter 1, I framed the main research question and the sub-questions of this thesis from the vantage point of human-computer interaction (HCI). In the review of HCI in Chapter 2, I emphasized the interdisciplinary nature of HCI and the often cyclical nature of artifact design. The chapter offered a literature review of themes that all have a bearing on the research problems in this thesis, including interactive music systems, artificial intelligence, agency, creativity, and music making. Chapter 3 offered a presentation of dynamical systems theory. Although the source field is physics, I made the case that the conceptual framework it offers can be useful to theorize about musical interaction—particularly in the context of human–computer interaction in interactive music systems.

In this chapter, I present a disciplinary triangulation framework that makes it possible to weave a comprehensive tapestry out of all of these threads. I start with a discussion of the intricate relationship between scientific research and interaction design, and identify Research through Design as an appropriate approach to bridge the two domains in a progressive manner. I subsequently describe how and why I settled on a methodological design based on four studies with an open-ended structure. The four studies are briefly outlined in this chapter. The details of each study are presented in subsequent chapters.

4.1 Research through Design

When it was still a nascent field in the 1980s, HCI was invested in the cognitivist paradigm of modelling the human user in information processing terms—also referred to as *cognitive ergonomics* (e.g. Long & Whitefield, 1989). For example, *GOMS* modelled the user's cognitive structure as "a set of *Goals*, a set of *Operators*, a set of *Methods* for achieving the goals, and a set of *Selections* rules for choosing among competing methods for goals" (Card et al., 1983, p. 140). The models deduced through theorizing about human minds were meant to yield methods by which to develop efficient, task-oriented computer software and hardware. By the end of the 1980s, several researchers began pointing out issues with cognitive models as the basis for interaction design. In the words of Carroll et al., "the ambition of theory-based design in HCI has been frustrated to a great extent" (J. M. Carroll et al., 1991, p. 74). Notably, there were very few examples of prevailing theory having any significant impact on design. Barnard (1991) attributed this lack to the "two-way conceptual traffic" required to bridge between the physical world of people and computers on the one hand, and the deduced "science representations" of cognitive models and process schematics on the other. Building upon

Long (1989), he proposed that any representation in the science base can only be mapped to and from the real world through "intermediary" representations, and that these representations are different in kind depending on the directionality of the mapping:

> [The] representations called upon for the purposes of software engineering will differ from the representations called upon for the purposes of developing an applicable cognitive theory. (Barnard, 1991, p. 104)

Barnard pointed out that the life cycle of theoretical inquiry and synthesis tended to lag behind the life cycle of products with which it seeks to deal (Barnard, 1991, p. 106), essentially rendering the models outdated or even irrelevant upon arrival. J. M. Carroll et al. (1991) called attention to the fact that deduction from scientific principles has seldom played a major role in technological innovation. Rather, they argued, technologies evolve through a process of development and redevelopment, and the tacit knowledge and skills embedded in design communities form a complex and dynamic relationship with artifacts that embody theory. Seen through this lens, a designed artifact is a "theory nexus" with emergent theoretical properties (J. M. Carroll & Kellogg, 1989). The contributions from J. M. Carroll and his colleagues were not intended to relegate theory to a secondary role in artifact design, but the notion of artifacts as "latent theory" sheds light on the tight coupling between the two. It also flips the conventional view of research methodology as a transfer of knowledge from theory to its application.

In hindsight, the insights shared by J. M. Carroll et al. read as an obituary for first wave HCI and its focus on the task-artifact cycle. The zeitgeist was changing, and second wave HCI brought with it a "turn to the social" (Rogers, 2012), as described in Chapter 2. However, the theorylike role of artifacts has reemerged several times later, particularly in third wave HCI. According to Magnusson (2009), for instance, systems of knowledge and representations are inscribed into technological artifacts—they are "epistemic tools" that "influence our thinking due to intrinsic mechanisms that can potentially be defined as cognitive" (p. 36). Leaning on the work of Baird (2004), Magnusson establishes the possibility of "thing knowledge", referring to artifacts bearing "knowledge that is not yet abstracted into theoretical forms" (Magnusson, 2009, p. 166). Curiously, Magnusson makes no reference to the "theory nexus" concept introduced by J. M. Carroll and Kellogg two decades earlier, although it seems strongly related to his theme. At any rate, Magnusson derives from these theoretical provisions the stance that the traditional division of science (*episteme*) and technology (*techne*) cannot be upheld, because they are "often inherent in the same cultural objects, and therefore influence our activities on both phenomenological and epistemological levels" (Magnusson, 2009, p. 166). This proposition coheres with Nelson and Stolterman's call for the reinstatement of *sophia*—the integration of *thought* and *action* through design—rather than following the split between science and craft (Nelson & Stolterman, 2003, p. 11).

This begs the questions: Should researchers be designers? How does scientific knowledge feed into design and how do designed artifacts affect the production of knowledge? Several researchers in HCI have referred to the concept "wicked problem" (Rittel & Webber, 1973, p. 161) to illustrate how research and design can be concomitant to each other when dealing with problems of a complex nature. Although the term is not native to HCI—it is borrowed from the domains of social policy making and city planning—it accurately describes a conundrum familiar to HCI designers: "The information needed to understand the problem depends on one's idea for solving it" (Rittel & Webber, 1973, p. 161). When working with wicked problems, the process of solving the problem is identical with the process of understanding its nature. Because the implicative links between problem solving and understanding never end, there is no "stopping rule". An approach that lends itself to addressing wicked problems is *Research through Design* (RtD). RtD employs methods and processes from design practice as a legitimate method of inquiry, and integrates knowledge and theories from several disciplines (Zimmerman et al., 2010). RtD researchers advocate an iterative approach to reframing research questions and design solutions, acknowledging that the goal is "a solution that is optimal for the current situation and not a focus on the discovery of truth" (Zimmerman et al., 2010, p. 311). This echoes Rittel and Webber's proposition that "solutions to Wicked Problems are not true–false, but good–bad" (Rittel & Webber, 1973, p. 162).

Literature related to RtD appears to dovetail with the notion of design outcomes as embodied knowledge. A study among researchers by Zimmerman et al. highlighted that one of the outcomes of the RtD approach are artifacts that are "a type of implicit, theoretical contribution"—a codification of the designers' tacit knowledge about the phenomena at play, their context, and how preferred future states are envisioned (Zimmerman et al., 2010, p. 314). According to Nelson and Stolterman (2003), the outcome of any design process is an *ultimate particular*. Ultimate particulars are, essentially, concrete and unique things or events. Through scientific inquiry, one can try to induce universal knowledge by reasoning about ultimate particulars—this is the discovery of truth. Conversely, one can reveal the nature of ultimate particulars through deduction from the universal—potentially leading to the understanding of reality. However, science is inadequate when it comes to the *creation* of an ultimate particular:

> There is no scientific approach for creating an ultimate particular because science is a process of discerning abstractions that apply across categories or taxonomies of phenomena, while the ultimate particular is a singular and unique composition

> or assembly. Creating that which is unique and thus particular, therefore, cannot be accomplished using a scientific approach. (Nelson & Stolterman, 2003, p. 31)

Elaborating further, Nelson and Stolterman suggest that design choices are "based on reason, but not made by reason"—design is a process that "includes imagination, intuition, feeling, and emotion as well" (Nelson & Stolterman, 2003, p. 98). The intention of the designer is neither to reveal the particular—*the real*—nor to reason the universal—*the true*. Rather, the intention is to envision the desirable—*the ideal* (Nelson & Stolterman, 2003, p. 36). Hence, design outcomes are concrete articulations of not only knowledge, but also value judgements—embodiments of what "ought to be" (Zimmerman et al., 2003; Zimmerman, Forlizzi, & Evenson, 2007; Waern & Back, 2017).

In their RtD approach, Waern and Back (2017) criticize the tendency for HCI research to focus on artifacts or singular systems. They argue that such perspectives lead to a technology-deterministic view of design knowledge, implying that technology tends to dictate its use. While adopting the notion of ultimate particulars produced by design as embodiments of design knowledge, they point out that ultimate particulars can be considered *activities* rather than *things*. By focusing on how humans act and interact with each other with and through artifacts, Waern and Back contend that the articulation of design knowledge as manifested in an activity can go beyond the singular artifact. While keeping in mind that an activity is still an ultimate particular, i.e. a singular event and not readily generalizable, an activity-focused RtD method nonetheless directs the research focus toward actor experiences within the context of the activity, rather than a narrow focus on experiences with a specific product. Such an approach differs from earlier activity-centered approaches to HCI such as situated action (Suchman, 1987) or activity theoretical approaches (Kaptelinin & Nardi, 2006), which primarily present analytical perspectives of sociotechnical milieus. As such, the latter approaches suffer the same theory–practice gap as identified earlier, and have largely fallen by the wayside in HCI (Rogers, 2012). The RtD activity-centered approach proposed by Waern and Back, on the other hand, frames activity through the lens of deliberate design, with the proviso that design only influences and does not determine the activity:

> Articulating design knowledge at the activity level opens up opportunities for addressing domains where design uses a multitude of resources, and where design and use are aligned and intermixed in a joint creative process (Waern & Back, 2017, p. 3399).

The above quote vaults us straight into the core theme of this thesis. In the next chapters, we will be dealing with the ultimate particulars that are the theme of this PhD: the human activity of engaging in the joint creative process of generating new ideas in the initial ideation stage of music making. No creative co-performance is like any other—they are all one-offs. The main research question asks how a mixed-initiative interactive music system can benefit such activities. This question eschews technological determinism. The premise laid is that these activities take place regardless of which artifacts (if any) are integrated in the process. There is also an implicit assumption that new designs are desirable. This assumption is rooted mostly in my personal experience as a musician, but also in a number of anecdotes about innovative musicians who eagerly try a hand at experimenting with new technology whenever they have a chance. I argue that a holistic, activity-centered design approach, where the entire activity of collaborative music making is taken into consideration, is familiar terrain for most composing musicians. Whereas focusing narrowly on an interactive music system may seem alienating, especially to musicians who do not view themselves as tech-savvy, the broader perspective of musical exploration using new tools and methods is not that big of a leap. Instead of primarily asking questions about interface functions and parameters, broader discussions regarding the overall experience of navigating a collaborative creative space—whether with other humans or with non-human agents—are probably more intriguing. The articulation of design knowledge at this level, especially when viewed from several perspectives, is likely to hold relevance beyond the theoretical implications of an artifact.

Before proceeding, let us first heed the advice of one of the participants in the study by Zimmerman et al. (2010), who noted that:

> [...] the first challenge for a design researcher taking on an RtD project is to verify that the problematic situation is indeed a "Wicked Problem" that requires a design inquiry approach, and not simply a complex problem that can more effectively be addressed through scientific or engineering methods of inquiry (p. 314).

If we return to the source of the term wicked problem—Rittel and Webber's "Dilemmas in a General Theory of Planning" (1973)—it quickly becomes clear that applying the term to interaction design problems is, in fact, not quite accurate. The term was originally coined to describe large-scale planning problems on societal or policy levels, such as "the location of a freeway, the adjustment of a tax rate, the modification of school curricula, or the confrontation of crime" (Rittel & Webber, 1973, p. 160). To the point, true wicked problems are ones where design decisions have irreversible consequences to people affected by them—this is one reason why they are "wicked". There are certainly aspects of wicked problems in the research problem of this thesis, but in my view, it is not necessary to qualify it as a wicked problem in order to justify an RtD approach. Leaning on Nelson and Stolterman (2003) and Waern and Back (2017), I posit that an RtD approach is justified because the challenge of designing a co-creative activity comprising humans and computers has no definitive problem formulation—it is a “slippery” problem where solutions feed back into the problem and redefines it in the process. The entire design process should be viewed holistically as an empirical contribution to the research focus, which is the ultimate particular activity of “mixed-initiative music making” as envisioned and iteratively revised throughout the design and evaluation process.

4.2 Triangulation

Research through Design has gained acceptance within HCI as a legitimate approach to developing theory. However, there is no consensus on a unified RtD methodology. Some critics have pointed to this as a weakness, and that the lack of a rigorous research methodology for RtD hinders the production of comprehensive and mature theoretical constructs (Zimmerman et al., 2010). Gaver acknowledges that “theory produced by research through design tends to be provisional, contingent and aspirational” (Gaver, 2012, p. 938). Rather than viewing this as a problem, however, he seizes on this to point out that design statements tend to be *generative* as opposed to *falsifiable*. As such, RtD should employ methods concerned with what might be instead of what is. This makes sense, considering the fact that design changes the context in which it operates, in a similar way to how social sciences can give rise to policies within the field of its studies, and also change it in the process. Generative disciplines are concerned with “ontological politics” in addition to epistemology (Gaver, 2012, p. 943), and thus not likely to converge to a singular paradigm:

> From this point of view, the reason that research through design is not convergent is that it is a generative discipline, able to create multiple new worlds rather than describing a single existing one. (Gaver, 2012, p. 943)

Adopting the stance of “creating new worlds” as opposed to describing an existing one goes well with the aim of this thesis. There is no denying that designing for the (ultimate particular) activity of collaborative music making is, essentially, a creative enterprise, where multiple outcomes are possible and where there are no apparent metrics other than subjective preferences and intuition to go by. The many-worlds stance promotes discursiveness rather than agreement. In the domain of musical creativity, discursiveness is a given—imagine how boring music would be if everyone agreed on how it should be made! If we accept that design outcomes are ultimate particulars, we must also accept that theory derived from a single design, on its own, can never be a mature theory, because a mature theory would be generalized knowledge induced from multiple particulars. On the other hand, viewing ultimate particulars as activities rather than single artifacts makes it possible to treat theoretical contributions from RtD as adding to a discourse on general categories of activity. In the case of this thesis, the general category is collaborative music making, with composition and improvisation as meta-categories.

In order to approach these phenomena comprehensively, it is appropriate to study them from different perspectives. In this thesis, I am concerned with both interface building and the psycho-social foundations of musical co-creativity, and therefore straddling the technical and behavioral HCI research domains (Hudson & Mankoff, 2014). Mackay and Fayard (1997) propose a framework for *triangulation* across disciplines to address complex design problems. Their use of this term is related specifically to the HCI domain, and different from how it is used in social sciences (e.g. Denzin, 1978). Disciplinary triangulation as advocated by Mackay and Fayard constitutes the adoption of both deductive and inductive scientific models from disparate disciplines to inform a design process. They argue that addressing individual problems using approaches from different disciplines should lead to results that are more robust and useful than single-discipline efforts. An important part of this process is to avoid creating a "Tower of Babel", which can easily happen if terminologies and knowledge models are not harmonized. When using the term triangulation, I am here on referring to this multi-disciplinary triangulation framework as opposed to the common notions of data triangulation or single-discipline mixed-methods approaches.

A relevant example of triangulation used in the domain of music composition is provided by Garcia (2014), who triangulated between observation, design and theory to "better understand composers, design new technology and evaluate its impact on composers' creative process" (p. 3). Garcia emphasized the uniqueness of different composers' working processes, and adopted the method of *participatory design* within a triangulation framework. In participatory design, the participants—the intended user group—are active contributors to a design process (Greenbaum & Kyng, 1991). The way they engage with design prototypes, and their feedback about the experience, influences subsequent design decisions. Garcia organized several studies where he observed and interviewed composers working with different software prototypes. Insights gained from one study informed the next one, and thus guided the research direction. Here, the triangulation framework is useful because of the theoretical revisions driven by the empirical findings in the iterative design process. Earlier theories are not subsumed—they become multifocal lenses. Rather than committing to a specific design philosophy from the outset, this approach facilitates the generation of knowledge that is integrated with the specialized practices of the composers participating in the studies.

In his PhD thesis, Mudd (2017) adopted a triangulation framework with a theoretical vantage point in nonlinear dynamics. An initial stage of exploratory studies resulted in the design of a set of different control interfaces that were used in a subsequent stage of lab-based comparative studies. The theoretical assumptions embodied in these artifacts related to the kinds of affordances present in interfaces with nonlinear dynamic control parameters as opposed to interfaces with linear control parameters. The comparative studies examined how the inclusion or exclusion of nonlinear dynamical processes affected how the musicians engaged with the interfaces (Mudd, 2017, p. 100). As a complement to this, Mudd conducted an ethnographically informed study of how musicians engage with musical tools in their own musical practices. In this approach, the focus was no longer specifically on digital interfaces, but related to a much larger range of instruments that were the musicians' own domains of expertise. The theoretical implications of this approach combined with the results from the lab-based comparative studies formed the basis of the main discussion in Mudd's thesis, and led to contributions informed by these different perspectives.

4.3 Design as an open-ended creative process

Two aspects of the multistep approaches of Garcia and Mudd described above particularly motivated me from the outset. One was their deferral of committing to specific design architectures at an early stage. The second was their reliance on significant contributions from practitioners in the target field. There are several other examples of such iterative participatory design approaches that influenced the design of my methodological framework. Perhaps the most relevant in the context of this thesis are the design-test cycles implemented by Swift in the problem domain of musicians jamming (Swift, 2012) and Martin's practice-led approach of letting series of musical rehearsals inform the design process (Martin, 2016).

The common denominator for all these projects is that artifact design is predominantly adapted to the activity rather than the other way around. As such, integrating the artifact into the activity becomes a new way of doing something the actors know well, as opposed to imposing upon them a new activity domain altogether. Another commonality is the fact that the designers are also practitioners in the target field, as is often the case in music-related HCI research. There are pros and cons to this. On the one hand, such research tends to arise from an experienced real need from a practitioner's perspective. Because understanding the problem domain is essential to a comprehensive design process, this is a positive vantage point. However, it is easy for a practitioner-researcher to assume that one's personal problem domain is applicable to the field at large. The projects I have cited above can serve as examples of good ways to avoid this pitfall. The inclusion of several perspectives and drawing from the experiences of both peers within and experts from outside one's own domain appears to help make the research problem more widely applicable and thus yield productive results with an acceptable degree of transferability to the target field.

In the previous section, I mentioned that designing for the ultimate particular activity of collaborative music making should be viewed as a creative enterprise as opposed to a strictly scientific one. There is support for this stance from Mudd, who claims that the design of creative tools can be *material-oriented*, by which he means taking advantage of the active agency of tools and cultivating their unpredictable aspects as valuable resources for exploration and experimentation (Mudd, 2017, p. 11). A material-oriented approach is opposed to a *communication-oriented* one, where the tool is viewed as an ideally "transparent" conduit for preconceived ideas. Mudd's point is that the design process "does not need to start from a clear idea of the final product, but can develop iteratively as an exploration of a particular medium" (Mudd, 2019, p. 130). Leaning on Frabetti (2017), he invokes the notion of a "moving target" to illustrate this situation:

> This process acknowledges the creative significance of what Frabetti calls the "moving target" issue within software engineering, that software design doesn't progress through set problems being examined and solved. Problems emerge and shift during the process. (Mudd, 2019, p. 130)

With these reflections in mind, I decided to follow the same guiding principle of letting successive research stages inform the design process from the view of practitioners of collaborative music making from several different genres. I had not begun developing any artifact at the inception of the research project. What I had was an abstract idea for an interactive music system that would serve the role of being a musical partner for creative brainstorming. Such a system would not replace the role of human musical partners, but would potentially bring about different kinds of ideas and be a useful substitute for human musicians in contexts where it is impractical to rely on the presence of other musicians. The fact that the COVID-19 pandemic occurred during the course of the research made the latter motivation more pressing. Although I did have some concrete ideas about how to implement the envisioned system, I started out by studying musicians who work collaboratively without focusing on technology at all. I planned four separate studies, where each study would inform both artifact design for and the detailed aspects of the next study. All of the studies except for the first one were deliberately semi-structured, because each successive study would become more structured only after concrete research questions crystallized through analyzing the results from the previous one.

The activity-oriented Research through Design approach outlined earlier in this chapter was proposed as a way of avoiding technology-determinism. On the other hand, I have also just presented an apparently opposing view—the material-oriented openness to letting the agency of tools influence design decisions. In fact, these are complementary rather than opposing attitudes. Mudd's (2019) material-oriented approach also focuses on the activity itself—the design and use of creative technologies. Arguably, his opposing notion of communication-orientation, where the tool is viewed as a conduit of preconceived ideas, is more technology-deterministic than the acknowledgement of tools as a defining part of an activity. In the former, the tool becomes a necessity. In the latter, it potentially presents unexpected possibilities.

4.4 The four-study plan

4.4.1 A prospective view

The four-study plan depicted in Figure 16 shows how activities and artifacts are designed in tandem. A new technological element is introduced for each successive study. The plan starts out with no artifact in the first study, and for each study, the technology becomes more complex and advanced. The computational influence in the envisioned scenarios increases for each step. The activity of co-performing with an agent—human or otherwise—remains the focal point throughout the four studies. The scenarios in studies 2, 3 and 4 are purposefully open-ended, because details about the experiments could only be filled in after the previous study was conducted and (at least partially) analyzed. Correspondingly, the particular artifact to be used in each study would be contingent on how much I could realistically program within the given period, and thus would also constrain the study scenario. The technology milestones are predominantly focused on interface functionality as it appears to the user. The underlying algorithms would need to support the envisioned functionalities as prescribed in these study scenarios. As such, this is very much a top-down approach to programming, as opposed to a bottom-up approach of starting from the foundations of the software architecture and seeing what kind of global functionalities may emerge from this process.

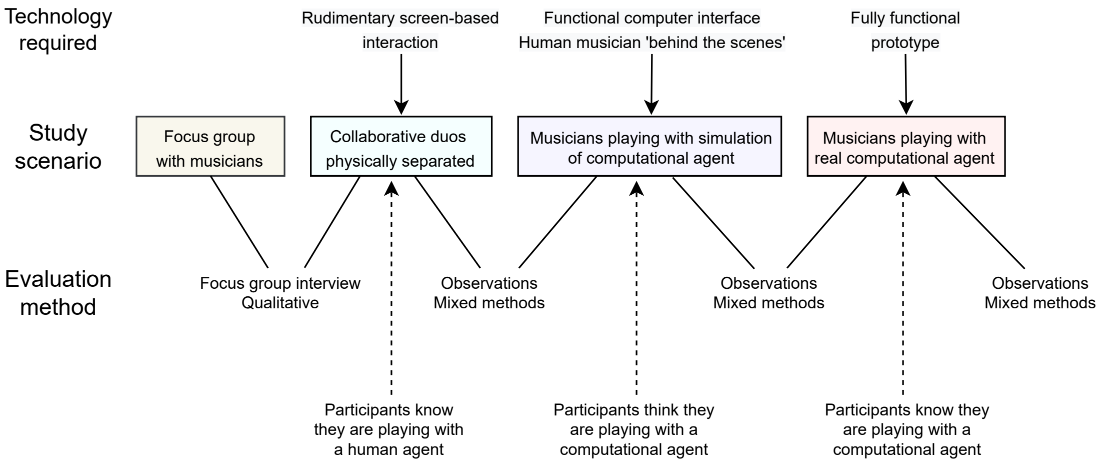


Figure 16. The four-study plan as envisioned before commencing with the first study.

4.4.2 Focus group

For the first study, I decided to organize a focus group interview with professional/semi-professional musicians or college/university level music students. In the invitation, I did not specify any genre, but set as a qualification that participants should be experienced in collective composition or song-writing, where improvising or jamming with ideas is typically an ideation method. The focus group participants were informed of my plans to develop an interactive music system, but this discussion would focus on their own experiences with developing ideas with other people in collaborative contexts. In particular, I was interested in how creative initiative is negotiated between collaborating musicians. This first study is detailed in Chapter 5.

4.4.3 Case study

The second study was planned as an observation study of four pairs of musicians in the setting of playing together in a collaborative song-making context. As a step toward simulating the scenario of interacting with a faceless agent, the musicians would be physically separated, with only instrumental audio contact with each other. I would also provide a rudimentary interface with which the musicians could send simple preset messages to each other. Whereas the first focus group had a broad theoretical vantage point in general dynamical systems concepts, the second study ended up with a narrower framework, which crystallized during the preparation, data collection, and early analysis stages of the first study. I call this the *convergence* vs. *divergence* framework (here on referred to as Framework A). In the analysis stage, I narrowed this even further by specifically adopting Cannone and Garnier's collective free improvisation framework of *collective sequences* and *articulation* (Canonne & Garnier, 2011; 2012), which is also based on dynamical systems theory.

Due to the difficulties of recruiting musicians during the COVID-19 related lockdowns, I decided to redefine the study as an instrumental case study of one musician duo. This turned out to be a blessing in disguise, because the analysis method was extremely time-consuming, and would have been difficult to achieve with four musician pairs. Because the analysis of this particular study took a very long time due to rich findings, it ended up informing not only the subsequent study, but also the final programming stages for the musical agent designed for the fourth and final study. The case study is the topic of Chapter 6, and the technical contributions to the fourth study are described in Chapter 8.

4.4.4 Wizard of Oz

The third study—described fully in Chapter 7—presented the first scenario where musicians were recruited to interact with a virtual agent. Unbeknownst to the participants, however, they were being recruited to a so-called *Wizard of Oz* prototype study (Kelley, 1983). The interface designed for the study represented a credible simulation of an interactive music system from the perspective of the user of the system. However, the musical agent "behind the scenes" was a human keyboard player pretending to be a computational agent (the "wizard"). The purpose of this arrangement was to obtain empirical data about how musicians experience co-creativity with a perceived computational agent before the implementation of the computational agent had begun in earnest. I reasoned that this would provide perspectives crucial to the activity-focused design process. As with the second study, my theoretical vantage point was once more dynamical systems, but this time I focused on a framework for the interactive behaviors of a believable computational agent. In particular, I wanted the agent to appear to change its behaviors in response to the participants in perceptible ways, without revealing any "human-ness". At the same time, these response types would need to be possible to implement in an actual computational agent at a later stage. To balance between these requirements, I adopted an *interactive behaviors* framework (Framework B) first proposed by Young and Bown (2010) and later extrapolated by Blackwell et al. (2012), who contend that a dynamical systems framework is advantageous to the implementation of such behaviors. Hence, I believed that a simulation—and later the implementation of these interactive behaviors would be a useful step in the direction of designing a system that would afford suitable interaction dynamics and support a co-creative environment.

4.4.5 User study

After the third study, there was a prolonged period of technological development. I ended up taking advantage of both the frameworks I had used in the two preceding studies. They provided means to conceptualize musical interaction from both the human and computational perspectives, and were integrated in a third hybrid framework (Framework C). A hypothesis that arose from this work became the main test case for the fourth and final study in this thesis. Two different implementations of the interactive music system were developed for a comparative user study, where the tradeoff between user control and system autonomy was a central premise. A null hypothesis claiming no significant difference between the two prototypes in terms of the user's creative engagement was made, against an alternative hypothesis predicting that increased system autonomy, despite affording less user control, would result in a deeper creative engagement. Additionally, I performed interviews, interaction logs were produced, and the participants of the study completed a home evaluation assignment a week after the study had taken place. The architecture of the interactive music system—dubbed *Spire Muse*—and the user study is described fully in Chapter 8.

4.4.6 A retrospective view

Due to the open-ended nature of the four-study plan, the design process was emergent as opposed to predefined. For example, bottom-up influences turned out to factor into the overall design process to a larger extent than described in the prospective view in 4.4.1—especially as important architectural choices needed to be made between studies 3 and 4. This will be described in Chapters 7 and 8.

Figure 17 shows a diagram that has been made retrospectively, and is based on the triangulation model proposed by Mackay and Fayard (1997). Here we can see how both theory and observations inform the design process. The overview does not go into details about goals, questions, methods, and approaches of analysis within the studies. This will be presented in Chapters 5–8, which correspond to each of the four studies.

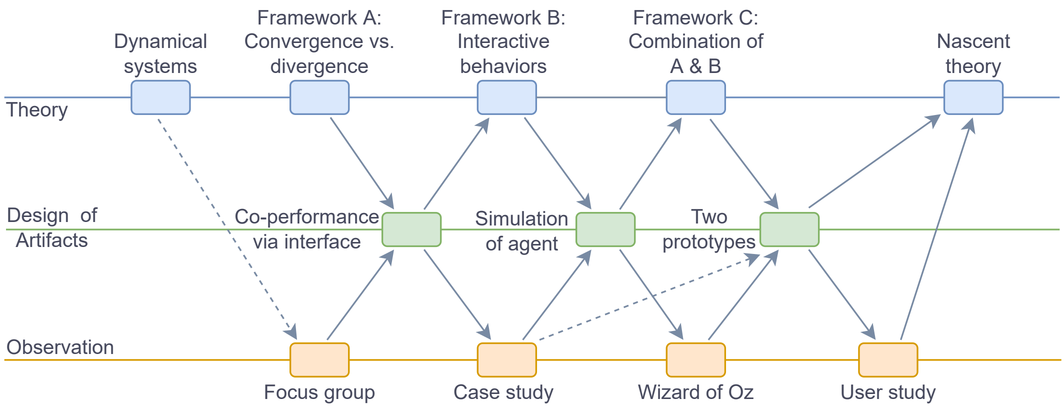


Figure 17. A retrospective overview of the triangulation process.

4.5 Ethical considerations

All of the four studies described above involve human subjects who impart thoughts and feelings about their musicianship and their attitudes toward issues such as creativity, aesthetics, intellectual property, technology, and more. For some people, this kind of information is deeply personal. I have abided by the general guidelines for research ethics issued by The Norwegian National Committees for Research Ethics (2016), where it is advised that people who participate in research, as informants or otherwise, shall be treated with respect, and that the research results have integrity, are available, benefits society, and does not cause harm. In this section, I will present the process of obtaining the ethical approval to conduct the four studies and discuss other ethical issues that were relevant to the research process.

4.5.1 Ethical approval

Norwegian research projects that collect personal information must be submitted to the Data Protection Official for Research at the Norwegian Social Centre for Research Data (NSD) for approval. NSD approved the project in June 2019 (see Appendix E).

For each of the four studies, the participants received information in writing about what their participation would involve (see Appendix A.2, B.2, C.2, D.2). The information was sent to the participants via email a few days prior to the study taking place, and included details about what kind of data would be recorded (video, audio, and interaction logs where applicable) and when (during interviews and during musical interaction sessions where applicable). They were informed that the research data would be treated anonymously, and that they could withdraw from the program at any time without any negative consequences. Because the four studies had different structures and different participation groups, I made independent information documents for each study. Furthermore, because participants in the first and last studies were both Norwegian and English speaking, I had to translate the two information documents pertaining to these studies from Norwegian to English. In the verbal introduction to each study, I repeated the information regarding what data would be recorded and reminded them that they could withdraw from the project at any point during or after the study without negative consequences. They were then handed a consent form that they could read through and sign before the studies commenced. During the studies, I informed the participants each time I started or stopped a video or audio recording.

4.5.2 Maintaining anonymity

The community of musicians from which I recruited participants is relatively small, so I was acutely aware that several details described in the results and analyses could contribute to jeopardizing their anonymity. In the first and third studies, I deemed it important to give the readers a sense of familiarity with the participants. Therefore, I gave all the participants in these two studies pseudonyms in order to preserve a sense of personality. This was especially important in the focus group study, where the personal dynamics between the participants were a vital part of how the conversation progressed. In the second study, I only referred to the participants as "the guitarist" and "the oboist". In the final study, the eight participants are simply referred to as P1–P8. In the first study, there are some quotes where the participants refer to works they have been involved with or fellow musicians with whom they have collaborated—no names mentioned. There is a chance that some readers close to one of the participants may suspect who is behind a pseudonym based on such details, but I believe the risk for this is very small.

After completing the third study (the Wizard of Oz study), I wanted to make a short documentation video of the process, including a clip from the interaction between one of the participants and the keyboard player who was simulating the computational agent (the wizard). At this point in the interaction, the participant was not aware that he was playing together with another human. In order to make this video, I sent an additional application to NSD for approval. NSD approved the request, and informed me that I would not need additional consent if the participants' voices were not included in the clip or they were not identifiable in any other way. If they were identifiable, I would need an additional written consent. I decided to take no chances, and obtained written consent from both the participant involved and the wizard. In the 2-minute long video, only the participant's back and arms are visible, but the wizard's face is clearly recognizable. He has seen this video and approves of the way he is portrayed. This video may be shown in future talks about how this study was conducted.

4.5.3 Other ethical considerations

In Chapter 7, I present the results of the so-called Wizard of Oz study, which involves a mild form of deception. The participants were recruited to interact with what they thought was a computational system. However, in the invitation I called it a "prototype", which in fact can involve humans simulating computational behavior. Thus, the invitation was not directly untruthful. Only halfway through the interview were they informed that the "system" was simulated by a human. I was aware that this could potentially lead to the participants feeling gullible or foolish when the set-up was revealed. I planned this section of the interview in such a way as to ensure that such feelings would be assuaged. I gave them plenty of time to react and make their feelings about this set-up known. I had prepared several arguments about how their opinions about the simulated system could be helpful for the research project, and that the Wizard of Oz method is used in HCI as a bona fide evaluation approach. Upon revelation, two of the participants did feel a passing sense of embarrassment. Fortunately, they took it with good humor, and none of them claimed their right to withdraw from the project.

In the third and fourth studies, the participants were placed in musical contexts they are unused to. Both in the introductions and during the studies, I made sure that they were well aware that I would not be evaluating their musical performance, but that I was interested in their interactions with the system prototypes. It is natural for humans to place the blame on oneself if some aspect of the technology they are dealing with does not function as expected. I emphasized the software they were interacting with is work in progress, and that it may not function optimally. In the few incidents where participants expressed insecurities about their own performance, I explained how design issues in the system were the likely reason for unsatisfactory musical results.

The interviews in Studies 2 and 3 were conducted and transcribed in Norwegian. Additionally, the interviews of half of the participants in Study 4 (P2, P5, P6, and P8) were also conducted and transcribed in Norwegian. The quotes I ended up using from these interviews have been translated to English. There is always a risk of misrepresenting someone's statements through inadequate or imprecise translation. I have done my best to capture the form and meaning from the original language. In Appendix F, an example of a translated dialog sequence is shown along with the original Norwegian dialog.

4.6 Summary

In this chapter, I have presented the overall methodological structure of the research reported in this thesis. I made a case for choosing a Research through Design approach, where methods and processes from design practice are used as guiding principles for knowledge production. Instead of a narrow focus on artifact design, I argued that an activity-oriented design approach is likely to yield contributions that are more applicable to the music field. A methodological framework involving a triangulation between theory, observations, and design was proposed. I put forward a four-study plan that was deliberately open-ended to allow for each subsequent study to be informed by findings from the previous ones. These studies will be the focus for the next four chapters. Finally, I discussed ethical considerations that I have taken into account in the design and execution of the four studies.

5 Study 1: Focus group

Despite the large amount of literature on music composition and improvisation, the initial explorative stage of collaborative music making emerging from improvisation—or jamming—is still a relatively poorly understood phenomenon. Several proposed models of the interaction dynamics in explorative improvisation were reviewed in Chapter 2. In Chapter 3, I introduced the possibility of using a conceptual framework based on dynamical systems theory as a vantage point for further investigation, with a particular focus on implications for the development of a computational agent that can engage creatively in collaborative music making sessions. In Chapter 4, I presented a methodological framework of triangulation between theory, observation, and design based on an activity-oriented Research through Design approach. This led to the design of four studies that are the topic of this and subsequent chapters.

The goal of the first study—described in this chapter—was to establish whether there are certain aspects of musical interaction and idea development that are common to a wide range of musicians. To this end, I organized a focus group for anyone who self-identifies as composing musicians with experience in collective music making, with an emphasis on improvising or jamming as a method to develop ideas. By learning more about the interaction dynamics between musicians in several stages of music making, the ulterior objective was to use findings from the focus group as a supplement to the theory introduced in previous chapters and hone in on a feasible conceptual model for an interactive music system that somehow appropriates these dynamics in its behavior. Throughout the study, I wanted to maintain a focus that would lead to a model that does not overtly impose genre-dependency and supports open-ended creative exploration.

5.1 Method

5.1.1 The focus group approach

In academic research, the standard method for qualitative data collection has been one-to-one interviews. Focus group interviews were first used by markets researchers to investigate consumer motives and product preferences, but has become more commonly accepted as a legitimate approach in social research in general in the past few decades (Kvale & Brinkmann, 2015). Compared to the more direct style of interviewing in one-to-one situations, focus groups take the form of a moderated conversation where the prime concern is to "create a permissive atmosphere for the expression of personal and conflicting viewpoints on the topics in focus" (Kvale &

Brinkmann, 2015, p. 150). According to Kvale and Brinkmann, focus group interviews can work well for exploratory studies where the researcher may want to discover different perspectives in a new domain. The dynamics of the collective interaction between people with different backgrounds and outlooks on the given topics may elicit "more spontaneous expressive and emotional views than in individual, often more cognitive, interviews" (Kvale & Brinkmann, 2015, p. 150).

I was convinced from the outset that a focus group would provide me with much more valuable data than if I had interviewed participants individually. In one-to-one interviews, the onus is on the interviewer to create a thematic trajectory throughout the conversation, and there is a clear power structure where the interviewee is seen as the research subject and therefore may feel obliged to answer in a way that they feel may be helpful to the interviewer. In a focus group, such a "pressure to perform" is much less prevalent. The researcher/moderator can take a step back and let the conversation between the participants flow in a more organic way. In fact, I argue that the creative group dynamics presented in Chapter 2 apply for focus groups in much the same way as collective improvisation (Sawyer, 1999). There is a larger chance for unforeseen, but important points to be brought forth when several different people "jam" on a topic. Furthermore, I do not view myself as an authority on the subject of collective music making in general. I accept that there are limits to how much one person can possibly know about a subject. As such, I believe that one-to-one interviews would expose my ignorance to particularities within different genres of music and possibly lead to a less-flowing style of verbal interaction. In a group context, on the other hand, there will usually be someone who can pick up and elaborate on points that that one counterpart would fail to do. As a researcher with a background as a musician and composer, I am in a unique position to guide a conversation between people who have spent large parts of their lives enhancing their music making skills in very different ways.

The drawbacks of focus group interviews are a reduced ability of the moderator to control the course of a conversation, and a tendency toward more chaotic interaction that could lead to challenges in transcribing the material (Kvale & Brinkmann, 2015). As I will demonstrate in the following sections, I was able to head off these challenges 1) by introducing a framework for turn-taking and thematic compartments in the conversation, and 2) by videotaping the conversation as well as relying on the audio recording.

5.1.2 Participants

In order to assure a level of consistency among the participants, the invitation (see Appendix A.1) described the target group as "professional/semi-professional musicians or music students on college/university level". One person who had recently finished their master's degree in music performance, but was no longer pursuing a professional career in music performance or composition, was admitted on the basis of experience, competence and interest in the subject. The group I ended up with comprised seven composing musicians with a commendable spread in terms of genre, age and gender. Among genres represented were jazz/improvisation, contemporary music, opera, rock, pop/singer-songwriter and electronic music. Participants were aged between 28 and 66 years old, and there were four males and three females in the group. An overview of the participants is shown below in Table 3.

David	Guitarist, laptop performer and composer. Relevant experience: Plays in music groups who make music collectively.
John	Oboe player, free jazz improviser and composer. Relevant experience: Works both collectively in music groups and performs solo with electronics.
Catherine	Contemporary opera composer, soprano and vocal improviser. Relevant experience: Improvises together with librettists and performers when developing material.
Lisa	Pianist, improviser and composer. Relevant experience: Plays in various groups who make music collectively.
Marcus	Electroacoustic music composer and sound artist. Relevant experience: Used to play bass in a rock band whose members made music collectively.
Marianne	Cellist and songwriter. Relevant experience: Usually writes music alone, but has also collaborated with musicians in the song-writing process.
Sebastian	Contemporary music composer. Relevant experience: Used to be part of an artistic group who made audiovisual performance art as a collective.

Table 3. Participants, their area of expertise and relevant experience. Pseudonyms are used.

5.1.3 Procedure

The focus group interview lasted for two hours, and was designed as a moderated conversation focusing on particular memories of collective music making sessions (see Appendix A.3). A week prior to the workshop, I sent the participants an email with a preparation task. The email encouraged them to reflect upon a specific memory of a musical collaboration—one where they started out with a simple idea and during the course of one session managed to make significant progress on this idea as part of an interchange between the collaborating partners. This memory would be their "story" going into the focus group session. The rationale behind giving the participants this preparation task was to mitigate the risk of participants spending an unnecessary amount of time ruminating through memories of past collaborations on the spot. I believed concrete memories of successful creative sessions would be the most fruitful vantage point for a conversation.

In the spoken introduction to the focus group, the overarching theme of collaborative music making was highlighted once more. I reminded them that they should be basing what they contribute on their experience as composers or songwriters. I emphasized that the focus would be on improvisation as a compositional method—a way to generate movement toward new musical ideas. I explained that there has been plenty of research on improvisation as performance, but not so much on improvisational approaches in composition. Finally, I underscored that this conversation would focus particularly on the ideation stage of collective composition—the transition between "playing around with ideas" to making more structured forms. To this end, the focus group interview would go systematically through four stages of this process:

- Original idea
- Obtaining a shared representation of idea
- Development of shared idea
- Result

Having a shared representation of an idea plays an important role in the success of a coordination task in group dynamics (Murray-Rust & Smaill, 2011; Canonne & Garnier, 2012). Some of the participants wanted me to clarify what was meant by "obtaining a shared representation" of an idea. After some back and forth, we decided that the informal expressions "being on the same page" or "being on the same wavelength" about an idea were apt informal descriptions in this context. I also pointed out that by "result" I did not mean a complete song or composition. "Result" in this context would be whatever formal structures this recalled session had produced. The most important parts of the conversation would revolve around the coming together around an idea and the development of this.

I wanted the group to have a visual representation of each other's stories in the form of a table on the whiteboard in the room. For each stage in the four-stage model, the participants took turns telling how the how ideas originated, how the ideas were communicated and became shared between the musicians, how they were developed, and what the result of the session was. I moderated the conversation with the help of a set of basic questions, sub-categorized according to the four sections, in order to keep a consistent format in the stories. While the participants were talking, I would write down key words on the whiteboard table. The table rows represented each participant's story in short form, and there were four columns representing the stages. This turned out to be a useful representation for both the participants and myself. Several times, someone would point to a cell while referring to their own or someone else's previous statements. This could be to elaborate, to revise or to clear up misunderstandings. The table in its final form is shown in Figure 18. In the first column, the (I) and (F) indicate whether the participant was the *initiator* of the idea, or if they were *following* someone else's initiative. In one case, the idea originated collectively, indicated as (F/I) in the table.

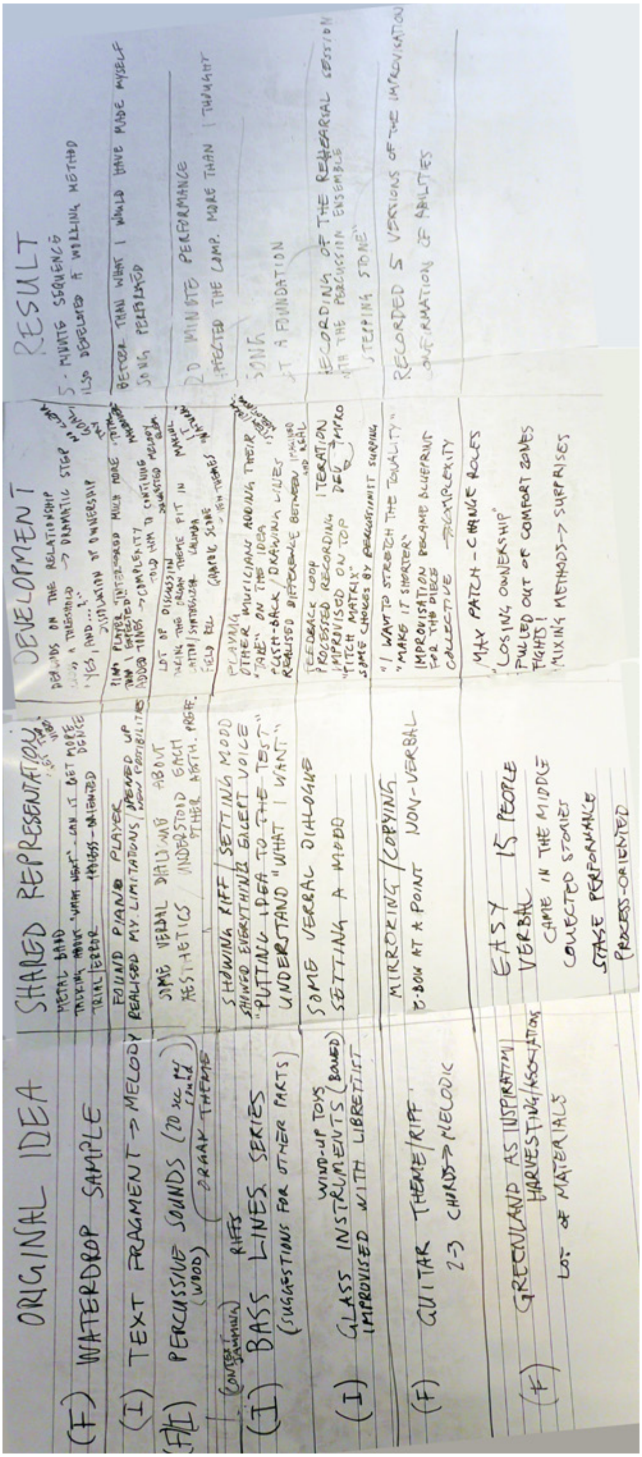


Figure 18. The whiteboard after the focus group session was over.

The focus group interview was both audio recorded and videotaped. I used the video recording to identify who was talking while transcribing the audio. This turned out to be very useful, particularly in parts of the discussion where several people were talking at the same time and interrupting each other. Watching their body language or facial expressions during the most such noisy and confusing moments made each speaker's intention more clear, and it made it possible to transcribe faithfully what each of them were saying. Without the video recording, some sections would probably have ended up not being transcribed. The video recording itself was not analyzed.

Transcribing is an interpretive process, and should be seen as a *translation* between the oral and written domains (Kvale & Brinkmann, 2015). The transcriber needs to make many decisions that go beyond simply turning sounds into letters. For example, it is normal for people to stop mid-sentence (or even mid-word) and start talking about something else. Sometimes it could be difficult to tell the difference between such leaps of thought and a poorly structured sentence. People also use many filler words ("um", "like", "sort of", "I don't know", etc.) conceivably to give themselves more time to finish a thought before continuing speaking. I have chosen to include most such filler words in order to preserve an impression of the formation of the speakers' thought processes as they unfolded. Sometimes the speech can speed up so much that it becomes nearly impossible to discern, especially if the speaker is excited. In extreme cases, the transcriber is forced to reconstruct the meaning behind the "babble" that is produced when syllables of the speaker get mixed up. In addition, an audio recording does not preserve body language, eye contact, and other visual cues that add meaning to the oral conversation. In some special cases, I have included such gestural information as annotations in parentheses. These annotations originate from my own recollections, but have been corroborated by the video recording where possible.

Because the participants were both Norwegian and other nationalities, the entire focus group interview was conducted in English. Only two of the participants were actually native English speakers. However, as English is a very common second language in Norway, this did not seem to hinder the participants' ability to express themselves in any significant way. When transcribing, I did not correct grammatical errors that were due to insufficiencies in English. However, in some of the quotes I ended up using, I have replaced some phrases with what I believe are more meaningful and understandable alternatives in square brackets. This has been done on only a few occasions.

The analysis of the interview transcription was done using an *inductive* coding approach, where the data itself gives rise to themes in a bottom-up manner (Saldaña, 2009). This constituted a departure from my original plan to follow a deductive approach to the analysis. In fact, I first coded the entire focus group conversation by categorizing the data according to dynamical systems concepts. While I did find that most of the participants' descriptions of their experiences with collective music making could be made to fit these concepts (e.g. finding common attractors, bifurcating from one attractor to the next, exploring phase spaces, etc.), I had a growing sense that I was force-fitting what they had shared to my preconceptions and simultaneously missing out on potentially interesting findings. Therefore, I decided to attempt an inductive coding approach to see how this would compare to the deductive approach. This change of methods was a revelation. I found that a data-driven analysis juxtaposed to a theory-driven moderated conversation provided a good balance between my presuppositions about the collective music making process from a dynamical systems perspective and themes important to the participants. I was therefore convinced to abandon the deductive approach, and rather look for themes that were not already apparent in the focus group invitation or in the prepared interview guide.

The inductive coding approach resulted in three main themes:

- Maintaining a process-oriented approach
- Attaining shared ownership
- Listening well

The themes and codes are available in Appendix A.4.

5.2 Results

As mentioned, I identified three key themes that appeared to resonate within the focus group. These are all themes that emerged spontaneously from the conversation as opposed to being prompted by my line of questioning in any clear way. In the following, I devote one section to each of these three themes.

5.2.1 Maintaining a process-oriented approach

With such a varied group of musicians, it is not surprising to find that there are large differences in terms of working methods and strategies invoked when developing ideas for compositions. One particularly important difference that became clear during the conversation was that the participants had varying notions of what constituted an "original idea" for a music making collaboration. Although the invitation focused on a bottom-up approach of collective music making (starting with a phrase or theme and collectively building material based on this), two of the participants shared stories where the collaboration started out with larger formal structures. This led to an interesting discussion about *process-oriented* versus *goal-oriented* collaborative approaches in music making. The first time the notion of a process-oriented approach came up was in an exchange between Sebastian and myself, whereupon John immediately picked up on the term to make a point from his own perspective:

> S: Yeah. The format. So we knew the format. But the point of the process rather than the idea-driven thing is to not know, and to let yourself be surprised.
>
> N: Yeah. Process-oriented. Okay.
>
> J: Yeah. You know, of course this process-orientation is the basis of everything, in a way…

This exchange shows that I proposed the term spontaneously, as a phrase to sum up Sebastian's point. This obviously resonated with John, and I intuitively sensed that several other participants could relate. Therefore, I reintroduced the term later, when David was introducing his story. He was commissioned by a metal band to "create some ambient textural thing" based on a single water drop sample. This was to be used as an introduction to one of their songs, but apart from that, the bandleader had no specific idea of what it should sound like. David insisted that they should work on this together. He sat with his laptop and had the bandleader give him feedback in real time as David developed the electronic material. The bandleader used abstract concepts such as "make it denser" or "do some glitchy stuff". In this manner, the one-second water drop sample quite quickly turned into a full five-minute piece of music. At one point while he was telling his story, I proposed:

> N: So this is also quite process-oriented as with (I point at Sebstian)…
>
> D: Totally, yeah.

Several other participants recounted similar stories. Catherine worked with an opera librettist on developing melodies and harmonic structure to a piece of text. The sole idea she began with was that she wanted to improvise using glass instruments. They began hitting, stroking and bowing the glass instruments, while Catherine was half-reading, half-singing words from the text. The recording of this session turned into the blueprint for the rehearsing score that she would bring to the entire ensemble weeks later.

John's story was of a recording session he had with two guitar players. They had no initial idea going into the session, but John quickly picked up on a short phrase that one of the guitar players was playing. He copied the phrase, whereupon the guitar player would move on to something else. John copied the next part, and the guitar player went back to the first phrase. While they were playing this musical game of tag, the second guitar player began layering with some sustained notes that would carry over from one segment to the next. John explained that he grew bored with the harmonic simplicity of what they were playing, and decided to "stretch the tonality". Within a relatively short period, they had a three-minute song, which they proceeded to record in several different versions.

While the notion of a process-oriented approach was forming during the focus group conversation, Marcus was the first one to contribute with the notion of a complementary goal-oriented approach. He told of coming to a band rehearsal with a collection of bass riffs and ideas for verse and chorus sections of a song. The band members began jamming together based on these ideas, but during the course of the rehearsal, some of the ideas morphed into something different. What emerged from the rehearsal was unmistakably a collective piece of music. However, seemingly implicitly acknowledging that breaking from a process-oriented mode was a necessary final step towards finishing a song, Marcus explained: "… We could become pretty goal-oriented at this point, if we wanted, like, a finished [song]".

Later, Marianne self-categorized her own approach as goal-oriented: "… in my very goal-oriented non-improvisational mind…" In formulating herself in this way, she implicitly defines a process-oriented approach as improvisatory. Indeed, her story is a good example of this distinction. She had brought in a pianist to accompany her on a song for her examination concert in composition. In her mind, the song was already written—she had even written a score for the pianist to play. However, the pianist insisted that he doesn't read scores—he only reads chord progressions. As they rehearsed, the pianist improvised over the chord progressions and gradually introduced new details to the composition. He also suggested structural changes. "He interfered a lot more than I thought he would", Marianne humorously admitted. In the end, even though it was originally her material, she felt that the song ended up being a co-composition. This was an unexpected, but positive experience.

All of the participants seemed to agree that when making music, the goal is always, as David put it, "a moving target". On a general basis (not just based on the session in focus), Marcus explained:

> [Often], setting a goal… helps me through some part of the work, and then I discard that goal because I realize I have a better goal now… Setting a goal, in a

> paradoxical way, sometimes just helps me focus on the current process I'm in. And then, when I'm, kind of, finished with that process I can just find another goal, or… refine the goal.

David thinks that not having a goal was an important part of why his collaboration was so effective:

> The cool thing about this process is neither of us really had any expectations going into it, so there wasn't an ideal we were working towards. It was, like, what can we do together, if we sit for a couple of hours in the studio, just, you know, work until we get somewhere that we like. So I think that was a big part of it as well, not having a clear goal.

He also pointed out that the result of this collaboration was not only the song itself, but they had discovered a fruitful working method that could be followed again in future collaborations. Several other participants told of similar experiences. John, for instance, said that the relative ease of how his experimental trio had produced a song, with hardly any verbal communication necessary, was in itself a confirmation of their competence as a group and of the framework for their aesthetics.

5.2.2 Attaining shared ownership

Another theme that arose spontaneously from the conversation was one related to idea ownership, and the effort going into attaining shared ownership of ideas. Marcus referred to this phenomenon as founded in a "cloud agreement":

> We [often had] this routine at the start and the end of each rehearsal; we were just jamming for like 20 minutes. … [This was most typically] just accepted to be in the moment and we never really tried to make something out of it, [it was] just gluing our taste and mental search in the same direction. So when I actually came to the rehearsal with a song, it was, kind of, already informed by all this… cloud of ideas which we built together with these impro sessions. So, that's an important context, because *then* when I showed everybody what they should do… reactions were based on the relation [to] this "cloud agreement" of musical ideas, and even group habits opposed to what can be found in this song idea.

Although the notion of a "cloud agreement" seems quite abstract, the other participants in the focus group seemed to understand intuitively the meaning of what Marcus was saying. All of the participants told stories where they had reached a level of creative collaboration where surrendering the idea to the collective proved beneficial. David summed this up:

> Whoever originated with the idea, maybe, had some sort of ideal of what… the end result should be to varying degrees. But none of them ended up there. So there's a certain point of letting go, or… being open to the idea of: "Ah, now, this is actually better than I could have expected". Or: "This is now augmented by this other person's idea." Sort of surrendering to that, just: "Cool, let's see what happens". I think that's something that, actually, all of these [stories] had in common to different degrees.

Sebastian's group had been conscious of the benefits of losing ownership of an idea, and had turned it into a method by swapping instruments as a means to this end:

> If you routinely have someone take your place and take someone else's place then you lose ownership. So that's a way of developing things… losing oversight, really. Which is kind of the point here because… that kind of method brings things in a place where no one person could see them go, and as it were it turned out better than we could imagine.

The benefits of staying attuned to the group mind and letting go of personal ambitions was also pointed out—once more—by David:

> So one of the things that I think was really nice about this process was that it was… always moving forward… like in improvised theater they have this "Yes, and…?" sort of thing, so it was, like: "That's cool, but what about if we do this, and move this forward?" And it was never, like: "No, no, not that, let's go back, let's do what we did before but make it…" […] And I think this point [that was made] about this… dissolution of ownership… is super-important, because… we protect these ideas… we have this thing that we've been thinking about for a long time and we want to preserve it at all costs, even if it doesn't really serve the greater good, in a way.

The transition from personal to collective idea ownership is not always easy, as attested by Lisa's account. As part of a residency, her trio found themselves in a room with a pipe organ. Although organ is not her instrument, they collectively decided that using an instrument that was so profoundly a part of the residency space they were given was a natural thing to do. Lisa came up with an organ theme that served as the backbone of the resulting piece of music. The entire working process ended up with getting the organ to work together with the other instruments, using the recurring organ theme as an anchor point. Lisa felt continuously challenged by one of her fellow musicians, who were more experienced with collaborative forms of music making: “What else can we do?”, “How can we turn it?”, “Can we play it a bit more?” Her contribution ended up being something she probably would never have done by herself. Although she said she had felt uncomfortable during the process, looking back she realized that using the instrument from the residency space, and the constant pushing by her fellow musician had made a positive impact on both the creative experience and on the resulting music. In a sense, the context had dictated the resulting work, and surrendering to this principle became an important part of gaining a sense of shared ownership.

5.2.3 Listening well

In an open-ended discussion toward the end of the focus group, the conversation turned to the theme of listening. It began with an anecdote told by Catherine:

> C: I did a concert in New York with two Argentine musicians, and they’re such fabulous improvisers. The first time I heard them I knew, these people think the way I think, I can hear them, how they’re processing the improvisation. So, the first time we played together, we recorded…
>
> N: And you could actually hear…
>
> C: I could hear…
>
> N: … how they think?
>
> C: Yeah, I hear how they think, structurally, motivically…
>
> N: Did you… Was that solely based on the sound of what they were doing, or was it also by looking at them, seeing…
>
> C: No, it was only based on the sound. Because… it was… a pianist and a drummer… And I loved their choices, and I could hear how they were listening to each other…

This exchange led me to ask how it is possible to hear how another person is listening. Catherine explained that it could be a combination of many factors in the back and forth between the musicians: Whether there is development in each other’s material, whether there is sensitivity to when that person needs space, whether there is need for support, or when there is an opening to come in with something new.

David also remarked that he often could tell if a musician is “listening well”, and openly posed the question: “How do we know when we feel heard?” He said that he had had this discussion with other improvising musicians for years, and still finds it difficult to understand. In the context of this discussion, Marianne made an illuminating distinction between the domains of improvisation and score-based music performance:

> For me… [with my] classical background… I recognize the listening part because if you play chamber music or in an orchestra, you listen a lot. But you listen for stuff that you already know is going to come, and I think it’s… difficult to understand… how improvisers listen when they… listen to stuff that they have never heard before, basically, on stage. It’s like… I get… blown away when I think about that, because listening can be very challenging, even when you actually read sheet music and you know exactly what they are going to play… and still you really have to concentrate. So listening to something new and responding to it is… on the other [end] of the scale.

She said she could tell that the pianist was “listening extremely well”—to the point that he was contributing with material she did not even knew she wanted, but which quickly became a natural part of the song.

Marcus noted that it is possible to hear how musicians listen not only from the sounds they are making, but also from the space in between the sounds: “If their silence is very pure, that kind of signals concentrated listening”.

According to John, the immediate aftermath of a session could also shed light on the depth of listening. In a recent collaboration, he was taken aback by the fact that the group leader had started talking about mundane issues right after the last note had been played, instead of reflecting upon the interaction they had just had. He found this disappointing, and thought it revealed a mind gap between the leader and the other musicians. He realized that they had been listening very differently. This realization “contaminated” his own reflexive thinking, whereupon he immediately began questioning his own assumptions about the session that had taken place.

5.3 Discussion

In this section, I will address the themes of process-orientedness, shared ownership, and listening as illuminated by the participants in the focus group. I dedicate a subsection per theme.

5.3.1 Going with the flow

Although the contrasting between process-oriented versus goal-oriented approaches was not a prepared theme for the focus group, I have identified it as one of the major themes arising from the 2-hour conversation. In my view, the terms do not appear to be a dichotomy. Rather, they seem to allude to a continuum between top-down and bottom-up working methods. The goal-oriented approach could be considered *top-down*—working from an idea of a song structure or a compositional ideal, and focusing collaborative efforts on developing material that fits this template. The process-oriented approach, on the other hand, is more open-ended. This is a *bottom-up* approach where the collaborators agree on a working method with no clear idea of any outcome. In this approach, ideas are grown from a "seed" (Pressing, 1984), and improvisation is a method powering this approach.

Clearly, some participants were unused to the notion of starting a collaboration entirely from scratch. However, Marianne's story of the "interfering pianist" demonstrates that collaborators can pick even presumed finished material apart, and process-oriented idea development may emerge from pieces detached from larger formal material. The shifting between process-oriented and goal-oriented approaches is perhaps more related to how far along a piece of music is in terms of completion, more than one method being more preferable to another. Undoubtedly, certain musical genres are more rule-based and have more rigid formal templates than others have. However, even within the confines of rules and formal etiquette, unexpected things can happen—especially when several musicians with different musical backgrounds interact. This corresponds with Sawyer's (1999) notion of emergent novelty as a primarily bottom-up process driven by group dynamics, as presented in Chapter 2.

My takeaway from these accounts is that improvisation tends to drive musical interaction into a process-oriented mode, even to the point of fragmenting existing forms. Hence, improvisation appears to have dual functions in collaborative music making. It can be used to "grow" material from the bottom and up, but may also serve as a performative wedge to break up consolidated formal structures and introduce new details with shared creative origin. Thus, improvisation is both a cultivator and a liquefier of form, depending on whether the composition is in an explorative or consolidative phase. The discussion concerning process-oriented versus goal-oriented collaborative approaches made this distinction stand out to me as a reasonable genre-independent principle. In Chapter 2, I drew on Benson (2003) to suggest that improvisation could be viewed as a catalyst for music making. This notion appears to hold water in light of the focus group participants' emphasis on the improvisatory aspects of a process-oriented approach in collaborative music making.

Improvisation-driven collaborative music making relies on a continuously shifting network of interpersonal relations, and musicians involved are required to make concessions on personal ambitions in order to reach a common goal. However, Mudd (2017) points out that the notion of a "goal" tends not to translate very well into the musical domain: "Often the 'goal' appears to be something slippery and possibly not even well defined to the musicians themselves" (p. 67). In fact, most of the focus group participants seemed more interested in the process-oriented approach of working with material in a bottom-up fashion, and seeing where this would lead. The deferral of a goal, paradoxically, can help create crucial movement toward achieving an outcome that is never explicitly formulated. Hence, the attitude that appears to be favored by the focus group participants in the initial explorative of collective music making is "going with the flow"—rather than forging it.

5.3.2 Decentering

The accounts that I have grouped under the second theme show that an inclusive attitude seems to be a crucial aspect in attaining shared ownership of an idea. Apparently, this includes accepting each collaborator's musicianship "as is". In terms of concrete musical strategies for collaboration, there was little talk about musicians giving music-theoretical directions to one another. In fact, in the cases where this came up, it was to point out that musicians would push back or refuse to take orders (e.g., "I don't do notes, I do my own stuff", the drummer refusing to open MIDI files with suggested grooves, the singer's "monopoly" of developing the lead melody, etc.). Examples of verbal directions given were mostly abstract, leaving it up to the recipient to interpret them on his or her own musical terms. For instance, Catherine's improvised session with the librettist using the glass instruments happened with little verbal dialog. She asked for certain things, like "sparse", "let them ring a long time", or "use these instruments", but apart from that they "just knew the mood". Rather than trying to guide each other explicitly with theory-laden suggestions, the participants mostly related anecdotes that revolved around influencing collaborators through musical actions. John copied the guitar player's theme to signal participation and acceptance of his lead, but "stretched" the tonality because he grew bored with the lack of complexity—both are examples of evaluative signaling (Pelz-Sherman, 1998). Marianne was primarily led by the pianist's elaborations around the chord progression to accept alterations to the song structure (although there were mutual instances of explicit musical direction as well). Musicians switching instruments, as in Sebastian's case, also demonstrates musicians accepting each other's differences and embracing the musical outcome.

Chapter 2 featured various conceptual models for improvised musical interaction revolving around strategies of musicians "pulling together" and "pushing apart". Strategies associated with these dynamics were identified as potentially being one of the driving forces behind the collective development of musical form. I introduced the theme of *negotiation* in the lead-off question in the second section of the interview—the one concerning obtaining a shared representation of an idea: *Was the idea easily shared, or was there negotiation before there could be a shared representation of that idea?* The topic of negotiation was clearly engaging to the participants, and appears to be engrained in the whole process of collective music making. However, I was reluctant to making it a theme in the analysis, because that would not be consistent with my intention of identifying themes independent from the ones introduced in the focus group invitation and questions. The theme of shared ownership, on the other hand, came from the participants, and arose from other contexts than the above question.

In fact, negotiation may be an inadequate term for describing the manner in which musicians decenter in order to reach collective subjectivity (Guattari, 1995; T. Davis, 2011) or a shared representation of the conceptual space (Murray-Rust & Smaill, 2011; Canonne & Garnier, 2012). What is normally associated with negotiation is the notion of "winning an argument" or gaining traction for one's own ideas. The general sense I got from the focus group, on the other hand, was different. The participants seemed to cherish the feeling they gained from leaving their egos aside and deliberately making space for other musicians to contribute with their input. This apparently takes practice. For musicians unused to the dynamics of collective improvisation and music making, an invitation to contribute may be misconstrued as *demands* rather than group-minded *gestures of inclusion*. They may feel put on the spot and forced to perform well, as Lisa's account of feeling pushed by her fellow musicians demonstrates. In general, the participants seemed to favor the inclusive and group-minded way of communicating, although getting to this point was not always without stress (Sebastian mentioned an incident where one collaborator had left the group in rage, only to come back a few days later). Experience does seem to matter. Asked whether she felt out of place or uncomfortable during the collaborative process, Catherine simply replied, "I'm comfortable being in the uncomfortable spot", giving the impression that having worked collectively with improvisation as a compositional method made her less fazed by the push and pull between the collaborators—it is an essential part of the collaborative music making process.

Marcus' notion of a "cloud agreement" is also interesting. He introduced the term as part of an elaborate point he was making about it being difficult to know how ideas originate. Any "original idea" is already informed by this "cloud agreement". Borgo (2005) makes a similar point in *Sync or Swarm* when problematizing what constitutes an initial gesture of an improvisation: "… in truth, that initial point is already implicated by feedback processes in a complex network dynamic" (p. 74). In other words, negotiation *does* take place within music collectives, but does not describe the actual musical communication very well. Negotiation happens in the social domain. Collaborating musicians form a social contract, and they are continuously negotiating the terms of this contract. There can be discussions about what is "yours" and "mine", forgetting to give space, worries about other people's opinions, bouts of stubbornness, and challenges that force musicians out of their comfort zones.

The "cloud agreement" appears to be an implicit understanding of the social playing field. In a well-functioning music collective, the individuals forming the collective may allow themselves to decenter in order to form collective subjectivity in the music making process. Collaborative creation is a remarkably complex undertaking, because each contributor has their own personal history, skill set, aesthetics, mood, and artistic vision going into a project. The decentering process apparently starts with the dispensing of excess personal pride and the channeling of this energy into a sense of shared ownership. Attaining this may feel emancipating and deeply gratifying when it happens, but getting there takes practice and requires knowing when to "kill your darlings". Being servile and always following the lead of others, on the other hand, does not amount to genuine collaboration. Decentering implies empathetic attunement, as proposed by Seddon (2005), but not leaving one's personal motivations behind altogether. There has to be a balance between giving and taking—between leading and following. These are profoundly social issues, where music happens to be the medium.

5.3.3 Inferring intentions

Music analysis focuses mostly on sound. This may seem obvious—music is, after all, *organized sound* (Varèse & Wen-chung, 1966). When experienced musicians create music together, however, they are not just listening for sounds—the above discussion about listening reveals that they could be listening for intentions. This aligns with Pelz-Sherman's notion of there being a semantic channel in the musical signal in addition to a sonic one (Pelz-Sherman, 1998). In the anecdotes about listening shared by some of the participants, there are examples of both formal, interactional, and evaluative cues being communicated between the musicians. Catherine's experience of understanding how the two Argentine musicians were thinking ("structurally, motivically") demonstrates an ability to infer the intentions about the formal development of the improvisation shared between the musicians, even though she had never met them before. There are both interactional and evaluative aspects implied by Marcus' point about certain "pure" types of silences signaling concentrated listening. Interactional—because the silence gives the other musicians space to develop their own material. Evaluative—because the signaling of deep concentration could also be understood as an affirmation of the current musical progression. Marianne's experience of the pianist's listening abilities could also be interpreted as evaluative signaling in that he came up with suggestions that she felt were improvements to the song's arrangement.

The discussion about listening reveals that *the way a fellow musician listens* has a definite bearing on the direction of the music. Anticipation—sensing what someone will do next—is an important part of improvised interaction. It is apparent that creative musical collaboration requires musicians to decode what they are hearing in a highly sophisticated manner—they are trying to discover the motivation behind the sounds and infer intentions about the development of the music. Conversely, the participants in the focus group revealed that they are actively listening for cues that can confirm that their own intentions are being heard in the semantic content of their musical signal. This bidirectional semantic signaling seems to be integral to the process of developing a piece of music in a collective manner.

5.3.4 Implications for modeling a mixed-initiative system

In the analysis of the focus group, I identified three themes that seemingly appealed to all the participants despite their different backgrounds. The implications for the findings related to these themes are compatible with the main goal of the study—to find common aspects of musical interaction and idea development across a wide range of musicians. I have formulated the themes as gerund phrases to highlight their function as acting toward an ideal of generating emergent novelty:

- *Maintaining a process-oriented approach*: The deferral of a goal and an attitude of "going with the flow" is most beneficial in the ideation stage of collective music making.
- *Attaining shared ownership*: Generative novelty tends to happen when individuals are able to decenter and form a collective subjectivity. This entails an attitude of reducing one's ego and overcoming personal pride.
- *Listening well*: The participants shared anecdotes revealing that they are constantly inferring intentions from the musical signal of their fellow musicians.

On the surface, devising an interactive music system with the appearance of such capabilities seems absurd. Are not all computers by their very nature process-oriented? What does it mean for a machine to decenter? How can a computational system tune in to the semantic content of a signal? Such anthropomorphizing can only lead to a conceptual impasse. Arguably, the challenge with computers is perhaps that any appearance of will in a computational system is often based on the preferences inscribed into the code by the designer. This is a crucial point, because if the designer and the user of the system have very similar aesthetic preferences (or, as often is the case, are the same person), the effect of overcoming contrast and reaching a middle ground that is co-creative disappears. This is acknowledged by Collins, who argues that interactive music systems that are "oppositional" and "pushes against" a human performer may lead to more inspirational human-computer encounters (Collins, 2010). This is also in line with the generative aspect of the models based on converging and diverging musical strategies presented in Chapter 2. In other words, a good co-creative system is probably one that appears to have a will of its own, perhaps even coming across as stubborn at times, but also can be responsive and give an impression of following or complementing input from the user.

The points made above are not new—after all, I made the point in Chapter 1 that a mixed-initiative interactive music system should be both reactive and proactive. The findings from the focus group, however, give a much more nuanced view of the reasoning behind such a statement. They can serve as guidelines throughout the design process and thus influence the architecture of the system. For example, the notion of "letting go" and surrendering to a collective subjectivity is an interesting dynamic to consider throughout the design stage. How might this be implemented as an effect in an interactive music system? Is it possible for a co-creative system to make its human partners feel heard, and how? In other words, the findings from the focus group also produced valuable questions for later studies.

5.4 Summary

In this chapter, I described the first of four studies, which focused fully on how human musicians interact when engaged in collective music making. A focus group consisting of seven musicians was organized with the goal of finding whether there are aspects of musical interaction and idea development that are common to a wide range of musicians. I have discussed these findings in relation to theory introduced in earlier chapters, and considered what they implicate for modeling a mixed-initiative interactive music system. The findings show that the following attitudes in general may increase the likelihood of generating emergent novelty in collective music making: 1) maintaining a process-oriented approach where goals are deferred in favor of "going with the flow", 2) attaining shared ownership by decentering and reaching a collective subjectivity, and 3) acutely listening for semantic content in the musical signal. All of these attitudes revolve around an active search for a co-creative middle ground between the aesthetic preferences of the individuals in the collective. This calls for an interactive music system with the appearance of a "will of its own"—a contrasting space from which it can diverge and converge to its human counterpart.

6 Study 2: Case study

In the previous chapter, I claimed that collective improvisation in an early, explorative phase of music making may have a cultivating function, where larger musical structures are emergent from interaction starting out with initial "seeds". Jamming—open-ended improvisation with no explicit goal—can be an efficient method of creating novel musical structures. In Chapter 3, emergence of musical form from improvisation was tied to the concept of self-organization in complex systems (Canonne & Garnier, 2011; van der Schyff et al., 2018). I offered several accounts of the high-level dynamics of this phenomenon through the prism of dynamical systems. In the case study that is the topic of this chapter, I wanted to examine these dynamics in terms of how improvising musicians converge and diverge (cf. the strategies of "pulling together" and "pushing apart" presented in Chapter 2).

The goal for this case study was to approach these questions in both a qualitative and quantitative manner. First, I performed a deep qualitative analysis of two musicians engaging in explorative improvisation using concepts based on dynamical systems. In this analysis, I primarily focused on the emergence of musically coherent sections of the improvisation and the manner in which shifts occurred between such sections. Second, I experimented with the unsupervised machine learning methods of autoencoding and clustering to see whether there are quantitative correlates to the identified sections and transitions. A bridge between perceived concepts and measurable patterns, if it exists, would be an important finding for the development of an interactive music system with the capability to detect such formal dynamics.

6.1 Method

6.1.1 Framing the case study

The original plan for the second study had been to observe four pairs of musicians engaged in music making sessions. The invitation letter for this study specifically asked for musical partners to participate as a duo (Appendix B.1). I reasoned that musicians who are used to collaborating would be better subjects for this particular study. This was in part informed by an observation from the first study, which showed that musicians with well-established partnerships are more likely to have dispensed with social insecurities and can focus more fully on musical communication. I defined the target group as "song-writing duos/composition partners who use creative interaction (jamming) as a method to develop musical ideas". The target group was not limited to a specific genre, but the invitation specified that tonal, acoustic-based instruments were preferable (i.e. not laptops, synthesizers or percussion instruments). This was in order to avoid making the analysis too complex.

Recruiting musicians for the study turned out to be difficult, and was exacerbated by the onset of the COVID-19 pandemic and the following lockdowns. Two duos contacted me, and correspondence was only upheld by one duo—a guitarist and an oboist. Based on this development, I decided to typify this particular research step as an instrumental case study of one improvising duo. A case study is a holistic study that uses multiple perspectives to explore the richness of real-life phenomena (Yin, 2009). An instrumental case study is one where a case is examined to provide insight into a phenomenon external to the case itself—the case is of secondary interest but facilitates our understanding of something else (Stake, 2003). Redefining the study as a case study allowed me to perform a much deeper and comprehensive analysis than what would have been the situation with four duos.

Fortunately, the duo I had recruited turned out to be very competent musicians who had collaborated for many years. They were mainly improvising partners, but they had also recorded composed material together. They were both men in their 60s, with backgrounds in jazz improvisation. Although the guitarist usually plays electric, he asked if he could bring his acoustic steel-stringed guitar for this session. I told him that this was, in fact, preferable. An electric guitar with effects could potentially be more difficult to transcribe and analyze.

The empirical gathering stage of the study had two parts. In the first part, the musicians were asked to jam freely for 20 minutes and see if they could make something together with no prior discussion of form and no explicit goal. The session was multitrack recorded with an audio track designated to each of the two instruments. The second part was an interview with both musicians together, where I would focus on transition points in the improvisation that led to new ideas. Combined, the improvisation session and the interview lasted 60 minutes.

The improvisation session took place in a controlled environment where opportunities for non-musical communication was reduced to a minimum. The musicians were placed in separate rooms—they could only hear each other. The rooms they were sitting in were adjacent to each other and interconnected with XLR patching, but with no window. They listened to each other's instrumental input through headphones. I sat behind a dividing screen in the same room as the oboist. This was the most practical solution because I needed to be in the same room as the sound card and recording equipment. I did not have access to a studio control room. The screen served a dual purpose. For one, I did not want my proximity to the oboist (being in the same room) to affect my position as an observer more than necessary. From behind the screen, I could hear the combined signals of the musicians through my own headphones, but I could not see them. Therefore, I was listening to them on equal terms. The screen also shielded me from the oboist's view. I did not want him to feel observed while playing.

The musicians were instructed not to speak or give any verbal cues to each other. This would have been possible through the instrument microphones. Instead, the musicians were provided with a laptop each, which was set up with a rudimentary interface they could use to signal very simple suggestions, such as "Stop—let's start over", "Let me start", "You start", "Let's start together—I'll count us in", etc. These commands could be sent between the musicians through the Open Sound Control (OSC) protocol (Wright, 2005). As for the counting in, I instructed the musicians to tap on their instruments instead of counting in with their voices. Once more, the reasoning behind this was to reduce social cues to a minimum. The human voice can reveal much more emotion than signaling through other media. It was also meant as a first test of observing how musicians relate to co-performing with an agent they cannot see, and by extension a first stage of the design process (cf. Figure 17 in Chapter 4). In this case, the musicians ended up not using the interface at all—they played for more than 26 minutes without even seeming to look at their computer screens. In the end, I needed to interrupt them and inform them that their time was up. Therefore, I will not mention the interface for the rest of the chapter, as it did not add any value to the study. Nonetheless, I still view the study as a stage in the design of the *activity* of mixed-initiative music making, as the factor of not seeing each other was a part of the simulation.

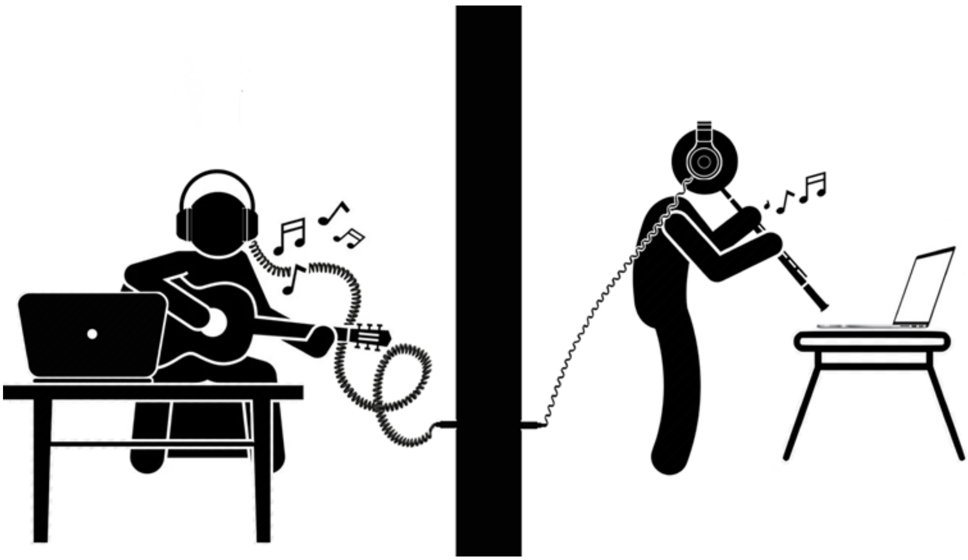


Figure 19. The musicians were separated and could only communicate musically or through preset commands via a laptop interface.

While they were playing, I made notes of places in the co-performance where I thought significant events were taking place. My working definition of "significant events" in this case included particularly well-flowing sequences in the improvisation, changes or breaks, and places where the musicians were apparently not communicating well. These notes would become the vantage point for the post-session interview. The interview was designed on the principle of stimulated recall—an introspective procedure through which subjects are prompted by a recording of an event to remember their concurrent thinking during that event (Lyle, 2003). Because the improvisation focused solely on sound, I decided to base the technique on an audio recording alone. Basing the interview on video recordings would imply following a split or dual screen, and I suspected that this would steal focus away from what mattered—the sound. Communicative body language had deliberately been filtered out of the study by separating the musicians.

Rather than going through the entire recording, which would add significantly to the interview length and probably lead to a disproportionate focus on less significant parts of the improvisation, I used the notes as markers for which parts of the recording to focus on. This decision was inspired by the notion of *Sonic Incidents*, proposed Caramiaux et al. (2015). Interviewing based on Sonic Incidents is an adaptation of the *critical incident technique*—a set of procedures in psychology designed to evoke memories related to particular moments experienced by the subject (Flanagan, 1954).

As I was transcribing the interview in the days following the study, I grew worried that the Sonic Incident technique had skewed the interview too much toward my personal interpretation of what constituted significant events. This unease grew as I listened back to the recording of the improvisation and discovered several other places in the recording I now realized could have been categorized as significant turning points, but that had escaped my attention at the time. Meanwhile, the musicians were asking me for a recording of the session, because they had been very happy with the way it had turned out musically. I made a rough mix of the recording and sent it to them. In the email, I politely asked them if they would be willing to send me a Word document each with a bullet point list of what they considered significant moments of transition in the improvisation, with short comments about why they considered them significant. Fortunately, they both happily complied with this request. In the analysis, these overviews turned out to be more useful reference points than the ones I had written down during the improvisation session. The combination of the interview transcription and the concise comments from these overviews were used to intersperse my musical analysis with the musicians' own reflections about the unfolding experience of playing the sequences I have focused on in the analysis.

Although the entire improvisation was more than 26 minutes, I decided to focus the analysis on the first 12.5 minutes. The musicians have a full pause at this point, and the second half of the session after this features extended techniques in both instruments, including multiphonics, key clicking, and using the instruments as percussive bodies. This was, in fact, not in line with my request in the invitation to avoid such extended techniques (which would make the analysis more difficult). Fortunately, the first half of the recording was more than enough for an analysis. From now on, "the improvisation" refers to these 12.5 minutes of the entire session. I ended up performing a deep qualitative analysis of the first six minutes of the improvisation, and a quantitative analysis of the full 12.5 minutes.

6.1.2 Qualitative analysis: Collective sequences and articulation

For the qualitative analysis of the first six minutes of the improvisation, I have used the model for collective free improvisation (CFI) proposed by Canonne and Garnier (2011; 2012) outlined in Chapter 3. As the musicians had no a priori agreement on any framework for the session, the co-performance can be characterized as *referent-free*, i.e. having no underlying formal scheme to guide the musical progression on an intermediate time scale (Pressing, 1984). The CFI model is based on concepts from dynamical systems theory. Canonne and Garnier describe two types of local structure (I allow myself some repetition from Chapter 3):

- A stable solution which can be seen as a "collective sequence"; this corresponds to a fixed point in the phase space of the system.
- An oscillating solution which can be seen as a phase of discoordination among the musicians; this corresponds to a limit cycle or chaotic behavior in the system.

Instead of applying the mathematical model that Canonne and Garnier (2011), devised in their first theoretical paper, this chapter will instead focus on the qualitative approach applied in a subsequent paper (Canonne & Garnier, 2012). Here, the authors applied concepts derived from the model in a study with improvising musicians. The mathematics behind the model were left out or implied, and the conceptual model was expanded because of the qualitative findings. A more detailed introduction of this conceptual model is necessary.

A *collective sequence* emerges when the musicians have converged to a common attractor in the musical stream. Although collective sequences are desirable in parts of an improvisation, musicians will also attempt to develop such a sequence to maintain interest. Complexity may thus build up to the point where the attractor is abandoned—ultimately ending the collective sequence. An unstable sequence may ensue, or a new attractor may lead the musicians into a new collective sequence. The transitioning between collective sequences—where musicians explore the phase space in search for new attractors—is called *articulation* (Canonne & Garnier, 2012).

Canonne and Garnier's study with improvisers revealed that the stable and unstable types of sequences both have "interesting" and "uninteresting" modalities. Thus, the authors claim that four kinds of sequences are possible:

- *Unstable and uninteresting*: Erratic, phase of discoordination.
- *Unstable but interesting*: Sense of getting somewhere, suspension or fragility.
- *Stable but uninteresting*: Low-complexity region.
- *Stable and interesting*: "Collective sequence".

Experienced improvisers tend to avoid low-complexity regions or "basins of attraction" (Borgo, 2005), which may result in a stable, but uninteresting sequences. Musicians may introduce instability by problematizing a texture experienced as too simple (Canonne & Garnier, 2012)—they engage in *problem-finding* as opposed to *problem-solving* behaviors (Sawyer, 2003). Free improvisers generally prefer to "surf the edge of chaos" (Borgo, 2005) at points where the musical stream may bifurcate and new attractors may be detected.

Improvisers engage in many different types of strategy depending on how they perceive the current situation (i.e. one of the four types of sequences described above):

- *Stabilization*. Repeating a pattern (ostinato or loop) or holding a sound/texture/pitch.
- "*Wait and see*". Hold back and see where the music is going.
- *"Playing along"*. Not keen on the idea, but playing along to be a part of the collective and support the others.
- *Densification*. Introducing complexity in the hope of provoking a transition or crystallization, by the dialectical confrontation of two contrasting elements, of a new, more exciting attractor.

Chapter 2 introduced several other models with more or less analogous strategies. I proposed they all could be grouped along an axis ranging from *converging* to *diverging* to the combined musical signal. Although I will mainly be leaning on Canonne and Garnier's model in this chapter, I have borrowed terms from some of these other sources where applicable. In all, they describe well the interaction dynamics of improvised co-performance—the push and pull between the musicians.

6.1.3 Quantitative analysis: Identifying segments and transitions

One of the questions this case study set out to answer is whether the perceived dynamics as presented through a dynamical systems model have quantifiable correlates in features extracted from the audio. If it were possible to find an acceptable level of correlation between what we perceive as salient features and qualitative shifts in the music on the one hand, and quantitative metrics on the other, this would suggest a way forward in creating a musical agent that can make autonomous behavioral choices on *meso* level timescales. According to Roads' classification of musical timescales, the meso level represents "groupings of sound objects into hierarchies of phrase structures of various sizes, measured in minutes or seconds" (Roads, 2001, p. 3).

The notion of musicians "identifying attractors" as presented in Canonne and Garnier's model comes across as subjective when placed out of context with the mathematical foundations of this model. Furthermore, they do not attempt to "reverse-engineer" the process and tie the concepts back to the mathematical equations from which the model is derived. At best, that would be a highly speculative approach to proving the validity of the model, because it would entail placing subjective features such as self-sensitivity, cognitive load and levels of boredom on a floating point scale. Clearly, the authors' intention is *corroboration* of the model, and not an exact validation of it. In other words, the model I have used in the qualitative analysis is applicable only in the perceptual domain. However, there are bound to be measurable patterns in the audio stream that the musicians are reacting to, and the assumption is that some of the mental representations they are forming are based on such patterns. Therefore, a correlation between percepts and patterns seems like a credible path to follow.

It is possible to extract a large variety of features from an audio signal. For this study, I have focused mainly on the harmonic and melodic dynamics, because the musicians, although free improvisers, were audibly attuned to each other in this respect (as the qualitative analysis will demonstrate). In particular, I have focused on a set of features called *chroma*—a 12-dimensional vector featuring the prominence of each of the notes in the standard chromatic scale within each given audio segment (from now on referred to as a *slice*). I performed a slicing of the recorded improvisation with an onset detection algorithm from the MuBu library (Schnell et al., 2009) in the *Max* graphical programming environment from Cycling '74[3]. This resulted in 742 audio slices with durations between 200 and 3000 milliseconds. I extracted chroma vectors from each of these slices. Inspired by so-called *pitch class transition matrices* in the MIDI domain, I devised a method to extract what I will call *chroma transition matrices* in the audio domain.

3 Cycling '74: Tools for sound, graphics, and interactivity. https://cycling74.com/

First, I discretized the chroma vectors from each slice so that only the most dominant pitches in each vector were activated and the rest ignored. I set the activation threshold at 0.4 (on a range from 0.0 to 1.0). The most dominant chroma features per vector (the ones above the threshold) were thus classified as ones, the rest as zeros. With this discretization, the transformed vector essentially becomes a like pitch class vector (see Table 4).

	Chroma vector	**Becomes**
Single note	0.11 0.78 0.15 0.21 0.19 0.27 0.31 0.14 0.39 0.18 0.12 0.26	0 1 0 0 0 0 0 0 0 0 0 0
Multiple notes	0.65 0.09 0.23 0.13 0.41 0.29 0.17 0.59 0.22 0.19 0.08 0.14	1 0 0 0 1 0 0 1 0 0 0 0

Table 4. The discretization of chroma vectors.

Next, I needed to make a decision about the temporal resolution at which I would be looking at local transitions between the pitches. Intuitively, a time range between 20 and 30 seconds seemed like a natural meso level time window. On average, this corresponded to 20 slices per window. Within each window, an algorithm I developed in Max counted the number of pairwise transitions between each "important" or "activated" note in the chromatic scale. This resulted in a 12×12 transition matrix per window. In order to smooth the effect of the windowing, the algorithm read a new window for every four slices (a so-called *hop size* of four). The improvisation was thus divided into 169 windows of 12×12 matrices.

I formed a collaboration with Sebastian Gregorius Winther-Larsen at the Department of Computational Physics at the University of Oslo, and we spent several months exploring different ways to detect relevant emergent patterns based on these chroma transition matrices. We found that a combination of autoencoding and K-means clustering yielded promising results. Ultimately, however, I find the results of the clustering confusing and inconclusive. This method should be seen as experimental, and the results in Section 6.3 are not submitted to any validation other than my own observations. I have nevertheless opted to publish these experimentations, because a byproduct of the process was the development of a windowing technique for tonal transitions that shows promise. In addition, I spent a considerable amount of time working with this quantitative analysis. Although the results are inconclusive, I surmised that a part of the RtD approach should be to show how I have worked. I have included several examples to show how the results varied depending on how the parameters were set.

As detailed in Chapter 2.3.2, an autoencoder is a type of artificial neural network that can learn efficient ways of encoding high-dimensional data in a lower-dimensional latent space in an unsupervised manner. Using autoencoding, the 144 dimensions of each transition matrix can be represented in a latent space with far fewer dimensions that represent the "type of tonal dynamics" happening within each window. In the method we used in our experiments, we disregarded the decoder part of the autoencoder algorithm. Instead, we extracted the features from the hidden layer and performed K-means clustering on these features in order to detect transitions in the music.

K-means clustering is one of many kinds of clustering techniques and was briefly mentioned in Chapter 2.3.2. The algorithm searches for a fixed number of clusters in a dataset by incrementally converging to the centroids of data points that are close together. The number of clusters (indicated by *k*) must be defined before running the algorithm. In cases where the ideal number of clusters is uncertain, there are algorithms that can help determine the most purposeful number of clusters. We have used the so-called *elbow method*, which refers to a characteristic "sharp point" (the *elbow*) in a graph that, simply explained, indicates the extent to which the number of clusters is likely to lead to over-fitting or under-fitting. In datasets where there are very clear clusters, the elbow will be sharp. In datasets with more even distributions of data points, the elbow will be weak or even undetectable. The Python code I used for the autoencoder and clustering algorithms is available in Appendix B.3.

6.2 Results of the qualitative analysis

As described in the previous section, the qualitative analysis focuses on the first six minutes of the improvisation. I made this decision after deliberating whether to focus on a deep analysis of a part of the improvisation or a more shallow analysis of the full 12.5 minutes. Ultimately, I opted for the former. I found a rich description to provide more qualitative value, but that such a detailed account of the whole improvisation would become too long and probably redundant. The quantitative analysis, described in the next results section, considers the entire improvisation (Section 6.3).

6.2.1 The opening: An accidental collective sequence?

Despite my careful instructions about not using any verbal cues, the oboe player used his voice once—to count themselves in. The manner in which he did this introduced a playful attitude even before they began playing. The counting in was completely arhythmical and was obviously meant to divert the guitar player from any sense of tempo whatsoever.

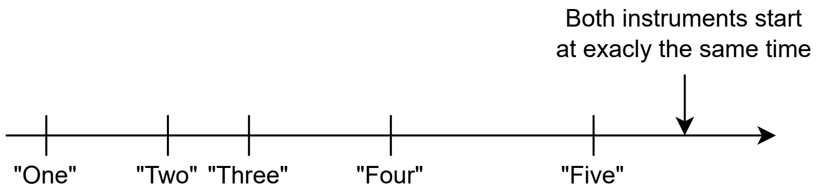


Figure 20. The oboe player's count-in was arhythmical and included an unconventional "five".

In spite of this overt disregard of tempo and convention, the musicians actually start at the same time. Within the first few seconds, they have locked in to the same tempo and tonality. Upon hearing the recording, the oboe player himself expressed surprise by how well coordinated they were from the start[4]:

> The decision about the arhythmical count-in, and to open in such a fanfare-like manner, was made on the spot. I was in a good mood. Just wanted to get started. The weird thing is that we still started completely precisely together. It is possible the count-in gave some timing indication anyway. I was, in fact, not aware that we had started so precisely until recently. I haven't analyzed how [the guitarist] also nailed it harmonically, but it is definitely strange.

Figure 21. The first four bars of the improvisation.

Figure 21 shows the transcribed score for the first four bars of the improvisation. Although it is an unplanned free improvisation (no prior communication about tonality, tempo or form was made), it is clear from this transcription that both musicians have found an attractor in the first few seconds of the co-performance. From an outside music theoretical perspective, it is tempting to claim that the attractor in this case is the tonality of a Dorian scale in C#, with

4 Via email.

a slow tempo at 67 beats per minute. However, when taken in context with the interaction that follows in the next several few minutes, it is debatable whether this can be claimed to be an intentional collective sequence.

If we look more closely at what is happening, we can assume that the first onset from both musicians is a “leap of faith”—they are doing the first thing that comes to mind. The oboe starts with an acciaccatura (grace note with emphasis on the main note) moving a minor third from a C# to an E. The guitar begins with a single note A#. Together with the oboe’s E this forms a salient tritone, creating harmonic tension which is open in terms of tonality. The lingering effect of the grace note hints at a diminished third in Bb. The guitar then moves a major third down to an F#. In the post-session interview, the guitarist claimed that this movement is a kind of “virus” for him—a typical “guitar thing” to do. The oboe briefly reinforces this note with a new grace note before going back to the sustained E. This appears to be a first “nod” from the oboist to the guitar player—perhaps a form of affirmation (i.e. “I hear you”). The oboe then breaks away from the E with a phrase that strongly suggests a C# minor scale, with clear emphases on the triad notes C#, E and G#. Together with the guitarist’s repeated A#, this skews the tonality toward a C# Dorian scale—although the musicians themselves may not be aware of this (they did not refer to scales in the interview). In the middle of the oboe’s phrase, the guitar acknowledges the E, going down by a tritone in the process. In the third bar, both instruments seem to decide to “land” their respective initial phrases. This apparently seals the Dorian scale completely—all of the scale’s seven notes have been “visited” by the combined phrases of the oboe and guitar, and the tonic C# is prominent in both the instruments as the first phrase concludes.

However, both musicians seem to want to leave an opening. The guitar plays a chord using the open B and E strings as a drone to supplement the C#. These three tones combined could be interpreted as an Em6. Meanwhile, the oboist is apparently not content staying on the tonic, instead seeking down to find the guitarist’s lowest tone—the B. Therefore, the opening phrase ends up as sounding most like a C# minor ninth with a prominent emphasis on the seventh. Although this opening only lasts for 12 seconds, it establishes the modality of the ensuing several minutes. The Dorian scale and the slow 4/4 meter is an attractor—at least for the oboist who seems to be more locked to this scale than the guitar player in the ensuing interactions.

If we look at Figure 22, we can see that the opening features several complementary movements between the instruments both melodically and rhythmically. This can help explain how they manage to converge harmonically and rhythmically so quickly. Whether intentionally or not, the musicians are engaging in calls, responses and mirroring on a very short time-scale, before moving in the same direction going into the third bar (red arrows). Furthermore, the figure shows that although they are playing in parallel, the instruments only happen to have note onsets at the same time twice (blue arrows). This forms a clear trajectory of a composite phrase starting decisively on the opening notes and resolving on the first downbeat of the third bar, with a tail in the oboe's voice. The rhythmical complementarity establishes a pulse driven by sixteenths during the second bar. The first real melodic parallel movement between the instruments coincides with the descending fifth interval in the oboe and the first polyphonic chord in the guitar—they both go down to the tonic.

Figure 22. Melodic calls and responses (red arrows) are prominent in the first bars. Meanwhile, the musicians only play simultaneous note onsets twice (blue arrows) in four bars.

6.2.2 Densification: The guitar goes rogue

Following this opening segment, the guitar once more plays the E-A# tritone interval—this time letting both notes resonate together—while the oboe responds with a G#-C#-E triad phrase. This reinforces the Dorian modality further, and establishes a kind of "division of labor" between the instruments, with the oboe playing the lead predominantly using notes in the range between the tonic and the fifth (still assuming a C# Dorian modality). The guitar, on the other hand, takes the role of "coloring" the harmonic palette. Within the first 30 seconds, he begins sneaking in notes outside of the established modality, thereby stretching the harmony and signaling a willingness to explore the phase space. These actions by the guitarist lay some doubt on whether the opening was deliberate harmonizing within a conventional frame or just happened by chance. At 0:26, for instance, he plays a short phrase in what seems to hint toward a D# major scale (or a D Locrian, G# Lydian, depending on what feels like the modal anchor point). He also alternates between playing the Dorian major 6th and the minor 6th, and between the minor and major 7th. This comes across as a deliberate ploy to create ambiguity. The oboist intermittently adopts the major 7th as a prominent leading note. The guitar gradually introduces competing or parallel tonalities, at one point strongly suggesting E minor with an emphasis on a raised 6th. However, he is intermittently "lured back" by the persistent lead by the oboe to the C# Dorian/harmonic minor. According to Canonne and

Garnier (2012), the guitarist can be described as alternating between "densification" and "playing along" strategies—he is both pulling to lead and pushed to being led. The guitarist gains adamance, and by 1:48, he seems to have discovered a new attractor, hinting at a B minor tonality—albeit without the B note featured (Figure 23).

Figure 23. At 1:48, the guitar player is starting to make large investments into the B minor tonality. The chords are my own interpretation of the underlying harmonic framework, and is based on the development the following 90 seconds.

A variation of this sequence makes a return at 2:14. Soon after, at 2:23 there is a slight pause in both instruments before the guitarist breaks into a sudden signal-like C#-B interval in the high-fret region. Post-session, both musicians have identified this call as a breaking point in the first part of the improvisation. The guitarist's call is a so-called "salient event" which functions as a "hook" for the oboist to follow, who seems to wake up to the guitarist's articulation attempts and immediately responds with an ostinato moving between B, A#/Bb and F#. This sounds as the oboe taking the guitarist's lead and moving to a stabilization strategy. The ostinato is giving the guitarist space to expand the phase space. The movement toward the B minor scale is by now quite clear in the guitar voice; meanwhile, the oboe once more breaks into a climatic upward movement in the C# Dorian scale—he "goes into the stratosphere", according to the guitarist. The guitar, meanwhile, climaxes with an intense repeated arpeggio in Em6, which is then left to resonate freely as the energy seems to run out of the sequence.

6.2.3 A "pastoral bridge": The second collective sequence

According to the analysis model, the simplest juxtaposition of two sequences is a so-called *cadential movement*, which happens when "all musicians have identified without any ambiguity that the ongoing sequence has come to its 'natural' ending" (Canonne & Garnier, 2012, p. 200). This is exactly what seems to happen at 3:16. After the previously described energetic climax and the sustained melodic pause, both musicians collectively appear to realize that something new must happen. As with the opening sequence, they begin in lockstep, and once more, they immediately find each other harmonically. The transition sounds pre-composed. This time, the B minor tonality is without ambiguity. Although the guitar has alluded to this tonality for more than a minute, it now sounds like a real modulation has taken place. The reason for this is that it is the first time the oboist is definitely playing in this tonality, and the guitarist is using the B as a bass for the first time. The tempo is markedly slower and contemplative. Listening back to the recording post-session, the guitarist describes this sequence as a "pastoral stretch". The oboist remembers well his own thinking in the sequence:

> I remember being preoccupied with doing more harmonic variation. I thought it was fun with the chords that [the guitarist] was putting in—and the rhythmical—a nice exchange between the instruments.

The "pastoral" collective sequence is shown in full in Figure 24. It has an ephemeral quality due to the descending pattern in the chords that the guitarist is introducing. He is moving away from the B minor tonality already in the second bar. The oboist is seemingly aware that the tonality is shifting—he waits for each next chord before resuming. He is feeling his way through a shifting tonal soundscape, led by the guitarist. Going into the final bar of the sequence, the oboist decides to hold a sustained note. Apparently, he is trying to identify the chords to know which way to resolve the tension. He ends up finally moving down a semitone. This turns out to be the fifth in an F# dominant 7th chord, with a suspended fourth resounding in the guitar voice. The * in the third bar in Figure 24 indicates that the chord is difficult to classify. Both the major and minor thirds are sounding at the same time. This is notated as a diminished fourth in the score.

Figure 24. The ephemeral "pastoral" sequence lasts for just over 20 seconds. The oboist is listening intently for tonal changes in the beginning of each bar.

Despite the short duration and all the tonal modulations, the close harmonic following by the oboist makes this sound like a collective sequence more than an articulation. It is ephemeral, yet stable. In fact, this sequence sounds like one of the most stable parts of the entire improvisation. It sounds, musically speaking, like a bridge.

6.2.4 Back to ambiguity

The main reason for the previously described session sounding like a bridge is that both instruments go back to what they were doing previously instead of cycling the new chord sequence. The guitarist resumes the arpeggios in Em6, with the open B and E strings as drones. In terms of tonality, he is mainly sticking with the B minor scale. The oboist returns to the melodic phrases in the C# Dorian modality. They are now clearly in separate "tonal camps", and the sense of tension and unresolvedness has returned. The most glaring example of this with a call and response at 3:42. The oboist plays a phrase in the now familiar C# minor triad range, reiterating variations of the phrase several times. The guitarist responds with a nearly identical phrase. However, it is shifted two semitones down—it is the same phrase in B minor. Whether intentional or not, they are communicating in different tonal frameworks.

Figure 25. The guitarist responds to the oboist's phrase, but in what appears to be a parallel tonality.

This and the following events strengthen the impression that they have adopted two parallel attractors. Overall, the entire ensuing sequence sounds quite the same as the first sequence of the improvisation. However, because the guitarist seems less intent on acknowledging the oboist's attractor, this time it sounds more like an "unstable but interesting" sequence rather than a collective one. Both musicians are comfortable in their own zones.

At around 5 minutes, the bridge-like "pastoral" sequence makes several partial returns. This time, however, it has a less stable quality, and the musicians are exploring the phase space to a much larger extent. It is still possible to discern roughly the same formal structure as in the first minutes, and this gives the entire first half of the improvisation an ABAB form.

The above analysis is obviously my own subjective description of the first six minutes of the improvisation. I have taken care not to present my own interpretations as the musicians' own inner mental representations. My use of "seems like", "appears to" and "sounds like" borders on the incessant, but it is necessary. It is debatable whether the musicians themselves would claim to have "found an attractor", or that they had "parallel attractors". However, they do seem to be in agreement about where the major transitions take place. This overall agreement will be further demonstrated in the discussion later in the chapter.

6.3 Results of the quantitative analysis

Earlier in this chapter, I introduced a potential method for quantifying the long-term tonal dynamics of musical co-performance. I will now demonstrate my attempts at applying this method on the audio recording of the improvisation that is the focus of this chapter. With deep learning models and clustering, it is difficult to know if one has arrived at an ideal model. Therefore, I will present a selection of the results of a few different models, interspersed with reflections about the level of consistency between the models and findings from the above qualitative analysis.

In the first model, the autoencoder had three hidden layers and a latent space of 20 dimensions—a moderate reduction from the 12 x 12 = 144 dimensions of each input matrix. In Figure 26, the values of the latent space are plotted as a function of time. The x-axis represents all of the 169 time windows. As explained earlier in the chapter, each time window contains 20 "slices" of audio events, with each slice corresponding to energy onsets in the audio signal lasting between 200 and 3000 milliseconds.

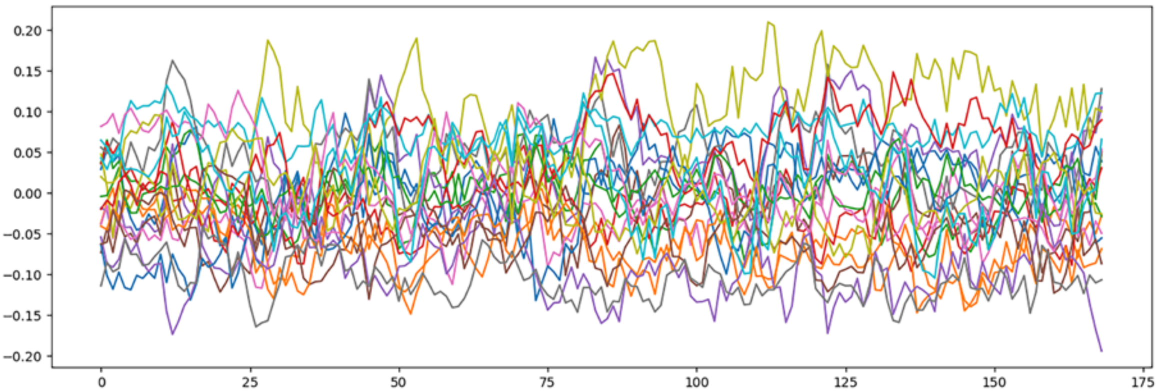


Figure 26. Graph E20 showing the progression of 20 dimensions in the encoded latent space of the autoencoding model.

In informal language, the graph shows the evolution of what the encoding algorithm "perceives" as the "20 most characteristic things happening". In the visual domain, a similar algorithm will typically detect edges and complexities in an input bitmap image, and disregard more uniform areas. In the case of the chroma transition matrices, we can assume that equivalent complexities are places where the tonal dynamics are undergoing some kind of change. The graph looks messy to the human eye, but we can see a couple of interesting-looking "bumps" at around index 25 and 50, a slightly different-looking mid-section, and relatively more variance toward the end of the improvisation.

The next step—clustering—was used to categorize these features into a fixed number of groups in terms of how they relate to each other. The elbow method indicated that 13 clusters might be most useful. In other words, the clustering algorithm was made to show 13 different types of constellations of these 20 dimensions. Figure 27 shows the result of this clustering process. The section of the improvisation that we performed a qualitative analysis on is featured in the 0–60 range in the x-axis. The first collective sequence equals indices 0–26, the articulation/transition happens at 27–35, and the pastoral bridge occurs at 36–40. The second "flavor of the pastoral section" is at around index 56. Furthermore, I summarized that the entire analyzed section roughly follows an ABAB form, but more unstable and with more variation in the second half.

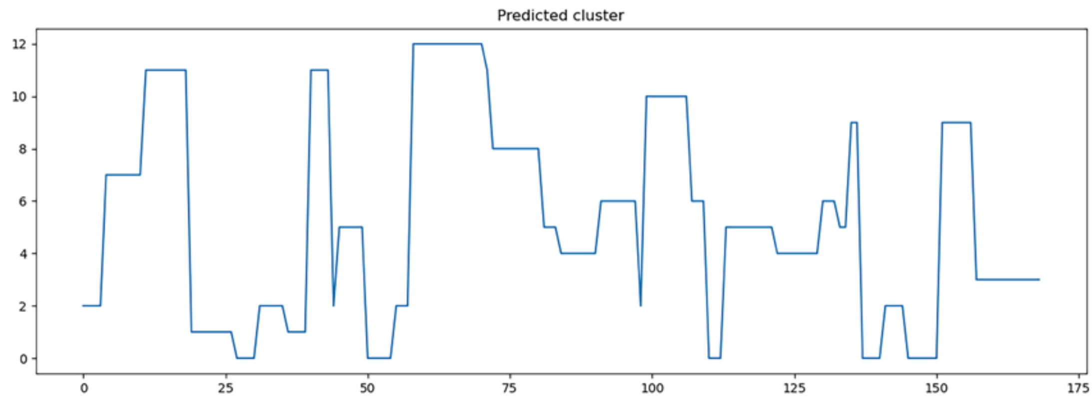


Figure 27. Graph E20C13 (13 clusters based on 20 encoded dimensions): The predicted cluster for each window in the 12.5-minute improvisation.

At first glance, these descriptions appear to have visible correlates in the provided graph—to a remarkable degree. However, I was acutely aware that I could be reading too much into this one graph. When I started studying the graph in further detail, I realized that the clusters do not actually give much information. For example, the graph makes it look like a large leap would imply a large qualitative shift, but this is actually not the case. There is no correlation between the step sizes and level of contrast between the musical sequences. However, there seems to be some correlation between the transition points in the graph and the changes described by the musicians.

I thought it would be interesting to see an even larger reduction of clusters. In fact, the elbow method was a little difficult to interpret, and it showed an indication of an elbow at four clusters as well. Figure 28 shows the result of the same encoded latent space reduced to just four clusters.

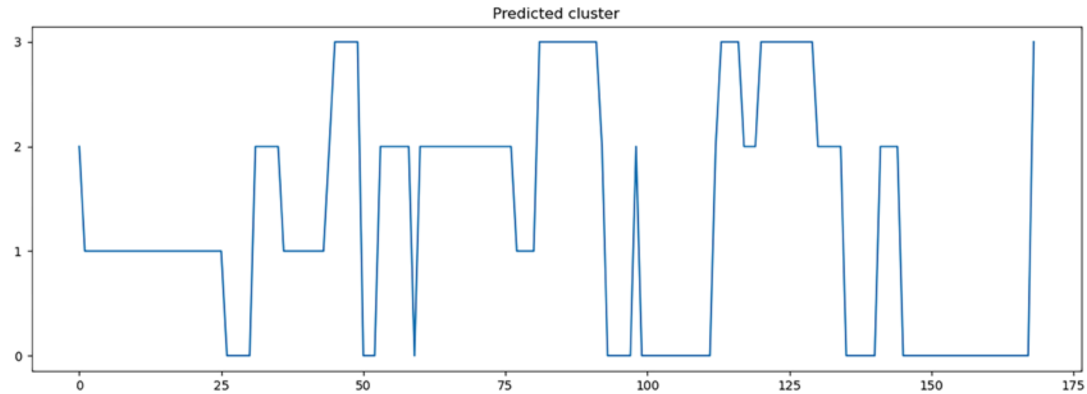


Figure 28. Graph E20C4: The same encoded latent space as above reduced to four clusters.

It looks different, but several transitions are occurring at roughly the same places. As with the previous graph, the clusters do not give much information, and there are several indicators that four is not enough clusters to make much qualitative sense. For example, the area of the "pastoral bridge" is categorized in the same cluster as the entire first collective sequence. This does not seem like a good clustering, considering I made a point out of the contrast between the two sequences. Additionally, the graph does not show a transition at the point where I identified an end of the pastoral bridge, which is roughly at index 40. Clearly, there are some qualities present in both the identified A and B parts that has made the latter clustering heap them together.

I wanted to know how making a significantly larger dimensionality reduction in the auteoencoding would impact the result. Figure 29 shows the result of autoencoding with a latent space of four dimensions.

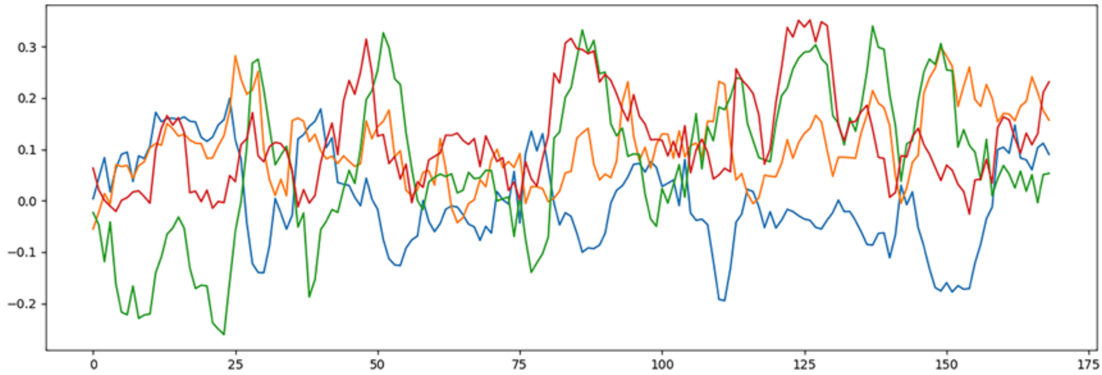


Figure 29. Graph E4: The result of autoencoding with a latent space of four dimensions.

With this low number of output dimensions, it is possible to discern the ABAB form in the 0–60 range even before the clustering. The areas around 35 and 55 are much narrower than in the rest of the first section. We know that these indices correspond to the "pastoral" B parts. Figure 30 shows the result of the clustering. The elbow method indicated a cluster count of six.

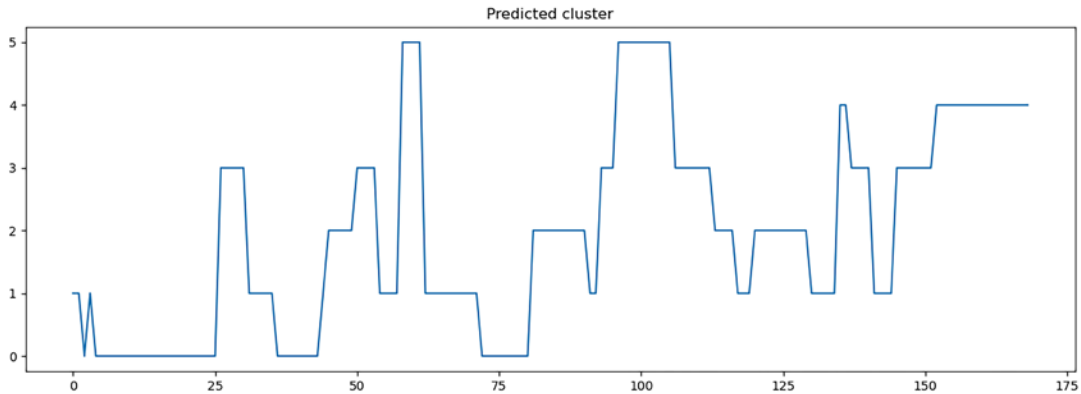


Figure 30. Graph E4C6: Six clusters based on four encoded dimensions.

As with the graph in Figure 6.10, I suspected that the amount of clusters could be too low. Once more, the first collective sequence is heaped together with the pastoral bridge. I tried increasing the amount of clusters to 15, which also had a small second "elbow" in the elbow method graph. The result of this graph is shown in Figure 31. This graph is also unconvincing. Again, the transitions are occurring in many of the "right" places, but the clustering does not correlate to the analysis in terms of the qualitative sameness of some of the sequences.

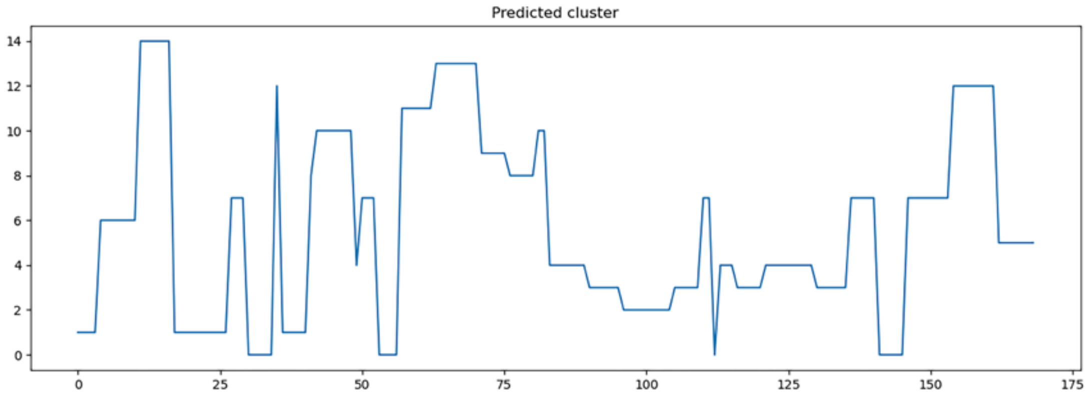


Figure 31. Graph E4C15: Fifteen clusters based on four encoded dimensions.

The examples above are only a selection of many other tests where I varied the number of latent spaces and clusters. They all showed the same general tendency. Many of the transition points were quite consistent, with a few of them featuring in all of the graphs. The manner in which the clustering part of the algorithm categorized sequences on the other hand was difficult to interpret.

6.4 Discussion

Applying Canonne and Garnier's model based on dynamical systems theory as a framework for analysis is a choice I made due to its focus on the converging and diverging dynamics of collective improvisation. What these dynamics consist of, however, is wide open to interpretation. The degree to which the different elements of a piece of music appear to be working in lockstep or in some kind of contrary fashion really depends upon which facets of the music are the focal point of the genre. In music based on functional harmonics, the notion of a home key works like an attractor. Moving away from the tonal center or introducing notes outside of the scale of the home key introduces tension. Arguably, these constraints are both natural and cultural. Other genres of music may place more significance on rhythmic or timbral constraints. There are probably countless other musical dimensions to consider. The usefulness of Canonne and Garnier's framework is due to its idealized and "agnostic" foundation (Canonne & Garnier, 2011). In its idealized form, the model assumes a hypothetical music signal with no emphasis on which aspects of the music signal are deemed important. Extended to actual practice, the complexity of the signal depends on what features are defined as significant:

> A cluster on the piano is more complex than a triad; a multiphonic on the clarinet is more complex than a traditionally-produced sound; a sub-division in septuplet is more complex than a subdivision in sextuplet; a stretch of music with very quick changes of pitches is more complex than a stretch of music with only one pitch. (Canonne & Garnier, 2011, p. 4)

In other words, the focal point depends on the type of music that is the subject of analysis. After having listened through the improvisation a few times, it became clear to me that the two musicians were relating to each other in terms of tonality, rhythmic variation and melodic themes. This was, of course, in some sense predetermined by the wording in my invitation (tonal instruments, avoid extended techniques, etc.). As early as in the planning stages of the study, I had already begun steering the analysis in a definite direction. The decision to show examples from the improvisation using musical notation instead of sequences of features extracted from the audio was a logical extension of this predetermined course. Musical notation is a visual representation that affords an effective visualization of motive dynamics (cf. Figure 22).

The post-session interview with the musicians revealed that the musicians had not been consciously reflecting upon their interactions in an analytical manner—they appeared to have been in a state of flow. As such, the comments they made during the interview provided sparse support for the chosen analysis method. When playing back selected sequences and posing questions in accordance with the Sonic Incident technique, they were mostly preoccupied with sharing how they had been thinking on abstracted and emotional levels. Examples of this are descriptions such as “pastoral stretch” and “going into the stratosphere” as offered by the guitarist. At one point, the oboist explained that he had decided to change his playing style from “points and dots” to “lines” in order to “tie things together”. I assumed he must have meant changing from staccato playing style to more sustained notes. Upon listening back to the particular sequence he was referring to while I was transcribing the interview, I realized that these “lines” must have been the sustained notes happening in the “pastoral stretches”. In that sense, the “tying together” could indeed be viewed as supporting the notion of convergence leading to a collective sequence.

The musicians were humorously self-critical throughout the interview, and commented on moments when they had fallen into old habits, such as the guitarist’s “virus”. He called this a “guitar thing” while pointing at his cerebellum, indicating a kind of brain parasite (my interpretation). Later, the guitarist poked fun at the oboist, who he apparently thought had fallen into some clichéd blues phrases. He joked that playing blues at a respectable music conservatory is surely something to be frowned upon. I understand these comments in the context of the ideal of free improvisation as referent-free, and they probably felt compelled to comment on obvious genre influences.

In hindsight, the immediate aftermath of such a session was probably not the opportune moment to garner the more technical kind of responses I was seeking for the analysis. The mood in the interview was mostly jocular—it was too soon for deeper reflection. Summed up, the interview transcription reads more like an emotional debrief than an analytical contribution. They were happy—almost euphoric—about having played together, but not focused on the how and why of the musical progression. What turned out to be a more substantial contribution to the analysis than the post-session interview was the emails the musicians sent me a week after the session. As described earlier, I asked the musicians to listen to the entire recording and send me a bullet-point list of what they deemed to be the most important transitions in the improvisation. These comments line up well with results from both my analysis and the transitions produced by the clustering graphs.

The notes I had taken while listening to the improvisation in real time, and upon which I based Sonic Incident technique interview, also did not turn out to be particularly useful. While there were certainly salient moments that we discussed based on these notes, they were structurally superficial compared with the musicians' own bullet point lists sent a week later. This is an interesting finding in itself. As a listener with an outsider perspective, I did not have the ability to hear the deeper structural dynamics unfolding in real time. At first, I was slightly disappointed that the musicians had apparently not identified transitions that aligned with my own notes. However, while I was performing the deep analysis on the first six minutes of the improvisation, I realized that I had placed too much significance on superficial events. I was distracted by ripples on the surface, and failed to see the waves.

The above issues notwithstanding, once I started to analyze the tonal tensions in depth, the framework turned out to be productive after all. I upheld the rationale that the choice of this particular analytical framework was an essential step in the overarching Research through Design approach. Zimmerman et al. (2010) write: "RtD allows researchers to rely on designerly activities as a way of approaching messy situations with unclear or even conflicting agendas; situations that are not well suited to other methods of inquiry" (p. 310). Or in the words of Nelson and Stolterman (2003): "Design takes place where there are no universal truths, no generalized solutions" (p. 257).

Viewing the interaction through the lens of tonal convergence and divergence is only one out of many ways to understand the music in this one particular case study. Consequently, the findings from this study cannot be presented as objective knowledge. Rather, they should be regarded as annotations from one instance (an ultimate particular) of the kind of activity that is the focus of this thesis. What this case study can offer is implications for the further design of a mixed-initiative interactive music system. As such, the findings provide support for the notion of viewing improvisation-driven music making from the perspective of converging and diverging tactics between agents. This has been documented with an in-depth description of a 6-minute stretch of free improvisation between two musicians, illustrated with several transcribed sequences, and further backed up by comments made by the musicians themselves. Although I can by no means claim I would find similar support in another co-performance with other musicians playing in a different genre, this case study has at least not weakened my assumption that similar dynamics are at play in open-ended collaborative music making in general. This taken together with research demonstrating the same tendencies (cf. Chapters 2 and 3) raises my confidence that the musical similarity axis ranging from converging to diverging as depicted in Figure 9 in Chapter 2 is a viable framework for further studies.

As mentioned earlier in this chapter, the quantitative method of autoencoding in combination with clustering yielded results that I found inconclusive. On the one hand, the graphs demonstrate a certain level of correlation between what the musicians describe as transitions and breakpoints in the graphs. On the other hand, there are many other unexplained breakpoints in the graphs, which I have not investigated further due to time limits. Furthermore, I did not find a model that consistently lines up with qualitative descriptions of what is happening within the sequences. Two sequences that sound qualitatively similar are not necessarily clustered together in the same category. This aspect of the clustering graphs seems quite arbitrary. However, the method's apparent capacity to identify transitions was interesting enough for me to look into further. I wanted to examine the extent to which the musicians' lists of perceived transitions lined up with the entire improvisation. Table 5 makes a comparison between the qualitatively perceived sequences and transitions (Column 2) and my comments on whether or not the graphs show any visual correlates (Column 3). Column 1 shows the time points. In Column 2, I have also included shorthand descriptions of my own analysis where applicable.

0:00	Oboist: "Fanfare-like opening, we find each other well." Me: First collective sequence, part "A".	Index 0. All graphs show a quick stabilization following an initial "turbulence".
1:11	Oboist: "I feel like the intro is over, and I think there should be some kind of melodic break."	Index 10. Graphs E20C13 and E4C15 show a transition.
2:23	Guitarist: "Break and introduction of something that develops into a pastoral stretch." Oboist: "I want to put in more energy and find another harmonic material—with a shorter ostinato-like theme."	Index 26–27. All graphs show a transition here.
3:16	Guitarist: "Pastoral stretch." Oboist: "Variation of harmony." Me: Second collective sequence, part "B".	Index 36. All graphs show a transition going into this part.
3:40	Uncommented by the musicians, but I have marked the end of the "pastoral bridge" at this point, and the beginning of an unstable version of the first collective sequence. Return to part "A" (but not quite).	Index 40. Only E20C13 shows a transition here.
4:50	Oboist: "We find a nice harmonic frame, but it's a little too nice." Me: A taste of the "pastoral bridge", but less stable. Return to part "B" (but not quite).	Index 56. All of the graphs have transitions around this point (not completely aligned).
5:56	Guitarist: "Transition to a more free tonal part."	Index 74. All graphs except E20C4 shows a transition happening at around 72.
6:50	Guitarist: "Octave ostinato in the oboe, transition to a static-rhythmical sequence. Oboist: "I find myself giving [the guitarist] enough space, and I can do something more rhythmical."	Index 87. All of the graphs have sustained "plateau".
8:45	Oboist: "I probably want some variation."	Index 121. It is unclear if the oboist points this out as a transition, but all graphs show a transition at around this point, followed by sustained cluster.
9:45	Oboist: "I want to give [the guitarist] more space, so I go over to playing an ostinato."	Index 138. A transition occurs in all the graphs.

Table 5. Time points, comments and comparisons to the graphs.

Figure 32 combines the transition points from all fours graphs in one representation. The coloring scheme makes it so that points with multiple occurrences get darker lines. For example, all four graphs has transition points at index 36 (the start of the pastoral bridge), which is shown as a dark line. Transitions that only happen in one graph are barely visible.

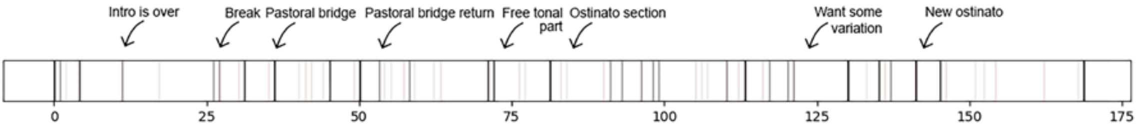


Figure 32. The transition points from the graphs combined.

Overall, I find the degree to which the experienced transitions and the graph breakpoints align interesting and moderately convincing. After spending several months researching the subject of quantitatively identifying longer-term dynamics in music, this method yielded results most closely reflecting the subjective experiences. Due to the inconclusiveness of the clustering, however, I ultimately began suspecting that the apparent success of identifying tonal transitions must be due to the underlying chroma transition matrices that formed the dataset, and that the ML algorithm was inadequately designed. Unfortunately, I did not have the opportunity to devise a new evaluation method to test this assumption within the scope of this part-study. I was, however determined to keep chroma transition matrices in mind for later experiments in the programming stages of the interactive music system development.

6.5 Summary

This chapter has covered an instrumental case study of two musicians—an oboist and a guitarist—engaged in a collaborative improvisation-driven music making session in order to examine the converging and diverging dynamics of collective music making described in the theory chapters. As a first step toward simulating the situation of engaging with a faceless interactive music system, the musicians were placed in separate rooms and could only communicate with musical signals or through preset commands on a rudimentary computer interface. The session was recorded, and the musicians were interviewed together immediately after the session. The interview focused on points of musical convergence and divergence using a Sonic Incident technique. Following an additional request, the musicians later sent me bullet point lists of what they deemed to be the most important transitions in the improvisation. The latter turned out to be the most useful data in addition to the audio recording of the session itself.

I analyzed the data both qualitatively and quantitatively. I performed an analysis of the first six minutes of the improvisation using Canonne and Garnier's (2011; 2012) dynamical systems model for collective free improvisation, which focuses on the concepts of convergence to collective sequences and articulation between sequences. Although this one case study cannot present generalized findings, the analysis shows that the musicians did indeed follow a structure as predicted by this model, indicating that the convergence vs. divergence framework is a viable one for subsequent studies. The quantitative analysis presented an experimental method based on extracting chroma transition matrices from meso level windows of the audio and submitting these matrices as a dataset for unsupervised machine learning. I tested whether an algorithm based on autoencoding and K-means clustering could detect sequences that are perceptibly stable and transitions between such sequences. The results were inconclusive. There seems to be a certain degree of consistency between transitions identified in the qualitative analysis, comments from the musicians, and the graphs produced by the ML algorithm. However, the clustering often categorized sequences identified as qualitatively different as belonging together. I suspect that the ML algorithm is flawed, and that the real promise for future experimentation lies in feature extraction using the chroma transition windowing technique.

7 Study 3: Wizard of Oz

In the first two studies, I was interested in learning more about the dynamics between human musicians in the early stage of making music together. The first study was a focus group with seven musicians who shared their experiences about the explorative phase of collective music making, where improvisation has a cultivating effect on the creative process. The second study was a deep analysis of two improvising musicians as they developed a piece of music in real time. By this point, I had formed a clearer idea of what an interface supporting explorative, open-ended experimentation with basic musical ideas could look like. The most important aspect of this interface is an emphasis on *communication* with the computational agent, not *control* of it. Before beginning software development in earnest, however, I wanted a user study to inform me how musicians would experience engaging in a mixed-initiative music making session using a prototype of this interface.

Beaudouin-Lafon and Mackay (2003) define a prototype as "a *concrete representation* of part or all of an interactive system" (p. 1007). In software engineering, prototyping is an established way to study the feasibility of a technical process, whereas in HCI it is used more to express new ideas and reflect on them. In this case, the psychological impact of the mixed-initiative context was more important than technical feasibility. I reasoned that bias plays an important part in improvising musicians' frames of mind. Therefore, I wanted to gauge the reactions of musicians playing with what they actually believed to be the envisioned interactive music system, even before it existed. This is possible using a so-called *Wizard of Oz method*. To simulate a computational system, a keyboard player was instructed to behave as a computational agent. This chapter describes how a Wizard of Oz study was set up and presents the results.

7.1 Method

7.1.1 The Wizard of Oz method

Successful prototypes are ones that support creativity, encourage communication and permit early evaluation. Prototypes can be *offline* or *online*. Offline prototypes do not require a computer, and include paper sketches, storyboards and videos. Online prototypes run on a computer, and include animations, interactive video presentations and programs mimicking the behavior of the envisioned system. Offline prototypes are inexpensive and quicker to design, less likely to constrain the designer's thinking, and can be created by a wide range of people, not just programmers. Sometimes, however, more sophisticated representations of the system may be required to inform the design process. The Wizard of Oz method, although considered offline by Beaudoin-Lafon and Mackay, involves the user interacting with a working representation of the interface. Behind the scenes, the developer follows what the user is doing, and creates an illusion of a working software program by manipulating the interface in a believable manner in real time. The term *Wizard of Oz* was coined by Kelley (1983) when he devised a method to simulate a natural language computer application for a user study. The term is based on the wizard from the 1939 movie of the same name, who turns out to be a frail old man behind a curtain operating a machinery that gives Dorothy and her companions the impression they are faced with a supernatural and menacing being.

There are several reasons why I believed that Wizard of Oz was the most informative method to gain the knowledge I needed for the development process. For one, I needed feedback about how musicians felt about interacting with the envisioned interface based on a realistic scenario. I feared that a hypothetical situation with musicians reacting to video demonstrations or obvious mock-ups would lead to missing the opportunity of catching genuine responses to interacting with a computational agent. If the users were aware of the simulation, I suspected their reactions would be different than interacting with what they thought was a "real" system. Actually playing an instrument together with an unknown agency is very different from imagining doing so. I also wanted to see how musicians used the envisioned interface while interacting with a *believable agent*—one that provides an illusion of life[5] (Bates, 1994). This could potentially reveal weakness in the design and help make the software development more focused on features that turns out to work well in the prototype study. I suspected that a less realistic scenario would skew the users' interactions with the interface, and potentially put me on a wrong path.

7.1.2 Structure of the study

In the following sections, I will refer to the prototype as "the system"—in keeping with the workshop participants' belief that they were interacting with a computational agent. When mentioning the keyboard player behind the scenes specifically, I will refer to him as "the wizard". The participants of the study will be referred to as "users". For the study, I had at my disposal the same two adjacent rooms used in the previous study (Chapter 6). I will refer to the two rooms as "the study room" (where the users tested the system) and "the operating room" (where the wizard played and I recorded the session).

5 This implies a double simulation. The goal was for the study participants to believe they were interacting with a computational system, albeit one that displays behavior that appears lifelike.

An XLR patching system made it possible to set up direct audio contact between the study and operating rooms. Users were only shown the study room. This room was prepared in such a manner that nothing could reveal the existence of the other operating room. A piano was placed along one wall (for participating pianists to use). A laptop computer was placed on a small table in front of where the users played. The system interface (described in the next section) was the only open application on this computer. The instrument microphone was not patched up to this computer. Instead, it went straight into the patching panel in the wall and into a sound card in the other room. This patching was obscured by the table and some items placed in front of the panel. Out of the same patching panel and into the user's headphones came the audio signal from the wizard's keyboard. These were the keyboard's native MIDI instrument sounds output as audio. The headphone patching was also obscured in a way that made it believable that it was coming from the user's computer sound card—presumably hidden from view. Summed up, in the study room, the user could only see a piano, the instrument microphone, and a small table with a computer laptop displaying the interface of the system.

In the operating room, the user's audio input went from the patching panel into the sound card's first channel, and the audio output from the wizard's MIDI keyboard went into the second channel. These channels were individually mixed for the wizard and the user, who both could adjust their own monitoring level (the wizard directly through the sound card interface, the user through the system interface). There was a second laptop computer in the operating room. This computer handled both recording the session and displaying the choices the user made in the system interface—essentially commands for the wizard to follow.

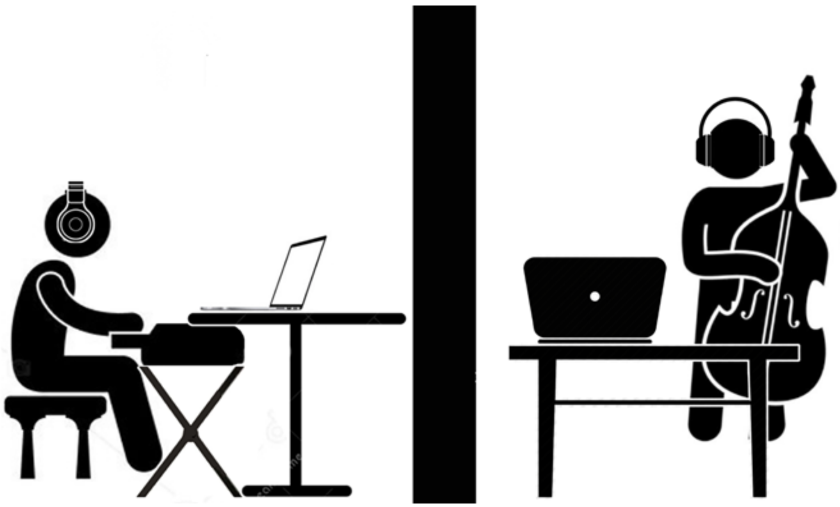


Figure 33. The set-up of the Wizard of Oz study.

Video equipment was set up in each room to film the processes in parallel. The video camera in the study room captured both the user and the computer screen. The camera in the operating room captured the wizard, the researcher (myself), and the computer screen. The video recordings served dual purposes:

- *Synchronization.* Seeing the wizard's and the user's interactions would make it easier to synchronize the audio and the interaction log during the analysis stage.
- *Back-up*. In case something went wrong with the audio recording, the video recording could capture the sound of the user's instruments, and the kind of interaction the wizard was engaging in on the keyboard.

The study was divided into five parts (see more details in Appendix C.3):

In part 1, I gave the user a brief introduction to the PhD project and outlined the agenda for this particular study. I also informed them how the video and audio recordings would be used. Then I read out the privacy statement, claiming full anonymity. I explained that I would be sitting in another room monitoring the recording. The user would then sign a statement of consent.

In part 2, I gave the user a brief introduction to the interface of the system. The explanation was formulated in a way that did not amount to telling untruths about the prototype. For instance, I did not refer to it directly as a "computational system" or a "machine". Instead, I used the term "prototype for the system". I also showed the user that explanations for the functionality of buttons on the interface were available as text by clicking on the "Learn more" links beside each button.

In part 3, the user engaged in a 15-minute session with the system. I explained that the user should experiment freely around a musical idea, and try to use the system to develop the idea through improvised interaction. Before leaving the room, I would make sure that the user had clicked the "Start session" button on the interface. I would then go to the operating room, and monitor the recording. The user could select which MIDI instrument that the system would be playing. The instrument choices were grand piano, electric piano, vibes, organ, synth pad and bass. The choice of using "cheap" native sounds from the MIDI keyboard was a deliberate ploy to mask any "humanness" in the wizard's playing (further explained in Section 7.1.3).

In part 4, I interviewed the user about the experience of playing with the system. The interview was semi-structured (see section 7.1.5), where the questions focused on eliciting immediate reactions to the interactive experience, the perceived autonomy of the system, how it differed to playing with musicians in more conventional setting, and the degree to which the interaction stimulated the user's creativity.

Part 5 began with a full disclosure of the Wizard of Oz method. After waiting for the user to express their immediate reactions to this revelation, I explained the rationale behind choosing this study method, and I asked them some additional questions (see Section 7.1.5). The remaining minutes of the interview were left open for additional comments or critique from the user.

7.1.3 The interface

As indicated in the introduction to this chapter, the basic idea for the interface I had in mind was an emphasis on communication as opposed to control. When human musicians make music together, they influence each other in various ways. However, they do not directly control each other's behavior (i.e. put electrodes into each other's brains and fire). My idea for an interface was inspired by this real-life context of musicians influencing each other, both through musical signals and through suggestions and requests that may or may not lead to desired kinds of behaviors. I expected that such an interface would lead to a focus on the performing agent instead of on the interface itself.

The interface used in the Wizard of Oz study was implemented in the Max graphical programming environment. The interface had two views. The front end was the user's view (Figure 34), and was a faithful representation of the envisioned design for the interactive music system's interface. In other words, this was the actual prototype. I will refer to this as the user's interface. The back end was the wizard's view (Figure 35), from now on referred to as the wizard's interface. The wizard's interface was designed to display the user's interface choices in large and easily discernable fonts on the computer screen in front of him. The wizard's interface also had a color-coding scheme to make it clear when the session was live ("IN SESSION"—green), when it was paused ("PAUSE"—yellow) and when in was offline ("NOT IN SESSION"—red). This was to ensure that the wizard never would make the mistake of playing when the system was not supposed to be live. The user's choices were sent to the wizard's interface via the UDP network protocol. This messaging is built into the udpsend~ and udpreceive~ objects in Max. The user's choices of MIDI instrument ware also sent via UDP messaging to the wizard's interface. From here, these choices were relayed to the MIDI keyboard. Therefore, the participants were controlling the wizard's keyboard MIDI instrument channel while he was playing.

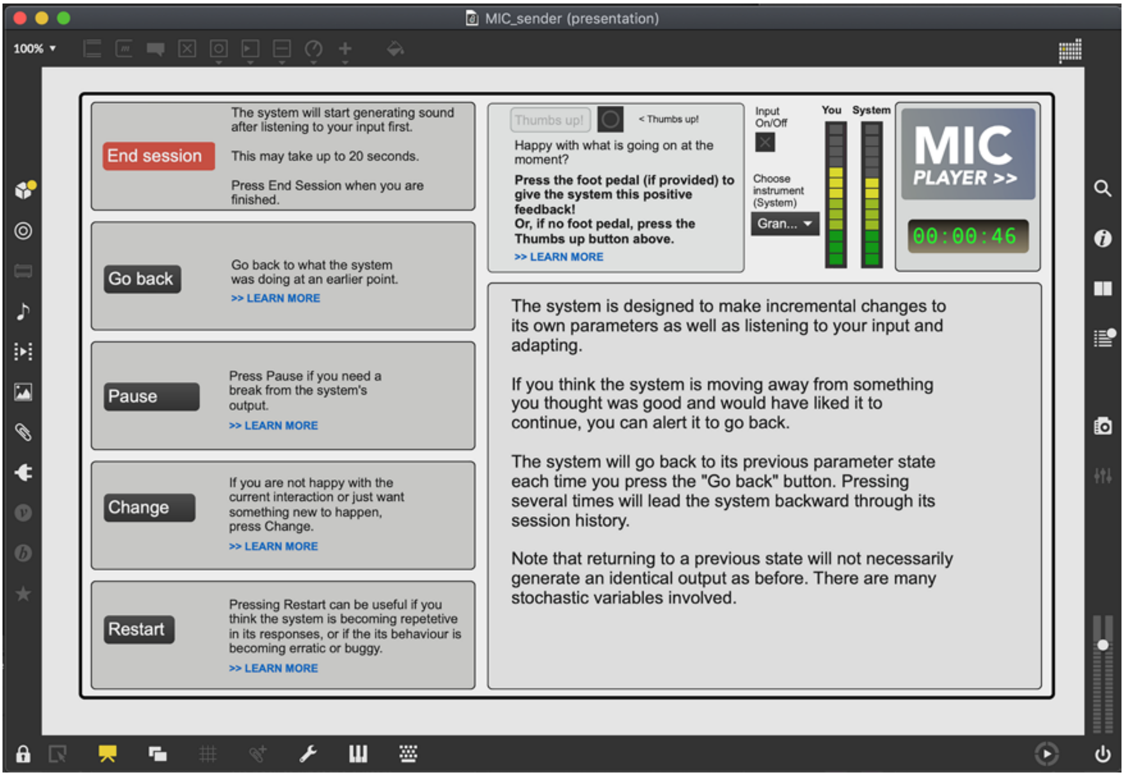


Figure 34. The user's view of the interface. At this time point, the user has just clicked the Go back button.

The user's interface—the prototype—had the following buttons: *Start session/End session* (toggle), *Go back*, *Pause/Continue* (toggle), *Change*, and *Restart*. Additionally, the user was provided with a foot pedal that could trigger a *Thumbs up* indicator. The functionality for each of these actions was explained to the user in "Learn more" links, which the user could click at any time during the session without triggering any action. The user was given time to read through these explanations before starting the session.

The explanation for each function is shown in Table 6. The left column shows the explanation that the user was given. The right column shows the explanation given to the wizard as part of his preparations for the Wizard of Oz sessions. In addition to these written explanations, I also made a video for the wizard one week prior to the study. This video demonstrated how the user and the wizard interfaces would work in parallel during a session.

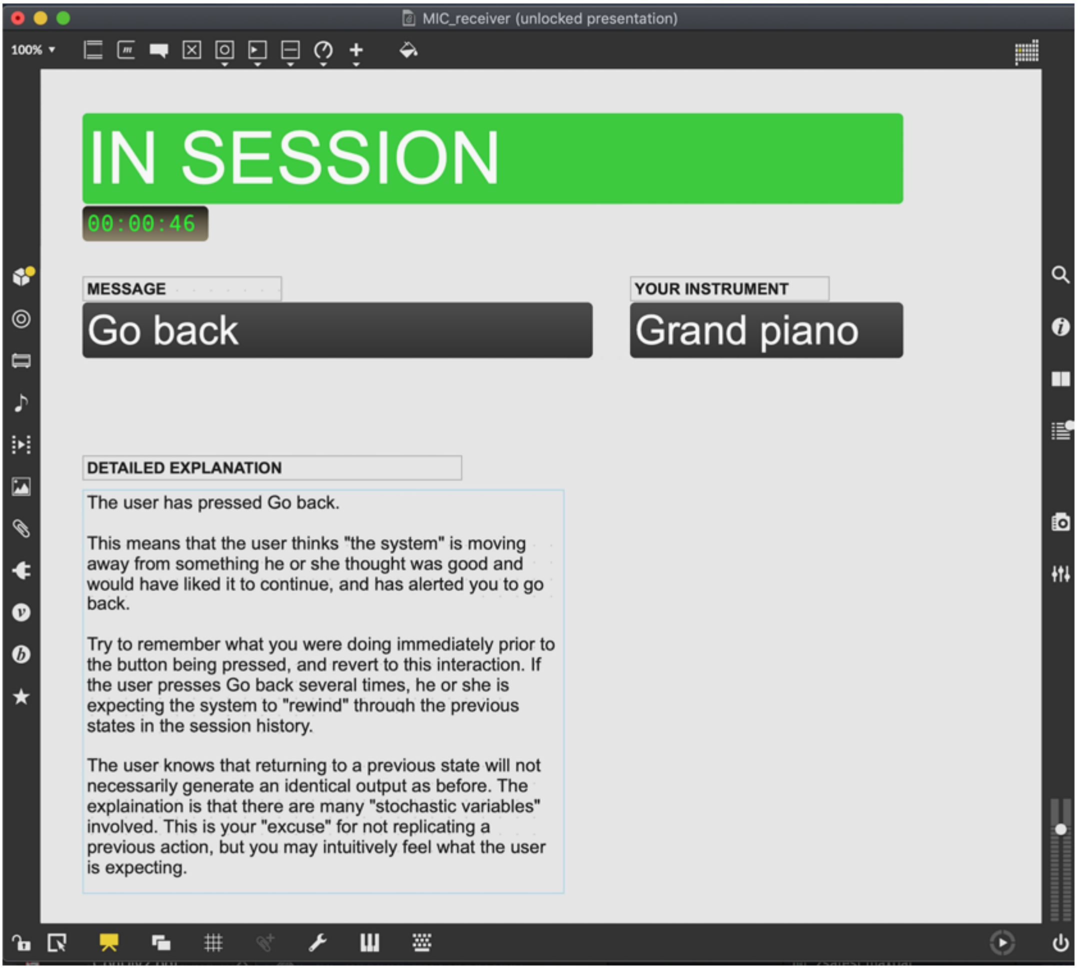


Figure 35. The wizard's view at the same time point. The Go back message is displayed in large fonts.

Explanation for the user	**Explanation for the wizard**
Start session	Start session
Give the system some time to adapt to your playing. If you have a specific idea (a theme/sequence), play it as a loop. Alternatively, adapt to the system's playing, if you like what you are hearing!	Let the user start playing first for at least 10–15 seconds. The user is told to give the "system" some time to adapt to his or her playing. They are told: *"If you have a specific idea (a theme/sequence), play it as a loop. Alternatively, adapt to the system's playing, if you like what you are hearing!"*
Go back	Go back
The system is designed to make incremental changes to its own parameters as well as listening to your input and adapting. If you think the system is moving away from something you thought was good and would have liked it to continue, you can alert it to go back. The system will go back to its previous parameter state each time you press the "Go back" button. Pressing several times will lead the system backward through its session history. Note that returning to a previous state will not necessarily generate an identical output as before. There are many stochastic variables involved.	The user has pressed Go back. This means that the user thinks "the system" is moving away from something he or she thought was good and would have liked it to continue, and has alerted you to go back. Try to remember what you were doing immediately prior to the button being pressed, and revert to this interaction. If the user presses Go back several times, he or she is expecting the system to "rewind" through the previous states in the session history. The user knows that returning to a previous state will not necessarily generate an identical output as before. The explanation is that there are many "stochastic variables" involved. This is your "excuse" for not replicating a previous action, but you may intuitively feel what the user is expecting.
Pause	Pause
It could be that you need a break, or time to figure something out for yourself and do not want to be disturbed. Upon pressing Continue, the system will remain silent until you start playing again. None of the parameters will have been changed during the break, so if you continue playing what you were playing just before the break, the system will interact in a similar way as where you left off (but not necessarily identically).	You have been muted because Pause has been pressed. It could be that the user needed a break, or time to figure something out for him or herself without being disturbed. When the user presses Continue, you will be unmuted. Let the user start playing first before you start interacting again. Your job is to keep the same interaction mode before and after the Pause (no parameter changes in the fictive "system").
Continue	Continue
(No new message)	Continue! You are now unmuted. Remember to let the user start playing first.

Explanation for the user	Explanation for the wizard
Change	Change
For new users, this will completely reconfigure the system's state in a random fashion. For returning users, the reconfiguration will be semi-random, but with weighting towards previously registered "attractor states". For example, if several "attractor states" show a tendency towards favorability for some particular parameter settings, there is a larger chance that the system will reconfigure with parameters closer to such settings. But this also depends on context, i.e. whether what you have currently been playing bears similarity to the contexts registered with the attractor states. Change can be pressed as many times as you like. As with Start session, give the system some time to orient itself and adapt to your playing before proceeding.	The user has pressed Change. The means that the user is expecting a complete change in interaction style. Just do "something completely different". According to the "system design", pressing Change will trigger a completely random change. For returning users, previously "liked" states (indicated by Thumbs up snapshots) will influence the direction of change. This is irrelevant for this experiment. Change can be pressed as many times as the user likes. As with Start session, let the user start playing first. Wait at least 10 seconds while listening, then start interacting.
Restart	Restart
In contrast to Change, pressing Restart will cause the system to ignore the session's history (previous states). It is essentially like starting a new session, except that the system retains the entire session history for machine learning purposes. It will also take a snapshot of the state at the time Restart was pressed, registering this as a "repeller state".	The user has pressed Restart. This is an indication that the user thinks the system is "stuck" and wants a clean slate. In contrast to Change, pressing Restart is supposed to cause the system to ignore the session's history (previous states). It is essentially like starting a new session, except that, according to the explanation, the system retains the entire session history for machine learning purposes. It will also take a snapshot of the state at the time Restart was pressed, registering this as a "repeller state".
Thumbs up!	Thumbs up!
This will be registered and entered into the machine learning algorithm. Pressing the foot pedal or the Thumbs up button will not alter the interaction, but a snapshot of the current parameter settings will be stored alongside features extracted from what you were playing prior to when the pedal was pressed (the "context"). The system accumulates data about "attractor states" for different users and different contexts.	The user has pressed Thumbs up, indicating that he or she is reacting positively to the current interaction. This is not supposed to affect the current interaction, but make a note of it and keep this in mind for the rest of the session. In the story of "the system", the Thumbs up will be registered and entered into the machine learning algorithm. A "snapshot" of the current parameter settings will be stored alongside features extracted from what the user was playing prior to when the button was pressed (the "context"). The system accumulates data about "attractor states" for different users and different contexts.
End session	End session
Session has ended!	Session has ended! Great work, you are free until the next session starts.

Table 6. Explanations for the prototype functionalities, as given to the user vs. the wizard.

The explanations provided in the above table reveals how my theoretical focus on dynamical systems influenced both the interface design and informed preliminary ideas about how to implement the functionality behind the interface. One of these ideas was to use reinforcement learning as a way to make the system learn the user's preferences through repeated use.

In the explanations, I was wary of crossing the line into untruthful deception. I found it acceptable to refer to the prototype as a "system" (the system in this case being a human being). In the places where I refer to a "machine learning algorithm" or other computational algorithms, I relied on the future tense in order to create ambiguity. These were envisioned functionalities, although the context would of course make the user assume that all of these functions had already been implemented.

7.1.4 Teaching the wizard how to behave

I recruited the wizard nine months before the actual study took place. The person I had in mind is a professional pianist, keyboard player, composer, technologist and music researcher. In other words, he was supremely qualified for the role of simulating a computational system. In the weeks prior to the study, we had several meetings where I meticulously explained the motivation for the study, and we discussed strategies for creating an environment that would make the user experience the study as realistic as possible.

It was important for me to keep a focus on interaction dynamics. The user would have to perceive clearly that the system was alternating between different kinds of behavior. For this purpose, I adopted four distinct behaviors proposed by Young and Bown (2010) as potentially emergent from engaging with an interactive music system (referred to as a *live algorithm*). The following behaviors and their explanation is paraphrased from Blackwell et al. (2012), and the same explanation was given to the wizard in writing two weeks prior to the study:

> **Shadowing**
> Shadowing involves a synchronous following of what the performer is doing, mapped into a different domain. In its simplest form, shadowing achieves strong but trivial participation, and little or no leadership, autonomy or novelty. However, even in this simple form, the appearance of coherence can have a strong effect for the co-performer, and can contribute to the sense of autonomy of the system, and the generation of novelty through its interactive affordances.

> **Mirroring**
> Mirroring involves some extraction of more abstract stylistic information or musical content from the performer, which is "reflected" back to the performer in novel ways. As with shadowing, the system predominantly takes the lead from the performer. This clearly demonstrates participation, and can contribute to a form of collaborative creativity through the opening up of new possibilities. As with shadowing, an appearance of autonomy comes with the sense that the musical output is coherent.
>
> **Coupling**
> Coupling refers to a system's behaviour that is largely driven by its own internal generative routines, which are perturbed in various ways by information coming from the performer. Coupling does not prescribe a specific behaviour, and may involve aspects of mirroring and shadowing (in the latter case the coupling would be tighter), but tends to refer to a situation in which the system can clearly be left to lead (by acting more independently of the performer), possibly to the detriment of the sense of participation (in which case we can think of the algorithm as "stubborn" or unresponsive).
>
> **Negotiation**
> Negotiation is defined as a more sophisticated behaviour that is related to coupling but is based on aspects of human cognition. A system that negotiates constructs an expectation of the collective musical output and attempts to achieve this global target by modifying its output. Since the collective musical output depends on the performer as well, negotiation, as the name suggests, may involve attempts to manipulate the behaviour of the performer, or equally, to adjust one's expectations in light of the direction of the music. (Blackwell et al., 2012, pp. 161–164)

The wizard rehearsed for the study on the basis of these explained interactive behaviors. Additionally, I gave him a written memo with instructions and things to keep in mind for the interaction. Essentially, this memo contained a set of reminders about how to be "a believable machine" (grounded in my own experience). Some of the instructions were:

- *Be insistent when first starting your response. This is the machine's current "state".*
- *If the input has a tempo, your response tempo should behave "overrulingly" when you have perceived/decided the tempo.*
- *If it seems like the user is trying to change the tempo, harmony or other parameters in another direction, adapt to this in discrete steps or "jerks", not gradually.*

- *It is better to stop entirely and wait for a new entry point if you are not sure about what to do at any given moment.*
- *Be attentive to musical genre, but try to appear as genre agnostic or generic.*
- *Do not think like a pianist. Your responses may be monophonic. This depends on the type of sound the participant chooses as preset.*
- *The transition between response types could happen abruptly.*

The entire memo is included in Appendix C.6.

7.1.5 Interviews

The semi-structured interview had two parts—one before and one after revealing that the users had been interacting with a human. The first (pre-revelation) part was the longest and most structured part, with the following questions prepared:

- *What was it like to play with the system?*
- *How did playing with an interactive music system differ from creative co-performance situations you are used to?*
- *How did it affect your creativity?*
- *Did you feel that there was some kind of negotiation about the initiative between yourself and the system?*
- *Did you feel that the system pulled you in a direction that you hadn't expected?*
- *Could you describe some positive and negative aspects about this tug-of-war, or initiative taking?*
- *Did the musical dialog lead to any new ideas? Examples?*

After going through these questions, I would reveal the simulation and explain the purpose of this set-up in the experiment. The second part of the interview had three questions prepared:

- *Did you at any point suspect that you were not playing with a machine?*
- *How would you have played differently if you knew it was a human?*
- *Are you relieved or disappointed that it was not a machine?*

I would keep this part of the interview open-ended, and give the user space to reflect upon the experience.

For transcribing and analyzing the interviews, I followed the same procedure as described in Chapter 5.1.3, with the exception that everything was done in Norwegian. When performing the inductive coding, I made codes in English, but kept the transcriptions in the original language. I have only translated the quotes and paraphrases used in the results section. The coding process produced three main themes:

- The impact of believing it was a machine
- Responsiveness vs. contrast
- Relating to an unknown other

After having completed the process of coding and annotating the interaction logs, described in the next section, I went back to the interview data and added one extra theme—Interface interaction—and moved a few of the relevant codes to this section. The themes and codes for this interview are available in Appendix C.4.

7.1.6 Interaction logs, behavior codes and annotations

The interaction logs produced CSV files (*comma separated values*) in the format *Timecode, Action*. This would produce overviews of the points at which the users had engaged in the buttons *Go back*, *Pause/Continue*, *Change*, *Restart*, and *Thumbs up*. As part of the analysis, I listened to the recordings and categorized which interactive behaviors (shadowing, mirroring, coupling, and negotiation) the wizard was engaging in around the points at which the users pressed these buttons. I added these categories as codes beside the actions in the log files. Of course, such categorization is open to interpretation, and sometimes ambiguous. However, since I knew that the wizard was well prepared and was instructed to make engage in behaviors according to these categories, I was able to categorize most cases quite confidently. For example, there were very few instances of real shadowing behavior (close following of the user's input), simply because this is not a very human behavior when the material is not known beforehand. This was probably the most difficult behavior for the wizard to simulate (and coincidently the easiest for a machine). Mirroring, on the other hand, was a behavior that was easy to categorize, and there were frequent cases of the wizard clearly mimicking or mirroring back the user's input. The behaviors coupling and negotiation were more difficult to discern. In cases where the wizard "diverged" from the user musically without appearing to be attentive to what the user was doing, I categorized this as coupling (largely driven by its own internal generative routines). I categorized sequences as negotiation if they included musical dialog of an interactive kind—a more sophisticated give and take than the other behaviors. Finally, beside each of these codes, I would write annotations explaining the musical interaction in shorthand. These annotations are the qualifying factors behind the behavior codes. In transition cases such as when the user had pressed the *Go back* or *Change* buttons, the annotations would include details about what happened before and after pressing the buttons.

7.1.7 Participants

The invitation to the study (Appendix C.1) explained that I was recruiting participants for a user test for an early prototype of an interactive music system developed for my PhD project. The target group was specified as musicians who:

- are familiar with collaborative music-making contexts (jamming for ideas)
- play a tonal acoustic instrument (not drums or percussion)
- are interested in contributing to the field of creative human-computer interaction

The invitation further expressed that I was not recruiting musicians from any specific genre, but participants “are requested to accept that the system is under development and has a limited number of interaction modes so far”.

The goal was to have four participants, and I did initially reach this target. Unfortunately, two days prior to the study, one of the participants let me know that he would not be able to show up. Despite intense efforts to find a replacement in time, the study went ahead with only three participants. They were all music students at the academy and in their twenties.

Jacob	Occupation: Bachelor music student (jazz) Instrument: Piano (originally a percussionist, see below) Background: Marching bands, orchestras and brass bands, pop, rock and jazz bands. Has worked with various forms of improvisation in duos and larger ensembles.
Andrew	Occupation: Bachelor music student (jazz) Instrument: Double bass Background: Jazz music. Interested in improvisation particularly the modern
Lisa	Occupation: Master music student (composition) Instrument: Piano Background: Plays in various groups who make music collectively (the same Lisa as from the first study—the focus group interview)

Table 7. The participants of the user study. Pseudonyms are used.

When Jacob expressed interest in participating with drums, I pointed out that I was looking for participants who play tonal instruments, as expressed in the invitation. I explained that the reason I wanted tonal instruments was that the system I was developing would focus mostly on extracting melodic and harmonic information from the audio signal, and that a focus on rhythmical elements would come at a later stage. While this was true, I was also secretly worried that it would be more difficult for the wizard to deal with both tonal and percussive instruments while maintaining a believable façade of being a computational agent. Therefore, I asked Jacob if he played any other instruments. He said he could play the piano.

7.2 Results

As explained in Section 7.1.5, the interview was divided into "pre-revelation" and a "post-revelation" parts. The lead-off question in the first part asked the users what it was like playing with the "system". The immediate reactions were mostly positive. "Fun", "exciting", "surprisingly absorbing", "organic" and "instructive" were among the adjective phrases used to describe the interaction soon after the session had ended. The opening question in the second part was whether they had suspected that they had been playing with a human. In response to this, none of the users suspected that they had been interacting with a human musician. Lisa had noticed that the system tended to have a delayed response when she pressed buttons on the interface, such as the *Go back* button. Despite this, she said she would never had suspected it was human.

Because the pre-revelation parts of the interviews were much longer than the post-revelation parts, most of the results presented in the following sections are genuine reactions to and reflections upon interacting with what they thought was a computational agent. In the cases where I present quotes or refer to statements made by the users from the post-revelation parts, I will explicitly point this out. Otherwise, the results are largely from the pre-revelation parts of the interviews.

7.2.1 Some notes on the interface interaction

Although the focus of this study is first and foremost the users' experiences of the activity of mixed-initiative music making, I will start with a brief summary of how the users engaged with the interface. The complete interaction logs along with behavior codes and annotations (explained in Section 7.1.6) are available in Appendix C.5.

The users used the interface very differently. Jacob was not very active with the interface. He pressed the *Thumbs up* pedal five times and changed MIDI instruments twice (from piano to vibes, then from vibes to synth pad). In the interview, he explained that he had chosen to see the session as a free improvisation, and was therefore prepared for long stretches of musical searching before getting to "something nice". This was based on his previous experience of improvisation, where it is common to accept that parts of the improvisation are an inevitable "transport stage". The times he pressed *Thumbs up* were all when the wizard was engaged in contrasting or independent behavior. I have categorized the wizard's behavior as either coupling or negotiation at these points.

Andrew was more active with the interface, but also only engaged with the *Thumbs up* foot pedal. A part of the reason for this could be that his instrument—the double bass—is unwieldy, and using the interface apart from the foot pedal would entail maneuvering around the instrument to get to the laptop. Nevertheless, he used the *Thumbs up* foot pedal 18 times. He also changed MIDI instruments once (from piano to synth pad) with about four minutes remaining of the session. The situations he responded positively to were mostly when the wizard was engaging in coupling or negotiation—a preference he also mentioned as "contrasting" in the interview. Only three of the *Thumbs up* indications were given after the wizard had responded to Andrew's gestures (mirroring). Interestingly, two of the *Thumbs up* indicators came after the wizard had stopped playing altogether. This could either be because Andrew appreciated the break, or it could be an indication of approval of a larger sequence preceding the pause.

Of the three users, Lisa engaged with the interface the most. She pressed the *Thumbs up* pedal 14 times, the *Go back* button six times, and the *Change* button once. She also changed MIDI instruments once (piano to bass). As with the other users, her *Thumbs up* responses were mostly given in situations when the wizard was engaging in coupling or negotiation—she described this as the system "going places" or "going other places". Like Andrew, Lisa also commented on her preference for these "other places" in the interview. She lamented the fact that the system "followed me too much", and was most positively surprised in the instances where it had stuck to more independent behavior for prolonged periods, thus giving her time to develop themes of her own. In fact, she apparently used the *Go back* button actively precisely for this reason. In two of these situations, when the wizard went back to the previous behavior, Lisa indicated her approval with the *Thumbs up* pedal and resumed what she had been playing. In one situation, she pressed the *Go back* button twice in a row. The wizard then had to remember how his behavior had changed over the course of a longer period. Judging by the audio recording, he appears to have remembered this correctly. After twenty more seconds, the wizard once more changed his playing style, whereupon Lisa pressed *Go back* for the third time in half a minute.

According to the interview, the one time Lisa used the *Change* button was due to a high-pitched piano sound that was painful in the headphones. The video recording supports this—she lifts the headphones away from her left ear while frowning and reaches for the *Change* button. The wizard then commences with a mirroring behavior in a lower register, whereupon she places the headphones back and presses the *Thumbs up* pedal.

7.2.2 The impact of believing it was a machine

All of the participants expressed that their focus had been influenced by the (presumed) absence of a human counterpart. In the first part of the interview—before the simulation debriefing—Jacob shared that his approach had been driven by curiosity about "what the machine wanted to happen". Because of this, he had not paid as much attention to musical rules as he normally would have: "I didn't try to play major or minor, sort of. That didn't matter to me". Based on what I have heard in the recording of Jacob's session, this attitude is obvious by the way he approached the piano, which as detailed earlier is not his main instrument. He did not play any traditional chords. Most of the time, he engaged in monophonic themes interspersed with occasional single or dual note accompaniment in the bass register. He appeared to favor atonal material, or perhaps it is more correct to say that tonality was, as he himself expressed, irrelevant.

After revealing the simulation, Jacob expressed both surprise and embarrassment. As a comment to how he would have played differently if he had known it was with a human, he told me that he would probably have been more prejudiced against "how that human had reacted". He also referred to the fact that piano is not his main instrument, and that he probably would not have had the courage to play at all if he had known that he would be playing with a human. He admitted that his attitude had been more analytical than musical. He also claimed that he had played mostly in order to provoke different reactions from the system: "[I was] just… very curious, you know, about how it would react."

In the first part of his interview, Andrew pointed out that playing with the system made him more aware of his own choices, and he was more focused on playing something that could lead to interactivity compared to when playing with people. He contrasted this to a normal context where he would be more focused on "playing what he hears". He explained that he had begun the session by trying out themes that were "relatively basic harmonic". He claimed that the system "didn't pick up on this", so he decided to move on to a more free improvisation style. This switch is quite clear in the audio recording and happens at around two minutes into the session. Once he had adapted a more carefree attitude toward tonality, he claimed that it did not feel that different to playing with a human. When asked post-revelation if he would have played differently if he had known that he was playing with a human, he replied that he did not think that would have made much of a difference. Especially toward the end of the session, he had felt it was "about the same as playing with a human".

For Lisa, the belief that she was playing with a computational agent gave her a heightened sense of security. While still under the impression she had played with a machine, she said:

> [It was] much safer, for me (laughs). I still think it's challenging to play with others, often. [This was] like playing with your best friend. It's, sort of, very safe then. So you could, in a way, do anything and it's still okay. So no pressure. That's perhaps the biggest difference.

She summarized the session as fun and "very instructive" due to the freedom this sense of security gave her: "I was able to check out a lot of things that I never would have done with people."

So far, I have summarized how the presumed absence of a human counterpart consciously affected the attitudes of the users. The interview also revealed the users' implicit assumptions about computational systems. For example, both Jacob and Lisa commented on the inferior quality of the synth sounds as if that was something to be expected from a computational system. As such, my ploy to use these sounds as a way to cover up the humanness of the wizard turned out to be successful.

Jacob was amused by the fact that the system would repeat things after using the Thumbs Up pedal. He surmised that this was a machine-like feature:

> J: You could hear… that was fun… when I liked something, you could hear that it uses those things again. What you, in a way (laughs)… Yeah, so that's cool that it comes back, sort of, in different… that's what it does, isn't it? A little?
>
> N: Yeah, it's kind of… like a mirroring thing…
>
> J: Yeah. Yeah, so that's cool. So, in a way you could hear that it's not a human, maybe.

Andrew's conclusion that the system was agnostic to tonality reveals an assumption about the musical capacity of computers, or lack thereof. The fact that he "gave up" on tonality early on is interesting when listening to the audio recording. According to the interaction log behavior coding, and annotations, the wizard is apparently engaging in a mirroring behavior at this stage in the interaction. He even attempts to repeat some of Andrew's intervals, although the wizard may not be aware of the exact harmonic context of the scale Andrew is playing on the double bass. In fact, the harmonic framework at this point is not easily discernable on the audio recording either, due to the sometimes tonally ambiguous sound of double bass notes in the lower registers. In reality, a dedicated computer algorithm would probably have been capable of classifying the tonality more quickly than many human musicians under similar listening conditions would.

Lisa expressed a similar sentiment to Andrew's in regards to the system's musicality. After first having expressed how she had enjoyed playing with the system because of the absence of social pressure, she added:

> And of course, on the other hand, the response I got was obviously often not human […], not so sensitive, perhaps. And it wasn't… it didn't feel like it was conscious, in a way. […] It did suggest contrasts and followed me, but it was a bit random and not that musical.

I asked Lisa how this realization affected her creatively. She replied that she needed to adapt to this context and explore what she could do together with the system's output, which was very fun for her. She felt that she could "explore width", which she does not usually do with other musicians: "[With people] I usually stay in one place and elaborate, in a way, where we find ourselves, while here I could go other places". As such, she said it was very creative.

7.2.3 Responsiveness vs. contrast

Judging by instances of *Thumbs up* in the interaction log and the behavior codes associated with these instances, the users showed a clear preference for the coupling and negotiation behaviors, which are more contrasting and independent compared with shadowing and mirroring (overt responses to the users' musical gestures). While these statistics were not available to me during the post-session interviews, their comments largely align with this impression. Two of the participants specifically mentioned that the least interesting parts of the session were when the system was being too imitative. The following section from my interview with Andrew illustrates this point:

> N: Did you feel that the system pulled you in a direction that you hadn't expected?
>
> A: Absolutely.
>
> N: Yeah. Could you describe some positive and negative aspects about this tug-of-war, or initiative taking? Um… first the positive.
>
> A: Well, the positive is that it's very organic, really. Or it reminds me of something that is very organic.
>
> N: Any downsides to it? I mean were there any situations where you felt that your initiative was not heard, in a way?

> A: Sometimes. But I think what I personally disliked the most was… I don't know if this answers your question, but often I could tell that it was very quick to imitate what I was doing, and played close to it. Instead of playing against it…
>
> N: Yeah. Okay.
>
> A: … and making a contrast like that. But there were a few places where it also pulled in a different direction. And I think that… those sections were really the most exciting.

An exchange with Lisa confirms a similar preference:

> N: While you were playing, did you reflect upon the system's responsiveness or, like, agency?
>
> L: Yeah, maybe that it was easily, in a way, tricked into going places. For better or worse. So, in a way, I was afraid to leave it because I knew it would follow me at some point.

However, Lisa also noted that the system seemed to have its own will, which she experienced both in positive and negative ways:

> N: Did you feel like there was some form of negotiation about the initiative between yourself…
>
> L: Yeah. And it was very, like, aggressive, in a way. Or, like, I expected it to just follow me. But when I understood that it went places I was like: "Okay, we can go there". And then I would, like, in a way, try to find something nice there, and then, in a way, make a choice: "Okay, I don't want this", in a way.
>
> N: Yeah, so sometimes it was like you were pulled in directions you hadn't expected?
>
> L: Yeah, yeah! Definitely.
>
> N: Could you describe some positive and negative aspects about this tug-of-war? You just mentioned that you hadn't expected…

> L: Yeah. It's, in a way, to repeat myself… the positive aspect was that I was taken places I never would have gone, which were very nice places and exciting to me. And very nice just to practice being with someone and try to go against it and stuff. And the negative was that, maybe sometimes I wanted to stay somewhere but then suddenly it would go somewhere else, and then: "No, I don't want to […] I want to stay here!"

Desire for contrast was also exemplified by Jacob in a part of the session where he seemed to be playing completely independently from the system. After a few minutes of playing without seeming to interact with the system sonically or through the interface, I became worried that he had forgotten about the experiment and was lost in his own playing—only listening to himself. I instructed the wizard to simulate a glitch by hanging on one tone until Jacob would be forced to press *Restart* and start listening again. However, he continued playing for a total of 75 seconds—seemingly unfazed by the jarring inertia of the hanging note. When the wizard finally gave up this tactic and started playing again, Jacob finally pressed the *Thumbs up* pedal, signaling that he was in fact listening. When I asked about this "glitch" in the post-session interview, Jacob exclaimed: "Yeah, but that was cool!" He described it as a drone-like quality that accommodated a build-up of intensity and energy.

7.2.4 Relating to an unknown other

The notion of mixed-initiative interaction—the synthesis of human and computational creativity—hinges upon the sharing of decision-making between the human user and the creative agent. Without some degree of autonomy granted to the system, it remains a creativity support tool. On the other hand, if the system is fully autonomous, the human becomes an operator of a computationally creative system. In a mixed-initiative creative interface, a certain shared sense of agency is a prerequisite. Therefore, it is of interest to take a closer look at the language the users used to describe how they related to the system.

Throughout the interviews, the users were clearly fumbling for adequate terms to describe the system. There are hints of anthropomorphism in several statements. As already mentioned, Jacob was curious about what the machine "wanted to happen". After saying this, he quickly added: "I don't know if it thinks like that, but…" Andrew apparently felt a similar need to qualify the way he anthropomorphized the system:

> I felt… at least toward the end […] that I was able to recognize, or predict the computer's choices to a certain degree. I managed to [anticipate how to play] considering

> who the computer… I'm saying… quote "who the computer" (makes quotation gesture) … if you know what I mean…

As seen in the previous section, Lisa expressed that it was "nice to practice being with someone". At one point, she inadvertently switched to a masculine pronoun when referring to the system: "He had these bass melodies…" I made a note of this and confronted her with this a few minutes later:

> N: It was interesting earlier when you talked about the system, you said "he"?
>
> L: Yeah… "he". Yeah, okay (laughs). I was definitely not thinking "she".

Jacob entered the session with the clearest conception of "being the boss": "I took on the role, like, I'm in charge, you know… for better or worse. I just went full speed ahead with what I wanted to do."

This attitude may have left little leeway for Jacob to reflect consciously upon the system's perceived autonomy. When probed whether he felt that any negotiation had taken place, however, he described moments where he had liked what the system was playing, and he had tried to play along. There were also times he found the system quite assertive and came up with material he had not expected. Curiously, he switched to a second-person perspective when describing how he reacted to this perceived assertiveness: "Of course, sometimes it was actually quite assertive, or like, came up with things I hadn't quite expected. That's true, it did. And then *you* react to that, of course" (my italics). However, when I further probed him about how he perceived the agency of the system and if he felt that there had been some kind of "battle" for the initiative in the interaction, he was adamant that he had been the one in charge. "I really didn't think of it in that way […] because of the attitude I had when I started playing, that I was kind of the boss".

The users' reactions to the revelation that they had been playing with a human provide compelling insights into how they mentally constructed their creative counterpart. Lisa, who had expanded on the "freedom" she felt in playing with a nonhuman, was visibly ruffled upon learning the truth. This is understandable when considering she had opened up about "going places" she would not have dared to go with musicians. She also acknowledged having been less prejudiced toward the more erratic behaviors of the system. Instead of "staying in place" and trying to provoke responses to her own input, she had let herself be led by what she deemed were "random" and "not so musical" contrasts. After regaining composure following the apparent unease at having exposed herself to another musician unknowingly, she acknowledged that a willingness to leave her own comfort zones and venture into unknown territory was the biggest difference from the co-performative situations to which she is accustomed.

In contrast to Lisa and Jacob, Andrew was not as self-conscious about his own playing after the disclosure of the Wizard of Oz experiment. He thought it was “awesome” that he had been playing with a human, and expressed some relief upon learning that programming a system equivalent to the one demonstrated in this study would require a tremendous amount of work: “Yeah, I’m glad to hear that, because this is a black box to me. I have no idea how anyone can conceptualize a framework for anything like this at all”.

7.3 Discussion

Despite the small sample size, the Wizard of Oz study produced very useful and interesting data.

7.3.1 The interface as a mediary

Although the interface was not the focal point of this study, I included an overview of how it was used. I would like to emphasize that the frequency of a user’s engagement with the interface is not indicative of their level of engagement in the overall user experience. Most of the interaction is designed to happen as an interchange of musical signals. As such, a user’s lack of focus on the interface is more likely an indication of deep involvement with the musical agent than a sign of general disengagement. Recall, for example, that the two musicians in the second study completely ignored the interface, but were so deeply engaged in the interaction that they exceeded their time limit and I ultimately had to interrupt them.

With these considerations in mind, the interface design worked as I had envisioned. All of the users were preoccupied with the nature of the “machine” they were interacting with, and appeared to view the interface as a mediary as opposed to an instrument or a control surface. The *Thumbs up* pedal was used the most, probably in part due to its hand-free feature. Another reason for the preference for the *Thumbs up* button could be due to a social custom of encouraging rather than criticizing fellow musicians when jamming or improvising. Although they thought they were jamming with a nonhuman agent, several statements made by the users imply that they brought this attitude to the study session. For example, Jacob expressed that he was prepared to work through less interesting sections musically as is customary in improvisation culture, and therefore did not engage with the *Go back* or *Change* buttons.

Andrew and Lisa both said that they adapted their musical approach in accordance with what they deemed was the musical context.

Out of the three users, Lisa was the only one who used any of the other buttons. Her use of the *Go back* feature worked as intended (kudos to the wizard who managed to remember his previous "states"). She only pressed the *Change* button when the interaction turned uncomfortable to the ear. The *Restart* button was not used at all. I think the intended functionality of this button is the least clear from a user perspective. During the analysis of this study, I have come to realize that it is, for now, a redundant function. Ending a session and starting a new one is essentially the same thing, although from a designer perspective a clean slate in an existing session could be a useful feature for keeping track of a global trajectory over the course of several attempted sub-sessions. In the scope of the first design iterations, however, the *Restart* button can be abandoned.

7.3.2 Freedom from judgement and freedom to explore

One particularly interesting finding in the study is the extent to which the users were willing to leave behind their usual comfort zones or modes of expression. All three users were quick to conclude that the system was tonally or harmonically agnostic, and they all paid less attention to normal musical conventions as a result of this assumption. In turn, at least two of the users (Jacob and Lisa) explicitly stated that they were emboldened by the notion of a faceless system that withheld judgement and allowed them to play—in the fun sense of the word—without having to worry about sounding good or being correct. They also admitted to having been less critical of the system's output than with human musicians. Apparently, this tradeoff—"less musicality" versus being free of bilateral judgement—was a prospect they found exciting. The assumption of "less musicality" also made the users positively surprised when the system appeared to respond in ways they had not expected.

Some of the findings from the first study—the focus group—are particularly relevant to this discussion. In Chapter 5, I summarized that maintaining a process-oriented approach of "going with the flow" and attaining shared ownership by decentering and reaching a collective subjectivity may increase the likelihood of generating emergent novelty in collective music making. The interviews following this Wizard of Oz study reveals that the users assumed precisely such attitudes when interacting with the system. They quickly abandoned premeditated strategies when confronted with a perceived lack of "musicality" on part of the system and embraced interaction on terms they seem to have interpreted as "common ground". They accepted the system "as is" despite apparent shortcomings. Furthermore, judging by the reactions of Jacob and Lisa, they had gone out on a limb because of the belief that they were not being evaluated by a human. Lisa expressed this figuratively as going places she would not have gone with people—except with a "best friend". Arguably, the shedding of inhibitions is an important part of any decentering effort. Habits are formed because learned actions nurture self-confidence. Exposing oneself to situations where new unlearned actions take place in a social context takes courage. Lisa felt that playing with the system was a practice in "being with someone". Using terms from previous chapters, I argue that this is essentially what is meant by decentering and reaching a collective subjectivity.

To sum up the above, the belief that they were playing with a computer helped make the users less self-conscious, and they were quicker to meet their counterpart on different terms than usual. An open question is whether this attitude would be preserved if the interface afforded more controllability. Without making any definite claims, I argue that a certain lack of control on the user's part may be what makes mixed-initiative music making possible. The more parameters that exist that enable the user to bring the computational agent closer to a preferred state, the more premeditation becomes a factor and the more the interface will resemble a tool or an instrument.

7.3.3 Dialog through contrast

I have not analyzed the sessions in this study as meticulously as the improvisation in the previous chapter, opting only to annotate what kinds of interactions took place around the points where the users engaged with the interface. These annotations show that most of the positive feedback occurred when the wizard was engaged in contrasting or oppositional behavior. The users' own statements confirm this to a considerable extent. Both Andrew and Lisa expressed preference for the contrasting behaviors and thought that the system tended to follow them too much. Furthermore, although Jacob was amused by instances where the system would mirror back phrases he had played, he also seemed content to spend sustained periods articulating themes without being complemented by the system. In particular, his reaction to the nearly two-minute long hanging note is illuminating. For one, Jacob might have been unaware of how non-interactive the described sequence sounded to external listeners. Secondly, he may have lost track of time and been fully absorbed in the music. The wizard reverting to a more normal, complementary style of playing coincided with what Jacob apparently felt was a climactic interactive moment, where the tension of the sustained drone was released. In all these examples, the dynamics of converging and diverging are described using different concepts. Lisa used the spatial concepts of "staying in places" versus "going places". Andrew expressed the dynamics as "playing close to" as opposed to "playing against" or "making a contrast". Jacob referred to a build-up of tension that was released. Once again, the dynamics of pulling together and pushing apart appears to be an important part of the collective music making experience.

The way that the users spoke about the system during the interviews reveal that they were both intrigued and unsure about the nature of the otherness they were dealing with. Judging by the language they used, there is a clear sense that they believed they had taken part in a dialog. On the one hand, they referred to the system as an agent with a capacity to influence the musical choices that they made. Interestingly, as the user who most adamantly claimed to be "in charge", Jacob inadvertently switched to a second-person perspective when admitting that such an influence had taken place ("*I* was the boss" versus an unexpected assertiveness that "*you* react to"). On the other hand, the users also emphasized their ability to influence the system, such as Lisa's claim that it was "easily tricked" into following her. The users' desire for contrast was accompanied by a need to understand "who" they were in a dialog with. In an exposition on the concept of being musical with an "other", Benson (2003) writes:

> A dialogue is only possible when each partner *both* holds the others in tension—that is, holds the other accountable—*and* feels the tension of accountability exerted by the other. As strange as it may sound, these "tensions" actually make the "freedom" of dialogue possible. Why that sounds strange is because we usually think of freedom as "negative freedom"—freedom *from* constraints. But what I have in mind here is "positive freedom"—freedom *for* genuine dialogue. Of course, in order to "feel that pull," one needs to be able to listen to the other. (p. 171)

When Andrew and Lisa complained that the system tended to be too imitative, and that they preferred the parts where it had offered forms of resistance or contrast, I interpret this as a longing for a "freedom for dialog". They were apparently listening for new voices—voices of otherness—and seemed genuinely interested in what their own responses to these voices would be. Benson posits that "to treat the other *as other* requires that I recognize the other as having a kind of claim on me" (Benson, 2003, p. 167). Following Benson's reasoning, this requires a stance of humility and an openness to encounter an other who, in Gadamer's words "breaks into my ego-centeredness and gives me something to understand" (Gadamer, 1997, p. 46).

On his part, Jacob was curious about what the machine "wanted", but seemingly wary of attributing it too much agency. Through subtle disassociation by directing the directionality of otherness away from himself (*I* became *you*), he suppressed the otherness of the nonhuman entity with which he thought he was interacting. According to Levinas (1996), the suppression of otherness is deeply engrained in Western philosophy (and by extension its culture) and is driven by a desire for autonomy. By insisting on being in charge and downplaying the system's agency, Jacob was seemingly motivated to play *in spite of* constraints rather than letting the constraints guide his performance. From the perspective of a listener, some parts of this particular session to me sounded more like dual monologs than a dialog. However, this does not detract from the intrinsic value Jacob may have felt while engaged in the interaction.

7.4 Summary

The third out of four studies used a Wizard of Oz method to better understand how musicians experienced engaging in a mixed-initiative music making session using a prototype of an envisioned interface. The musicians (the *users*) were led to believe that they were playing together with a computational agent when in reality there was a human keyboard player playing the role of the computer. The keyboard player (the *wizard*) was instructed to engage in four different interactive behaviors: *shadowing* (a close following of the user's input), *mirroring* (a reflecting back of the user's input in novel ways), *coupling* (more independent behavior perturbed by input from the user), and *negotiation* (a higher level behavior that may involve attempts to modify the behavior of the user or to adapt its own behavior according to the direction of the music).

The implementation of the study was successful, and all three users believed they were playing with a computational agent. The interface worked as intended—as a mediating layer rather than a control surface. The buttons were used as envisioned with the exception of the *Restart* button, which was not used at all. I will not include this button in the next iterations of the interface. The interviews with the users revealed that they had taken a different attitude to musical interaction than they normally do with human musicians. They had lower expectations to the musicality of the system, and therefore felt less constrained by musical rules. A perceived freedom from judgement allowed two out of the three musicians to feel less self-conscious about their own performance. The carefree attitudes the users were able to assume concur with two of the attitudes identified as engendering emergent novelty in the first study—a process-oriented approach and an awareness of the collective. Thus, repeated exposure to mixed-initiative music making could be good practice to habituate such attitudes in general.

The users expressed a clear preference for the system's more contrasting or oppositional behaviors. This desire for contrast was also evident in the way they signaled approval through the use of the *Thumbs up* pedal during the sessions. Judging by the language they used when describing the system, the users were both intrigued and unsure about how and to which degree they should attribute agency to the system. They all intermittently used anthropomorphic terms when describing the system's will and capacity to influence and be influenced by the musical direction. Two of the users openly admitted this mutual influence and were excited by the prospect of engaging in a creative dialog, whereas one user insisted that he had been in charge.

The user study described in this chapter has demonstrated that the user interface that was tested affords mixed-initiative music making, and that the technical implementation of the four interactive behaviors of shadowing, mirroring, coupling, and negotiation is a viable path forward. The study also showed that a genuine musical dialog with a computational agent is not solely dependent on technological factors, but requires the human user to relax the notion of control and give space for the agency of the system to manifest itself collectively.

8 Study 4: Spire Muse development and evaluation

As I embark on this chapter, I emphasize that the fourth study of this thesis was conducted after a prolonged period of software development. Whereas Studies 1–3 were conducted successively during a 5-month period, Study 4 happened 18 months after Study 3. This chapter will begin with a review of the development period, where I used findings from the first three studies combined with experimentation of other interactive music systems to design what became a prototype for a mixed-initiative interactive music system: *Spire Muse*. I will also provide an overview of the system architecture and interface before presenting the user study.

During the software development phase, I formed a working hypothesis that relinquishing control and giving up a degree of agency to the system potentially could lead to a more creative experience on the user's part than when being able to control—and hence predict—most aspects of the system's behavior. This hypothesis was supported by provisional conclusions based on findings from Studies 1–3. For example, I surmised from Study 1 (Chapter 5) that the attitude of seeking a collective creativity with the possibility of emergent novelty suggests that a mixed-initiative interactive music system should appear to have a "will of its own"—a contrasting space from which it can diverge and converge to its human counterpart. Study 2 (Chapter 6) showed that the musicians in a case study engaged in behaviors indicating that converging and diverging dynamics indeed may be essential to collective music making practice. Study 3 (Chapter 7) revealed that musicians who believe they are interacting with an autonomous computational system are willing to accept music making on different terms than when they play with humans, and that this could lead to pleasant creative surprises. Thus, the above-mentioned hypothesis supported by these findings formed the basis for the fourth and final study in this thesis. After the development of Spire Muse, a second more manually controllable version of the prototype was developed for a comparative user study, where the tradeoff between user control and system autonomy was a central premise.

I have previously presented Spire Muse in a paper at NIME (Thelle & Pasquier, 2021). Some concepts behind the system and the architecture are reformulations from this paper. Throughout this chapter, I will clearly indicate such instances by referring to the NIME paper. All text related to Spire Muse and the concepts behind it are my own contributions and part of my PhD work. Pasquier's role as co-author in the NIME paper relates to an underlying agent architecture based on self-organizing maps (MASOM) first presented by Tatar and Pasquier (2017). When writing about these elements of the software, I will refer to the latter paper. Thus, the distinction between Spire Muse and MASOM will be made explicitly clear throughout the chapter.

8.1 Designing Spire Muse

The name Spire Muse is based on the notion of growing music from a seed, with improvisation having a cultivating effect. Hence, "spire" is a reference to the definition of the word as "a sprout, shoot, spike, blade, tapering stalk of grass". The contention is that music creation starts with a *spire*—a phrase or a sound object. Musicians—in*spire*d by its sound, re*spire* life into compositions by improvising around the idea, adding layers, growing complexity. Seemingly, the music takes a life of its own—it a*spire*s to grow. In the following sections, I will present a chronological account of how Spire Muse was developed, from its conceptual foundation through to a fully functional interactive music system.

8.1.1 Synchronization of frameworks

In Chapter 6, I used a framework of *convergence vs. divergence* to analyze the formal dynamics of two musicians engaged in a free improvisation session. Specifically, I adopted Canonne and Garnier's (2011; 2012) model for collective free improvisation—derived from dynamical systems theory—in order to focus on the strategies the musicians applied to form collective sequences and articulate between sequences. More generally, however, the convergence vs. divergence framework includes several different models of collective music making, all of which revolve around interactive strategies focused on "pulling together" and "pushing apart" (Thelle & Pasquier, 2021). Figure 2 in Chapter 2 shows these strategies mapped onto a musical similarity axis ranging from converging to diverging. In Chapter 7, I applied an *interactive behaviors* framework as a guide for the keyboard player who simulated the computational agent in the Wizard of Oz study. The behaviors—shadowing, mirroring, coupling, and negotiation—are envisioned as potentially emergent in interactive music systems rather than directly programmed (Blackwell et al., 2012, p. 161). The simulation of distinct behaviors was an important part of the experiment because it allowed me to focus on how the users reacted to the interaction dynamics.

The two frameworks have some similarities. They both involve interactive strategies in music making, and the strategies within each of the frameworks appear to have ranges appertaining to a level of togetherness between the agents that contribute to the musical signal. However, the frameworks actually belong to different domains, and there are good reasons why they should not be thought of as interchangeable. The convergence vs. divergence framework is a conglomeration of mostly ethnographically derived models explaining how musicians apply strategies for collective music making. Although Canonne and Garnier's model for collective free improvisation is derived from dynamical systems theory (Canonne & Garnier, 2011), it was corroborated and elaborated upon using an ethnographic study (Canonne & Garnier, 2012). Hence, the convergence vs. divergence framework belongs to the real-world domain of collective music making, and may be applied in explanatory or predictive ways. On the other hand, the interactive behaviors framework belongs to the domain of interactive music systems and is devised as a prescriptive or conceptual framework for design. The former framework can be applied to *understand practice,* whereas the latter is a potential framework to *inform design*, as illustrated in Figure 36.

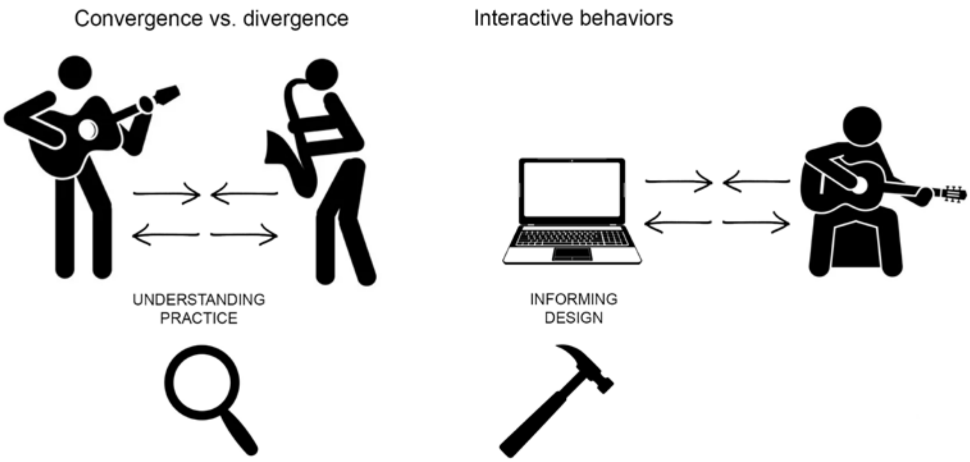


Figure 36. Convergence vs. divergence as a framework for understanding practice, interactive behaviors as a framework for informing design.

In Chapter 4, I recounted the realization in first-wave HCI that theory rarely has any significant impact on design. This was suggested as being attributable to the need for "intermediary" representations which differ in kind depending on whether they are called upon for software design or for modeling phenomena (Long, 1989; Barnard, 1991). Keeping this pitfall in mind, the advantage of the four-study plan and the use of different frameworks in the second and third studies (Chapters 6 and 7) becomes quite clear. First analyzing the real-world phenomenon of collective music making using an explanatory framework in one study, and thereafter simulating a software interface using a conceptual framework for computational interactive behaviors in the next study made it straightforward to maintain an awareness of the gap between the two domains. While the frameworks are not interchangeable, they can be synchronized. Taking a page from J. M. Carroll (1990), the convergence vs. divergence framework may be viewed as a "discovery representation" to be used as *a design rationale for an artifact*, whereas the interactive behaviors framework may be seen as an "application representation" which is *a collection of user-interaction scenarios* (p. 323).

Upon delving into research on interactive music systems and considering what kind of agent architectures could support mixed-initiative music making with emergent behaviors corresponding to shadowing, mirroring, coupling, and negotiation, it became clear that negotiation is different in kind from the other three behaviors. Whereas the other three behaviors in theory could be embedded as different modes in the software itself, negotiation is a type of behavior that would depend mostly on how the computational and human agents interact and influence each other. As such, negotiation could be viewed as the "meeting space" where the computational agent trades decision making with the human user. Based on this, the interactive behaviors framework was refined by conceptualizing shadowing, mirroring, and coupling as interactive modes along a *system autonomy axis* ranging from *reactive* to *proactive* (Thelle & Pasquier, 2021). In this updated framework, negotiation is envisioned as happening when the system switches between these three modes, either autonomously based on what it "hears" in the collective input or through influence through the user interface. Now viewed as an interface-layer behavior that requires the sharing of decision-making, negotiation does not map directly onto the autonomy axis and is placed above the other behaviors in Figure 37.

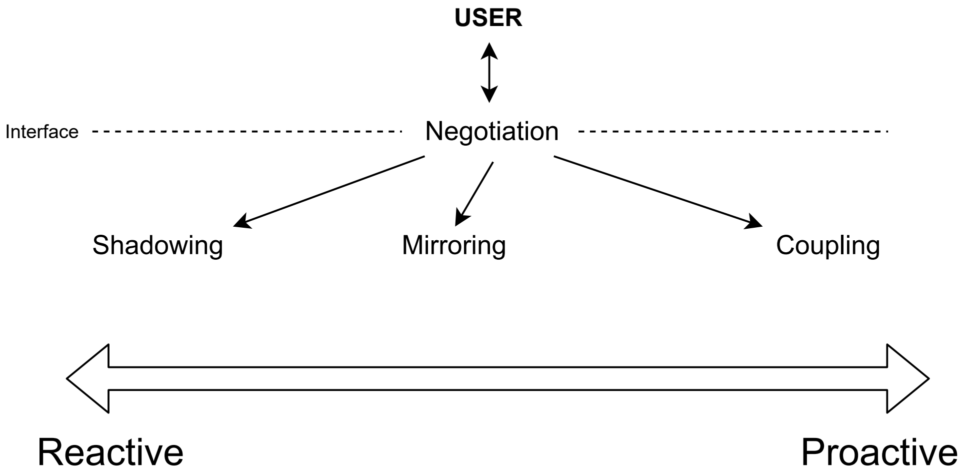


Figure 37. Interactive behaviors mapped onto the system autonomy axis.

In Figure 38, the system autonomy axis is superimposed on the musical similarity axis in a two-dimensional diagram (Thelle & Pasquier, 2021). This is to demonstrate that although the axes are derived from frameworks that belong to different domains, there is a tendency toward parallelity. The loose correlation is illustrated by displaying the interactive behaviors diagonally. Computational interactive behaviors that are more reactive will also tend to generate converging musical results, and vice versa, proactive or autonomous behaviors will tend toward diverging musical output. It should be emphasized that this is an *expected tendency* of the emergent interaction, and by no means meant as a rule. For example, it is possible to imagine scenarios where imitative or unison behavior could be viewed as diverging from a collective sequence where agents have been complementing each other in a counterpointed fashion. In addition to superimposing the interactive behaviors on the convergence vs. divergence framework, the diagram also shows how the principle of mixed-initiative music making may emerge through the shifting of interactive system modes. The reactive mode of shadowing is envisioned as generating interaction where the computational agent will tend to follow the human user, whereas the more proactive coupling mode can be seen as leading to interaction where the user will tend to follow the agent.

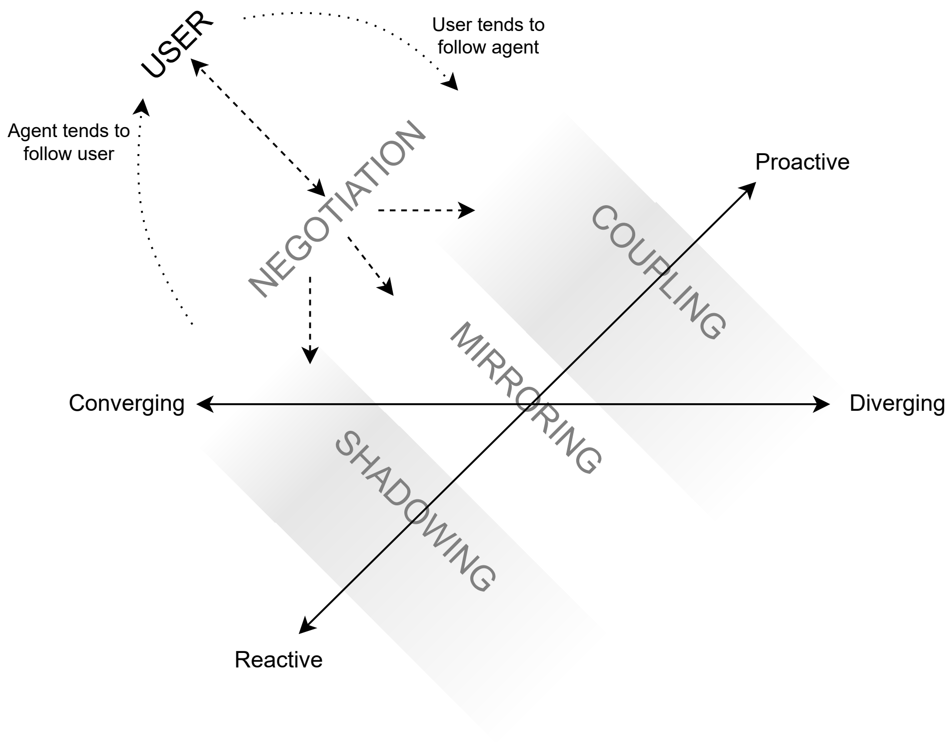


Figure 38. The similarity and autonomy axes combined.

I wanted to use the interface from the Wizard of Oz study as the vantage point for designing an interactive music system proper. The design task thus became to develop a system architecture that would support the emergence of the shadowing, mirroring, and coupling behaviors while the user's interaction with the system—both musically and through the interface—would constitute negotiation. As mentioned in the introduction to this chapter, the experimental design phase that finally led to the prototypes presented in this chapter lasted for 18 months. Of course, this period was rife with trial and error, which I will not focus on. However, I will briefly recount the process I went through of analyzing some existing interactive music systems before developing Spire Muse. The review of existing interactive music systems and different machine learning methods in Chapter 2 provides the necessary background for the following sections of this chapter.

8.1.2 Review of related work

Prior to programming the agent architecture, the following three premises were in place:

1. *Audio inputs and outputs.* The user should be able to interact with the system based on the sound of their own instrument, and the system output should be the unprocessed sound of that input combined with the system's own sounds generated on basis of the interaction.

2. *Interface.* The user should be able to communicate with the computational agent based on the principles of the interface used in the Wizard of Oz study.

3. *Interactive modes.* Based on the updated framework presented in the previous section, the computational agent should be able to make autonomous decisions about which of the three interactive modes (shadowing, mirroring, and coupling) is the most appropriate given the current interaction. The interface should enable the negotiation of these transitions through the feedback buttons that signal approval (*Thumbs up*), force reversions to previous modes (*Go back*), or request change (*Change*).

In a preliminary research phase before programming, I studied several existing interactive music systems to gain a better understanding of what kind of algorithms could work for my purposes. I became particularly interested in the approaches of two research groups. The Music Representations Team at IRCAM have championed a scheme referred to as *Symbolic Interaction*, which combines the principles of real-time contextual listening and corpus-based sequence modeling (Assayag, 2014). Central to the scheme is the notion of stylistic imitation of musical material that can be learned both offline and online with the aid of fast-working algorithms such as the Factor Oracle (FO). The other research group whose output caught my attention was Metacreation Lab for Creative AI at Simon Fraser University. In particular, their research revolving around the notion of Musical Metacreation (MuMe) provided guidance in the conceptualization of a mixed-initiative agent architecture. MuMe is defined as "a subfield of computational creativity that focuses on endowing machines with the ability to achieve creative musical tasks" (Pasquier et al., 2017).

As part of my familiarization of the FO algorithm, I downloaded and experimented with two FO-based systems: OMax and Somax (variably written as SOMax, SoMax, or SOMAX). I found playing with OMax instantly gratifying. It is possible to live record one's own audio input and interact with "stylistic reinjections" of this input, or one can load one or several sound files and interact with recombinations of this material. The recombined material can be processed in a number of ways on multiple channels, so the combined output can quickly grow complex. As the novelty of this form of interaction wears off, however, I recognized some key issues with the platform. While the FO's recombinations are fun to play with, they also quickly become repetitive and relentless unless one actively works with the interface in a conductor-like manner. For variation, FO buffers need to be emptied and refilled, audio streams manually stopped and started, and processing actively managed. In short, OMax is as much an instrument as a player. Judging by videos I have seen of live performances, there is always a person operating OMax on a laptop while instrumentalists are interacting with the system. As such, OMax is not, strictly speaking, an autonomous computational agent. Musical decisions are made by the human operator of OMax. Another issue with OMax is that it does not include any machine listening. It does not "know" what is happening in the audio input.

This latter point is also addressed by Bonnasse-Gahot (2014): "Its choices are only based on internal considerations" (p. 5). The incapacity of OMax to synchronize with an external stream due to this shortcoming is the motivation behind Somax, which builds on OMax but includes machine listening to make it more reactive to the current musical environment (Bonnasse-Gahot, 2014). Hence, Somax not only follows the internal Markovian continuation principle of the FO algorithm, but allows jumps to places in the corpus that the machine listening algorithm deems is relevant based on the input stream. The machine listening extracts harmonic, melodic, and rhythmical features from the input stream. The rhythmical features may be used to synchronize the pulse of the machine output so that it matches that of the user's input. The interface includes a weighting function that can be used to inform the degree to which the agent should be listening to melodic or harmonic features in the user's input stream, or itself (i.e. maintaining internal coherence according to the FO). Furthermore, a *note to note* preset ensures that, when selected, Somax only responds for each onset in the user's input—a welcome possibility for variation. These are only the main features of Somax—there are many more features which are described partly by Bonnasse-Gahot (2014). Since then, the software has been updated as Somax2, which I have not tried.

Combined, the Somax extension the OMax algorithm constitutes a different experience altogether. The variation afforded by the balance between external and internal influences, and the more reactive option of machine output being triggered by instrument onsets makes Somax behave more like a duet partner without the explicit need for another person to control the interface to provide that variation. However, the jumps caused by external influence tend to sound somewhat arbitrary. One reason for this could be that the machine listening context is quite local. Sometimes, such breaks can be pleasantly surprising. Most of the time, however, I was left with the sense that the user input disturbs the internal logic of the FO more than actually changing its behavior. To me, it does not seem like the machine is following the user—it is just skipping to another part of its own "mind" and continuing from there.

In the description of interactive behaviors in Chapter 7, the coupling behavior was described as "largely driven by its own internal generative routines, which are perturbed in various ways by information coming from the performer" (Blackwell et al., 2012, pp. 162–163). This is precisely the kind of interaction I experienced when playing with the FO-generated output of Somax. My conclusion was therefore that FO could potentially be an algorithm supporting the coupling mode of a mixed-initiative system, and that switching to a coupling mode supported by FO would probably give users the sense of the computer taking the initiative or being proactive in the interaction. Furthermore, I found the stylistic reinjection principle of presenting the user with recombinations of their own input to be incompatible with the notion of performing with an "other". This is not a criticism of the principle itself—I found it pleasing and creative in its own right. However, it does not present combinatory surprises in the same way as performing with another agent. As such, the potential use of FO would need to be based on corpora of material performed by other agents than the user. Herein lay another challenge I found with the Symbolic Interaction scheme as implemented in many of the OMax-based systems, including Somax, ImproteK, PyOracle, and others. The FO only loads a buffer (a sequence of symbols) at a time. For all practical purposes, this buffer is usually one file (an audio or a MIDI file) from a potentially large collection of files. The FO is agnostic to anything other than what is loaded into the buffer at any time. In other words, the FO has no global knowledge.

Based on my experiences with OMax and Somax, I surmised that FO could be a potential foundation for the coupling mode in the mixed-initiative interactive system, but that I would need other schemes to drive the shadowing and mirroring behaviors. On top of this, I would need:

- some kind of global organizing principle for a large corpus of musical material, beyond the manual selection of individual sound files as exemplified in the Symbolic Interaction scheme, and

- some kind of "meta agent" that could monitor both the input from the user, make decisions about which interactive mode may be most appropriate, and (if the FO is used) autonomously load buffers from the corpus into the FO based on the current interaction.

It was these two requirements that led me to discover the research at the Metacreation Lab. For example, the notion of a "meta agent" autonomously choosing operational agents or subroutines is established by this research group under the term *curator agent* (Tatar & Pasquier, 2018). Incidentally, this is akin to Beyls' (2018) concept of a Policy Agency recruiting from a "pool of potential agents", introduced in Chapter 2.4.3. Further, the need for an organizing principle for musical objects is duly addressed by Tatar and Pasquier (2017), who devised an agent architecture called MASOM (Musical Agent based on Self-Organising Maps). Rather than being developed as an out-of-the-box interactive music system, MASOM offers a barebones corpus organization and sequence encoding approach that leaves room for other designers to build their own musical agents based on these organizing principles. I realized that, with a bit of reworking, MASOM could be used as the training module for the system I was about to design.

8.1.3 Restructuring and extending the MASOM training module[6]

MASOM was originally designed to be used for electroacoustic and electronic music performance, and has been used in several works featuring improvised noise music, acousmatic music, live electronics together with instrumental performers, and audiovisual installations (Tatar et al., 2018). It also formed the foundation of a gibberish language agent relying on a latent space of syllables collected from the audio of speakers of several languages (Boersen et al., 2020). In dialog with Kıvanç Tatar and Philippe Pasquier, who originally introduced the architecture (Tatar & Pasquier, 2017), I redesigned MASOM's training module to optimize it for instrumental corpora. In this section, I will provide an overview of the training module with a particular focus on new features I added for the system to be compatible with the notion of switching between different interactive modes. For more details about the original MASOM architecture, readers may refer to Tatar and Pasquier (2017).

As the name indicates, MASOM's main organizing feature is a self-organizing map (SOM, cf. Chapter 2.3.2), which provides a global knowledge representation of the corpus' audio content. Additions to the training process also contribute to descriptors for tempo and meso time scale harmonic dynamics, as will be explained below. Implemented in the Max graphical

6 Parts of this and the following sections were presented in the NIME paper "Spire Muse: A Virtual Musical Partner for Creative Brainstorming" (Thelle & Pasquier, 2021).

programming environment, the training module takes a corpus folder containing an optional number of audio files and proceeds to conduct the following steps[7]:

1. Every audio file in the folder is segmented into series of slices. The length of these slices may vary depending on the detection of loudness onsets, signifying new sonic events. A duration range indicating a minimum and a maximum length for each slice can be set before segmentation. For my experiments detailed in the rest of the chapter, I have used a minimum length of 200 milliseconds and a maximum length of 3 seconds. The segmentation process is performed with MuBu (Schnell et al., 2009). This process does not create new audio files—rather; the process creates a list of timecodes, indicating the beginning of each slice within each audio file. The audio files are left as is in the corpus folder, on which the timecodes can function as lookup coordinates.

2. Each audio slice is labeled with a feature vector (cf. Chapter 2.3.1). In all, there are 55 features within each feature vector. The first is *duration*, which is calculated by subtracting the onset time of the current slice from the onset time for the next slice (or from the length of the entire audio file in the special case of the last slice in each file). The remaining features are extracted using objects from the PiPo toolkit (Schnell et al., 2017) which is integrated with MuBu. The features are calculated by taking the mean and standard deviation of feature values for all FFT frames within each slice. I found that relatively large FFT frame and hop sizes (8192/512 samples) yielded more reliable melodic and harmonic data. The features are (mean and standard deviation): *loudness* (2), *mel frequency cepstrum coefficients* (MFCC) (26), *fundamental frequency* (2), and *chroma* (24). The chroma features were not extracted in the original MASOM—I have added them to strengthen the musical agent's capability to orient itself harmonically as well as melodically. The inclusion of chroma features serves two functions. Firstly, it reinforces the melodic classification of slices containing one note. Equally important, it minimizes pitch errors introduced in slices that happen to contain several notes. The average pitch of two or more notes yields a single pitch that is musically out of context. However, the chroma features are discrete and can reveal the presence of several notes within one slice. Hence, there is a better chance for slices with similar harmonic content to be clustered together in the self-organizing map, even in cases where the derived pitch misrepresents the tonality.

3. The next step is a new inclusion altogether, namely the extraction of chroma transition matrices. As described in detail in Chapter 6.1.3, the pre-analyzed corpus transition

7 For most of the experimentation and the subsequent user study, I have used a corpus containing 180 acoustic guitar solos in different keys, in the styles of jazz, blues, rock, funk, bossa nova, and singer-songwriter (Xi et al., 2018).

matrices have index numbers that are pointers to the file and timecode of the 20-slice windows from where they are derived. Experiments showed that when playing back one of the original corpus files through the input, machine listening and analysis modules would correctly display index numbers corresponding to the audio locations being played back by comparing sequences of chroma vectors in the input with the transition matrices. The algorithm behind this will be presented in the next section. Moreover, playing back other audio input not present in the corpus would produce pointers to audio in the corpus that sounds quite similar and complementary. This of course depends on the size of the corpus. If the corpus is sparse, the closest matching sequence may not be as similar as a closest match in a large and varied corpus.

4. The next step is also a new inclusion. The tempo (BPM) for each audio file in the corpus is derived from a Python script via OSC. The tempo data helps make the generative FO playback (introduced in the next section) more aligned with the audio files' original tempos. For corpus material that is not tempo-based, the script will still attribute a perceived tempo. Although redundant, forcing a grid on atemporal material does not seem to have a negative impact—only minor time adjustments are made. These small adjustments are extremely important for tempo-based material, but not significant for other material. Therefore, the grid is used for all material, and there is no need to create a dichotomy in the training process.

5. This step is the training of the SOM. This is done using the ml.som object from the ml.* machine learning toolkit which is available in the Max platform (Smith & Garnett, 2012). The labels (feature vectors) for all the slices in the corpus are treated as the training dataset. The resulting SOM is an encoded representation of all events in the corpus—a topographical abstraction of "what" divorced from "when". On average, the number of nodes created in the SOM is approximately one-sixth the size of the number of audio slices.

6. After the SOM has been created, each audio slice is assigned to a SOM node based on a best matching unit function (cf. BMU, Chapter 2.3.2). Hence, similar slices are clustered together at these nodes. As pointed out in Chapter 2.3.2, input vectors are unevenly distributed across the SOM, and some nodes may be empty.

7. The final part of the training is a procedure where each audio file in the corpus is encoded as a sequence of SOM nodes—each audio slice is represented by its SOM index. This is a form of lossy encoding, because many different audio slices may be represented by one SOM node.

Steps 1, 5, 6, and 7 are original MASOM processes. Steps 3 and 4 are my contributions. Step 2 is a modified version of the original labeling process. In all, the training module does much more than train a SOM. It provides a multifaceted analysis of the given corpus—one which yields computational views of the corpus on local and global levels, and which adds descriptors for tempo and meso time scale harmonic dynamics.

As part of the training, a data folder is produced containing the following lookup files:

- A list of feature vectors as labels for all the audio slices in the corpus
- A list of chroma transition matrices corresponding to the audio file and temporal window (20 slices, cf. Chapter 6.1.3) from which the matrices were derived
- A list of BPMs for each audio file
- A SOM where all audio slices in the corpus are mapped to the SOM nodes
- A list of the audio files encoded as SOM node indices

In the next section, I will finally present Spire Muse—the mixed-initiative interactive music system I developed on the basis of the first three studies. I will also present how Spire Muse utilizes the various knowledge representations derived from the extended MASOM training process.

8.2 The Spire Muse architecture and interface

The key feature I had planned for Spire Muse was the autonomous selection of the interactive modes shadowing, mirroring, and coupling based on the current interaction. Variance between these modes could give the impression of the system sometimes following the user, and sometimes taking the initiative. To identify what the user is playing, the system would require a machine listening and feature extraction unit to gather data from the input in real time. This data would in turn be relayed to the decision-making unit making operational choices. For analyzing the input on the sound object time scale (Roads, 2001), I relied on a real-time version of the segmentation and feature extraction process used in the training. In other words, the input data is sliced and formatted in the same way as the training data, making it straightforward to compare input data with corpus material. The input data units are readable both in the audio slice format (for comparisons with the original audio slices) and in SOM node format (for comparisons with the SOM nodes). The reasons for this will be explained below.

To add flexibility and variation, I integrated the possibility for the user to set *influence parameters* (see Figure 39), where the listening module can be directed to give some groups of features more weight than others when comparing input and corpus units. Experimentation showed that this alters the subsequent matching algorithm considerably. The four influence parameters are *rhythmic*, *spectral*, *melodic*, and *harmonic*. The rhythmic parameter weights the duration feature. Setting the rhythmic parameter high and the rest low will make the matching algorithm search for material in the corpus that follows the timing of the input closely, but disregards the other features. The spectral parameter weights the MFCC features. The melodic parameter focuses on the fundamental frequency, and the harmonic parameter weights the chroma features. The influences can be set with sliders, so any combination of relative influence is possible.

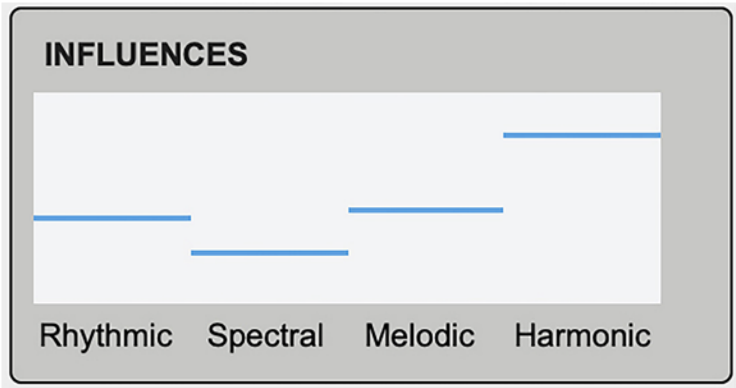


Figure 39. The influence parameters.

To extract input features on the meso level time scale (Roads, 2001), the listening module keeps a rolling buffer of the last 20 onsets from the input and extracts chroma transition matrices using a real-time version of the algorithm described in Chapter 6.1.3. As mentioned earlier in this chapter, experiments demonstrated that this produces pointers to time windows in the corpus containing sequences tending toward similar harmonic dynamics as in the input—at least as similar as the corpus allows. These pointers are produced by comparing the currently detected chroma transition matrix derived from the input with the whole list of matrices in the corpus. The corpus matrix with the lowest Euclidian distance measure between the feature vectors is taken as the closest match, and its index number becomes the output of the matching algorithm. Extracting and comparing chroma transition matrices only focuses on one feature group, so the influence parameters have no bearing on this part of the machine listening process.

Before proceeding to explain how these input features contribute to the agent's decision-making, I will present how the interactive modes have been implemented in Spire Muse. As described earlier, the shadowing behavior is described as "a close following of the user's input". In order to achieve this, I devised a *concatenative synthesis* method based on a unit selection technique using the matching algorithm outlined above. Continuously playing back audio slices from the corpus that are the closest matches to the audio input ensures a shadowing-like behavior in the system's output. Concatenative synthesis is an extensively researched topic in computer music (e.g. Schwarz, 2000; Schwarz et al., 2006). In Spire Muse, the unit selection is done using the zsa.dist object from Zsa.Descriptors (Malt & Jourdan, 2008). Additionally, the influence parameters come into play—closest matches vary depending on how these are set.

SOM nodes are not looked up in shadowing mode. Instead, instances from the input are compared directly to the feature vectors that label the audio slices in the corpus. Looking up audio slices directly creates a better contrast to the mirroring mode, which, as described below, does look up SOM nodes. Direct slice matching makes sense when attempting to create an impression of an agent that follows the user as closely as possible. I found on the other hand that outliers in the SOM nodes weaken this effect to a certain degree.

Sparsity in some areas of the feature space may yield discrepancies between the input and respective slice matches. Rather than being unwelcome artifacts, these discrepancies tend to make sense musically. The harmonic influence is useful here because harmonically related events have similar chroma profiles. However, a high-quality input signal is important. Direct input signals work much better than microphone signals, which capture room acoustics as well as the instrument itself. The signal and data flow for shadowing mode is depicted in Figure 40.

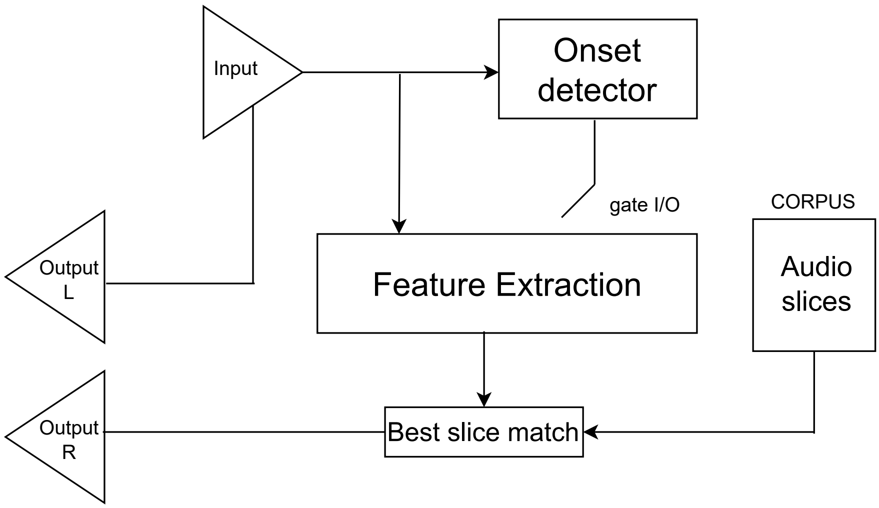


Figure 8.5. Input vs. corpus matching in shadowing mode.

Video 1 demonstrates an example of how Spire Muse responds in shadowing mode. The interface has been redesigned since this video was recorded, but the underlying algorithm is the same.

Video 1. A free improvisation session in shadowing mode. [2.35].

In mirroring mode, I designed the musical agent to engage in reflexive interaction. Unlike the shadowing mode, the agent does not respond to input immediately, but listens to longer phrases and attempts to respond with similar phrases. Upon receiving input, the agent starts building a list of closest SOM matches based on audio slices from the input stream. Accumulated SOM lists are expedited after eight beats, adjusted according to a tempo detection object listening to the input. This SOM list is subsequently loaded into a FO (more details about the FO object below). The playback of the FO lasts for as many nodes as the length of the list that loaded it. For eight beats after the FO is initiated, SOM list gathering is inactive, corresponding roughly to the length of the agent's response. This creates a sense of back and forth between the user and the agent. This process iterates for as long as the mirroring mode is active.

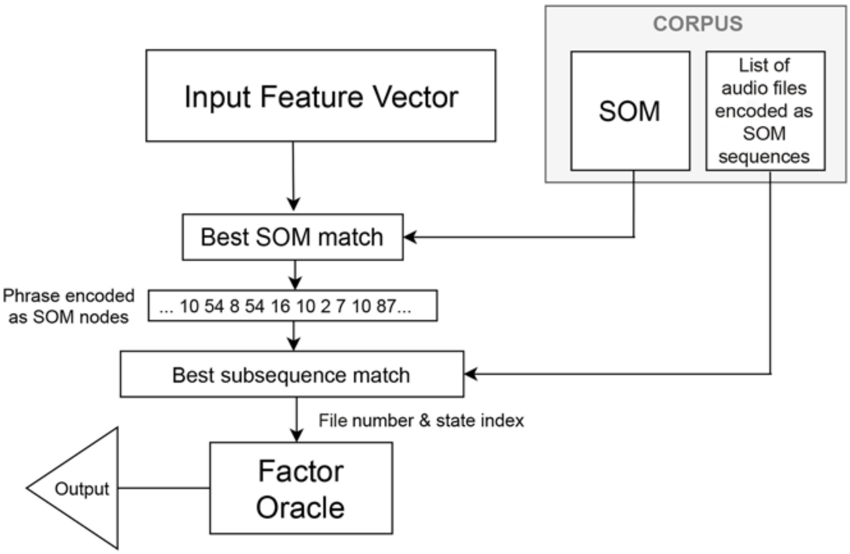


Figure 41. Input vs. corpus matching in mirroring mode.

Video 2 is a demonstration of mirroring mode. As with Video 1, the interface has gone through cosmetic changes since the video recording.

Video 2. Improvising in mirroring mode. [3.55].

As mentioned earlier in this chapter, I considered FO to be a good candidate for sound generation in coupling mode. The MASOM repository includes a runtime module featuring the Max external factorOracle (A. J. Wilson, 2016), which can load SOM sequences from the training process into its buffer and play back recombinations according to the FO algorithm. I decided to use the factorOracle object, but not in the same way as proposed in the runtime example patch. A SOM node may potentially have several audio slices associated with it, and all slices do not necessarily belong to the audio file that is represented by the given encoded sequence. One SOM node may figure in the SOM sequences representing multiple audio files in the corpus. Hence, only one slice is actually the original segment in the sequence represented by the SOM indices. The sound generation process in the example patch would select an audio slice from the current SOM node at random, resulting in a sequence model that frequently would play back contextually inappropriate substitutes of the audio slices from the original sequence. While this could probably work quite well for some types of corpus material, I found it not to work very well with the instrumental corpora with which I experimented. Therefore, I implemented a selection algorithm that ensured that only audio slices from the original audio sequences would be played back. There would still be a recombination, as many SOM nodes generally tend to be featured several times in the same sequence even though they represent different audio slices. In other words, the FO principle was preserved, but a more radical-sounding mix of source material was avoided.

When in coupling mode, the user is "coupled" to a FO, which is played back continuously. Left unperturbed, the FO iteratively queries its next state, thereby taking on an autonomous style that may coerce the user to follow the musical agent's lead—similar to the self-listening mode of Somax. However, the agent also listens to the user and attempts to align with the input by intermittently loading new FOs from other parts of the corpus or by jumping to new states within the same FO. This is reminiscent of the jumping caused by outside influence in Somax, the difference being that the agent "sees" the entire corpus and not only other locations within the loaded FO (in Somax all FOs are loaded manually). As explained earlier, the input buffer for this part of the machine listening is 20 input slices—corresponding to the window length of the chroma transition matrices that were built during training. Hence, Spire Muse's "harmonic attention span" is relatively long—between 15 and 30 seconds depending on the length of the slices currently being analyzed.

The SOM sequence that is automatically loaded from the corpus into the FO is selected based on a combination of two criteria:

- *Meso time scale harmonic dynamics:* A chroma transition matrix of the past 20 input onsets is compared with corresponding matrices built from the corpus. Corpus audio files associated with the top ten matches are contenders for affecting an FO change.
- *Tempo similarity:* A list of tempos associated with the corpus audio files that are within plus/minus 10 BPM of the currently detected tempo is gathered.

If one or more same corpus audio files feature in both these groups, the FO will load the highest scoring match (using a cosine distance measure in the zsa.dist object) and initiate the change. After a change, the input buffer will start building anew, so changes will be no more frequent than at least the time it takes to fill the buffer.

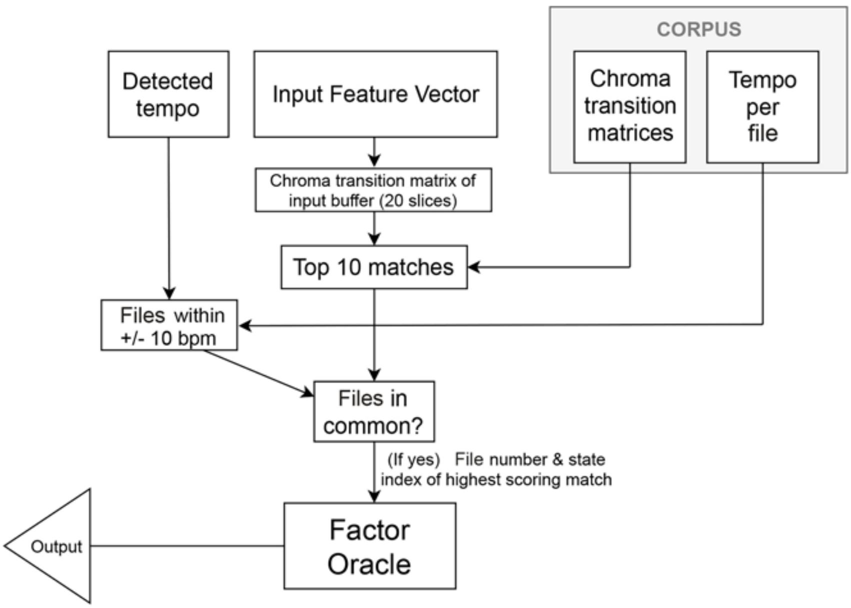


Figure 42. Input vs. corpus matching in coupling mode.

In Video 3, a demonstration of coupling mode can be seen.

Video 3. Interaction in coupling mode (automated FO changes disabled). [3.26]

Now that the processes that underlie the interactive modes shadowing, mirroring, and coupling are known, it is possible to explain how the autonomous decision-making takes place. I would like to emphasize that the solution in this first version of Spire Muse is still at an experimental stage. Shadowing is implemented as the baseline behavior of the system. It is the initial default mode at the start of a session, and also the fallback mode if the mirroring and coupling modes do not meet the qualifications for activation. The mirroring and coupling modes, meanwhile, are designed as competing for activation. The coupling mode is activated each time the two criteria for FO change mentioned above (tempo within plus/minus 10 BPM among the 10 closest matching chroma transition matrices) are met. However, the mirroring mode trumps the coupling mode if the eight most recent onsets have many close matches in the corpus. Informally explained, this means that a phrase level equivalent to the input is present in the corpus, triggering a "memory burst". Technically speaking, the zsa.dist object continuously outputs the Kullback-Liebler distance between 1) the SOM feature vector extracted from the input and 2) the best-matching SOM feature vector in the corpus. If the mean of the eight most recent distance measures is lower than 20th percentile of the equivalent distance measures of the 20 most recent SOM nodes (the meso level time scale window), this is interpreted as the local memory structure taking precedence over the longer-term memory structures. I arrived at this solution after several months of experimentation, and it seems to result in interaction with an acceptable distribution between the three different modes. In short, the formula can be described as follows:

- If no short-term (8 onsets) or long-term (20 onsets) memory structures in the corpus are considered close enough, then shadowing mode.
- If the criteria for loading a new FO are met, signifying a co-occurrence of tempo and harmonic similarity between corpus subsequences and the last 20 inputs onsets, then coupling mode.

- If the latest 8 onsets have close-matching equivalent instances in the corpus, and the mean of the distance scores is lower than the 20th percentile of the distance scores of the 20 last onsets, then mirroring mode.

When the mirroring or coupling modes are activated, they are sustained for at least 10 seconds before automated mode shifts are allowed to occur. This is to avoid a jittery interaction style where mode shifts occur too frequently. Video 4 demonstrates a session where automation is activated. Here, the interface has been updated to its current design.

Video 4. Demo session featuring automation and use of the negotiation panel. [5.17]

As mentioned in Section 8.1.1, negotiation is imagined as an interface level behavior that influences the choices of the other behaviors. Hence, the negotiating interface functions as a counterweight to the agent's automated behaviors. Figure 43 shows the interface of Spire Muse. As with the interface used in the Wizard of Oz study, it features the buttons *Go back*, *Pause/Continue*, *Change*, and *Thumbs up*. *Go back* forces the agent to its previous mode. This backtracking can be repeated. The agent tracks its own history, which also includes FO song (file) changes. *Pause* will mute the agent, but it is still listening. This is useful if the user needs time to figure out something in his or her playing without interruption. Upon pressing *Continue*, the session will proceed based on the most recent listening. *Change* will force the agent away from its current state. For now, this resets the interactive mode, influences, and FO song selection randomly. The *Thumbs up* button signals to the agent that the user is enjoying the current interaction, and stays in the same state for the next 20 seconds. Apart from this, the influence parameters are also a new feature compared with the interface from Study 3, along with a panel showing which mode is currently activated and a panel for dropping the trained agent folder. The interface has kept one significant feature from MASOM, namely the visualization of the SOM. A small dot shows the corresponding SOM node for each audio slice played back.

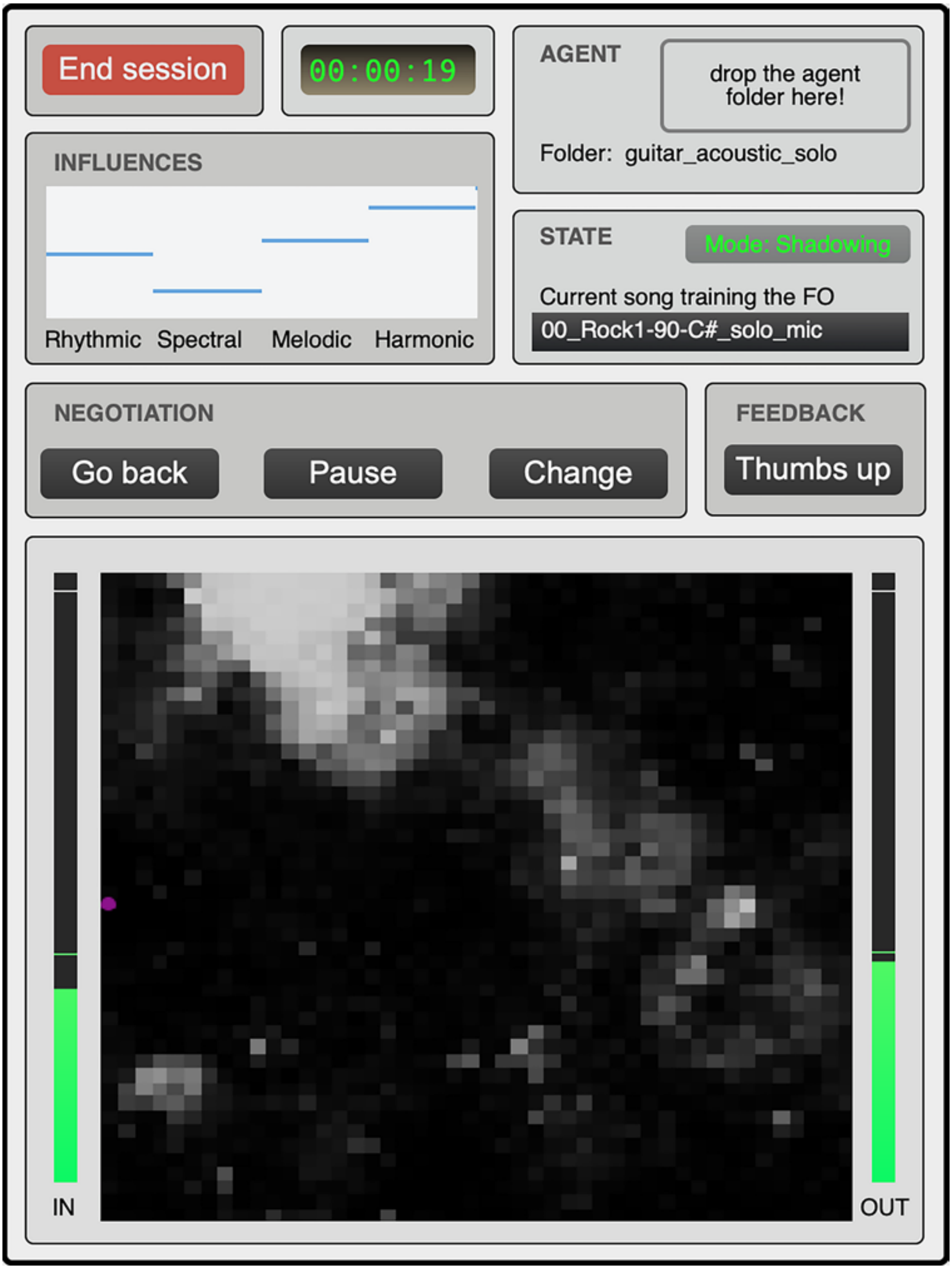


Figure 43. The Spire Muse interface.

The Spire Muse system and the interface was under continuous development up until the user study in October 2021. Hence, some of the descriptions of the system differ from the paper presented at the NIME (New Interfaces for Musical Expression) conference earlier that year (Thelle & Pasquier, 2021). The rest of this chapter is dedicated to presenting the method, results and discussion of the fourth and final study in this thesis—the evaluation of Spire Muse.

8.3 Method

In order to examine the tradeoff between system autonomy and user control outlined in the introduction to this chapter, two different prototypes of Spire Muse were created for a comparative user study. The prototype presented in the previous section will from here on be called the *Auto Modes* prototype. This prototype switches modes autonomously based on what it "hears" in the input, but can be indirectly influenced to change its behavior. In the *Manual Modes* prototype, on the other hand, the user may select interactive modes manually and thus be more directly in charge of the direction of the interaction. I hypothesized that engaging with the Auto Modes prototype, despite being less predictable and harder to control, may lead to a more creative experience for the user than engaging with the Manual Modes prototype.

In this section, I will first discuss challenges connected with evaluating co-creativity, acknowledge these challenges, but ultimately argue that it is possible to sidestep these controversies in this particular user study by focusing on user experience. I subsequently present a mixed methods study design involving the collection of data from surveys, interviews, interaction logs, and self-evaluation by the users, followed by both statistical and qualitative analyses. Throughout all the sessions in the user study, I used a corpus containing 180 acoustic guitar solos in different keys, in the styles of jazz, blues, rock, funk, bossa nova, and singer-songwriter (Xi et al., 2018). The reason for using the same acoustic guitar solo corpus as I had done in most of the testing was twofold. Firstly, it was a corpus I had become familiar with, and which had turned out to work with several different kinds of input. Secondly, I needed a corpus with a versatile type of instrument that could work in different genres and together with different instruments without appearing out of place.

8.3.1 Evaluating co-creativity as an experience

As discussed in Chapter 2.5.2, a mixed-initiative system should be considered a type of co-creative system. A case study conducted by Jordanous (2017) has demonstrated that "people are significantly less confident at evaluating the creativity of a whole co-creative system involving computational and human participants, compared to the (already tricky) task of evaluating individual creative agents in isolation" (p. 159). This difficulty could be attributed to what Bown (2015) refers to as "the 'islands of creativity' view", which refers to the misconception that creativity is something that occurs within individual actors (human or computational). This conflicts with the holistic view of creativity as emergent from the interaction between several agents, as presented in Chapter 2.

In the context of this thesis, however, the discourse regarding the attribution of creativity is a sidetrack. The main research question of this thesis is how a mixed-initiative interactive music system can *aid human musicians* in the initial ideation stage of music making. This chapter purports to examine the creative experience of the (human) users of the Spire Muse prototypes. As such, there is no need to evaluate creativity from an objective standpoint, insofar as that is at all possible. It is clear from the research questions that I am interested in learning about user experience (UX). This is an important distinction to make. Although Spire Muse falls into the category of co-creative systems, the focus on UX means that in terms of evaluation, methods from the field of creativity support tools (CST) are more relevant than methods devised to evaluate computational creativity or co-creativity. This is the case despite my contention that the computer is more like a partner than a tool in mixed-initiative creative interfaces. I am not evaluating the creativity of anyone or anything—I am interested in how creative the human user *feels* when engaging with the co-creative system. This is what Hassenzahl et al. (2000) refer to as the hedonic quality of the system (cf. Chapter 2.2.2). Furthermore, as emphasized in Chapter 4, it is the user's experience of the *activity* of mixed-initiative music making that is the focus of this study. Swift (2013) maintains that in UX parlance, jamming falls into the category of "an experience"—with a well-defined beginning and end (p. 59).

According to a meta-analysis of NIME (New Interfaces for Musical Expression) proceedings from 2012 to 2014, system evaluations from the user's perspective to understand UX are usually performed using qualitative methods (Barbosa et al., 2015). This can lead to issues related to validity, especially in cases where researchers use self-developed questionnaires without providing items or statistical validations (Bargas-Avila & Hornbæk, 2011). Hence, there are clear benefits associated with adopting standardized evaluation methods. On the other hand, questionnaires often miss out on useful details that users may provide in qualitative interviews. Therefore, this user study will both adopt a standardized survey for a statistical approach and a qualitative approach based on semi-structured interviews, in addition to interaction logs and self-evaluations by the users.

8.3.2 The prototypes

As mentioned, the Auto Modes prototype is the same system as presented in Section 8.1.4, with only one slight modification. Comparing the screenshots in Figure 44 with the screenshot in Figure 43 will reveal that the influences labels are selectable in the latter, and that the currently selected label is highlighted. This is because the influences are redesigned as discreet presets instead of continuous parameters. I wanted to reduce the amount of options for the users to experiment with the influences, as this was not the main focus of the study. Therefore, the user could only choose between four influences presets, which would skew the machine listening toward the rhythmic, spectral, melodic, and harmonic features respectively. Apart from this, the interface featured to the left in Figure 44 is identical to the one in Figure 43.

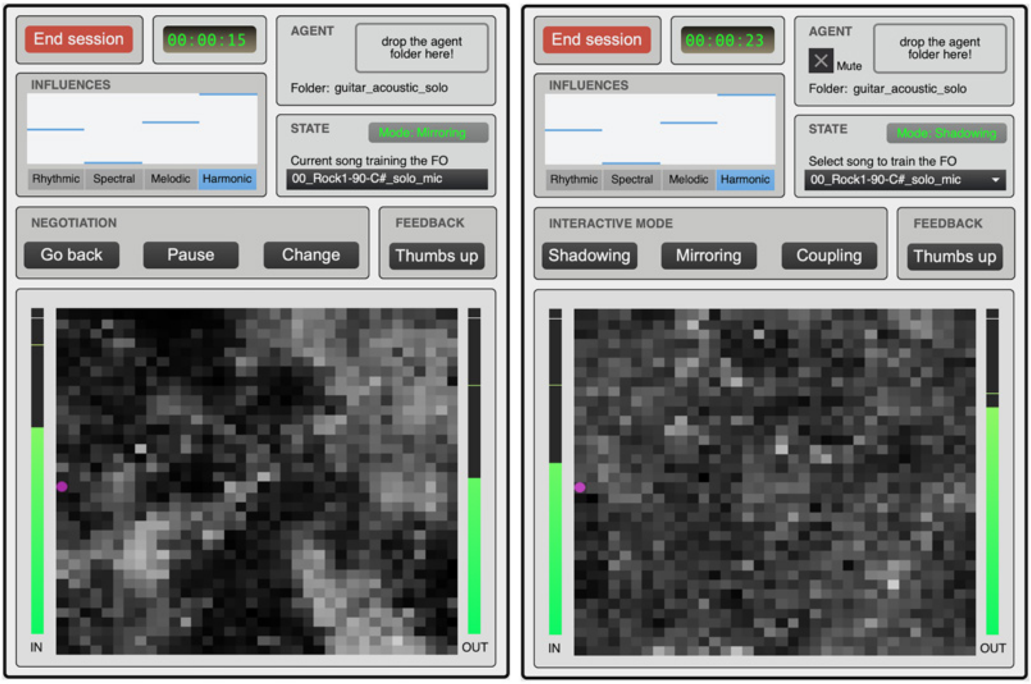


Figure 44. The Auto Modes (left) and Manual Modes (right) prototypes.

The most important difference between Auto Modes and the Manual Modes prototypes is the panel in the middle. Instead of a Negotiation panel, the Manual Modes interface has an Interactive Modes panel where the user can manually select which interactive mode they would like the system to engage in. As seen in Figure 44, the *Go back*, *Pause/Continue*, and

Change buttons in Auto Modes are exchanged with *Shadowing*, *Mirroring*, and *Coupling* in Manual Modes. Under the hood, the automation between modes is deactivated in Manual Modes. Another difference is that when in coupling mode, the FO buffer is no longer loaded automatically based on what the machine listening hears in the input. Instead, the user must select from a dropdown list which file from the corpus should be used to train the FO. The small white downward-pointing triangle in the State panel's FO display in Manual Modes indicates the dropdown menu option. In Auto Modes, this display does not react to user clicks, and only displays the system's automated choices. As such, the coupling mode in the Manual Modes prototype functions according to the OMax principle—it is only engaged in following its own internal routine and pays no attention to the input. It is up to the user to select a file that most closely matches the kind of material he or she wants to hear in the system's generated output. In other words, the principle of mixed-initiative interaction is abandoned in the Manual Modes prototype.

To the users participating in the study, the *Thumbs up* button was introduced as follows: "*Thumbs up* alerts the agent that it is doing something that the user finds particularly engaging. This will be bookmarked for future reference." This wording was chosen to disguise the fact that the button has an additional function in Auto Modes—as explained in Section 8.1.4, it makes the agent stay longer in the current state. However, in the user study, the most important function for the *Thumbs up* button was its contribution to the interaction log, as the button could be used in the analysis as an indicator for which parts of the sessions the users particularly enjoyed. If the users had known that the button also actively influences the interaction in Auto Modes, and only provides feedback to the researcher in Manual Modes, the effect could be that they would have used the button differently with the two prototypes. Therefore, the influencing function in Auto Modes was not explained to the users, and it is safe to assume that they engaged with this button on equal terms when using both prototypes.

A final difference between the two prototypes is that Manual Modes features a mute/unmute toggle in the Agent panel to make up for the lack of the *Pause/Continue* button that features in the Negotiation panel in Auto Modes. Despite their differences in placement and appearance, they have the exact same function. The main concern was to keep the interfaces as visually similar as possible while giving prominence to the differences that matter, namely the negotiation/system automation abstraction in Auto Modes and the interactive modes/user direction abstraction in Manual Modes. A discreetly placed button made it possible to keep the possibility to silence the system's response while keeping the same number of buttons in the Negotiation/Interactive Modes panels.

In order to minimize the need for users to interrupt their own playing while engaging with the interfaces, I designed a footswitch system for both prototypes. Four footswitches were mapped to the four most prominent buttons in the interfaces as seen in Figure 45.

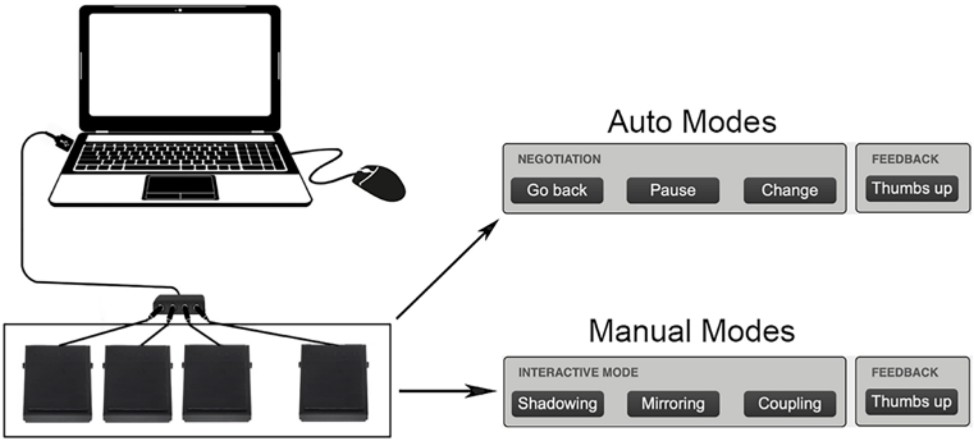


Figure 45. Four footswitches mapped to the four most prominent buttons in both prototypes.

To ensure that users would not be confused by the change of functionalities when interacting with the different prototypes, I made laminated labels that could easily be swapped between sessions, as seen in Figure 46.

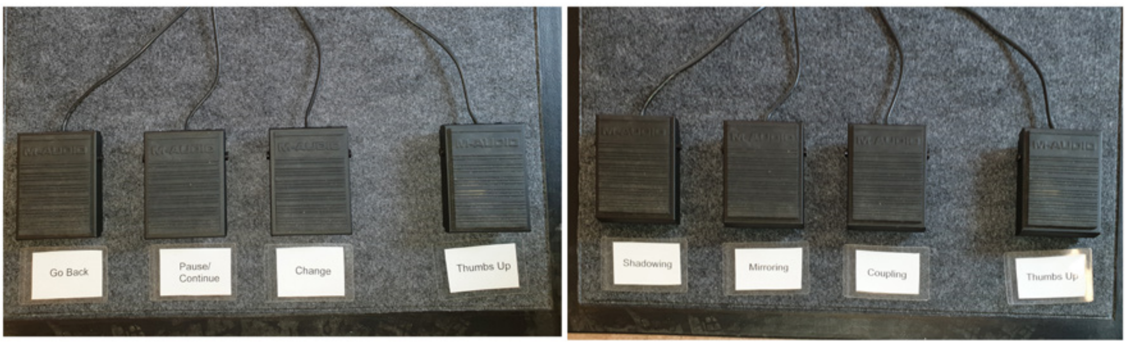


Figure 46. Swappable laminated labels for the pedals.

8.3.3 Participants

A statistical analysis component in the study required a minimum of eight participants for there to be a realistic possibility of attaining a result with statistical significance in a paired t-test comparing the two prototypes. The target group was defined as "musicians who use collaborative experimentation as a method to develop musical ideas (improvising/jamming)". Participants from any musical genre were welcome. The invitation (see Appendix D.1) further specified the following:

However, due to the way in which the system "listens" it is crucial that the instrument is tonal. This means that percussionists/drummers are not in the target group for this study. It is also an advantage if the instrument can play outside of the bass register, because the machine listening is better in the mid and higher registers.

As I did not want the invitation to sound too technical, this formulation was chosen to avoid being too specific about the chroma transition matrices and the fact that instrument microphones tend to pick up more room acoustics with bass instruments, potentially making the input signal more difficult for the machine listening component to analyze. The latter was based upon observations when trying different instruments with Spire Muse.

In the invitation, the length of the user study was estimated to be around 90 minutes. Additionally, participants would receive a home assignment where they could listen through recordings of their creative sessions with the prototypes and evaluate the interaction in real-time using a web application. As a token of gratitude for would-be participants, the invitation promised a gift card worth 300 Norwegian kroner to all participants who completed the user study.

Invitations were distributed through multiple channels via intermediaries at my own institution, other colleges and universities, and music interest organizations. Fortunately, I managed to get the eight participants I needed. The gender balance was far from ideal—seven males and one female. Although unfortunate, it is not surprising considering that males still far outnumber females in music technology education, and most of the participants were music technology students.

Rather than give the participants pseudonyms as in the previous studies, I have decided to call them P1–P8 (participants 1–8). Table 8 gives an overview of the participants, their instrument, and their current occupation.

P1	Laptop/synth	Electronic music performer/composer Recently graduated music technology Master's student
P2	Electric guitar	Pop/media performer/composer (film, TV, etc.) Music technology Master's student
P3	Saxophone	Jazz music performer/composer Music technology Master's student
P4	Electric guitar	Electronic/experimental/avant-garde performer/composer
P5	Electric guitar	Rock/metal/electronic music performer/composer Session musician
P6	Piano	Pop/media performer/composer Assistant professor/lecturer in music pedagogy
P7	Vocals	Singer-songwriter performer/composer Music neurology researcher
P8	Electric guitar	Blues/rock/jazz performer/composer Music technology Master's alumni Acoustician

Table 8. The eight participants in the user study.

A few days prior to the user study, each participant received an email with additional practical information regarding the study, including a plan for the 90-minute study as follows:

1. Introduction (5 minutes)

2. Prototype 1 (35 minutes)
    (a.) Introduction and learning session (15 minutes)
    (b) Creative session (10 minutes)
    (c) Questionnaire and interview (10 minutes)

3. Prototype 2 (35 minutes)
    (a.) Introduction and learning session (15 minutes)
    (b) Creative session (10 minutes)
    (c) Questionnaire and interview (10 minutes)

4. Comparisons, final interview (15 minutes)

They also received some information about the COVID-19 precautions I had put in place, as strict measures were still upheld at my institution. All devices would be cleaned with anti-bacterial wipes before each study, and I would keep a distance of at least one meter at all times. I would also rapid test myself in the morning for each day that user studies were conducted to ensure that I was negative.

8.3.4 Structure of the study

As mentioned, the user study was estimated to last 90 minutes for each participant, but the average completion time turned out to be approximately 75 minutes. The overall structure of the study was shown in the previous section. In the following, I will provide a more detailed overview into the structure of the study, and present how the participants were informed of what to do before each part of the study.

1. Introduction (5 minutes)
   After a brief word of welcome, I handed out a one-page information sheet (see Appendix D.4) about the purpose of the study (including a brief introduction of the Auto Modes and Manual Modes prototypes), the study process, which parts would be recorded and filmed, and about the participant's right to pull out of the study at any point with no negative consequences. I read this information aloud while the participant could read along. Subsequently, they were handed a consent form they would sign before commencing with the study.

2. Prototype Auto Modes or Manual Modes (35 minutes)
   The order of the prototypes Auto Modes or Manual Modes was randomized for different participants so as to minimize the influence of the order. As Lazar et al. (2010) point out, a learning effect could make participants perform better toward the end of a study, while an opposite impact could be that fatigue could make them less focused. Randomized order of prototypes is a common way to minimize the impact of such effects in comparative studies. Hence, the first prototype test started with either one of the two.
   (a.) Prototype introduction (5 minutes)
   The basic concept for both prototypes was explained as being a virtual partner for creative brainstorming. I explained the principle of corpus-based concatenative synthesis (using audio fragments from the corpus and putting them back together as remodeled sequences), the interactive modes (shadowing, mirroring, and coupling), and the influence presets (rhythmic, spectral, melodic, and harmonic). I then explained the functionality of the prototype, which because of the randomized order could be either Auto Modes or Manual Modes.

   (b) Learning session (5 minutes)
   The participant got five minutes to play with the system and try out its different functions. I told them that they would be given a chance to ask questions that could arise in a conversation after the learning session. While the participant was

engaged in the learning session, I would sit quietly at my desk across the room without appearing to pay attention. I would normally be preparing the iPad for the upcoming questionnaire at this point.

(c) Short conversation (5 minutes)
This conversation was to ensure that the participant had understood the concept and design of the interface. I went through each function and cleared up any misconceptions. I then explained the purpose of the upcoming creative session. I read aloud the following:

Think of it as a jam session. You and a fellow musician are improvising loosely around an idea. Start with a short musical phrase. It could be an idea you have for a song. A motive, theme or riff. Gauge the musical agent's response and let it develop from there. You do not need to stick to the theme if you find your interest wandering to something more interesting. Keep in mind that you do not have to show any "musical result" at the end of the session. We are interested in your experience of the creative session.

The participant was told that the creative session is the actual study, and hence would be audio and video recorded. I explained that after starting the session, I would leave the room for 10 minutes and not be listening in.

(d) Creative session (10 minutes)
Before leaving the room, I would make sure that the video and audio were recording, and that the correct Max session (Auto Modes or Manual Modes) was loaded. There were two video cameras. One camera recorded the screen, as I did not want to spend CPU power on screen recording software. The other camera captured the participant and the computer screen in one frame, in case I needed it to synchronize audio and interaction events during analysis (this turned out to be important). After leaving the room, I would stay outside for nine and a half minutes before quietly entering the room. The participants would generally notice me, but keep playing until the interface timer showed 10 minutes, then stop.

(e) Questionnaire (3 minutes)
After ending the session, I handed the participant an iPad with a questionnaire containing 10 questions, with two pages containing five questions each. The survey was based on the Creativity Support Index (CSI), which measures several

dimensions of creativity support (Cherry & Latulipe, 2014). The specifics of this survey will be explained in Section 8.2.5.

(f) Semi-structured interview (7 minutes)
Before starting the interview, I informed the participant that in the following conversation would be recorded. Although defined as a semi-structured interview, I ended up following the structure of the interview questions without exception. These are detailed in Section 8.2.6.

3. Prototype Manual Modes or Auto Modes (35 minutes)
The procedure of this part of the study was identical to 2 (a–f), but with the other prototype. As mentioned, this could be either Auto Modes or Manual Modes depending on the randomized order. In general, following the procedure a second time would normally take a few minutes less, because the participant would be quicker to understand the design concepts in the introduction and learning session.

4. Comparisons (15 minutes)
   (a.) Questionnaire (5 minutes)
   The final questionnaire was part 3 of the CSI survey, which is called the Paired-Factor Comparison Test. As with the previous questionnaires in parts 2e and 3e of the study, this questionnaire was also administered on an iPad. The Paired-Factor test is explained fully in Section 8.2.5.

   (b) Semi-structured interview (5 minutes)
   In the final interview, I asked the participant questions about how the experience of playing with the two prototypes differed from each other. The participant could also share concluding thoughts or reflections. The specific questions I asked will be listed in Section 8.2.6.

   (c) Home assignment explanation (5 minutes)
   Before ending the user study, I would remind the participant that I would be sending a home assignment within a week. I would show them an example of a web application especially designed for the purpose of real-time evaluation of audio. The specifics of the home assignment and web application are detailed in Section 8.2.7.

A more detailed guide for Study 4 is available in Appendix D.3.

8.3.5 Statistical survey

The goal of the statistical survey was to test the following null hypothesis (H0) and alternative hypothesis (H1):

- *H0: There is no significant difference between the prototypes in regards to the users' creative experience.*
- *H1: Interacting with the Auto Modes prototype does lead to a more creative experience than interacting with the Manual Modes prototype.*

H1 has been formulated as my working hypothesis elsewhere in this chapter. To reiterate, my claim is that Auto Modes may feel more creative despite being less predictable and harder to control. In this study, the independent variable was the prototype, and the dependent variable was the Creativity Support Index (CSI) score for each participant in the study for both prototypes. The null hypothesis would be rejected if the mean of scores for the Auto Modes prototype were significantly higher than the mean of scores for Manual Modes. The statistical comparative analysis was done using a one-tailed paired t-test.

The CSI is a quantitative, psychometric survey designed for evaluating the ability of a system or tool to assist a user engaged in creative work (Cherry & Latulipe, 2014, p. 1). It was designed with the intention of it being an additional evaluation metric that could be used together with other of methods of evaluation, as I have done in this study. The original CSI measures six dimensions (factors) of creativity support: Exploration, Expressiveness, Immersion, Enjoyment, Results Worth Effort, and Collaboration. Due to irrelevance, I decided to remove the Collaboration factor from the survey and adapt the score calculation for five instead of six factors. Cherry and Latulipe (2014) discuss this option, because they acknowledge that evaluating how people collaborate with a system or tool may seem out of place for applications that are designed to be specifically for single users. They advise against removing this factor, because the inclusion of the Collaboration factor makes it possible to have one standardized survey that fits both collaborative and non-collaborative systems or tools. Notwithstanding, I decided to remove it for two reasons:

1. I was worried that the statements related to the Collaboration factor were confusing to a degree that they would affect the participants' motivation to take the survey seriously.

2. These are early prototypes, and they are not intended to be compared with other products. Rather, I needed a well-established and thoroughly validated survey for internal comparisons between Auto Modes and Manual Modes.

Ultimately, I decided that the CSI without the Collaboration factor would suit the requirements for this user study very well.

With the Collaboration factor removed, there are 10 statements on the CSI. Each agreement statement is answered on a scale of "Highly Disagree" (1) to "Highly Agree" (10). In Table 9, the statements are shown grouped under their respective dimensions.[8]

Enjoyment 1. I would be happy to use this system on a regular basis. 2. I enjoyed using the system.
Exploration 1. It was easy for me to explore many different ideas, options, designs, or outcomes, using this system. 2. The system was helpful in allowing me to track different ideas, outcomes, or possibilities.
Expressiveness 1. I was able to be very creative while doing the activity inside this system. 2. The system allowed me to be very expressive.
Immersion 1. My attention was fully tuned to the activity, and I forgot about the system that I was using. 2. I became so absorbed in the activity that I forgot about the system that I was using.
Results Worth Effort 1. I was satisfied with what I got out of the system. 2. What I was able to produce was worth the effort I had to exert to produce it.

Table 9. The 10 Agreement Statements on the modified CSI.

In deployment, the factor names are not shown, and the participant does not see the statements grouped by factor (Cherry & Latulipe, 2014, p. 6). The layout and use of sliders shown in Figure 47 is modeled on the CSI tool designed by Cherry and Latulipe. The sliders do not display the score as numbers, but numbers 1–10 are registered when the participant sets the slider according to their evaluation. The participants were administered the Agreement Statements immediately the sessions with each prototype, as presented in Section 8.2.4.

8 In the original CSI, each statement refers to the "system or tool". In order to avoid wordiness in the statements, I have referred to just the "system".

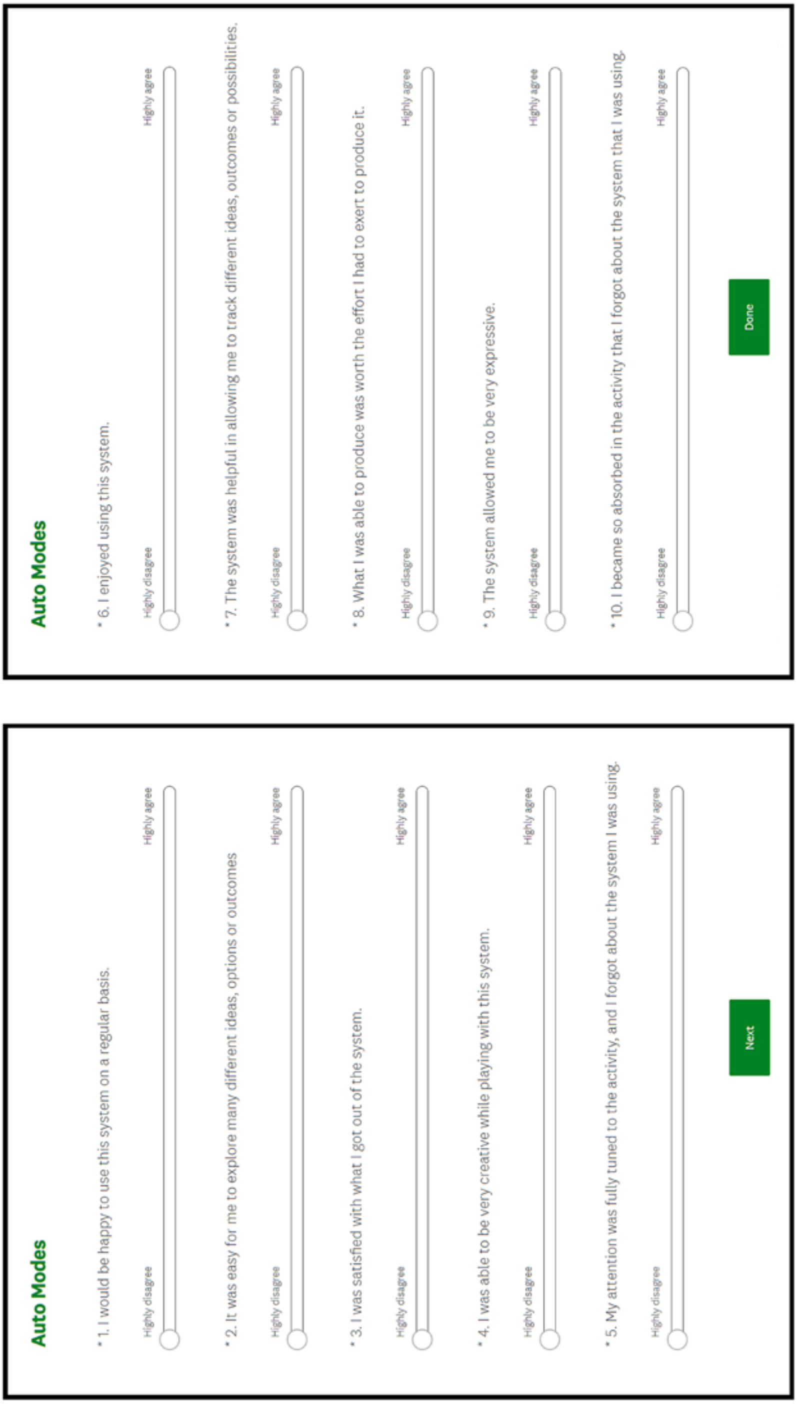


Figure 47. The Agreement Statements as they appeared to the participants over two pages on the iPad.

The Paired-Factor Comparison Test is meant to provide weighting to the score obtained from the Agreement Statements. The test has 10 comparisons (instead of 15 for the original CSI that includes the Collaboration factor). For each pair, participants need to choose one factor description in response to the following statement: “If you could play with these or similar systems in the future, it would be more important to be able to…”

1. Be creative and expressive

2. Become immersed in the activity

3. Enjoy using the system

4. Explore many different ideas, outcomes, or possibilities

5. Produce results that are worth the effort I put in

This was the final survey that the participants received after having played with both prototypes. The factors are paired up in all possible combinations, thus systematically making the participant create a weighting score for the five factors. Each time a factor chosen as the most important out of a pair, it gets one weighting point.

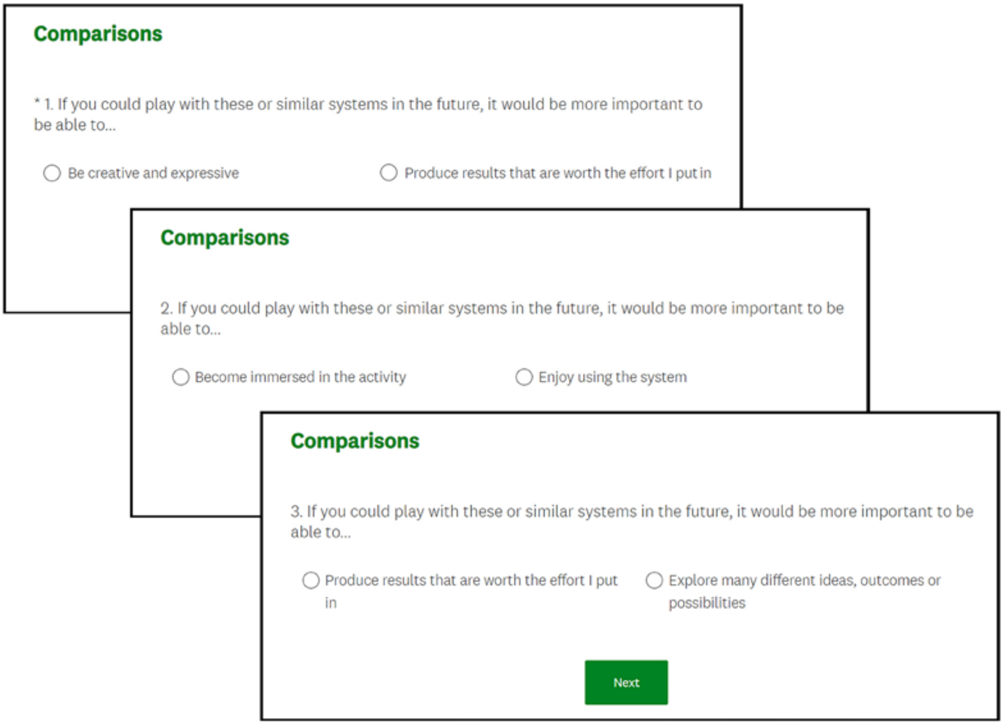


Figure 48. Examples of some of the Factor-Pairs. Each Factor-Pair was displayed as a single page on the iPad.

$$\begin{aligned}\textbf{CSI} = \Big[\, &(\text{Expressiveness1} + \text{Expressiveness2}) * \text{ExpressivenessCount} \; + \\ &(\text{Exploration1} + \text{Exploration2}) * \text{ExplorationCount} \; + \\ &(\text{Enjoyment1} + \text{Enjoyment2}) * \text{EnjoymentCount} \; + \\ &(\text{Immersion1} + \text{Immersion2}) * \text{ImmersionCount} \; + \\ &(\text{ResultsWorthEffort1} + \text{ResultsWorthEffort2}) * \text{ResultsWorthEffortCount}\Big] \, / \, \mathbf{2.0}\end{aligned}$$

Figure 49. Equation for scoring the CSI.

Figure 49 shows the equation for calculating the CSI with five factors. The Factor-Pair counts on the right side reflect how important the participant thinks these factors are in creativity support. For example, if a participant consistently chooses Expressiveness as more important than the other factors it is paired with, it will get a maximum possible count of 4, and the rating from the Agreement Statements related to these factors (Expressiveness1 and Expressiveness2) will affect the total score to a large degree. To calculate the CSI score, each factor subtotal is first multiplied by its factor comparison count. Then, these are summed and divided by two for an index score, out of a highest possible score of 100.

8.3.6 Interviews

The study included three short interviews—one after each prototype session and one final interview after the sessions. The interviews were defined as semi-structured, but they turned out to be very on point and followed the structure of the questions without many digressions. The prototype interviews were identical, and I had prepared the following questions:

- *Can you describe what it was like to play with the system using three adjectives?*
- *What strategies did you employ to engage with the system creatively?*
- *Did these strategies lead to any interesting results? Examples?*
- *How did the context affect the way you play your instrument?*
- *Did you feel in charge most of the time, did you feel like you were mostly following the system's lead, or was it a combination of both?*
- *How did you feel about that balance, was it useful or did you want something else?*
- *What did you miss about the interaction?*

The final interview was in the Comparisons part of the study, and had the following questions prepared:

- *Which prototype do you prefer to play with? Why?*
- *Which prototype do you feel most creative with?*
- *Which prototype feels most like a musical partner to you?*
- *Could you share any other reflections about the prototypes you tried today which may not have come to light in the questionnaires or in the previous interview formats?*

In contrast to the interviews in the previous studies in this thesis, these interviews are meant to be qualitative supplements in the deductive approach of testing the hypotheses related to the tradeoff between user control and system autonomy. Therefore, I have decided to present the results from these interviews in a straightforward and chronological fashion following the structure of the questions. Thoughts and reflections offered by the participants provide a counterweight to the quantitative results from the statistical survey.

8.3.7 Interaction logs

The Max patches running the Auto Modes and Manual Modes prototypes included a subroutine that registered events in a list throughout the sessions. At the end of each session, all of these events were concatenated and saved to a CSV (comma-separated values) file. The resulting file had two sections: one for user actions (events caused by the user interacting directly with the interface) and one for all state changes. The user actions part had the format *Timecode, Action*, as seen in Figure 50 below.

0	Start session
72670	Melodic influence
121270	Thumbs up
146470	Go back
209570	Change
355870	Change
363670	Spectral influence
535270	Rhythmic influence
599470	Change
639370	End session

Figure 50. Readout of user actions.

The states part of the log kept track of all parameters and added a new line for every state change, whether they were caused by user actions or autonomous changes in the case of the Auto Modes prototype. The format for this part of the log was: *Timecode*, *Agent Folder Name*, *Interactive Mode* (1=Shadowing, 2=Mirroring, 3=Coupling), *Rhythmic Influence* (0.0–1.0), *Spectral Influence* (0.0–1.0), *Melodic Influence* (0.0–1.0), *Harmonic Influence* (0.0–1.0), *Audio File Index*, *Audio File Name*, *Quantizing Resolution* (1=1/8, 2=1/16, 3=1/32). Agent Folder Name is the name of the folder containing the trained corpus. Audio File Index and Audio File Name is the index and name of the audio file currently loaded into the FO. The Quantizing Resolution was permanently set to 1/16 for the user study, and was not a selectable option in the interface. Figure 51 shows an example of a readout of states in a session. Many of the parameters stay the same, because any change in one parameter will cause a new line containing all parameters to be printed.

0	guitar_acoustic_solo	1	0.500	0.000	0.600	1.000	51	01_Rock2-85-F_solo_mic	2
44726	guitar_acoustic_solo	2	0.500	0.000	0.600	1.000	51	01_Rock2-85-F_solo_mic	2
87726	guitar_acoustic_solo	1	0.500	0.000	0.600	1.000	51	01_Rock2-85-F_solo_mic	2
102026	guitar_acoustic_solo	1	1.000	0.300	0.200	0.400	51	01_Rock2-85-F_solo_mic	2
139126	guitar_acoustic_solo	3	1.000	0.300	0.200	0.400	51	01_Rock2-85-F_solo_mic	2
377526	guitar_acoustic_solo	3	0.400	1.000	0.200	0.400	51	01_Rock2-85-F_solo_mic	2
381426	guitar_acoustic_solo	3	0.400	1.000	0.200	0.400	61	02_BN1-147-Gb_solo_mic	2
385826	guitar_acoustic_solo	3	0.400	1.000	0.200	0.400	59	01_SS3-98-C_solo_mic	2
571826	guitar_acoustic_solo	3	0.500	0.000	1.000	0.400	59	01_SS3-98-C_solo_mic	2
580726	guitar_acoustic_solo	3	0.500	0.000	1.000	0.400	20	00_Rock2-142-D_solo_mic	2
587126	guitar_acoustic_solo	3	0.500	0.000	1.000	0.400	9	00_Funk2-119-G_solo_mic	2

Figure 51. Readout of states in a session.

8.3.8 Home assignment

In order to capture the participants' evaluation of their own performances with the prototypes, I commissioned a design company to make a web application where the participants could listen back to the sessions and simultaneously rate the perceived quality of the interaction using a screen-based slider. Here, I will present a short description of the procedure of collecting the evaluation data from the participants.

After mixing down the sessions to stereo MP3 tracks, I could upload them to a WordPress-based web platform. Using any one of the audio filenames as a Javascript suffix in the URL would load a custom page where this particular track could be played and rated by the participant. Hence, I could send an email to the participant with an introductory assignment text followed by links for evaluating the two sessions. It was important to make the application as simple and minimally time-consuming as possible to ensure that all participants would go through with the evaluation assignment. Ideally, it would not take any longer than the time it takes to listen to the sessions. A screenshot of the web application is shown in Figure 52.

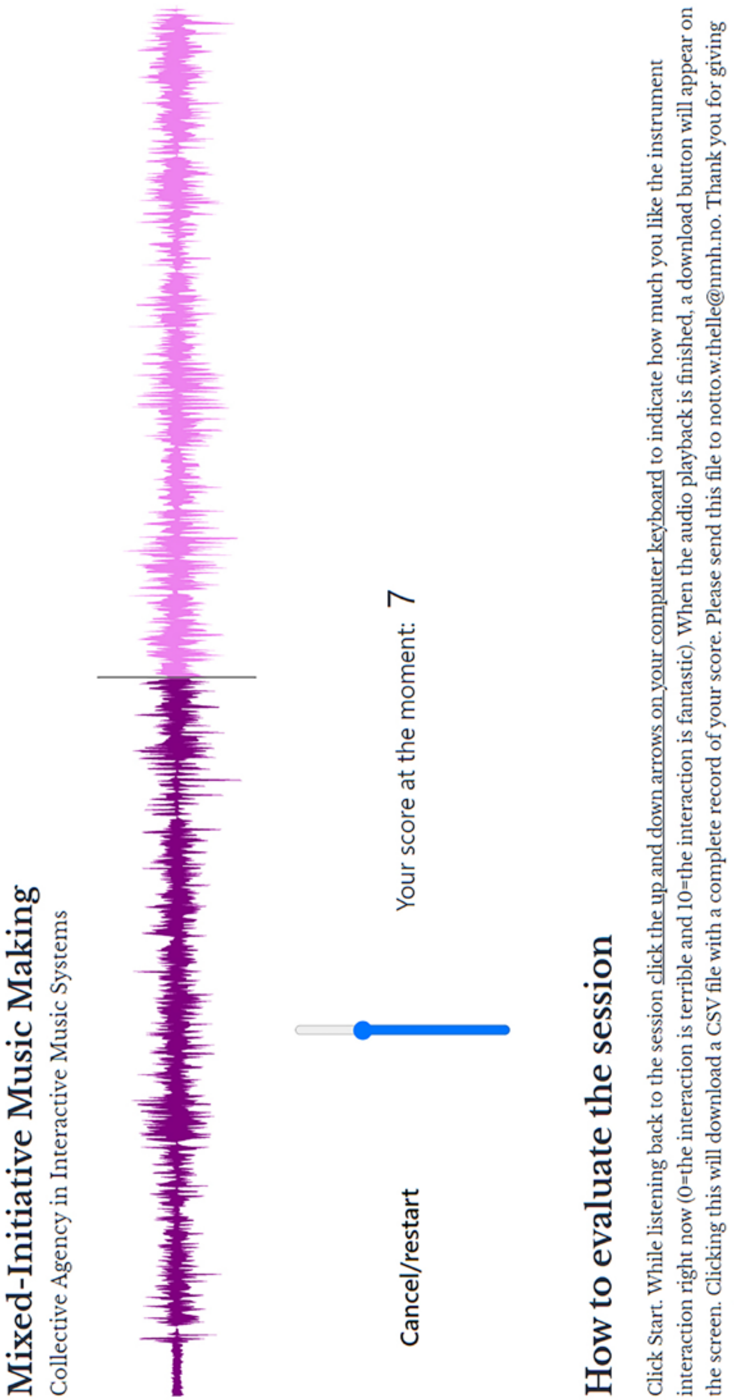


Figure 52. Screenshot of the evaluation web application.

An explanation of how to evaluate the session was available on the same page as the application itself. After pressing a start button, the audio would start playing back and a cursor would show the progression along a waveform representation of the session. While listening back to the session, the participant could rate the interaction in real-time, either by using the up or down arrows on the computer keyboard or by handling the web-based slider directly. The score was an indication of how much the participant liked the instrument interaction now, running from 0 (the interaction is terrible) to 10 (the interaction is fantastic). The running score would be displayed to the right of the slider. The slider would be at score 5 (the middle) as a default at the start of the session. The procedure was designed as a one-shot operation, so it was not possible to fast-forward or skip backward in the audio file. The only other option was to cancel the operation and start from the beginning. Upon finishing the audio playback, a download button would appear on the screen. Clicking this would download a CSV file with a complete record of the participant's scores, sampled once every second. After going through the evaluation of both sessions, the participant could then send me an email with the two CSV files as attachments. With the CSV files, I could produce graphs showing timelines of how well all participants thought the interaction worked for the duration of both sessions. Each session lasted 10 minutes, so the whole operation would take just over 20 minutes per participant.

8.4 Results: Statistical survey

Table 10 shows the total CSI scores for Auto Modes and Manual Modes. The individual scores for all the Agreement Statements and the Paired-Factor Comparison Test are available in Appendix D.5.

Participant	Instrument	Auto Modes	Manual Modes
P1	Laptop/synth	56.0	61.0
P2	Electric guitar	22.0	47.0
P3	Saxophone	32.0	67.5
P4	Electric guitar	62.0	66.0
P5	Electric guitar	55.5	62.5
P6	Piano	20.5	44.0
P7	Vocals	28.0	39.5
P8	Electric guitar	48.0	16.5

Table 10. All the CSI scores.

The mean score for all participants was 40.5 for Auto Modes and 50.5 for Manual Modes. Figure 53 shows the box plots for both prototypes.

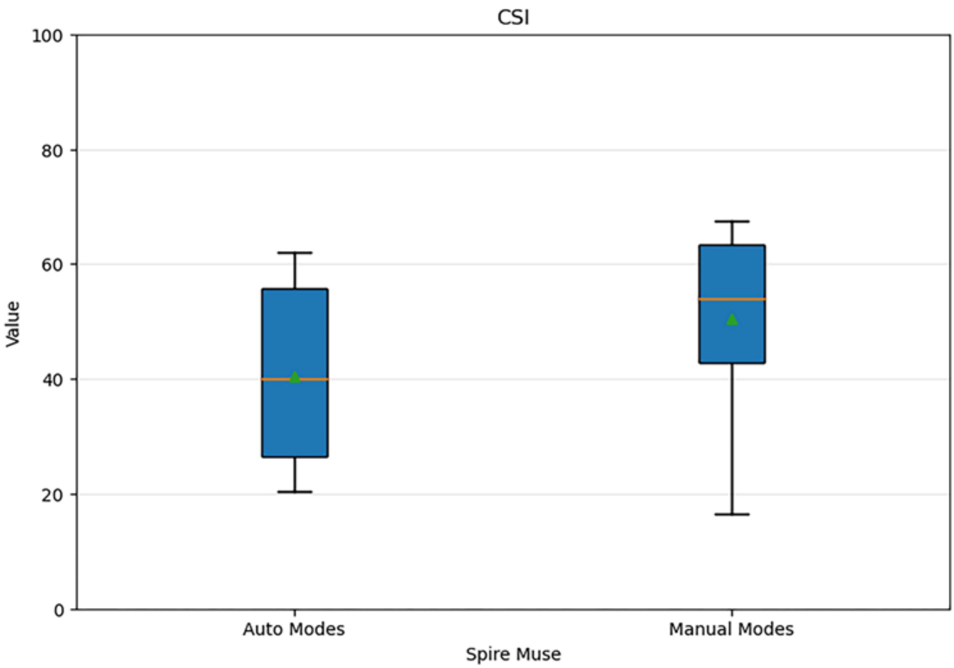


Figure 53. Box plots for Auto Modes and Manual Modes.

Results of the one-tailed paired-t test with a significance level set to 0.05 (5%) indicated that there is a non-significant mean difference between Auto Modes (M = 40.5, SD = 16.7) and Manual Modes (M = 50.5, SD = 17.4), t(7) = 1.4, p = 0.898.

Parameter	Value
P-value	0.898
t	1.401
Sample size (n)	8
Average of differences ($\bar{x}_d$)	10
Standard deviation of differences (S_d)	20.185
Normality p-value	0.315
A priori power	0.653
Post hoc power	0.00172
Skewness	-1.145
Skewness shape	Potentially symmetrical
Excess kurtosis	2.277
Kurtosis shape	Potentially mesokurtic
Outliers	-31.5

Table 11. Key statistical values from the t-test.

Since the p-value > 0.05, H0 cannot be rejected. A non-significance result cannot prove that H0 is correct, only that the null assumption cannot be rejected. Judging by the CSI scores, it is clear that the general tendency for the results is that the effect is in the opposite direction from the alternative hypothesis, hence the high p-value.

The low sample size made this test vulnerable to outliers in the data. There is one clear outlier, namely P8's low score for Manual Modes (16.5), which goes against the tendency for all other participants. This is not an error, as P8 was clearly not satisfied with his Manual Modes session and stated a clear preference for the Auto Modes prototype (more details in the next section). Notwithstanding, the low sample size is problematic and hence the power of the test is questionable.

When editing and mixing down the session recordings, I noticed that the guitarists had generally been better at adapting to the Auto Modes prototype than the other instrumentalists. Therefore, I think it is interesting to include the results from the four guitarists as a subgroup. As seen in the box plot for the guitarists in Figure 54, the mean CSI score for Auto Modes (46.9) is almost equal to the score for Manual Modes (48.0). A sample size of four is too low to perform a valid t-test, but there is a clear tendency showing that the guitarists generally rated Auto Modes higher than the other instrumentalists did.

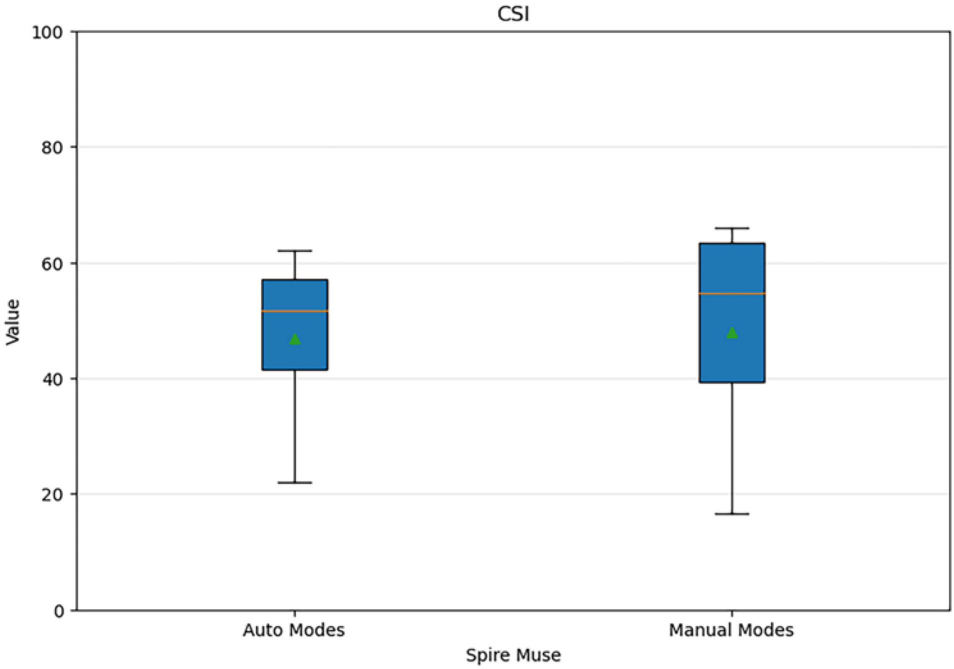


Figure 54. Box plot for the Auto Modes and Manual Modes scores for the guitarists.

8.5 Results: Interviews

At the beginning of both post-session interviews, I asked participants to describe what it was like to play with the system using up to three adjectives. Table 12 shows the replies given for Auto Modes and Manual Modes respectively.

	Auto Modes	Manual Modes
P1	Surprising	Confusing Eye-opening
P2	Fun Chaotic Fumbling	Fun Bouncy Playful
P3	Difficult Not permissible Frustrating	Creative Experimental Inventive
P4	Fun Challenging Revealing (about oneself)	Playful Thought provoking Inspiring Fun
P5	Uncanny (not in a bad sense) Synthetic (not in a bad sense) Disturbing (not in a bad sense)	Fun Fascinating
P6	Demanding Unclear Exciting	Challenging Easier (than AM) Fun
P7	Challenging Interesting Experimental	Interesting
P8	More independent (than MM) More inventive (than MM) Creative	Limiting Interesting Potential

Table 12. Immediate reactions to playing with the prototypes.

P5 was keen to emphasize that his adjectives "uncanny", "synthetic", and "disturbing" after the Auto Modes session were not meant as something negative. This was an interview performed in Norwegian, but he used the English word "uncanny", which I immediately assumed to be a reference to "uncanny valley"—the feeling of unease that may be evoked by witnessing something nonhuman appear close to human (Mori et al., 2012). The exact exchange was as follows:

> N: Could you describe what it was like to play with the system using three adjectives?
>
> P5: Yeah. What's the word in Norwegian… uncanny… I don't know what it's called in Norwegian.
>
> N: Yeah, I know what you mean.
>
> P5: Yeah (laughs). But not in a negative sense. But… synthetic, is what I'm thinking. And… disturbing, I think.

N: (laughs)

P5: But these are not negative words.

N: No, I see what you mean (laughs).

8.5.1 Creative strategies

The participants employed different strategies to engage with the prototypes creatively. Six participants (P1, P2, P3, P6, P7, and P8) started their first sessions with more or less premeditated strategies based on simple and repeated motifs as a method to gauge the system's response. Three of these participants (P2, P6, and P7) were quick to point out that they abandoned their initial strategies because they were not getting a response they felt was adequate. These all had Auto Modes as the first prototype. P2 started with "some simple things", but thought the system was too eager to "skip along to entirely new landscapes" and he felt insistent on trying to "pull it back". He later discovered that by thinking less rhythmically, it became easier to accept the system's contributions, and there were fleeting moments he thought were "really fun". P6 was more critical. He also started "with some simple ideas", but felt that the system was responding "out of tune and out of style". He tried responding to this by playing more freely. There were moments where he felt "there was something there", but these moments passed by very quickly. By the end of the session, he was left with a feeling of "not being there at all". P7 began the first session by trying to sing consistently, and then she realized that she should try to listen as well. She discovered that in some modes, it was not working: "It was, of course, responsive, (but) not… like, not creative, and… or, like, initiating." She then adopted a strategy of copying the system's output, and she tried to adapt her own style according to how the influences were set. For example, she started to play the mouth trumpet when the spectral influence was high, and sang rhythmically and in a staccato style when the rhythmic influence was high. One thing she said at this point in the interview later made me realize that she may have been under the impression that setting the influences herself was not an option in Auto Modes, and that she instead was waiting for random changes to the parameters caused by the *Change* button: "I don't really remember seeing, like, harmonic high. And so I didn't really focus on that too much. But in general, I sing more melodically, perhaps." In other words, an apparently missed opportunity to manipulate directly the influence presets seems to have had a decisive effect on her strategic choices throughout the Auto Modes session.

The three participants (P2, P6, and P7) described above who started with Auto Modes and faced challenges gave the prototype very low CSI scores—a combined mean of 23.5—only a little more than half the total mean for the prototype. They were generally more satisfied with the ensuing Manual Modes session. P2 expressed that “we were much better friends this time”. He used the shadowing mode to hear if the system played along, and found this to be a playful form of interaction. Then he switched to coupling, where he became so engaged that he “didn’t have time to give it a thumbs up”. When the coupling mode got too repetitive, he would switch back to shadowing. He would spend most of the session alternating between the two modes. P6 spent the Manual Modes session trying to understand the concept of the interactive modes. He admitted that he did not get much out of this, but he discovered that the system responded better when the rhythmic or melodic influence presets were selected, expressing that then: “We used […] the same language, or […] it had some kind of tacit knowledge. While the other one (Auto Modes) was totally, like… totally like an alien”. In her second session with Manual Modes, P7 explained that she liked being able to switch between the modes. For her as a singer, she felt like coupling was particularly nice, because it reminded her of the sort of interaction she would expect from a musician partner: “I felt like there was more harmony structure to work with”. She also switched the influences a few times, but stuck mostly with the melodic preset. Additionally, she enjoyed shadowing: “There were some times where it was like: ‘Oh, this is super cool!’” Mirroring was more challenging, but interesting at times.

P1, who also began with Auto Modes, was far more favorably inclined to his first session. He began with two kinds of motifs. One was “more chord-based, and sort of staccato”, which he thought would be a good way to get the system going. The other was arpeggiated style motifs with just singular notes. He switched between these two approaches a few times. He felt that he was getting some interesting results in the early middle of the Auto Modes session while he was using the chords. The system started to do “some really nice little licks over the top”. He would carry on doing these chords while the system “played the lead, which was really cool”. This lasted for about a minute, and he felt “really into the system then”. Then he decided to change, and “it took a while to get going, you know, it felt like we were feeling each other out”. In his second session with Manual Modes, P1 began with similar tactics as in the first session (going between staccato chords and arpeggiator style melodies). However, he found that he struggled more with the shadowing and mirroring modes in this session. He pointed out that he seemed to get better responses when playing dissonant chords or abrupt changes in pitch than when playing things that were “sonically pleasing” and staying in the same note range. I theorized that this could be due to a lack of attack in his synth sound, and that perhaps the machine listening was not catching the onsets properly. In this second session, he really enjoyed the coupling mode, claiming it “came out straight off”.

P3 started out with Manual Modes:

> First, I just started basically by playing some patterns in a scale that I had already in my mind, and then continue on that scale. And it was okay to follow with the same patterns, and also staying in the same scale with the system.

When asked if this strategy lead to some interesting results, he explained that he particularly enjoyed the system's response when the rhythmical and spectral influences were selected: "I thought, because it was really questioning the different aspects of the patterns I put in first, and then doing this really… randomize the rhythmical pattern, but which were creating something interesting… new ideas". He applied a similar approach of playing patterns and waiting for responses in the second session with Auto Modes, but complained that "the software moved elsewhere" when he responded.

P8 also started with Manual Modes, with a strategy of "keeping it simple and playing some repetitive phrases that the system perhaps could better understand". He thought the coupling mode complimented what he was playing quite well, but he had the sense that he was playing along with the system more than that the system was interpreting anything (which is true for coupling mode in the Manual Modes prototype). In his second session with Auto Modes, he tried a few different things, while keeping it simple. He tried some chords, and also attempted taking on a bass playing role. He thought it worked very well when he was playing chords: "It was almost as if the system could figure out and… play solo over the chords" (in other words, he could hear the meso time scale harmonic dynamics algorithm actually working). He thought this resulted in some "quite coherent ideas". P8 was the only participant who rated Auto Modes higher than Manual Modes.

Notably, the two participants (P4 and P5) who started their sessions without claiming to have had any premeditated musical ideas gave both prototypes CSI scores that were much higher than the mean (62.0 and 55.5 for Auto Modes, 66.0 and 62.5 for Manual Modes respectively). Their initial strategies differed from the "simple and repetitive" approaches adopted by the rest of the group. Both started with Manual Modes in their first sessions:

> I kind of, I wanted to treat it very much like I would treat a human instrumentalist. So that was kind of interesting to see, like, if I just like stopped playing for a little while, like what, you know, will it… if I give it some space as a soloist, I did want to hear what it's got to express. That was nice. And I sort of recognized that the times when it wasn't responding in a way that I would have expected from a human performer, I kind of had the same response that I've had in the past when

> you play with a human performer and then, like: "What are you doing, man?" Like, kind of, like: "Are we playing together?" Like: "Are you listening to me? Are you part of this?" So, in a way my response […] was sort of an unconscious reaction but it was still very, like… I started treating it very much like a human performer, which is surprising. (P4)

> I didn't think that much. Because there was a nice, like, call and response. So, really, I just sat there and played and I thought… yeah, it was like playing with another guitar player in many ways. And… I wasn't thrown by the timing or anything like that. So I didn't think so much about what I was playing, in a way. And I think there was nice harmonizing even when out of key. It wasn't floundering, sort of. (P5)

In their second sessions with Auto Modes, the attitudes of P4 and P5 were somewhat reversed. P4 went from a standpoint of being slightly bemused by some of the system's responses in the first session to essentially non-judgmental in the second, while P5 seems to have gone the opposite way:

> I tried to not really interact with it at all, I think, until near the […] last couple of minutes, when I started changing things […]. I kind of just wanted to focus on trying to make some music, I guess, which was interesting also, because I didn't have to feel responsible for any, sort of… anything other than myself. […] I noticed that I was […] listening quite differently than in the first example. So I think that was the biggest difference was that I just, kind of, if I knew I didn't have to worry about anything, and I wasn't going to until, yeah, so maybe seven or eight minutes. And then I was like: "Okay, like, let's push things a little bit". Or like: "Let's try and find something new". (P4)

> I was battling for the direction. That was the fun thing about it. That you sort of have to demand your space, and then the computer, or the system does that as well. And both do it with varying degrees of success. That was the very fascinating thing about it, and fun. (P5)

8.5.2 Initiative balance

When asked what they felt about the initiative balance, the participants replied the following for Auto Modes:

> Probably I'd say it's a combination of both. (P1)
>
> I think probably it was mostly me trying to take control, but I guess there was a combination… sometimes. (P2)
>
> This time I was more trying to follow the system. (P3)
>
> I think it was a combination again. […] It's kind of interesting because, like, I could also see how… I… this is a weird way to answer this question but I could imagine a scenario where I felt like I was in control the whole time, or I felt like I was leading the whole time. But I didn't […]. When I heard something change, I felt like okay, this is… this is not me making a change. This is not my decision. (P4)
>
> Quite equal. I think. (P5)
>
> I tried to make it a combination, I guess. But it felt very unclear what was going on. (P6)
>
> There's certainly times I felt like: "Oh, it responded to me, that's really cool". But then […] sometimes it would change and I wouldn't know if it was like: "Woah, did I make a mistake and that's why it's moving on"? Or like: "It's responding, of course, to what you're doing". But it was hard to tell in what ways was responding, but yeah, it was using the influences and analyzing. I don't know. Yeah, yes, I definitely felt like both were true. Like, there were times where I could tell it had responded to what I had done and then times where I was completely responding and relying on what it had done. (P7)
>
> This time it was a little bit like a combination. But I think still the system controls a bit itself. So you still have to play a little bit along with the system. But when you have found something you like and that can be interpreted well, it feels as if it follows what I am playing. (P8)

The participants had varying opinions about this balance. P2, P3, P6, and P7 expressed a desire to be more in control, and this corresponds with the low scores they gave the Auto Modes prototype. However, P7 acknowledged the usefulness of being challenged: "It's nice to be creative as well and be forced into new territories". P1 felt quite content with the possibility to cycle through possibilities with the *Change* pedal. When he played the chords, he felt he was taking on a passive role and leaving the system to lead. On the other hand, having the pedals at his disposal reminded him that this is a tool, and he wants to be able to control it. In that sense, he felt like he had the final word. P4, P5, and P8 were quite happy about the balance, but P4 commented that it could be dependent on mood: "But I could also see how, myself on a different day, or somebody else would just be, like: 'Whatever. This is my time'".

As for the initiative balance in Manual Modes, several participants claimed it was more a matter of delegating the initiative. According to P1, the system was struggling to keep up in shadowing and mirroring, while with coupling he was letting the system take the lead. P2 expressed sentiments along the same lines:

> It was a combination now, definitely. A little bit in a way where I could kind of choose it as well. Because if I went back to shadowing, for example, I felt as if, well, I was very much in control. And then with coupling it was fun to let it control.

P4 was surprised by how much he decided to follow the system. He thought this could be because it was a (familiar) guitar sound, which made it easier to know what to expect. On the other hand, he was expecting a "tighter mimicking" in the shadowing mode, and was surprised by the unexpected amount of sounds that appeared independent of what he was doing:

> It wasn't always like: "Oh, we're doing this together". Which was actually quite nice. Like, it wasn't just […] an intelligent delay, right? Like, it wasn't just… there was some other input there. So […] I surprised myself in how much I was, sort of, following, in a way. Like: "What are you trying to say"?

After his Manual Modes session, P5 felt that he was "the one who called the shots". About that balance, he said he wished that the system could be more in control and take more direction. P8 also felt that he was in charge, and would have wanted another kind of balance. He experienced the way Manual Modes worked as more like "filling in the background" where it "threw in some snippets" that felt somewhat disjointed. With this kind of interaction, he would have preferred another sound than guitar—something that would have worked better as background material.

8.5.3 Comparing Auto Modes and Manual Modes

Table 13 shows what the participants responded in the final interview when asked to compare different aspects of the prototypes.

	Preferred to play with	Most creative with	Most like a partner
P1	Auto Modes	Auto Modes	Auto Modes
P2	Manual Modes	Manual Modes	Manual Modes
P3	Manual Modes	Manual Modes	Manual Modes
P4	Manual Modes	Manual Modes	Auto Modes
P5	Auto Modes	Manual Modes	Auto Modes
P6	Manual Modes	Manual Modes	Manual Modes
P7	Manual Modes	Auto Modes	Manual Modes
P8	Auto Modes	Auto Modes	Auto Modes

Table 13. Comparing different aspects of the prototypes.

The participants gave some interesting qualifiers and reservations when answering these questions, some of which are worth presenting here. In relation to the question about which prototype the participants preferred to play with, here are some of the comments:

> I think I prefer the first one (Auto Modes). Maybe it's the simplicity of the options of just having Next so you don't even need to know what it's doing you just you just have a Next button. (P1)

> The manual. With that one I felt I had more control, plain and simple. (P2)

> The Manual Modes because I thought it was more responding. […] Not responding, but it was more going in a way I prefer subjectively. (P3)

> Well, I think I preferred the manual version. But I also don't know if that's just because it was the first one I did today. […] Things are, at least for me, things are nice when they're new and shiny. So […] everything was fresh. I enjoyed them both. But I think, yeah, if I had to pick one, I'd pick the manual one. But, yeah, it may just be because it was the first one today. (P4)

> I think the last one (Auto Modes), because it adds something that I didn't bring. It takes more direction and catches me off-guard. (P5)

> The manual one because I have more control over which mode and, like, at which point in the improvisation I wanted to switch, I guess. (P7)

> I prefer playing with the automatic one. I feel a bit overwhelmed by having too many parameters to deal with. (P8)

Following are some of the comments made in relation to the question with which prototype the participants felt most creative:

> I definitely got more, like, lost in the first… in the manual mode. More, sort of, in flow. But I think… I don't know, I think the auto mode has a lot of potential as well, because you don't, at least for me, because I would only have to worry about playing my own instrument and not steering anything else. And I think, yeah, on another day […] that might have been more engaging for me. (P4)

> Yeah, the first (Manual Modes) I was more creative, maybe. And, of course, that was my first session as well. […] But it was more positive… more exciting with the second one. (P5)

> Challenging question. I answered the auto one because maybe it pushed me into places where I wouldn't normally go. But depends how you're defining creative and expressive as well. (P7)

Even more nuances became apparent in the answers to the question about which prototype feels most like a musical partner:

> The first one did (Auto Modes). Yeah. Oh well, I mean the second one felt like you know, a musical partner that's, like, way elsewhere. Playing the virtuoso. (P1)

> Yeah, maybe the auto mode? Like I said, it was… I started listening in a different way. And I think that that felt much more musical […]. Yeah, I was listening to [it] the more musical way instead of, I think, with the manual mode, you're sort of listening for confirmation, then like, I've changed something, and I can hear it. Whereas in auto mode, you're reacting to a change that you've heard. Yeah. And then […] you maybe you look for confirmation on the interface or whatever. And that, to me, I think is […] closer to […] the music making experience, I guess. (P4)

> The manual one because it had… I could choose the coupling setting. And to me as a vocalist, the coupling setting was the most realistic to who I'd be usually playing with. (P7)

Summed up, the participant's answers and reflections on those answers reveal complexities and nuances that are well worth revisiting in the ensuing discussion.

8.6 Results: Interaction logs and self-evaluations

The following pages show the results of the interaction logs and the evaluations that the participants did in the home assignments. Figures 55–62 show all the most relevant data for each participant on one single page. Color coding schemes are used to display the influence settings and interactive modes per session. I have only included actions that the participants used for each session, which explains why there is an unequal number of lanes in the action sections. I have left out some data from the interaction logs. For example, the figures show when participants have initiated FO changes, but not which corpus files have been selected in these changes. The Evaluation sections show the participants' own scores generated in the home assignments.

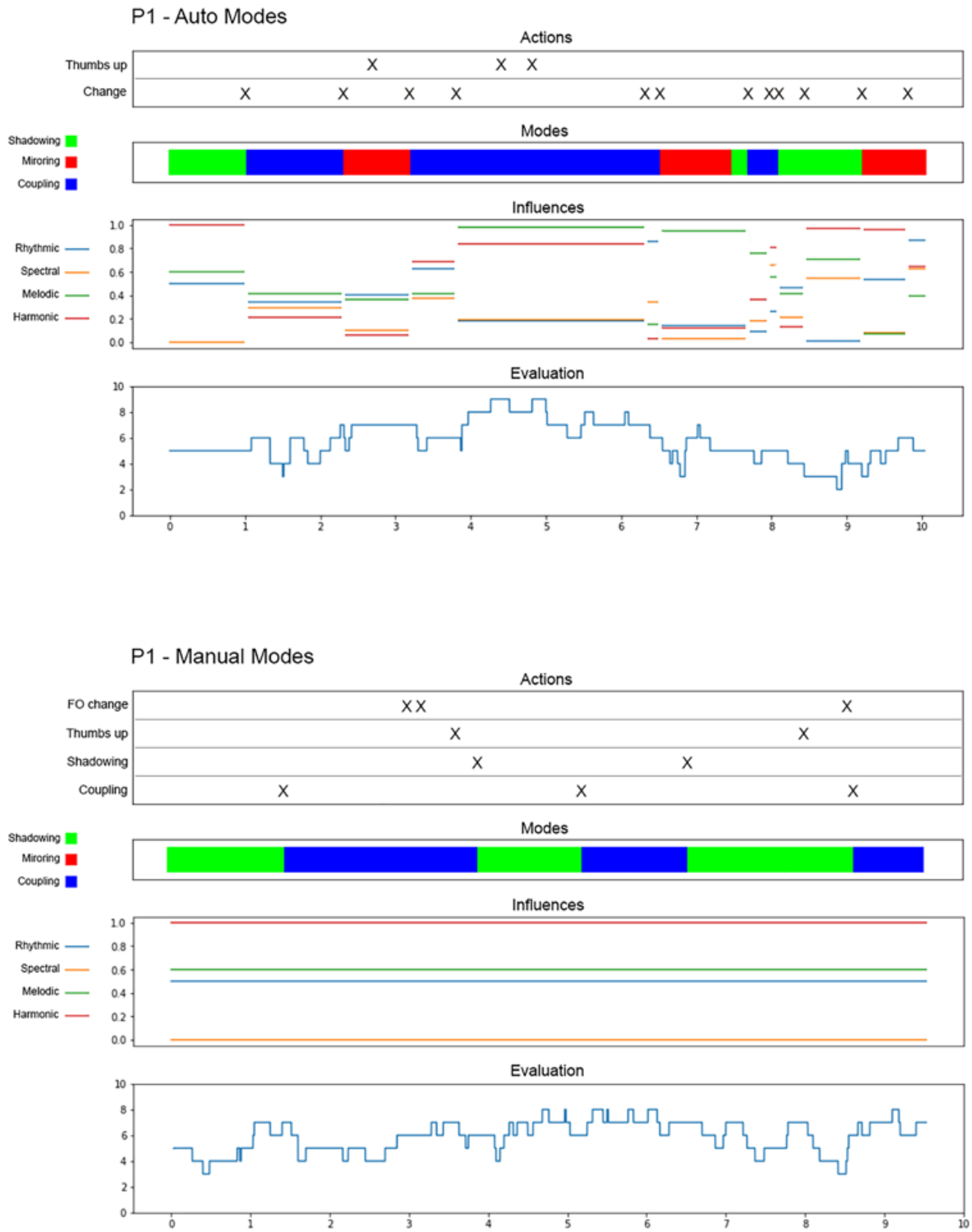


Figure 55. Interaction logs and evaluations for P1.

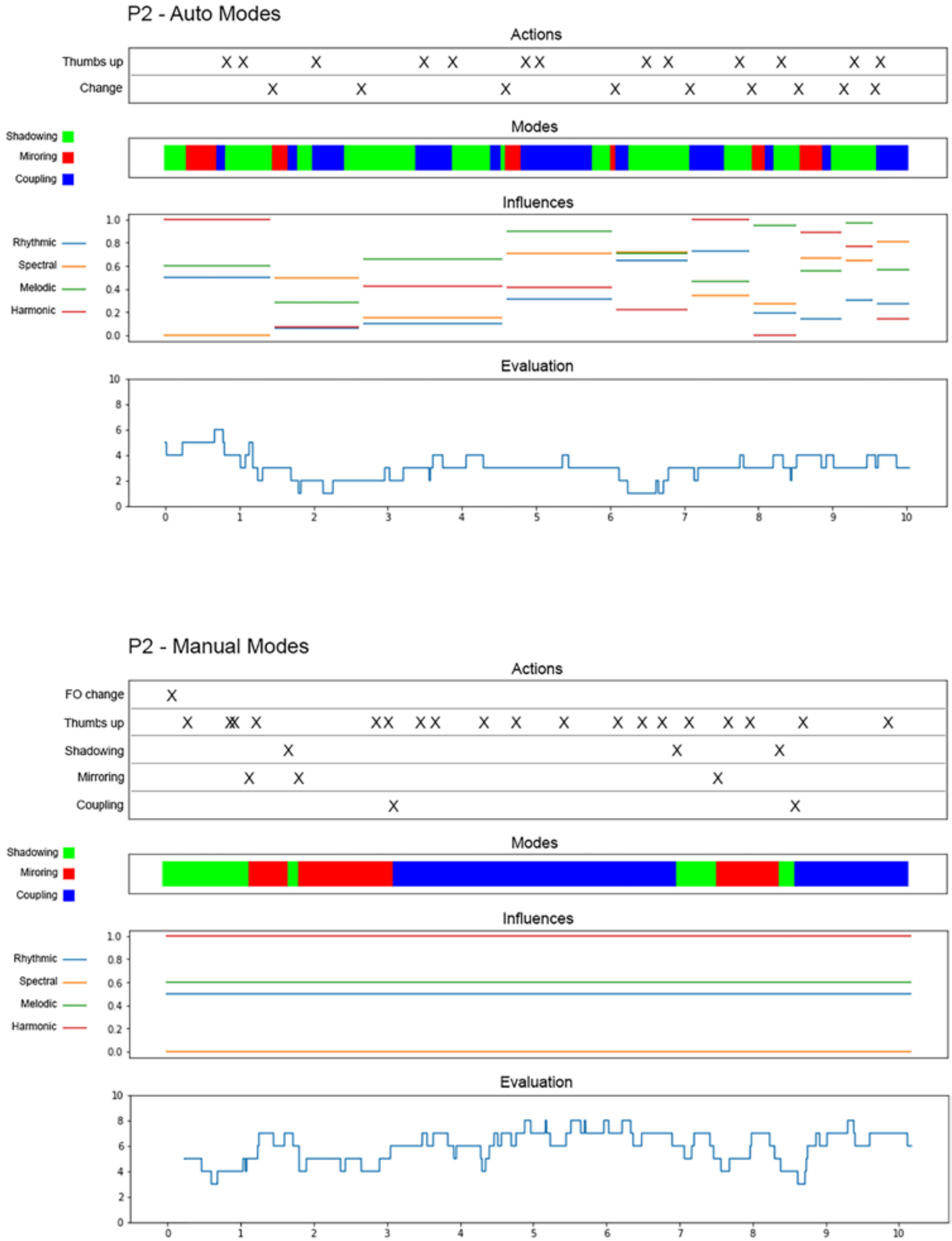


Figure 56. Interaction logs and evaluations for P2.

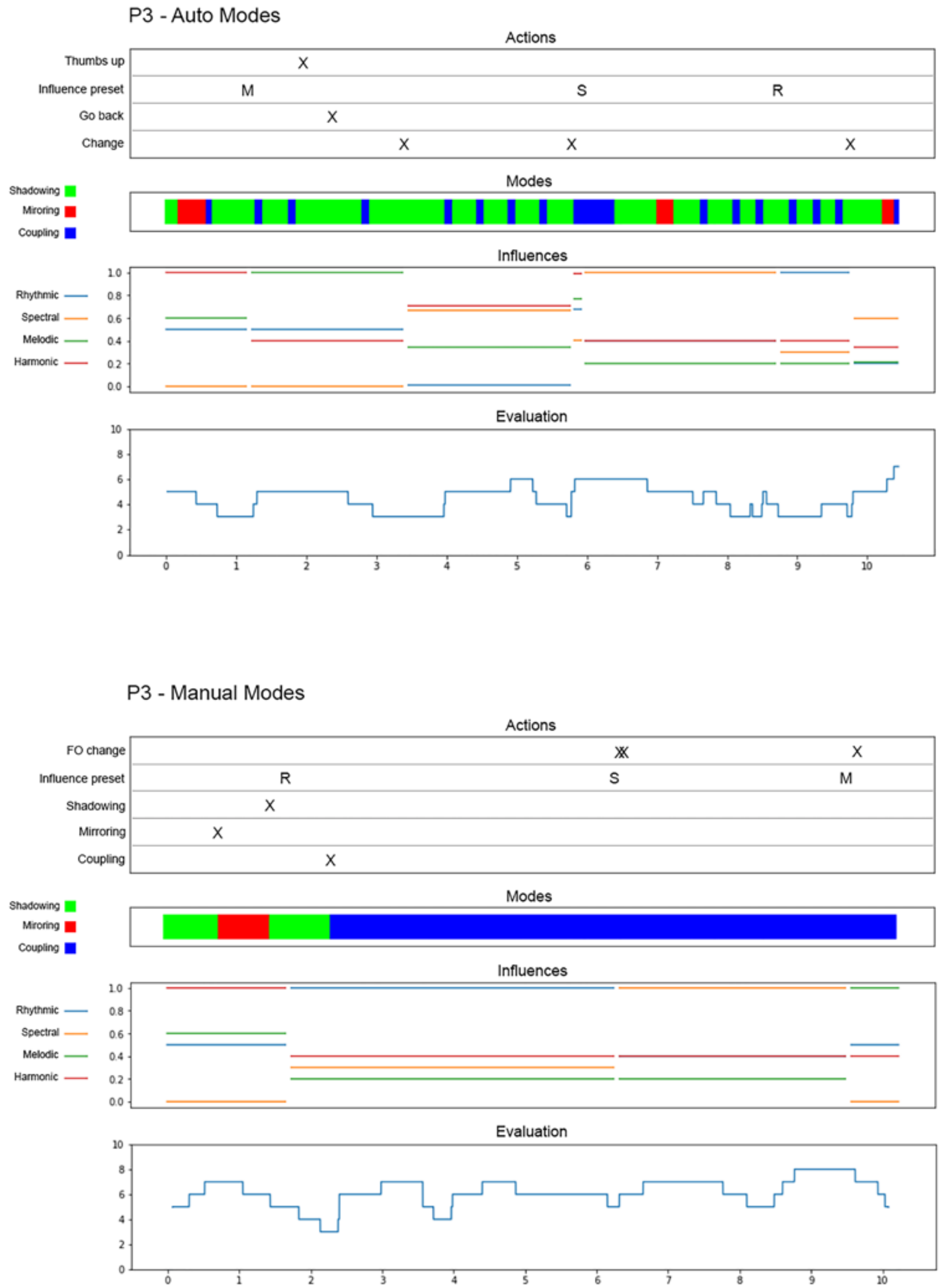
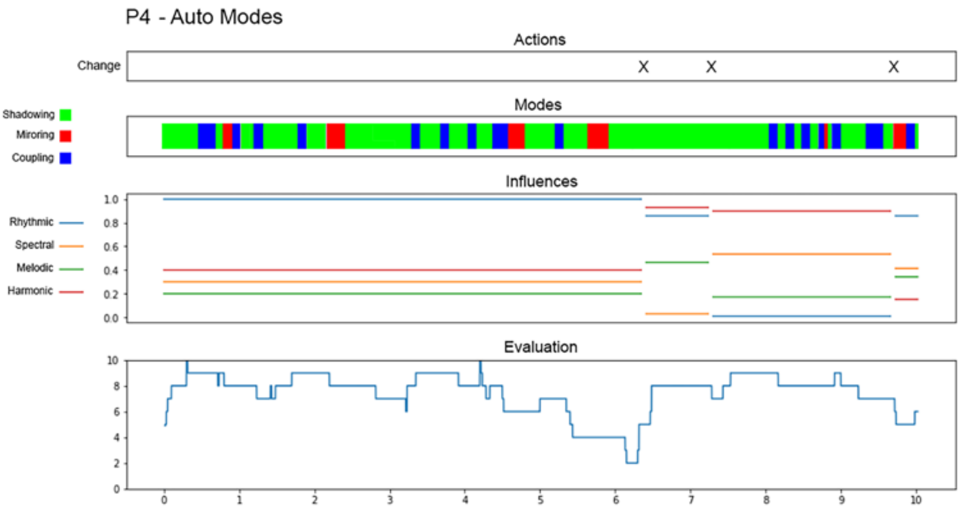


Figure 57. Interaction logs and evaluations for P3.



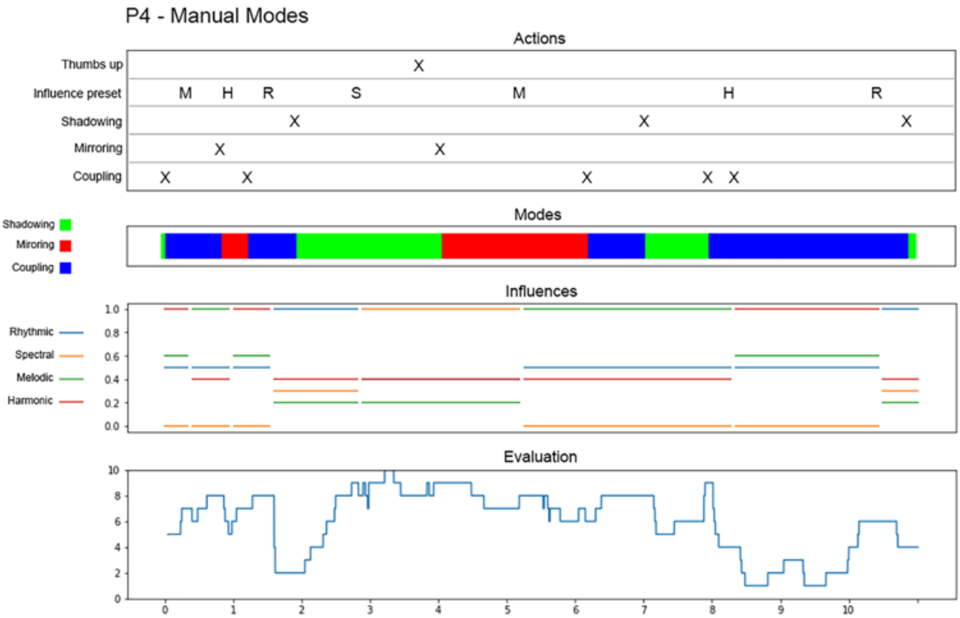


Figure 58. Interaction logs and evaluations for P4.

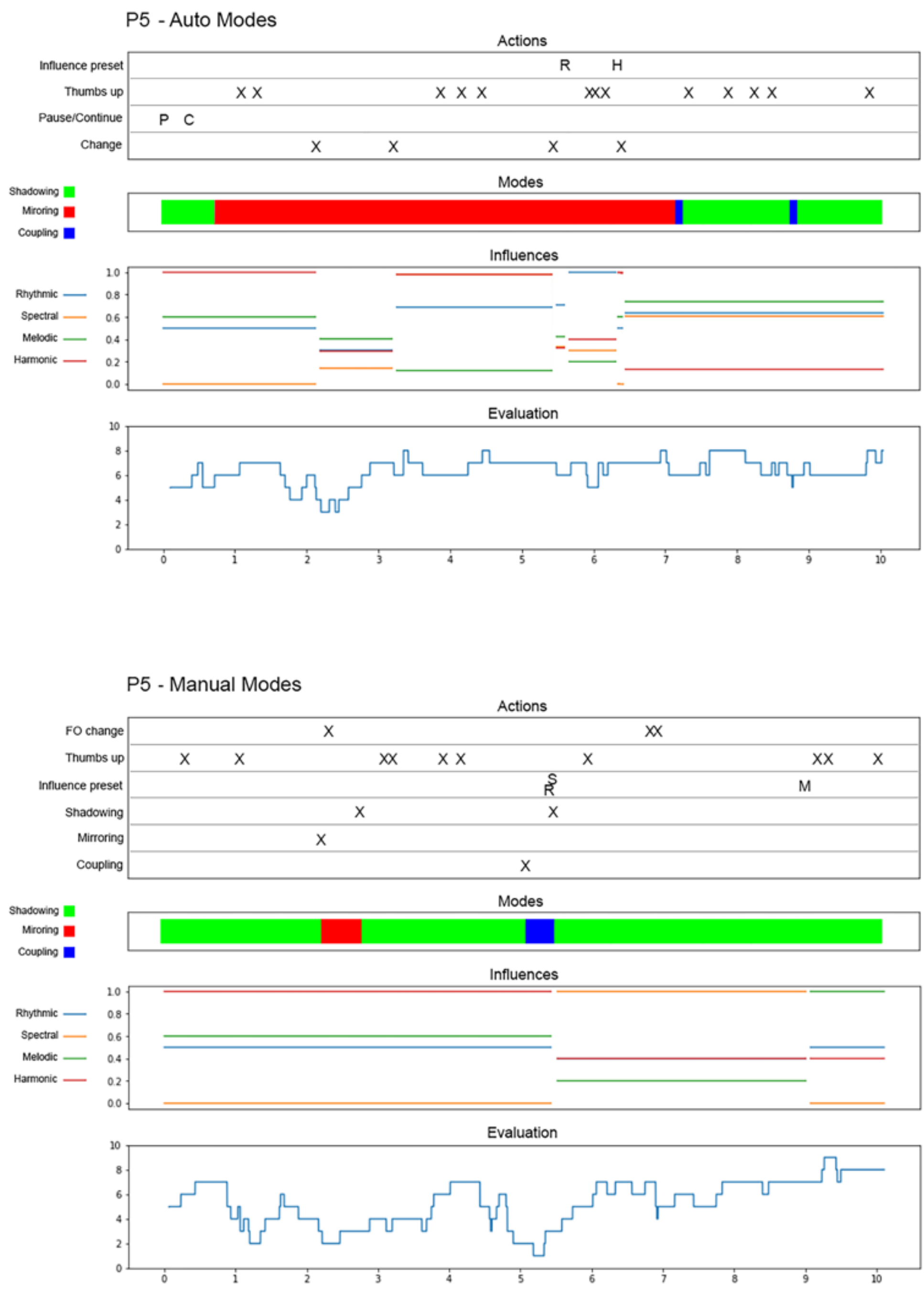


Figure 59. Interaction logs and evaluations for P5.

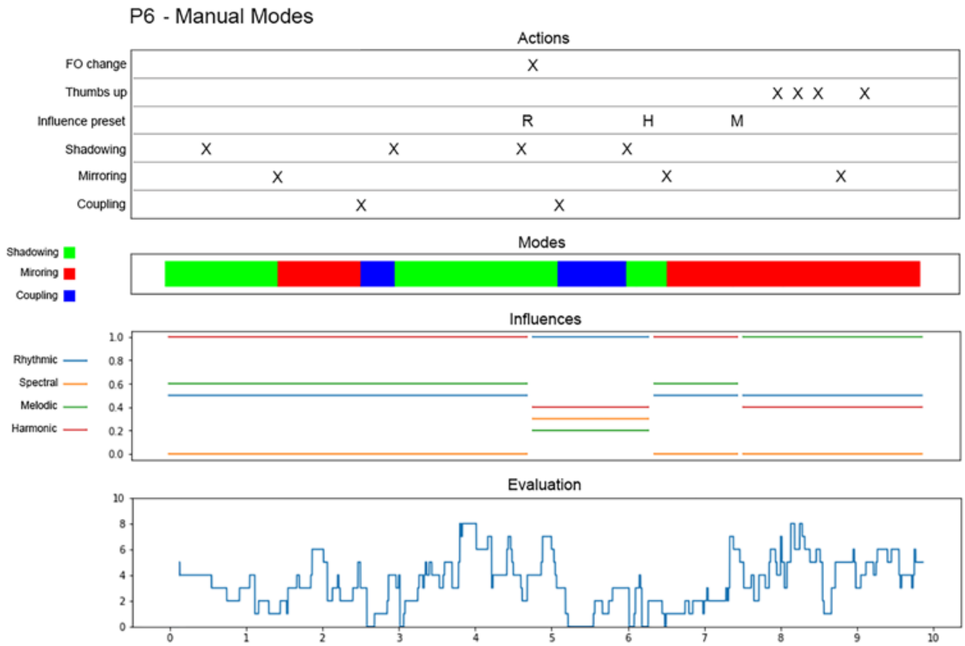


Figure 60. Interaction logs and evaluations for P6.

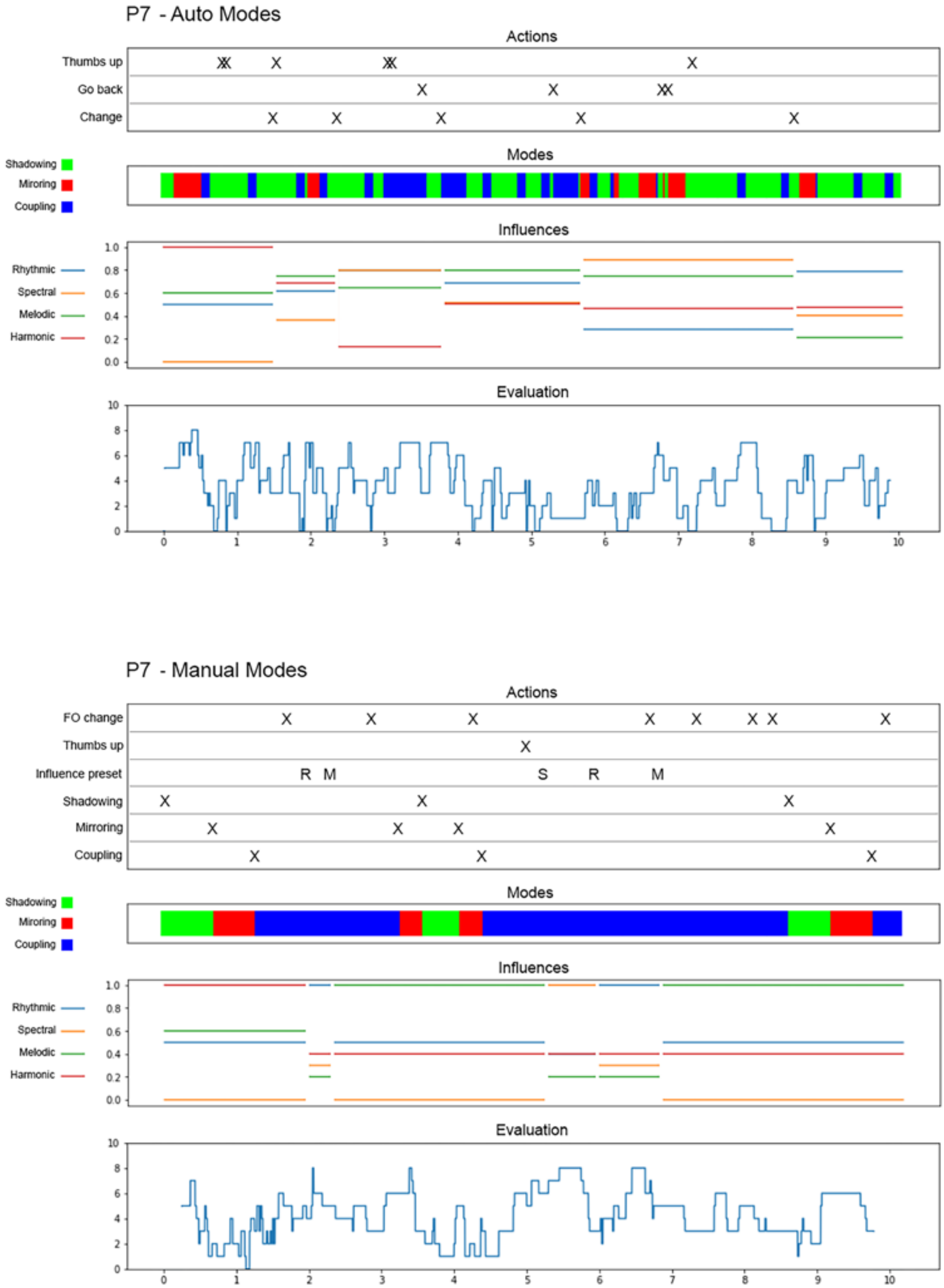
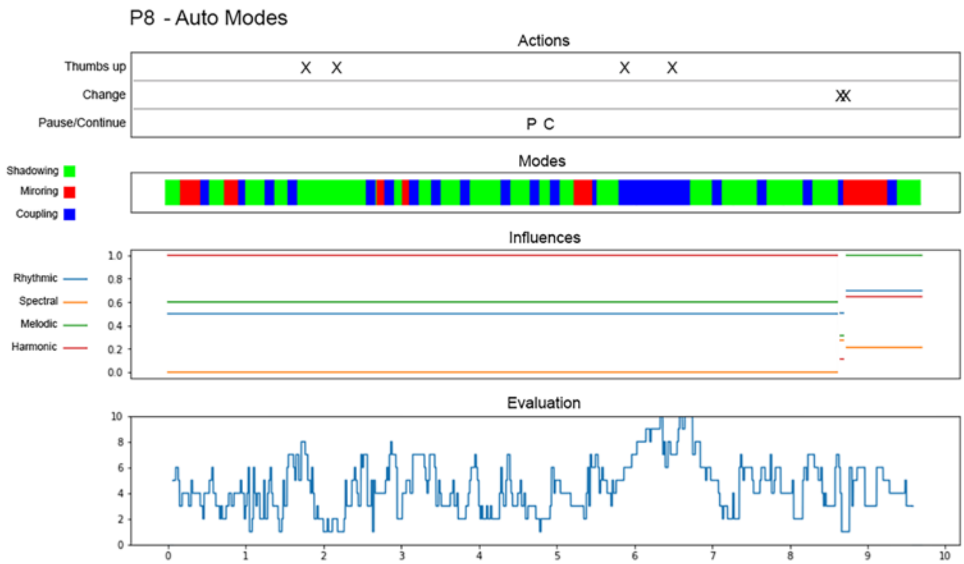


Figure 61. Interaction logs and evaluations for P7.



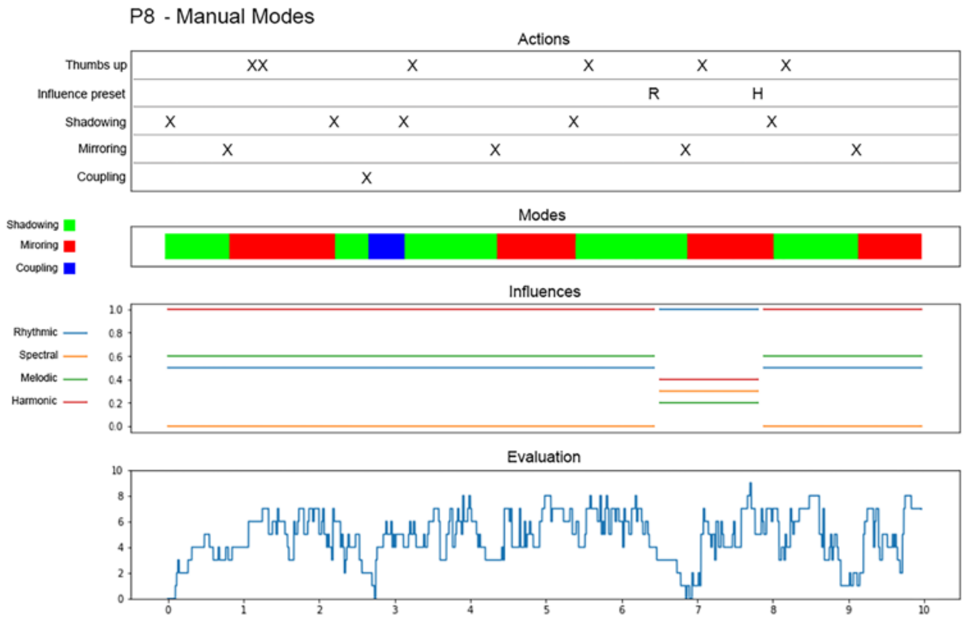


Figure 62. Interaction logs and evaluations for P8.

In Auto Modes, there were large individual differences in how the modes were activated. This, of course, depends on how well the participant's instrument and playing style matches the material in the corpus. Many of the sessions are characterized by quite short activation periods in coupling mode. As explained in Section 8.1.4, I had set a minimum activation period of 10 seconds for mirroring and coupling modes (unless mirroring trumps coupling or vice versa) in order to avoid an overly jittery response. For some participants, the activation criteria for coupling mode seems to have only been met intermittently, resulting in repeated triggering of the mode without it being sustained for more than 10 seconds at a time. Hence, some participants were justified in thinking that the system moved on too quickly. In total, the distribution among interactive modes for the Auto Modes prototype was:

- Shadowing: 55%
- Mirroring: 20%
- Coupling: 25%

In the sessions with the Manual Modes prototype, where the participants could choose which modes to be in, the distribution was:

- Shadowing: 40%
- Mirroring: 21%
- Coupling: 39%

To get an indication of how instantly gratifying the different modes were for the participants, the frequency of the use of the *Thumbs up* pedal was calculated as weighted percentages relative to the total amount of time the modes were active. In Auto Modes, these percentages were 39% for shadowing, 23% for mirroring, and 38% for coupling. In Manual Modes, the percentages were 41% for shadowing, 32% for mirroring, and 27% for coupling.

The evaluation sections are the results of the home evaluation assignments. I waited one week after each user study before sending participants links to the evaluation platform. The rationale for this was that I wanted their memories of the sessions and how they rated the creative experience to be weakened enough for them to listen to their interactions with the prototypes with fresh perspectives. As explained in Section 8.2.8, the evaluation scores were generated in real-time while the participants listened to the recordings and moved a scoring slider up and down. Figure 63 shows box plots comparing the combined evaluation scores for the Auto Modes and Manual Modes sessions for each participant.

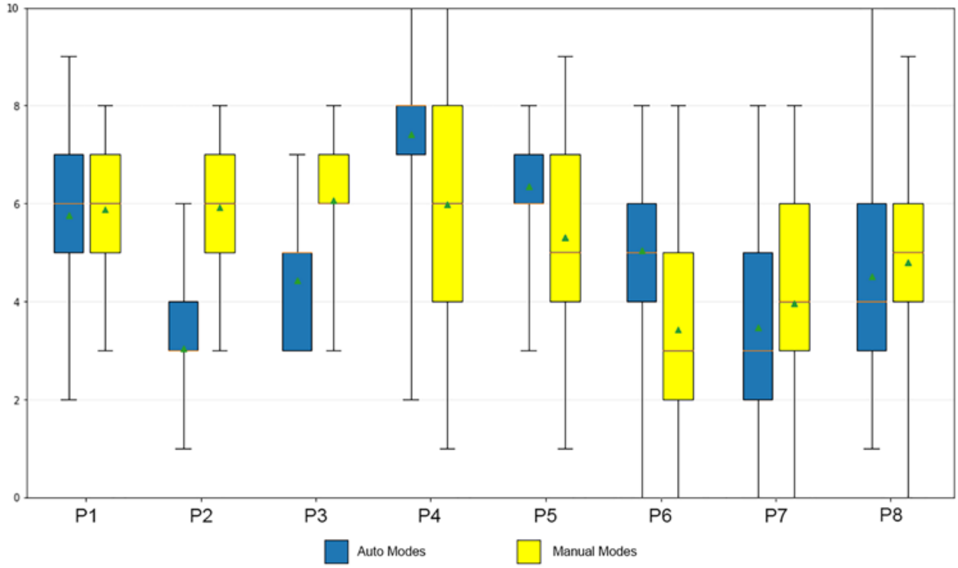


Figure 63. Box plots showing the combined evaluations scores for all participants.

In the box plots, the green triangles represent the mean scores, and the orange lines are the median scores. The boxes are the so-called interquartile range, i.e. "where most of the action is". The lines below and above the boxes represent the lower and upper 25% of the scores (technically, these are called the first and fourth quartiles). I will use the colloquial terms "worst parts" and "best parts" of the interactions for these ranges. Table 14 shows the distribution of which modes the system was in for these worst and best parts of the interaction for Auto Modes and Manual Modes respectively.

	Auto Modes			**Manual Modes**		
Mode	Shadowing	Mirroring	Coupling	Shadowing	Mirroring	Coupling
Best parts	47%	15%	38%	34%	12%	54%
Worst parts	63%	20%	17%	49%	17%	34%

Table 14. Distribution of modes for the best and worst parts of the interactions.

The interaction logs and evaluation scores represent rich sources of data with the potential of supporting much deeper analysis. Due to space considerations, I will limit myself to the results presented here and move on to discussing them in relation to the results from the statistical survey and interviews.

8.7 Discussion

The results from the statistical survey, interviews, interaction logs, and participant evaluations presented in the previous sections are examined here and linked to the study-specific hypothesis regarding the tradeoff between user control and system autonomy outlined in the introduction to the chapter. The working hypothesis was that relinquishing control and giving up a degree of agency to the system potentially could lead to a more creative experience on the user's part than when being able to control—and hence predict—most aspects of the system's behavior. In the statistical survey, this was reformulated as a null hypothesis claiming no difference in creative experience between the two prototypes Auto Modes and Manual Modes, against an alternative hypothesis that interacting with Auto Modes does lead to a more creative experience than interaction with Manual Modes. The broader implications of the results in relation to the main research question of this thesis and the theoretical underpinnings of all the four studies are reserved for Chapter 9.

8.7.1 First impressions, reflections, and later evaluations

The combined results of Study 4 are inconclusive, but at the same time both surprising and interesting. The statistical survey yielded a result where the null hypothesis was not rejected—there was no significant difference in the creative experience of participants interacting with Auto Modes and Manual Modes. In fact, the high p-value indicates an effect going the opposite direction of the alternative hypothesis. If I had performed a so-called right-tailed t-test with an alternative hypothesis predicting a more creative experience with Manual Modes, the p-value would have been 0.1—still non-significant but coming close to significant. Obviously, I had not anticipated such a tendency. Based on my prolonged period of experimenting with the Auto Modes version of the prototype, it seemed much more probable to me that the participants would either rate the prototypes equally or feel more creative with Auto Modes.

A second surprise was how much the participants changed their views of their interactions with the prototypes from the initial reactions to the evaluations a week later. As early as in the final interviews when comparing the prototypes, some of the participants were saying things that appeared at odds with the CSI scores—or at least much more nuanced than the scores would indicate. A week later, five of the participants scored the interactions in a manner that seemed to contradict their immediate reactions after the sessions. P4 and P5, whose CSI scores were quite evenly balanced between the prototypes but with a slight edge to Manual Modes, gave Auto Modes much higher scores while listening to the interactions a week later. Even more dramatically, P6 rated the interaction between himself and the Auto Modes prototype decisively higher than with Manual Modes, despite having said that he thought the

Auto Modes session had been like interacting with "an alien". An equally significant reversal in the opposite direction came from P8, who had given Manual Modes the lowest of all CSI scores and was in no doubt that he preferred Auto Modes in the final interview. However, the home evaluation shows overall higher scores for the interaction with Manual Modes. In fact, he commented on his change of opinion in the email containing the evaluation files:

> I think […] the program in Recording 1 (Manual Modes) is quite flexible, even though it appeared more random when I played with it. The fact that it can also identify rhythmical information makes it good at recreating the contours of melodies, even though not always with the right notes. After hearing the recordings I am actually not sure which program I would prefer as a song-writing tool—Recording 1 vs. Recording 2—even though I had a quite strong preference for Recording 2 that day.

P7 also expressed a change of opinion after hearing the recordings. Her CSI scores immediately following the sessions showed a strong preference for the creative experience with Manual Modes. When evaluating the recordings a week later, Auto Modes came closer to Manual Modes than she had expected. In the email, she commented on this and blamed it on fatigue: "Wow I feel like I can really hear my fatigue level rising through the session cause the first recording (Auto Modes) seemed to sound more in tune and interesting than the second recording."

Another good example of the inconclusive nature of the results came from P1, who in the final interview comparing the two prototypes quite clearly stated that he preferred Auto Modes, felt more creative with it, and thought it felt more like a partner than Manual Modes. Despite this, his CSI score for Manual Modes was slightly higher than for Auto Modes. When his listening evaluations came back a week later, they showed once more that he rated the sessions quite equally, with a slightly higher mean score for Manual Modes. A closer look at the scores, however, reveals that the Auto Modes session contained both the best and the worst parts of the interactions. In my personal experience, obtaining even short periods of peak performance is worth prolonged periods of forgettable performance. In his own words, P1 explained that Auto Modes felt more like a partner to him because he felt that he connected better with it than Manual Modes, which he jokingly claimed was "playing the virtuoso". Incidentally, P8's evaluation also had a higher best parts rating in Auto Modes, despite the mean being somewhat lower than in Manual Modes.

Drawing any conclusions from such diverging data is speculative at best. However, the data does indicate something very important that I failed to take into account: First impressions are unreliable. I had spent more than a year experimenting with these interactive modes, and had familiarized myself with how they work. The participants had five minutes with each prototype to get a sense of how they work, before delving into ten-minute improvised sessions that were audio recorded and filmed. Immediately after these sessions, they were handed questionnaires with questions related to how these systems aided them creatively. That is a lot to take in. Being confronted with new situations is challenging. Upon reflection, it is not at all surprising that some semblance of being in control is useful in situations like this.

The fact that more than half of the participants had drastically altered impressions of the creative sessions when listening to them later is a very useful finding. It indicates that they managed to be creative and produce musical results they themselves found interesting in circumstances they described as difficult or even uncomfortable. This is, in fact, promising. My working hypothesis was that non-control could be creatively productive. Non-control could also be confusing and uncomfortable in the moment. When asked to rate their own creative experience immediately after the sessions, it is reasonable to assume that any feeling of discomfort may dominate the emotional response. In Study 1, Catherine—an experienced improviser—claimed that she is "comfortable being in the uncomfortable spot". I think this statement captures an acceptance of non-control. To some extent, this sort of attitude could be an inherent part of some people's personality. However, I personally believe it is an attitude that most people can cultivate through repeated exposure to unknown situations. As such, repeated exposure to interactive music systems such as Spire Muse could be a good way to train oneself in "letting go". In Study 3, Lisa did make a claim to this effect when she claimed that playing with the "system" (the wizard) was "good practice being with someone" and "being taken to places I never would have gone".

Looking back at my first interactions with Spire Muse (which essentially is the same as the Auto Modes prototype), I remember being quite dissatisfied with how the system responded and by my own performances early on. As I made incremental changes to the algorithm, I thought the system began to respond better and that my own performances improved. In light of the results of this study showing this opinion shift in the space of a week based only on one encounter with the prototypes, I am uncertain whether my experience of gradual improvement of the system can be attributed solely to technical changes in the algorithm. Perhaps I changed as well. Perhaps I grew accustomed to a new form of interaction—a form I found increasingly rewarding in a creative sense. In short, perhaps Auto Modes is an acquired taste. In future studies, it would therefore make sense to conduct longitudinal studies of musicians using Auto Modes and Manual Modes, and examine the development of their assessment over time.

8.7.2 Delegating control

P2 and P3 were the only two participants whose CSI scores, interview answers, and listening evaluations all decisively pointed in the same direction—toward Manual Modes. They both complained that Auto Modes had moved on too quickly once it started to get interesting. Their Auto Modes session histories show frequent, but short forays into coupling mode before snapping back to shadowing. P2 described that he was trying to "pull back" the system (curiously, he did not use the Go back pedal at all, but used the *Thumbs up* more than any other participant). Both P2 and P3 explained that they preferred the ability to control when to switch interactive modes. Judging by their Manual Modes session, they preferred prolonged periods of coupling mode interaction. Interestingly, coupling in Manual Modes is the only case where the user's musical input has no bearing on the system's output at all—machine listening is ignored. P2 did not even initiate any FO buffer changes while in coupling, and P3 only did so twice in seven minutes. They were, in fact, playing together with a non-listening automaton for most of their Manual Modes sessions, yet these are the parts that generated the highest evaluation scores.

This observation led me to further examine the Manual Modes session of P7—the third participant who claimed a preference for being able to control the interactive modes directly. She also chose to stay in coupling mode most of the time. Most of the *Thumbs up* indications occurred in these sequences, and they also correspond with many of the best parts of the evaluation scores. From the interviews, we also know that P1 preferred the coupling mode parts of the Manual Modes sessions. Based on these observations, we can conclude that four of the participants preferred Manual Modes because they could *choose* to let the system be autonomous and take the lead. Comments made by P2, P3 and P7 also indicate that the problems they experienced with Auto Modes was that the instances of coupling mode were not sustained for long enough periods for them to develop a coherent musical sequence. Meanwhile, some participants, most notably P4 and P5, cherished the unpredictable facets of Auto Modes. P5 emphasized that the struggle of "battling for the direction" was precisely why he enjoyed playing with Auto Modes. I assume that his description of Auto Modes as "uncanny" in a positive sense related to this prototype's capacity to engage in such a "battle of wills".

Based on these deliberations, I am forced to admit a weakness in the premise of my experiment. I have stringently classified Auto Modes as the more autonomous and less controllable prototype, and Manual Modes as the more controllable one. However, we now know that many of the participants preferred Manual Modes because they could choose to let the system stay in the more autonomous mode. Essentially, they were delegating control. A consequence of this realization is that I cannot place much value in the results of the statistical survey. On the other hand, the interviews, interaction logs, and participant evaluations have provided findings that make me optimistic about future studies. I interpret them as indicative of support for the benefits of a mixed-initiative approach to designing interactive music system. However, better evaluation methods need to be developed.

8.7.3 Reliability issues

From the outset, I knew that the prototypes would respond somewhat differently depending on the instrument. I had taken care to exclude percussive instruments, which have little or no pitched harmonic spectra and would cause quite random responses. I also pointed out that the prototypes would respond better in the mid to high registers, as explained earlier in this chapter. With these measures, I hoped that the differences in machine listening would be within an acceptable range. Unfortunately, the machine listening algorithm did perform quite poorly with some of the instruments. P6 commented that he thought the system performed "out of tune and out of style". This was a fully justified comment, as the system responded significantly worse with piano as input. I have later concluded that the attack of the piano is too chaotic to be captured by a normal microphone—I should have used a piezo microphone. The system response was noticeably more erratic in P6's sessions. He also claimed that the system performed better with rhythmic and melodic presets. This also makes sense given the chaotic spectral content in the piano input. The rhythmic preset focuses on the timing of onsets and disregards timbre and pitch, while the melodic preset focuses mostly on the fundamental frequencies. Curiously, however, in the evaluation of the Manual Modes session, the rating drops to zero immediately following his selection of the rhythmical influence setting and stays very low until he selects the melodic mode preset. This indicates that something that is experienced as instantly gratifying does not necessarily translate to good or creative interaction.

In Section 8.3, I included the CSI results of the guitar players as a subgroup to make an important point. All participants played with the same corpus. Because this was based on acoustic guitar, it is reasonable to assume that the guitar players would have input instances that overall had closer matches in the corpus. Consequently, we can also assume that the chroma transition matrices worked more accurately than with instruments that produced poorer matches. I have no means to document such a correlation, but judging by the accounts of P4, P5, and P8, they were quite satisfied with the musical relevance of the system's responses in general. P2, however, was not impressed by the system's autonomous choices. In future tests, a way to increase reliability would be to invite only one instrument group.

Another source of unreliability is the impact of the influences. I deemed influence settings to be important enough to include despite acknowledging that they may affect test reliability. For example, some participants may have been fortunate to find influence settings that triggered good responses early enough to develop a constructive interaction, whereas others may have been less fortunate. Visually, there are several interesting correlations between changes in the influence settings and large shifts in the evaluation scores. This is purely speculative, but mentioned as an acknowledgement of a reliability issue. Additionally, there was at least one case where a participant was seemingly unaware of the option to change influence settings when in Auto Modes. As mentioned, P7 described looking at the influence settings and trying to adapt to what she saw there, while apparently hoping for "harmonic high" to match her voice. In my view, leaving the influence weights completely static or hidden would have increased reliability. Notwithstanding, it would have done so to the detriment of the overall interactive experience. Therefore, I am not convinced it would have improved the experiment.

8.8 Summary

This chapter has focused on the development and the evaluation of Spire Muse—the mixed-initiative interactive music system I have developed on the basis of the theoretical framework and the first three studies of this thesis. Starting with a working hypothesis claiming that relinquishing control and giving more agency to the system could lead to more creative experiences, I devised a comparative study with two prototypes to test this hypothesis. Both of the prototypes were implemented with three interactive modes—shadowing, mirroring, and coupling—featuring behaviors ranging from reactive to proactive. Auto Modes was designed as the more autonomous prototype that switches between modes based on what it "hears" in the user's input. Manual Modes, on the other hand, leaves it up to users to select interactive modes directly.

Eight participants were recruited to the study, which consisted of several parts. After being introduced to the prototypes, they performed with each of them for ten minutes. Following each session, they filled out questionnaires and were interviewed about the experience. At the end of the study, they were asked to compare the creative experiences with the prototypes. Finally, a week following the study, they received links to a web application where they could rate audio recordings of their performances using a score slider.

The results were analyzed using several different methods. In a statistical survey, a null hypothesis claiming no difference in the creative experience between the two prototypes was tested against an alternative hypothesis proposing that interacting with Auto Modes does lead to a more creative experience than interacting with Manual Modes. The null hypothesis was tested using the mean Creativity Support Index (CSI) score as the dependent variable in a paired t-test. The null hypothesis was not rejected. In fact, the results of the test showed an effect tending toward the opposite of the alternative hypothesis—only one out of the eight participants gave Auto Modes a higher CSI score than Manual Modes. The interviews revealed many more nuances. For example, half of the participants deemed Auto Modes to be more like a musical partner than Manual Modes.

Surprisingly, the web-based evaluations showed that many of the participants had radically different impressions of the performances one week later. The results were still inconclusive. Some participants gave Auto Modes much higher scores than in the post-session survey, while the participant who had preferred the creative experience with Auto Modes in the first place was no longer certain which one he preferred as a creative tool. A closer examination of the interaction logs revealed that most of the participants who stated a preference for the Manual Modes prototype did so because they could choose when to delegate control to the system. In fact, most of the highest evaluation scores occurred when coupling mode was selected, which in Manual Modes is a non-listening automaton. This discovery lead me to conclude that there was a weakness in the premise of the statistical survey, because the majority of the participants who preferred Manual Modes—the purported more controllable prototype—did so because they could choose to delegate control to the system for more sustained periods. In Auto Modes, on the other hand, the proactive coupling mode tended to be activated for shorter periods. Overall, the combined results show that there is substance to the original working hypothesis that delegating agency to the system may lead to more creative experiences. New methods need to be devised to examine this further.

9 Discussion

In this chapter, I will discuss implications of the findings from the four studies presented in Chapters 5–8 within the framework of addressing the research questions of this thesis. I will do so by drawing on the theoretical underpinnings presented in Chapters 2–4. I start the chapter with a short summary of the findings from the four studies, so readers can have these readily available as a quick reference. The main research question asked how a mixed-initiative interactive music system could aid human musicians in the initial ideation stage of music making. Although the challenge posed herein has a technological element, I argued that a deeper understanding of the human activity of collective musical interaction is crucial to be in a position to take on this challenge adequately. For this reason, I resume by first reflecting on the vantage point of dynamical systems theory as my approach to the sub-questions related to understanding the unfolding dynamics between musicians engaged in the activity of creative musical interaction, and to which extent such an understanding is transferable to modeling an interactive music system. Finally, the main research question is discussed in light of findings from the studies, and in relation to the concepts of agency and creativity introduced in Chapter 2. After each section treating the research questions, I conclude with concise answers as a summary of the preceding discussions.

9.1 Summary of findings

Study 1 offered a comprehensive view of the initial, exploratory stages of collective music making as experienced by seven musicians and composers from a range of different genres. I found that the following attitudes in general may increase the likelihood of generating emergent novelty in collective music making: 1) maintaining a process-oriented approach where goals are deferred in favor of "going with the flow", 2) attaining shared ownership by decentering and reaching a collective subjectivity, and 3) acutely listening for semantic content in the musical signal. All of these attitudes revolve around an active search for a co-creative middle ground between the aesthetic preferences of the individuals in the collective. I concluded that this would call for an interactive music system with the appearance of a "will of its own"—a contrasting space from which it can diverge and converge to its human counterpart.

Study 2 examined the converging and diverging dynamics of two musicians engaged in an improvised music making session, where the musicians were placed in separate rooms as a measure to minimize visual cues and simulate the facelessness of interacting with a computer. The analysis showed that the musicians were able to converge to collective sequences and articulate transitions to new ones through strategies of "pulling together" and "pushing apart". Specifically, Canonne and Garnier's (2011; 2012) dynamical systems model for collective free improvisation was used to perform this analysis. Consequently, the notions of converging and diverging became an important part of the design rationale for the prototypes in Studies 3 and 4. A quantitative analysis of the interaction led to the discovery of a windowing technique using chroma transition matrices as a promising method for detecting harmonic dynamics on the meso time scale.

Study 3 provided invaluable insights into how three musicians experienced interacting with what they thought was a computational musical agent through an interface prototype. Behind the scenes, the agent was actually a human musician pretending to be a computer, and who was instructed to engage in four interactive behaviors—shadowing, mirroring, coupling, and negotiation. In post-session interviews, the participating musicians expressed a clear preference for the system's more contrasting or oppositional behaviors. They claimed to have taken a different attitude to musical interaction than they normally do with human musicians. They had lower expectations to the musicality of the system, and therefore felt less constrained by musical rules. A perceived freedom from judgement allowed two out of the three musicians to feel less self-conscious about their own performance. The carefree attitudes the users were able to assume concur with two of the attitudes identified as engendering emergent novelty in Study 1—a process-oriented approach and sensibility to a collective subjectivity. Thus, repeated exposure to mixed-initiative music making could be good practice to habituate such attitudes in general.

In Study 4, I hypothesized that relinquishing control and giving more agency to the computational agent in an interactive music system may lead to a more creative experience. A comparative study with two prototypes was devised to test this hypothesis, with one prototype (Auto Modes) designed to be more autonomous than the other (Manual Modes). Both prototypes featured the interactive modes shadowing, mirroring, and coupling, with behaviors ranging from reactive to proactive. The Auto Modes prototype switched between these modes automatically, whereas participants could select the modes directly with the Manual Modes prototype. The results of a statistical survey, interviews, interaction logs, and evaluations by the participants were inconclusive. In general, the participants first thought interacting with Manual Modes was somewhat more creative, but the statistical survey showed no significant difference. However, the evaluations showed a large shift of opinions a week later, with three participants now showing a strong preference for their interactions with Auto Modes. By seeing the interaction logs and evaluations together, it also became clear that the participants who did prefer Manual Modes tended to stay in coupling mode, in which the computational agent is completely autonomous. Thus, in spite of the inconclusive results, there is substance to the original working hypothesis that delegating agency to the system may lead to more creative experiences. New methods need to be devised to examine this further.

9.2 Reflections on the dynamical systems approach

In Chapter 1.2, I suggested that the main challenge underpinning the research question is not technological, but human. Although extensively researched, the dynamics between human musicians when engaged in creative musical interaction remains poorly understood. To understand how a mixed-initiative interactive music system can help human musicians, it is useful to have a comprehensive understanding of how human musicians interact creatively, and how such interactions tend to evolve over time. Therefore, I posed the following sub-questions:

- *What can be learned about the interaction dynamics between musicians in the ideation stage of collective music making?*
- *To which degree can these interaction dynamics serve as a model for an interactive music system?*

I chose dynamical systems theory (DST) as a conceptual tool to navigate between these human-oriented sub-questions and the more technologically oriented main research question. I proposed that the framework could function as a metaphor for the high-level dynamics emerging from a complex system of interactions including both physical and psychological phenomena. In Chapter 8.1.1, I introduced a difference between using conceptual frameworks to *understand practice* and *inform design*. This was, in fact, essentially a preview of my answers to the questions posed above, and which I will treat more thoroughly in the following.

Early on in this thesis, I devoted an entire chapter (Chapter 3) to DST. I argued that using DST as a conceptual framework could be useful, because it provides a set of interlinked concepts related to development or change with strong metaphorical parallels to music making. While acknowledging the figurative status of DST applied in a different domain, "many of the affinities and oppositions are carried along in the transfer of meaning" (Kittay, 1987, p. 154), and could offer a holistic view that may be lacking in other conceptual frameworks.

Astute readers may have noted that DST has gradually receded into the background throughout the four studies presented in Chapters 5–8. I was acutely aware of this development, and it is a matter of fact that DST ended up being applied less explicitly in the design of the prototypes than I had anticipated at the outset of the development phase. For example, no DST methods were embedded in the software algorithms. However, the RtD approach and a methodological framework of triangulation between theory, observations and design was adopted precisely because it assumes that theory is malleable. As I pointed out in Chapter 4 when reviewing Garcia's (2014) PhD thesis, the triangulation framework is useful because theoretical revisions are driven by the empirical findings in the iterative design process. This approach facilitates the generation of knowledge that is integrated with the specialized practices of the musicians participating in the studies. Therefore, offering readers an insight into how I started out with a broad focus on a large theoretical structure and gradually chiseled out more narrow derivations of this vantage point is an appropriate way to convey how I have worked throughout the thesis. My methodological framework and the software prototypes did not appear out of thin air. There is no doubt that DST did have a decisive impact on both the design of the first three studies and on the conceptual framework for the prototypes used in the fourth study.

As explained in Chapter 8, DST has influenced the development of the prototypes used in Study 4 in two indirect ways. As a discovery representation, DST was used as a *design rationale* for the prototypes. As an application representation, a DST approach led me to a collection of *user-interaction scenarios* that were implemented in the software (J. M. Carroll, 1990, p. 323). In Study 2, the convergence vs. divergence framework was used to understand the practice of collective music making. As explained in Chapter 6, I performed an analysis of two musicians engaged in a collective music making session using the work of Canonne and Garnier (2011; 2012), who developed a dynamical systems model of collective free improvisation and later elaborated upon this model in an ethnographical study of improvising musicians. In Study 3, I applied an interactive behaviors framework with four interactive behaviors—shadowing, mirroring, coupling, and negotiation (Blackwell et al., 2012)—as a guide for the keyboard player in a Wizard of Oz study to simulate a computational agent engaging in shifting interaction dynamics. Three of these behaviors (shadowing, mirroring, and coupling) were later implemented as interactive modes in the prototypes used in Study 4, where negotiation was defined as emergent from the interface between the human user and the computational agent. The interactive behaviors framework was thus used to inform the design of the prototypes. The main reason behind my choice of the interactive behaviors framework was the promotion by Blackwell et al. (2012) of a dynamical systems approach to implement the interactive behaviors in question. Although I ended up not using a dynamical systems approach directly in the design, it was my awareness of a potential DST approach that led me to adopt the interactive behaviors framework in the first place.

It is safe to claim that I would not have ended up with the same software if it were not for my orientation toward DST. The frameworks derived from my DST orientation provided me with a language to formulate the requirements for a mixed-initiative interactive music system. The goal of simulating the "push and pull" of interactive dynamics in collective music making crystallized during my studies of DST. In short, this would have been a very different thesis in the absence of DST as a theoretical vantage point.

However, I will not argue that DST is the best or the only way to approach the research question of this thesis. I found DST methods to be very challenging to implement in practice when building the software. For example, I spent several months experimenting with methods such as calculating the fractal dimension of musical signals to detect variances in complexity, and delay coordinate embedding to detect possible attractors present in a given input. Without going into further detail about these attempts, my conclusions were that these methods did not translate very well to what I perceived were musically significant events on time scales relevant for the development of musical form. Music is not a purely physical event—it is psychophysical. Ultimately, heuristic approaches such as the use of chroma transition matrices and the statistical approach to system-initiated changes between interactive modes described in Chapter 8.1.4 turned out to be more productive, and this was arrived at after a prolonged period of experimentation. While DST helped me in formulating problems well, it was more difficult to design solutions with it.

Circling back to this thesis' sub-questions, I can now offer two concise answers based on the above discussion:

What can be learned about the interaction dynamics between musicians in the ideation stage of collective music making?

I was able to gain a comprehensive understanding of interaction dynamics between musicians engaged in collective music making through the lens of DST. A conceptual framework derived from DST, based on the work of Canonne and Garnier (2011; 2012), focused particularly on how musicians converge to collective sequences and articulate transitions to new ones by engaging in different strategies of diverging and exploring the phase space in search of new attractors. This is, admittedly, a very narrow use of DST to understand one aspect of collective music making. This narrowing down was necessary in order to make the project feasible within the scope of this PhD. However, DST is a rich framework to apply in different ways that have inter-coherence due to their common source.

To which degree can these interaction dynamics serve as a model for an interactive music system?

During the prolonged period of developing the prototypes used in Study 4, I gradually moved away from using DST to model interaction dynamics. While the metaphorical remapping of DST concepts onto the target field of collective musical interaction was an effective way to understand the interaction dynamics involved, I found DST methods were difficult to apply as a modeling approach to design. Instead, I used a conceptual framework of interactive behaviors ranging from reactive to proactive that could be more loosely associated with the converging and diverging strategies observed in the human domain of collective musical interaction. I used heuristic approaches to arrive at machine listening algorithms that would result in a type of decision-making that I thought were representative of the interaction dynamics I had observed in the preceding studies.

9.3 The benefits of a mixed-initiative interactive music system

The prototypes used in Study 4 are in a sense also answers to the two sub-questions discussed in the previous section. In reference to the theorylike role of artifacts discussed in Chapter 4, the answers are embedded as "thing knowledge" (Baird, 2004), as "theory nexuses" (J. M. Carroll & Kellogg, 1989), or as "epistemological tools" (Magnusson, 2009). As such, the prototypes became the main tools used to address the main research question of this thesis:

- *How can a mixed-initiative interactive music system aid human musicians in the initial ideation stage of music making?*

Using two different prototypes in a study that examined the tradeoff between user control and system autonomy allowed me to gain different perspectives on the importance of a mixed-initiative interface. Auto Modes was designed as the "most mixed-initiative" prototype where the musical agent is capable of taking the initiative autonomously, whereas in Manual Modes, the user is in control and delegates the initiative when that is desirable. In the following discussion, I will attempt to tie the benefits of mixed-initiative interaction back to the concepts of agency and creativity presented in Chapter 2.

9.3.1 A creative blind date

In the discussion section of Chapter 7, I touched upon the notion of otherness, and referred to Levinas (1996) when suggesting that an ingrained suppression of otherness may come in the way of developing a musical dialog. To see an other is like a leap of faith, with a perceived risk of losing the individual self. It "requires that I recognize the other as having a kind of claim on me" (Benson, 2003, p. 167) so that it "breaks into my ego-centeredness and gives me something to understand" (Gadamer, 1997, p. 46). It is hard to see an other when focused on preserving oneself. One of the findings in Study 3 was that the nonhuman agency of the "machine" did appear to have a mellowing effect on the sense of self-preservation, and guards were let down. Mixed-initiative interactive music systems can offer users opportunities to engage musically with an other that does not pass judgement on their aesthetical choices. While the system's choices in turn sometimes may appear de-aestheticized or alien, the freedom from judgement affords a type of interaction that may reduce social inhibitions that often arise between human musicians. Findings from both Studies 3 and 4 suggest that human musicians may surprise themselves if they accept "machine aesthetics" as a feature and see where this takes them musically. Several study participants discovered that there turned out to be some method to the apparent madness when they adapted to what they heard and attempted to go along with it. It became a "dance of agency" (Pickering, 1995) where "material and human agencies are mutually and emergently productive of one another" (Pickering, 1993, p. 567).

Part of the reason why I believed that Auto Modes would feel more creative than Manual Modes was because if the user were in charge of the musical agent's decision-making, I thought this sense of an other would fail to materialize. As explained, I had grown accustomed to the unexpected twists and turns of Auto Modes. I appreciated how it provided me with challenges in the form of a "bisociation of previously unrelated matrices" (Koestler, 1964). However, I had not sufficiently taken into account that coupling mode in Manual Modes may also provide a sense of otherness, and the fact that this mode was so popular is a testament to the participants' willingness to surrender to a collective subjectivity—in this case to the agency of the Factor Oracle renditions of the corpus. In hindsight, I realize that both the Auto Modes and Manual Modes prototypes have agencies of the conditional kind (the capacity to produce unintended effects) and the delegated kind (the capacity to realize intentions delegated to them by somebody or something else) (Kaptelinin & Nardi, 2006).

In Study 3, Lisa claimed to have been "taken places" she never would have gone with a human. This statement speaks volumes about the psychological challenges musicians face each time they expose themselves to other human musicians. As mentioned, there is an element of risk involved in improvised co-performance, and hence it takes courage to thrive creatively in the face of such risk. The fact that Lisa's claim was made before I revealed that she had, in fact, been playing with a human makes it even more intriguing. She described the session as "good to practice being with someone". Apparently, the experience gave her the opportunity to explore her own creativity in relation to an unknown agency. In Study 4, P7 expressed similar sentiments when explaining that she felt more creative in the Auto Modes session "because maybe it pushed me into places where I wouldn't normally go". In contrast to Lisa, P7 had felt this process as a struggle, and was quite dissatisfied with the session. Upon hearing it again a week later, however, she was more positive and described her own responses as interesting. In terms of Boden's (1990) creativity categories (combinatorial, exploratory, and transformational), the above accounts are testimonies of exploratory creativity—the traversal of a conceptual space. Colloquially speaking, they were performing "outside of their comfort zones" and pushing boundaries for what they deemed acceptable. According to Lisa, the lack of a human counterpart was what allowed her to traverse as far as she did.

These and several other findings from Studies 3 and 4 indicate that playing with a mixed-initiative interactive music system could be described as a "creative blind date". The anonymity provided by the faceless system afforded different kinds of interaction than they were used to with humans. From such a perspective, the musical outcome may be subordinate to the activity of creative discovery and developing a sense of trust in venturing into the musical unknown. When playing with a nonhuman, the notion of failing loses potency. There is no right and wrong—no one is there to care. This does not mean that the interactive music system could be anything, or that musicality loses importance. The function of an artifact as a "theory nexus" (J. M. Carroll & Kellogg, 1989) or "epistemological tool" (Magnusson, 2009) becomes very important. The designer is a latent actor, like a genie in a lamp, waiting to materialize. In a sense, the participants of Study 4 were interacting with me, and with the encoded dynamical system I had spent so much time creating. I am loathe to admit that I felt a sense of pride when P1 described getting "really into the system", when P4 and P5 claimed that there were sequences where they experienced it was like playing with human musicians, and when P8 sensed that the system could "figure out" the harmonic framework and put together "quite coherent ideas". The participants were also interacting with the agency of the musicians represented in the corpus (Xi et al., 2018). In short, the notion of collective agency was reified in the prototypes, and the premise loftily stated in the title of this thesis was substantiated in practice.

Study 1 highlighted the importance of maintaining a process-oriented approach and a willingness to lose ownership of one's ideas when involved in creative collective music making. I argued that these attitudes could engender emergent novelty—a concept introduced by Sawyer (1999) as the type of idea generation associated with group improvisation. Study 3 demonstrated that mixed-initiative interactive music systems may be particularly well suited to promote such attitudes in the user. Anonymous human-computer co-creative spaces afford both freedom from judgement and freedom to explore, and goals and intellectual property are notions that recede into the background. Acceptance of nonhuman agency and willingness to cede control to this agency is transferable to the practice of decentering (Guattari, 1995; T. Davis, 2011) and finding a common aesthetic ground with other humans. Repeated exposure to such interactions—whether with human or computational agents—may aid musicians in their music making pursuits in general. It may offer different forms of resistance, help musicians break out of habits and promote spontaneity. None of the interactive music systems that exist to date, including the prototypes developed in this thesis, are anywhere close to being human-like. Nor do they need to be. The nonhuman factor is precisely why users may more easily ignore their own egos and focus on what music could be.

9.3.2 Developing a relationship

In 2002, Ben Schneiderman proclaimed "the old computing is about what computers can do, the new computing is about what people can do" (Shneiderman, 2002, p. 2). This thesis has focused on what computers and humans can do together, based on a firm belief that interactive music systems can be creative partners as opposed to mere tools. Specifically, I have committed myself to understanding how initiative taking between people works in the domain of music making, and attempted to design software that can make choices autonomously and sometimes take the creative initiative. Study 3 demonstrated that musicians involved preferred when the "system" engaged in contrasting behaviors on its own initiative, and regretted the fact that the system followed what they were doing too much. Study 4 also showed that the participants generally liked that the system engaged in autonomous behaviors. However, some of the participants were clearly not satisfied with the pace of turn taking in Auto Modes, where the frequent coming and going of coupling mode by all accounts seems to have been the most problematic issue.

Clearly, the decision-making algorithm in the Auto Modes prototype is not as sophisticated as the "algorithm" of the human brain—even a brain that is trying to simulate the relative crudeness of a digital computer. However, the fact that the decision-making algorithm did make choices that pleased some of the participants is promising. It is also encouraging to see that several participants were more accepting of the interaction with Auto Modes upon listening to the recordings later. Several researchers that I have referred to in earlier chapters have pointed out the automation of high-level control as the largest challenge in the development of digital musical instruments or interactive music systems (e.g. Magnusson, 2009; Eigenfeldt, 2014; Martin, 2016). Based on what I have come across in the literature, a lot of the research in this area seems focused on making computational decision-making as human-like as possible. After working with Spire Muse and analyzing the results of Study 4, I believe such a focus is one-sided. Humans need to learn to play with machines as well. George Lewis is still performing with his Voyager system, which he built almost four decades ago (Lewis, 2000). In my view, it is still one of the most artistically impressive mixed-initiative interactive music systems. Although the software algorithm in the system is complex, it is not particularly advanced by today's standards. There is no machine learning involved. The most advanced aspect is the quality of the interaction between Lewis and the Voyager system. Together, they have developed a musical style that sounds interesting and vital. George Lewis is a phenomenal musician who has learned how to play well with a machine. In my view, this is the main reason behind the longevity of the project.

The most profound lesson I have learned from Study 4 is that it takes time to develop a sensibility for interacting with a new form of agency. What may sound alien at first may actually be worth revisiting. The disparity between what is deemed acceptable in the short and long terms could be tied back to the concepts of performative and memetic agency (Bown et al., 2009). As explained in Chapter 2, performative agency refers to the here and now of musical performance, whereas memetic agency refers to the influences of software on musical styles over historical time. To continue the blind date metaphor, it may be wise to not let first impressions ruin what could turn out to be a relationship. If "there is something there", as P6 put it after playing with Auto Modes, that something may grow. It may not be "the one", but you never know unless you try. Metaphors aside, Martin (2016) called for qualitative longitudinal evaluations that go beyond first impressions to capture the evolving perspectives of musicians playing with digital musical instruments. This approach would be useful for Spire Muse as well.

After the user study, I have made one major modification to the software. Because some participants preferred Auto Modes and others preferred Manual Modes, I decided to package them together into one. Now, pressing the tabulator key on the computer keyboard will cause the system to switch between the two interfaces. Thus, it is now possible to "take the reins" by switching to Manual Modes, and "let go" by switching to Auto Modes. Personally, I have found this to be extremely useful in further experimentation. As a now advanced user of the system, I find that Manual Modes works best when working with new corpora, while Auto Modes tends to be more exciting and creative when I have learned how the system tends to interact based on a given corpus. This further bespeaks my ignorance in assuming that new users would immediately take to Auto Modes as the most creative prototype.

I will conclude this section as I did in Section 9.2, this time by offering a short answer to the main research question based on the preceding discussion:

How can a mixed-initiative interactive music system aid human musicians in the initial ideation stage of music making?

A mixed-initiative interactive music system offers musicians freedom from judgement and freedom to explore their own creativity in relation to an unknown agency. Social factors make these kinds of freedom difficult to attain with other musicians. Hence, playing with interactive music systems can lead to different kinds of musical interaction than can be achieved between people. An acceptance of machine aesthetics may lead to surprising creative results. Repeated exposure to mixed-initiative interactive music systems could help cultivate attitudes that are valuable for collective music making in general, such as maintaining a process-oriented approach and accepting the loss of idea ownership.

9.4 Summary

In this chapter, I addressed the research questions of this thesis in light of theory and findings. I began by presenting a summary of findings from the four studies presented in Chapters 5–8. The sub-questions of this thesis were discussed in the context of reflecting on how my theoretical vantage point of dynamical systems theory developed and narrowed down during the course of Studies 1–4. Although I ended up not using specific methods from DST in the software itself, the framework had a large influence on the first three studies and on the design of the prototypes used in Studies 3 and 4. As a discovery representation, DST was used as a *design rationale* for the prototypes. As an application representation, a DST approach led me to a collection of *user-interaction scenarios* that were implemented in the software. Finally, I discussed the implications of the studies' findings as a way to address the thesis' main research question. Playing with a mixed-initiative interactive music system could be described as a "creative blind date", where the musical outcome may be subordinate to creative discovery and exploring new forms of musical expression. Study 4 demonstrated that first impressions are unreliable. Potential users may need time to develop relationships with mixed-initiative interactive music systems and grow accustomed to the new forms of interaction afforded by them. Therefore, future research should focus on the long-term development of the relationship between musicians and mixed-initiative interactive music systems.

10 Conclusion

In this thesis, Research through Design (RtD) was introduced as a design-oriented approach to knowledge production. Zimmerman et al. (2010) suggest that RtD tends to contribute to so-called *nascent theory*, which may highlight new sets of relationships between phenomena that require more rigorous research before maturing into established theory. This is the case for this thesis as well. An activity-centered design approach, which I adopted from Waern and Back (2017), has maintained a focus on the *activity* of mixed-initiative music making as opposed to *artifacts*. Hence, the theoretical findings—discussed in the previous chapter—are also oriented toward this activity, whereas artifacts are presented in this chapter as methodological or technological contributions. Further, I will review some of the limitations of the methodology and study design, and present some areas for future work.

10.1 Contributions

Chroma transition matrix windowing technique
In Chapter 6.1.3, I described a technique for following harmonic development by extracting chroma transition matrices from audio slices over meso time scale windows. This proved to be an efficient way to find harmonically appropriate sequences in the corpus based on what is happening in the input. One participant in Study 4 specifically commented that he sensed the system was listening and responding with solos that were appropriate given what he was playing.

Wizard of Oz method for testing interactive music systems
In Chapter 7.1, I presented an elaborate ploy to make study participants believe they were interacting with a computational musical agent. Although the Wizard of Oz method is well known in HCI, I have not come across it being used to test musical response types in interactive music systems. I found this to be very helpful in understanding how users may potentially react to the envisioned interactive behaviors before beginning to design the system architecture. I think the fact that the participants of Studies 3 and 4 expressed some of the same sentiments in reaction to the interactive behaviors suggests that the user-interaction scenarios were successfully implemented in the final prototypes. The findings from Study 3 certainly contributed to this accomplishment.

Web application for real-time evaluation of audio recordings
In Chapter 8.2.8, I described how I commissioned the design of a web application that enabled study participants to easily rate their interactions with the prototypes they had played with. While I cannot be certain that such a tool does not already exist, I could not find any that were as straightforward as the one I used in Study 4. The fact that all of the participants went through with the listening home assignments—a total of 160 minutes of audio rated second by second—suggests that the tool was efficient and user-friendly. The data it produced was extremely useful. For example, I would never have known which parts of the interaction the participants actually liked and disliked without this data, and I would not have been able to conclude that coupling mode tended to be more popular.

A mixed-initiative interactive music system
The development of Spire Muse is described in Chapter 8.1. After Study 4, I integrated Manual Modes with Auto Modes, and it is now possible to switch between these interfaces by pressing the tabulator key on the computer keyboard. While I consider Spire Muse a work in progress, it is a fully functional interactive music system, and it has now also been used live in an improvisation concert. The Spire Muse software is available on GitHub: http://www.github.com/sirnotto/SpireMuse.

10.2 Limitations and future work

There were two problems with the design of Study 2, described in Chapter 6.1. The first problem was that the Sonic Incident technique did not work as well as I had hoped. I took notes of what I deemed were important transitions during the improvised session between the participating musicians. In the interview, I played back these sequences and let them describe how they remember thinking about the performance during these transitions. While they gave some interesting comments, I later realized that I might have misjudged what were the most crucial transitions in the interaction. Thankfully, I was able to get much more useful data by requesting the participants to listen to the recording and sending me a list of what they thought were the most interesting transitions in the session, together with short comments on why they regarded them as qualitative transitions. The second problem was that the quantitative analysis using autoencoding and clustering as described in 6.1.3 was most likely methodologically flawed. I decided to present the results despite my insecurities about this, because the resulting graphs did show some consistencies in tracking transitions between sequences. I concluded that this was most likely due to the quality of the underlying data—the chroma transition matrices.

In Study 4, the statistical survey presented in Chapter 8.3 suffered from a low sample size. Eight participants was the minimum requirement, but the power of the t-test would have been more satisfactory with a few more participants. Unfortunately, recruiting was difficult as this was during COVID-19. Furthermore, in Chapter 8.6.3, I listed some potential reliability issues. In particular, the fact that the participants played different instruments and hence got responses from the prototypes with varying degrees of appropriateness was problematic. In future tests, I will make sure that all participants play the same instrument group. Seeing as the guitars apparently got the more appropriate responses, I would most likely invite guitarists for a first test. Additionally, it is quite clear that the Creativity Support Index displays a degree of arbitrariness in the results. For example, the more detailed comparisons in Chapter 8.4.4 showed a much more nuanced picture, with Auto Modes frequently mentioned as both more creative and more like a partner. Another example is that P1 seemed quite certain that he preferred Auto Modes and felt more creative with it. The fact that the CSI still showed a higher score for Manual Modes is quite odd. Perhaps he "calibrated" differently between the tests, and forgot how he had scored after the first session. It is likely that such arbitrariness will be less of an issue in surveys with large sample sizes. In this survey with only eight participants, however, such aberrations may be decisive.

In Chapter 8, one of the most striking findings was how much the participants' opinions about their interactions with the prototypes changed from their first immediate reactions to their evaluations a week later. If I had anticipated such a shift, I would have designed the study quite differently. In future studies, it would still be interesting to capture participants' immediate reactions by conducting post-session interviews. I also think the one-week delay before listening to the recording and evaluating worked well. However, I would design a longitudinal study where participants are invited to have regular new sessions with the prototype over a longer period to see how they develop their playing styles and sensibilities in relation to the musical agent. Such a study would probably give valuable data that could be used to further develop the software.

As for the Spire Muse software, I am already involved in a new artistic research project where Spire Muse is one of the musical agents used. The project is called *Co-Creative Spaces* and is funded by Arts Council Norway. The project follows four musicians through a six-month long music making process, and aims to shed light on issues and explore possibilities related to new forms of musical co-creation where artificial intelligence is part of the creative cycle. The study will not focus on software evaluation, but will most certainly give new insights into how musician develop their music making skills and sensibilities in relation to the technologies they apply in the process.

10.3 Concluding remarks

Throughout working with this thesis, the interdependence between humans and technology has been on full display, and I have frequently pondered upon the adage: “We shape our tools and thereafter our tools shape us” (Culkin, 1967). Sometimes I wonder if the opposite might be equally true. Technology is certainly human-made, but humans are also technology-made. We are born into and grow up with technologies that define us—we co-evolve. Sometimes, this thought is frightening. How will artificial intelligence shape us? Will the machines take over? When presenting my project, I sometimes sense this unease. Why should we want to make music with machines? Are people not enough? A bone flute carbon-dated 40,000 years ago begs to differ.

I am taking the initiative.

11 Bibliography

Abattouy, M., Alertz, U., Al-Hassani, S., Ambrosetti, N., & Aytes, A. (2016). *Allah's automata: Artifacts of the Arabic-Islamic Renaissance (800–1200)* (S. Zielinski Ed.). Hatje Cantz.

Albert, J. V. (2013). *Interactive Musical Partner: A system for human/computer duo improvisations.* [PhD thesis, Louisiana State University and Agricultural and Mechanical College]. LSU Doctoral Dissertations. https://digitalcommons.lsu.edu/gradschool_dissertations/3582

Albert, R. S., & Runco, M. A. (1999). A history of research on creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 16–31). Cambridge University Press. https://doi.org/10.1017/CBO9780511807916.004

Allauzen, C., Crochemore, M., & Raffinot, M. (1999). Factor Oracle: A new structure for pattern matching. In Pavelka, J., Tel, G., Bartošek, M. (Eds.), SOFSEM'99: *Theory and practice of informatics* (pp. 295–310). Springer. https://doi.org/10.1007/3-540-47849-3_18

Assayag, G. (2014). *Creative symbolic interaction.* [Key note talk]. 40th International Computer Music Conference and 11th Sound and Music Computer Conference (ICMC / SMC joint conference), Athens, Greece. https://hal.archives-ouvertes.fr/hal-01146734

Assayag, G., Bloch, G., Chemillier, M., Cont, A., & Dubnov, S. (2006). *OMAX brothers: A dynamic topology of agents for improvisation learning.* [Conference paper]. AMCMM '06: Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia, Santa Barbara, CA, USA (pp. 125–132). https://doi.org/10.1145/1178723.1178742

Bader, R. (2018). Musical instruments as synchronized systems. In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 171–196). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-55004-5_10

Baird, D. (2004). *Thing knowledge: A philosophy of scientific instruments.* University of California Press. https://doi.org/10.1525/9780520928206

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning.* Duke University Press Books. https://doi.org/10.1215/9780822388128

Barbosa, J., Malloch, J., Wanderley, M. M., & Huot, S. (2015). *What does "Evaluation" mean for the NIME community?* [Conference paper]. Proceedings of the international conference on new interfaces for musical expression, Baton Rouge, LA, USA (pp. 156–161). http://doi.org/10.5281/zenodo.1179010

Bargas-Avila, J. A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: *A critical analysis of empirical studies of user experience.* [Conference paper]. CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada (pp. 2689–2698. Association for Computing Machinery. https://doi.org/10.1145/1978942.1979336

Barnard, P. (1991). Bridging between basic theories and the artifacts of human-computer interaction. In J. M. Carroll (Ed.), *Designing interaction: Psychology at the human-computer interface.* (pp. 103–127). Cambridge University Press.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM, 37*(7), 122–125. https://doi.org/10.1145/176789.176803

Bausor, R. (1994). Qualitative dynamics in economics and fluid mechanics: A comparison of recent applications. In P. Mirowski (Ed.), *Natural images in economic thought: Markets read in tooth and claw* (pp. 109–127). Cambridge University Press. https://doi.org/10.1017/CBO9780511572128

Beaudouin-Lafon, M., & Mackay, W. E. (2003). Prototyping tools and techniques. In A. J. Julie & S. Andrew (Eds.), *The human-computer interaction handbook* (pp. 1006–1031). L. Erlbaum Associates Inc.

Beer, S. (1959). *Cybernetics and management.* English Universities Press.

Benson, B. E. (2003). *The improvisation of musical dialogue.* Cambridge University Press. https://doi.org/10.1017/CBO9780511615924

Berkowitz, A. (2010). *The improvising mind: Cognition and creativity in the musical moment.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199590957.001.0001

Beyls, P. (1988). *Introducing Oscar.* [Conference paper]. International Computer Music Conference, Cologne, Germany (pp. 219–230). https://quod.lib.umich.edu/i/icmc/bbp2372.1988.024/1

Beyls, P. (1991). Chaos and creativity: The dynamic systems approach to musical composition. *Leonardo Music Journal, 1*(1), 31–36. https://doi.org/10.2307/1513118

Beyls, P. (2018). *Motivated learning in human-machine improvisation.* [Conference Paper]. Proceedings of the International Conference on New Interfaces for Musical Expression, Blacksburg, VA, USA (pp. 238–243). http://doi.org/10.5281/zenodo.1302565

Biles, J. A. (1994). *GenJam: A genetic algorithm for generating jazz solos.* [Conference paper]. International Computer Music Conference, Aarhus, Denmark. https://quod.lib.umich.edu/i/icmc/bbp2372.1994.033/1

Bishop, C. M. (2006). *Pattern recognition and machine learning.* Springer New York.

Blackwell, T. M., & Bentley, P. (2002). *Improvised music with swarms.* [Conference paper]. Proceedings of the 2002 Congress on Evolutionary Computation, Honolulu, HI, USA (pp. 1462–1467). https://doi.org/10.1109/CEC.2002.1004458

Blackwell, T. M., Bown, O., & Young, M. (2012). Live algorithms: Towards autonomous computer improvisers. In J. McCormack & M. d'Inverno (Eds.), *Computers and creativity* (pp. 147–174). Springer-Verlag.

Boden, M. A. (1990). *The creative mind: Myths and mechanisms.* George Weidenfeld and Nicolson Ltd.

Boersen, R., Liu-Rosenbaum, A., Tatar, K., & Pasquier, P. (2020). *Chatterbox: an interactive system of gibberish agents.* [Conference paper]. Proceedings of international symposium of electronic art 2020 (ISEA 2020), Montreal, Quebec, Canada (pp. 55–62). https://www.isea-archives.org/docs/2020/ISEA2020-Programme-Catalogue.pdf

Bonnasse-Gahot, L. (2014). *An update on the SOMax project.* [Internal report]. IRCAM.

Borgo, D. (2005). *Sync or swarm: Improvising music in a complex age.* The Continuum International Publishing Group Inc.

Born, G., Lewis, E., & Straw, W. (2017). *Improvisation and social aesthetics.* Duke University Press. https://doi.org/10.1215/9780822374015

Bown, O. (2011). Experiments in modular design for the creative composition of live algorithms. *Computer music Journal, 35*(3), 73–85. https://doi.org/10.1162/COMJ_a_00070

Bown, O. (2014). *Empirically grounding the evaluation of creative systems: Incorporating interaction design.* [Conference paper]. Proceedings of the Fifth International Conference on Computational Creativity, Ljubljana, Slovenia (pp. 112–119). http://computationalcreativity.net/iccc2014/proceedings/proceedings-pdf/

Bown, O. (2015). *Attributing creative agency: Are we doing it right?* [Conference paper]. Proceedings of the Sixth International Conference on Computational Creativity, Park City, UT, USA (pp. 17–22). http://computationalcreativity.net/iccc2015/proceedings/1_3Bown.pdf

Bown, O. (2018). Performer interaction and expectation with live algorithms: experiences with Zamyatin. *Digital Creativity, 29*(1), 37–50. doi:10.1080/14626268.2018.1432663

Bown, O., Eldridge, A. C., & McCormack, J. (2009). Understanding interaction in contemporary digital music: from instruments to behavioural objects. *Organised Sound, 14*(2), 188–196. https://doi.org/10.1017/S1355771809000296

Bretan, M., Weinberg, G., & Heck, L. (2017). *A unit selection methodology for music generation using deep neural networks.* [Conference paper]. Proceedings of the Eighth International Conference on Computational Creativity, Atlanta, GA, USA (pp. 72–79). https://doi.org/10.48550/arXiv.1612.03789

Brown, A. R., Gifford, T., & Voltz, B. (2017). Stimulating creative partnerships in human-agent musical interaction. *Computers in Entertainment, 14*(2), 1–17. https//doi.org/10.1145/2991146

Brown, C., & Bischoff, J. (2002). *Indigenous to the net: Early network music bands in the San Francisco Bay area.* Crossfade. http://crossfade.walkerart.org/brownbischoff/index.html

Bryson, J. (1995). The reactive accompanist: Adaptation and behavior decomposition in a music system. In L. Steels (Ed.), *The biology and technology of intelligent autonomous agents.* Springer.

Bødker, S. (1989). A human activity approach to user interfaces. *Human-Computer Interaction,* 4(3), 171–195. https://doi.org/10.1207/s15327051hci0403_1

Candy, L., & Edmonds, E. (2002). Modeling co-creativity in art and technology. In E. A. Edmonds, L. Candy, T. Kavanagh & T. T. Hewett (Eds.), *Proceedings of the 4th conference on creativity & cognition* (pp. 134–141). Association for Computing Machinery. https://doi.org/10.1145/581710.581731

Canonne, C., & Garnier, N. (2011). A model for collective free improvisation. In C. Agon, M. Andreatta, G. Assayag, E. Amiot, J. Bresson & J. Mandereau, *Mathematics and computation in Music* (pp. 29–42). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21590-2_3

Canonne, C., & Garnier, N. B. (2012). *Cognition and segmentation in collective free improvisation: An exploratory study.* [Conference paper]. Proeedings of the 12th International Conference on Music Perception and Cognition and 8th Triennal Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki, Greece (pp. 197–204). https://doi.org/10.13140/2.1.3534.9445

Caramiaux, B., Altavilla, A., Pobiner, S. G., & Tanaka, A. (2015). *Form gollows sound: Designing interactions from sonic memories.* [Conference paper]. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, South Korea (pp. 3943–3952). https://doi.org/10.1145/2702123.2702515

Caramiaux, B., & Tanaka, A. (2013). *Machine learning of musical gestures.* [Conference paper]. Proceedings of New Interfaces for Musical Expression, Daejeon & Seoul, South Korea (pp. 513–518). http://doi.org/10.5281/zenodo.1178490

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Lawrence Erlbaum Associates.

Carroll, E. A., & Latulipe, C. (2012). *Triangulating the personal creative experience: self-report, external judgments, and physiology.* [Conference paper]. Proceedings of Graphics Interface 2012, Toronto, Ontario, Canada (pp. 53–60). https://dl.acm.org/doi/10.5555/2305276.2305286

Carroll, J. M. (1990). *Infinite detail and emulation in an ontologically minimized HCI.* [Conference paper]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Seattle, WA, USA (pp. 321–328). https://doi.org/10.1145/97243.97303

Carroll, J. M., & Kellogg, W. A. (1989). *Artifact as theory-nexus: Hermeneutics meets theory-based design.* [Conference paper]. CHI '89: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Seattle, WA, USA (pp. 7–14). https://doi.org/10.1145/67449.67452

Carroll, J. M., Kellogg, W. A., & Rosson, M. B. (1991). The task-artifact cycle. In J. M. Carroll (Ed.), *designing interaction: Psychology at the human-computer interface* (pp. 74–102). Cambridge University Press.

Chadabe, J. (1997). *Electric sound: The past and promise of electronic music.* Prentice Hall.

Cherry, E., & Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Trans. Computer-Human Interaction, 21*(4), 1–25. https://doi.org/10.1145/2617588

Collins, N. (2008). *Reinforcement learning for live musical agents.* [Conference paper]. Proceedings of ICMC 2008 International Computer Music Conference, Belfast, Northern Ireland, UK. http://hdl.handle.net/2027/spo.bbp2372.2008.005

Collins, N. (2010). *Contrary motion: An oppositional interactive music system.* [Conference paper]. Proceedings of New Interfaces for Musical Expression, Sydney, Australia (pp. 125–129). http://doi.org/10.5281/zenodo.1177747

Collins, N. (2011). *LL: Listening and learning in an interactive improvisation system.* Technical report, University of Sussex.

Colton, S., & Wiggins, G. A. (2012). *Computational creativity: the final frontier?* [Conference paper]. ECAI'12: Proceedings of the 20th European Conference on Artificial Intelligence, Montpellier, France (pp. 21–26). https://dl.acm.org/doi/10.5555/3007337.3007345

Conklin, D. (2003). *Music generation from statistical models.* [Conference paper]. Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences, Aberystwyth, Wales, UK.

Cont, A., Dubnov, S., & Assayag, G. (2007). Anticipatory model of musical style imitation using collaborative and competitive reinforcement learning. In M. V. Butz, O. Sigaud, G. Pezzulo, G. Baldassarre (Eds.), *Anticipatory behavior in adaptive learning systems. ABiALS 2006. Lecture Notes in Computer Science (4520),* 285–306. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74262-3_16

Cope, D. (1991). *Computers and musical style.* A-R Editions, Inc.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* Harper Perennial Modern Classics.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention.* Harper Collins.

Culkin, J. M. (1967). *A schoolman's guide to Marshall McLuhan.* Saturday Review.

Dacey, J. S. (2011). Historical conceptions of creativity. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (2nd ed.). Academic Press.

Davis, N. (2015). *An enactive approach to facilitate interactive machine learning for co-creative agents.* [Conference paper]. Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition, Glasgow, United Kingdom (pp. 345–346). https://doi.org/10.1145/2757226.2764773

Davis, T. (2011). Towards a relational understanding of the performance ecosystem. *Organised Sound, 16*(2), 120–124. https://doi.org/10.1017/s1355771811000069

Day, R. H. (1993). Nonlinear dynamics and evolutionary economics. In R. H. Day & P. Chen (Eds.), *Nonlinear dynamics and evolutionary economics* (pp. 18-41). Oxford University Press.

Dean, R. T. (2003). *Hyperimprovisation: Computer-interactive sound improvisation.* A-R Editions.

Demos, A. (2013). T*he Perception of Movement through Musical Sound: Towards a Dynamical Systems Theory of Music Performance.* [PhD thesis, University of Connecticut]. Doctoral Dissertations. 155. https://opencommons.uconn.edu/dissertations/155

DeNora, T. (2000). *Music in everyday life.* Cambridge University Press.

Denzin, N. K. (1978). *Sociological methods: A sourcebook.* McGraw-Hill.

Deterding, S., Hook, J., Fiebrink, R., Gillies, M., Gow, J., Akten, M., Smith, G. C., Liapis, A., & Compton, K. (2017). *Mixed-initiative creative interfaces.* [Conference paper]. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, USA (pp. 628–635). https://doi.org/10.1145/3027063.3027072

Dourish, P. (2001). W*here the action is: The foundations of embodied interaction.* The MIT Press.

Dubnov, S., & Assayag, G. (2005). *Improvisation planning and jam session design using concepts of sequence variation and flow experience.* [Conference paper]. 2nd Sound and Music Computing Conference (SMC2005), Salerno, Italy. https://doi.org/10.5281/zenodo.849279

Dubnov, S., Assayag, G., & Cont, A. (2011). *Audio Oracle analysis of musical information rate.* Proceedings of 2011 IEEE Fifth International Conference on Semantic Computing, Palo Alto, CA, USA (pp. 567–571). https://doi.org/10.1109/ICSC.2011.106

Eigenfeldt, A. (2014). *Generating structure - Towards large-scale formal generation.* [Conference Workshop Paper]. Musical Metacreation, Papers from the 2014 AIIDE Workshop, October, Raleigh, NC (pp. 2–9). https://www.aaai.org/ocs/index.php/AIIDE/AIIDE14/paper/view/9012

Eno, B. (2004). Generating and organising variety in the arts. In C. Cox & D. Warner (Eds.), *Audio cultures: Readings in modern music.* Continuum International Publishing Group.

Farbood, M., & Schöner, B. (2001). *Analysis and synthesis of Palestrina-style counterpoint using markov chains.* [Conference paper]. Proceedings of International Computer Music Conference.

Fiebrink, R. (2011). *Real-time human interaction with supervised learning algorithms for music composition and performance.* [PhD, Princeton University]. ACM Digital Library. https://dl.acm.org/doi/10.5555/2125776

Fiebrink, R., & Caramiaux, B. (2018). The machine learning algorithm as creative musical tool. In R. T. Dean & A. McLean (Eds.), *The Oxford handbook of algorithmic music.* https://doi.org/10.1093/oxfordhb/9780190226992.013.23

Flanagan, J. C. (1954). The critical incident technique. P*sychological Bulletin, 51*(4), 327–358. https://doi.org/10.1037/h0061470

Frabetti, F. (2017). Interview with Mark Fell. *Cesura/Acceso 2.* Cesura//Acceso.

Fuller, S. (1994). Making agency count: A brief foray into the foundations of social theory. *American Behavioral Scientist, 37*(6), 741–753. https://doi.org/10.1177/0002764294037006003

Gadamer, H.-G. (1997). Reflections on my philosophical journey. In L. E. Hahn (Ed.), *The Philosophy of Hans-Georg Gadamer.* Open Court.

Garcia, J. (2014). *Supporting music composition with interactive paper.* [PhD thesis, Université Paris Sud - Paris XI]. HAL theses. https://tel.archives-ouvertes.fr/tel-01056992

Gaver, W. W. (1991). *Technology affordances.* [Conference paper]. CHI '91: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA (pp. 79–84). https://doi.org/10.1145/108844.108856

Gaver, W. W. (2012). *What should we expect from research through design?* [Conference paper]. CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA (pp. 937–946). https://doi.org/10.1145/2207676.2208538

Gibson, J. J. (1979). T*he ecological approach to visual perception.* Houghton, Mifflin and Company.

Gifford, T., Knotts, S., McCormack, J., Kalonaris, S., Yee-King, M., & d'Inverno, M. (2018). Computational systems for music improvisation. *Digital Creativity,* 29(1), 19–36. https://doi.org/10.1080/14626268.2018.1426613

Gleick, J. (1987). *Chaos: Making a new science.* Penguin Books.

Grawe, K. (2004). *Psychological therapy.* Hogrefe.

Greenbaum, J., & Kyng, M. (1991). *Design at work: Cooperative design of computer systems.* Lawrence Erlbaum Associates.

Guattari, F. (1995). *Chaosmosis: An ethico-aesthetic paradigm* (P. Bains & J. Pefanis, Trans.). Indiana University Press (Orginal work published 1992).

Hamanaka, M., Goto, M., Asoh, H., & Otsu, N. (2003). *A learning-based jam session system that imitates a player's personality model.* [Conference paper]. IJCAI'03: Proceedings of the 18th international joint conference on Artificial intelligence, Acapulco, Mexico (pp. 51–58). https://dl.acm.org/doi/abs/10.5555/1630659.1630667

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers, 22*(5), 353–362. https://doi.org/10.1016/j.intcom.2010.04.002

Hassenzahl, M., Platz, A., Burmester, M., & Lehner, K. (2000). *Hedonic and ergonomic quality aspects determine a software's appeal.* [Conference paper]. CHI '00: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, The Hague, The Netherlands (pp. 201–208). https://doi.org/10.1145/332040.332432

Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., & Verplank, W. (1992). *ACM SIGCHI curricula for human-computer interaction.* Association for Computing Machinery. https://doi.org/10.1145/2594128

Hiller, L. A., & Isaacson, L. M. (1959). *Experimental music: Composition with an electronic computer.* McGraw-Hill.

Horvitz, E. (1999). *Principles of mixed-initiative user interfaces.* [Conference paper]. CHI '99: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Pittsburgh, Pennsylvania, USA (pp. 159–166). https://doi.org/10.1145/302979.303030

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A. M., Hoffman, M., Dinculescu, M., & Eck, D. (2018). *Music transformer: Generating music with long-term structure.* [Conference paper]. 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA.

Hudson, S. E., & Mankoff, J. (2014). Concepts, values, and methods for technical human–computer interaction research. In J. S. Olson & W. A. Kellogg (Eds.), *Ways of knowing in HCI* (pp. 69–93). Springer New York.

Hunt, A., & Kirk, R. (2000). Mapping strategies for musical performance. In M. M. Wanderley & M. Battier (Eds.), *Trends in gestural control of music* (pp. 231–258). IRCAM.

Hunt, A., Wanderley, M. M., & Paradis, M. (2002). *The importance of parameter mapping in electronic instrument design.* [Conference paper]. Proceedings of the 2002 conference on New interfaces for musical expression, Dublin, Ireland (pp. 1–6). https://dl.acm.org/doi/10.5555/1085171.1085207

Hutchins, E. (1995). *Cognition in the wild.* The MIT Press.

IMA (1983). MIDI *musical instrument digital interface specification 1.0.* International MIDI Association.

Jensenius, A. R. (2007). *Action-sound: Developing methods and tools to study music-related body movement.* [PhD, University of Oslo]. DUO Research Archive. http://urn.nb.no/URN:NBN:no-18922

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., & Leman, M. (2010). Musical gestures: Concepts and methods in research. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 12–36). Routledge. https://doi.org/10.4324/9780203863411

Jones, D., Brown, A. R., & d'Inverno, M. (2012). The extended composer. In J. McCormack & M. d'Inverno (Eds.), *Computers and creativity* (pp. 175–203). Springer-Verlag. https://doi.org/10.1007/978-3-642-31727-9

Jordanous, A. (2017). *Co-creativity and perceptions of computational agents in co-creativity.* [Conference paper]. Proceedings of the Eighth International Conference on Computational Creativity, Atlanta, Georgia, USA (pp. 159–166). https://kar.kent.ac.uk/61658/

Kaplan, D., & Glass, L. (1995). *Understanding nonlinear dynamics.* Springer-Verlag.

Kaptelinin, V., & Nardi, B. A. (2006). *Acting with technology: Activity theory and interaction design.* The MIT Press.

Karimi, P., Maher, M. L., Grace, K., & Davis, N. (2018). *Evaluating creativity in computational co-creative systems.* [Conference paper]. Proceedings of the Ninth International Conference on Computational Creativity, Salamanca, Spain. https://doi.org/10.48550/arXiv.1807.09886

Kellert, S. (2008). *Borrowed knowledge: Chaos theory and the challenge of learning across disciplines.* University of Chicago Press. https://doi.org/10.7208/chicago/9780226429809.001.0001

Kelley, J. F. (1983). *An empirical methodology for writing user-friendly natural language computer applications.* CHI '83: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA (pp. 193–196). https://doi.org/10.1145/800045.801609

Kirchhoff, E., & Matheson, C. (1997). Chaos and literature. *Philosophy and Literature, 21*(1), 28–45. https://doi.org/10.1353/phl.1997.0006

Kittay, E. (1987). Metaphor: *Its cognitive force and linguistic structure.* Clarendon.

Koestler, A. (1964). T*he Act of Creation.* Hutchinson & Co.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE, 78*(9), 1464–1480. https://doi.org/10.1109/5.58325

Kvale, S., & Brinkmann, S. (2015). *InterViews: Learning the Craft of Qualitative Research Interviewing* (3rd ed.). SAGE.

Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By.* University of Chicago Press.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory.* Oxford University Press.

Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. S*ystems practice, 5,* 379–393. https://doi.org/10.1007/BF01059830

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research methods in human-computer interaction* (2 ed.). Morgan Kaufmann.

Levinas, E. (1996). Trancendence and height. In E. Levinas (Ed.), *Basic philosophical writings* (pp. 11–32). Indiana University Press.

Lewis, G. E. (2000). Too many notes: Computers, complexity and culture in "Voyager". *Leonardo Music Journal, 10,* 33–39. https://doi.org/10.1162/096112100570585

Lewis, G. E. (2017). From network bands to ubiquitous computing: Rich Gold and the social aesthetics of interactivity. In G. Born, E. Lewis, & W. Straw (Eds.), I*mprovisation and social aesthetics* (pp. 91–109). Duke University Press. https://doi.org/10.1215/9780822374015-005

Liapis, A., Yannakakis, G. N., Alexopoulos, C., & Lopes, P. (2016). Can computers foster human users' creativity? Theory and praxis of mixed-initiative co-creativity. *Digital Culture & Education, 8*(2), 195–214. https://doi.org/10.1007/978-3-319-42716-4_11

Long, J. (1989). Cognitive ergonomics and human computer interaction: An introduction. In J. Long & A. Whitefield (Eds.), *Cognitive ergonomics and human computer interaction* (pp. 4–34). Cambridge University Press.

Long, J., & Whitefield, A. (1989). *Cognitive ergonomics and uman computer interaction.* Cambridge University Press.

Loughran, R., & O'Neill, M. (2018). *Is computational creativity domain-general?* In F. Pachet, A. Jordanous & C. Leòn (Eds.), *Proceedings of the ninth international conference on computational creativity (iccc'18)* (pp. 112–119). Association for computational creativity.

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal, 29*(6), 861–878. http://www.jstor.org/stable/1502138

Mackay, W. E., & Fayard, A.-L. (1997). *HCI, natural science and design: a framework for triangulation across disciplines.* [Conference paper]. DIS '97: Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques, Amsterdam, The Netherlands (pp. 223–234). https://doi.org/10.1145/263552.263612

Magnusson, T. (2009). *Epistemic tools: the phenomenology of digital musical instruments.* [PhD thesis, University of Sussex]. Sussex Research Online. http://sro.sussex.ac.uk/id/eprint/83540/

Mainzer, K. (2004). *Thinking in complexity: The computational dynamics of matter, mind and mankind* (4th ed.). Springer-Verlag.

Malt, M., & Jourdan, E. (2008). *Zsa.Descriptors: A library for real-time descriptors analysis.* [Conference paper]. 5th Sound and Music Computing Conference (SMC2008), Berlin, Germany. https://doi.org/10.5281/zenodo.849571

Marriott, M. (1976). *Hindu transactions: Diversity without dualism.* University of Chicago, Committee on Southern Asian Studies.

Martin, C. P. (2016). A*pps, agents, and improvisation: Ensemble interaction with touch-screen digital musical instruments.* [PhD thesis, Australian National University]. Open Research Library. https://doi.org/10.25911/5d78d600cc591

Masterpasqua, F., & Perna, P. A. (1997). *The Psychological meaning of chaos: Translating theory into practice.* American Psychological Association. https://doi.org/10.1037/10240-000

Meusburger, P. (2009). Milieus of creativity: The role of places, environments, and spatial contexts. In P. Meusburger, J. Funke, & E. Wunder (Eds.), *Milieus of creativity. Knowledge and space, vol. 2* (pp. 97–153). Springer. https://doi.org/10.1007/978-1-4020-9877-2_7

Minsky, M. (1986). *The society of mind.* Simon & Schuster.

Miranda, E. R., & Wanderley, M. M. (2006). *New digital musical instruments: control and interaction beyond the keyboard.* A-R Editions.

Misselhorn, C. (2015). *Collective agency and cooperation in natural and artificial systems: explanation, implementation and simulation.* Springer Publishing Company, Incorporated.

Mitchell, T. (2006). *The discipline of machine learning.* [Technical report]. Carnegie Mellon University, School of Computer Science, Machine Learning Department.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Mudd, T. (2017). *Nonlinear dynamics in musical interactions.* [PhD thesis, The Open University]. Open Research Online. https://doi.org/10.21954/ou.ro.0000cc07

Mudd, T. (2019). Material-oriented musical interactions. In S. Holland, T. Mudd, K. Wilkie-McKenna, A. McPerson & M. Wanderley (Eds.), *New directions in music and human-computer cnteraction* (pp. 123–133). Springer, Cham. https://doi.org/10.1007/978-3-319-92069-6

Murray-Rust, D., & Smaill, A. (2011). Towards a model of musical interaction and communication. *Artificial Intelligence, 175*(9), 1697–1721. https://doi.org/10.1016/j.artint.2011.01.002

Nelson, H. G., & Stolterman, E. (2003). *Design way: Intentional change in an unpredictable world.* The MIT Press. http://www.jstor.org/stable/j.ctt5hhcsk

Nielsen, J. (1993). *Usability engineering.* Academic Press.

Nika, J., & Chemillier, M. (2012). *Improtek: integrating harmonic controls intoimprovisation in the filiation of OMax.* [Conference paper]. Proceedings of the International Computer Music Conference (ICMC), Ljubljana, Slovenia (pp. 180–187). https://doi.org/10.13140/2.1.5009.8564

Norman, D. A. (1988). *The psychology of everyday things.* Basic Books.

Norman, D. A. (1999). Affordance, conventions, and design. *Interactions, 6,* 38–43. https://doi.org/10.1145/301153.301168

Osborn, A. F. (1953). *Applied imagination: Principles and procedures of creative thinking.* Scribner.

Ott, E., Sauer, T., & Yorke, J. A. (1994). *Coping with chaos: Analysis of chaotic data and the exploitation of chaotic systems.* John Wiley & Sons, Inc.

Pachet, F. (2003). The Continuator: Musical interaction with style. *Journal of New Music Research, 32*(3), 333–341. https://doi.org/10.1076/jnmr.32.3.333.16861

Pasquier, P., Eigenfeldt, A., Bown, O., & Dubnov, S. (2017). An introduction to musical metacreation. *Computers in Entertainment, 14*(2), 1–14. https://doi.org/10.1145/2930672

Pasteur, L. (1854). "Dans les champs de l'observation le hasard ne favorise que les esprits préparés". [Lecture, University of Lille]. Accessed 31 December 2021. https://en.wikiquote.org/wiki/Louis_Pasteur

Pelz-Sherman, M. (1998). *A framework for the analysis of performer interactions in western improvised contemporary art music.* [PhD, University of California]. Scribd. https://www.scribd.com/doc/224601338

Pérez y Pérez, R. (2018). *The computational creativity continuum.* [Conference paper]. Proceedings of the Ninth International Conference on Computational Creativity, Salamanca, Spain. http://ilitia.cua.uam.mx:8080/jspui/handle/123456789/449

Pickering, A. (1993). The mangle of practice: Agency and emergence in the sociology of science. *American Journal of Sociology,* 99(3), 559–589. https://doi.org/10.1086/230316

Pickering, A. (1995). T*he mangle of practice: Time, agency and science.* University of Chicago Press.

Pickering, A. (2010). *The cybernetic brain: Sketches of another future.* University of Chicago Press.

Pressing, J. (1984). Cognitive processes in improvisation. In W. R. Crozier & A. J. Chapman (Eds.), *Advances in Psychology, 19,* 345–363. https://doi.org/10.1016/S0166-4115(08)62358-4

Pressing, J. (2001). Improvisation: Methods and models. In J. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition.* Oxford University Press.

Rammert, W. (2008). Where the action is: Distributed agency between humans, machines and programs. In U. Seifert, J. H. Kim, & A. Moore (Eds.), *Paradoxes of interactivty: Perspectives for media theory, human-computer interaction, and artistic investigations* (pp. 62–91). transcript Verlag. https://doi.org/10.14361/9783839408421-004

Ravikumar, P. T., & Wyse, L. (2019). MASSE: *A system for music action selection through state evaluation.* [Conference paper]. Proceedings of the Tenth International Conference on Computational Creativity, Charlotte, NC, USA (pp. 132–139). http://computationalcreativity.net/iccc2019/papers/iccc19-paper-35.pdf

Reynolds, G. H. (1991). Chaos and the Court. *Columbia Law Review, 91*(1), 110-117. https://doi.org/10.2307/1122858

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155–169. https://doi.org/10.1007/bf01405730

Roads, C. (1985). *Composers and the computer.* W. Kaufmann.

Roads, C. (1996). *The computer music tutorial.* The MIT Press.

Roads, C. (2001). *Microsound.* The MIT Press.

Rogers, Y. (2012). *HCI theory: Classical, modern, and contemporary.* Morgan Claypool.

Rowe, R. (1992). *Interactive music systems: Machine listening and composing.* The MIT Press.

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Pearson Education.

Saldaña, J. (2009). *The coding manual for qualitative researchers.* Sage.

Sawyer, R. K. (1999). The emergence of creativity. *Philosophical Psychology, 12*(4), 447–469. https://doi.org/10.1080/095150899105684

Sawyer, R. K. (2003). *Improvised dialogues: Emergence and creativity in conversation.* Ablex.

Schnell, N., Röbel, A., Schwarz, D., Peeters, G., & Borghesi, R. (2009). *MuBu and friends: Assembling tools for content based real-time interactive audio processing in Max/MSP.* [Conference paper]. Proceedings of the International Computer Music Conference (ICMC 2009), Montreal, Canada (pp. 423–426). http://hdl.handle.net/2027/spo.bbp2372.2009.096

Schnell, N., Schwarz, D., Larralde, J., & Borghesi, R. (2017). *PiPo, a plugin interface for afferent data stream processing operators. [Conference paper].* Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR, Suzhou, China (pp. 361–367). https://doi.org/10.5281/zenodo.1416912

Schwarz, D. (2000). *A system for data-driven concatenative sound synthesis.* [Conference paper]. Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx00), Verona, Italy (pp. 97–102). https://hal.archives-ouvertes.fr/hal-01161115

Schwarz, D., Beller, G., Verbrugghe, B., & Britton, S. (2006). *Real-time corpus-based concatenative synthesis with CataRT.* [Conference paper]. Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06), Montreal, Canada (pp. 279–282). https://hal.archives-ouvertes.fr/hal-01161358

Schön, D. A. (1993). Generative metaphor: A perspective on problem-setting in social policy. In A. Ortony (Ed.), *Metaphor and thought* (2 ed.), (pp. 137–163). Cambridge University Press.

Seddon, F. A. (2005). Modes of communication during jazz improvisation. *British Journal of Music Education, 22*(1), 47–61. https://doi.org/10.1017/S0265051704005984

Shneiderman, B. (2002). *Leonardo's laptop: Human needs and the new computing technologies.* The MIT Press.

Shneiderman, B. (2007). Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM, 50*(12), 20–32. https://doi.org/10.1145/1323688.1323689

Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: The integration of product, person, and process perspectives. *Psychol Bull, 129*(4), 475–494. https://doi.org/10.1037/0033-2909.129.4.475

Skarda, C. A., & Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences, 10*(2), 161–173. https://doi.org/10.1017/S0140525X00047336

Smith, B. D., & Deal, W. S. (2014). *ML.* Machine learning library as a musical partner in the computer-acoustic composition flight.* [Conference paper]. Proceedings of the 2014 International Computer Music Conference, Athens, Greece (pp. 1285–1289). http://hdl.handle.net/2027/spo.bbp2372.2014.197

Smith, B. D., & Garnett, G. E. (2012). *Unsupervised play: Machine learning toolkit for Max.* [Conference paper]. Proceedings of the International Conference on New Interfaces for Musical Expression, Ann Arbor, MA, USA. http://doi.org/10.5281/zenodo.1178419

Stake, R. (2003). Case studies. In N. K. Denzin (Ed.), *Strategies of qualitative inquiry* (2 ed.) (pp. 134–164). Sage.

Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: Entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception and Performance, 35*(6), 1811–1832. https://doi.org/10.1037/a0014510

Strogatz, S. H. (2015). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering* (2nd ed.). Westview.

Suchman, L. A. (1987). *Plans and situated actions: the problem of human-machine communication.* Cambridge University Press.

Surges, G., & Dubnov, S. (2013). F*eature selection and composition using PyOracle.* [Workshop paper]. 2nd International Workshop on Musical Metacreation (MUME 2013), Boston, MA, USA (pp. 144–121).

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2 ed.). MIT Press.

Swift, B. (2012). *The design of a smartphone-based digital musical instrument for jamming.* [PhD thesis, Australian National University]. Open Research Library. https://openresearch-repository.anu.edu.au/handle/1885/151432

Swift, B. (2013). Chasing a feeling: Experience in computer supported jamming. In S. Holland, K. Wilkie, P. Mulholland, & A. Seago (Eds.), *Music and human-computer interaction.* Springer. https://doi.org/10.1007/978-1-4471-2990-5_5

Tatar, K., & Pasquier, P. (2017). *MASOM: A musical agent architecture based on self-organizing maps, affective computing, and variable markov models.* [Conference paper]. Proceedings of the 5th international workshop on musical metacreation (MUME 2017), Atlanta, GA, USA.

Tatar, K., & Pasquier, P. (2018). Musical agents: A typology and state of the art towards Musical Metacreation. *Journal of New Music Research, 48*(1), 56–105. https://doi.org/10.1080/09298215.2018.1511736

Tatar, K., Pasquier, P., & Siu, R. (2018). *REVIVE: An audio-visual performance with musical and visual AI agents.* [Conference paper]. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada. https://doi.org/10.1145/3170427.3177771

The Norwegian National Committees for Research Ethics. (2016). General guidelines for research ethics. https://www.forskningsetikk.no/en/guidelines/general-guidelines/

Thelle, N. J. W., & Pasquier, P. (2021). *Spire Muse: A virtual musical partner for creative brainstorming.* [Conference paper]. New Interfaces for Musical Expression (NIME 2021), Shanghai, China. https://doi.org/10.21428/92fbeb44.84c0b364

Thom, B. (2000). *BoB: An interactive improvisational music companion.* [Conference paper]. AGENTS '00: Proceedings of the fourth international conference on Autonomous agents, Barcelona, Spain (pp. 309–316). https://doi.org/10.1145/336595.337510

Tschacher, W. (1997). *Prozessgestalten.* Hofgrefe-Verlag.

Tschacher, W., & Scheier, C. (1997). Complex psychological systems: Synergetics and chaos. In *The psychological meaning of chaos: Translating theory into practice.* (pp. 273–298). American Psychological Association.

Tuomela, R., & Miller, K. (1988). We-intentions. *Philosophical Studies, 53*(3), 367–389. https://doi.org/10.1007/BF00353512

van der Schyff, D., Schiavio, A., Walton, A., Velardo, V., & Chemero, A. (2018). Musical creativity and the embodied mind: Exploring the possibilities of 4E cognition and dynamical systems theory. *Music & Science, 1.* https://doi.org/10.1177/2059204318792319

Varèse, E., & Wen-chung, C. (1966). The liberation of sound. *Perspectives of New Music, 5*(1), 11–19. https://doi.org/10.2307/832385

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need.* [Conference paper]. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. https://doi.org/10.48550/arXiv.1706.03762

Waern, A., & Back, J. (2017). *Activity as the ultimate particular of interaction design.* [Conference paper]. CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA (pp. 3390–3402). https://doi.org/10.1145/3025453.3025990

Walker, W., Hebel, K., Martirano, S., & Scaletti, C. (1992). *ImprovisationBuilder: Improvisation as conversation.* [Conference paper]. International Computer Music Conference, San Jose, CA (pp. 190–193).

Wallas, G. (1926). *The art of thought.* Harcourt, Brace and Company.

Walton, A. E., Richardson, M. J., Langland-Hassan, P., & Chemero, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Frontiers in psychology, 6,* 313–313. https://doi.org/10.3389/fpsyg.2015.00313

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine.* MIT Press.

Wilson, A. J. (2016). *factorOracle: An extensible Max external for investigating applications of the Factor Oracle automaton in real-time music improvisation.* [Conference paper]. MUME 2016 - The Fourth International Workshop on Musical Metacreation. https://academicworks.cuny.edu/ny_pubs/114/

Wilson, G. B., & MacDonald, R. A. R. (2016). Musical choices during group free improvisation: A qualitative psychological investigation. *Psychology of Music, 44*(5), 1029–1043. https://doi.org/10.1177/0305735615606527

Wright, M. (2005). Open Sound Control: An enabling technology for musical networking. *Organised Sound, 10*(3), 193–200. https://doi.org/10.1017/S1355771805000932

Wu, Y. (2018). *Design and evaluate support for non-musicians' creative engagement with musical interfaces.* [PhD, Queen Mary University of London]. Queen Mary University of London Theses. http://qmro.qmul.ac.uk/xmlui/handle/123456789/49923

Xi, Q., Bittner, R. M., Pauwels, J., Ye, X., & Bello, J. P. (2018). *GuitarSet: A dataset for guitar transcription.* [Conference paper]. 19th International Society for Music Information Retrieval Conference, Paris, France (pp. 23–27). https://doi.org/10.5281/zenodo.3371780

Yin, R. K. (2009). *Case study research: Design and methods.* Sage.

Young, M. (2008). NN Music: Improvising with a "living" computer. In R. Kronland-Martinet, S. Ystad & K. Jensen (Eds.), *Computer Music Modeling and Retrieval. Sense of Sounds. CMMR 2007. Lecture Notes in Computer Science, vol 4969.* Springer. https://doi.org/10.1007/978-3-540-85035-9_23

Young, M., & Bown, O. (2010). Clap-along: A negotiation strategy for creative musical interaction with computational systems. In D. Ventura, A. Pease, R. Pérez y Pérez, G. Ritchie & T. Veale, *Proceedings of the International Conference on Computational Creativity* (pp. 215–222). Departament of Informatics Engineering, University of Coimbra.

Zicarelli, D. (1987). M and Jam Factory. *Computer Music Journal, 11*(4), 13–29. https://doi.org/10.2307/3680237

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). *Research through Design as a method for interaction design research in HCI.* [Conference paper]. CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, CA, USA (pp. 493–502). https://doi.org/10.1145/1240624.1240704

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). *An analysis and critique of Research through Design: towards a formalization of a research approach.* [Conference paper]. DIS '10: Proceedings of the 8th ACM Conference on Designing Interactive Systems, Aarhus, Denmark (pp. 310–319). https://doi.org/10.1145/1858171.1858228

Appendixes

Appendix A:

Study 1 Material

A.1 Study 1: Invitation

Would you like to participate in the project

"Mixed-Initiative Composition: Collective Agency in Interactive Music Systems"?

Participants are wanted for a workshop in connection with a research project where the objective is to study how creative initiative is negotiated and shared between individuals in collective composition processes. The workshop is planned for **Thursday, October 17 at 10 a.m.** at the Norwegian Academy of Music, and will last for between 2 and 3 hours.

Objective
Ideas for music compositions often emerge in the context of improvisational interactions between musicians. In this PhD project, I will research how creative initiative is negotiated through dynamic interactions between musician-composers. Describing how ideas develop in such contexts can be very challenging due to the intrinsic knowledge and intuitive choice-making that musicians apply in the process. In this workshop, I will focus on human-to-human interaction, but in the longer term I am interested in how interactions may be modelled in interactive music systems comprised of both human and computer agents.

This invitation to participate relates to the first of four workshops that will shed light on the topic in different ways.

Who may participate?
The target group for this workshop are professional/semi-professional musicians or music students on college/university level. The most important qualification is the experience of collectively composing music/making songs with other musicians (e.g. by improvising/jamming with ideas at rehearsals or similar contexts).

What does participation imply?
Participation implies joining a focus group with 8-16 participants. The format will switch between structured questions and moderated group discussions. The session will be recorded (sound and video).

I would like to participate – what do I need to do?
Send an email to nottot@nmh.no with the following details:

- Name
- Music background
- A short description of why you are interested in participating

Privacy
The information will be treated confidentially and in accordance with privacy regulations. All data will be anonymized in the final PhD thesis, og no information that may identify individuals will be published. All information published will have a relevance to the project's theme.

Best regards,

Notto J. W. Thelle
PhD fellow, Norwegian Academy of Music

A.2 Study 1: Consent form

DECLARATION OF CONSENT

"Mixed-Initiative Composition: Collective Agency in Interactive Music Systems"

Thank you for participating in the workshop on October 17, 2019. In order to allow data collected from the focus group interview to be treated by the researcher, please sign the declaration of consent below.

Objective
Ideas for music compositions often emerge in the context of improvisational interactions between musicians. In this PhD project, I will research how creative initiative is negotiated through dynamic interactions between musician-composers. Describing how ideas develop in such contexts can be very challenging due to the intrinsic knowledge and intuitive choice-making that musicians apply in the process. In this workshop, I will focus on human-to-human interaction, but in the longer term I am interested in how interactions may be modelled in interactive music systems comprised of both human and computer agents.

Project owner
The research project is part of a PhD at the Norwegian Academy of Music, who is responsible for the project.

Implications of participation
Participation implies joining a focus group with around eight participants. The format will switch between structured questions and moderated group discussions. The session will be recorded (audio and video).

Participation is voluntary
Consent may be retracted at any point with no reasons given. All details about you will be anonymized. Withdrawal from the workshop or later retraction will not have any negative consequences for you.

Your privacy – how we store and treat the data
We will only use information given by you for the objective stated in this document. Information will be treated confidentially and in accordance with privacy regulations.

- Persons with access to the data at the Norwegian Academy of Music: Notto J. W. Thelle (project leader) and Professor Sidsel Karlsen (supervisor).
- Your name and contact details will be substituted with codes and stored on a list separate from the rest of the data. All data will be stored on the institution's encrypted research server (Box).

All data will be anonymized, and no information that may identify individuals will be published. All published information will be relevant to the project's topic.

What happens to personal details when the research project ends?
According to the plan, the project will terminate in August 2021. All personal information will be anonymized at the end of the project.

Your rights
As long as you can be identified in the data material, you have the right to:
- request access to what personal information is registered about you,
- have personal details about you edited,
- have personal details about you deleted,
- access a copy of your personal details (data portability), and
- file a complaint to a privacy ombudsman or to the The Norwegian Data Protection Authority (Datatilsynet) about the treatment of your personal details.
-

What gives us the right to treat personal details about you?
We treat details about you on the basis of your consent.

On behalf of the Norwegian Music Academy, NSD – the Norwegian Centre for Research Data has deemed that the treatment of personal data in this project is in accordance with privacy regulations.

Where can I find out more?
If you have any questions regarding the research project, or would like to take advantage of your rights, please contact:
- The Norwegian Academy of Music by Notto Johannes Windju Thelle, nottot@nmh.no, tlf. 988 82 613
- Our privacy ombudsman: personvernombud@nmh.no
- NSD – Norwegian Centre for Research Data, on e-mail (personverntjenester@nsd.no) or telephone: 55 58 21 17.

Best regards,
Notto J W Thelle
Project leader

Declaration of consent

I have received and understood the information about the project *Mixed-Initiative Composition: Collective Agency in Interactive Music Systems*, and have been given the opportunity to ask questions. I consent to:

- ☐ participating in a focus group interview and discussion
- ☐ having the entire session recorded (audio and video)

I consent to the treatment of my personal details until the project terminates in August 2021.

(Signed by project participant, date)

A.3 Study 1: Focus group guide

STUDY 1 – FOCUS GROUP GUIDE

Top goals:

- Understand if there is a terminological level of abstraction where interaction types can be discussed regardless of musical genre
- Understand if musician-composers with varying backgrounds share experiences of collective musical ideation that may be generalized using the same terminology
- Classify what musicians think of as their own musical "safe zones", and what kind of musical interactions they feel comfortable with engaging in
- Understand how they go about making trajectories away from these spaces to explore new ideas
- Understand how they experience interactional surprises and moments of radical change during interactions
- Get information about how to narrow scope of subsequent workshops (e.g., genre)
- Get information about previously unconsidered perspectives that may affect the scope of subsequent workshops

Longer term/ulterior goals:

- Provide some evidence that musicians' experiences are well characterized by dynamical systems theory
- Collect data to support an architecture for interactive music systems based on the principles of dynamical systems theory

1. Greeting and presentation of project

Brief summary of what the project is about.
Remind the participants that they should be basing what they contribute on their experience as composers or songwriters. As stated in the invitation, there is often an improvisational element in the creation of music. However, we will not be talking about improvisation as performance – there is no audience. In this context, by improvisation I mean a compositional method, a way to generate movement toward new musical ideas.
There has been plenty of research on improvisation as performance. Not so much on improvisational approaches in composition.

2. Short introduction round

- Name
- Musical background (experience, instrument(s), etc)
- Reason for interest

3. Roundtable presentation of events prepared by the participants

One round per theme: original idea, obtaining shared representation, development, result.
Display the different events symbolically on whiteboard or flip-overs during the presentations.
After each round: see if there is a way to place the different contributions into categories.

ORIGINAL IDEA

- What was the original idea? (a melodic theme, a phrase, a rhythmical pattern, a sound, a sample loop, a harmonic progression, a riff, etc.)

OBTAINING SHARED REPRESENTATION OF IDEA

- Was the idea easily shared (followed without instructions), or was there some negotiation before there could be a shared representation of that idea?

Notes to self:

- Identifying an attractor (convergence problem)
- Negotiate differences in understanding
- What are the attractors?
- Verbal or non-verbal exchange?
- How do you know that you have a shared representation?
- Minimal conditions for co-experience (visual cues or no?)

DEVELOPMENT OF SHARED IDEA

- How did the idea become "something more"? (new elements introduced, how, when, and by whom)
- Were you "pulled out from your safe zone" during the interaction?
- Did you pull your collaborator(s) out of his/her/their safe zone(s)?
- Were there any surprises during the interaction?

Notes to self:

- What are "elements"?
- (Could be) transition from one attractor to another (articulation problem)
- Maintain the current idea and explore it (going deep) or to initiate change to explore trajectories that may lead to a reframing of musical opportunities (going broad).
- Maintain strategy: settling on a groove, looping, repeating a musical phrase, a sequence, a riff, etc.
- = attractor (something with a certain amount of gravity where things have "fallen into place" and it may be hard to imagine "a way out")
- Once you "got each other", what then? Avoid something static... how?
- Leading/following (does the experience change depending on which role?
- Combinatorial (unexpected combination) or interactional (unusual instrument technique) surprise?
- "Game-changer"
- Lack of awareness during optimal performance: "creator-witness phenomenon"
- When do you stop to re-negotiate?

RESULT

- How was the musical result different from your original expectations?

> **Notes to self:**
>
> *- By "result", I do not mean a complete composition. I mean the sense of having made a qualitative step from a simple idea to a more complex arrangement. Something in the direction of becoming a composition. You go home afterward and have a clearer idea of where this is headed.*

4. Discussion

Go through the themes once more, this time as a collective discussion.

Are there any commonalities?
Genre-dependencies in types of interaction?

Tell the participants that it is fine to use metaphorical language at this point.

Carefully introduce the concepts behind dynamical systems theory but using informal substitute terminology (e.g. "gravitate toward something" instead of "attractor", "radical change" instead of bifurcation point.

Draw diagrams/symbolic representations during discussion to illustrate points (participants may also step forward and contribute)

Gauge responses, make note of agreements/disagreements and moderate the discussion, in keeping with the themes mentioned above

5. Summarize the focus group, last round of comments

A.4 Study 1: Themes and codes

Study 1 – Themes and codes

Theme 1: Maintaining a process-oriented approach

Codes

- Original idea
    - Melody from text
    - Water drop sample
    - Organ theme
    - Greenland as conceptual theme
    - Libretto
    - Guitar theme
    - Series of bass riffs

- Process-oriented approach
    - Growing from idea
    - Improvising as method
    - Goal-oriented vs process-oriented
    - Context dictated result
    - Discovering better goals is a goal in itself
    - Moving target
    - Not having a goal as something positive
    - Result can be a working method
    - Result is a stepping stone
    - Result not just song but foundation for future collaboration

- Strategies
    - Complimenting
    - Contrasting
    - Copying
    - Adding complexity
    - Avoiding boredom
    - Iteratively trying and writing down or recording
    - Present idea and gauge response
    - Trial and error
    - Verbal dialog
    - Abstract word giving direction
    - Mimicking an instrument
    - Switching roles
    - Recurring theme as anchor
    - Redefining instructions
    - Setting the mood
    - Using methods from other fields

Theme 2: Loss of ownership

Codes

- Fear of other people's opinions
- Losing ownership as method
- Shared ownership
- Being a computer
- Go out of comfort zone
- Challenged by other musicians
- Crossing threshold between politeness and being dramatic
- Following an initiative
- Forgetting to give space
- Giving parts to other musicians
- Dramatic shift
- Harsh dynamic
- Inclusive attitude
- Mutual initiative
- Negotiations about form
- Not important to follow a task
- Red lines
- Point of letting go or surrendering
- Stubbornness
- Taking the initiative
- Collaboration made the result better
- Surprising details

Theme 3: Listening well

Codes

- How do you know when you feel heard
- I could hear how they listened
- I could hear how they think
- Listening for things never heard
- Listening well
- Pure silence signals concentrated listening
- Recognizing other peoples mind state
- Context dictates process
- Cloud agreement - informed by previous jams
- Personal reactions are crucial

Appendix B:

Study 2 Material

B.1 Study 2: Invitation

Would you like to participate in the research project

"Mixed-Initiative Composition: Collective Agency in Interactive Music Systems"?

Are you a part of a song writing duo? Or do you play in a band and make your own songs collectively?

This is an invitation to participate in a research project where the objective is to study how creative initiative is negotiated through dynamic interactions between individuals in collective music composition processes. You are invited to a workshop lasting for approximately 60 minutes, where you and a music partner are asked to collaborate on developing a novel musical idea while being filmed. The workshop includes a post-session interview.

Objective

Ideas for music compositions often emerge in the context of improvisational interactions between musicians. This PhD project examines how creative initiative is negotiated through dynamic interactions between musicians. Describing how ideas develop in musically creative contexts can be very challenging due to the intrinsic knowledge and intuitive choice making that musicians apply in the creative process.

In this workshop, the focus is on human-human interaction. However, the workshop is part of a larger study to shed light on whether there are aspects of such interaction that can be modelled in an interactive music system comprising both humans and machines. Is it possible to create a "virtual jamming partner"?

Who is responsible for the project?

The research project is part of a PhD at the Norwegian Academy of Music, who is the project owner.

Who may participate?

The target group for this workshop is song-writing duos/composition partners who use creative interaction (jamming) as a method to develop musical ideas. The target group is not limited to a specific genre.

What does participation entail?

If you choose to participate in the project, you and a partner will be asked to join a 60-minute workshop in two parts.

In, the first part of the workshop, you and your musical partner will collaborate on a musical idea for 20-25 minutes. This interaction will start with a simple idea (phrase, riff, collection of chords) and develop into something more complex, as the first steps toward a composition (as far as you get in the limited time at disposal).

There are some criteria that will make the context quite constrained. However, these constraints may be a fun challenge for musicians to try:

- The musicians will use their main instrument, an acoustically-based instrumnt (i.e. not a laptop, but electrically amplified instruments are welcome)
- The musicians will not see each other during the session. They will sit in separate room, and only be able to make contact through simple commands on a computer screen (this is in order to reduce extra-musical cues normally given through verbal dialog and visual contact)
- The musicians will start with a simple musical theme, and collaborate to develop this into something more composed, via improvised interaction (jamming) and exchange of commands on a computer
- In order to make transcription easier, it is preferable that the musicians work with non-complex timbres (avoid multiphonics, noise or other extended instrumental techniques)

In the second part of the workshop, there will be an interview where I will focus on critical parts of the session where the interaction led to qualitative changes that led to new ideas.

Both the musical session and the interview will be recorded (sound and video), and will be analyzed later.

Privacy
The information will be treated confidentially and in accordance with privacy regulations. All data will be anonymized in the final PhD thesis, and no information that may identify individuals will be published. All information published will have a relevance to the project's theme.

What happens with the data when we finish the research project?
The project is planned to be finalized by August 2021. All personal information will be anonymized by the end of the project.

Best regards,

Notto J. W. Thelle
PhD fellow, Norwegian Academy of Music

B.2 Study 2: Consent form

DECLARATION OF CONSENT

"Mixed-Initiative Composition: Collective Agency in Interactive Music Systems"

Thank you for participating in the workshop on December 10, 2019. In order to allow data collected from the focus group interview to be treated by the researcher, please sign the declaration of consent below.

Objective
Ideas for music compositions often emerge in the context of improvisational interactions between musicians. In this PhD project, I will research how creative initiative is negotiated through dynamic interactions between musician-composers. Describing how ideas develop in such contexts can be very challenging due to the intrinsic knowledge and intuitive choice-making that musicians apply in the process. In this workshop, I will focus on human-to-human interaction, but in the longer term I am interested in how interactions may be modelled in interactive music systems comprised of both human and computer agents.

Project owner
The research project is part of a PhD at the Norwegian Academy of Music, who is responsible for the project.

Implications of participation
Participation implies joining in a 90-minute workshop in two parts together with a musical partner. In the first part, you will be involved in a 20-25 minute creative interaction where the objective is to work collaboratively on developing a musical idea. The details of this has been provided separately. The second part will be an interview where the focus will be on the emergence of new ideas during the interaction. The session and interview will be recorded (audio and video).

Participation is voluntary
Consent may be retracted at any point with no reasons given. All details about you will be anonymized. Withdrawal from the workshop or later retraction will not have any negative consequences for you.

Your privacy – how we store and treat the data
We will only use information given by you for the objective stated in this document. Information will be treated confidentially and in accordance with privacy regulations.

- Persons with access to the data at the Norwegian Academy of Music: Notto J. W. Thelle (project leader) and Professor Sidsel Karlsen (supervisor).
- Your name and contact details will be substituted with codes and stored on a list separate from the rest of the data. All data will be stored on the institution's encrypted research server (Box).

All data will be anonymized, and no information that may identify individuals will be published. All published information will be relevant to the project's topic.

What happens to personal details when the research project ends?
According to the plan, the project will terminate in August 2021. All personal information will be anonymized at the end of the project.

Your rights
As long as you can be identified in the data material, you have the right to:
- request access to what personal information is registered about you,
- have personal details about you edited,
- have personal details about you deleted,
- access a copy of your personal details (data portability), and
- file a complaint to a privacy ombudsman or to the The Norwegian Data Protection Authority (Datatilsynet) about the treatment of your personal details.
-

What gives us the right to treat personal details about you?
We treat details about you on the basis of your consent.

On behalf of the Norwegian Music Academy, NSD – the Norwegian Centre for Research Data has deemed that the treatment of personal data in this project is in accordance with privacy regulations.

Where can I find out more?
If you have any questions regarding the research project, or would like to take advantage of your rights, please contact:
- The Norwegian Academy of Music by Notto Johannes Windju Thelle, nottot@nmh.no, tlf. 988 82 613
- Our privacy ombudsman: personvernombud@nmh.no
- NSD – Norwegian Centre for Research Data, on e-mail (personverntjenester@nsd.no) or telephone: 55 58 21 17.

Best regards,

Project leader

Declaration of consent

I have received and understood the information about the project *Mixed-Initiative Composition: Collective Agency in Interactive Music Systems*, and have been given the opportunity to ask questions. I consent to:

- ☐ engaging in a creative musical interaction together with your musical partner
- ☐ participating in an interview together with your musical partner
- ☐ having the entire session recorded (audio and video)

I consent to the treatment of my personal details until the project terminates in August 2021.

(Signed by project participant, date)

B.3 Study 2: Code

```
# # Add K-means Clustering Layer to Autoencoder

# ## Import data
import os
os.environ['KMP_DUPLICATE_LIB_OK']='True'
from matplotlib import pyplot as plt
import numpy as np


data = np.load("pctm_chroma.npy")
data = data.astype(np.float64)
data.shape

# ## Create and Train (Auto)encoder

from keras.models import Model
from keras.layers import Dense, Input
from keras.optimizers import SGD

# Setup simple model
init = "glorot_uniform"
act = "relu"
dims = [data.shape[1] * data.shape[2], 100, 100, 50, 20]
n_stacks = len(dims) - 1
input_img = Input(shape=(dims[0],), name='input')
# internal layers in encoder
x = input_img
for i in range(n_stacks-1):
    x = Dense(dims[i + 1], activation=act, kernel_initializer=init, name=
f"encoder_{i}")(x)

encoded = Dense(dims[-
1], kernel_initializer=init, name='encoder_%d' % (n_stacks - 1))(x)  # hi
dden layer, features are extracted from here
x = encoded
for i in range(n_stacks-1, 0, -1):
        x = Dense(dims[i], activation=act, kernel_initializer=init, name=
'decoder_%d' % i)(x)

# output
x = Dense(dims[0], kernel_initializer=init, name='decoder_0')(x)
decoded = x
encoder = Model(inputs=input_img, outputs=encoded, name='encoder')
autoencoder = Model(inputs=input_img, outputs=decoded, name='AE')

new_data = np.zeros_like(data)
for i, bilde in enumerate(data):
    new_data[i] += data[i]/np.linalg.norm(data[i])
```

```python
# It is necessary to reshape and scale data
x = new_data.reshape(new_data.shape[0], -1)
# x = np.divide(x, 156)

# Training hyperparameters
optimizer = SGD(lr=0.1, momentum=0.9)
epochs = 1000
batch_size = 10

# Actual training (If necessary)
autoencoder.compile(optimizer=optimizer, loss="mse")
autoencoder.fit(x, x, batch_size=batch_size, epochs=epochs)

# Optional store
autoencoder.save(f"autoencoder_{dims[-1]}_{epochs}")
autoencoder.save_weights(
    f"autoencoder_{dims[-1]}_{epochs}_ckpt"
)
encoder.save(f"encoder_{dims[-1]}_{epochs}")
encoder.save_weights(
    f"encoder_{dims[-1]}_{epochs}_ckpt"
)

# Optional model loading
from keras.models import load_model
autoencoder = load_model(f"autoencoder_{dims[-1]}_{epochs}")
autoencoder.load_weights(f"autoencoder_{dims[-1]}_{epochs}_ckpt")

# ## Elbow Method

from sklearn.cluster import KMeans
from scipy.spatial.distance import cdist

distortions = []
inertias = []
mapping1 = {}
mapping2 = {}
cluster_range = range(2, 20)

for n in cluster_range:
    # Building and fitting the model
    kmeans_model = KMeans(n_clusters=n)
    kmeans_model.fit(x)
```

```
    distortions.append(
        sum(
            np.min(
                cdist(
                    x,
                    kmeans_model.cluster_centers_,
                    'euclidean'), axis=1
            )
        ) / x.shape[0])

    inertias.append(kmeans_model.inertia_)

plt.plot(cluster_range, distortions)
plt.title("Distortion")
plt.show()


plt.plot(cluster_range, inertias)
plt.title("Inertia")
plt.show()

kmeans_model = KMeans(n_clusters=(13))
y_pred = kmeans_model.fit_predict(x)

plt.figure(figsize=(15,5))
plt.plot(y_pred)
plt.title("Predicted cluster")
plt.show()


plt.figure(figsize=(15,5))
plt.plot(encoder.predict(x))
plt.show()

for i, cluster in enumerate(kmeans_model.cluster_centers_):
    plt.figure()
    plt.plot(cluster, label=i)
    plt.title(f"cluster: {i}")
plt.show()
```

Appendix C:

Study 3 Material

C.1 Study 3: Invitation

Mixed-Initiative Composition: Collective Agency in Interactive Music Systems

Ideas for music compositions often emerge in the context of improvisational interactions between musicians. This PhD project examines how creative initiative is negotiated through dynamic interactions between musicians, and applies this knowledge in developing a model for an interactive music system designed specifically for the first ideation stage of a composition project. The type of context that has inspired this project is, for instance, a musician presenting a new idea to fellow musicians at a rehearsal, or the mutual agreement between rehearsing musicians to start jamming and "see what ideas pop out". Hence, this project is thematically placed in the borderline between composition and improvisational performance.

An early prototype of the system has been developed, and I am recruiting participants for a user test. Participants are asked to bring their own musical instrument and present a musical idea (a musical theme/melody/short progression/riff), and thereafter improvise with the responses provided by the system).

Who may participate?
The target group for this workshop are music students on college/university level, or semi-professional/professional musicians who make their own music. The most important qualification is the experience of collectively composing music/making songs with other musicians (e.g. by improvising/jamming with ideas at rehearsals or similar contexts).

I am not recruiting musicians from any specific genre, but participants are requested to accept that the system is under development and has a limited number of interaction modes so far.

What does participation imply?
Joining in the experiment implies bringing your own musical instrument and participating in a session lasting for one hour at the Norwegian Academy of Music. The session includes:

- 5-10 minutes introduction
- 15 minutes interaction with the system
- 30 minutes interview

Both the musical interaction and the interview will be recorded and transcribed. These transcriptions and annotations will be included in the empirical material for the project.

Privacy
The information will be treated confidentially and in accordance with privacy regulations. All data will be anonymized in the final PhD thesis, and no information that may identify individuals will be published. All information published will have a relevance to the project's theme.

Best regards,

Notto J. W. Thelle
PhD fellow, Norwegian Academy of Music

C.2 Study 3: Consent form

DECLARATION OF CONSENT

"Mixed-Initiative Composition: Collective Agency in Interactive Music Systems"

Thank you for participating in the workshop on January 30, 2020. In order to allow data collected from the focus group interview to be treated by the researcher, please sign the declaration of consent below.

Objective
Ideas for music compositions often emerge in the context of improvisational interactions between musicians. This PhD project examines how creative initiative is negotiated through dynamic interactions between musicians, and applies this knowledge in developing a model for an interactive music system designed specifically for the first ideation stage of a composition project. The type of context that has inspired this project is, for instance, a musician presenting a new idea to fellow musicians at a rehearsal, or the mutual agreement between rehearsing musicians to start jamming and "see what ideas pop out". Hence, this project is thematically placed in the borderline between composition and improvisational performance.

An early prototype of the system has been developed, and I am recruiting participants for a user test. Participants are asked to bring their own musical instrument and present a musical idea (a musical theme/melody/short progression/riff), and thereafter improvise with the responses provided by the system).

Project owner
The research project is part of a PhD at the Norwegian Academy of Music, who is responsible for the project.

Implications of participation
Joining in the experiment implies bringing your own musical instrument and participating in a session lasting for one hour at the Norwegian Academy of Music. The session includes:

- 5-10 minutes introduction
- 15 minutes interaction with the system
- 30 minutes interview

Both the musical interaction and the interview will be recorded and transcribed. These transcriptions and annotations will be included in the empirical material for the project.

Participation is voluntary
Consent may be retracted at any point with no reasons given. All details about you will be anonymized. Withdrawal from the workshop or later retraction will not have any negative consequences for you.

Your privacy – how we store and treat the data
We will only use information given by you for the objective stated in this document. Information will be treated confidentially and in accordance with privacy regulations.

- Persons with access to the data at the Norwegian Academy of Music: Notto J. W. Thelle (project leader) and Professor Sidsel Karlsen (supervisor).
- Your name and contact details will be substituted with codes and stored on a list separate from the rest of the data. All data will be stored on the institution's encrypted research server (Box).

All data will be anonymized, and no information that may identify individuals will be published. All published information will be relevant to the project's topic.

What happens to personal details when the research project ends?
According to the plan, the project will terminate in August 2021. All personal information will be anonymized at the end of the project.

Your rights
As long as you can be identified in the data material, you have the right to:
- request access to what personal information is registered about you,
- have personal details about you edited,
- have personal details about you deleted,
- access a copy of your personal details (data portability), and
- file a complaint to a privacy ombudsman or to the The Norwegian Data Protection Authority (Datatilsynet) about the treatment of your personal details.

What gives us the right to treat personal details about you?
We treat details about you on the basis of your consent.

On behalf of the Norwegian Music Academy, NSD – the Norwegian Centre for Research Data has deemed that the treatment of personal data in this project is in accordance with privacy regulations.

Where can I find out more?
If you have any questions regarding the research project, or would like to take advantage of your rights, please contact:
- The Norwegian Academy of Music by Notto Johannes Windju Thelle, nottot@nmh.no, tlf. 988 82 613
- Our privacy ombudsman: personvernombud@nmh.no
- NSD – Norwegian Centre for Research Data, on e-mail (personverntjenester@nsd.no) or telephone: 55 58 21 17.

Best regards,

Project leader

Declaration of consent

I have received and understood the information about the project *Mixed-Initiative Composition: Collective Agency in Interactive Music Systems*, and have been given the opportunity to ask questions. I consent to:

- ☐ engaging in a musical interaction with a prototype for an interactive music system
- ☐ participating in an interview about the experience
- ☐ having the session (audio and video), and the interview (audio only) recorded

I consent to the treatment of my personal details until the project terminates in August 2021.

(Signed by project participant, date)

C.3 Study 3: Interview guide

Interview guide

Study 3 – Interactive session and interview

1. **Part 1- Introduction (10 minutes)**

   - Introduction/short presentation of project
   - Go through the agenda
   - Explain what the participant will do in the experiment
   - Explain what the audio and video recordings and the following interview will be used for, and explain the participant's privacy rights

2. **Part 2 – Explanation of interactive music system (10 minutes)**

   - In the study room there is a computer and a sound card. Participants bring their own instruments (exception for pianists, there is a piano in the room). The purpose of the experiment is for the participant to improvise with what they think is a functioning prototype of the interactive music system that I have introduced.

   Note: In reality, the system is a simulation, and the sounds coming out of the system is played by a keyboard player situated in an adjecent room. This method is called "Wizard of Oz", and is a useful way to get feedback on user experience before a system has been fully developed. The keyboard player can hear what the user is playing, and can follow which buttons the user is pressing on the interface.

   - Explain the functionality of each button on the interface. These explanations are also available as text via Help buttons on the interface.

   - Explain that the participant should experiment freely around a musical idea, and try to use the system to develop the idea through improvised interaction.

3. **Part 3 – Improvised session (15 minutes)**

   - Leave the room and let the participant improvise
   - Go to the other room where the keyboard player is, and stay on alert in case situation arise where he needs advice on what kind of musical response to give to the participant

4. **Part 4 – Interview part 1 (10 minutes)**

Questions
- What was it like to play with the system?
- How did playing with an interactive music system differ from creative co-performance situations you are used to?
- How did it affect your creativity?
- Did you feel that there was some kind of negotiation about the initiative between yourself and the system?
- Did you feel that the system pulled you in a direction that you hadn't expected?
- Could you describe some positive and negative aspects about this tug-of-war, or initiative taking?
- Did the musical dialog lead to any new ideas? Examples?

5. **Part 5 – Revelation and final comments (7 minutes)**

Reveal that this had been a simulation, and explain the purpose of this set-up in the experiment. Give the participant time to regain composure after the revelation.

Question
- Did you at any point suspect that you were not playing with a machine?
- How would you have played differently if you knew it was a human?
- Are you relieved or disappointed that it was not a machine?

Keep this part of the interview open-ended, and give the participant space to reflect upon the experience.

C.4 Study 3: Themes and codes

Study 3 – Themes and codes

Theme 1: Interface interaction

Codes

- Didn’t use the interface much
    - Used to extended transport stages
- Wanted to play rather than use the interface

Theme 2: The impact of believing it was a machine

Codes

- Effects/impact on user
    - Triggers a response from me
    - Lost track of time
    - Became more focused on my own choices
    - Became more focused on interactivity
    - Could predict the computer’s choices
- Affordances
    - Good practice being with someone
    - Good for testing ideas
    - Led to new ideas
    - More explorative/taken to places
    - Improvisational more than compositional
    - Most fun with the kind of stuff I never do
    - Surprised by how the machine picked up and reused details
- Context dependency/assumptions
    - Played differently believing it was a machine
    - Would have elaborated more (“stayed in place”) with human
    - Different context - different attitude
    - Different from playing with a human
    - Safer than playing with human - less pressure
    - Less searching with humans
    - Curious about the machine's response
- Bias/assumptions
    - Didn't respond to tonal or harmonic stuff
    - Didn’t mind apparent “glitch”
    - Can tell it's not human
- Post-revelation reflections
    - Would have dared less if I knew it was human
    - Would be more prejudiced if I knew it was a human
    - Fooled
    - Didn't suspect it was a human
    - Wish it was real
    - Dad will be disappointed that it wasn't a machine

    - Would like to hear the recording
    - Like playing with a human
- Statements about technology
    - Music automation is coming
    - I have belief in computers

Theme 3: Responsiveness vs. contrast

Codes

- Preferred contrasting responses
- Responsive - picks up from me
- Imitated me too much instead of complementing
- Easy to trick the machine - followed too easily
- More interactive than expected
- Responsive - picks up from me

Theme 4: Relating to an unknown other

Codes

- Power balance
    - Machine was assertive at times
    - I took the lead
    - I followed the machine's lead
    - Negotiation took place
- Tried to understand the machine
- No common reference
- The computer – “he”
- The computer – “who”
- Leaving comfort zone

Miscellaneous

Codes

- Immediate reactions – pros
    - Organic
    - Instructive
    - Fun
    - Exciting
    - Surprising
    - Surprisingly absorbing
    - Responsive
- Immediate reactions – cons
    - Some sounds were better
    - Bad sounds
- Black box

C.5 Study 3: Interaction logs, behavior codes, annotations

Time	**Action**	**Instrument**	**Behavior ***	**Details**
00:00:00	Start	Piano		
00:01:32	ThumbsUp	Piano	Coupling	User playing haphazardly, no real theme, wizard starts hammering atonally in bass register
00:05:09	ThumbsUp	Vibes	Negotiation	User seemingly just playing clusters to provoke a reaction, wizard starts varying widely between very high and low registers, quite atonally
00:07:52	ThumbsUp	Synth pad	Coupling	Wizard combines very high and very low sustained notes while user plays up and down a semitone in mid-register
00:09:04	ThumbsUp	Synth pad	Negotiation	User has had the semitone theme going for a long time, supplemented with bass notes, wizard starts playing cluster chord in very high register combined with lower notes
00:11:55	ThumbsUp	Synth pad	Coupling	Wizard has simulated a "glitch" for a long time, holding single note while user rambles on in his "own world". Once the wizard switches to lower register, the user approves.
00:15:05	End	Synth pad		

* As interpreted by myself, a very rough categorization

Time	Action	Instrument	Behavior *	Description
00:00:00	Start	Piano		
00:00:26	ThumbsUp	Piano	Mirroring	Wizard copies user's gesture
00:00:46	ThumbsUp	Piano	Mirroring	Wizard copies user's gesture
00:01:21	ThumbsUp	Piano	Coupling	Wizard diverges, contrasts with sustained upward motion
00:02:07	ThumbsUp	Piano	Negotiation	Wizard picks up on theme introduced, keeps it going as the user diverges into staccato repetetive rhythm
00:02:26	ThumbsUp	Piano	Negotiation	Wizard slows down from monotone repetetive rhythm adopted from user, lower energy
00:02:50	ThumbsUp	Piano	Coupling	Wizard falls into a new, slower monotone rhythm while user starts playing melody with sustained notes
00:03:12	ThumbsUp	Piano	None, silence	Wizard stops playing, opening up space
00:03:49	ThumbsUp	Piano	Coupling	After a long break, wizard starts playing somethin complementary/contrasting to the user
00:04:20	ThumbsUp	Piano	Mirroring	After a hint of shadowing note by note (not same tones), wizard quadruples speed and breaks into an a kind of accompaniment
00:05:50	ThumbsUp	Piano	Negotiation	The bass goes a bit crazy and staccato, and after a short break the wizard break into a complementary and contrasting groove as a response
00:06:23	ThumbsUp	Piano	Negotiation	The craziness reaches a climax, the wizard plays in the upper and lower registers simultaneously
00:06:50	ThumbsUp	Piano	Coupling	The user stops playing, listens to the wizard winding down from the crazy climax, coming to a stop
00:09:22	ThumbsUp	Piano	Coupling	Wizard is playing something independently of the user, with a similar level of energy
00:10:58	ThumbsUp	Piano	None, silence	After having hammered semi-rhythmically on the keyboard for 10-15 seconds, the wizard stops playing for a while
00:11:36	ThumbsUp	Synth pad	Coupling	The user has changed to synth pad, and the wizard stays silent for half a minute before laying some sustained cluster chords
00:12:48	ThumbsUp	Synth pad	Coupling	The user stops and seems to realize that the synth pad is "rumbling" in the bass register (may not have heard this before)
00:13:31	ThumbsUp	Synth pad	Coupling	The user starts playing in the higher registers as a complement to the wizard's bass rumbling, enjoys it
00:14:13	ThumbsUp	Synth pad	Coupling	Wizard holds two sustained notes, a minor third several octaves apart
00:15:30	End	Synth pad		

* As interpreted by myself, a very rough categorization

Time	Action	Instrument	Behavior *	Descritpion
00:00:00	Start	Piano		
00:01:03	ThumbsUp	Piano	Coupling	After some mirroring behavior, the wizard adds some very low bass tones and begins to diverge
00:01:15	ThumbsUp	Piano	Shadowing	The wizard imitates the user's monotone repetitive gesture, a minor ninth interval apart
00:02:50	GoBack	Piano	Mirroring to coupling	User is playing some light arpeggios in the higher registers, wizard compliments with some rhythmic overlays in the same register, then adds a deep bass tone, user want to og back
00:02:57	ThumbsUp	Piano	Coupling to mirroring	Wizard reverts to the high-registered overlays, user approves of the wizard's reversal
00:03:52	Change	Piano	Coupling	The user stops playing, wizard continues incessantly in the higher register. According to the interview, the sounds were loud and painful to the ear.
00:04:10	ThumbsUp	Piano	Mirroring	The user start playing some complex chords in the mid-register, the wizard plays for a while and responds in a slightly lower register
00:07:14	GoBack	Bass	Negotiation	Wizard started playing long deep sustained notes, then broke into fast staccato, user wants to og back
00:07:18	ThumbsUp	Bass	Negotiation	Wizard goes back to the deep sustained bass notes, user approves
00:08:05	ThumbsUp	Bass	Coupling	Wizard playing some quite independent bass lines along with user's improvised piano melodies
00:08:45	ThumbsUp	Bass	Mirroring	Wizard switches to staccato bass, seemingly responding to the user's gestures
00:09:20	ThumbsUp	Bass	Coupling	Wizard plays a repetetive staccato downward theme in parallel with user's meandering improvisation
00:10:39	ThumbsUp	Bass	Negotiation	User plays incesantly staccato in the higher register, wizard gradually joins with similar instensity in the higher (bass) register, parallel improvisation
00:10:56	ThumbsUp	Bass	Negotiation	User goes down to the lower register, wizard stays in the same higher register
00:11:24	ThumbsUp	Bass	Negotiation	Wizard suddently shifts to a very deep register, same instensity (negotiation)
00:11:35	ThumbsUp	Bass	Negotiation	Wizard doubles, a return of the high register at the same time as the low register instense staccato
00:11:56	ThumbsUp	Bass	Negotiation	Wizard now with the continuing staccato in bass with "stabs" in the treble, they are both seemingly in a flow state, perhaps wizard has forgotten the machine-role
00:12:16	GoBack	Bass	Negotiation	Wizard winds down to a monotone repetetive note in the bass, then start slowing down, apparently the user wants back
00:12:21	GoBack	Bass	Negotiation	Wizard goes back to the monotone repetetive note in the bass, apparently the user wants to og further back, wizard goes to the higher register staccato
00:12:43	GoBack	Bass	Negotiation	Wizard goes up some octaves to a very light register and varies with some lower register, apparently the user wants the wizard to og back
00:14:51	GoBack	Bass	Coupling	After having been playing intermittent chords in a relatively high register, the wizard switches to a deep register with longer notes, apparently the user wants back
00:15:14	ThumbsUp	Bass	Coupling	The wizard goes back to the chords, but at some point starts leaving them sustained instead of staccato
00:15:23	End	Bass		

* As interpreted by myself, a very rough categorization

C.6 Study 3: Wizard instructions

Your task is to be a "believable machine".

You need to use a combination of your musical talents and social capacities to adapt to the participating musician’s input in a way that will not disclose that you are a human.

In order to build up this illusion, the participant will have a limited number of quite standard MIDI instruments to choose from, with a limited dynamical range.

At the same time, you need to keep in mind that the participant believes that he or she is interacting with a prototype of an interactive music system with promising tendencies.

If the participant is left with the feeling that there were a few segments of the interaction that «gave something back”, that is a good result.

THINGS TO REMEMBER FOR THE INTERACTION

- Let the participant start with a theme/riff. Wait at least 4 or 5 seconds before responding to a first initiative.

- Be insistent when first starting your response. This is the machine’s current “state”.

- If the input has a tempo, your response tempo should behave «overrulingly» when you have perceived/decided the tempo.

- If it seems like the participant is trying to change the tempo, harmony or other parameters in another direction, adapt to this in discrete steps or «jerks», not gradually.

- It is better to stop entirely and wait for a new entry point of you are not sure about what to do at any given moment.

- Be attentive to musical genre, but try to appear as genre agnostic or generic.

- Do not think like a pianist. Your responses may be monophonic. This depends on the type of sound the participant chooses as preset.

- The envisioned prototype should skip stochastically between various response types. For example, it can jump from providing chords/harmonies to responding melodically.

- The transition between response types could happen abruptly.

- You do not need to be «good». But you need to be creative, and stubbornly stick to what you are doing at any given moment.

- You have no memory of earlier interactions. Everything that happens in the co-performance happens as if for the first time, even when it is a repetition.

Appendix D:

Study 4 Material

D.1 Study 4: Invitation

Would you like to jam with a computer?

This is an invitation to participate in the research project «Mixed-Initiative Composition: Collective Agency in Interactive Music Systems». If you play an instrument (or sing) and you enjoy experimenting with new ideas in collaboration with others, this is your chance to try a new system where “the other” is a computer system.

The study
Ideas for music compositions often emerge in explorative interactions between musicians. Such contexts are prone to yielding surprises, and ideas one normally would not have thought of in isolation can pop up while co-performing. In this PhD project, I have researched how creative initiative is negotiated through such interactions, and during the course of the project, I have developed an interactive music system that can co-perform with musicians. The system responds to what it “hears” in the form of accompaniment or continuation of musical phrases.

I am now arranging a final user study where this system will be evaluated by musicians. The duration of the study is approximately 90 minutes.

All participants who complete the user study will receive a **NOK 300 gift card.**

The research project is part of a PhD at the Norwegian Music Academy.

Target group
The target group for this study is musicians who use collaborative experimentation as a method to develop musical ideas (improvising/jamming). Participants from any musical genre are welcome. However, due to the way in which the system “listens” it is crucial that the instrument is tonal. This means that percussionists/drummers are not in the target group for this study. It is also an advantage if the instrument can play outside of the bass register, because the machine listening is better in the mid and higher registers.

If you are interested, but unsure whether your instrument is suitable, just sign up anyway. The study needs as many participants as possible!

What does participation imply?
Participation implies setting aside 90 minutes for attending the study at the Norwegian Academy of Music. During the study, you will get the opportunity to test two different versions of the interactive music system (10 minutes for each), followed by short interviews. You will also be asked to fill out a survey about the level of creative engagement you experienced while playing with the system.

The music sessions and interviews will be recorded. The performance will not be evaluated qualitatively by myself, but you will be asked to listen through the recording as a home assignment, and evaluate how creative you think the different parts of the interactive sessions were using a web-based app.

Privacy
The information will be treated confidentially and in accordance with privacy regulations. All data will be anonymized, and no information that may identify individuals will be published. All information published will have a relevance to the PhD project’s theme.

Best regards,

Notto J. W. Thelle
notto.w.thelle@nmh.no

D.2 Study 4: Consent form

DECLARATION OF CONSENT

"Mixed-Initiative Composition: Collective Agency in Interactive Music Systems"

Thank you for participating in the workshop on September 17, 2021. In order to allow data collected from the interviews to be treated by the researcher, please sign the declaration of consent below.

Objective
Ideas for music compositions often emerge in the context of improvisational interactions between musicians. This PhD project examines how creative initiative is negotiated through dynamic interactions between musicians, and applies this knowledge in developing a model for an interactive music system designed specifically for the first ideation stage of a composition project. The type of context that has inspired this project is, for instance, a musician presenting a new idea to fellow musicians at a rehearsal, or the mutual agreement between rehearsing musicians to start jamming and "see what ideas pop out". Hence, this project is thematically placed in the borderline between composition and improvisational performance.

Project owner
The research project is part of a PhD at the Norwegian Academy of Music, who is responsible for the project.

Participation is voluntary
Consent may be retracted at any point with no reasons given. All details about you will be anonymized. Withdrawal from the workshop or later retraction will not have any negative consequences for you.

Your privacy – how we store and treat the data
We will only use information given by you for the objective stated in this document. Information will be treated confidentially and in accordance with privacy regulations.

What happens to personal details when the research project ends?
According to the plan, the project will terminate in December 2021. All personal information will be anonymized at the end of the project.

Your rights
As long as you can be identified in the data material, you have the right to:

- request access to what personal information is registered about you,
- have personal details about you edited,
- have personal details about you deleted,
- access a copy of your personal details (data portability), and
- file a complaint to a privacy ombudsman or to the The Norwegian Data Protection Authority (Datatilsynet) about the treatment of your personal details.

What gives us the right to treat personal details about you?
We treat details about you on the basis of your consent.

On behalf of the Norwegian Music Academy, NSD – the Norwegian Centre for Research Data has deemed that the treatment of personal data in this project is in accordance with privacy regulations.

Where can I find out more?
If you have any questions regarding the research project, or would like to take advantage of your rights, please contact:

- The Norwegian Academy of Music by Notto Johannes Windju Thelle, nottot@nmh.no, phone no. 988 82 613
- Our privacy ombudsman: personvernombud@nmh.no
- NSD – Norwegian Centre for Research Data, on e-mail (personverntjenester@nsd.no) or telephone: 55 58 21 17.

Best regards,

Project leader

Declaration of consent

I have received and understood the information about the project *Mixed-Initiative Composition: Collective Agency in Interactive Music Systems*, and have been given the opportunity to ask questions.
I consent to:

- ☐ engaging in a musical interaction with a prototype for an interactive music system
- ☐ participating in an interview about the experience
- ☐ having the sessions (audio and video), and the interview (audio only) recorded

I consent to the treatment of my personal details until the project terminates in December 2021.

____
(Signed by project participant, date)

D.3 Study 4: Detailed study guide

WORKSHOP 4
SPIRE MUSE USER STUDY

1. Introduction (5 minutes)

An introduction about the study process will be given in the form of an information sheet.

Exact formulation of text sheet (the researcher will read the following out load or paraphrase its content in Norwegian, line by line). The participant may read along if they wish.

Purpose of the study.
- The purpose of this study is to understand how different interfaces will affect the user's creative engagement with the musical agent.

You will be testing two different prototypes. In one of them, the interactive modes can be selected manually—I call this the "Manual Modes" prototype. The other prototype switches modes autonomously based on what it "hears" in the input, but it can be indirectly influenced to change its behavior—I call this the "Auto Modes" prototype.

Study process. The study includes three main parts. It will take approximately 90 minutes, during which time you are free to opt out at any point.

Part one:
- You will be introduced to the first prototype and engage in a learning session, followed by a longer jam session.
- At the end of the learning session, there will be a short conversation about the prototype's functionalities. The purpose of this conversation is to ensure that you have understood the functionalities, and that there are no misunderstandings before the main creative session.
- The purpose of the main creative session is to start playing an original musical idea (a phrase/theme/riff) and improvise with the musical agent. Keep an open mind and see how the idea develops as a co-creative experiment.
- After the main session, you will be asked to fill out a questionnaire, followed by a short interview.

Part two:
- The same process as Part 1, only with the second prototype.

Part three:
- A final questionnaire comparing the two prototypes
- Final short interview, reflections on the differences between the prototypes

You will also be given a home assignment in a few days, which is to subjectively rate the quality of the two main sessions. This will be further explained at the end.

Consent form. (video and audio recording, interaction log, anonymous)

End of information sheet

2. Prototype "Auto Modes" or "Manual Modes" (35 minutes)

The order of the prototypes (named "Auto Modes" and "Manual Modes" respectively) will be randomized for different participants so as to minimize the influence of the order. Video and audio recordings will be made, and participant's interactions with the software will be logged.

Prototype introduction (5 minutes)
The following information will be provided to each participant

1. <u>The basic concept.</u> Both prototypes have the same objective: To be a virtual musical partner for creative brainstorming.

   - *Corpus-based re-synthesis*
   The sound material is a large library of acoustic guitar performances. The audio has been sliced up in very short fragments, and the musical agent uses different methods to put these fragments together in new ways, depending on the musical context and the user's input. So, what you are hearing are real instruments, but the performances are remodeled.

   - *Interactive modes*
   The musical agent has three interactive modes: *shadowing, mirroring* and *coupling.* In shadowing mode, the agent tries to imitate what the user is doing musically. In mirroring mode, the agent listens to longer phrases and reflects its own "interpretations" of these phrases back to the user. In coupling mode, the agent is more independent of the user's input. Instead, it models its output on one of the songs in the corpus.

   - *Influence presets*
   The user can influence which types of musical parameters the agent should listen to: *harmonic, melodic, spectral* or *rhythmic*. If, for instance, the harmonic preset is selected, the agent's responses will be more adapted to the user's harmonic content than other factors.

2. <u>The design of the interfaces.</u> Auto Modes is designed to behave more autonomously than Manual Modes.

   - *Explanation for prototype Manual Modes*
   Based on the user's input, the musical agent chooses its interactive modes autonomously. The user can either go along with what is happening or try to change its output through a negotiating panel with four main functions:
   Go back will force the agent to go back to its previous state.
   Pause/Continue can be used if the user needs a break from the output. The agent will still listen, but be silent.
   Change will force the system to switch to another mode, randomly selected.
   Thumbs up alerts the agent that it is doing something that the user finds particularly engaging. This will be bookmarked for future reference.
   Four footswitches are mapped to these four functions, in the same order and layout as in the interface. This frees the user to negotiate with the system while playing their instrument.

The influence presets can be selected with the computer mouse.

- Explanation for prototype Manual Modes
The user chooses the agent's interactive modes directly.
Shadowing: The agent tries to imitate what the user is doing musically.
Mirroring: Phrases are reflected back in novel ways.
Coupling: The agent does its own thing, modelled on a song in the corpus.
In coupling mode, the user needs to select which song the agent should use as its generative model from a dropdown menu, using the computer mouse.
Thumbs up alerts the agent that it is doing something that the user finds particularly engaging. This will be bookmarked for future reference.
Four footswitches are mapped to the three interactive modes and the Thumbs up button, in the same order and layout as the in the interface. This frees the user to control the system's modes while playing their instrument.
The influence presets can be selected with the computer mouse.
The agent can be muted (show toggle button) if you wish to pause its input momentarily.

Learning session (5 minutes)

Instructions for the participant:
Play with the system and try out its different functions. Try to gain an understanding of what they do. At the end of the five minutes, I will ask you to formulate in your own words what the different functions do before moving on to the creative session.

Note to the researcher:
Sit quietly in the corner. If the participant has any questions, politely let them know that questions will be answered after the learning session. The point of this brief session is for the participant to discover affordances alone, without being influenced.

Short conversation (5 minutes). This is to ensure that the participant has understood the concept and design of the interface, and that they are made aware of the purpose of the creative session ahead.

Go through each function and let the participant explain what they think their function is. Clear up any misconceptions.

Creative session explanation:
We are ready for the main creative session.
Think of it as a jam session. You and a fellow musician are improvising loosely around an idea. Start with a short musical phrase. It could be an idea you have for a song. A motive, theme or riff. Gauge the musical agent's response and let it develop from there. You do not need to stick to the theme if you find your interest wandering to something more interesting. However, keep in mind that you do not have to show any "musical result" at the end of the session. We are interested in your experience of the creative session.

After starting the session, I will leave the room for 10 minutes. I will not be listening in while you are playing.
Good luck!

Creative session (10 minutes)

Very important! Make sure video and audio are recording, and double check that the correct Max session is loaded. Start session and exit room.

Questionnaire (3 minutes)

Enter the room and end the Max session. Administer the questionnaire to the participant (while they are filling out the questionnaire, double-check that the Max interaction log has been written to file).

<u>Questionnaire (Creativity Support Index, part 1)</u>

The questionnaire will be administered on an iPad. The following 10 statements will be presented over two pages. The order of the statements will be randomized on a per-page basis. The sliders are floating point scores between 0.0 and 10.0.

Note to reader: The format of the questionnaire as presented in this document is not representative of what it will look like in the end. The intended layout of the Creativity Support Index will be used.

Page 1

1. I would be happy to use this system on a regular basis.
   Highly disagree **–O---------------------- –** Highly agree

2. It was easy for me to explore many different ideas, options or outcomes using this system.
   Highly disagree **–O---------------------- –** Highly agree

3. I was satisfied with what I got out of the system.
   Highly disagree **–O---------------------- –** Highly agree

4. I was able to be very creative while playing with this system.
   Highly disagree **–O---------------------- –** Highly agree

5. My attention was fully tuned to the activity, and I forgot about the system I was using.
   Highly disagree **–O---------------------- –** Highly agree

Page 2

6. I enjoyed using this system.
   Highly disagree **–O---------------------- –** Highly agree

7. The system was helpful in allowing me to track different ideas, outcomes or possibilities.
   Highly disagree **–O---------------------- –** Highly agree

8. What I was able to produce was worth the effort I had to exert to produce it.
   Highly disagree **–O---------------------- –** Highly agree

9. The system allowed me to be very expressive.
   Highly disagree **–O---------------------- –** Highly agree

10. I became so absorbed in the activity that I forgot about the system that I was using.
   Highly disagree **–O---------------------- –** Highly agree

Semi-structured interview (7 minutes)

Overall experience

- Can you describe what it was like to play with the system using three adjectives?
- What strategies did you employ to engage with the system creatively?
- Did these strategies lead to any interesting results? Examples?
- How did the context affect the way you play your instrument?
- Did you feel in charge most of the time, did you feel like you were mostly following the system's lead, or was it a combination of both?
- How did you feel about that balance, was it useful or did you want something else?
- What did you miss about the interaction

3. Prototype “Manual Modes” or “Auto Modes” (35 minutes)

Same procedure as 2, but with the other interface. Introducing the prototype may take a little less time than in 2, because some of the common concepts have already been introduced.

4. Comparisons (15 minutes)

As with the previous questionnaires, this will also be administered on an iPad.

Questionnaire (5 minutes)

The first part of the questionnaire is a direct comparison between the two prototypes. The second part is final part of the Creativity Support Index questionnaire (weighting the dimensions for each participant).

Please note: The format of the questionnaire as presented in this document is not representative of what it will look like on the iPad. The Creativity Support Index layout will be used.

1. Please choose which interface you feel the following statements are most appropriate to:

	Auto Modes	Manual Modes
I enjoyed myself most		
I explored more ways of playing		
I felt I was more expressive		
I became more absorbed with the activity		
I felt more creative		
I felt more satisfied with the result		

2. If you could play with these or similar systems in the future, it would be more important to be able to…

Be creative and expressive	
Produce results that are worth the effort I put in	

3. If you could play with these or similar systems in the future, it would be more important to be able to…

Become immersed in the activity	
Enjoy using the system	

4. If you could play with these or similar systems in the future, it would be more important to be able to…

Produce results that are worth the effort I put in	
Explore many different ideas, outcomes or possibilities	

5. If you could play with these or similar systems in the future, it would be more important to be able to…

Enjoy using the system	
Be creative and expressive	

6. If you could play with these or similar systems in the future, it would be more important to be able to…

Produce results that are worth the effort I put in	
Enjoy using the system	

7. If you could play with these or similar systems in the future, it would be more important to be able to…

Become immersed in the activity	
Be creative and expressive	

8. If you could play with these or similar systems in the future, it would be more important to be able to…

Enjoy using the system	
Explore many different ideas, outcomes or possibilities	

9. If you could play with these or similar systems in the future, it would be more important to be able to…

Explore many different ideas, outcomes or possibilities	
Become immersed in the activity	

10. If you could play with these or similar systems in the future, it would be more important to be able to…

Become immersed in the activity	
Produce results that are worth the effort I put in	

11. If you could play with these or similar systems in the future, it would be more important to be able to…

Be creative and expressive	
Explore many different ideas, outcomes or possibilities	

Interview (5 minutes)

- Which prototype do you prefer to play with? Why?
- Which prototype feels most like a musical partner to you?
- Could you share any other reflections about the prototypes you tried today which may not have come to light in the questionnaires or in the previous interview formats?

Home assignment explanation (5 minutes)

In a few days, I will send you a link to an online listening assignment. The page you will be directed to will feature the recording of each of the two creative sessions you played today.

I hope you can take the time to listen back to the two sessions and simultaneously rate the perceived quality of the interaction using a screen-based slider. This will produce a graph showing a timeline of how well you think the interaction works for the duration of the sessions. This final subjective evaluation will be an important part of the results. It will take you approximately 25 minutes to complete this home assignment.

The email will contain everything you need to know about how to complete the assignment.

Thank you so much for your time today!

D.4 Study 4: Information sheet

SPIRE MUSE USER STUDY

Purpose of the study.
- The purpose of this study is to understand how different interfaces will affect the user's creative engagement with the musical agent.

You will be testing two different prototypes. In one of them, the interactive modes can be selected manually—I call this the "Manual Modes" prototype. The other prototype switches modes autonomously based on what it "hears" in the input, but it can be indirectly influenced to change its behavior—I call this the "Auto Modes" prototype.

Study process. The study includes three main parts. It will take approximately 90 minutes, during which time you are free to opt out at any point.

Part one:
- You will be introduced to the first prototype and engage in a learning session, followed by a longer jam session.
- At the end of the learning session, there will be a short conversation about the prototype's functionalities. The purpose of this conversation is to ensure that you have understood the functionalities, and that there are no misunderstandings before the main creative session.
- The purpose of the main creative session is to start playing an original musical idea (a phrase/theme/riff) and improvise with the musical agent. Keep an open mind and see how the idea develops as a co-creative experiment.
- After the main session, you will be asked to fill out a questionnaire, followed by a short interview.

Part two:
- The same process as Part 1, only with the second prototype.

Part three:
- A final questionnaire comparing the two prototypes
- Final short interview, reflections on the differences between the prototypes

You will also be given a home assignment in a few days, which is to subjectively rate the quality of the two main sessions. This will be further explained at the end.

Consent form. (video and audio recording, interaction log, anonymous)

D.5 Study 4: CSI individual scores

CSI Results - Auto Modes

AGREEMENT STATEMENTS

Participant	ResultsWorthEffort1	Exploration1	Enjoyment1	Expressiveness1	Immersion1	Enjoyment2	Exploration2	ResultsWorthEffort2	Expressiveness2	Immersion2
P1	5	6	7	7	5	6	5	7	5	5
P2	2	1	1	1	1	7	2	7	3	2
P3	2	3	3	3	8	3	3	3	3	3
P4	6	7	5	7	6	7	5	8	7	6
P5	6	5	5	6	4	8	5	7	5	5
P6	2	2	2	3	1	4	3	1	1	1
P7	3	5	1	6	1	2	3	3	3	1
P8	6	6	6	5	4	6	3	5	5	5

CSI Results - Manual Modes

AGREEMENT STATEMENTS

Participant	ResultsWorthEffort1	Exploration1	Enjoyment1	Expressiveness1	Immersion1	Enjoyment2	Exploration2	ResultsWorthEffort2	Expressiveness2	Immersion2
P1	5	8	7	6	4	7	8	4	6	5
P2	5	6	7	4	3	8	4	10	3	3
P3	4	8	3	8	8	4	6	4	8	8
P4	7	7	6	5	7	8	5	8	7	7
P5	5	6	7	8	7	8	5	8	5	7
P6	5	6	6	2	5	6	4	4	3	4
P7	5	6	2	4	2	5	4	5	4	2
P8	2	2	4	1	1	3	2	2	1	2

PAIRED FACTOR COUNTS

Participant	Creativity count	Immersion count	Exploration count	Enjoyment count	RWE count
P1	2	3	3	1	1
P2	4	2	3	1	0
P3	3	2	3	1	1
P4	2	4	2	2	0
P5	3	1	3	0	3
P6	3	3	1	3	0
P7	3	1	0	3	3
P8	3	4	2	1	0

Appendix E:

NSD Approval

NSD NORSK SENTER FOR FORSKNINGSDATA

Vurdering

Referansenummer

455790

Prosjekttittel

Mixed Initiative Composition: Collective Agency in Interactive Music Systems

Behandlingsansvarlig institusjon

Norges musikkhøgskole / NordART - Arne Nordheim-senteret

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Notto Johannes Windju Thelle, nottot@nmh.no, tlf: 98882613

Type prosjekt

Forskerprosjekt

Prosjektperiode

01.08.2019 - 30.11.2021

Vurdering (3)

01.10.2021 - Vurdert

NSD har vurdert endringen registrert 31.08.21.

Vi har nå registrert 30.11.2021 som ny sluttdato for behandling av personopplysninger. Vi gjør oppmerksom på at ytterligere forlengelse ikke kan påregnes uten at utvalget informeres om forlengelsen.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Kontaktperson hos NSD: Karin Lillevold
Lykke til videre med prosjektet!

10.06.2021 - Vurdert

NSD har vurdert endringen registrert 19.05.21.
Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 10.06.21. Behandlingen kan fortsette.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Kontaktperson hos NSD: Karin Lillevold
Lykke til videre med prosjektet!

03.06.2019 - Vurdert

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 03.06.2019. Behandlingen kan starte.

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde:

https://nsd.no/personvernombud/meld_prosjekt/meld_endringer.html

Du må vente på svar fra NSD før endringen gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 31.08.2021.

LOVLIG GRUNNLAG
Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER
NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

Box er databehandler i prosjektet. NSD legger til grunn at behandlingen oppfyller kravene til bruk av databehandler, jf. art 28 og 29.

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp underveis (hvert annet år) og ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet/pågår i tråd med den behandlingen som er dokumentert.

Lykke til med prosjektet!

Kontaktperson hos NSD: Karin Lillevold
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Appendix F:

Translation example

Translation example

Translation

N: Did you feel that the system pulled you in a direction that you hadn't expected?

A: Absolutely.

N: Yeah. Could you describe some positive and negative aspects about this tug-of-war, or initiative taking? Um… first the positive.

A: Well, the positive is that it's very organic, really. Or it reminds me of something that is very organic.

N: Any downsides to it? I mean were there any situations where you felt that your initiative was not heard, in a way?

A: Sometimes. But I think what I personally disliked the most was… I don't know if this answers your question, but often I could tell that it was very quick to imitate what I was doing, and played close to it. Instead of playing against it…

N: Yeah. Okay.

A: ... and making a contrast like that. But there were a few places where it also pulled in a different direction. And I think that… those sections were really the most exciting.

Norwegian (original)

N: Følte du at systemet dro deg i en retning som du ikke hadde forutsett?

A: Absolutt.

N: Ja. Kunne du beskrive noen positive og negative sider ved denne dramkampen, eller initiativtakingen? Eh, først positive.

A: Nei, det positive er jo at det er veldig organisk, egentlig. Eller det minner meg på noe som er veldig organisk.

N: Noen negative sider ved det? Altså, var det tilfeller hvor du følte at ditt initiativ ikke ble hørt, på en måte?

A: Til tider. Men jeg tror det jeg kanskje personlig reagerte mest på var... jeg vet ikke om dette svarer på spørsmålet ditt... men ofte skjønte jeg når jeg spilte at den var veldig kjapp på å etterligne det jeg gjorde, og spille opp mot det. I steden for å spille fra det...

N: Ja, nettopp.

A: ... og lage en kontrast sånn. Men det var et par punkter hvor den også dro i en annen retning. Og det synes jeg at... de partiene synes nesten var mest spennende.

How might artificial intelligence change the way humans conceive musical ideas? Notto J. W. Thelle's thesis focuses on jamming-oriented approaches to music making and examines how setting up an environment where humans and computers share the creative initiative might affect the music-making experience. Starting from a theoretical vantage point of dynamical systems, the gradual development of such an environment is realized through four studies using a Research through Design approach within a methodological framework of triangulation between theory, observation, and design.

Whereas the aim of the first and second studies is to understand the creative dynamics between human musicians, the third study employs a Wizard of Oz methodology in which a keyboard player behind the scenes simulates a computer system playing along with human musicians. Based on findings from these first three studies, *a mixed-initiative interactive* music system developed by Thelle forms an integral part of the final study.

Combined, the studies show that a mixed-initiative interactive music system offers musicians freedom from judgement and freedom to explore their own creativity in relation to an unknown agency. Social factors make these kinds of freedom difficult to attain with other musicians. Hence, playing with interactive music systems can lead to different kinds of musical interaction than can be achieved between people. An acceptance of machine aesthetics may lead to surprising creative results. Repeated exposure to mixed-initiative interactive music systems could help cultivate attitudes that are valuable for collective music making in general, such as maintaining a process-oriented approach and accepting the loss of idea ownership.

Notto J.W. Thelle (b. 1974) is a musician, musicologist, and music technologist. Since 2022, he heads the Makerspace at Oslo Metropolitan University.

Find other publications and download them
https://nmh.no/en/research/publications

Norwegian Academy of Music
NMH Publications 2022:4

9 788278 533048